

# A Comparison of Different Prediction Models in the Progression of Ocular hypertension to Primary Open Angle Glaucoma

Eng. M. Waly

El Shorouk Academy, High Institutes of Engineering, Systems and Biomedical Engineering Department

Amr. Sharawy, PhD.

Cairo University, Faculty of Engineering, Systems and Biomedical Engineering Department

K.Wahba, PhD.

Cairo University, Faculty of Engineering, Systems and Biomedical Engineering Department

## ABSTRACT

The issue of risk assessment in glaucoma has received increasing attention in the past few years. Predictive models are in order to estimate the risk that patients with ocular hypertension will develop to primary open angle glaucoma (POAG) if left untreated. These models are based on classification techniques on the risk factors. Classification is accomplished using conventional risk factors besides retinal nerve fiber layer (RNFL) thickness. It was found that RNFL is sensitive to glaucomatous damage by using different classification algorithms in order to reach to best prediction model.

We have applied the Decision tree (DT), Fuzzy logic and Neural Network to the glaucoma problem. The performances of the various classifiers are compared by the area under the receiver operating characteristics curve (AUROC) and the accuracy. The decision tree classifier gives the best result with accuracy 80% for the training dataset, 68.7% for testing data set with AUROC 0.868.

## Keywords

Glaucoma, primary open angle glaucoma, retinal fiber layer, generative and discriminative classifiers.

## 1. INTRODUCTION

Glaucoma is the most common optic neuropathic process affecting human and the second most common cause of blindness world-wide [1], [2]. The prevalence and pattern of disease varies in different regions of the world. It is estimated that 66.8 million people are affected by glaucoma worldwide and 6.7 million people are bilaterally blinded due to the disease. It is a disease in which progressive loss of retinal ganglion cells is characterized by a recognizable pattern of both visual function loss and optic nerve head pallor and excavation [1-4]. If untreated, the natural course of the disease is towards blindness, or at least significant visual loss disability [4].

Glaucoma is called the “sneak thief” of sight because it often goes undetected in its earliest stages figure 1. About 50% of individuals in North America with glaucoma are not aware of it, and the problem is worse in less medically advanced portions of the world [5-6].

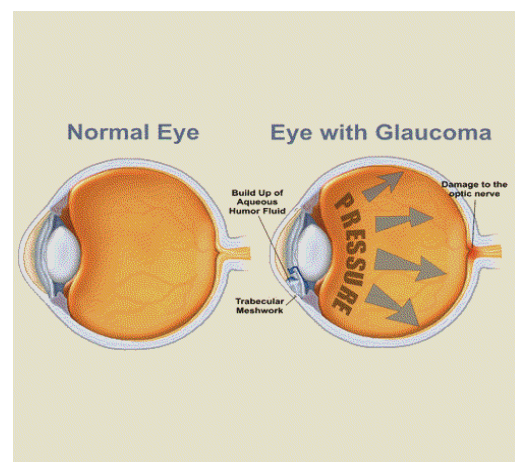


Figure.1 Normal eye and eye with glaucoma

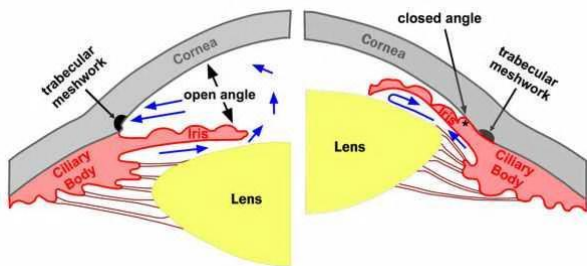
There are several reasons why glaucoma goes undetected [7]:

- Few programs are directed towards screening exclusively for glaucoma.
- Glaucoma usually does not affect both eyes with the same degree of severity at the same time, most patient don't know their condition as there is no clear symptoms or complain in early stages of the diseases.
- The progression of glaucomatous damage is usually slow. So, it is difficult for individuals to notice such gradual changes.
- Glaucoma primarily affects peripheral vision until the advanced stages of the disease process. As the field of the vision of the both eye complete each other.

Glaucoma can be defined as optic nerve neuropathy associated with changes in optic nerve cup/disc ratio and field of vision pulse increasing or decreasing in intraocular pressure (IOP), low tension glaucoma (LTG) in which all criteria of glaucoma are present but with low IOP according to normal intraocular pressure in the general population, ocular hypertension (OHT) refers to patients who have high IOP without affecting of field of vision or cup/disc ratio.

There are two main types of glaucoma figure 2: primary open-angle glaucoma (POAG), in which there is no anatomical blockage to the trabecular meshwork, and angle-closure glaucoma (ACG), in which an anatomical blockage prevents access to the trabecular meshwork. To distinguish

between these causes, it is necessary to perform gonioscopy to determine whether the angle is open or closed [7-9].



**Figure.2 Open and closed angle glaucoma**

In this study we will be interested in estimating the risk of progression of cases of OHT to POAG.

Primary open angle glaucoma, which affects almost 2% of the world population and accounts for most of the glaucoma cases. Although the pathophysiology of POAG is not precisely known, its causes are clearly multi-factorial. It is a result of multiple interactive genetic and environmental effects [10].

In this paper we will seek to develop a considerably powerful model that can be used to give the most accurate prediction for the progression of OHT to POAG by making a comparison between three different approaches, namely decision trees, fuzzy logic, and artificial neural networks.

## 2. MATERIALS

### 2.1 Risk factors for glaucoma development

The development of predictive models requires a series of complex steps which initially involve the acquisition and analysis of data from one or multiple longitudinal studies that have carefully followed patients over time. A critical step is the identification of the risk factors associated with the outcome one wants to predict. A few large, prospective, longitudinal studies have provided evidence with regard to the risk factors for conversion from ocular hypertension to glaucoma.

From these studies, two were randomized clinical trials, the Ocular Hypertension Treatment Study OHTS [11] and the European Glaucoma Prevention Study (EGPS) [12]. These two studies have provided the basis for the development and validation of nowadays available prediction models for glaucoma development. Both studies have evaluated a large number of predictive factors for their potential association with the risk of conversion to glaucoma. When pooled analyses of the OHTS and EGPS data were conducted, only five baseline factors were identified as significantly associated with the risk of conversion to glaucoma [13]:

Intraocular pressure (IOP), Central corneal thickness, Age, The measurement of the vertical cup/disc ratio of the optic nerve and The visual-field pattern standard deviation (PSD)

Table 1 shows relative risks for the baseline predictive factors found to be significantly associated with the risk of developing glaucoma in these two studies. These predictive factors have been incorporated into predictive models to

estimate the risk of conversion from ocular hypertension to glaucoma.

