Accurate prediction of response to Interferon-based therapy in Egyptian patients with Chronic Hepatitis C using machine-learning approaches

Mahmoud ElHefnawi[1]; Mahmoud Abdalla[2]; Safaa Ahmed[1]; Wafaa Elakel[3]; Gamal Esmat[3]; Maissa Elraziky[4]; Shaima Khamis[1]; Marwa Hassan[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
[4] Professor of endemic medicine & Hepatology. Cairo University

Abstract

Hepatitis C virus' patients with genotypes 1 & 4 have break-even response rates to Pegylated-Interferon (Peg-IFN) and Ribavirin (RBV) treatment. Furthermore, the incompliance to the treatment because of its high cost and related unfavorable effects makes its prediction of paramount importance. By using machine-learning techniques, a significantly accurate predictive model constructed to predict Egyptian patients' response based on their clinical and biochemical data. The model uses Artificial Neural Networks (ANN) and Decision Trees (DT) to achieve this goal. Two-hundred patients treated with Peg-IFN and RBV; 83 responders (41%), and 117 non-responders (59%) retrospectively studied to extract informative features and train the Neural Networks and Decision Trees. Optimization done by using six different Neural Network architectures, starting with an input layer of 12 neurons, a hidden layer of 70 to 180 neurons and an output layer containing a single neuron. For decision Trees (DTs), the CART classification algorithm was used. Six DTs with two classes, pruning levels of 9, 11, 13, and 17, and nodes from 45 to 61 investigated. Among the 12 features in the study, the most statistically significant informative features were the patient's Histology activity index, fibrosis, viral-load, Alfa-feta protein and albumin. Validation of the models on a 20% test set was then performed. The best and average accuracy for the ANN and DT models were 0.76 and 0.69, and 0.80 and 0.72 respectively. Sensitivity and specificity were 0.95 and 0.39, and 0.89 and 0.78 respectively. We conclude that decision trees gave a higher accuracy in predicting response and would help in proper therapy options for patients.

Keywords
Data mining; Decision tree; HCV; response to therapy; Peg-interferon, machine-learning

Introduction

Hepatitis C is an infectious disease affecting the liver. It caused by the Hepatitis C Virus (HCV) [1]. The current treatment regime, Pegylated Interferon-α (Peg-IFN-α) and Ribavirin (RBV) is long and difficult, requiring months of weekly injections, with serious side effects ranging from flu-like symptoms to depression and autoimmune disorders [2]. Numerous studies in recent years have proposed markers for predicting HCV patient response to therapy [3]. Although combination therapy with Peg-

978-0-7695-4799-2/12 $26.00 © 2012 IEEE
DOI 10.1109/ASONAM.2012.140

IFN-α and RBV has been the recommended treatment for CHC patients, many patients will not be cured by treatment. In addition to limited efficacy, it well known that some patients often stop completing therapy because of the high cost and significant unfavorable reactions [4]. As a result, it would be highly desirable to determine effective parameters on the combination treatment response and predict the possible outcome of therapy to distinguish responders from non-responders [5]. The aim of the study was to assess whether virological response could predicted accurately before using combination therapy to treat the patients.In chronic viral hepatitis, the data of 200 patients treated with a combination of Peg-IFN and RBV was retrospectively analyzed; 83 (0.415%) patients responders and 117 (0.585) non-responders.

Combination treatment with Peg-IFN and RBV considered the standard treatment for hepatitis C patients in Egypt. Unfortunately, the Peg-IFN plus RBV regimens have a number of drawbacks. Intolerable side effects necessitate pre-maturely stopping treatment or dose reductions [5]. The SVR rates patients chronically infected with HCV-4 responded favorably to Peg-IFN-α-2a/RBV therapy with higher SVR rates than those reported for genotype 1[6].

Material and Methods

The study included 200 Hepatitis C patients with genotype 4 at Cairo University Hospital who were treated with combined therapy PEG-IFN-α and RBV for 48 weeks.

Patients who show clearance of the virus after 48 weeks considered responders, while those who did not show clearance considered non-responders

.

