Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics, Istanbul, Turkey, May 27-29, 2006 (pp463-468)

Intelligent "Health Restoration System ": Reinforcement Learning Feedback to Diagnosis and Treatment Planning

O. D. KARADUMAN [1], A. M. ERKMEN [2], N. BAYKAL[3]
Informatics Institute [1], Electrical and Electronics Department [2]
Middle East Technical University
İnönü Bulvarı, 06531, Ankara
TURKEY

Abstract: - In this study we develop a decision support architecture that evaluates pathology findings for defining levels of chronic hepatitis B, and models patients' clinical stages for assisting treatment decisions. It is a learning system that generates a feedback to pathological diagnostic as well as to the clinical decision making by using reinforcement learning techniques. The system receives reinforcement from the patient as a consequence of undertaken actions during a treatment plan. This received information leads system to learn from experiences such as the patient's response to the treatment and evaluations of related parties (pathologist and clinicians).

Key-Words: - Treatment Planning, Diagnostic Decision, Reinforcement Learning, Markov Decision Process

1 Introduction

Chronic liver disease including hepatitis B is quite common in the world. Hepatitis is a liver inflammation that may causes damage to hepatocytes. The severity may range from healthy carrier to decompensated cirrhosis. There are several studies on the diagnosis of hepatitis [1,2,3,4,5,6]. These studies are mainly focused on the diagnosis and the prognosis of the disease. However, defining the severity level of hepatitis and evaluating alternative treatments are important components of medical care processes that do not exist, yet, in the literature but still need to be considered.

The aim of our study is to imbed such component in an intelligent decision making architecture. Consequently we develop a hybrid methodology for supporting diagnostic decision and treatment planning processes in chronic hepatitis B, and provide a novel structure for evaluation of the patient response to alternative treatments by staging disease severity. This architecture has gained more robustness in its performance by a reinforcement learning two level feedback system that helps pathologist (second level) diagnostic tuning and clinicians (first level) clinical treatment tuning according to the patients need by monitoring and evaluating their response to treatment. In the clinical level feedback, results of treatment are evaluated and a reinforcement signal, indicating the change of the patient's health level, is produced. Based on this signal, a feedback is send both to the treatment planning module in the first level to modify treatment administration and clinicians opinions about the patient. At the second level, a reinforcement feedback is also provided to the pathologist to modify their diagnostics on such disease cases if their performance is found to be poor.

We thus create a cascaded architecture (figure 1) composed of 1) a Fuzzy Inference System to assess pathology grading and staging, 2) an Artificial Neural Networks to learn and classify the severity of disease, 3) a Markov Decision Process to suggest an optimal policy for treatment. As illustrated in Figure 1, the proposed

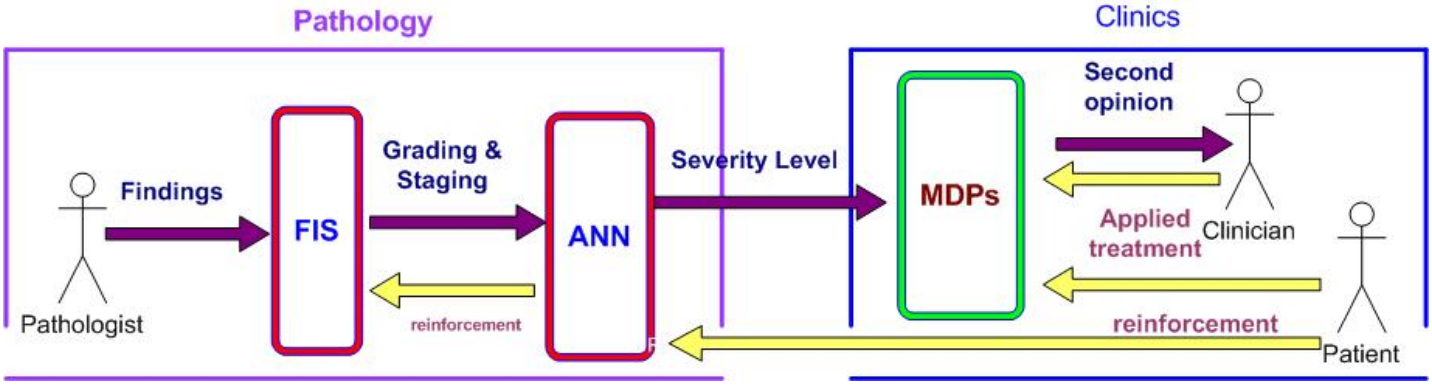

Figure 1: System Components

architecture integrates by bridging the diagnostic decision of the pathologist to the treatment decision of the clinician.