**Table 1. OHTS vs EGPS - risk factors**

	OHTS observation N = 819		EGPS placebo N = 522	
	HR	95% CI	HR	95% CI
Age (per decade)	1.12	0.91-1.39	1.39	1.01-1.91
IOP (per mmHg)	1.22	1.12-1.32	1.10	0.97-1.26
CCT (per 40 μm thinner)	2.03	1.61-2.55	2.12	1.51-2.97
PSD (per 0.2dB greater)	1.19	0.97-1.45	1.06	0.96-1.17
Vertical C/D ratio (per 0.1 larger)	1.27	1.14-1.43	1.26	1.03-1.53

Below, we review some of the evidence with regard to the predictive value of risk factors reported to be associated with glaucoma development.

#### 2.1.1 Intraocular pressure

In the pooled analysis of the OHTS and EGPS control groups (1319 patients followed without treatment), 1 mmHg higher baseline IOP was associated with a 9% higher risk of developing POAG (Hazard Ratio HR = 1.09; the 95% confidence interval CI = 1.03–1.17), after adjustment for other predictive factor [13]. It is important to note that even for this pooled analysis; the 95% confidence interval for the development of POAG was still relatively large, ranging from 1.03 to 1.17. That is, each 1 mmHg increased IOP could be associated with a 3% to 17% increased risk [14].

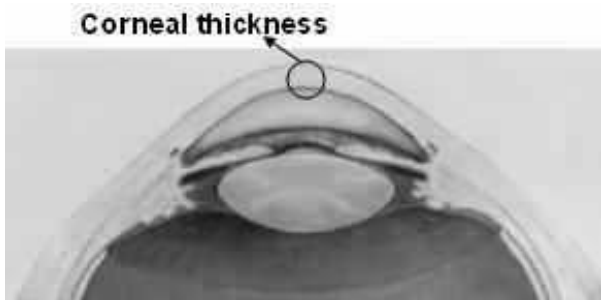
Because of the known association between elevated intraocular pressure (IOP) and glaucoma, IOP measurement has been considered a fundamental part of glaucoma evaluations for many years. Today, IOP remains important as a diagnostic indicator, but it is not the same critical component it used to be in the past. Lowering of IOP is unrelated to the diagnosis, although it remains the only modifiable risk factor in treatment [15].

Historical and current data have demonstrated that lowering IOP in confirmed cases of glaucoma can reduce the risk of increased damage and functional loss. Unfortunately, we now know that most untreated glaucoma patients in Asian countries have IOP levels that are considered to be within normal limits (normal- or low-tension glaucoma) [16-18]. Also, a large percentage of glaucoma patients, in general, have IOP measurements within normal limits for a single measurement [6]. Additionally, using IOP as a method of screening for glaucoma has less likelihood of diagnosing glaucoma than other clinical information such as optic disc and nerve fiber layer appearance, visual field properties, central corneal thickness, and age [19].

#### 2.1.2 Central corneal thickness (CCT)

Corneal thickness is another factor that has been associated with the risk of conversion from OHT to glaucoma figure 3. IOP as assessed by applanation tonometry may be overestimated or underestimated in thick or thin corneas, respectively [20-25]. A considerable subset of patients classified as having ocular hypertension may simply have

thicker than average corneas that result in an overestimation of what is likely a normal, true IOP. SO,



**Figure.3 Central Corneal thickness**

OHT patients with thicker corneas may be at a lower risk for glaucoma development. In fact, the OHTS showed that CCT was a powerful predictor of development of primary open-angle glaucoma among ocular hypertensive eyes [11]. Eyes with CCT of 555  $\mu\text{m}$  or less had a threefold greater risk of developing glaucoma compared with participants who had CCT of more than 588  $\mu\text{m}$ . A 40  $\mu\text{m}$  thinner cornea was associated with a 71% increase in the risk of conversion to glaucoma among OHTS patients in a multivariate model adjusting for other risk factors. Similar results were found by the EGPS, with a 40  $\mu\text{m}$  thinner cornea being associated with a 32% increase in the risk of conversion to glaucoma in the multivariate model [22].

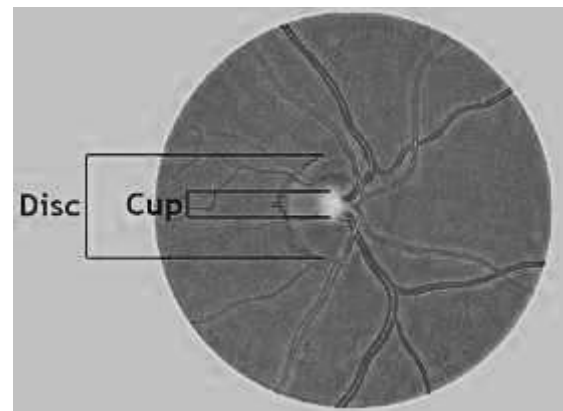
OHT patients with thinner corneas had a higher prevalence of abnormalities on this test compared to patients with thicker corneas. This additional evidence for the association between thinner corneas and the development of glaucomatous functional and structural damage supports the importance of considering central corneal thickness in the assessment of risk for the development of glaucoma in patients with ocular hypertension. The mechanism by which CCT influences the risk of developing glaucoma has not been completely established. Although the effect of corneal thickness could potentially be attributed to an artifact of tonometric measurements, it is possible that CCT could be a marker for biomechanical and structural characteristics of ocular tissues, which may influence the risk of development of glaucomatous neuropathy. Eyes with thinner corneas could have a particular structural susceptibility that would make them more prone to develop glaucomatous damage. Further studies are necessary to evaluate this hypothesis [23-24].

