The Decision tree Mode for Prediction the Response to the Treatment in Patients with Chronic Hepatitis C

M. Hassan[1]; M. I. Abdalla[2]; S. R. Ahmed[1]; W. Akil[3]; G. Esmat[3]; S.Khamis[1]; M. ElHefnawi[1]

[1] Systems and Information Department, Engineering Division, National Research Center; Cairo, Egypt
[2]Electronics and Communication Department Faculty of Engineering, Zagazig University, Zagazig, Egypt
3 Hepatology Department, Cairo University Hospitals, Cairo, Egypt
engmarwa_3@yahoo.com; maddalla13356@hotmail.com; mahef@aucegypt.edu; shaimaa.tarek07@yahoo.com

Abstract: Interferon plus Ribavirin is the standard treatment for chronic hepatitis C is often accompanied by adverse side effects and unfortunately fails in almost half of cases. The ability to predict such failures previous to treatment could save a great deal of pain and expense for the patient with HCV. Decision tree with CART classification algorithm was developed to forecast response to therapy with 200 chronic hepatitis C patients. The data was divided where 150 cases were used to create the classifier and 50 cases for validation. HAI (hepatitis activity index), Fibrosis, ALT, age, and Gender were used as predictors for response of therapy. The overall classification error was 20% and 80% was the best accuracy, sensitivity and specificity were 0.80, 0.89% and 78%, respectively were results from validation stage as the following, true cases are 40 from 50 (total number of validation cases) and false cases 10 (represent error 20%). This model will help the physicians to build a decision before patient undergo treatment.

[M. Hassan; M. I. Abdalla; S. R. Ahmed; W. Akil; G. Esmat; S.Khamis; M. ElHefnaw. **The Decision tree Mode for Prediction the Response to the Treatment in Patients with Chronic Hepatitis C.** New York Science Journal 2011; 4(7): 69-79]. (ISSN: 1554-0200). http://www.sciencepub.net/newyork.

Keywords: Data mining; Decision tree; HCV; Virological response**;** Peg-interferon

1. Introduction

Hepatitis C is an infectious disease affecting the liver, caused by the hepatitis C virus (HCV) The infection is often asymptomatic, but once established, chronic infection can progress to scarring of the liver (fibrosis), and advanced scarring (cirrhosis) which is generally apparent after many years [1]. An estimated 130 million people worldwide [1] and nearly 4 million in the United States are chronically infected with HCV, leading to liver damage and increased risk of hepatocellular carcinoma. In the United States, 10,000 deaths each year are attributed to chronic HCV infection [2]. Chronic hepatitis C (CHC) affects more than 170 million individuals worldwide and is a chronic liver disease characterized by infection with the hepatitis C virus (HCV) persisting for more than six months [3]. The current treatment regime, pegylated IFN-α and ribavirin, is long and difficult, requiring months of weekly injections, with serious side effects ranging from flu-like symptoms to depression and autoimmune disorders[4]. Numerous studies in recent years have proposed markers for predicting HCV patient response to therapy [4]. Markers maybe based on viral factors, such as viral load, and genotype, host factors, such as age, gender, body mass index (BMI), ALT, AST fibrosis and cirrhosis.

Although combination therapy with pegylated interferon-alpha (peg-IFN-alpha) and ribavirin (RBV) has been the recommended treatment for CHC patients, many patients will not be cured by treatment. In addition to limited efficacy; it is well known that some patients often stop completing therapy, because of the high cost and significant unfavorable reactions.

As a result, it would be highly desirable to determine affect parameters on the response of treatment (the peg-IFN-alpha and RBV) and predict the possible outcome of therapy to distinguish responders from non responders [3].

There is no recognized standard interpretation system and different systems can produce different results from the same genotype [5]. Several groups have explored the use of bioinformatics to predict the therapy outcome. For example, artificial neural networks (ANN), decision tree (DT), support vector machines (SVM) or phenotype matching in relational databases have all been used to predict phenotype from genotype [5]. The aim of the study was to assess whether virologic response could be predicted accurately before therapy in patients treated with (pegylated) interferon-_ in combination with ribavirin for 24 to 48 weeks. One of the most widely used markers to predict response of therapy (therapy outcome) is PCR test (definition), which involves the measurement of markers, HAI, Fibrosis, ALT, age, and, gender, which, in combination have a high predictive value for the forecast responders and non responders. The correlation of these markers with therapy outcome involves a formula that was derived through logistic regression using data from 200 patients. The data were also classified for the construction of a classifier using decision tree, which were constructing using the CART algorithm.

The previous studies identified several variables that correlate with a greater chance of SVR, where achieved 70% accuracy of predict response and non response by artificial neural network. The progress in the field of informatics and its large use for decision making has led to the development of novel techniques related to machine learning using the Artificial Intelligence, which includes decision tree [1]. In chronic viral hepatitis data are lacking. Our study built-in 200 patients, 158(mal), and 42 (female) were treated with IFN plus RIB was retrospectively analyzed with the aim to predict the response to the treatment. 83 (0.415%) patients responders and 117 (0.585) non responders. six decision tree (DTs) consist of 2 classes, 9, 11, 13, and 17, pruning level, and nodes from 45 to 61were designed. Where, the chosen features for the decision tree were the patient's Hepatitis activity index, fibrosis, and Alanine Amino Transferase). DT model generated a The best and average accuracy were 80% and 75.2, respectively, sensitivity and specificity were 89% and 78% respectively, The correlation coefficient was 0.40, and the ROC AUC was 0.99.