The system receives semi-quantitative tissue observation from the pathologist; assesses the grading and the level of severity information by utilizing the fuzzy inference system (FIS). It carries out the final diagnostic decision by artificial neural networks (ANN), and also grading the severity of the disease handles vagueness in the nature of such decision making by the combination of FIS and ANN.

2 Methods

Learning from experience, and self tuning is an important aspect of the designed system. The basic novelty of our architecture is to provide a complete diagnostic and treatment planning system, especially having a learning feedback to clinicians and pathologist from the responses of the patient to treatment using reinforcement learning techniques. This feedback is a multi resolutional loop, not only furnishing data to modify the clinical treatment process but also to modify diagnostics at the pathology level for increased efficiency in similar future cases. Initial system parameters are based on a-priory domain knowledge and are subsequently automatically tuned by the learning mechanism of our structure, as new cases arise accompanied with success and failures of taken diagnostic and treatment decisions.

There are four families of learning methods in the literature, characterized by the differences in information source used for learning [7]. In our work we use reinforcement learning techniques that do not use an explicit teacher or supervisor, and are based on an internal evaluator, or critic, capable of evaluating the dynamic system performance [8]. In our model an ANN produces this evaluation. Since, the ultimate goal of the medical practice is to leverage a patient's health state, we base reinforcement on signs of body responses such as test results as well as evaluations of patients and clinicians within a treatment program.

The main strategy in reinforcement learning is to estimate the utility of taking actions in states of the world. Utilities that refer to success of the treatment are assigned by the experts. Learning has to find an optimal behavior (meaning a policy) that optimizes an evaluation function expressed in terms of reinforcement. There are various algorithms for learning to generate an optimal policy, such as adaptive heuristic critic, TD(λ) and Q-learning [9]. In this study we implement a Q-learning algorithm that considers discounted sum of reinforcements coming from taking an action in a state (planning a treatment for a state of a disease).

In hepatitis B biopsy is the gold standard for evaluation of liver disease. Pathology results are useful not only in diagnosis but also in grading disease severity and staging disease progression. However the reading of the liver biopsy remains quite subjective. [10]. Therefore fuzzy methods are used to address this problem of vague knowledge due to the linguistic nature of the domain [11]. A fuzzy rule base system is implemented to infer the grading and generate the state information of the patients' health. However, there are no explicit rules available in literature for classifying the severity level of chronic hepatitis B. Therefore in our work, the proposed system learns to classify this severity using an ANN trained by experimental data. Received from a FIS, this cascaded FIS and ANN subsystem generates the levels of severity of the disease as its output.

In our study we represent clinical state changes of the patient using MDPs to model the sequential treatment decision making under uncertainty, taking into account both outcomes of current treatment regimen and future treatment planning. Treatment planning methods involves the ability to predict the interplay between the natural history of disease and the effects of clinical actions over a prolonged time. Most common formalisms of this problem are decision tress and influence diagrams, temporal belief networks, Markov Decision Process (MDPs) and Partially Observable Markov Decision Process (POMDPs). [12,13,14].

3. Proposed Architecture

Our system architecture is composed of four cascading modules (figure 1) namely FIS performing the pathology grading and staging of diagnostic decision, ANN classifying severity levels of chronic hepatitis B on the patient, MDPs representing clinical states of patients and suggesting treatment policy, and lastly reinforcement learning that enable the system to learn from experienced cases and sends two level feedback that tunes treatment planning of clinicians as well as diagnostic performance of pathologist through reinforcement obtained by patient responses to the given treatment.

3.1 FIS for Pathology Grading and Staging

Grade defines the severity of the hepatitis B disease. Liver tissue is evaluated according to following criteria: Degree of portal inflammation, Limiting Plate damage due to the inflammation, Hepotositic necrosis (confluent or focal), Fibrosis (from mild to cirrhosis), and Extend of the inflammation

Pathologists describe their observations with a linguistic term (fuzzy label) such as 'some', 'most', 'continuous',

etc. There are several histological classification schemes and scoring systems used in chronic hepatitis reporting [15]. Modified Ishak HAI scoring and grading system is one of them. In this study we use HAI, and with the help of a pathology expert, we transform the semi-quantitative descriptions of the HAI scoring system into if-then rules of our FIS module. HAI system has four grading and one staging score that we represented in our system if-then rules, such as; "*If 'limiting plate damage' is mild and 'extend of the inflammation in portal areas' is focal then 'Per portal or perispetal interface hepatitis' grade is 1*"