### 2.1.3 Age

There is strong evidence that older age is an independent risk factor for the progression of ocular hypertension and glaucoma. Older age has been reported as a risk factor for the development of glaucoma in patients with ocular hypertension in multiple longitudinal studies. Several population-based studies have also found that the incidence of POAG increases with older age. Both the OHTS and the EGPS found that older patients with ocular hypertension had an increased risk of conversion to glaucoma over time [14].

### 2.1.4 Cup/ disc ratio and pattern standard deviation of visual field

The OHTS as well as the EGPS and several other longitudinal studies have found that certain indicators of structural and functional integrity at baseline are predictive factors for development of overt glaucomatous optic neuropathy or visual field defects in the future. Two indices that have consistently been associated with higher risk of developing glaucoma are the vertical cup/disc ratio and the visual-field PSD, both measured at the baseline visit. Their assessment proves to be helpful in predicting which patients are more likely to develop clinically important stages of disease in the future, and their inclusion in predictive models is justified. Both vertical cup/disc ratio and PSD were significantly associated with the risk of developing glaucoma in the multivariate model combining OHTS and EGPS datasets. A 0.1 increase in vertical cup/disc ratio was associated with a 19% higher chance of developing glaucoma figure 4 [14].



**Figure.4 Cup/ Disc Ratio**

For the PSD, a 0.2 dB increase in the baseline PSD value was associated with a 13% increase in risk.

### 2.1.5 Retinal Nerve fiber Layer (RNFL)

RNFL assessment for glaucoma diagnosis and follow-up has several distinct advantages over current diagnostic approaches. It was demonstrated over 30 years ago that RNFL defects are the earliest sign of glaucoma. Since then, numerous studies have found that RNFL defects occur prior to visual field loss [26-28]. A study by Sommer et al. [27] found that 88% of ocular hypertensives who developed glaucoma had RNFL defects at the time the visual field defect was detected with standard automated perimetry (SAP). Furthermore, 60% of these patients had RNFL defects that were present 6 years prior to visual field defect. The evaluation of the RNFL is important for glaucoma diagnosis as RNFL damage often occurs earlier than can be detected with visual fields and even before optic nerve head damage. Red-free RNFL photography has many advantages, but the subjective interpretation of the results and the practical limitations of the method limit its usefulness [28].



## 2.2 Data collection

The combination between the structural and functional techniques for detecting glaucoma can improve detection, so practitioners must determine whether the patient has glaucoma or is at risk of developing glaucoma. The clinician must combine many aspects of information, including diagnostic test results, medical history, family history, clinical observations, prior examination findings and other related information [29-31].

The present study is based on analysis of prospectively collected data from randomly selected healthy individuals from a defined catchment area and glaucoma patients followed at the department of ophthalmology of Egypt Air hospital. The study includes about 398 study participants older than 40 years.

Each study participant underwent a comprehensive ophthalmic evaluation including review of medical history, best corrected visual acuity, slit lamp bio-microscopy, intraocular pressure measurement with Goldmann applanation tonometry gonioscopy, dilated slit lamp fundus examination with a 90-D lens, SAP using the Swedish interactive threshold algorithm (SITA) (Humphrey field analysis II, Carl Zeiss Meditec, Inc). Study participants underwent ocular imaging with commercially available optical coherence tomography (Cirrus OCT, Software 4.0, CarlZeiss, Meditec, Inc) in order to measure RNFL thickness.

Exclusion criteria for both groups included unreliable visual fields, angle abnormalities on gonioscopy; any diseases other than glaucoma that could affect the visual fields, and medications known to affect visual field sensitivity. Subjects with a best-corrected visual acuity worse than 20/40, spherical equivalent outside 6.5 diopters and a cylinder correction of 3.0 diopters were excluded. Poor quality stereoscopic photographs of the optic nerve head served as an exclusion ground for the glaucoma population. A family history of glaucoma was not an exclusion criterion.

Inclusion criteria for the glaucoma category were based on optic nerve damage and not visual field defects. The classification of an eye as glaucomatous or normal was based on the consensus of masked evaluations of two independent graders of a stereoscopic disc photograph.

All photographic evaluations were accomplished using a stereoscopic viewer (Asahi Pentax Stereo Viewer II) illuminated with color corrected fluorescent lighting. Glaucomatous optic neuropathy (GON) was defined by evidence of any of the following: excavation, neuro-retinal rim thinning or notching, nerve fiber layer defects, or an asymmetry of the vertical cup/disc ratio of 0.2. Inconsistencies between grader's evaluations were resolved through adjudication by a third evaluator. Inclusion criteria for the normal category required that the subjects have normal dilated eye examinations, open angles, and no evidence of visible GON. Normal optic discs had a cup-to-disc ratio asymmetry  $\leq 0.2$ , intact rims, and no hemorrhages, notches, excavation, or nerve fiber layer defects. Normal subjects had

intraocular pressure (IOP)  $\leq 22$  mm Hg with no history of elevated IOP. Excluded from the normal population were suspects with no GON and with IOP  $\geq 23$  mm Hg on at least two occasions. These suspects are part of a separate study on classification of stratified patient populations..

## 3. METHODS

Although it can be difficult to combine the large amount of data provided by currently available glaucoma detection techniques in a meaningful way, machine learning classifiers (MLC's) can accomplish this objectively. The application of machine learning methods in medicine for automated classification is a common practice after all [31-33].

Considering different modern classification methods competing for this task is usually insufficient to compare diagnostic performance. In clinical applications the definition of an examined disease is often complex and different examination methods are used [34]. For illustration we use available data of clinical study on early detection of glaucoma.

### 3.1 Discriminative and Generative Classification:

In a binary classification problem, we are given a training dataset  $\{x_i, y_i\}$ ,  $i = 1, \dots, N$  where  $x_i \in \mathbb{R}^D$  (which could contain both continuous and discrete entries) is the input and  $y_i = \pm 1$  is the output label. When performing classification, one approach is to first model the class-conditional probability  $p(x|C_{\pm})$  for each class  $C_{\pm}$ , and then employ Bayes' rule.

Under the Cox axioms [32], Bayes' rule is the only consistent way to manipulate beliefs and plausibility, if they are represented by real numbers. Classification using (1) is also

known as the generative model, since the probability of

$$P(C_{\pm}|\mathbf{x}) = \frac{p(\mathbf{x}|C_{\pm})P(C_{\pm})}{p(\mathbf{x})} \quad (1)$$

generating the data point "x" is first modeled. This effectively reduces the classification problem to that of modeling the class-conditional probability distribution  $p(x|C_{\pm})$  for the two classes.

