Early Detection of Cardiovascular Disease in Patients with Chronic Kidney Disease using Data Mining Techniques

Avijit Kumar Chaudhuri[1], Arkadip Ray[2], Anirban Das[3], Prasun Chakrabarti[4], Dilip K. Banerjee[5]

[1]Department of Computer Science and Engineering, Techno Engineering College Banipur, Habra, India, [2]Department of Information Technology, Government College of Engineering & Ceramic Technology, Kolkata, India, [3]Department of Computer Science & Engineering, University of Engineering & Management, Kolkata, India, [4]Institute Distinguished Senior Chair Professor, Techno India NJR Institute of Technology, Udaipur- 313003, Rajasthan, India, [5]University Research Professor, Seacom Skills University, India Kendradangal, Bolpur, District: Birbhum, West Bengal, India, PIN: 731236

[1]c.avijit@gmail.com, [2]arka1dip2ray3@gmail.com, [3]anirban-das@live.com, [4]drprasun.cse@gmail.com, [5]dkbanrg@gmail.com

Abstract— **A constant obstacle for doctors is the high prevalence of cardiovascular disease (CVD) in patients with chronic kidney disease (CKD). Increasing efforts have been made to jointly treat patients with heart and kidney disease, as shown by an increasing number of basic research and clinical investigations concerning CVD in CKD. Typical risk factors for CVD are common in CKD, such as age, blood pressure (bp), hypertension (htn), and blood sugar (sg). Standard risk factors tend to be the major contributors to CVD in patients with mild to moderate CKD. However, in patients with advanced CKD, non-traditional CKD-specific risk factors (e.g. Potassium level in blood) are more prevalent than in the general population, contributing, in addition to traditional risk factors, to the high burden of CVD in CKD. However, in patients with CKD, CVD often remains underdiagnosed and undertreated. Nevertheless, CVD still remains under control and care in patients with CKD. Researchers in this paper aims to predict the probability of CVD from CKD by using various popular data mining techniques and definitively propose a decision tree and by using Random Forest analysis to test its specificity and sensitivity to achieve concrete results with sufficient precision.**

Keywords*— *Chronic Kidney Disease (CKD), Cardio Vascular Disease (CVD), Glomerular Filtration Rate (GFR), Decision Trees (DT), Logistic Regression (LR), Random Forest (RF).

1. Introduction

Long-term illness is known to be a chronic disease. By the definition of the U.S. National Center for Health Statistics, chronic disease has a minimum period of three months or longer. Chronic diseases in global mortality and morbidity are now becoming a major factor. So far, only developed countries are struggling with chronic disease-related health issues. At present, 4 out of 5 chronic diseases demise occurs in low and middle income-oriented countries. In India, chronic disease deaths reported in 1990 amounted to 3.78 million (40.4% of all deaths). This figure is expected to grow to 7.63 million in 2020 (66.7% of all deaths) [1]. Historically, earlier chronic disease prevention programs focused primarily on obesity, diabetes mellitus, and cardiovascular disease (CVD), but increased incidence of chronic kidney disease (CKD) progressed to end-stage renal disease (ESRD) and subsequently, early and advanced renal replacement therapy (RRT) and economic responsibility [2].

Data mining is mainly used to classify and forecast diseases in health care surveillance. There are various data mining techniques available to identify and forecast diseases. The aim of using data mining techniques is to undertake extensive research in support of an intelligent health-monitoring program to build a new Novel Ensemble Health Care Decision Support System [3].

Data mining [4] is a constructive method that can be practically used in service to recover unfamiliar knowledge from a huge dataset. Data Mining has three key aspects, Classification or Clustering, Sequence Analysis, and Rules of Association. Anyone or collection of the above categories can be used based on the purposes. Classification or clustering examines data collection and conducts a set of classification rules that can be used to identify future data.

Most of the classical machine learning algorithms for classifying datasets can be performed very competently for uniformly distributed datasets. However, these classical algorithms show less or poor learning performance at the time of classification for uneven and scattered data sets that have variation in the class labels. To obtain the accuracy and certainty of classification algorithms, we use the combined collection of algorithms to achieve reasonable accuracy. This method of merging various types of algorithms is commonly referred to as an ensemble [5]. The real motivation of predictions in medical and health care data mining is to uncover patterns of patient data for earlier disease identification and to promote health management [6]. The increasing volume of derivable information of details related to patient health and medical history provides a gold mine that can be used to recognize the condition of different parts of the human body. Data mining's main influences are to create a health care informatics (HCI) research scope [7].

