

An Ensemble Multi-Model Technique for Predicting Chronic Kidney Disease

N. Komal Kumar¹, R. Subhashree¹, S. Swetha¹

¹Department of Information Technology,

Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India..

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 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.

Keywords:CKD, C4.5, AdaBoost, ensemble, machine learning.

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 six sections, In section II the related work is presented, Section III gives the description of the dataset and its attributes, the proposed framework is presented in section IV, Experimental analysis and findings are presented in section V and finally the conclusion and future outlook in section VI.

2. RELATED WORKS

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 K [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. Daniel L. 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. Claudia 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. Austin G 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. G. 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 a glomerular filtration rate leading to hypertension, anemia and mineral bore disorders. Kuo-Liong 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. Tai-Shuan 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. DATASET DESCRIPTION

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

No	tribute	lue type
1	Age	Age in years
2	Blood Pressure	numeric
3	Specific Gravity	nominal
4	Hemoglobin	nominal
5	Hematuria	nominal
6	Red blood cells	nominal
7	White Cell	nominal
8	White 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	Peripheral Edema	nominal
24	Anemia	nominal
25	Class	ckd or notckd

4. 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.

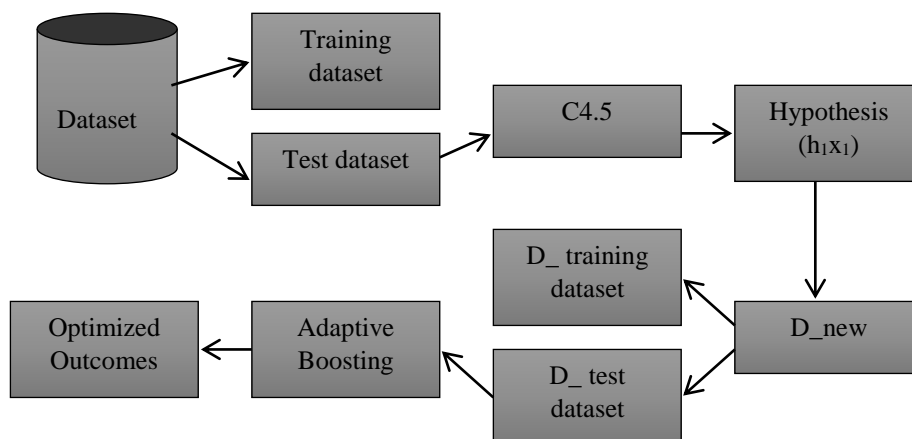


Figure 1. ABC4.5 framework

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

```

Input: Tset = s1, s2, ..., sn and Fsam = x1, x2, ..., xp, i
Output: Fbest = (h1x1, h2x2, ...)
begin
  for each Fsam
    find normgain ratio on Fsam
  Fsam_best = High(normgain ratio)
  do For
    Dnode = split(Fsam_best)
    Fbest = recur(Dnode)
  end

```

Algorithm 2: Optimized Classifier

```

Input: h1x1, h2x2, ...
Output: Opclass
begin
  initial weight w = w1,1, w2,1, ...
  b = 1 ... B
  Opclass = hbx
  choose minimizer αb
  do For
    Fb(x) = Fb-1(x) + αb hbx where b = 1, 2, 3, ...
  Opclass = FB(x) where b = 1, 2, 3, ...
  end
  update weights wi,b+1 for all i

```

5. 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.

