

An ensemble multi-model technique for predicting chronic kidney disease

Komal Kumar N¹, R. Lakshmi Tulasi², Vigneswari D³

^{1,3}Department of Information Technology, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, India

²Department of Computer Science and Engineering, RVR & JC College of Engineering, India

Article Info

Article history:

Received Feb 6, 2018

Revised Oct 8, 2018

Accepted Oct 26, 2018

Keywords:

AdaBoost

C4.5

CKD

Ensemble

Machine learning

ABSTRACT

Chronic Kidney Disease (CKD) is a type of lifelong kidney disease that leads to the gradual loss of kidney function over time; the main function of the kidney is to filter the waste in the human body. When the kidney malfunctions, the wastes accumulate in our body leading to complete failure. Machine learning algorithms can be used in prediction of the kidney disease at early stages by analyzing the symptoms. The aim of this paper is to propose an ensemble learning technique for predicting Chronic Kidney Disease (CKD). We propose a new hybrid classifier called as ABC4.5, which is ensemble learning for predicting Chronic Kidney Disease (CKD). The proposed hybrid classifier is compared with the machine learning classifiers such as Support Vector Machine (SVM), Decision Tree (DT), C4.5, Particle Swarm Optimized Multi Layer Perceptron (PSO-MLP). The proposed classifier accurately predicts the occurrences of kidney disease by analysis various medical factors. The work comprises of two stages, the first stage consists of obtaining weak decision tree classifiers from C4.5 and in the second stage, the weak classifiers are added to the weighted sum to represent the final output for improved performance of the classifier.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Komal Kumar N,

Department of Information Technology,

Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College,

Chennai, Tamil Nadu, India.

Email: komalkumarnapa@gmail.com

1. INTRODUCTION

Chronic Kidney Disease (CKD) comprises of decreased renal function from a placid to acute kidney failure. CKD mostly occurs in the elderly population and progressive loss in younger population, over 30% of the populations with 65 years have this stable disease [1]. CKD is also kindred with a high risk of Cardio Vascular Disease (CVD). The malfunction of the kidney totally deals with the glomerular filtration rate (GFR) [2]. The glomerular filtration rate represents the filtration rate at which the kidney operates to filter the wastes in the blood. The stages of the Chronic Kidney Disease are classified into five stages, the first stage is the normal kidney damage with a GFR of (>90 mL/min/1.73 m²), the second stage is the Mild reduction in GFR (60-89 mL/min/1.73 m²), the third stage contains two phases of moderate reduction, Moderate reduction in GFR (45-59 mL/min/1.73 m²), and Moderate reduction in GFR (30-44 mL/min/1.73 m²), the fourth stage is Severe reduction in GFR (15-29 mL/min/1.73 m²) and the fifth stage is the kidney failure where the filtration rate is <15 mL/min/1.73 m². In stage 1 and 2, the GFR doesn't conclude the diagnosis since the GFR are at normal levels, but these can be identified from the albumin excretion ratio, historical kidney transplantation, and abnormalities. In stages 1 to 3 there will not no signs of CKD, but when the stage crosses to 4 and 5 the body experiences significant changes, the changes are respect to the metabolic acidosis, alterations, anemia and even electrolyte imbalance, signs of the metabolic acidosis include malnutrition in

protein, weakness in muscle and loss in body mass and the alterations include Hypertension, Pulmonary edema, and Peripheral edema. Anemia is also associated with the CKD by fatigue, impaired immune function and increases CVD mortality. The CKD can be diagnosed by urinalysis, lipid profile, complete blood count, and level of serum albumin.

With the expeditious development in the medical diagnosis, ensemble classifiers can play a crucial role in predicting and diagnosing at early stages. The rest of the paper is organized into three sections, in Section 2, research methods are presented, Section 3 gives the result and the analysis which includes proposed methodology, experimental analysis and its findings, and finally the conclusion and future outlook in Section 4.