Treatment with Peg interferon pulse ribavirin is considered the standard treatment for hepatitis C patient in Egypt but unfortunately the peg interferon plus ribavirin regimens have a number of drawbacks. Intolerable side effects necessitate pre-maturely stopping treatment and dose reductions in another Moreover, the drug regimen is very expensive (~US$26000 for a 48 week course), which means that most patients in countries such as Egypt, which has 10–12 million infected individuals, cannot afford this therapy[6] . patients chronically infected with HCV-4 responded favorably to PEG-IFN-{alpha}-2a/ribavirin therapy with SVR rates higher than those reported for genotype 1 with PEG-IFN-{alpha}-2a/ribavirin (46-58%) or PEG-IFN-{alpha} 2b/ribavirin (48%) 15-18 but lower than SVR rates in genotypes 2 and 3.

The ability to accurately predict the response of patients to antiviral therapy is great interest. In general, predictors may be clinical, biochemical or histological. They can be assessed before therapy is started (pre-treatment predictors) or during therapy (on-treatment predictors).

Some of their as viral and others are host factors such as: (Age; Gender; Body mass index; Albumin; Alanine Amino Transferase; Aspartate Amino Tansferase; Alfa-Feto Protein; Histology Activity Index; Viralload; Genotype; Fibrosis stage, and Cirrhosis.

Responses to therapy were defined as SVR and non-SVR including non response and relapse.SVR means undetectable HCV RNA at the end of treatment and after a further 24 weeks of follow –up (72 week in total) .relapse mean undetectable HCV –RNA at the end of treatment but positive after a further 24 weeks of follow –up .non –response mean detectable HCV –RNA during and the end of therapy [7].

2. Material and Methods

2.1. Patients

The study included 200 Hepatitis C patients with genotype 4 at Cairo University Hospital who were treated with combined therapy interferon-Alfa and ribavirin for 48weeks .Patients that shows clearance of the virus after 48 weeks were considered as responder and those who didn't show were considered as non responder.

2.2. Data preprocessing

Data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user which will be the neural network in our .our experiment used Ishak scoring which has achieved best train to DT than METAVIR the neural network take to two sets of data on for training and the other for testing so that 150 record were assigned for training and 50 record for testing the model ,which include 12 features(Age; Gender; Body mass index; Albumin ;Alanine Amino Transferase; Aspartate Amino Tansferase; Alfa-Feto Protein; Histology Activity Index; Viral load; Genotype; Fibrosis stage, and Cirrhosis) .The order of the pre-processing steps is important. One should avoid as much as possible the elimination of patients form the analysis during data pre-processing, and try to eliminate uninformative features first. If feature selection is performed first, even without using sophisticated methods for missing data imputation, the number of eliminated cases is smaller. The feature selection methods were performed in three steps.

1. Cleaning. Unimportant and problematic features and patients were removed.
2. Ranking. The remaining features were stored and ranks were assigned based on importance.
3. Selecting. The subset of features to use in subsequent models was identified.

In data cleaning, we always removed or excluded from the analysis the following variables: variables that have all missing values; variables that have all constant values; Variables that represent case ID. Variables that have more than 70% missing values; Categorical variables that have a single category; Counting for more than 90% cases; Continuous variables that have very small standard deviation; (Almost constants); Continuous variables that have a coefficient of variation CV < 0:1 (CV = standard deviation/mean); and Cases that have missing target values [8].

For ranking the features, "predictor" an important step of feature selection, also important step of feature selection, also important for understanding the biomedical problem, we used a simple but effective method which considers one feature at a time, To see how well each feature alone predicts the target Variable. For each feature, the value of its importance is calculated as (1 - p), where p is the p value of the corresponding statistical test of association between the runner feature and the target variable. The target variable was categorical with two or more categories for our problem, and the features were numerical, binary, and categorical. For categorical variables, the p value was based on Pearson's Chi-square. While the Non-categorical variables, p values based on the F statistic are used.

The importance of features (1−p) which calculated from P as explained above, and sorted first by p value in the ascending order or by descending (1-P) is recorded in Table 1 and Table 2.

Table1. Chi square test on Categorical parameters

Parameter	P -value	(1-P)
Fib stage	0.003	0.99
Cirrhosis	0.034	0.97
Age	0.188	0.82
Genotype	0.371	0.63
Gender	0.743	0.26

Table 2. F test on NON Categorical

Parameter	P -value	(1-P)
Viral load	< 0.0001	> 0.999
BMI (Body mass index)	< 0.0001	> 0.999
Albumin	< 0.0001	> 0.999
HAI (Histology Activity Index)	< 0.001	> 0.998
ALT (Alanine Amino Transferase)	< 0.001	> 0.998
AFP (Alfa-Feto Protein)	< 0.007	> 0.993
AST (Aspartate AminoTansferase)	< 0.016	> 0.984

3.3. Regression analysis

Regression analysis is a statically tool for the investigation of relationships between several variables where it includes any techniques for modeling and analyzing these variables, when the focus is on the relationship between a dependent and one or more independent variables. More specifically, regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed [9].

3.3.1 Multiple Regressions between all important parameters and study outcome

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. It is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables [9]. Plainly, study outcomes are affected by Varity of factors in addition to value of viral load, factors that were aggregated into the noise term is BMI, ALT, AST, AFP, HAI, viral load, fibrosis stage, Albumin, Cirrhosis, age, Gender, and Genotype, so we used multiple regression, is a technique that allow additional factors to enter the analysis separately. So that the effect of each can be estimated. It is valuable for quantifying the impact of various simultaneous influences upon a single dependent variable.