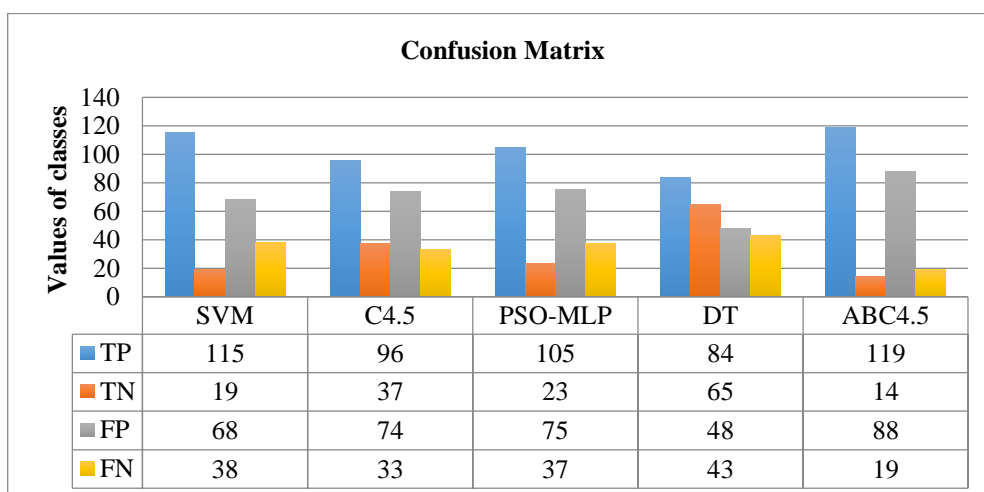


Figure 2. Confusion Matrix of Classifiers

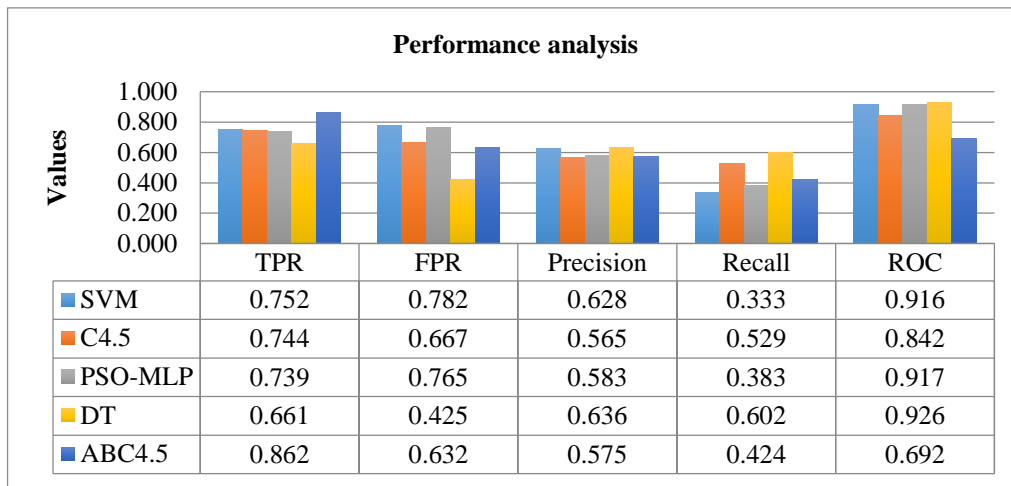


Figure 3. TPR, FPR, Precision, Recall, and ROC

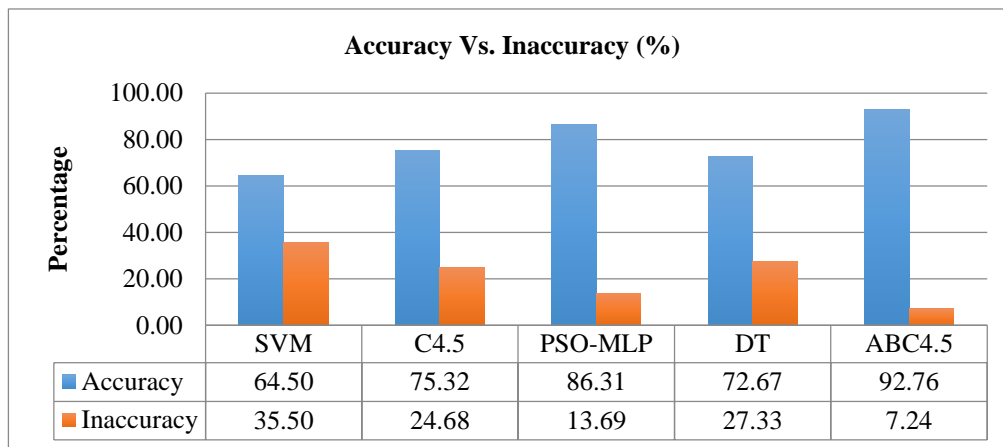


Figure 4. Accuracy Vs. Inaccuracy of the classifiers

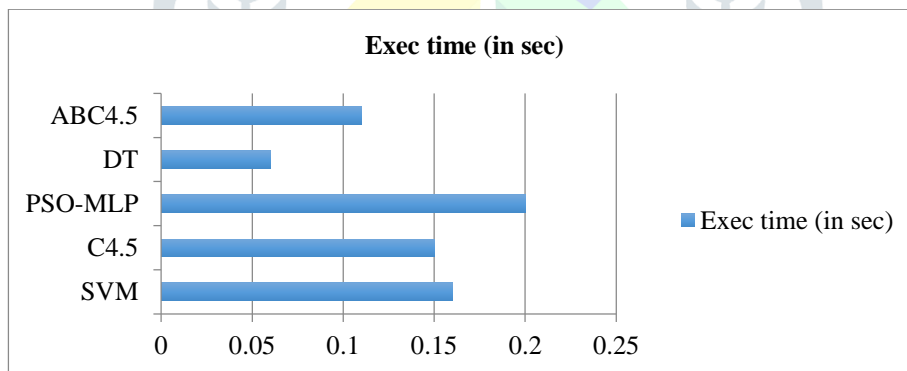


Figure 5. Average execution time of the classifiers

6. CONCLUSION AND FUTURE OUTLOOK

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."* 18(10):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.* 2013. 3:1-150.
- [3] Michelle M. Estrella, Rulan S. Parekh, "The Expanding Role of APOL1 Risk in Chronic Kidney Disease and Cardiovascular Disease", *"Seminars in Nephrology"*, Volume 37, Issue 6, Pages 520-529, ISSN 0270-9295, 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"*, Volume 52, Issue 4, Pages 575-587, ISSN 0029-6465, ISBN9780323552844,2017.
- [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"*, ISSN 0889-8529, 2017.
- [6] Daniel L. Galvan, Nathanael H. Green, Farhad R. Danesh, "The hallmarks of mitochondrial dysfunction in chronic kidney disease", *"Kidney International"*, Volume 92, Issue 5, Pages 1051-1057, ISSN 0085-2538, 2017.
- [7] Michelle A. Hladunewich, "Chronic Kidney Disease, and Pregnancy", *"In Seminars in Nephrology"*, Volume 37, Issue 4, Pages 337-346, ISSN 0270-9295, 2017.
