E-Anfis to Diagnose the Progression of Chronic Kidney Disease

Subhashini R, Jeyakumar M K

Abstract: Chronic renal failure is not well explored. In this study, an artificial intelligence technique is proposed for overcoming the occurrence of local minima and local maxima in diagnosing the progression of kidney disease. An AI technique, a mixture of ALO and ANFIS, E-ANFIS (Enhanced Adaptive Neurofuzzy Inference Systems) is introduced. Normally back propagation is used in ANFIS, but in proposed using new optimizer ALO. The performance of ANFIS is improved by utilizing the Ant Lion Optimizer. This enhanced ANFIS used to diagnose the progression stage of the CKD. The proposed technique was executed in Matlab/Simulink platform and compared with the existing techniques ANFIS, fuzzy, and ANN. Performance evaluation is assessed in terms of accuracy, recall, precision, F-measure and specificity. The obtained results showed that the newly introduced E-ANFIS is the best algorithm when compared to other involved existing algorithms.

Index Terms: Ant Lion Optimizer, Adaptive Neurofuzzy Inference System, data mining, E-ANFIS, GFR, microalbuminuria.

I. INTRODUCTION

Chronic kidney disease is a decline in kidney function due to any type of diabetes mellitus, abnormal blood pressure, glomerulonephritis, congenital abnormalities in the kidneys, or genetic reasons [1]. CKD leads to lack in removing wastes and extra fluids from the body. The hormone level is imbalanced in the body and not able to maintain the body's balance of acid and base [3]. According to the severity of the disease, CRF is classified into five stages based on the Glomerular Filtration Rate (GFR). Stage 1 represents kidney damage with normal or increased GFR and final stage 5 referred as kidney failure wherein there is total loss of kidney function. At this stage, most of the people need dialysis [1] [2]. The problems relevant to CRF may happen gradually, over a long period of time without symptoms and they may end at end stage renal failure. Early identification and treatment are therefore useful in preventing the progression of the disease. The kidney disease progression considered as a function of various factors including GFR, urine microalbumin, serum sodium, serum potassium, serum uric acid, blood urea, total protein, serum albumin [1], [2] [4]. Among these, microalbuminuria (30-300 mg/day) is an earlier sign of chronic kidney disease [5]. In recent years, early diagnosis of the disease determines the appropriate time to apply medical treatments for CKD received great attention among physicians. Researchers through studies try to diagnose CKD in patients as early as possible and to control the risk factors of the disease progression like high blood pressure, proteinuria, and hyperphosphatemia [6], [7]. Based on the evaluation, different models were developed to predict progression. However, they cannot accurately predict the variations of GFR [8]. This paper is further organized as: Section 2 represents related work, Section 3 provides materials and methods, Section 4 is modelling of E-ANFIS technique based on ant lion optimizer, Section 5 is the result and discussion, and finally Section 6 provides the conclusion.

II. SURVEY OF RELATED RESEARCH WORKS

Aguilar et al. analysed the factors associated with CKD on 105 patients. The CKD related factors were age more than 65, sex, presence of cardiovascular disease, anemia, and overweight with BMI>30. Their work showed that age and anemia both were the strongest factors relevant with CKD [9]. There has been more number of studies on GFR variations among different CKD patients. Artificial Intelligent and machine learning techniques have been increasingly used in disease forecasting.

Gaspari et al. (2004) derived from 12 prediction equations by plasma iohexol clearance in a group of 91 renal transplant patients [10]. They found that all models overestimate renal function. Brier et al. (2003) compared neural networks with logistic regression in prediction of delayed graft function (DGF) in renal transplant patients. They evaluated the results of neural network with logistic regression and founded higher sensitivity of logistic regression in the prediction of DGF (91 versus 80%), neural network was sensitive to prediction for DGF (66 versus 47%) [11]. Hussain et al. (2011) given an tool for detecting cancer using support vector machines (SVM). They evaluated the performance of the new method with remaining classification methods. Accordingly, SVM improved in its performance [12]. Recently, fuzzy methods, especially expert systems have been increasingly used in prediction of diseases. It seems like employing this method with clinical tools for diagnosis of diseases and the condition may reduce diagnostic errors. Fuzzy inference technique is accurate. ANFIS is based on neural networks concepts. ANFIS network is proposed by Jang et al. [13]. This is a network equivalent to Takagi-Sugeno fuzzy system. Learning is continuous update of parameters. ANFIS is a hybrid algorithm in which back propagation algorithm is used to update fundamental factors [14].

