

## **Risk Level Prediction of Chronic Kidney Disease Using Neuro-Fuzzy and Hierarchical Clustering Algorithm (s)**

Kerina Blessmore Chimwayi, Noorie Haris, Ronnie D. Caytiles\*  
and N. Ch. S. N Iyengar\*\*

*School of Computer Science and Engineering, VIT University, Vellore, T.N., India,*

*\* Multimedia Engineering department, Hannam University, Daejeon, Korea*

*\*\*Professor, Department of Information Technology, Sreenidhi Institute of  
Science and Technology, Hyderabad, India*

*kchimwayi@gmail.com, noorienittoor@gmail.com, rdcaytiles@gmail.com,  
srinannarayanach@sreenidhi.edu.in*

### **Abstract**

*Chronic Kidney Disease (CKD) is usually characterized by a gradual loss of the functioning which the kidney does over time due to various factors. Early prediction and treatment save the kidney and halts the progress of CKD. CKD disease is being viewed as global public health issue for the past decade. The greatest threat for this deadly disease is developing countries where getting therapy is very expensive. The importance of predicting individuals who are at risk of CKD as well as applying clustering techniques cannot be underestimated since these can modify the progression of the disease. Identifying the silent killer disease early offers best opportunities for implementing possible strategies for lessening the probability of kidney loss. Neuro-fuzzy algorithm is applied to determine the risk of CKD in patients. Predictions done using neuro-fuzzy gave an accuracy of 97 percent. Using selected features, prediction for CKD disease is done so as to identify the risk. The results of the prediction are clustered to identify the percentage of patients with a high risk of having kidney disease who have a higher probability of being diabetic. Using hierarchical clustering three clusters formed show that there is a strong relationship between chronic kidney and diabetes.*

**Keywords:** *Nuero-fuzzy, CKD, Diabetes, Clustering, ANFIS, Random Forest*

### **1. Introduction**

In this century, healthcare big data analytics is inevitable due to the huge amounts of data that has variety, velocity, veracity and volume as their core characteristic. This healthcare data is being collected each second from various healthcare organizations all over the world. Utilizing big data for healthcare analytics offers great benefits to all stakeholders in healthcare which include healthcare providers, payers, patients and management. With terabytes of health data information being generated, especially through various electronic health records of huge hospitals, great insight about high risk disease predictions and much better ways of diagnosing them are acquired. One problem in the healthcare domain is the CKD that the world faces now. Huge amount of data is being generated related to CKD, from which various insights can be drawn for an effective solution through decision making among the health care sectors.

CKD is damage to kidneys due to factors related to lifestyle changes. Nowadays due to exposure to environmental changes, there are changes even observed in health which many do not recognize due to busy lifestyles. CKD can be caused due to lack of water consumption, smoking, improper diet, loss of sleep and many other factors. Researches also state that being diabetic is a highest risk which by time leads to kidney failure. CKD is unique in its nature among most diseases since it is mostly discovered when it is in the

final stages of progression whereby it will be much risky as well as expensive to treat due to being in the final stage called kidney failure.

In the domain of healthcare, this paper aims in building a model for risk level prediction in CKD considering all of the symptoms and causes contributing to it. The symptoms are the attributes that will define different stages of kidney diseases. Based on the different stages, one can classify a set of patient records to identify to which class of kidney disease a patient may belong to. On classifying patients, it results in easy recognition of the dominant attributes of CKD. Certain solutions can be provided with respect to the dominant attributes to avoid progression of CKD.

To construct a model on risk prediction of kidney disease, various machine learning techniques can be applied and then their performance can be compared with respect to accuracy, specificity and sensitivity of the models. Before application of any machine learning technique, there is a need of doing feature selection to understand the dominant attributes. Feature selection method called random forest is used to achieve selection of dominant attributes. This paper is mainly concerned with the use of machine learning techniques namely neuro-fuzzy systems and clustering which is termed unsupervised learning. These techniques will predict the class of kidney disease to which a patient can belong to, based on the symptoms. Finally, a risk analysis can be made based on all kinds of machine learning techniques applied.

## 2. Literature Survey

Dhifaf Azeez *et al.*, [1] proposed a model which does categorize patients who are in the emergency department by developing an intelligent triage system in an emergency department. Triage system is where there are a number of patients waiting for the treatment because of disease outbreaks. Triage system is integrated to the neural network system and also to the neuro-fuzzy system that can classify by capturing the signs of the patients and their appearance. This will reduce the time for examining the emergency of patients staying in the queue. Proposed work is used to examine the feasibility of ANFIS which is not explored previously in categorizing the patients.

Medical data was extracted from the Objective Primary Triage Scale (OPTS) data sheet from the Emergency Department that consists of categorical, text and continuous data. It was found from the accuracy results that ANN worked better in comparison to ANFIS for triage prediction. It has been concluded that ANN fits the output better when compared with ANFIS for new unseen data.

