

## The effects of the underlying disease and serum albumin on GFR prediction using the Adaptive Neuro Fuzzy Inference System (ANFIS)

Jamshid Nourozi<sup>1\*</sup>, Mitra Mahdavi Mazdeh<sup>2</sup>, Seyed Ahmad Mirbagheri<sup>3</sup>

Received 4 Apr 2014; Accepted 3 May 2014

### ABSTRACT

**Introduction:** Kidney disease is a major public health challenge worldwide. Epidemiologic data suggest a significant relationship between underlying diseases and decrease in Glomerular Filtration Rate (GFR). Clinical studies and laboratory research have shown that the mentioned parameter is effective in development and progression of the renal disease per se. In this study, we used learning-based system based on the neural network concepts.

**Method:** To predict GFR and propose an intelligent method with few errors (about 3%), we need to prognosticate the course and severity of the kidney disease in patients with chronic kidney disease using limited data and information. Adaptive neuro fuzzy inference system (ANFIS) used in the present study is based on the model proposed by Jang, and all laboratory (creatinine, calcium, phosphorus, albumin) and underlying disease caused by chronic kidney disease (CKD) were reviewed.

**Results:** It has been shown that the rate of GFR decreases in patients with diabetes and glomerulopathy was faster than other causes. Furthermore, serum albumin level less than 4.5gr/dl with diabetes was also associated with higher risk of rapid GFR loss.

**Conclusion:** Therefore, it seems that this modeling of fuzzy variables with error less than 3.5% in some cases and create a fuzzy inference system model that presents the complex relationships between the laboratory input variables and GFR as simple linear models.

**Keywords:** GFR, ANFIS, Underlying kidney disease, Albumin, CKD

► Please cite this paper as:

Nourozi J, Mahdavi Mazdeh M, Mirbagheri SA. The effects of the underlying disease and serum albumin on GFR prediction using the Adaptive Neuro Fuzzy Inference System (ANFIS). J Health Man & Info. 2014;1(3):46-50.

### Introduction

Chronic kidney disease (CKD) is a common but treatable disease. Early diagnosis of this "silent disease" is possible through primary statistical information and advanced mathematical methods. Despite the activities to determine the prevalence or risk factors of CKD, complementary studies are required to elucidate its epidemiology, especially in multi-ethnic populations.

Determining the prevalence of the disease, risk factor of progression in populations at risk, and use of intelligent methods, especially the neuro-fuzzy inference, have gained importance in recent years (1, 2). Many formulas have been developed in recent years to estimate the renal function using the patients' biochemical, demographic, and anthropometric data. Gaspari et al compared 12 available formulas for predicting GFR with Iohexol clearance as a reference and concluded that all formulas overestimated renal function while the Walser and MDRD formulas were preferred over others for accuracy and low bias (3).

Brier et al used artificial neural network to predict the occurrence of delayed graft function in transplant renal function in comparison with traditional logistic regression models on 304 renal transplant patients. Their results showed that logistic regression was more sensitive to prediction of no DGF (91 vs. 70%) while the neural network was more sensitive to prediction of yes for DGF (56 vs 37%). The sensitivity and specificity of the artificial neural network was 63.5% and 64.8%, respectively (4). Polat and Günes used a combination of principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS) to enhance the accuracy of the diagnosis of diabetes and obtained 89.47% accuracy. PCA is a statistical method used to decrease the variables that affect diabetes. The results of decreasing the variables were fed to NAFIS as input data and the disease status was regarded as the output (5). In another study, they used PCA and ANFIS to diagnose diseases related to the lymph system. The calcification accuracy of the system was 88.83% (6).

<sup>1</sup>Environment and Energy Department, Science and Research Branch, Islamic Azad University, Tehran, Iran

<sup>2</sup>Nephrology Research Center, Tehran University of Medical Sciences

<sup>3</sup>Civil Engineering Department, K. N. Toosi University of Technology, Tehran

\*Corresponding Author: J Nourozi, Environment and Water Quality Department, expert Hamedan Regional Water Co, Hamedan, Iran, P. O. Box 65159-1-4481, Tel: +98-81-32216717 And +98 - 9183509332 ; Fax: +98-81-38820225 . E-mail: nourozj@gmail.com .