Data Preprocessing

Data preprocessing transforms the data into a format that would be more easily and effectively comprehended by the machine learning computational techniques. The Ishak inflammation scoring criteria was used which provided better training to the ANNmodel than METAVIR. The neural network takes to two sets of data; one for training and the other for testing. For training, 150 records were assigned and for testing 50 records were assigned. The model includes the following 12 features: Age; Gender; Body Mass Index; Albumin; Alanine Amino Transferase; Aspartate Amino Transferase; Alfa-Feto Protein; Histology Activity Index; Viral load; Genotype; Fibrosis stage, and Cirrhosis. The order of the preprocessing steps is important. One should avoid the elimination of patients from the analysis during data pre-processing as much as possible, and try to eliminate uninformative features first.

1. Cleaning: unimportant and problematic features and patients removed.
2. Ranking: the remaining features were stored and ranks assigned based on importance.
3. Selection: the subset of features to be used in the subsequent models was identified [7].

For each feature, the value of its importance calculated as (1 - P); where P is the value of the corresponding statistical test of association between the runner feature and the target variable. For categorical variables, the P value based on Pearson's Chi-square, while for non-categorical variables the P value based on the F test. Based on the features' importance, the

values sorted first by P value in ascending order and descending order for (1 - P) as recorded in table 1 and table 2. Once the database was completed, all parameters passed through normalization test. This means that there is a positive linear relationship between the data columns and target column [8]. The ANN model is then constructed. The development steps in this study outlined in the flow chart in Figure1.

Artificial Neural Network and Forecasting Treatment Efficacy in Chronic Hepatitis

An ANN is a paradigm that uses interconnected artificial neurons and mathematical models in order to represent complex and nonlinear relationships between input and output variables [9]. The most common type of ANN consists of three layers: an "input" layer, connected to a layer of "hidden" units, which is connected to a layer of "output" units. The activity of the input units represents the raw information that is fed into the network [10]. Feed Forward Back Propagation Neural Network model can used for class prediction based on a small number of predictors. Training a network by Back-propagation involves three stages: Feed forward of the input training pattern; Back-propagation of the associated error; Adjustment of weights [11].

Decision Trees' principle and Algorithm

The basic idea of DT is very simple. We want to predict a response or class Y from Inputs X1, X2,…, Xp. 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 [12].

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 performed to select an attribute with lowest entropy. Information gain (IG) used as a measure of entropy difference (H) when an attribute contributes the additional information about class C [13].

$$\text{Entropy} = H(C) = -\Sigma p(c) \log p(c) , c \in C \quad (1)$$

$$\text{Remainder} = H(C|Xi) = -\Sigma p(x) \Sigma p(c|x) \log p(c|x), x \in Xi, c \in C \quad (2)$$

$$IGi = \text{Entropy} - \text{Remainder} \quad (3)$$

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' [14]. Table 5 shows different decision tree models, which we have generated. The nodes represent questions about the attribute values or ranges of values. The edges represent the possible answers that link question nodes with other nodes down the tree, which in turn 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 [9].

Results

The data was divided into 150 cases for training and 50 for validation; the best ANN and DT have given the maximum accuracy

0.76 and 0.80 respectively.Within table 5, the performance of the best ANN and DT is shown : the mean square error (MSE), predictive values accuracy, sensitivity, specificity, and AUC. As accuracy, sensitivity and specificity increase, the MSE decreases and the model gives high performance. The ANN model gives sensitivity and specificity values that diverse from 88% to 92% and from 32% to 64% whilst that of the DT model gives sensitivity and specificity ranging from 88% to 92% and from 32% to 64%, respectively. As for the predictive positive and negative values, the values varied from 57.0% to 71.0% and from 69.0% to 84.0%, respectively, when using the ANN model. Nevertheless, the DT generated values from 35.3% to 55.6% and from 81.25% to 97.0%, respectively. The diagnostic accuracy for ANN 68% (DT6) to 80% (DT1) model rose from 62% (ANN6) to 76% (ANN1).

The preprocessed parameters used for training and validation for the six feed forward back-propagation neural networks: architectures consisted of 75, 80, 90, 125, 150 and 180 hidden nodes in one layer. The results shows that network with 180 nodes, has the best performance while the one with 70 has the least. In table four the sensitivity, specificity, predictive values and diagnostic accuracy of the best ANN and DT is shown. Six DT architectures evaluated with three attributes: fibrosis score, ALT, and HAI). They consist of two classes, pruning levels of 9, 11, 13, and 17 were used, and nodes from 45 to 61. The result shows that the decision tree with 61 nodes has the best performance, while the decision tree with 45 has the least. In table5, the sensitivity, specificity, predictive values and diagnostic accuracy of the six DTs recorded.