Subhashini R*, Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, India.
Dr. Jeyakumar M K, Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, India.

Retrieval Number: F2304037619/19©BEIESP
Journal Website: www.ijrte.org

526

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

If we can predict the renal function worsening, we can manage this disorder. An appropriate parameter should be considered for disease worsening. The microalbuminuria is the parameter, which detects the progression of kidney disease at an earlier stage [15]. The other additional parameters considered in this paper is GFR, serum sodium, serum potassium, serum uric acid, blood urea, total protein, and serum albumin. No other efficient method proposed in the past for predicting CKD worsening time [16]. The objective is to provide a reliable method with good accuracy in healthcare system.

III. MATERIALS AND METHODS

A. Data Collection

The datas of the present study were the renal failure test records of diagnosed CKD patients from Dr. Jeyasekharan Medical Trust, Kanniyakumari during January 2014 – December 2017. The new parameter included is urine microalbumin. All the procedures were approved by the committee of Dr. Jeyasekharan Medical Trust. A total of 900 CKD patient's lab data were collected.

B. Input Selection

E-ANFIS is used in the proposed study to predict GFR values. The GFR value is calculated by MDRD equation.

All variables were used as continuous to have a good training. Seven variables were influencing parameters of GFR [17]. These variables included urine microalbumin, serum sodium, serum potassium, serum uric acid, total protein, blood urea, and serum albumin. These variables were taken as the inputs of the predicting model. The existing work excluded the urine microalbumin lab data. In this proposed study we included this as an additional attribute. The correlation between the considered variables and GFR values were calculated using Pearson correlation coefficients technique. Pearson correlation coefficients test was used to determine the most significant input variables and this was used because of the continuous nature of the variables.
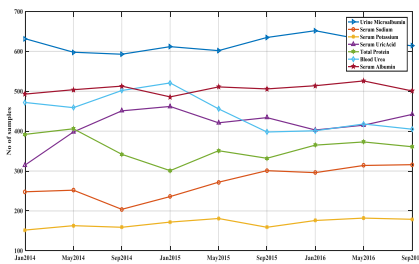

Fig. 1 GFR ranges with other attributes

Table 1: GFR and attribute variable correlation

	Urine microalbumin	Serum sodium	Serum potassium	Serum Uric acid	Total Protein	Blood Urea	Serum Albumin
Jan-14	632	248	152	315	392	472	493
May-14	598	252	163	398	406	459	504
Sep-14	593	204	159	451	342	502	513
Jan-15	612	236	172	462	301	521	486
May-15	602	272	181	421	351	456	511
Sep-15	635	301	159	434	332	398	506
Jan-16	652	296	176	403	365	401	514
May-16	629	314	182	415	373	418	526
Sep-16	624	316	179	442	361	405	501

Fig. 1 and table 1 represent the correlation coefficients between the inputs and output GFR at 4-month interval. Of the 7 inputs, microalbumin is more correlated with the output. Therefore, we considered the urine microalbumin as a new input for modelling the technique for diagnosing the progression.

In the next step, the GFR values were predicted at 4-, 8-, and 12-month intervals using E-ANFIS network model. The real data during a four-year period were collected at 4-month intervals. Therefore, the GFR values were predicted for three sequential 4-month intervals at 4-, 8-, and 12-month intervals.

C. Building Training and Test Datasets

The first step is to train all neural networks into training and test datasets. Training datas used for optimization of weights. Testing data used for quality and forecasts. The test datasets are normally selected among 25 to 35% of the original data. In this work, 30% of the data were selected for test data, remaining 70% were used as for training.

D. Fuzzification of Input Variables

Neurofuzzy classifier in MATLAB was used to fuzzify input variables and to establish the rule base.