2. RESEARCH METHOD

In this section, the work related to the analysis, diagnosis, and prediction of Chronic Kidney Disease of past years is presented. Michelle M [3] found the role of variants of the APOLI gene causing various forms of kidney disease progression in African descent. Jessica [4] proposed management strategies which include in finding new potential biomarkers for monitoring the diabetes mellitus disease and also the new potential therapies for slowing down the progression of the diabetes mellitus and chronic kidney disease. The treatment of Chronic Kidney Disease associated with dyslipidemia is presented in [5], where the statins have a moderate effect on the CKD and addition of Ezetimibe to a statin may prevent CKD. Galvan [6] proposed the overview of the role of mitochondrial dysfunction in chronic kidney disease in the development of diabetic nephropathy and the analysis found the mitochondrial targets which improve the treatment of chronic kidney disease. Michelle A [7] discuss the factors of the progression of the renal dysfunction. The work in [8] estimates the predictive ability of cardiovascular disease in patients with chronic kidney disease. Two common risk scores were used in evaluating the predictability of the cardiovascular event. The results conclude that the probability of atherosclerotic cardiovascular events in patients with CKD regardless of renal function, albuminuria and previous cardiovascular events. Pontillo *et al.* [9] investigated whether CKD273, a urinary biomarker in predicting the glomerular filtration rate. In the investigation, the urinary biomarker CKD273 predicts the stage 3 CKD. The work in [10] proposed a prediction of autonomic neuropathy in chronic kidney disease in stage 5, where the work carried out in two stages, the first stage contained a set of questionnaires and followed by the testing of postural hypotension. The results concluded that the distribution of autonomic neuropathy with the questionnaire method was higher than the postural hypotension. Stack [11] proposed the measures of CKD which improves cardiovascular disease prediction, the measures concluded that both GFR and ACR are the casual cause of cardiovascular disease. Bilancio [12] proposed a predictive model for cardiovascular disease in case of kidney transplants, in the analysis and 34 variables were investigated and the confidence interval and hazard ratio were calculated, the diabetic nephropathy predicated 91.2% of cardiovascular disease. The patterns, prediction and progression of the chronic kidney disease are presented in [13], patterns such as trajectory and non-trajectory, progressions such as glomerular filtration rate leading to hypertension, anemia and mineral bone disorders. Chien [14] proposed a predictive model for the beginning of chronic kidney disease where a clinically based model is constructed by the biomedical parameters such as receiver operating characteristics curve and the optimal cutoff value. Lai *et al.* [15] aimed in clarifying the association between the infectious disease and CKD. Logistic models were used for prevalence ratio of HCV chronic infections for CKD, the results conclude that HCV viral load and HCV genotype 2 are the strong predictors of chronic kidney disease.

3. RESULTS AND ANALYSIS

We experimented our proposed ABC4.5 technique on chronic kidney disease dataset collected from a private hospital in Tamil Nadu. The dataset contained eleven numeric and fourteen nominal attributes, containing 400 observations with two classes' ckd and notckd. The detailed attribute information is presented in Table 1.

Table 1. Attribute, Class, and Value of CKD Dataset

S. No	Attribute	Value type
1	Age	Age in years
2	Blood Pressure	Numeric
3	Specific Gravity	Nominal
4	Albumin	Nominal
5	Sugar	Nominal
6	Red blood cells	Nominal
7	Pus Cell	Nominal
8	Pus Cell Clumps	Nominal
9	Bacteria	Nominal
10	Blood Glucose Random	Numeric
11	Blood Urea	Numeric
12	Serum Creatinine	Numeric
13	Sodium	Numeric
14	Potassium	Numeric
15	Hemoglobin	Numeric
16	Packed Cell Volume	Numeric
17	White Blood Cell Count	Numeric
18	Red Blood Cell Count	Numeric
19	Hypertension	Nominal
20	Diabetes Mellitus	Nominal
21	Coronary Artery Disease	Nominal
22	Appetite	Nominal
23	Pedal Edema	Nominal
24	Anemia	Nominal
25	Class	ckd or notckd

3.1. Proposed methodology

The objective of this paper is to propose an ensemble technique for Chronic Kidney Disease. The work contains two stages; in the first stage the dataset is preprocessed and subjected to C4.5 classifiers, in the second stage the output classifiers are optimized using the adaptive boosting technique. The framework is shown in Figure 1.