However, it has always been difficult to model  $p(x|C_{\pm})$  accurately. The naive Bayes' classifier [33] assumes independence between the components of the input. Modeling  $p(C_{\pm}|x)$  through  $p(x|C_{\pm})$  is known to be inefficient [34], as it generally requires the estimation of more parameters. Take the example of performing classification by classical linear discriminant analysis (LDA): modeling the two classes of data with Gaussian densities of same variance but different means. It takes  $D(D+1)/2 + D + D$  parameters to calculate in this approach. The resulting classifier is well known to be a linear discriminant function  $U(x)$  which only needs  $D+1$  parameters.

$$U(x) = w_0 + \sum_{i=1}^D w_i x_i \quad (2)$$

Where  $w_i$  are the components of the weight vector  $w$ .

For a dataset of finite size, this means that we have fewer data points for each parameter in the generative approach. Unless



the equi-variance assumption fits well to the data, the classical LDA will have low efficiency for the sole purpose of classification. On the other hand, logistic regression [35] makes fewer assumptions about the classes and is generally more robust against outliers and noise in the data. Another weakness of the generative approach is that the model parameters are usually optimized by maximum-likelihood (ML) estimation [36]. It is widely believed that discriminative classifiers are to be preferred since the discriminative criterion is more closely related to the classification error.

The above arguments suggest that we may be better off using the discriminative approach in which the posterior probabilities  $p(C_{\pm}|x)$  are directly estimated. Logistic regression is a well-known example of the discriminative approach and is widely used in medical research. Decision trees, such as CART [37] or C4.5 [38], are another kind of discriminative classifier. Recently, attention has shifted to neural-network-type classifiers [39], [40] and the support vector machine (SVM) [41]. In some of these classifiers the estimation of the posterior probabilities is unnecessary. The classifier simply returns the label "y" by applying discrimination functions on the input. The advantage of discriminative classifiers is that they concentrate on the decision boundary and, hence, are usually strong against irrelevant outliers in the training data. However, they provide less insight into the structure of the data space and it is difficult to handle data containing missing entries.

## 3.2 Classifiers

### 3.2.1 Decision Tree

Decision tree learning is one of most widely used and practical methods for inductive inference, it is a method for approximating discrete-valued target functions in which the learned function is represented by decision tree. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of instance. Each node in the tree specifies a test of some attribute of instance and each branch descending from the node corresponds to one of possible values for this attribute. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, and then moving down the tree branch corresponding to the value of attribute, this process is then repeated for the sub-tree rooted at the new node [42, 43].

Most algorithms that have been developed for learning decision trees are a variation on a core algorithm that employs a top-down, greedy search through the space of possible decision trees. This approach is exemplified by Iterative Dichotomiser 3 (ID3) [42] and its successor C4.5 [43].

The basic algorithm ID3 starts decision tree learning by constructing trees top-down, beginning with the question "Which attribute should be tested at the root of the tree?" To answer this question, each instance attribute is evaluated using a statistical test to determine how well does it alone classify the training data? Then the best attribute is selected and used as a test attribute at the root node of the tree. We define a statistical property called information gain (IG) that measures

how well a given attribute separates the training data according their target classification [43].

$$IG = E_{\text{before}} - E_{\text{after}} \quad (3)$$

Where E is the Entropy, which measures the uncertainty associated with a random variable, it characterizes the (im) purity of an arbitrary collection of data.

Given a collection "S", containing positive and negative data of some target concept, the entropy of "S" relative to this boolean classification is

$$E(S) = -P_+ \cdot \text{Log}_2(P_+) - P_- \cdot \text{Log}_2(P_-) \quad (4)$$

Where  $P_+$  is the proportion of the positive data in "S" and  $P_-$  is the proportion of the negative data in "S" [Note: the entropy is "1" when the collection contains an equal number of positive and negative count. If the counts are unequal, the entropy is between 0 and 1] [44].

At each node of the tree, this calculation is done for each feature, and the feature with the largest IG is chosen for the split, this process continues iteratively until the end.

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm [44].

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set of already classified samples. Each sample is a vector which represents attributes or features of the sample. The training data is augmented with a vector which represents the class to which each sample belongs.

C4.5 made a number of improvements to ID3. Some of these are [43]:

- Handling both continuous and discrete attributes . In order to handle continuous attributes, C4.5 creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it.
- Handling training data with missing attribute values C4.5 allows attribute values to be marked as '?' for missing. Missing attribute values are simply not used in gain and entropy calculations.
- Handling attributes with differing costs.
- Pruning trees after creation C4.5 goes back through the tree once it has been created and attempts to remove branches that do not help by replacing them with leaf nodes.

### 3.2.2 Fuzzy logic

In 1982 Marr established two principles [45]:

#### I. Principle of Least Commitment

Do not do something that may later have to be undone.

#### II. Principle of Graceful Degradation

Degrading the data will not prevent the delivery of at least some of the answer. Which means no matter how the quality of data still some answers can be extracted from it, both

principles are very important in classification and decision-making processes for expert systems: the first principle is consistent with the continuous degree of belonging to fuzzy sets (equivalent with continuous degree of truth in fuzzy logic), and ensures the conservation of uncertainty until a crisp (binary) decision becomes necessary. The second principle asks for robust methods/algorithms to be used. The implementation of these two principles can be expressed in a natural way by using the fuzzy paradigm and classifications approaches [46].

Diagnosis as a medical activity will state if a patient suffers from a specific disease; and if the answer is yes, the specialist will provide a specific treatment. Despite the difficulties, the diagnosis of glaucoma is solved for the majority of cases. An important challenge for an ophthalmologist remains on the evaluation of the risk of occurrence and the prediction of progression to establish the suitable follow up and treatment accordingly. A major concern is the reliability of the diagnostic tools used by the physician. There is usually low confidence in these rules mainly due to their negative prediction rate. One of the glaucoma characteristics is that it can be “triggered” in very short periods of time (one hour for example) and without notice – which makes evident the challenge facing any attempt to predict it. Our goal is to face this challenge in developing a machine that can evaluate more precisely the risk factors [46].