E. Creating a Fuzzy Rule Base for E-ANFIS

The fuzzy rules are generated using the membership functions of input variables. Total 200 rules (6*8+3*8+3*8+3*8+2*8+2*8+2*8+2*8+2*8) are created in the rule base and used to estimate GFR values. Table 2 shows membership function for the considered variables in this study. Figure 2 shows the E-ANFIS architecture of the predicting model used in the proposed work.
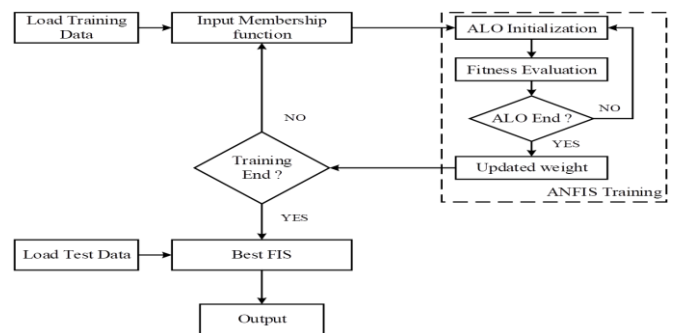

Fig. 2: Proposed Architecture of E-ANFI

IV. SECTION

Modelling of proposed E-ANFIS technique:

The structure of fuzzy inference system is a model that maps input variables to input membership functions, output variables to output membership functions, and the output membership functions to decision associated with the output.

Fuzzy Logic Algorithmic steps:

Input:

ur ac, ur ma, sr Na, sr K, tp, sr ua, sr al

Output:

Chronic kidney disease progression- Progression, Non-progression.

Begin

Input crisp value

Set fuzzy model with if-then rules.

Assign fuzzy member for the input variables.

Introduce ALO

Output variable is predicted using input variable.

Generate rules

Calculate Guassian Membership function

$$Gaussian\ (x;c,\sigma) = e^{\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2}$$

where c is function centre and σ is function width.

Rule base Construction.

Defuzzification of the fuzzy value. The Centre of Sum method for defuzzification

$$z^{\chi} = \frac{\int_z z \sum_{k=1}^{n} \mu_{c_k} z\, dz}{\int_z \sum_{k=1}^{n} \mu_{c_k} z\, dz}$$

Where n is the number of fuzzy sets, z is the number of fuzzy variables, c is the membership function.

End

Table 2: Description of the data attributes and membership function:

	Normal	Mild	Moderate	Less Severe	Severe	Failure	Membership Function
GFR /ml/	>90	60-	45-59	30-44	15-29	<15	6
Urine Micr	<30	30-	---	---	---	>300	3
Serum	135-145	--	145-1	---	---	<120,>160	3
Serum P	3.5-5.0	--	5.1-6	---	---	>7,<3.5	3
Serum	2.4-6.0	--	---	---	---	>6	2
Blood	14-23	--	---	---	---	>14-23	2
Total protei	6.0-8.0	--	---	---	---	>8.0	2
Serum al	3.4-5.4	--	---	---	---	>5.4	2

Table 2 shows the attribute details and its membership function with output variable. The Gaussian membership function is used in all the fields. The final output variable is GFR. Number of rules framed for the system is 200. The Mamdani system is used for designing the E-ANFIS.

E-ANFIS technique disallowed the local minima and local maxima for receiving an optimal solution. The ALO algorithm is used to compute the input for E-ANFIS algorithm. ANFIS is a grouping of ANN and fuzzy logic. Fuzzy logic has visions into the development of precise quantitative analysis. The parameters decide the shape of membership function. Fuzzy Logic Toolbox in Matlab gives the information about the membership function. Fuzzy modelling procedure learns information about the data set. From the given data set, the proposed algorithm E-ANFIS system is tuned using ALO optimizer. This makes the fuzzy to learn from the model. ALO is applied to adjust the occurrence of local minima and local maxima.

Identifying the progression of renal disease is helpful in maintain the graft construction for dialysis. The E-ANFIS technique is an effective optimization technique.

The trained E-ANFIS was used for estimating GFR at 4, 8, and 12 months. The GFR changes are followed by E-ANFIS. Then the GFR function was estimated for the 4-month period. The results are presented as the value of GFR at 4-month interval. This shows the relationship between input and output variables. In the same way, GFR values for 8 months and 12 months were predicted. E-ANFIS is able to predict GFR and find out the progression with an improved accuracy.

V. RESULT AND DISCUSSION

This paper introduced an innovative methodology for diagnosing the progression of the chronic renal disease. This method is developed by hybriding ALO and ANFIS. Matlab 2016a version is used for the proposed methodology. The parameters used for validation are accuracy, precision, recall, specificity, and F-Measure.