M. Zhamim Hossain *et al.*, [2] proposed a big data healthcare framework through the use of voice pathology assessment (VPA) as a case study. In this system, two features which include MPEG-7 low-level audio as well as the interlaced derivative pattern have been used for processing voice and speech signals. Support vector machines and a Gaussian model are used as the classifiers. This worked also includes VPA system which shows its efficiency as far as accuracy and time requirements are considered.

With the new era of the internet of things (IOT), electronic health records, genomic data, behavioral data and public health data are generating huge amounts of data which are rich with hidden patterns in which big data health analytics can be exploited to bring evidence and insights which will reduce the average cost of healthcare as well as yield improved outcomes through smarter decisions.

Dr. S. Vijayarani and Mr.S.Dhayanand [3] predicted the type of chronic kidney disease into four different types of diseases using MATLAB to implement ANN and SVM. The major aim was to compare the performance of these techniques in predicting CKD in terms of accuracy and performance. ANN networks proved to be superior to SVM machines in terms of accuracy.

Manish Kumar [4] in his work predicted the risk of chronic kidney disease by comparing numerous algorithms which were implemented using WEKA tool. The

researcher concentrated on applying various algorithms using WEKA tool then comparing these with the numerous methodologies applied. The researcher further used tenfold cross validation to classify each classifier. The only methodology that was detailed in explanation is that of Random forest for prediction yet the conclusion and analysis went on to discuss about other techniques that were not on the waiting list.

K. R. Anantha Padmanaban and G. Parthiban [5] in their work had an aim of predicting the early detection of chronic kidney disease also known as chronic renal disease for diabetic patients with the help of machine learning methods. In their research they suggested a decision tree to arrive at concrete results with desirable accuracy by measuring its performance to its specification and sensitiveness. The tool that was used with the diabetes dataset is WEKA. Naïve Bayes was compared with decision trees in predicting the risk of CKD in diabetic patients.

Adler Perotte *et al.* [6] implemented risk prediction on CKD progression with by using unrelated electronic health records accompanied with time series analysis. Here the researchers focused on using heterogeneous data sources for calculating the risk of progression from stage 3 to 5. The paper is more statistical at the same time no design was implemented for the paper.

Pushpa M. Patil [7] wrote a review paper on prediction of chronic kidney disease. From their review majority of researchers implemented the work using WEKA tool. Whilst the majority of researchers in trying to find the best solution, implemented the project using WEKA, some used MATLAB tool. Studies prove and show that for the prediction and diagnosis of this deadly CKD, these two tools were used in comparison to R which is good for data analysis and easy integration with big data technologies such as Mongo and Hadoop. In the review methodologies highlighted include Multilayer Perceptron, Logistic Regression, and Support Vector Machines (SVM) among other techniques.

Neha Sharma, Er and Rohit Kumar Verma [8] aimed at improving the prediction of kidney disease in old ages by using MATLAB to implement neuro-fuzzy algorithm which they said is better than existing probabilistic methods. In this paper they state that since this area is dealing with patient life, the use of if then rules reduces accuracy as compared to the use of mathematical models which they proposed. The researchers used the ANFIS (Adaptive Neuro-Fuzzy Inference Systems) model for implementation in MATLAB.

Abushariah, M.A.M *et al.* [9] Worked on research whereby they emphasized on designing an Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and ANFIS Approaches. The project was implemented using MATLAB with a dataset from UCI.

Abeer Y. Al-Hyari *et al.* [10] worked on a system which diagnosed renal failure, the main aim being to make it easy to identify CDK. In their research they compared three techniques which include artificial neural networks (ANN), Naïve Bayes and Decision trees using WEKA tool.

Naganna Chetty *et al.*, [11] in their research, the Role of Attributes Selection in Classification of Chronic Kidney Disease Patients, classification models with different classification algorithms which included wrapper subset and best first search method were explored in order to predict and classify CKD and non CKD patients. From their results it was evident that they obtained better accuracy on a dataset where they applied feature selection on the dataset as compared to when they did not apply any attribute selection mechanism.

A.Q. Ansari and Neeraj Kumar Gupta [12] proposed a computationally intelligent system which integrated neural networks and fuzzy logic for predicting the risk of Coronary Heart Disease (CHD). So as to automate CKD diagnosis, a simulation was done. Results of the research suggested that hybrid systems are suitable for identifying high or low cardiac risk.

From the analysis and review of work being done with respect to healthcare domain, many machine learning techniques are being implemented to predict possibility of diseases. This shortcoming will be tackled by a different approaches whereby the researchers propose to implement ANFIS for prediction of CKD and risk determination then clustering the output of the prediction to identify meaningful clusters of diabetic patients from the CDK outcome. Identification of relationships and clusters within high risk CKD patients to other chronic diseases such as sugar is therefore important. This is due to the fact that CKD is highly associated with a higher risk of complications which are related to those diseases.