Assume:

Y=Study outcome, x=constant, α=albumin, β= afp/afp_ul, λ=viral load, η =fib stag, δ=HAI, and ε Minimum sum square error.

$$Y= x + a^{*} \alpha + b^{*} \beta + c^{*} \lambda + d^{*} \eta + e^{*} \delta + \varepsilon$$

Values of coefficient a, b, c, d, and e can be delivered readily and evaluated easily on a computer. We used med calc to calculate these values x= 0.169, a= 0.089, b= -0.069, c= 0.024, d= -0.086, e= 0.019, and ε=0.2, where ε = (sum of slandered error) /5=0.055, as ε is small value, the high correlation was satisfied between Study outcome and input features (albumin, afp/afp_ul, viral load, fib stag , and HAI). Values of standard error recorded as the result of med calc in table3.

Once the database was completed and all parameters passed through normalized test before using data sheet, data test based on estimating correlation between parameters and study outcome (Target), which have values from 0.9 to 1 these mean that there is a positive linear relationship between the data columns and target column, the constructed ANN should be started. Finally, the development steps in this study are outlines in this flow chart Figure 2.

Table 3: Regression equation

Independent variable	coefficient	Std. error
Constant(x)	0.169	_
Albumin(α)	0.089	0.082
afp/afp_ul (β)	- 0.069	0.079
Viraload(λ)	0.024	0.064
Fib_stag (η)	- 0.086	0.032
HAI(δ)	0.019	0.018

Table 4: List of variables used by the five DT

Field name	Description of variables	Values and Code
Age	years (rang)	20-58
Gender	Gender	M(158)=1,F(42)=0
BMI	Body mass index(rang)	16.84 - 43.15
Genotype	HCV Genotype(rang)	0:Non4 genotype (25),4:4 genotype(175)
AST	Aspartate Amino Transferase	0.01-0.29
ALT*	Alanine AminoTransferase	0.78-7.05
Cirrhosis	Absent or present cirrohsis	0:No(172),1:Yes(28)
Fibrosis score*	Score of fibrosis Activity(rang)	0 – 6
Albumin	Albumin(rang)	2.5-5.3
Viremia	Viral load (copies/ml)	0.006-5.050
AFP	Alfa feta protein	0.02-3.26
HAI*	Histology Activity Index	1-15
Result	Therapy result	1:response,0: Non-response

* The key parameters in our model

2.4. Statistical analysis

A pretreatment database of 12 variables was created containing 9 variables from the blood chemistry test [albumin, aspartate amino transferase, alanine amino transferase, Viral load, Histology Activity Index (HAI), genotype, fib stag, Cirrhosis and alpha-fetoprotein (AFP)], and 3 variables for patient characteristics (age, gender and body mass index). Based on this database, the recursive partitioning analysis algorithm referred to as decision tree analysis was implemented to define meaningful subgroups of patients with respect to the possibility of achieving SVR. Decision tree analysis is a family of nonparametric regression methods. Software is used to automatically explore the data to search for optimal split variables and to build a decision tree structure [10]. For the analysis, the entire study population was evaluated to determine which variables and cutoff points yielded the most significant division into 2 prognostic subgroups that were as homogeneous as possible for the probability of SVR. Thereafter, the same analytic process was applied to all newly defined subgroups. A restriction was imposed on the tree construction such that the procedure stopped when either no additional significant variable was detected. For this analysis, the data mining software med calc was used for multivariate logistic regression analysis [10].

3. Decision Trees

3.1 Definition

Decision trees, either classification or regression trees, are Linear nonparametric models represent the relationship between a continuous response variable and one or more predictor variables (either continuous or categorical) in the form $y = X \beta$, where:

Y is an n-by-1 vector of observations of the response variable.

X is the n-by-p design matrix determined by the predictors.

β is a p-by-1 vector of unknown parameters to be estimated.

4.2 The basic principle of Prediction Trees

The basic idea is very simple. We want to predict a response or class Y from Inputs X1; X2; ;; X p. We do this by growing a binary tree. At each internal node in the tree, we apply a test to one of the inputs, say Xi. Depending on the outcome of the test, we go to either the left or the right sub-branch of the tree. Eventually we come to a leaf node, where we make a prediction. This prediction aggregates or averages all the training data points which reach that leaf. Figure 1 should help clarify this [11].

C&RT, a recursive partitioning method, builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification) [12].

3.3 Features of Classification Trees

The data file reports the HAI, fibrosis and ALT of patients and target (study outcome) which have two classes (responders and non responders) [12].

The purpose of the analysis is to learn how we can discriminate between the two cases of study outcome, based on 12 attributes which have been discussed previously in material and method part, but three only (HAI, fibrosis and ALT) have satisfied the highest accuracy, which used to build the final decision tree after inputting all 200 layers into CART algorithm in figure 5.1. Discriminated function analysis will estimate several linear combinations of predictor variables for computing classification scores (or probabilities) that allow the user to determine the predicted classification for each observation. A classification tree will determine a set of logical if-then conditions (instead of linear equations) for predicting or classifying cases [13].