The information discovered can be pragmatic in perceiving how the body of the patient is responding with several reports of medical tests. Most recent health and medical care studies have anticipated many diseases to be predicted earlier. For example, hepatic cancer prognosis, cardiovascular disease prediction [8], breast cancer recurrence identification [9], dermatological disease prediction, diabetes prediction [10], prediction of hepatitis C virus (HCV), and many more. CKD has now turned out to be a dangerous cause of death in recent days because of the drastic change in citizens' daily lifestyle. In health and medical science, kidney disease is an imperceptible and enormous problem. A large number of patients with such diseases require special attention and care. In such cases, medication and treatment are also non-trivial. Therefore, their prompt and accurate prediction is the ultimate requirement that can be beneficial for patients to recover from the adversity of such diseases. Our current research work meets the objective to estimate the probability of assimilating kidney disease from a given set of patient data. However, CVD is often diagnosed and treated in patients with CKD. For clinicians, they must recognize that CKD patients are a group at excessive risk for CVD and cardiovascular events growth. Moreover, it is found that potassium is one of the most important factors for CKD from the DT analysis of the CKD dataset with all variables.

2. CKD and Heart Disease

Patients with CKD are found to have high CVD potential [11]. Brilliant first detailed the relationship between CKD and CVD in [12]. Deficiency in renal function can increase the risk of CVD from two to four times [13]. CKD is known to be present when at least three months apart in two events the disabled function of the kidney is established [14]. The approximate glomerular filtration rate (eGFR) can be determined using serum creatinine and the state of collaboration in the epidemiology of chronic kidney disease (CKD-EPI) [15]. The measurement of proteinuria is based on the ratio between urinary albumin and creatinine [15]. CKD is organized in 5 levels of Glomerular Filtration Rate (GFR) and three phases of proteinuria [14]. Similar work has shown that a higher occurrence of CVD is associated with low eGFR and increased albuminuria. Among patients with stage 3 CKD, cardiovascular mortality among patients with stage 4 CKD was double higher and triple higher than in healthy patients with renal function [16][17]. Double rates increase in patients with eGFR the risk of congestive heart failure (CHF), atrial fibrillation, stroke, coronary artery disease (CAD), and peripheral artery disease (PAD). Two large cohort studies showed a significantly reduced lifespan for patients with stage 3B CKD (17-year shorter survival) and stage 4 CKD (25-year shorter survival) compared to normal kidney function subjects [18]. Patients with CKD and CVD have a higher mortality rate (58-71%) compared with patients with CVD and normal renal function (22-27.5%) [19].

For instance, compared with traditional cardiovascular risk factors such as diabetes mellitus and hypertension, the effect of CKD on CVD risk appears to be higher as the observed decrease in life expectancy for middle-aged patients with diabetes mellitus and hypertension is approximately 8 and 3 years respectively ([20]-[23]). Hypertension is a crucial factor in the susceptibility to CKD enhancement. Hypertension in subjects with CKD has been shown in [24] to cause more CVD than in patients with normal kidney function [24]. The prevalence of left ventricular hypertrophy (LVH) in patients with CKD is increased, particularly in patients with eGFR.

There is a significantly increased risk of CVD and cardiovascular mortality associated with CKD. In particular, studies of CKD patients demonstrated an elevated relative risk of coronary heart disease (CHD), heart failure (HF), and stroke compared to those without CKD. Direct comparisons in CVD risk disparities in patients with and without CKD, however, have not been well established, which may direct the prioritization of specific therapies to improve the poor prognosis in this high-risk population. Authors use various machine learning methods to tackle such knowledge gaps. The authors found age and hypertension as the highest score variables in both datasets from the study of several machine learning techniques and their ensemble in CKD and CVD datasets and can recognize that CKD patients are a high-risk group for developing CVD and cardiovascular events.

3. Significance of Potassium in Heart Failure

Hyperkalemia, typically described as blood potassium exceeding 5.0 mmol / L, maybe a routine clinical exercise problem in patients with heart failure (HF) and maybe a severe condition [25][26]. Hyperkalemia with high potassium levels (e.g., greater than 6.0 mmol/L) may cause cardiac arrhythmias and death [27], but even potassium levels greater than 5.0 mmol/L are associated with increased mortality in patients admitted to the acute care hospital and patients with heart failure [28][29]. There is little knowledge of the incidence of hyperkalemia in real-world HF patients and related outcomes [30][31].