### 3. Algorithms

In clustering a group of objects that belong to the same class is termed as a cluster. Similar objects are grouped in one cluster and in another cluster contains the dissimilar objects. The process of clustering is to create classes of similar objects by abstracting objects into a group. Analysis of cluster involves partitioning dataset into groups based on similarity among them followed by assigning labels to the groups formed. Clustering is adaptable to changes and helps derive useful features that differentiates groups.

A fusion of neural networks with fuzzy logic is generally defined as a system trained using a particular learning algorithm which is derived from neural network foundations. Learning in this case is done on local information whereas modifications are done locally. This approach combine artificial neural networks (ANN) with Fuzzy Rule Based Systems (FRBS). FRBS is laid upon the structure of ANN and the learning algorithm is used to adapt FRBS parameters which can be membership functions. Adaptive Neural Fuzzy Inference System (ANFIS) method for implementing neuro-fuzzy on a medical charges dataset is considered.

#### 3.1. Neuro-fuzzy

Neuro-fuzzy is a popular algorithm in the field of artificial intelligence that can be used in predictions of certain data in every domain that we work in. ANFIS works on the basis of training the neural network which is tuned by the membership function parameters on integrating with the fuzzy system. Neuro-adaptive learning techniques have been used to tune membership function parameters. In general it can be said that ANFIS is similar to that of fuzzy inference system in terms of functionality. Sugeno and Tsukamoto models are represented by ANFIS. Hybrid learning algorithm is adopted by ANFIS.

The assumption is that the FIS has got two inputs which are a and b and an output c. On considering a model called sugeno fuzzy model, it will generate the following two rules:

First Rule:

If a is  $R_1$  and b is  $S_1$  then  $g_1 = x_1 a + y_1 b + z_1$ .

Second Rule:

If a is  $R_2$  and b is  $Q_2$  then  $g_2 = x_2 a + y_2 b + z_2$ .

Each of the working of the layers in ANFIS architecture can be explained as follows:

First Layer :

- $Z_{L,j}$  which is output from the jth node of a layer named L.
- Here all the nodes are called adaptive and they have a node function associated to them

$Z_{L,j} = \mu_{P_i}(a)$  for  $j = 1, 2$ , or

$Z_{L,j} = \mu_{Q_{i-2}}(a)$  for  $j = 3, 4$

- a (or b) are the input nodes j and  $P_j$  (or  $Q_{j-2}$ ) are linguistic labels associated with this node.
- $Z_{L,j}$  is the membership function of fuzzy set ( $P_1, P_2, Q_1, Q_2$ ).

$$\text{Membership function } \mu(a) = \frac{1}{1 + \left| \frac{a-r_j}{p_j} \right|^{2q_j}} \quad (1)$$

$P_j, q_j$  and  $r_i$  are the premise parameters in this case.

Second Layer:

- Here we have all nodes that are fixed and have a label which is named p.
- Output receives a product of the input for all the coming  $s_i$ . Product of all the incoming signals are forwarded to the output.

$$Z_{2,j} = W_j = \mu P_j(a) \cdot \mu Q_j(b) \text{ for } j = 1, 2$$

- All nodes resemble and carry the firing strength of each rule.
- T-norm operator can be used like AND operator.

Third Layer:

- In this layer we have fixed nodes that are labelled as Norm.
- These node calculate the ratio of the  $j^{\text{th}}$  rule's firing strength in line with the sum of all rules' firing strengths.

$$Z_{3,j} = \bar{W}_j = \frac{W_j}{W_1 + W_2}, j = 1, 2 \quad (2)$$

- Normalized firing strengths represents the output.

Fourth Layer:

- All nodes belonging to this layer are adaptive nodes associated with a node function:

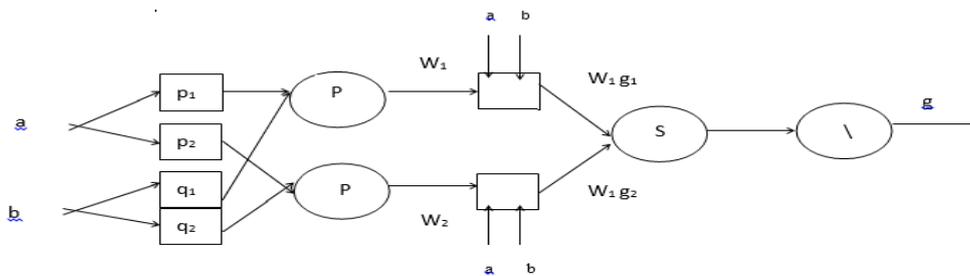
$$Z_{4,j} = W_j G_j = W_j(x_j a + y_j b + z_j)$$

- Layer 3 produces the normalized firing strength,  $\bar{W}_j$ .
- The parameter set of node is  $\{x_j, y_j, z_j\}$  and these termed as consequents.