3.4 Algorithm

To generate classification tree, we used CART for all the experiments mentioned in this manuscript. It works as follows: to partition the data at each stage of tree, a test is performed to select an attribute with lowest entropy. Information gain (IG) is used as a measure of entropy (H) difference when an attribute contributes the additional information about class C [12].

$H(C) = -\Sigma\ p(c) \log p(c)$, $c \in C$ (1)

$H(C| Xi) = -\Sigma\ p(x)\ \Sigma\ p(c| x) \log p(c| x)$, $x \in Xi$, $c \in C$ (2)

$IG\ I = H(C) - H(C| Xi)$ (3)

In equation (1), p(c) is the probability that an arbitrary sample belongs to class 'c'. Equation (2) shows the entropy after observing the attribute Xi for the class 'c' and p (c| x) is the probability that a sample in attribute branch Xi belongs to class 'c' [15]. Table.5 shows different decision tree models, which we generated during experiments.

Where nodes represent questions about attribute values or ranges of values and edges represent the possible answers that link question nodes with other nodes down the tree, which represent further questions. Nodes at the bottom of the tree represent classes: the class of an object satisfying all the questions associated to the nodes in the path from the top question node to the bottom class node.

In our experiment the first step was to identify major variables which characterize the disease and affect prediction in order to define specific attributes for our DT. The following inputs were included in the model (table3) Age, gender, BMI, ALT, AST, AFP, HAI, viral load, fibrosis stage, Albumin, HCV genotype, and cirrhosis (preprocessing).

In a first hypothesis, elimination of cases with missing Values was not taken into account. Which were replaced with the most probable among those available; on the contrary, for numerical variables the mean value of available values, grouped in bases according to the output (positive/negative), was taken up. So we considered 200 cases, 172 with a positive result (responders) and 28 with a negative result (no responders). A correct construction of DTs based on the mixed database which facts necessary for the training phase and testing of six DTs [15].

We designed the 6 decision tree with (3 to 9) attributes from 12 (Age, gender, BMI, ALT, AST, AFP, HAI, viral load, fibrosis stage, Albumin, HCV genotype, and cirrhosis).

Within production, the structure of DT For the training and test of the decision tree in figure has been designed in the Mat lab ver.3.9 program (Scientific Software). This program enabled building DTs.

4. Results

4.1. Diagnostic performance of Six Decision tree

By finishing follow up, 83 of patients resulted responders (0.415) where as 117(0.585) were non-responders. We divided data to 150 cases for train and 50 for validation; the best decision tree gives the maximum accuracy %80.

Within table 5, the performance of the 6 decision tree is proved by (mean square error (MSE), predictive values accuracy, sensitivity, specificity, and AUC and ROC curve). As accuracy, sensitivity and specificity increase, But MSE decrease the model gives high performance. Sensitivity and specificity have diverse from 88% to 92% and from 32% to 64%, respectively. Relating to the predictive positive (following probability of treatment response) and negative values (following probability of no response to treatment), they varied from 35.3% to 55.6% and from 81.25% to 97.0%, respectively. The diagnostic accuracy rose from 68% (DT1) to 80% (DT6). By using Mat lab, Sensitivity, specificity, positive and negative likelihood ratio were calculated with mathematical equations. These parameters also automatically calculated through ROC curve analysis in Med Calc and give equivalent values.

Evaluations:

// the body mass index is:

$$BMI = \frac{Weight(kg)}{[height(m)]^2}$$

// the model accuracy, positive predictive, negative predictive values, sensitivity, and specificity can be computed as:

$$Accuracy = \frac{True\ positive(TP) + True\ negative(TN)}{False\ Negative(FN) + False\ positive + True\ positive(TP) + True\ negative(TN)}$$

$$positive\ predictive\ value = \frac{true\ positive(TP)}{True\ positive(TP) + Faulse\ positive(FP)}$$

$$Negative\ predictive\ value = \frac{True\ negative(TN)}{True\ Negative(TN) + Faulse\ negative(FN)}$$

Where,

Sensitivity (true positive rate) is probability that a test result will be positive when the disease is present.

Specificity (true negative rate) is probability that a test result will be negative when the disease is not present.

Accuracy is (True rate) is probability that a test result will be negative and positive.

The preprocessed data used for training and validation for six DT with 3to 12 different attributes .six DT architecture was evaluated with three attributes (Fibrosis score, ALT, and HAI) consist of 2classes, 9, 11, 13, and 17, pruning level, and nodes from 45 to 61.the results shows that decision tree with 61 nodes has the best performance which indicated in figure, while the decision tree with 45 has the least. In table5 the sensitivity, specificity, predictive values and diagnostic accuracy of the 6 DTs are shown [9].

Table 5. Performance of 6 DTs

Decision tree number	TP	TN	Positive predictive value %	Negative predictive value%	Sensitivity %	Specificity %	Accuracy %	AUC %
DT1	8	32	47.0	97.0	88.9	77.5	80	83.2
DT2	8	30	44.4	93.8	77.8	75.0	76	78.4
DT3	8	28	44.4	90.6	70.0	74.4	74	72.2
DT4	10	26	55.6	81.25	60.0	76.5	72	68.2
DT5	7	28	38.9	87.5	60.0	71.8	70	66.9
DT6	7	27	35.3	84.8	38.9	84.4	68	62.8

4.2 ROC curve analysis

The diagnostic performance of a test is evaluated using Receiver Operating Characteristic curve analysis (ROC) curves, which can be used to compare the diagnostic performance of two or more laboratory or diagnostic tests [9].