FIS rule based is composed of 30 rules where
R_i : ith rule of rule base
S_N: input variables
L_{ij}: linguistic term (fuzzy label) of input variable Sj in rule Ri. Its membership function denoted by μ_L
Y_N: output variables, grades and scores

FIS uses Mamdani type trapezoids and minimum operator as a fuzzy implication operator, and max-min operator for composition while centroid method is used for defuzzification. Matching degrees of fired rules are obtained from:

$$\alpha \text{ Ri (S)} = \max[\min(\mu L1 (S), \mu L1' (S)), \ldots, \min(\mu Ln (S), \mu Ln' (S))] \quad (1)$$

FIS outputs calculated according to the firing strength of the rules are:

$$Y(S) = \Sigma \alpha \text{ Ri (S)} \quad (2)$$

Figure 2 demostrates the decision surface for periportal or Perispetal Interface Hepatitis grading.

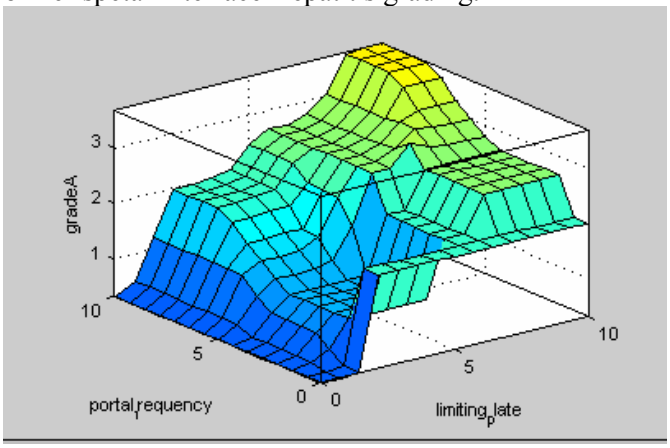

Figure 2: Decision Surface for grading A

3.2 ANN for Classifying Severity Level

Severity of chronic level of the disease is one of the main input data for clinicians. Severity is estimated by using grading and state values of the disease. We use an ANN with supervised training in cascade with FIS to classify levels of chronic hepatitis B. ANN inputs are the different grades and stage of the diagnosed disease while the outputs are chronic disease is absent, minimum, mild, medium or severe.

3.3 MDPs Formalization

Clinical state of the patient is formalized as MDPs. MDP is a stochastic control process and formally corresponds to a 4-tuple (S, A, P, R), [7]

1) S is a finite set of process states (patient states)

States are either level of disease or patients' reaction to the therapy.

Set of states in this study is: S:{Chronic Absent, minimum, mild, medium, severe, Complete Response, Partial Response, Breakthrough Response, Transient Response, Transplant Accepted, Transplant Rejected, Death }. Initial state is as received from the ANN module.

2) A is a finite set of actions (diagnostic and treatment procedures)

In this study, the action set represents the available treatment types such as: {Wait, Interferon Treatment, Lamivudine Treatment, Adefovir Treatment, Transplant} Action sets gives possible treatment actions for each disease level state.

3) P: SxAxS→[0, 1] is a set of transition probabilities between states that describe the dynamics of the modeled system

At each time step t, the agent observes the current state S_t and selects and action A_t from the set of possible actions corresponding to that state. When action is triggered the system state at the next time step is S_{t+1} with the probability $P_{S_t,S_{t+1}}(A_t)$. Transition probabilities among states are:

$$P^a_{SS'} = \Pr\{ S_{t+1} = s' \mid S_t = s,\ A_t = a\} \quad (3)$$

4) R: SxAxS → R denotes a reward (cost) model that assigns rewards to state transitions and models payoffs associated with such transitions.

Each state has an assigned reward value R [-100,100]. Expected value of next reward is:

$$R^a_{SS'} = E\{ R_{t+1} \mid S_t = s,\ At = a,\ S_{t+1} = s' \} \quad (4)$$

3.4 Reinforcement Learning

In this reinforcement learning model, the agent is connected to its environment: patient responses to treatment via perception and action. Perception is based on observations of physician and test results such as biochemistry, serology, or pathology. Actions are treatment options of the clinician, such as drug treatments, transplantation or just waiting. And environment is defined as health of the patient. At initial step, state of patient's health (chronic level of hepatitis B) is determined by classification of test results. Based on this information, physician suggests an action. Clinician is free to select treatment procedure either taking into account the system suggestion or determining

independently another policy. On each step of interaction the physician receives an input that is some indication of the current state of the patient. Then he/she chooses an action, treatment, to be administered. The action changes the state of the patient, and the value of this state transition is communicated back, evaluated by a scalar reinforcement signal, r.