Sengur made an attempt to determine the normal and abnormal states of cardiac valves using PCA, artificial neural networks, and fuzzy k-nearest neighbor. Their detection system had three stages. The first stage included pre-processing of the data (normalization and noise analysis). The second stage was features extraction through PCA and the third stage was determination of the status of the cardiac valves in 215 participants using artificial neural networks and KNN with a sensitivity of 95.9% and specificity of 96% (7).

We conducted the present study to evaluate the possible effect of underlying diseases on the rate of GFR decrease in patients with chronic kidney disease.

## Method

The files of the CKD patients who were visited in the nephrology clinic of Imam Khomeini Hospital, Tehran, Iran, were used during 2002-2011. The criteria of diagnosing CKD were small sized kidney on ultrasound and  $15 < \text{GFR} \leq 60$  cc/min/m<sup>2</sup> for more than 3 months and supervision by the clinic for at least 6 months. Then, clinical and laboratory data were extracted according to the Helsinki Treaty and the set of data were developed.

The patients were divided into two groups of those who were under regular visit and control group who did not follow the schedule of their visits. The first group included 389 patients who were visited at least every six months in the clinic. The control group included 76 patients who were not visited for at least one year and did not follow the instructions and directions.

The patients' information and their laboratory data including Ca, P, Cr, GFR, and Albumin were recorded in each visit until 2011. The process of care and performing laboratory tests stopped if the patient was prescribed to start dialysis, renal transplantation, GFR decrease to less than 15 cc/min/1.73 m<sup>2</sup>, or patient death.

The aim of this project was to predict the course of the progression of loss of renal function with regard to the underlying disease; therefore, the focus of the predictions was on the changes of this variable using the Adaptive Neuro Fuzzy Inference System (ANFIS). It is a learning-based system based on the neural networks concepts. According to the method used in Jang's doctorate thesis, this system is similar to the Takagi-Sugeno fuzzy inference system (8). Factors of layer 1 and layer 4 are of learner type. Factors of the first layer determine membership functions. Factors of layer 4 determine the first-order estimated function. It should be noted that ANFIS training algorithm is a hybrid algorithm in which ordinary least squares (8) algorithm is used to update coefficients of output functions ( $f_i$ ), while error back propagation algorithm is used to update fundamental factors of the system (9).

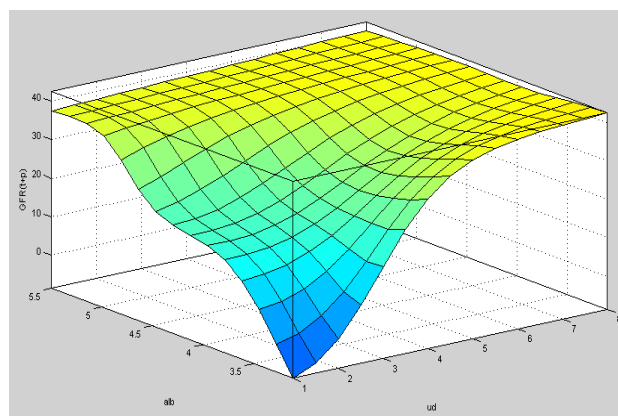
In this study, due to the continuous nature of the variables, Pearson's correlation coefficient was used. Table 1 shows the mean correlation coefficient between each input parameter and patient GFR as output for the periods of 6, 12, and 18 months.