Table 3: E-ANFIS information

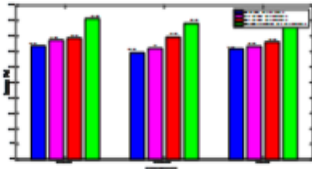| Type	:Mamdani
NumInputs	:8
InLabels :sr ma, sr Na, sr K, sr ua, tp, bl ur, ur al, gfr	
NumOutputs	:1
OutLabels	:Progression/Nonprogression
Numrules	:200
Optimizer	:ALO
DefuzzMethod	: COS

A. Accuracy

Number of correct output from all output made. Accuracy is not the be-all and end-all metric to use when selecting the best model. Accuracy is the traditional way to measure the performance of a system but equally weighs the positive and negative results, which may not be desirable in an informal retrieval system, as the number of negative results can vastly outweigh the number of positive results. Thus other parameters also considered.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} *100$$

Retrieval Number: F2304037619/19©BEIESP
Journal Website: www.ijrte.org

528

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fig. 3 Accuracy

Table 4 Accuracy

	Jan	May	Sep
ANN	73.67	69.33	71.67
Fuzzy	77.33	72	73.33
ANFIS	78.67	79.33	76.33
E-ANFIS	91.33	88.33	89.33

From Fig. 4 and Table 4 it is known that the proposed algorithm given a better accuracy. This is because existing ANFIS uses back propagation algorithm. Back propagation algorithm considers the nearby best fit and the problem is local minima, but there may be better solution at a distant point. Proposed algorithm rectifies the problem by its exploratory behaviour which help in local optima and local maxima avoidance and with its exploitation behaviour it converges rapidly towards the global minimum and global maximum.

B. Precision

It can be very precise but inaccurate, also be accurate but imprecise. Precision talks about how precise the model is out of those predicted positive, how many of them are actual positive. A false positive means the patient without the disease progression is identified as progression. This will create unwanted chaos in treating the patient. So, the proposed algorithm should be more précised.

$$\Pr ecision = \frac{TP}{TP + FP} *100$$
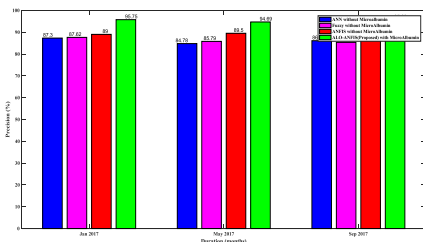
Fig. 4 Precision

Table 5 Precision

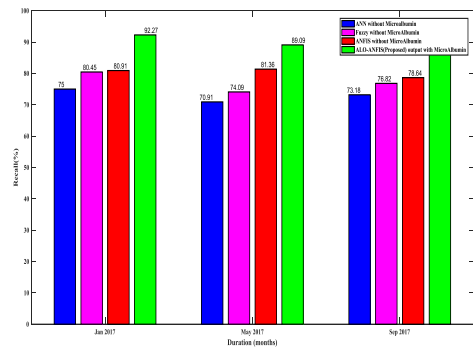|	Jan	May	Sep
ANN	87.3	84.78	86.1
Fuzzy	87.62	85.79	85.35
ANFIS	89	89.5	87.82
E-ANFIS	95.75	94.69	95.63

From Fig. 4 and Table 5, it is known that the

proposed algorithm is more précised than the existing algorithms. ANFIS requires large amount of labelled data in order to provide proper result. The newly added attribute urine microalbumin helps in taking decision, thus the proposed E-ANFIS gives more precision rate.

C. Recall

If a disease progressive patient diagnosed as non progressive, the cost associated with false negative will be high as the patient is left untreated. It is necessary to prove that the proposed algorithm improves the percentage of recall.

$$\mathrm{Re}call / Sensitivity = \frac{TP}{TP + FN} *100$$

Fig. 5 Recall
Table 6 Recall

	Jan	May	Sep
ANN	75	70.91	73.18
Fuzzy	80.45	74.09	76.82
ANFIS	80.91	81.36	78.64
E-ANFIS	92.27	89.09	89.55

Figure 5 and Table 6 shows the recall percentage. Network paralysis occurs in existing algorithms when the weights are adjusted from very low to very high and vice versa. The proposed E-ANFIS takes many repeated presentations of the input patterns and the weights are needed adjusted before the network is able to settle down into an optimal solution. This improves the percentage of recall in the proposed algorithm.