In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test [9].

5.2.1. Area under the ROC curve (AUC) with standard error and 95%.

Table 6. Indicate Results of the best DT

	Expected value(Target)	
ANN Output(y)	1	0
1	TP=8	FP=9
0	FN=1	TN=32

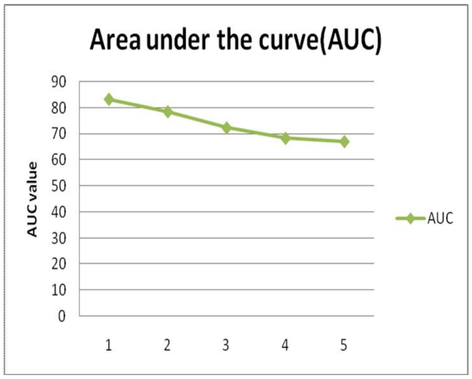


Figure 1. Area under the Curve of 6 DTs

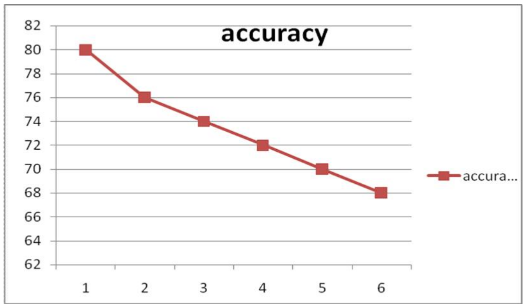


Figure 2. Accuracy of 6 DTs

When the variable under study cannot distinguish between the two groups, where there is no difference between the two distributions, the area will be equal to 0.5 (the ROC curve will coincide with the diagonal). When there is a perfect separation of the values of the two groups, there is no overlapping of the distributions, the area under the ROC curve Equals 1 (the ROC curve will reach the upper left corner of the plot the 95% confidence interval for the area can be used to test the hypothesis that the theoretical area is 0.5 [9].

From our result we have calculated the Regression between accuracy and area under the curve (AUC) of DTs as following:
Y=31.3648+0.586x, where significance level (p 0.001) and slandered error =2.1237.