3.4.1 Learning in the Treatment Planning

The action choice in each state is determined by a policy. Policy is a mapping from states to actions and is here denoted by π . Value of a state under a policy π is donated by $V\pi$ (S) and calculated with an evaluation function that represent the sum of all primary reinforcement functions received during a finite number of time steps.

$$V\pi(S) = E\pi [\Sigma^{n}_{t=0} \gamma_t r(S_t, \pi(S_t)) | S_0=S] \quad (5)$$

Model of optimality determines how an agent should take the future into account for the decision it makes about its present behavior. Finite horizon, infinite horizon or average reward models can be used for such optimality [16]. In our approach, we make use of immediate reward. A policy π is defined to be better then other policy π 'if expected return is greater than π' for all states. Value of taking action in state s under policy π is denoted by $Q\pi(S,A)$. Under optimal policy:

$$Q^*(S,A) = E \{ r_{t+1} + \gamma V^*(S+1) | S_t=s, A_t=a \} \quad (6)$$

In our model, Q value gives us how to behave optimally in a specific state. For any state of the patient, there can be numerous alternative treatment decisions. Treatment decision is taken by considering both the immediate reward (r_{t+1}) and value of any possible successive state (V*). Suggestion of system is based on these calculated Q values.

The immediate reward is a scalar signal which is received when action is done. In our model immediate response is produced by an ANN. Inputs of that ANN are biochemistry and serology test results, clincians and patient evaluations.

3.4.2 Learning in Classification of the Diagnosis

Disagreements often occur between pathologist and clinician. Chronic level classification being subjective, it generally varies from school to school. Clinician might argue that, result of classification is inconsistent with other test results. Therefore, he/she might take the liberty to plan the treatment based on different initial state. In such cases where clinical self initiative is taken and is followed by successful treatment results being, there is a positive feedback to the diagnostic level for tuning the diagnostic classification system. The diagnostic level ANN tunes for better diagnostics of future cases based on the successful experience of clinical self initiatives.

If we denote result of diagnostic classification by s and clinicians initiative s′ for any case j, the desired and actual outputs of ANN can be defined as $d_j(n)= s'$ and $y_j(n)=s$ respectively.

Weights of ANN are updated as follows:

$$\Delta w = - \eta \sigma E(n) / \sigma wij(n) \quad (9)$$

By means of updating weights, pathological diagnosis classification is tuned with the reinforcement learning feedback obtained from clinical treatment, and the system learn to harmonize pathologist's and clinician's views and successful experiences.

4. Results

4.1 Pathology Test Decision Making

Test results are carried in Gülhane Military Medical Academy (GATA) Hospital, Ankara. Diagnostic decisions are carried out using 16 patient cases. However, only two of them could be tracked down to clinical treatment level. Pathology findings under the close guidance of pathologist in GATA were collected for the 16 cases. Observations with fuzzy expressions were evaluated by the fuzzy inference systems.

The resulting grade and stage scores are then classified by the ANN. Table 1 gives the results of pathology grading by FIS and chronic severity level classification by ANN of two patients who are tracked down to the clinical treatment level. In following part of this section, we will present optimal treatment policy evaluation and two level reinforcement learning feedback for tuning system parameters based on these two patient data.

Table 1: Classification of Chronic Level

	Periportal/periseptal interface hepatitis	Confluent Necrosis	Focal Spotty necrosis	Portal Inflamation	Chronic Level
Patient 1	4	0	1	4	Severe
Patient 2	2	0	0	1	Minimal

4.2 Clinical Decision Making

4.2.1 Optimal Behavior Model

Optimal behavior model provides the information on how the agent should take the future into account in the decision it makes about its present behavior. In the clinical decision making human agent (clinician) evaluates the initial state of hepatitis B disease and generates feasible actions as future possibilities. As mentioned in the previous section of 'Proposed Architecture', there are 12 distinct states and 5 available actions for the treatment of the chronic Hepatitis B patients. Finite horizon model with 3 future steps is

applied for the optimal behavior selection. In the model, V gives the optimal value function for each state and P gives the optimum policy. Model run with 0.009 learning rate which indicates a more future benefit oriented approach. Obtained V and P matrices are presented in Table 2.