**Table 1.** The mean correlation coefficient between input parameters and the GFR

Input name	Correlation coefficient between input and GFR
Creatinine	-0.6458
Calcium	0.1983
Phosphorus	-0.14556
Serum Albumin	0.4958
Underlying Disease	-0.3861

To implement ANFIS, inputs and outputs should be determined based on the number of lag periods ( $p$ ).  $P$  reveals the number of periods for which GFR prediction should be carried out. In the present study,  $p$  means any 6-month interval. For example, if  $p = 1$ , GFR prediction is performed for subsequent 6 months. If  $p = 3$ , GFR prediction is performed for subsequent 18 months, all data in the dataset were approached as input variables and GFR was selected as output variable.

A total of ten inputs were considered for GFR prediction. Some selected parameters did not necessarily show a good relationship with the output (GFR), so other more important inputs were selected. In addition to the removal of insignificant parameters, input number reduction increased the accuracy of prediction and better training of the model. In each step with the data as input of the period of  $t$ , GFR ( $t+p$ ) with the developed model of ANFIS was predicted. (Fig-6) Finally, underlying diseases were put into the model and summarized in the results as diabetes and non-diabetes.



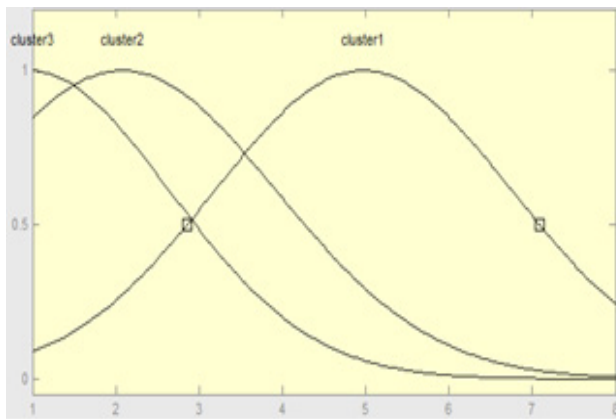
**Figure 1.** GFR function and its relationship with underlying disease and albumin variables for the 6-month period

## Graph 1- The structure of the study with regard to input and output variables

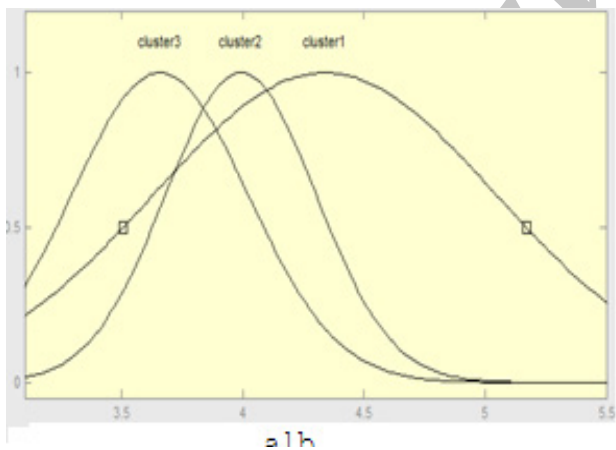
Training data were used to optimize the weights and other parameters in the model. The test data were used to evaluate the quality of estimates and forecasts. In all

further processing and modeling the test dataset have not been used for training models. Test data actually simulate the model in the case where there is no information about the future.

Test data were selected randomly so that all data shared an equal chance in the test data set. In this study, 40% of the data was used as test and the remaining as training data. Genfis 2 code in MATLAB was used to fuzzify input variables and to establish the rule base. Genfis 2 uses Fuzzy C-Means Clustering technique to fuzzify variables. The membership functions are Gaussian. According to the clinical dataset recorded from patients, fuzzy clustering of albumin and underlying variables are shown for a period of 12 months (Figures 2, 3).