- [8] Santiago Cedeño Mora, Marian Goicoechea, Esther Torres, "Cardiovascular risk prediction in chronic kidney disease patients", *"Nefrología" (English Edition)*, Volume 37, Issue 3, Pages 293-300, ISSN2013-2514, 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"*, Volume 2, Issue 6, Pages 1066-1075, ISSN 2468-0249, 2017.
- [10] Ali AbdulmajidyyabAllawi, "Prediction of autonomic neuropathy in chronic kidney disease (stage 5) Iraqi patients (case-control study)", *"Diabetes & Metabolic Syndrome: Clinical Research & Reviews"*, Volume 11, Supplement 2, Pages S969-S973, ISSN 1871-4021, 2017
- [11] Austin G Stack, "Common measures of chronic kidney disease improve vascular risk prediction", *"The Lancet Diabetes & Endocrinology"*, Volume 5, Issue 9, Pages 675-677, ISSN 2213-8587, 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"*, Volume 49, Issue 9, Pages 2092-2098, ISSN 0041-1345, 2017.
- [13] David Collister, Thomas Ferguson, Paul Komenda, NavdeepTangri, "The Patterns, Risk Factors, and Prediction of Progression in Chronic Kidney Disease: A Narrative Review", *"Seminars in Nephrology"*, Volume 36, Issue 4, Pages 273-282, ISSN 0270-9295, 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"*, Volume 123, Issue 9 Pages 836-846.e2, ISSN 0002-9343, 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"*, Volume 92, Issue 3,Pages 703-709, ISSN 0085-2538, 2017.