Renal dysfunction patients have an increased risk of hyperkalemia [26] and a complex association of cardio-renal syndrome between heart and kidney dysfunction [32] with proof of CKD in over half of HF [33] patients. Additionally, some medicines commonly used to treat HF may contribute to hyperkalemia [34] by interfering with kidney potassium excretion, as well as angiotensin-converting enzyme inhibitors (ACEIs), angiotensin receptor blockers (ARBs), and potassium-sparing diuretics such as spironolactone and eplerenone [30][35]. To understand the potential effects of recent drug therapies for hyperkalemia [27][36], it is important to assess the actual burden of hyperkalemia among HF patients and evaluate patient characteristics, current treatment procedures, and clinical outcomes in real-world settings.

4. Methodology

The CKD dataset used in this research was obtained from the UCI machine learning dataset. The dataset contains pathological data of 400 people from the Southern part of India. There is a total of 24 features present in the dataset, most of which are physiological and clinical. Table 1 lists various parameters and their data types. The missing values of all attributes have been replaced by the arithmetic mean of the numerical and discrete integer values of all instances in the preprocessing stage of data.

Table 1. Description of CKD Dataset

Sl. No.	Attribute	Descriptions	Values
1	age	Age of the patient when diagnosed (numerical)	Years
2	bp	Blood Pressure (numerical)	mm/Hg
3	sg	Specific Gravity (nominal)	1.005,1.010,1.015,1.020,1.025
4	al	Albumin (nominal)	0,1,2,3,4,5
5	su	Sugar (nominal)	0,1,2,3,4,5
6	rbc	Red Blood Cells (nominal)	normal, abnormal
7	pc	Pus Cell (nominal)	normal, abnormal
8	pcc	Pus Cell clumps (nominal)	present, not present
9	ba	Bacteria (nominal)	present, not present
10	bgr	Blood Glucose Random (numerical)	in mgs/dl
11	bu	Blood Urea (numerical)	in mgs/dl
12	sc	Serum Creatinine (numerical)	in mgs/dl
13	sod	Sodium (numerical)	in mEq/L
14	pot	Potassium (numerical)	in mEq/L
15	hemo	Hemoglobin (numerical)	in gms
16	pcv	Packed Cell Volume (numerical)	nominal
17	wc	White Blood Cell Count (numerical)	in cells/cmm
18	rc	Red Blood Cell Count (numerical)	millions/cmm
19	htn	Hypertension (nominal)	yes, no
20	dm	Diabetes Mellitus (nominal)	yes, nao
21	cad	Coronary Artery Disease (nominal)	yes, no
22	appet	Appetite (nominal)	good, poor
23	pe	Pedal Edema (nominal)	yes, no
24	ane	Anemia (nominal)	yes, no
25	class	Class (nominal)	ckd, notckd

The CVD dataset is collected from https:/www.kaggle.com/amanajmera1/framingham-study-dataset. Information is linked to hospital patients. The subjects were randomly selected in this CVD study as a sample of 4240 patients who went for medical examinations. Table 2 displays the biometric data obtained during the physical examination of the following factors.

Table 2. Description of CVD Dataset

Sl. No.	Attributes	Description	Range of Values	Mean	Standard Deviation
1	Age	Age at exam time in years	Continuous	49.5801	8.5729
2	Male	Male or Female	0 = Female; 1 = Male		
3	Education	Education of the patient	1 = Some High School; 2 = High School or GED; 3 = Some College or Vocational School; 4 = college		
4	currentSmoker	At present smoker or not	Value 0 for no smoking; value 1 for smoking		
5	cigsPerDay	Smoking habits - Average no. of cigarettes/day	Continuous	9.0059	11.9225
6	BPMeds	Blood Pressure medications	Value 0 for not taking any Blood Pressure medications; value 1 for already in Blood Pressure medications		
7	prevalentStroke	Fasting blood sugar > 120 mg/dl	0 = false; 1 = true		
8	prevalentHyp				
9	diabetes	Diabetes present or not	0 = No; 1 = Yes		
10	totChol	Total amount of cholesterol present in blood	mg/dL	236.6995	44.5913
11	sysBP	Systolic blood pressure	mmHg	132.3546	22.0333
12	diaBP	Diastoloc blood pressure	mmHg	82.8978	11.9104
13	BMI	Body Mass Index	Weight/Height (kg/m$^{2)}$	25.8008	4.0798
14	heartRate	Beats/Min (Ventricular)	Continuous	75	12.0254
15	glucose		mg/dL	81.9637	23.95433
16	TenYearCHD	Heart disease present or not	0 = No; 1 = Yes		