According to the Marr’s “Principle of Least Commitment” we need to preserve as much as possible the natural embedded uncertainty in medical approaches (due of the natural complexity of the human); and according to his “Principle of Graceful Degradation” we need to build a robust system. And a natural way to fulfill these natural medical requirements is to use fuzzy sets and fuzzy “if-then” rules [47], [48].

Our design of the Fuzzy Glaucoma Diagnosis and Prediction System is based on a fuzzy inference system that matches some input values with a fuzzy diagnostic model (designed as a collection of fuzzy IF-THEN rules) to assign a risk factor and/or progression estimation. It uses a process like the one described in Figure 5.

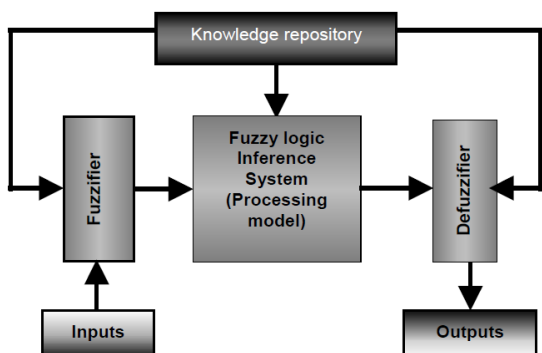


Figure.5 Fuzzy logic decision infrastructure

Figure.5 illustrates the basic flow of information of a fuzzy logic decision mechanism.

The knowledge repository contains a set of linguistic variables defined as a quintuple of the following form  $\langle X, T(X), U, G, M \rangle$ , where X is the name of the variable (inputs or outputs), T(X) is the set of linguistic terms for X, each of these terms has associated a fuzzy set in U, the Universe of discourse. U is the range of all possible values for this linguistic variable. The syntactic rule G is the grammar for generating the terms in the term set T(X). M is a semantic rule used for associating each linguistic term from T(X) with its meaning (membership function). The linguistic variables are the “vocabulary” that the fuzzy rules use to express the mapping from inputs into outputs.

For example, we can define “X”, the Intraocular Pressure (IOP), as a linguistic variable where the set term could be defined as  $T(IOP) = \{Low, Normal, High\}$

Each term in T (IOP) can be associated to a fuzzy set of values in the Universe of discourse  $U = [0, 45]$  (measured in mm of Hg).

Data and facts of glaucoma diagnosis and prediction (modeling environment) are transformed from a numerical level to the conceptual framework of fuzzy sets.

Low might be interpreted as “a pressure above 0 mm Hg and around 11mm Hg”; Normal as “a pressure around 16.5 mm Hg” and High as “a pressure around 21 mm Hg and below 45 mm Hg”. Every term can be described as fuzzy sets whose membership functions are like the ones drawn in Figure 6.

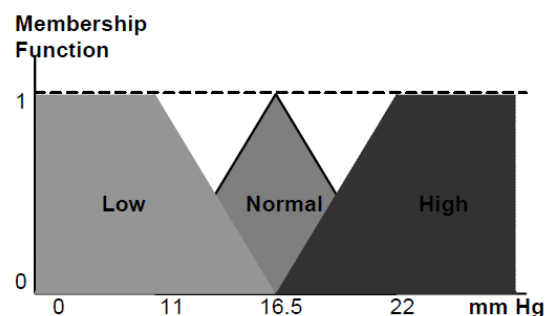


Figure.6 Fuzzy sets to characterize the linguistic variable IOP

The processing module is the algorithmic part of the schema, and its results are converted by the output interface (using some defuzzification technique) and returned to the modeling environment.

Fuzzy IF-THEN Rules (the knowledge representation selected for this project) are “extracted” from an expert’s knowledge and experience in a particular field. In some specific cases it is possible to “obtain automatically” such rules from data. This is not the case for our Fuzzy Glaucoma Diagnosis and Prediction System, due to the complexity of the diagnosis risk evaluation and prediction processes. In our case, it is essential to work close with medical specialists in order to obtain the knowledge necessary to build a complete set of IF-THEN rules; and to “confront” these rules with “mathematical tools” for verification and validation where appropriate.

An incremental development, a close relation to the ophthalmologists and a well-documented progressive work



were the foundation for the design of a process to create the Fuzzy IF-THEN Rules that will be used in the fuzzy Glaucoma Diagnosis and Prediction System.

### 3.2.3 Artificial Neural Network (ANN)

Multilayer perceptrons are often used for classification problems in medicine [49], [50].

The MLP is a layered neural network, i.e. its processing units, so-called neurons, are arranged in layers 1 to n, where layer 1 denotes the input layer and layer n denotes the output layer. Layers 2 to (n – 1) are called hidden layers. Multilayer perceptrons are feedforward networks, i.e. a neuron in layer 'a' can send a signal to a neuron in layer 'b' only if  $b > a$ . The connections between the neurons are weighted. The computational flexibility of MLPs results from the variability of these weights. Training is done by changing the weights in a way that the desired output for a given input is reached. Back-propagation is an efficient training method for MLPs [51]. This is a gradient-descent method, where the derivative of the neuron's output is needed. Therefore, the threshold function that was used in single-layer perceptrons is not appropriate. A similar function which is continuously differentiable, is given by the logistic function

$$f(z) = (1 + e^{-z})^{-1} \quad (5)$$

The limits of this function are 0 and 1 for  $z \rightarrow -\infty$  and  $z \rightarrow \infty$ , respectively. To prevent over-fitting of MLPs, weight decay can be used. With this variant of the learning algorithm, large weights are avoided and decision boundaries are smoothed. MLP scan separate nonlinearly separable data [51].

All MLPs consisted of 2 hidden layers with tangent hyperbolic transfer functions and an output layer of one neuron with a logistic transfer function that provided the MLP output. The number of neurons in the hidden layers was chosen based on the type of input used in order to achieve the best performance as judged by the results derived from the 10-fold cross validation procedure.

Our ANNs were constructed with the MATLAB Neural Network toolbox version 7 and trained with the scaled conjugate gradient algorithm described by Moller [52].

ANNs were trained and tested with the 10 fold cross-validation procedure, to reduce bias from training and testing on the same individuals, while fully utilizing our data set. Data were randomly divided into ten subsets, each containing test data from an approximately equal proportion of glaucoma patients and healthy individuals.