Comparing the ROC curves of 6 ANNs and 6 DTs, it is clear that, the ANN1 and DT1 with output Y1, are closer to the top left of ROC Curve and satisfy the higher AUC. This means that the ANN1 and DT1 satisfy the best ROC curve and the highest values of sensitivity and specificity of ANN, as in figure3 [15].

Discussion

ANN gives the best accuracy with five features: Albumin, AFP, Viral load, fibrosis score and HAI. DT however generated acceptable results with that with only three features: ALT, Fibrosis score and HAI). The final step is calculating the output (y), calculating the MSE. After this, the best ANN and DT selected. The percentage of responders and non-responders to the treatment was 58% and 42%, respectively. This resulted was generated using the ANN and DT which gave the best performance: ANN1 and DT1. They used for predicting the therapy outcome. As AUC, increases as accuracy increase that indicated in Figure3 the comparison between the results of the different ANNs and DTs can summarize in Figure 4 and Figure 5. Our system contributes valuable guidance to response classification specialists. It does so by utilizing machine-learning techniques. The process of training artificial neural networks and decision trees, which consequently brings benefits in aspects of patient, hospital, and healthcare system.

References

[1] N.N. Zein, “Ongoing Challenges in the Management of Chronic Hepatitis C Challenges in HCV Therapy,” English Journal, vol. 12, 2008.

[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, vol. 3, 2009, pp. 1-8.

[3] 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, vol. 10, Jan. 2004, pp. 2101-9.

[4] M.M. ElHefnawi, S. Zada, and I. a El-Azab, “Prediction of prognostic biomarkers for Interferon-based therapy to Hepatitis C Virus patients: a metaanalysis of the NS5A protein in subtypes 1a, 1b, and 3a.,” Virology journal, vol. 7, Jan. 2010, p. 130.

[5] S.S. Lee and A.A. Abdo, “Predicting antiviral treatment response in chronic hepatitis C : how accurate and how soon ?,” Journal of Antimicrobial Chemotherapy, 2003, pp. 487-491.

[6] M. Sherman, S. Shafran, K. Burak, K. Doucette, W. Wong, N. Girgrah, E. Yoshida, E. Renner, P. Wong, and M. Deschênes, “Management of chronic hepatitis C: consensus guidelines.,” Canadian journal of gastroenterology = Journal canadien de gastroenterologie, vol. 21 Suppl C, Jun. 2007, p. 25C-34C.

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

[8] Mat lab help

[9] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi, “Usefulness of an Artificial Neural Network for Differentiating Benign from Malignant Pulmonary Nodules on High-Resolution CT : Characteristic Analysis,” Lung Cancer, 2002, pp. 657-663.

[10] A. Drăgulescu, A. Albu, C. Gavriluță, and K. Menyhardt, “Statistical Analyses and Artificial Neural Networks for Prognoses in Hepatitis C,” vol. 3, 2006, pp. 71-79.

[11] O. Article, “Artificial Neural Network Analysis for Prediction of Headache Prognosis in Elderly Patients,” Journal of Medicine (Cincinnati), vol. 39, 2009, pp. 5-12.

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

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

[14] Muhammad Umer Khan et al "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.

[15] Med calc program.

Tables

Table 1:

Chi-square Test on Categorical Parameters

Parameter	P –value	(1-P)
fib_stag	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 Parameters

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

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 ANNs and DTs

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 AminoTransferase	0.01-0.29
ALT+	Alanine AminoTransferase	0.78-7.05
Cirrhosis	Absent or present cirrhosis	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 DT model
* The key parameters in ANN model

Table5
Performance of ANN1 and DT1

Neural Network Number	TP	TN	Positive predictive value %	Negative predictive value%	Sensitivity %	Specificity %	Accuracy %	AUC %
ANN1	22	16	71	84	88	64	76	76
DT1	8	32	47.0	97.0	88.9	77.5	80	83.2