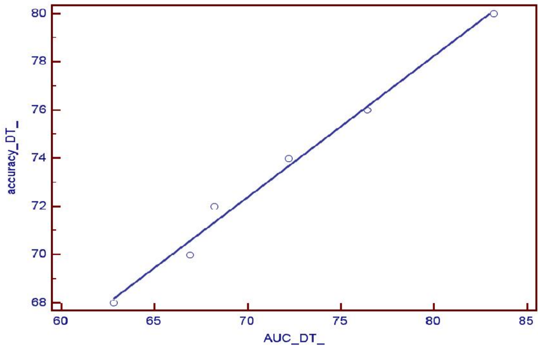
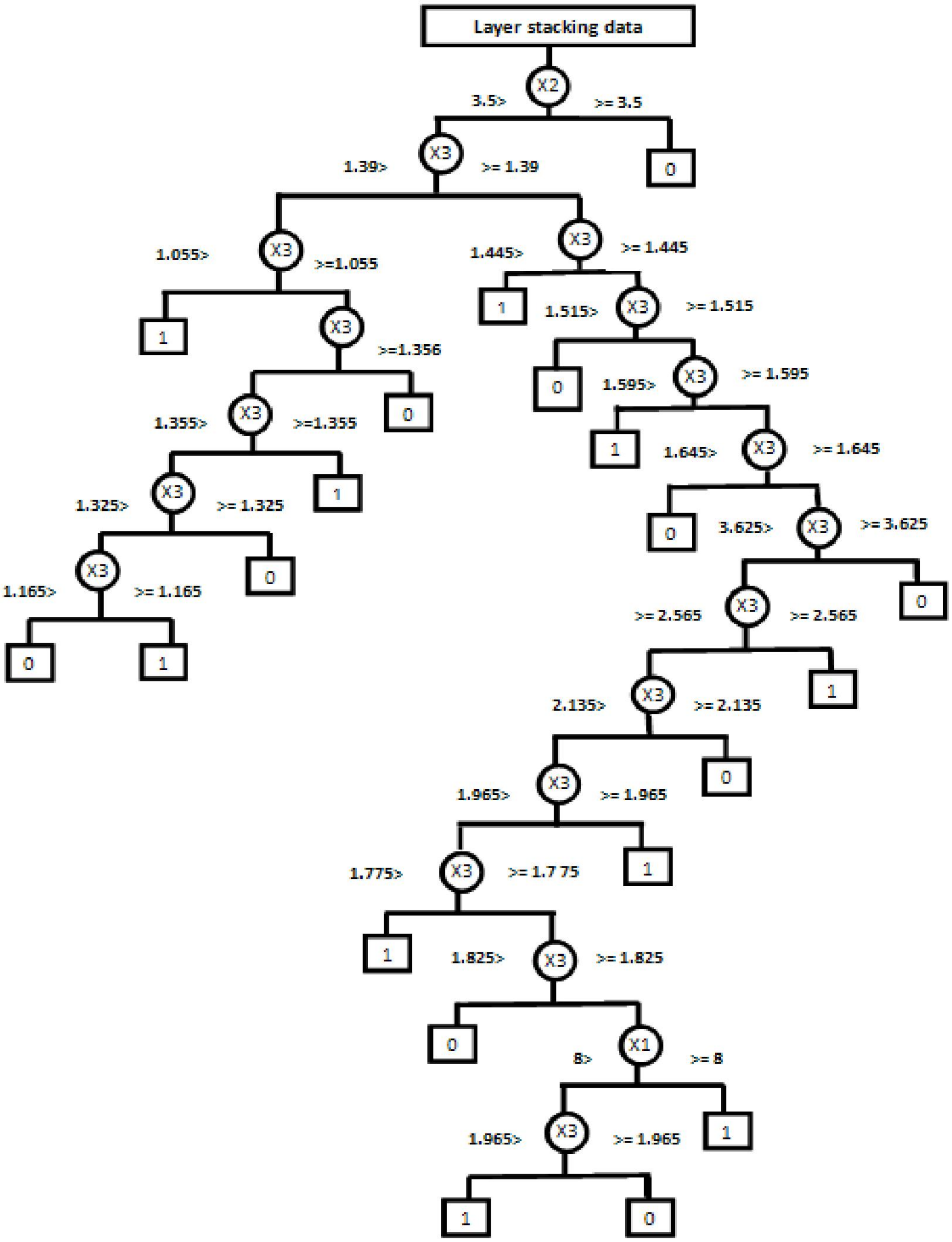
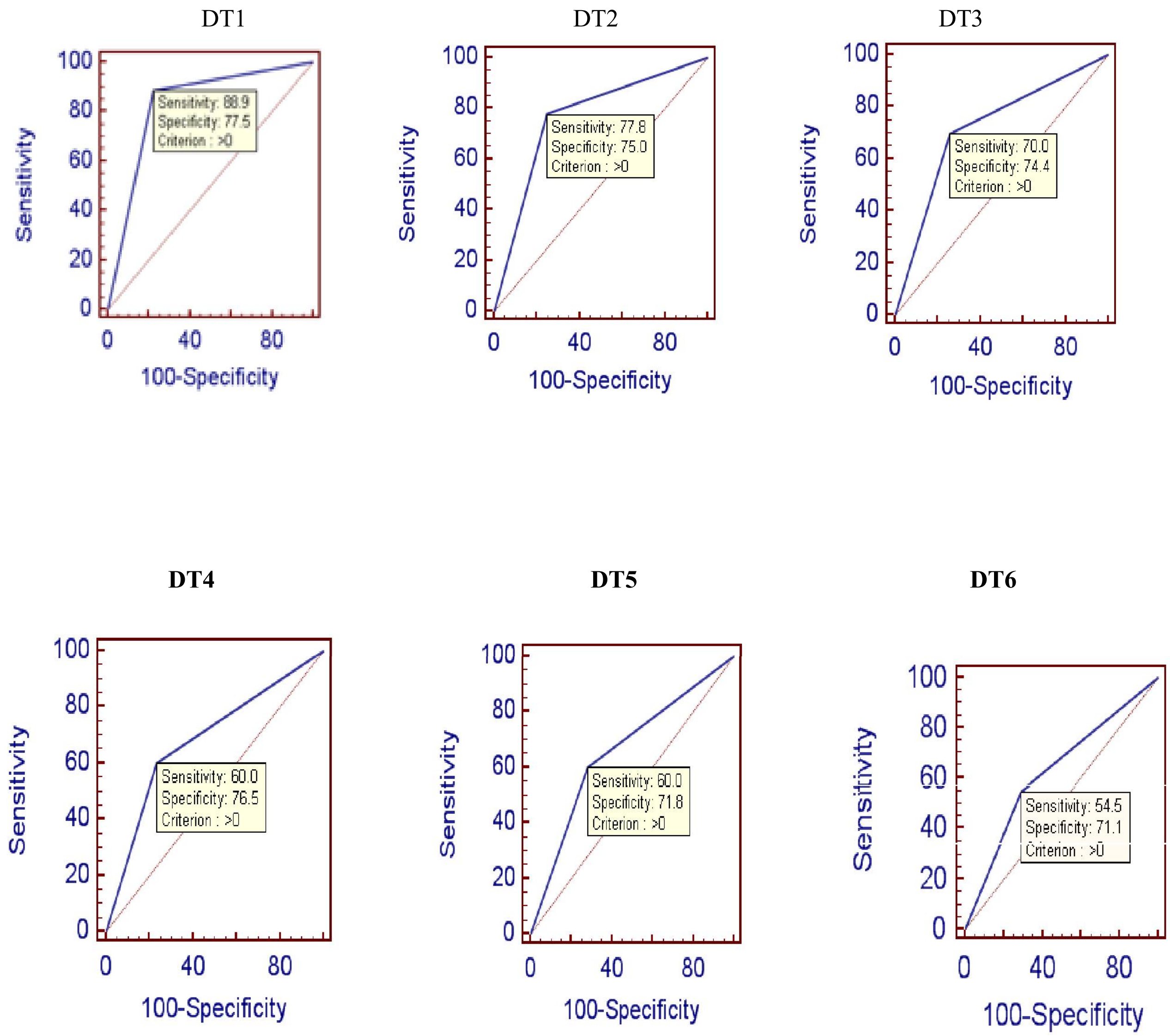