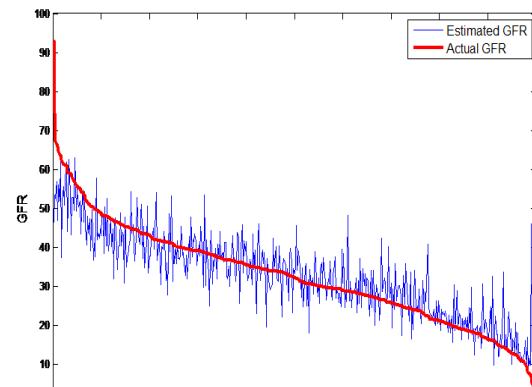


**Figure 2.** The fuzzy functions selected for the underlying disease input

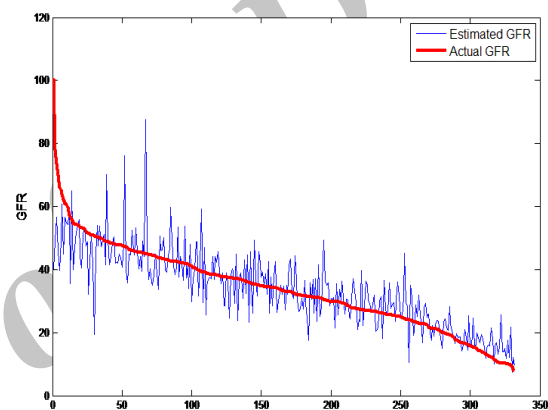


**Figure 3.** The fuzzy functions selected for albumin

During modeling, GFR was considered after a time period (6 months) with input variables in Figure 1. The graphs of function of the ANFIS model for test and training data are depicted in Figures 4 and 5. Accordingly, ANFIS is able to predict the trend of GFR changes with an acceptable accuracy.



**Figure 4.** Comparison of the estimated and real GFR for training dataset for a 6-month period



**Figure 5.** Comparison of the estimated and real GFR for test dataset for a 6-month period

Normalized Mean Square Error (NMSE) was used to reach a better conclusion regarding the error of estimation of the model and the results were used to compare the models.

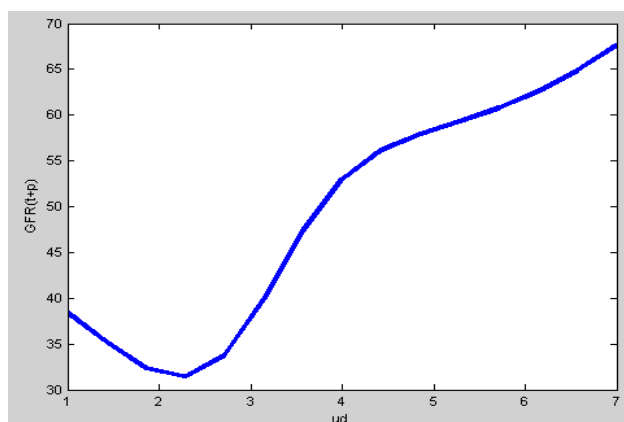
## Results

### Prediction of GFR with regard to the underlying disease and albumin in CKD patients

By changing the framework of modeling, the impact of the underlying disease on GFR prediction was studied. To consider the underlying disease in predictions, we used creatinine, calcium, and phosphorus variables. In addition, we included the variables of underlying disease (Ud) and albumin (alb). The error of the models was its minimum, (3%).

Figure-6 shows the simultaneous effects of underlying disease and albumin. We found that when the underlying disease was diabetes and serum albumin was below 4.5gr/dl, GFR was far lower in the next period.

The ANFIS model showed similar results in the control group with 245 observations. When the underlying disease was diabetes, GFR was much lower in the next period (Figure7).



**Figure 6.** The impact of underlying disease on GFR in the control group

Normalized mean square error (NMSE) (Table 2) indicated favorable results in the prediction of GFR variable. NMSE is superior because it shows the error value as a percentage in the range of 0-100.

**Table 2.** Error criterion

Error	Value
Mean Square Error (MSE)*	36.14
Mean Absolute Error (MAE)**	4.40
Normalized MSE (NMSE)***	3.10%

$$* \text{MSE} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}, \quad ** \text{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}, \quad *** \text{NMSE} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i)^2} \times 100$$

## Discussion

The most important finding of this study was the effect of diabetes as an underlying disease on the rate of renal function decrease. As this method has shown an appropriate prediction power in other diseases, it seems that it can aid in the medical management of patients with CKD, as well.