Fifth Layer:

- Here we have a node which is single and also fixed with a S. All the incoming signals are given to output and they are given considering that they are summations of all signals.

$$\text{Output of layer 5} = Z_{5,j} = \sum_j \bar{W}_j g_j = \frac{\sum_j W_j g_j}{\sum_j W_j} \quad (3)$$



**Figure 1. Neuro-Fuzzy Architecture**

### 3.2. Hierarchical Clustering

In this type of analysis, each case of the CKD dataset starts as its own cluster. The clusters are combined on at a time till a single cluster is formed.

In this case  $Y = \{y_1, y_2, y_3, \dots, y_n\}$  are CKD data points.

- 1) The first step is a disjoint clustering having level as zero and having a sequence  $n = 0$  number.
- 2) Second step involves finding the minimum distance between pair for clusters within the current cluster, for example pair (e), (f), according to the distance  $d[(e), (f)]$ .

- (f) =  $\min d [(i), (j)]$  and in the current cluster the minimum distance is of all is considered.
- 3) In this stage sequence number is incremented as:  $n = n + 1$ . Upon finishing this step clusters (e) and (f) are merged into a solitary cluster forming the next clustering n. The level of clustering is established to  $L (n) = d [(e), (f)]$ .
  - 4) For CKD dataset the matrix for distance is updated, and it is represented by D, whereby it is done deleting columns and the rows belonging to clusters (e) and (f) and at the same time a row is added as well as the column which corresponds to the new cluster. Distance between the new cluster, denoted (e, f) and old cluster (l) is therefore given by the following equation:  $d [(l), (e, f)] = \min (d [(l), (e)], d [(l), (f)])$ .
  - 5) Repeat the steps 2 to 4 till all clusters have been merged to form one cluster of patients else stop.

### 3.3. Random Forest

Random forest (RF) is a well-known ensemble learning method which is being applied in different fields which include high-dimensional classification and pattern recognition. When using RF many single individual decision trees will be created utilizing classification as well as regression algorithms. CART as an example of rule based systems splits a node by recursive partitioning process with regard to the no and yes answer which is from predictors. The class purity is maximized at each step of rule generation within the two resulting subsets. Based on the independent rules each subset is then further split. To measure the data partition's impurity CART uses Gini index. Maximizing the difference of heterogeneity is the main aim of CART, but it usually stumbles when real world datasets are presented dealing whereby high error of prediction comes as a result of over fitting. Bagging is a technique used in RF which provide a solution by deriving classifiers for data with high dimensions in a faster way. Through voting within the ensemble, accuracy of decision for classification has to be obtained from the individual classifiers.

## 4. Problem Definition

It is noted that CKD has been a global health concern therefore there is greater need to understand the relationship between CKD and other diseases through prediction of risk and clustering the results since it will play a greater role in developing public health policies that improve outcomes. Fig 1 below showing the relationship of CKD to chronic diseases such as diabetes. Among patients with chronic diseases there is an increased risk of complication related to those diseases.

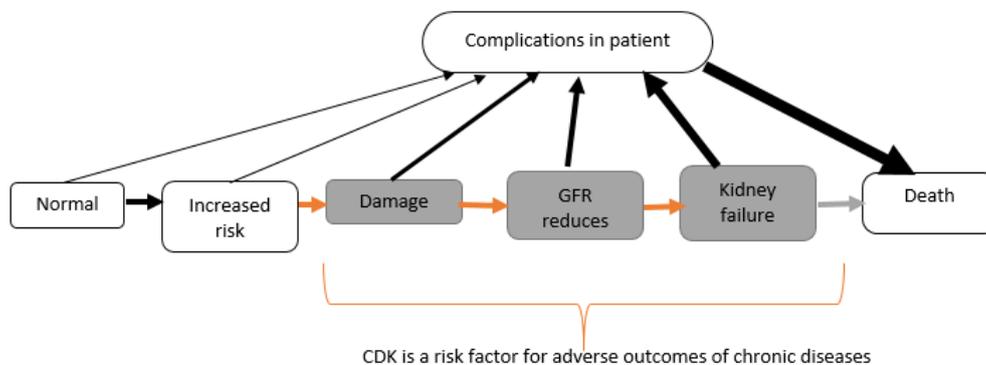


Figure 2. Relationship of CKD with Other Diseases

## 5. Design Methodology

The figure below resemble the design of the ANFIS system for the prediction, risk identification and clustering for the chronic kidney disease dataset.