4A. Random Forest Analysis

1. For CKD Dataset

Random forest techniques used on the dataset of CKD data show that the bp, htn, su, and age are relatively significant factors with hemo, pcv and ba as the contradictory variables (Figure 1).

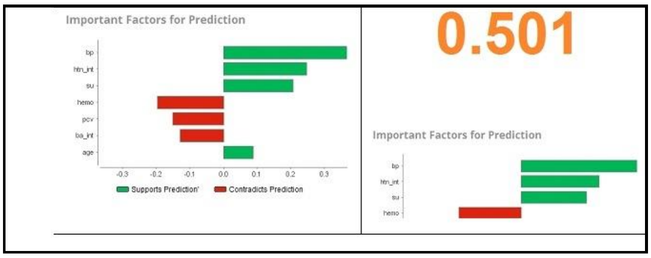


Fig. 1 - RF Analysis for CKD Dataset

2. For CVD Dataset

Random forest techniques used on the dataset of cardio vascular disease data show that the glucose, sysBP, age, diaBP, currentSmoker and prevalentHyp are relatively significant factors without the contradictory variables (Figure 2).

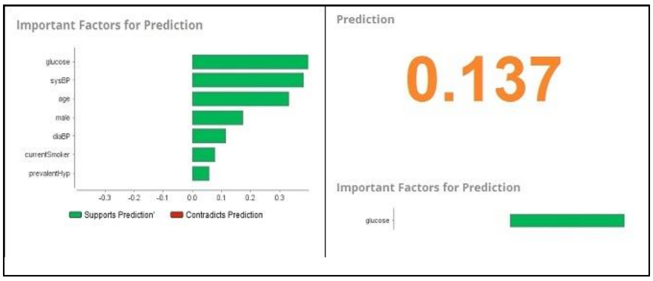


Fig. 2 - RF Analysis for CVD Dataset

4B. Decision Tree (DT) Analysis

1. For CKD Dataset

The tree diagram provides a graphical representation of the tree structure. This tree diagram (Figure 3) shows that:

- The age factor is the best predictor of CKD using the CHAID test.
- 18.8% of patients may have CKD if their age <= 24.
- This is referred to as a terminal node since it has no child nodes.
- If age > 39.0 and age <= 47.0, then 13.3% patients may have CKD.
- This is referred to as a terminal node since it has no child nodes.
- If age > 47.0 and age <= 54.0, then htn will be the next best predictor.
  - If htn = 1.0, then 100.0% patients may have CKD.
  - If htn = 0.0, then 8.3% patients may have CKD.
  - This is referred to as a terminal node since it has no child nodes.
- If age > 59.0 and age <= 69.0, then sc will be the next best predictor.
  - If sc <= 1.20 or missing, then 5.3% patients may have CKD.
  - If sc > 1.20, then 100.0% patients may have CKD.
  - This is referred to as a terminal node since it has no child nodes.
- If age > 69.0, then 12.5% patients may have CKD.
- If the value of age is missing, htn will be the next best predictor.
  - If htn = 1.0, then 98.4% patients may have CKD.
  - This is referred to as a terminal node since it has no child nodes.
  - If htn = 0.0, then pot will be the next best predictor.
    - If pot <= 4.30, then 78.6% patients may have CKD.
    - If pot > 4.30 and pot <= 5.20, then 54.5% patients may have CKD.
    - If pot > 5.20 or missing, then 100.0% patients may have CKD.
    - This is referred to as a terminal node since it has no child nodes.

The decision tree with all variables predicts the accuracy of 92.0% as shown in Table 3. The factors found to be relatively significant in decision tree analysis (i.e., age followed by htn, sc and pot) do not vary with results from RF analysis in both the findings.