Table 2: Optimal Value Function and Policy (discount rate: 0.009)

	V=				P=		
Absent	271.0000	190.0000	100.0000	0	1	1	1
Minimum	160.6820	102.2000	41.0000	0	1	1	1
Mild	87.3425	47.4500	20.0000	0	1	1	1
Medium	115.7398	50.5145	0	0	2	2	1
Severe	177.8700	121.1700	58.8000	0	2	2	2
Comp.R	270.0000	189.0000	99.0000	0	1	1	1
Partial R.	0	0	0	0	1	1	1
Breakt. R	0	0	0	0	1	1	1
TransientR	149.0765	80.5100	9.5000	0	3	3	3
Trnspl. A.	0	0	0	0	1	1	1
Trnspl.R.	0	0	0	0	1	1	1
Death	0	0	0	0	1	1	1

1: wait
2: interferon treatment
3: Lamivudine treament

According those results of Table 4, best policies for disease 'absent', 'minimal' and 'mild' initial states are to 'wait'. And for disease 'medium' and 'severe' states 'Interferon treatment' is suggested. If transient response is observed 'lamivudine treatment' is given. For example, under the optimal policy, a patient with medium hepatitis should be treated with interferon treatment, and if any partial response observed treatment should continue with lamivudine treatment.

4.2.2 Reinforcement Learning

Case 1: Treatment Planning System Performance Tuning: Feedback to Clinical Level

In this case, MDPs is tuned by a reinforcement learning that is implemented by using Q learning rules. Q learning maps states to actions such that, in each state, there is a Q-value associated with each action. The definition of Q-value is the sum of the (possibly discounted) reinforcement received when performing the associated action and then following the given optimal policy thereafter. The transition rule of this Q learning is:

Q(state, action) =

$R(state,action) + \gamma \, Max[Q(next\ state, all\ actions)]$ (10)

In the Q matrix, row represent states and columns represent actions. In our case, patient 1' state is 'severe' then we search values of possible action in that state. Table 3 presents this row of the Q matrix which presents action values for 'severe' state, calculated with the 0.009 learning rate. Action values for the 'severe state' is calculated by addition of rewards matrix (R) to the maximum value of **Q** for all action in the next states multiplied by learning parameter γ (10).

Table 3:

Q severe=

0	**58,8**	25,5	-20,4	48
wait	**interferon**	iamivudine	adefovir	transplantation

As it can be observed from Table 6, suggested action for this 'severe' state is 'interferon treatment' with 58.8 action value. However the domain expert disagreed and decided to apply policy 3: 'lamuviden treatment'. After that treatment is applied he/she receives reinforcement from the patient as an immediate response to treatment in the form of a reward (r(severe, lamuviden)=0,2) via ANN. Then Q values are updated with this reinforcement information as follows:

Q(severe, lamuviden treatment)= r(severe, lamuviden) + 0.009 maxQ(complete R, breakthrough R; partial R)

The new Qsevere row is presented in Table 4.

Table 4:

Qsevere=

0	58,8	**26,93**	-20,4	48
wait	interferon	**lamivudine**	adefovir	transplantation

Consequently treatment planning system at the clinical level has learned from experiences, receiving a positive reinforcement value; Q matrix for chronic severe state with 'lamuviden treatment' is thus a slightly higher value 26.93 with the corresponding row, through the first level reinforcement feedback.

Case2: Diagnostic Classification Performance Tuning

In the second case, feedback is send to both clinical treatment and pathology level. In this situation patient 2's chronic level is classified as 'minimum' by the diagnostic system. However, the clinician disagrees: When she evaluates other findings such as serology test results, she is convinced that this second patient is at level Mild. Medical treatment is based on her judgment, and system produce treatment suggestion with Mild initial state such as yielding the raw of table 5 in the new Q matrix:

Table 5:

Qmild =

20	1.4	0	0	0
wait	interferon	lamivudine	adefovir	transplantation

Suggested action for this initial state is 'wait' (20). However the clinician chooses 'interferon treatment'. After that treatment is applied to the patient, an immediate reinforcement signal is received via the ANN as (r=0.4) hence being a positive reward. This generates a feedback which is sent to update the treatment policy as:

Q(mild, interferon treatment)= r(severe, interferon) + 0.009 maxQ(complete R, partial R, medium)

New Q value is as shown in Table 6:

Table 6:

Qmild =

20	**2,7**	0	0	0
wait	**interferon**	lamivudine	adefovir	transplantation

Similar to case 1, positive reinforcement increases the value of applied treatment and 'interferon treatment' is thus slightly updated to a higher value 2.7 through the first level reinforcement feedback.