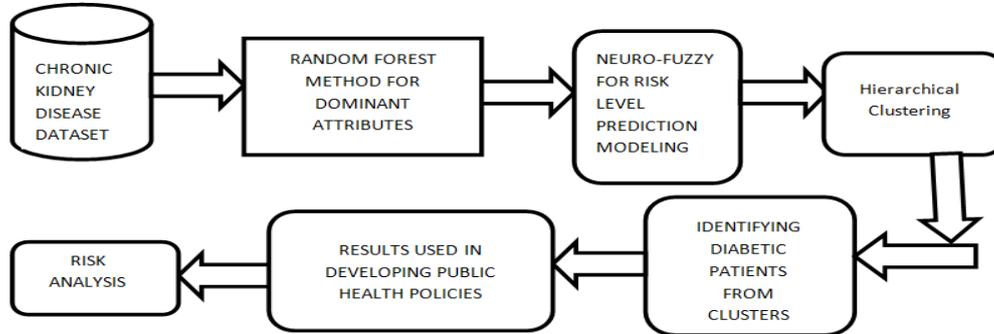


Figure 3. Methodology Design

## 6. Implementation Procedure

Using big data healthcare analytics, we can utilize huge amounts of patients' data so as to contribute in providing the right intervention to the right patients' at the most appropriate time. Proposed work consists of extracting essential attributes using random forest feature selection method. Following a thorough feature selection the data will be used for risk level prediction of kidney diseases using ANFIS. The applied techniques will predict the classification of patients based on different stages of kidney diseases with respect to the symptoms. The results from implementing ANFIS will be clustered using hierarchical clustering to identify valuable clusters.

The dataset for CKD is obtained from UCI Repository having 400 instances and 25 attributes on which the above mentioned techniques are to be applied. R tool will be used for the implementation of risk level prediction of CKD.

### 6.1. Dataset Collection

Upon definition of the problem, the appropriate dataset was chosen. CKD dataset has got a lot of useful variables which are key and necessary for the identification of the disease in patients. The dataset was initially in Attribute-Relation File Format (ARFF) which is the format for WEKA and was hence transformed into a suitable comma separated value format for use in R. Each and every feature in the Table 1 has a contribution to the class label being CKD or NOTCKD.

Table 1. Dataset Attributes

Attribute	Description
Age in years	Number of years
Blood pressure of patient	Given in mm/Hg
Specificity Gravity	Ranges from 1005 to 10025 (the higher the risk)
Albumin	Range is 0 to 5 (the higher the better)
Sugar level	5 levels indicating severity

Red Blood Cells	Is abnormal or normal
Pus Cell	Is normal or not normal (high number lead to urinary tract)
Pus Cell clumps	Can be present or not present
Bacteria	Can be present or not present
Blood Glucose	It is in mgs/dl
Blood urea	It is in mgs/dl
Serum creatinine	High level is not good
Sodium	It is measured in mEq/L
Potassium	It is measured in mEq/L
Haemoglobin	Less than 15 is kidney failure
Packed Cell Volume	This is numerical
White Blood Cell Count	This is numerical cell count
Red Blood Cell Count	Should not be higher or less than normal
Hypertension	It is categorical (yes or no)
Diabetes (mellitus)	It is categorical (yes or no)
Artery disease (Coronary)	It is categorical (yes or no)
Appetite	Is it poor or good (yes or no)
Pedal Edema	It is categorical (yes or no)
Anemia	It is categorical (yes or no)
Class	Given as ckd or notckd

## 6.2. Data Pre-Processing and Feature Selection

To obtain accurate results which are not misleading, sufficient data pre-processing should be done. Many healthcare datasets usually have missing values, noisy and inconsistent data. All these contribute to the quality of data, low quality data results in extremely poor machine learning results. The data was cleaned and missing values were replaced with column means.

The data under consideration initially had 25 attributes (14 nominal and 11 numeric) describing early stages of CKD in Indians. Of the 400 instances of the dataset, 150 are not having CKD and 250 are having CKD.

Random Forest (RF) feature selection is implemented to give a reduced feature set for model development. As shown in Figure 4 below, RF represents the importance of each feature by using mean decrease accuracy and gini. Accuracy in this case is measured in terms of the impact that each feature in the dataset is having directly on accuracy of the model and hence unimportant variable have less effect on ANFIS model. Using Mean

decrease gini on the other hand gives variables represented in way such that the most important attribute is at the top at the same time giving this as a dot on an x-axis and it uses gini impurity on each node during the growing of the trees. The first ten features are used for ANFIS model development that plots the membership functions for each attribute as in Figure 5.