A study by Sivasankar on 2230 patients with the right iliac fossa pain showed the efficacy and accuracy of the fuzzy logic rule based classifier in the diagnosis of the severity of appendicitis (9).

Hann et al performed a retrospective study on a number of patients with prostatic carcinoma who underwent staging radical prostatectomy and pelvic lymphadenectomy, showing the superiority of the artificial neural networks to nomograms in predicting the pathologic stage (10). Other models have been evaluated in different areas, even for predicting the cause of infants' crying (11). It seems that this mathematical method can be used to describe ambiguities and enhance the precision of predictions. Using intelligent and machine learning models in this

research made earlier diagnosis and treatment of CKD possible; it can be employed as an important strategy to lower its burden. It seems that mathematical models are capable of identifying risk factors which may contribute to progression of the disease, and thus help the physicians and the caregivers.

This model makes modeling of fuzzy variables possible and therefore, predictions can be more precise and can be modeled appropriately in such a way that error is less than 3.5% in some cases.

The ANFIS model, in addition to predicting GFR, produces a fuzzy dataset which presents the complex relations of the input laboratory variables and GFR as simple linear models.

The results of our study are compatible with the findings of a doctorate thesis by Tian Min MA who predicted survival in haemodialysis patients with the use of artificial intelligent models (neural network and neuro fuzzy models), regression and algebraic formulas and it seems that the use of this model (neuro fuzzy) has a high validity (12). However, we recommend studies in shorter time periods (less than 6 months) since the analyses are more comprehensive in comparison with those conducted in shorted time periods.

## Acknowledgement

The authors would like to thank Dr. Tofigh Allahvirenloo, for his collaboration and technical assistance during the research. This project was supported financially by a grant from Tehran University of Medical Sciences.

## References

1. Mahdavi-Mazdeh M. Why do we need chronic kidney disease screening and which way to go? Iranian journal of kidney diseases. [Review]. 2010;4(4):275-81.
2. Song Q, Kasabov N. TWNFI--a transductive neuro-fuzzy inference system with weighted data normalization for personalized modeling. Neural networks : the official journal of the International Neural Network Society. [Comparative Study Research Support, Non-U.S. Gov't]. 2006;19(10):1591-6.
3. Gaspari F, Ferrari S, Stucchi N, Centemeri E, Carrara F, Pellegrino M, et al. Performance of different prediction equations for estimating renal function in kidney transplantation. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2004;4(11):1826-35.
4. Brier ME, Ray PC, Klein JB. Prediction of delayed renal allograft function using an artificial neural network. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2003;18(12):2655-9.
5. Polat K, Güneş S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing. 2007;17(4):702-10.
6. Polat K, Güneş S. Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting

- pre-processing and ANFIS. Expert Systems with Applications. 2007;33(3):636-41.
7. Sengur A. An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. Computers in biology and medicine. 2008;38(3):329-38.
  8. Jang JSR. Neuro-Fuzzy Modeling: Architectures, Analyses and Applications. Berkeley CA 94720: University of California; 1992.
  9. Jang J-SR, Sun C-T, Mizutani E. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. Automatic Control, IEEE Transactions on. 1997;42(10):1482-4.
  10. Sivasankar E, Rajesh RS. Knowledge Discovery in Medical Datasets Using a Fuzzy Logic rule based Classifier. Electronic Computer Technology Internationalconference, IEEE. 2010:208 - 13.
  11. Han M, Snow PB, Brandt JM, Partin AW. Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma. Cancer. [Evaluation Studies]. 2001;91(8 Suppl):1661-6.
  12. MA TM. Local and Personalised Modelling for Renal Medical Decision Support Systems, Auckland University. 1996 ed. Neural network design. Boston, USA: PWS Pub Co.; 2012.

Archive of SID