D. Specificity

In medical diagnosis, specificity is the ability of a test to correctly identify those without the disease (true negative rate). In this proposed work true negative is the number of patients who undergoes conservative care.
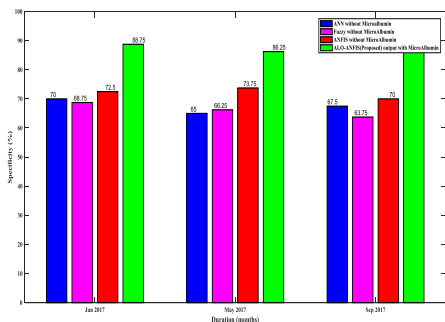
$$Specificity = \frac{TN}{TN + FP} *100$$

Fig 5. Specificity

Table 7 Specificity

	Jan	May	Sep
ANN	70	65	67.5
Fuzzy	68.75	66.25	65.75
ANFIS	72.5	73.75	70
E-ANFIS	88.75	86.25	88.75

Figure 6 and Table 7, it is proved that the proposed E-ANFIS performance better than other existing techniques. In proposed algorithm, number of fuzzy rules are more than the existing ANFIS and are automatically generated in Matlab. This help in decision making through learning from training with the set of new rules.

E. F-measure

Figure 7 and Table 8 provides the F-Measure details. F1 score might be a better measure to use if we need to seek a balance between Precision and Recall and there is an uneven class distribution (large number of actual negative). In binary classification, when dataset is imbalanced, model is not able to generalize only by giving better accuracy
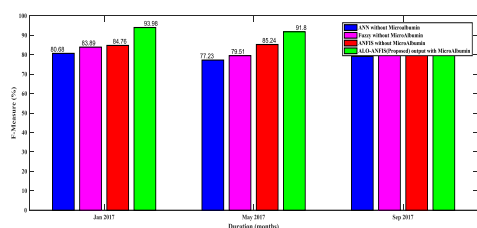

Fig. 7 F-Measure

Table 8 F-Measure

	Jan	May	Sep
ANN	80.68	77.23	79.12
Fuzzy	83.89	79.51	80.86
ANFIS	84.76	85.24	82.97
E-ANFIS	93.98	91.8	92.49

In most classification problem, it is difficult for human experts to provide sufficient information required to generate fuzzy rules. Proposed method uses learning capability of ANFIS along with ALO and appropriate membership functions. F-score judges the quality of the algorithm.

VI. CONCLUSION

E-ANFIS was developed for modelling the renal failure progression with an additional attribute urine microalbumin. The model predict the GFR for 4-, 8-, and 12-month intervals. Existing ANFIS model uses back propagation algorithm as an optimizer, which has the problem of local minima and local maxima occurrence, but the proposed E-ANFIS overcomes this problem by obtaining global minima and global maxima. Network paralysis occurs in existing ANFIS, but it is rectified in new model E-ANFIS by generating automatic rules equivalent to membership function. Existing ANFIS model is slow in convergence. This convergence problem is also solved in proposed model by adding a new labelled data urine microalbumin. The proposed algorithm proved as an acceptable method. The comparative analysis has done with ANN, fuzzy, and ANFIS. This proposed algorithm can be used in other application area also, example: stock market to analyse the movement of the stocks by considering the following inputs: 52 weeks high, 52 weeks low, day's low, day's high, open price, close price etc. In this study, the number of rules generated by E-ANFIS is more as the number of membership function of the variables are high, so the computational time may be increased. This can be avoided by considering the most influencing factors of GFR.

ABBREVIATION

CKD-Chronic Kidney Disease, GFR-Glomerular filtration rate, ALO-Ant lion optimizer, ANFIS-Adaptive neurofuzzy inference system, E-ANFIS-Enhanced adaptive neurofuzzy inference system, CRF-Chronic renal failure, BMI-Body mass index, MDRD-modification of diet in renal disease, ANN-Artificial Neural Network, ur ac- uric acid, ur ma- urine microalbumin, sr Na- serum sodium, sr K-serum potassium, tp- total protein, sr ua-serum urea, sr al-serum albumin.