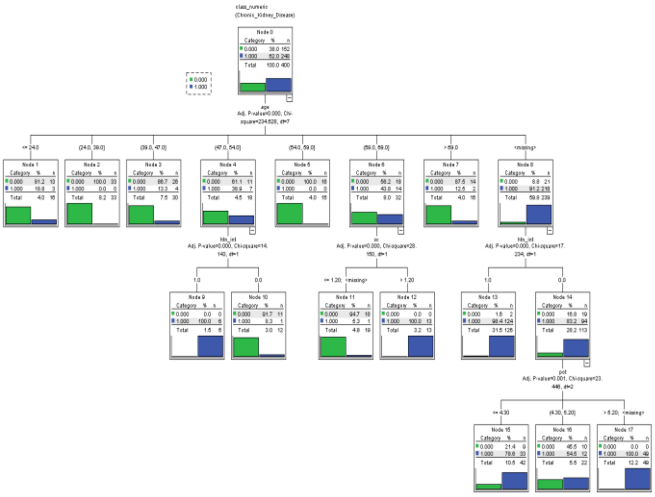


Fig. 3 - Outcome of DT Analysis Carried Out on All Variables of CKD Dataset

Table 3. Prediction Accuracy of DT Analysis Carried Out on All Variables of CKD Dataset

Classification			
Observed	**Predicted**		
	0	**1**	**Percent Correct**
0	131	21	86.20 %
1	11	237	95.60 %
Overall Percentage	35.50 %	64.50 %	92.00 %
Growing Method: CHAID			
Dependent Variable: class_numeric (Chronic_Kidney_Disease)			

2. Decision Tree Analysis for CKD Dataset Without Contradictory Variables

The results of the DT analysis are exactly the same as all variables in all respects.

3. Decision Tree Analysis of CKD Dataset Considering Relatively Important Variables Determined from RF Analysis

Figure 4 shows the outcome of DT analysis carried out on the Relatively Important factors determined by RF Analysis.

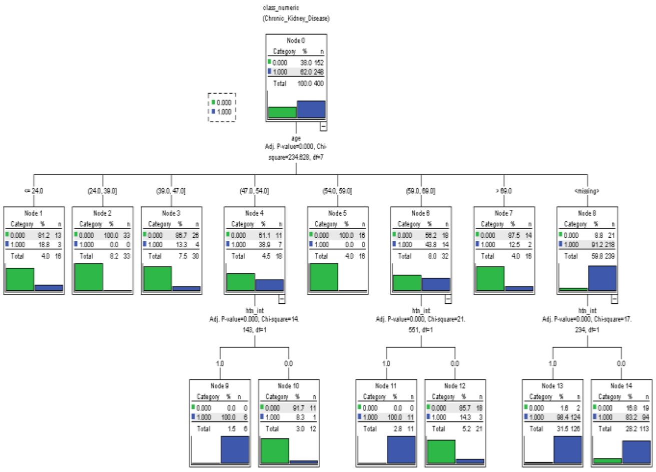


Fig. 4 - Outcome of DT Analysis Carried Out on Relatively Important Variables Determined from RF Analysis of CKD Dataset

The tree diagram is a graphic representation of the tree model. This tree structure (Figure 4) illustrates that:

• Using the CHAID method, age factor is the best predictor of CKD.

• If age <= 24.0, then 18.8% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If age > 39.0 and age <= 47.0, then 13.3% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If age > 47.0 and age <= 54.0, then htn will be the next best predictor.

• If htn = 1.0, then 100.0% patients may have CKD.

• If htn = 0.0, then 8.3% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If age > 59.0 and age <= 69.0, then htn will be the next best predictor.

• If htn = 1.0, then 100.0% patients may have CKD.

• If htn = 0.0, then 14.3% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If age > 69.0, then 12.5% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If age is equal to missing, then htn will be the next best predictor.

• If htn = 1.0, then 98.4% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

• If htn = 0.0, then 83.2% patients may have CKD.

• This is referred to as a terminal node since it has no child nodes.