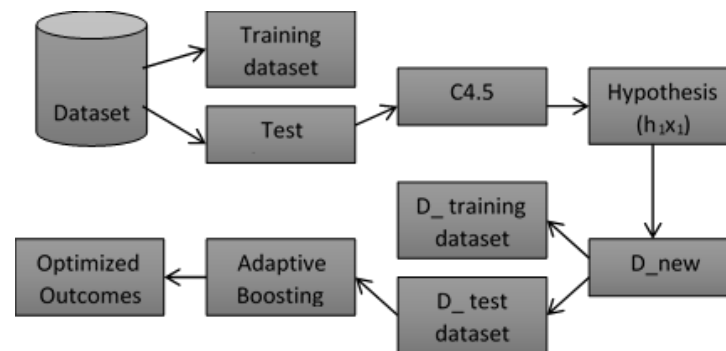


Figure 1. ABC4.5 framework

The main steps in ABC framework for Chronic Kidney Disease (CKD) are as follows:

- Step 1. In this step, the attribute values in the dataset are preprocessed to remove any unwanted and missing values.
- Step 2. The preprocessed data is subjected to 10-cross-validation and a percentage split of 60%.
- Step 3. With the percentage split, the preprocessed dataset is divided into training and test dataset, where the test dataset is subjected to C4.5 classifier; the C4.5 tree classifier produces a hypothesis.
- Step 4. The produced hypothesis forms a new dataset called as D_{new} , this new dataset is also subjected to the percentage split of 60%, the new test dataset is given as an input to the adaptive boosting where the weak learners from the C4.5 classifier are tweaked in favor of misclassified instances with their weighted sum.
- Step 5. The final outcome represents boosted classifier for Chronic Kidney Disease.

```

Algorithm 1: Hypothesis generation using C4.5
1  Input:  $Tset=s_1, s_2, \dots, s_n$  and  $Fsam= x_{1,b}, x_{2,b}, \dots, x_{p,i}$ 
2  Output:  $Fbest= (h_1x_1, h_2x_2, \dots)$ 
3  begin
4  For each  $Fsam$ 
5  Find normgain ratio on  $Fsam$ 
6   $Fsam\_best= High$  (normgain ratio)
7  end For
8   $Dnode= split$  ( $Fsam\_best$ )
9   $Fbest= recur$  ( $Dnode$ )
10 end
    
```

```

Algorithm 2: Optimized Classifier
1  Input:  $h_1x_1, h_2x_2, \dots$ 
2  Output:  $Opclass$ 
3  begin
4  Initial weight  $w=w_{1,1}, w_{2,1}, \dots$ 
5  For  $b=1 \dots B$ 
6   $Wclass= h_bx$ 
7  Choose minimizer  $\alpha_b$ 
8  end For
9   $F_b(x)=F_{b-1}(x)+ \alpha_b h_bx$  where  $b=1,2,3 \dots$ 
10  $Opclass= F_b(x)$  where  $b=1,2,3 \dots$ 
11 end
12 Update weights  $w_{i, b+1}$  for all  $i$ 
    
```

3.2. Experimental analysis and findings

This section contains the experimental analysis and findings of the proposed ABC technique for Chronic Kidney Disease (CKD). The confusion matrix containing true positive, true negative, false positive and false negative values were obtained for classifiers under analysis such as SVM, C4.5, PSO-MLP, DT and the proposed ABC4.5. These values obtained from the confusion matrix of the various classifiers were used to calculate the performance metrics such as True Positive Rate (TPR), False Positive Rate (FPR), accuracy, precision, recall, and receiver operating characteristics (ROC). Confusion matrix of the classifiers are shown in Figure 2. Performance metric comparison is shown in Figure 3. Accuracy and Inaccuracy are shown in Figure 4 and the average execution time of the classifiers under analysis is shown in Figure 5 respectively.