Second feedback is send to the pathology classifier ANN. In this case, the input vector is the grading scores of the FIS: x_i: [2,0,0,2]

Actual output of the ANN is: $y_j(n)$=Medium

As a result of the received treatment feedback, the desired output becomes:

$d_j(n)$= Mild

This new case is added to the training set of the ANN and the classifier is retrained. Weights are updated with:

$\Delta w = - \eta \sigma E(n) / \sigma wij(n)$

Therefore, clinical experience obtained from clinical treatment is feedback to the classification of diagnosis and both treatment and diagnostic decision modules are tuned.

5 Conclusion

Our system not only integrates diagnostic decision and treatment planning problems making health restoration a continuous activity from the diagnostic carried out by the pathology to the treatment planning of the clinician; both, diagnostic and treatment planning are tuned and harmonized by a reinforcement feedback resulting from the patient response to treatment.

Our system learns from experiences via two level feedbacks that acts as a critic based on reinforcement learning. There, Q learning is used that assigns a value of taking an action for a particular state. When there is a disagreement between pathology and clinic decisions, clinician's decision is monitored and if it is successful, a reward feedback is returned to pathology diagnostic module. As a consequence the system has more robustness learning from cases involving both the treatment decision and the diagnostic decision.

References:

[1] Lin W, Tang J. DiagFH: An Expert System for Diagnosis of Fulminant hepatitis. Computer-Based Medical Systems: Fourth Annual IEEE Symposium. 1991: 330-336.

[2] Bomfadin C C, Adlassnig K P , Kreihsl M, Hatvan A, Horak W. A www Accessible Knowledge Base for the Interpretation of Hepatitis Serologic Tests. International Journal of Medical Informatics. 1997: v. 47: 57-60.

[3] Gamper J, Nejdl. Abstract Temporal Diagnosis in Medical Domains. Artificial Intelligence in Medicine. 1997: v. 10: 209-234.

[4] Buscher H P, Engler Ch., Führer A., Kirschkle S., Puppe, F. HepatoConsult: a knowledge-based second opinion and documentation system. Artificial Intelligence in Medicine. 2004: 205-216.

[5] Jajoo R, Mital D, Haque S, Srinivasan S. Prediction of Hepattis C using Artificial Neural Network. Seventh International Conference on Control, Autmation, Robotics and Vision. 2002: 1545-1550.

[6] Bonfa I, Maioli C, Sarti F., Milandri G L, Dal Monte P R. HERMES: an Expert System for the Prognosis of Hepatic Diseases. Proceedings of the IEEE. 2004: V 92 Issue 11:1759 - 1779

[7] Jouffe L, Fuzzy Inference System Learning by Reinforcement Methods. IEEE Transactions on System Man and Cybernetics. Part C: Applications and Reviews 1998; 28-3:338-355.

[8] Berenji H R, Khedkar P. Learning and Tuning Fuzzy Logic Controllers Through Reinforcements. IEEE Transactions on Neural Networks 1992; 3-5: 724-740.

[9] Kaelbling L P, Littnab M L. Reinforcement Learning: A Survey. Journal od Artifical Intelligence Research 1996; 4:237-285.

[10] Lin X, Sun Y, Horng M, Guo X. Computer Morphometry for Liver Fibrosis Using An Automatic Image Analysis System. 18 th Annual Conference of the IEEE Engineering in Medicine and Biology Society. Amsterdam 1996: 683-684.

[11] Innocent, P.R., John, R.I. Computer aided fuzzy medical diagnosis. Information Sciences. 2004. 162: 81-104.

[12] Magni, P., Quaglini, S. , Marchetti, M., Barosi, G. Deciding when to intervene: a Markov decision process approach. International Journal of Medical Informatics.2000: 60: 237-253.

[13] Peek N.B. Explicit temporal models for decision-theoretic planning of clinical management. Artificial Intelligence in Medicine 1999; 15: 135-154.

[14] Hauskrecht M, Fraser H. Planning treatment of ischemic heart disease with partially observable Markov decision processes. Artifical Intelligence in Medicine 2000; 18:221-244.

[15] Okafor O., Ojo S. A comparative analysis of six current histological classification schemes and scoring systems used in chronic hepatitis reporting Revista Espanola de Patologia. 2004: vol 37. n 3

[16] Sutton, R. S. Reinforcement learning : an introduction. Cambridge, Mass. MIT Press, 1998.