Table6.Results of the best ANN model

	Expected value(Target)	
ANN Output(y)	1	0
1	TP=22	FP=9
0	FN=3	TN=16

Table7.Results of the best DT

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

Figures

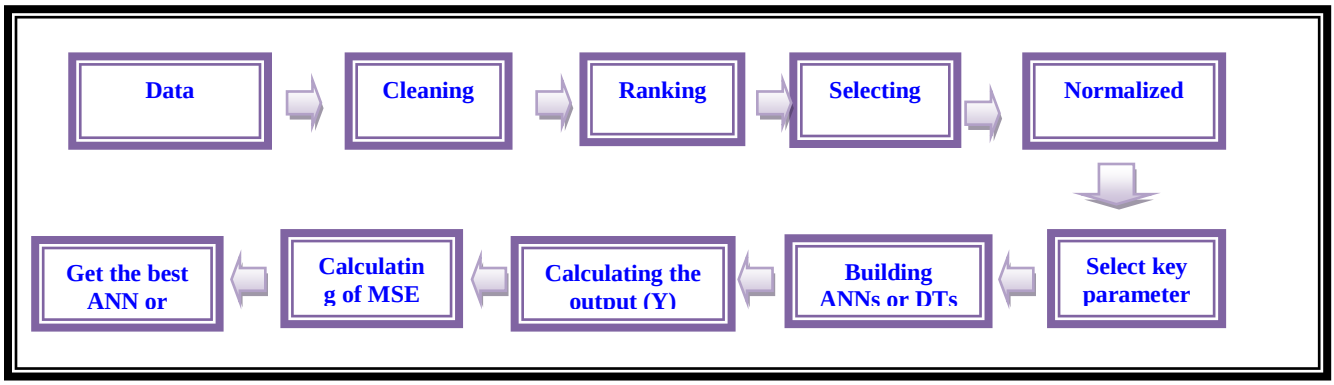

Figure 1: Flow chart showing steps for finding the best ANN or DT model for predict of response

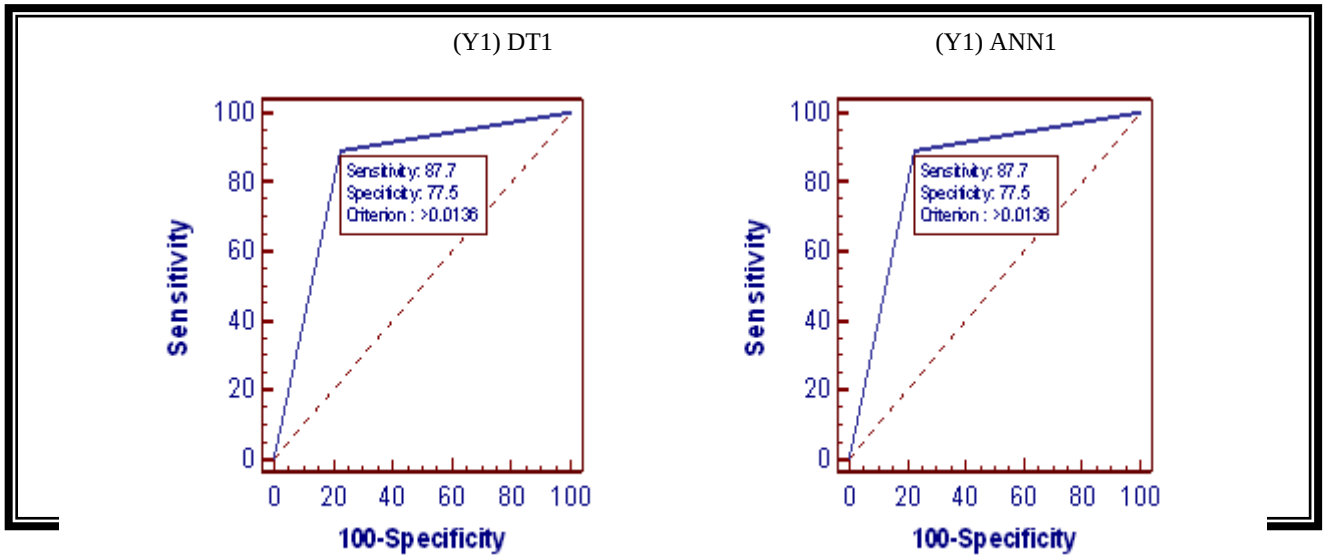

Figure 2: Comparison of the best ROC curve, values of Sensitivity and Specificity of ANN1 and DT1