Table 4. Prediction Accuracy of DT Analysis Carried Out on Relatively Important Variables Determined from RF Analysis of CKD Dataset

<table>
<tr><th colspan="4">Classification</th></tr>
<tr><th rowspan="2">Observed</th><th colspan="3">Predicted</th></tr>
<tr><th>0</th><th>1</th><th>Percent Correct</th></tr>
<tr><td><b>0</b></td><td>131</td><td>21</td><td>86.20 %</td></tr>
<tr><td><b>1</b></td><td>13</td><td>235</td><td>94.80 %</td></tr>
<tr><td><b>Overall Percentage</b></td><td>36.00 %</td><td>64.00 %</td><td>91.50 %</td></tr>
<tr><td colspan="4"><b>Growing Method:</b> CHAID</td></tr>
<tr><td colspan="4"><b>Dependent Variable</b>: class_numeric (Chronic_Kidney_Disease)</td></tr>
</table>

The DT analysis with relatively significant RF variables predicts 91.5 percent precision depicted in Table 4. For decision tree analysis, the variables considered to be relatively important are age followed by htn.

4. For CVD Dataset

Figure 5 shows the outcome of DT analysis carried out on all the factors of CVD dataset. The study predicts the rules of association as follows:

• age is the best factor for the detection of CVD.

• For age <= 41, the next best factor to predict the disease is cigsPerDay.

• If cigsPerDay <= 19.0 or missing, then 3.1% patients may have CVD.

• If cigsPerDay > 19.0, then sysBP is the next best predictor.

• This is referred to as a terminal node since it has no child nodes.

• If sysBP <= 133.0, then 6.5% patients may have CVD.

• If sysBP > 133.0, then 20.6% patients may have CVD.

• For age > 41 and age <= 46, the next best factor to predict the disease is currentSmoker.

• If currentSmoker = 0.0, then 5.7% patients may have CVD.

• If currentSmoker = 1.0, then prevalentHyp is the next best predictor.

• This is referred to as a terminal node since it has no child nodes.

• If prevalentHyp = 0.0, then 9.1% patients may have CVD.

• If prevalentHyp = 1.0, then 16.0% patients may have CVD.

• For age > 46 and age <= 54, the next best factor to predict the disease is male(gender).

• If male = 1.0, then sysBP is the next best predictor.

• This is referred to as a terminal node since it has no child nodes.

• If sysBP <= 123.5, then 11.4% patients may have CVD.

• If sysBP > 123.5 and sysBP <= 162.0, then 27.3% patients may have CVD.

• If sysBP > 162.0, then 60.6% patients may have CVD.

• If male = 0.0, then diabetes is the next best predictor.

• This is referred to as a terminal node since it has no child nodes.

• If diabetes = 0.0, then 10.5% patients may have CVD.

- If diabetes = 1.0, then 30.8% patients may have CVD.
- For age > 54 and age <= 61, the next best factor to predict the disease is prevalentHyp.
- If prevalentHyp = 0.0, then the next best factor to predict the disease is male.
- This is referred to as a terminal node since it has no child nodes.
- If male = 1.0, then 23.3% patients may have CVD.
- If male = 0.0, then 11.7% patients may have CVD.
- If prevalentHyp = 1.0, then the next best factor to predict the disease is cigsPerDay.
- This is referred to as a terminal node since it has no child nodes.
- If cigsPerDay <= 8.0, then 25.3% patients may have CVD.
- If cigsPerDay > 8.0, then 41.1% patients may have CVD.
- For age > 61, the next best factor to predict the disease is prevalentStroke.
- If prevalentStroke = 0.0, then the next best factor to predict the disease is prevalentHyp.
- This is referred to as a terminal node since it has no child nodes.
- If prevalentHyp = 0.0, then 21.8% patients may have CVD.
- If prevalentHyp = 1.0, then 34.6% patients may have CVD.
- If prevalentStroke = 1.0, then 100% patients may have CVD.

Table 5 shows the accuracy level of DT analysis. The decision tree with all variables estimates 85.1% accuracy. The variables found to be relatively significant in the analysis of DT (i.e. age followed by cigsPerDay, currentSmoker, male, prevalentHyp, prevalentStroke, sysBP and diabetes) do not differ in both findings with the results of RF analysis.