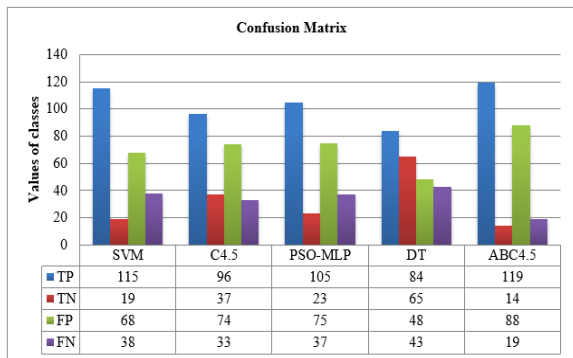


Figure 2. Confusion matrix of classifiers

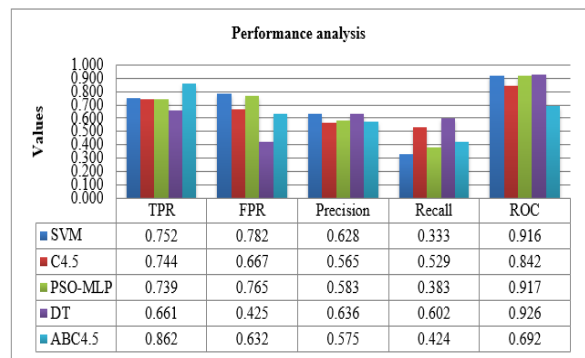


Figure 3. TPR, FPR, precision, recall, and ROC

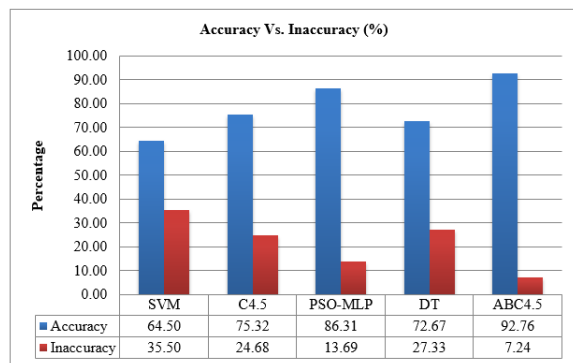


Figure 4. Accuracy vs. inaccuracy of the classifiers

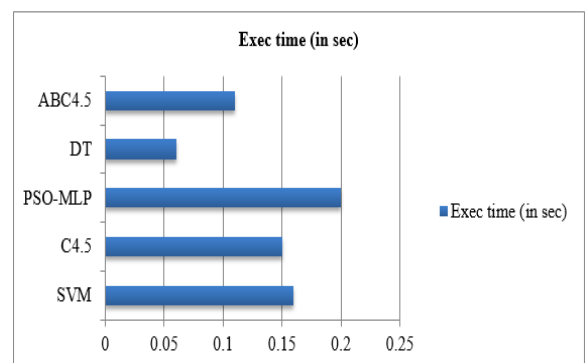


Figure 5. Average execution time of the classifiers

In the analysis of predicting chronic kidney disease, Support Vector Machine (SVM) achieved an accuracy of 65% which is less than the C4.5 classifier with 75% in prediction. Particle Swarm Optimization with Multi-Layer Perceptron [16] achieved an accuracy of 86% which is quiet less than our proposed

ABC4.5 93%. Decision Tress managed to predict with an accuracy of 72%. Proposed hybrid classifier took only 0.2s in prediction which is less than the PSO-MLP; hence the researchers can save time in prediction. The accuracy achieved by our proposed ABC4.5 is compared with the two hybrid classifiers such as PSO-MLP [16] and SVM [14], the results obtained for predicting chronic kidney disease is found to be astonishing.

4. CONCLUSION

In this analysis, SVM, C4.5, PSO-MLP, DT, and ABC4.5 classifiers were implemented on a Chronic Kidney Disease (CKD) dataset. The proposed ABC4.5 achieved an accuracy of 92.76%, which is higher than the classifiers with an execution time of 0.12 sec in detecting Chronic Kidney Disease (CKD). Applying hybrid ensemble learning algorithms on a Chronic Kidney Disease dataset for improved performance can do further outlook.