REFERENCES

1. R. Subhashini, M.K. Jeyakumar, "OF-KNN Technique: An Approach for Chronic Kidney Disease Prediction" International Journal of Pure and Applied Mathematics, vol. 116, No. 24, pp. 331–348, 2017.
2. R. Subhashini, M.K. Jeyakumar, "Performance Analysis of Different Classification Techniques For the Prediction of Chronic Kidney Disease" Journal of Advanced Research in Dynamical and Control Systems, vol. 9, no 16, pp. 1515-1533. 2017.
3. John M. Arthur, Elizabeth G. Hill, "Evaluation of 32 urine biomarkers to predict the progression of acute kidney injury after cardiac surgery," HHS Public Access, vol. 85, no. 2, pp. 431–438, 2014.
4. Doreswamy, Hemanth K S, "Hybrid data mining technique for knowledge discovery from engineering materials data sets," International journal of database management system, vol. 3, no. 1, pp. 167-179, 2011.
5. Kamna Solanki, Parur Berwal, "Analysis of Application of Data Mining Techniques in Healthcare," International Journal of Computer Applications, vol. 148, no. 2, pp. 16, 2016.

Retrieval Number: F2304037619/19©BEIESP
Journal Website: www.ijrte.org

530

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

6. K. R. Anantha Padmanaban, G. Parthiban, "Applying Machine Learning Techniques for predicting the Risk of Chronic Kidney Disease," Indian Journal of Science and Technology, vol. 29, no. 9, pp. 1-5, 2016.
7. Pius Toma, K Sunil Kumarb, "Cadaver Kidney Demand Forecasting and Classification Modelling of Kidney Allocation," Procedia Technology, vol. 25, pp. 1162-1169, 2016.
8. M. Ali Yadollahpour, "Applications of Expert Systems in Management of Chronic Kidney Disease," Oriental Journal of Computer Science and Technology, vol. 7, no. 2, article 306-315, 2014.
9. Arif Khan, Shahadat Uddin, "Adapting graph theory and social network measures on healthcare data – a new framework to understand chronic disease progression," Læknabladid, ACT Australia, vol. 16, pp. 1–7, 2016.
10. Theofilos Papadopoulos, Magdalena Krochmal, "Omics databases on kidney disease," Clinical Kidney Journal, vol. 9, no. 3, pp. 343–352, 2016.
11. S. Vijayarani, S. Dhayanand, "Kidney Disease Prediction Using SVM And ANN Algorithms," International Journal of Computing and Business Research, vol. 6, no. 2, pp. 1-12, 2015.
12. KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification: "NKF KDOQI Guidelines", pp. 1-18, 2016.
13. A.A. Ojugo., A.O. Eboka., R.E. Yoro., M.O. Yerokun and F.N. Efozia, "Hybrid Model for Early Diabetes Diagnosis," Second International Conference on Mathematics and Computers in Sciences and in Industry, vol. 35, no. 10, pp. 55-66, 2015.
14. Olga LaszczySka, Milton Severo, "Electronic Medical Record-Based Predictive Model for Acute Kidney Injury in an Acute Care Hospital," European Federation for Medical Informatics, vol. 10, pp. 810-812, 2016.
15. Yutaka Hatakeyama, Hiromi Kataoka, Noriaki Nakajima, Teruaki Watabe, Yoshiyasu Okuhara, "Baseline estimation for Serum Creatinine for definition of Acute Kidney Injury," IEEE, ICIS 2016.
16. Thiyagarajan C, Ananda Kumar, "A Survey on Diabetes Mellitus Prediction using Machine Learning Techniques," International Journal of Applied Engineering Research, vol. 11, no. 3, pp. 1810–1814, 2016.
17. Anu Batra, Usha Batra, Vijendra Singh, "A review to predictive methodology to diagnose chronic kidney disease," IEEE, vol. 16, no. 2, pp. 2760–2763, 2016.

AUTHORS PROFILE

Subhashini R is a Research Scholar in the Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu, India. She has published 3 research papers related to Kidney disease and data mining.

Dr. M. K. Jeyakumar is working as Professor in the Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu, India. He has 25 years of teaching experience including 14 years of research experience in the field of Mobile Computing and Image Processing. He published more than 75 peer review research articles and one book chapter.