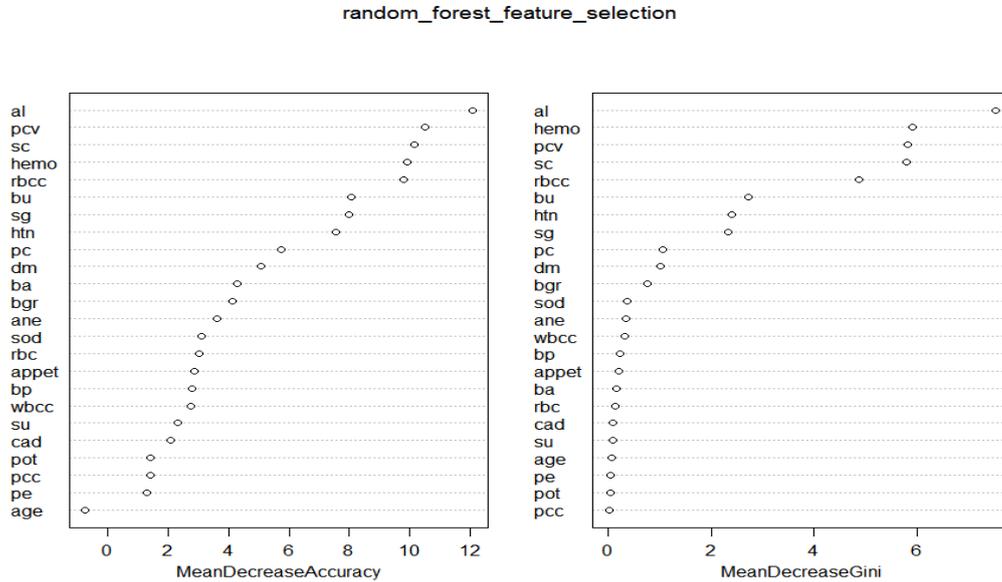


Figure 4. Random Forest Feature Selection

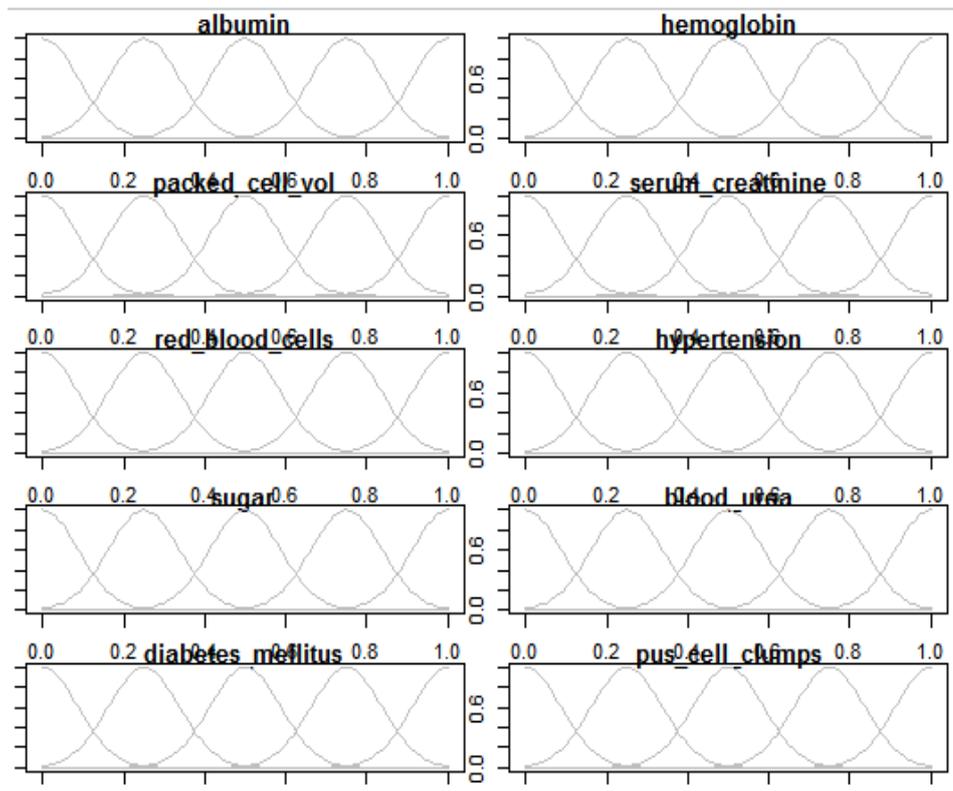
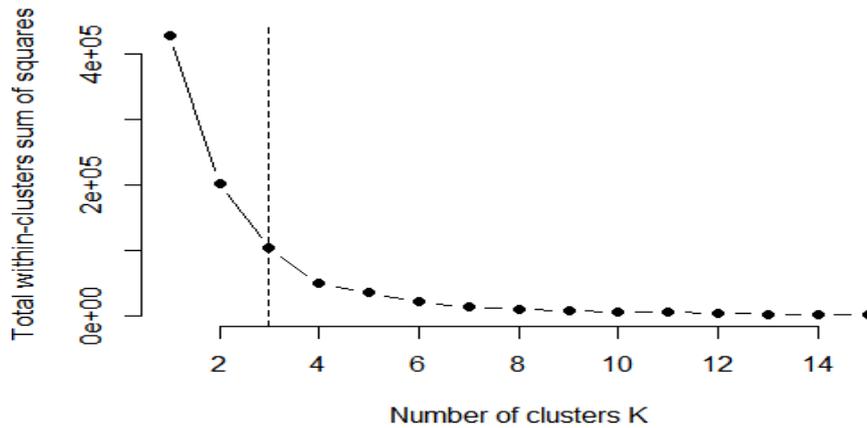