Figure 3. The relation between accuracy and AUC in DTs



The role of credit in empowerment of rural women



DT1

DT2

DT3

DT4

DT5

DT6

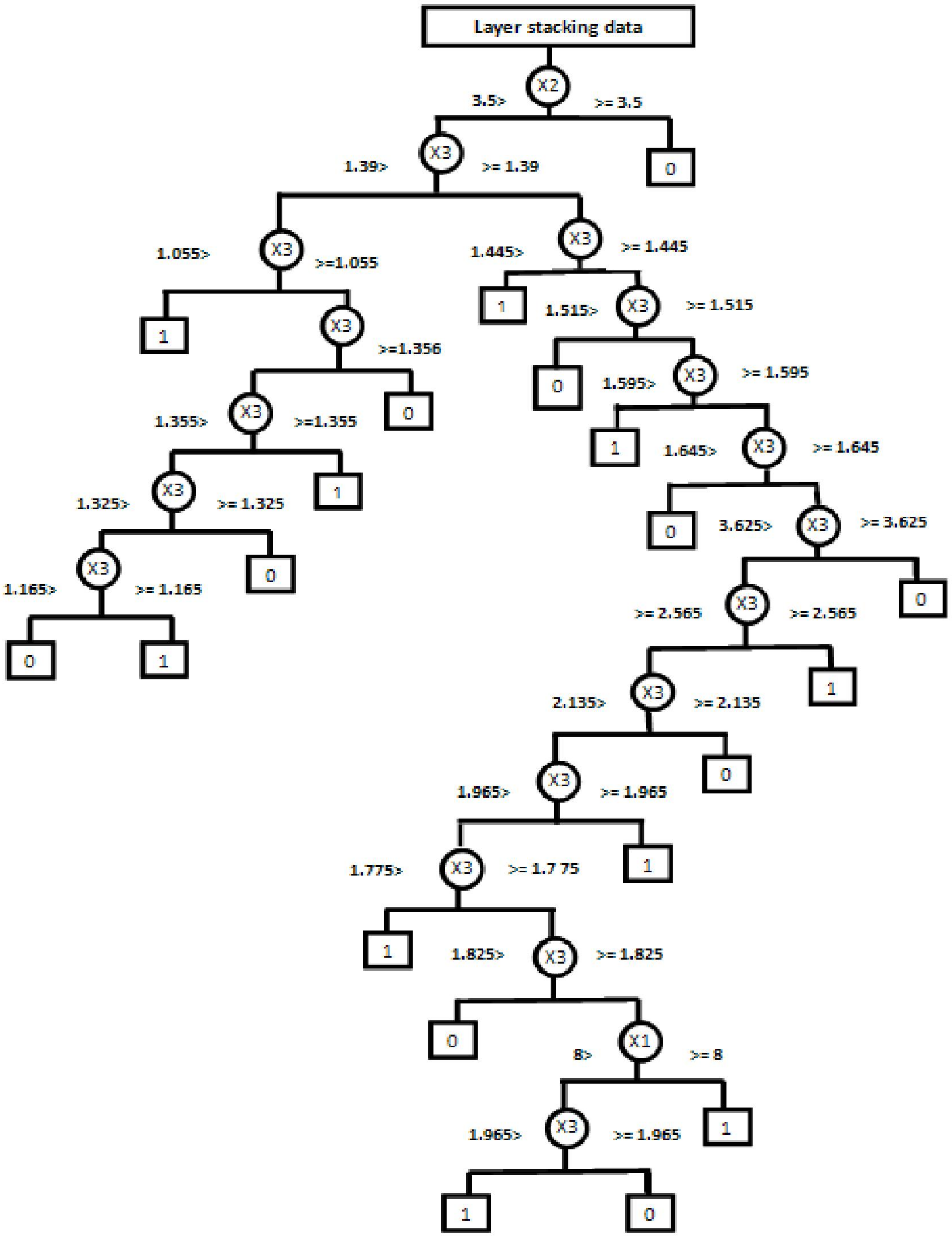
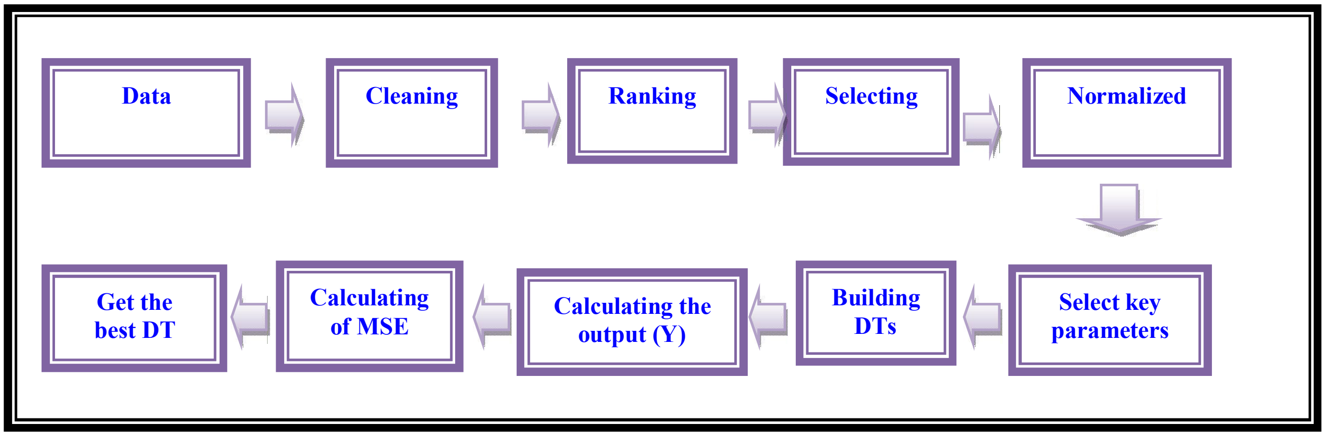


Figure 5. The decision tree constructed by CART algorithm.
(Where: X1 = HAI; x2 =fibrosis; x3=ALT.)