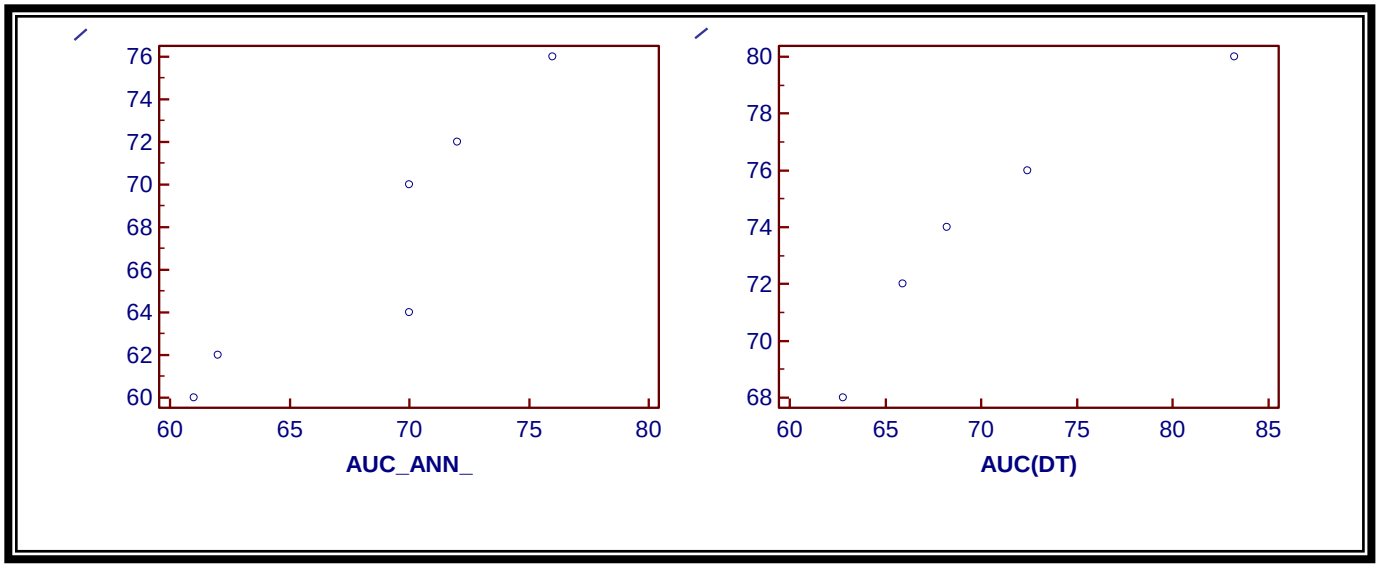

Figure 3: The relation between AUC and accuracy of 6 ANN and 6 DT.

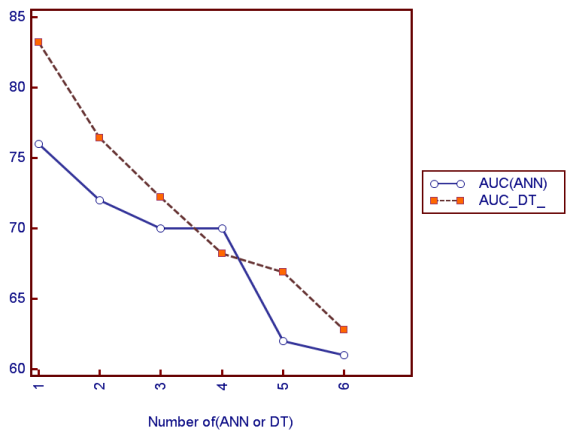

Figure 4: Comparison of the AUC values of 6 ANNs and AUC of 6 DTs.

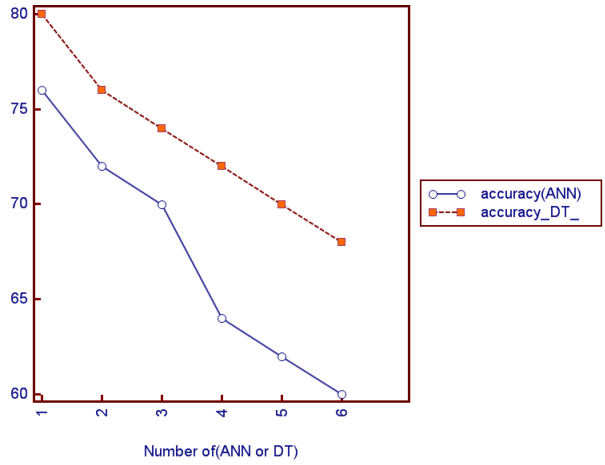

Figure 5: Comparison of the accuracy of 6 ANNs and 6DTs