Figure 5. Plot of the Gaussian Membership Functions

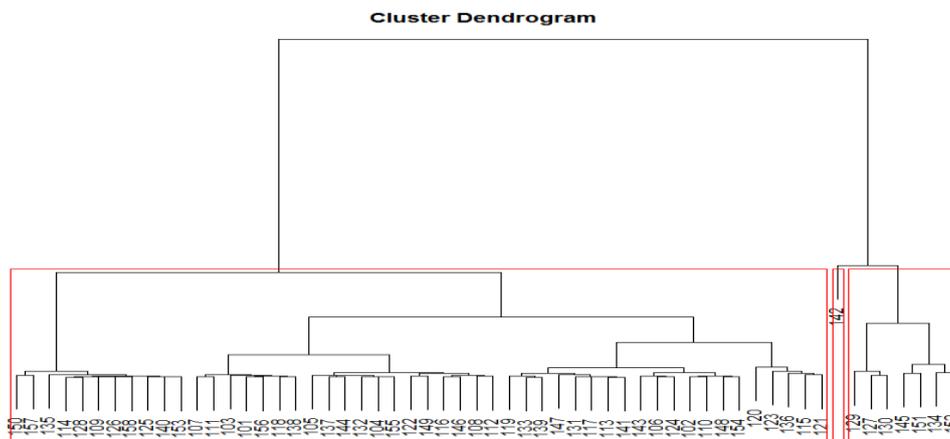
### 6.3. Hierarchical Clustering from ANFIS Results

Authors determine the number of clusters using nbclust. Using the elbow method, it is seen that the optimal number of clusters for the ANFIS predicted results is 3 as shown in Figure 7. This shows therefore that to find the maximum similarity the k of hierarchical clustering is set to 3. This will enable identification for diabetic and other chronic diseases clusters.



**Figure 6. Determining Optimal Number of Clusters**

The dendrogram below show the clusters formed by using hierarchical clustering.



**Figure 7. Dendrogram of Predicted Results**

The figure below show the cluster details.

```
> clusterCut <- cutree(clusters, 3)
> rect.hclust(clusters,k=3)
> table(clusterCut, clust$pred)

clusterCut  1  2
            1  0 50
            2  0  7
            3  1  0
```

**Figure 8. Clusters Formed as Per Class**

```
> aggregate(clust,by=list(cluster=clusterCut),median)
  cluster albumin hemoglobin packed_cell_vol serum_creatinine red_blood_cells hypertension
1      1      0      14.65           46           0.9           5.1           0
2      2      3      10.30           33           5.3           3.6           1
3      3      4       3.10            9          13.3           2.1           1
  sugar blood_urea diabetes_mellitus pus_cell_clumps pred
1  1.02     37.5           0           1     2
2  1.01    125.0           1           0     2
3  1.01    309.0           1           0     1
> |
```

**Figure 9. Analysis of Clusters Formed**

## 7. Results and discussions

Using neuro-fuzzy system, predictions were done with an overall accuracy of 97 percent. The results of the prediction showing the risk of any patient having CKD given the ten features selected from feature selection are clustered using hierarchical clustering. The 3 clusters formed identify a group of patients who have higher risk of other related chronic diseases. Cluster 1 shows that those in this cluster are similar in nature in that they have high hemoglobin, high sugar level, low blood urea, high packed cell, high red blood cell count, with pus cell clumps and without having diabetes mellitus form cluster 1 and these have CKD.

The majority in cluster 2 have sugar which is a little bit lesser than those in cluster 1. However because the sugar level is 1.01 they form cluster 2 which is also classified as having CKD. Those with lower serum creatinine also formed a cluster. It has been shown from the results of this study that a patient with hypertension, having diabetes mellitus and high sugar level there is a greater risk of kidney failure.

### 7.1. Confusion matrix

The confusion matrix for the neuro-fuzzy algorithm on application of algorithm on the dataset gave the following results of classifying the risk as indicated in Table .2.

**Table 2. Confusion Matrix**

Actual	Predicted	
	1	2
1	14	0
2	0	44

**Table 3. Accuracy Measures for ANFIS**

Accuracy	100%
Sensitivity	100%
Specificity	97%

The error rate for prediction is found to be 0. Accuracy of the neuro-fuzzy model can be summarized as in table.3.

## 8. Conclusions and Future Work

The results of this research can be added to the domain of healthcare and can be used for providing suggestions in the domain by making it easy for healthcare professionals in diagnosis and treatment of patients as well as for identifying relationships within diseases that patients have. Future work should mainly focus on implementing more big data

oriented tools and techniques which makes the process much faster and effective. The greatest challenge in healthcare domain is the data, provided enough and appropriate data is available, many applications can be implemented which take the healthcare industry to an advanced level.

## Acknowledgements

The author Dr. N. Ch. S. N. Iyengar is thankful to Dr. P. Narasimha Reddy, Executive Director, SNIST and Dr. G. Viswanathan, Chancellor, VITU for their support and encouragement.