One subset was used to test classification performance while the remaining nine subsets were used for training purposes. In our ANN, one out of the nine training subsets was reserved for early stopping of the ANNs in order to avoid over-fitting. We additionally used bagging [53] of the remaining eight subsets to create the training sets used by the ANN ensemble. During training, this process was iterated, each time using a different subset as the early stopping set, until all the data subsets had been used to both train and stop the training of the ensemble. We further iterated the training process using each time a different test subset, so that all data could be used both for training and testing the classifiers, and averaged the test

results in order to produce a single performance measure for each ANN.

## 4. RESULTS AND DISCUSSION

The sample was split into a training sample consisting of 200 patients and a test sample consisting of 198 patients. The demographic data are shown in Table 2. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. In the field of artificial intelligence, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class [54].

Table 2. The Demographic data

Parameter	Min	Max	Mean	Standard deviation
Age	40	75	55.89	10.91
IOP	23	30	26.695	1.702
CUP/Disc	0.1	0.7	0.351	0.148
CCT	480	629	577.4	35.435
PSD	1	2.2	1.325	0.256
RNFL	-10	-0.06	-2.347	1.486

Sensitivity, specificity and accuracy are widely used statistics to describe a diagnostic test. In particular, they are used to quantify how good and reliable a test is. Sensitivity evaluates how good the test is at detecting a positive disease. Specificity estimates how likely patients without disease can be correctly ruled out. The ROC curve is a graphic representation of the relationship between both sensitivity and specificity, and it helps to decide the optimal model. Accuracy measures how correct a diagnostic test identifies and excludes a given condition. The accuracy of a diagnostic test can be determined from sensitivity and specificity with the presence of prevalence [54].

There are several terms that are commonly used along with the description of sensitivity, specificity and accuracy. They are true positive (TP), true negative (TN), false negative (FN), and false positive (FP). If a disease is proven present in a patient, and the given diagnostic test also indicates the presence of disease, the result of the diagnostic test is considered TP. Similarly, if a disease is proven absent in a



patient, and the diagnostic test suggests that the disease is absent as well, the test result is TN. Both true positive and true negative suggest a consistent result between the diagnostic test and the proven condition (also called standard of truth). However, no medical test is perfect. If the diagnostic test indicates the presence of disease in a patient who actually has no such disease, the test result is FP. Similarly, if the result of the diagnosis test suggests that the disease is absent for a patient with disease for sure, the test result is FN. Both false positive and false negative indicate that the test results are opposite to the actual condition.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (6)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (7)$$

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP}) \quad (8)$$

## 4.1 Training Results

### 4.1.1 Decision tree

Table 3 shows the confusion matrix with overall accuracy = 80% and Table 4 shows the result of True positive rate virus False positive rate in order to plot ROC.

Table.3 The confusion matrix of DT for training data set

Actual values		Predictive Values		
		Low	Mid	High
Low	62	38	0	
Mid	0	58	2	
High	0	0	40	

Table.4 The TPR viruses FPR of DT for training data set

	TP rate	FP rate
Low	0.62	0
Mid	0.966	0.373
High	1	0.0125

### 4.1.2 Fuzzy logic

Table 5 shows the confusion matrix with overall accuracy = 77% and Table 6 shows the result of True positive rate virus False positive rate in order to plot ROC.

Table.5 The confusion matrix of Fuzzy logic for training data set

Actual values		Predictive Values		
		Low	Mid	High
Low	56	44	0	
Mid	0	60	0	
High	0	2	38	

Table.6 The TPR virus FPR of Fuzzy logic for training data set

	TP rate	FP rate
Low	0.56	0
Mid	1	0.38
High	0.95	0

### 4.1.3 Neural network

Table 7 shows the confusion matrix with overall accuracy = 77% and Table 8 shows the result of True positive rate virus False positive rate in order to plot ROC.

Table.7 The confusion matrix of Neural network for training data set

Actual values		Predictive Values		
		Low	Mid	High
Low	62	38	0	
Mid	0	57	3	
High	0	5	35	

Table.8 The TPR virus FPR of Neural network for training data set

	TP rate	FP rate
Low	0.62	0
Mid	0.95	0.307
High	0.875	0.0187

## 4.2 Test Results

### 4.2.1 Decision tree

Table 9 shows the confusion matrix with overall accuracy = 68.7% and Table 10 shows the result of True positive rate virus False positive rate in order to plot ROC.

Table.9 The confusion matrix of DT for testing data set

Actual values		Predictive Values		
		Low	Mid	High
Low	12	21	1	
Mid	0	37	36	
High	0	6	87	





**Table.10** The TPR virus FPR of DT for testing data set

	TP rate	FP rate
Low	0.353	0
Mid	0.5	0.165
High	0.935	0.345

#### 4.2.2 Fuzzy logic

Table 11 shows the confusion matrix with overall accuracy = 65.1% and Table 12 shows the result of True positive rate virus False positive rate in order to plot ROC.

**Table.11** The confusion matrix of fuzzy logic for testing data set

Actual values		Predictive Values		
		Low	Mid	High
Low	2	30	0	
Mid	0	46	27	
High	0	12	81	

**Table.12**The TPR virus FPR of fuzzy logic for testing data set

	TP rate	FP rate
Low	0.0625	0
Mid	0.6301	0.2763
High	0.871	0.2571

#### 4.2.3 Neural network

Table 13 shows the confusion matrix with overall accuracy = 67.6% and Table 14 shows the result of True positive rate virus False positive rate in order to plot ROC.