4. Discussion

The one hundredth of responders to IFN plus RIB treatment is still near to the ground, while its cost and side effects are elevated. Therefore, the possibility to predict patient's response to the above treatment is of the utmost importance. So far no author has explored the role of DTs as a tool for prediction of treatment response in Egyptian patients with chronic hepatitis C. Several viral and host factors have been associated with non-response [14]. Many studies have demonstrated that in non-responders, some interferon-stimulated genes were upregulated before treatment. Those findings associated to clinical, biochemical and histological data may help detect responders before starting any treatment. This is a very important issue because the standard treatment is physically and economically demanding [14]. Different elements are associated with Non-response: (i) viral factors, (ii) host factors [14]. The goal of this study is to identify factors that can help to predict the response to anti-HCV therapy.

In our show study where data of patients were treated with IFN alfa plus RIB included 12 features. Then passes pre-processing steps, then through chi-square and ANOVA F testes of association between the runner feature and the target variable. It cleared from these testes there are five key parameters have the highest effect on the therapy outcome, after that Building network, Calculating output (y), Calculation of MSE and, finally, get the best performance (DT1), the percentage of response and no response to treatment was58% and 42%, respectively. DTs have proven to be a good tool for forecasting.



Acknowledgements

I thank God who paved the way and only by his will everything can be achieved. And I'm deeply grateful to my advisor Prof.Dr.Mahmoud Ibrahim, Professor of Communication science, Zagazig University, and Dr.Mahmoud EL-Hefnawy, Professor of Bioinformatics, Information and Systems Department National Research center for continuous guidance, support and scientific supervision that enabled me to accomplish this study.

Corresponding author
M. Hassan
Systems and Information Department, Engineering Division, National Research Center, Cairo, Egypt
engmarwa_3@yahoo.com;

References

[1] P.A. Maiellaro, R. Cozzolongo and P. Marino, "Artificial Neural Networks for the Prediction of Response to Interferon Plus Ribavirin Treatment in Patients with Chronic Hepatitis C.", Current Pharmaceutical Design, 2004, 10, 2101-2109.

[2] M.H. Omran, S.S. Youssef, W.T. El-garf, and A.A. Tabll ,"Phylogenetic and Genotyping of Hepatitis C Virus in Egypt.",. Microbiology, 3:1-8.

[3] Chun-Hsiang Wang, Yuchi Hwang, and Eugene Lin2, "Pharmacogenomics of chronic hepatitis C therapy with genome-wide association studies", Journal of Experimental Pharmacology.

[4] Thomas S. Oh and Charles M. Rice,"Predicting response to hepatitis C therapy", Center for the Study of Hepatitis C, The Rockefeller University, New York, New York, and USA.

[6] Dechao Wang, Brendan Larder, Andrew Revell, Julio Montaner, Richard Harrigan, Frank De Wolf, Joep Lange, Scott Wegner, Lidia Ruiz, Marıa Jesus Perez-Elıas, Sean Emery, Jose Gatell, Antonella D'Arminio Monforte, Carlo Torti, Maurizio Zazzi, Clifford Lane," A comparison of three computational modeling methods for the prediction of virologiresponse to combination HIV therapy", Artificial Intelligence in Medicine (2009) 47, 63—74.

[7] T Asselah, I Bieche, S Narguet, A Sabbagh,I Laurendeau,M-P Ripault, N Boyer, M Martinot-Peignoux, D Valla,M Vidaud, P Marcellin, "Liver gene expression signature to predict responseto pegylated interferon plus ribavirin combination therapy in patients with chronic hepatitis C", Gut 2008;57:516–524. doi:10.1136/gut.2007.128611.

[8] A.G. Floares, "Artificial Intelligence Support for Interferon Treatment Decision in Chronic Hepatitis B," Engineering and Technology, 2008, pp. 110-115.

[9] Mat lab help.

[10] Masayuki Kurosaki , Naoya Sakamoto, Manabu Iwasaki, Minoru Sakamoto, Yoshiyuki Suzuki, Naoki Hiramatsu , Fuminaka Sugauchi, Hiroshi, Yatsuhashi, Namiki Izumi, "Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis," J Gastroenterol (2011) 46:401–409.

[11]http://www.stat.cmu.edu/~cshalizi/350/lectures /22/ lecture-22.pdf

[12]http://www.statsoft.com/textbook/classification-and-regression-trees/

[13] Eltahir. M. Elhadi. , Nagi. Zomrawi, "Object-based land use/cover extraction from QuickBird image using Decision tree," Nature and Science, 2009; 7(10).

[14] Tarik Asselah, Emilie Estrabaud, Ivan Bieche, Martine Lapalus, Simon De Muynck, Michel Vidaud, David Saadoun, Vassili Soumelis, and Patrick Marcellin., "Hepatitis C: viral and host factors associated with non-response to pegylated interferon plus ribavirin, "Liver International ISSN 1478-3223.

[15] Muhammad Umer Khan, Jong Pill Choi, Hyunjung Shin and Minkoo Kim., "Predicting Breast Cancer Survivability Using Fuzzy Decision Trees for Personalized Healthcare.," 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008.

6/11/2011