## References

- [1] D. Azeez, M. A. Ali, K .B. Gan and I. Saiboon, "Comparison of adaptive neuro-fuzzy inference system and artificial neural networks model to categorize patients in the emergency department", Springer Plus 2013.
- [2] M. Zhamim Hossain and G. Muhammad, "Big Data Voice Pathology Assessment Framework", IEEE Access, (2016).
- [3] Dr. S. Vijayarani<sup>1</sup> and S.Dhayanand, "Kidney disease prediction using SVM and ANN algorithms", International Journal of Computing and Business Research (IJCBR) ISSN (Online), vol. 6, no. 2, (2015), pp. 2229-6166.
- [4] M. Kumar, "Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm", "International Journal of Computer Science and Mobile Computing", vol. 5, no. 2, (2016), pp. 24-33.
- [5] K. R. Anantha Padmanaban and G. Parthiban, "Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease", Indian Journal of Science and Technology, DOI: 10.17485/ijst/2016/v9i29/93880, vol. 9, no. 29.
- [6] A. Perotte<sup>1</sup>, R. Ranganath, J. S. Hirsch<sup>1</sup>, D. Blei and N. Elhadad<sup>1</sup>, "Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis", Oxford University press.
- [7] N. Sharma and Er. Rohit Kumar Verma, "Prediction of Kidney Disease by using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 9, September 2016, ISSN: 2277 128X.
- [8] M.A.M. Abushariah, A.A.M. Alqudah, O.Y. Adwan, and Yousef, "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches", Journal of Software Engineering and Applications, pp. 1055-1064.
- [9] A.Y. Al-Hyari, M. Al-Tae and M. A. Al-Tae, "Clinical Decision Support System for Diagnosis and Management of Chronic Renal Failure", IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), (2013).
- [10] N. Chetty, K.S. Vaisla and S.D. Sudarsan, "Role of Attributes Selection in Classification of Chronic Kidney Disease Patients", IEEE, (2015).
- [11] A. Q. Ansari and N. K. Gupta, "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System".
- [12] Welling Healthcare Private Limited available online at <http://homeopathyclinic.co.in/undiagnosed-chronic-kidney-disease-ckd-india/>.
- [13] J. Sun and C. K. Reddy, "Bigdata Analytics for Healthcare", Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX, (2013).
- [14] J. Jang, C. Sun and E. Mizutani, "Neuro-fuzzy and Soft Computing", Prentice Hall, (1999).
- [15] P. M. Patil, "Review on Prediction of Chronic Kidney disease Using Data Mining Techniques", International Journal of Computer Science and Mobile Computing, vol. 5, no. 5, (2016), pp. 135.

## Authors



**Kerina Blessmore Chimwayi**, she did her B. Tech in Electronic Commerce from Harare Institute of Technology, Harare, Zimbabwe and is currently pursuing M. Tech in Computer Science & Engineering (with Specialization in Big Data Analytics) at VIT University, Vellore. Her research interests include Machine Learning, Deep Learning and Spatial Data Mining in Healthcare.



**Noorie Haris**, she did her B.tech in Computer Science and Engineering from MES College of Engineering, Kuttipuram, Kerala and is currently pursuing M.tech in Computer Science and Engineering with specialization in Big Data Analytics at VIT University, Vellore. Her research interests include advanced analytics using big data technologies.



**Ronnie D. Caytiles**, he had his Bachelor of Science in Computer Engineering- Western Institute of Technology, Iloilo City, Philippines, and Master of Science in Computer Science- Central Philippine University, Iloilo City, Philippines. He finished his Ph.D. in Multimedia Engineering, Hannam University, Daejeon, Korea. Currently, he serves as an Assistant Professor at Multimedia Engineering department, Hannam University, Daejeon, Korea. His research interests include Mobile Computing, Multimedia Communication, Information Technology Security, Ubiquitous Computing, Control and Automation.



**N. Ch. S. N. Iyengar**, (b 1961), currently Professor, Information Technology, Sreenidhi Institute of Science (SNIST) and Technology, Yamnapet, Ghatkesr, Hyderabad-501301, Telangana, India He is a doctorate in both Applied Mathematics and Computer Science and Engineering. His research interests include Agent-Based Distributed Computing, Intelligent Computing, Network Security, Cloud Computing, Big Data Analytics and Fluid Mechanics. He had 32+ years of experience in teaching to B.Tech and M.Tech students. He guided 12 Ph.Ds, 5 M. Phils and 75 + M.Tech Projects apart from authoring several textbooks. He had 200+ research publications in reputed peer reviewed international journals along with students. He organized many conferences/workshops and continuing education programmes He served as Keynote speaker/ / Invited speaker //PCM/reviewer for many International conferences. He serves as an Editor in chief/Guest editor /Editorial board member for many international journals. He is the professional member of many bodies.