**Table 13.** The confusion matrix of Neural network for testing data set

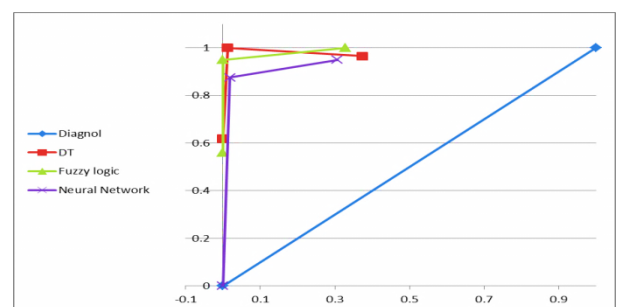
Actual values		Predictive Values		
		Low	Mid	High
Low	17	15	0	
Mid	4	34	35	
High	0	10	83	

**Table 14.** The TPR virus FPR of Neural network for testing data set

	TP rate	FP rate
Low	0.53	0.024
Mid	0.465	0.152
High	0.892	0.333

The comparison of classifiers relies on some measure of the performance of the recognition system. One performance measure that we used is based on the receiver operating characteristic (ROC) curve. In general, the area under the ROC curve (AUROC) gives the probability that the output value from one randomly chosen glaucoma eye, for instance, will be greater than the output value from one randomly chosen normal eye. Given two overlapping two-dimensional curves representing the distribution of the output value for each class (e.g., glaucoma eyes and normal eyes), the sensitivity (i.e., the percentage of glaucoma eyes classified as glaucoma), and specificity (i.e., the percentage of normal eyes identified as normal) will vary as the identification threshold is moved from one end of the curve to the other. The graph of the sensitivity on the y axis vs. 1" specificity on the x axis is the ROC curve. The AUROC indicates with one number the performance of the classifier. Chance alone is equivalent to an area of 0.5 and an area of 1.0, which is equivalent to perfect accuracy. Confidence intervals for the AUROC permit comparison of the performance of different classifiers.

It is also possible to measure performance by sensitivity at fixed levels of specificity or simply by diagnostic accuracy (captured as a single percentage as shown in Figure7).



**Figure 7.** ROC Curve

## 5. CONCLUSION

Of course, there are additional questions to answer. For instance what MLC techniques are best for what types of data? This question is complicated by the continuous development of new MLC techniques that may be better suited for the task. Because MLCs are very adaptable to data structure, it would be ideal to identify techniques that are



generalizable to many glaucoma related datasets. Such techniques should be able to combine information from multiple tests that generate data with varying distributions. The results of the present study suggest that the demographic predictive factors for the development or progression of glaucoma can be used to create a predictive model based on decision tree technique, which is the best technique due to it has overall accuracy 80% for the training set and about 68.7% for the testing set with AUROC 0.868 (Table 15).

**Table.15 Comparison between three classifiers**

Classifier	Accuracy in training data set	Accuracy in testing data set	AUROC
DT	80%	68.7%	0.868
Fuzzy Logic	77%	65.1%	0.815
Neural network	77%	67.7%	0.844

We can conclude that the decision tree gives the best accuracy regarding the training, as well as the testing data set, due to the nature of the input data.

The combination between the information from the optic disc measurement and RNFL measurement will provide better prediction.

Much work has been done employing developing machine learning techniques for solving the glaucoma classification problem during the last decade, automated techniques which based on supervised learning can perform as well as accurate at identifying glaucoma and detection it's progression .

## 6. RECOMMENDATIONS

The use of predictive models in clinical practice has several limitations. Predictive models are based on restricted populations of patients that were selected based on strict inclusion and exclusion criteria and that may not be representative of all patients seen at everyday clinical settings. Use of these models should be restricted to those patients who are similar to the ones included in the studies used to develop and/or validate it. It is also important to emphasize that although predictive models can provide a more objective evaluation of risk, their use does not replace the judgment of a clinician when making management decisions.

Also, it is important to emphasize that current risk calculators for glaucoma have been designed to estimate the risk of

development of the earliest signs of disease, which do not necessarily have an impact on the quality of vision of the patient. Finally, as more evidence regarding risk factors for disease development and progression accumulates, newer and better refined predictive models will be developed that should replace current existing ones.

## 7. FUTURE WORK

The future researches can be focused on development a classifier which give high accuracy for the prediction of progression of POAG. This classifier may be result from hybrid technique by using different classifiers or repeating the same classifier.

## 8. REFERENCES

- [1]. Quigley HA, Number of people with glaucoma worldwide, Br J Ophthalmol, 1996;80(5):389–93.
- [2]. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol 2006 Mar;90(3):262–7
- [3]. Tielsch JM, Sommer A, Katz J, et al., Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey, 1991;266:369–74.
- [4]. Mitchell P, Smith W, Attebo K, Healey PR, Prevalence of open angle glaucoma in Australia. The Blue Mountains Eye Study, Ophthalmology, 1996;103:1661–9.
- [5]. Dielemans I, VingerlingJR, Wolfs RCW, et al., The prevalence of primary open-angle glaucoma in a population-based study in the Netherlands. The Rotterdam Study, Ophthalmology 1994;101:1851–5.
- [6]. Topouzis F, Wilson MR, Harris A, et al., Prevalence of open angle glaucoma in Greece. The Thessaloniki Eye Study, Am J Ophthalmol, 2007;144:511–19.
- [7]. G. E. Trope, Glaucoma: A Patient's Guide to the Disease (Univ. of Toronto Pr., 2001)
- [8]. J. J. Kanski, J. A. McAllister, Glaucoma: A Colour Manual of Diagnosis and Treatment (Butterworths, London, Boston, Singapore, Sydney, Toronto, Wellington, 1989
- [9]. Kanski JJ. CLINICAL OPHTHALMOLOGY. Oxford: Butterworth Heinemann, 2003.(Guideline Ref ID: KANSKI2003) page 102:135.
- [10].Tielsch, J.M., Katz, J., et al. A population-based evaluation of glaucoma screening: the Baltimore eye survey. Am. J. Epidemiol., 1991 134(10): 1102–1110.
- [11].Gordon, M.O., Beiser, J.A., et al. The ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma Arch. Ophthalmol., 2002, 714–720; discussion 829–830.
- [12].Miglior, S., Pfeiffer, N., et al. Predictive factors for open angle glaucoma among patients with ocular hypertension in the European glaucoma prevention study. Ophthalmology, 2007 114(1): 3–9.