REFERENCES

- [1] O'Hare AM., *et al*, "Age Effects Outcomes in Chronic Kidney Disease," *J Am SOC Nephrol*, vol. 18 (10), pp. 2758-65. Oct 2007.
- [2] "Kidney Disease: Improving Global Outcome (KDIGO) CKD Workgroup. KDIGO 2012 Clinical Practice and Guidelines for the Evaluation and Management of Chronic Kidney Disease," *Kidney Int Suppl.*, vol. 3, pp. 1-150, 2013.
- [3] Michelle M. Estrella, Rulan S. Parekh, "The Expanding Role of APOL1 Risk in Chronic Kidney Disease and Cardiovascular Disease," *Seminars in Nephrology*, vol. 37, issue 6, pp. 520-529, 2017.
- [4] Jessica K. Yakush Williams, "Management Strategies for Patients with Diabetic Kidney Disease and Chronic Kidney Disease in Diabetes," *Nursing Clinics of North America*, vol. 52, issue 4, pp. 575-587, 2017.
- [5] Dragana Lovre, Sulay Shah, Aanu Sihota, Vivian A. Fonseca, "Managing Diabetes and Cardiovascular Risk in Chronic Kidney Disease Patients," *Endocrinology and Metabolism Clinics of North America*, 2017.
- [6] Daniel L. Galvan, Nathanael H. Green, Farhad R. Danesh, "The Hallmarks of Mitochondrial Dysfunction in Chronic Kidney Disease," *Kidney International*, vol. 92, issue 5, pp. 1051-1057, 2017.
- [7] Michelle A. Hladunewich, "Chronic Kidney Disease, and Pregnancy," *In Seminars in Nephrology*, vol. 37, issue 4, pp. 337-346, 2017.
- [8] Santiago Cedeño Mora, Marian Goicoechea, Esther Torres, "Cardiovascular Risk Prediction in Chronic Kidney Disease Patients," *Nefrología (English Edition)*, vol. 37, issue 3, pp. 293-300, 2017.
- [9] Claudia Pontillo, Zhen-Yu Zhang, Joost P. Schanstra, "Prediction of Chronic Kidney Disease Stage 3 by CKD273, a Urinary Proteomic Biomarker," *Kidney International Reports*, vol. 2, issue 6, pp. 1066-1075, 2017.
- [10] Ali Abdulmajid dyab Allawi, "Prediction of Autonomic Neuropathy in Chronic Kidney Disease (stage 5) Iraqi Patients (Case-control Study)," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 11, pp. S969-S973, 2017
- [11] Austin G Stack, "Common Measures of Chronic Kidney Disease Improve Vascular Risk Prediction," *The Lancet Diabetes & Endocrinology*, vol. 5, issue 9, pp. 675-677, 2017.
- [12] G. Bilancio, M. Celano, V. Cozza, F. Zingone, G. Palladino, M. Cirillo, "Early Prediction of Cardiovascular Disease in Kidney Transplant Recipients," *Transplantation Proceedings*, vol. 49, issue 9, pp. 2092-2098, 2017.
- [13] David Collister, Thomas Ferguson, Paul Komenda, Navdeep Tangri, "The Patterns, Risk Factors, and Prediction of Progression in Chronic Kidney Disease: A Narrative Review," *Seminars in Nephrology*, vol. 36, issue 4, pp. 273-282, 2016.
- [14] Kuo-Liong Chien, Hung-Ju Lin, Bai-Chin Lee, "A Prediction Model for the Risk of Incident Chronic Kidney Disease," *The American Journal of Medicine*, vol. 123, issue 9, pp. 836-846.e2, 2010.
- [15] Tai-Shuan Lai, Mei-Hsuan Lee, Hwai-I. Yang, "High Hepatitis C Viral Load and Genotype 2 are Strong Predictors of Chronic Kidney Disease," *Kidney International*, vol. 92, issue 3, pp. 703-709, 2017.
- [16] Komal Kumar.N, Vigneswari.D, B A S Roopa Devi, "MSO-MLP Diagnostic Approach for Detecting DENV Serotypes," *International Journal of Pure and Applied Mathematics*, vol. 118 (5), pp. 1-6, 2018.

BIOGRAPHIES OF AUTHORS



Komal Kumar N is currently working as an Assistant Professor in the Department of Information Technology, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamil Nadu, India. His research interest includes machine learning, data mining, and cloud computing.



Dr. R. Lakshmi Tulasi is currently working as an Associate Professor in the Department of Computer Science and Engineering, RVR & JC College of Engineering, Guntur, Andhra Pradesh, India. Her research interests include Information Retrieval, Semantic web.



Vigneswari D is currently working as an Assistant Professor in the Department of Information Technology, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamil Nadu, India. Her research interests include Data Mining, Machine Learning.