- [13].Gordon, M.O., Torri, V., et al. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology*,2007 ,114(1): 10–19.
- [14].Miglior, S. Risk factors for development of glaucoma. *European Ophthalmology review* , 2007, 105(1): 4–8.
- [15].Gordon, M.O. and Kass, M.A. The ocular hypertension treatment study: design and baseline description of the participants. *Arch. Ophthalmol.*, 1999, 117(5): 573–583.
- [16].Friedman DS, Wilson MR, Liebmann JM, et a. An evidence-based assessment of risk factors for the progression of ocular hypertension and glaucoma. *Am J Ophthalmol* 2004 Sep;138(3 Suppl):S19-31.
- [17].Anderson DR, Drance SM, Schulzer M; Collaborative Normal-Tension Glaucoma Study Group. Factors that predict the benefit of lowering intraocular pressure in normal tension glaucoma. *Am J Ophthalmol* 2003 Nov;136(5):820-9.
- [18].Shiose Y. The aging effect on intraocular pressure in an apparently normal population. *Arch Ophthalmol* 1984 Jun;102(6):883-7.
- [19].Detry-Morel M, Zeyen T, Kestelyn P, et al. Screening for glaucoma in a general population with the non-mydiatic fundus camera and the frequency doubling perimeter. *Eur J Ophthalmol* 2004 SepOct;14(5):387-93.
- [20].Ehlers, N., Bramsen, T., et al. Applanation tonometry and central corneal thickness. *Acta Ophthalmol.* 1975,53(1): 34–43.
- [21].Whitacre, M.M., Stein, R.A., et al. The effect of corneal thickness on applanation tonometry. *Am. J. Ophthalmol.*2004,115(5): 592–596.
- [22].Herndon, L.W., Choudhri, S.A., et al. Central corneal thickness in normal, glaucomatous, and ocular hypertensive eyes. *Arch. Ophthalmol.*,1997,115(9): 1137–1141.
- [23].Copt, R.P., Thomas, R., et al. Corneal thickness in ocular hypertension, primary open-angle glaucoma, and normal tension glaucoma. *Arch. Ophthalmol.*, 1999,117(1): 14–16.
- [24].Doughty, M.J. and Zaman,M.L. Human corneal thickness and its impact on intraocular pressure measures: a review and meta-analysis approach. *Surv.Ophthalmol.* 2000,44(5): 367–408.
- [25].Brandt, J.D. The influence of corneal thickness on the diagnosis and management of glaucoma. *J. Glaucoma*, 2001, 10(5 (Suppl 1)): S65–S67.
- [26].Sommer A. Retinal nerve fiber layer. *Am J Ophthalmol.* 1995;120:665–667
- [27].Sommer A, Katz J, Quigley HA, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol.*1991;109:77–83
- [28].Sommer A, Miller NR, Pollack I, et al. The nerve fiber layer in the diagnosis of glaucoma. *Arch Ophthalmol.* 1977;95:2149–2156.
- [29].Caprioli j. Discrimination between normal and glaucomatous eyes. *Invest ophthalmolvis sci.* 1992;33:153-159
- [30].Mardin CY. And Peter A. Improving glaucoma diagnosis by combination of perimetry and HRT measurements. *J Glaucoma.*2006;15:299-305.
- [31].Shah NN. Combining structural and functional testing for detection of glaucoma. *ophthalmology.* 2006;113:1593-1602.
- [32]. R. T. Cox, *The Algebra of Probable Inference.* Baltimore, MD: Johns Hopkins Univ. Press, 1961.
- [33].N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine Learning*, , 1997, vol. 29, pp. 131–163.
- [34].T. Mitchell, *Machine Learning.* New York: McGraw-Hill, 1997
- [35].D. R. Cox and E. J. Snell, *Analysis of Binary Data*, 2nd ed. London, U.K.: Chapman and Hall, 1989.
- [36].L. K. Saul and D. D. Lee, “Discriminative mixture modeling,” in *Advances in Neural Information Processing Systems 14.* Cambridge, MA: MIT Press, 2002
- [37].L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees.* New York: Chapman and Hall, 1993.
- [38].J. R. Quinlan, *C4.5: Programs for Machine Learning.* San Mateo, CA: Morgan Kaufmann, 1993.
- [39]. C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford, U.K.: Clarendon, 1995.
- [40]. B. D. Ripley, *Pattern Recognition and Neural Networks.* Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [41].V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York: Springer-Verlag, 2000.
- [42].Quinlan, J.R. Learning efficient classification procedures and their application to chess endgames. In R.S. Michalski, J.G. Carbonell & T.M. Mitchell, (Eds.), *Machine learning: An artificial intelligence approach.* Palo Alto: Tioga Publishing Company (1983a).
- [43].Quinlan, J.R. Decision trees and multi-valued attributes. In J.E. Hayes & D. Michie (Eds.), *Machine intelligence 11.* Oxford University Press (in press) (1985b).
- [44].T. Sharp. Implementing decision trees and forests on a gpu. In *ECCV*, 2008. 2, 5
- [45].D. Marr, *Vision*, W. H. Freeman and Company, San Francisco, California, 1982
- [46].N. Varachiu, *A Fuzzy Shapes Characterization for Robotics*, Lecture Notes in Computer Science, vol.1625, *Computational Intelligence - Theory and Applications*, Editor: B. Reusch, Sprienger-Verlag, Berlin, Heidelberg, New York, 1999, pp. 253-258
- [47].M. Uliuru, *Fuzzy logic in diagnosis: possibilistic network*, invited chapter in *Fuzzy Logic* (J. Baldwin, Ed.) John Wiley & Sons, ISBN o471962813, 1996



- [48].Fuzzy Control Manager, Version 1.5.4, TransferTech GmbH, Braunschweig, Germany, 2001
- [49].Heckerling PS, Gerber BS, Tape TG, Wigton RS.Entering the black box of neural networks. *Methods Inf Med* 2003; 42 (3): 287-296.
- [50].Seuchter SA, Eisenacher M, Riesbeck M, Gaebel W, Köpcke W, other members of the A.N.I. Study Group. Methods for predictor analysis of repeated measurements: application to psychiatric data. *Methods InfMed* 2004; 43 (2): 184-191.
- [51].Rumelhart DE, Hinton GE, Williams RJ.Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. *Parallel distributed processing (Vol 1)*. Cambridge, MA: MIT Press; 1986. pp 318-362.
- [52].Moller MF: A Scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*,1993,6:525-533
- [53].Haykin SS: *neural networks: a comprehensive foundation*. Upper saddle River,1999.
- [54].Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". *Remote Sensing of Environment* 62 (1): 77–89. doi:10.1016/S0034-4257(97)00083-7