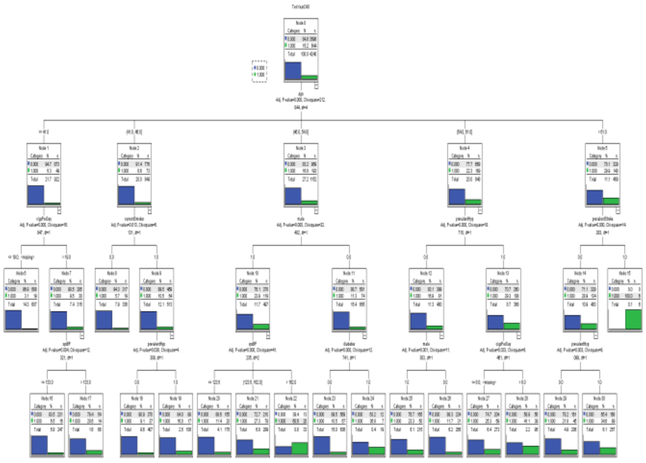


Fig. 5 - Result of Analysis of DT on All CVD Dataset Factors

Table 5. Predictive Precision of DT Analysis Conducted on All CVD Dataset Variables

Classification			
Observed	**Predicted**		
	0	**1**	**Percent Correct**
0	3583	13	99.60 %
1	618	26	4.00 %
Overall Percentage	99.10 %	0.90 %	85.10 %
Growing Method: CHAID			
Dependent Variable: TenYearCHD			

5. Decision Tree Analysis of CVD Dataset Considering Relatively Important Variables Determined from RF Analysis

Figure 6 shows the outcome of the Decision tree analysis carried out on Relatively Important Variables determined from RF Analysis. The study predicts the association rules as follows:

- age is the best factor for the detection of CVD.
- For age <= 41, the next best factor to predict the disease is currentSmoker.
- If currentSmoker = 0.0, then the next best factor to predict the disease is male.
- If male = 1.0, then 5.1% patients may have CVD.
- If male = 0.0, then 0.9% patients may have CVD.
- This is referred to as a terminal node since it has no child nodes.
- If currentSmoker = 1.0, then the next best factor to predict the disease is sysBP.
- If sysBP <= 133.0, then 5.7% patients may have CVD.
- If sysBP > 133.0, then 14.7% patients may have CVD.
- This is referred to as a terminal node since it has no child nodes.
- For age > 41 and age <= 46, the next best factor to predict the disease is currentSmoker.
- If currentSmoker = 0.0, then 5.7% patients may have CVD.
- This is referred to as a terminal node since it has no child nodes.
- If currentSmoker = 1.0, then prevalentHyp is the next best predictor.
- This is referred to as a terminal node since it has no child nodes.
- If prevalentHyp = 0.0, then 9.1% patients may have CVD.
- If prevalentHyp = 1.0, then 16.0% patients may have CVD.
- For age > 46 and age <= 54, the next best factor to predict the disease is male(gender).
- If male = 1.0, then sysBP is the next best factor to predict the disease.
- This is referred to as a terminal node since it has no child nodes.
- If sysBP <= 123.5, then 11.4% patients may have CVD.
- If sysBP > 123.5 and sysBP <= 162.0, then 27.3% patients may have CVD.
- If sysBP > 162.0, then 60.6% patients may have CVD.

  - If male = 0.0, then prevalentHyp is the next best predictor.
  - This is referred to as a terminal node since it has no child nodes.
    - If prevalentHyp = 0.0, then 9.3% patients may have CVD.
    - If prevalentHyp = 1.0, then 15.3% patients may have CVD.
- For age > 54 and age <= 61, the next best factor to predict the disease is prevalentHyp.
  - If prevalentHyp = 0.0, then the next best factor to predict the disease is male.
  - This is referred to as a terminal node since it has no child nodes.
    - If male = 1.0, then 23.3% patients may have CVD.
    - If male = 0.0, then 11.7% patients may have CVD.
  - If prevalentHyp = 1.0, then 29.3% patients may have CVD.
  - This is referred to as a terminal node since it has no child nodes.
- For age > 61, the next best factor to predict the disease is prevalentHyp.
  - If prevalentHyp = 0.0, then the next best factor to predict the disease is diaBP.
  - This is referred to as a terminal node since it has no child nodes.
    - If diaBP <= 69.0, then 40.0% patients may have CVD.
    - If diaBP > 69.0, then 17.5% patients may have CVD.
  - If prevalentHyp = 1.0, then the next best factor to predict the disease is male.
  - This is referred to as a terminal node since it has no child nodes.
    - If male = 1.0, then 45.7% patients may have CVD.
    - If male = 0.0, then 31.9% patients may have CVD.

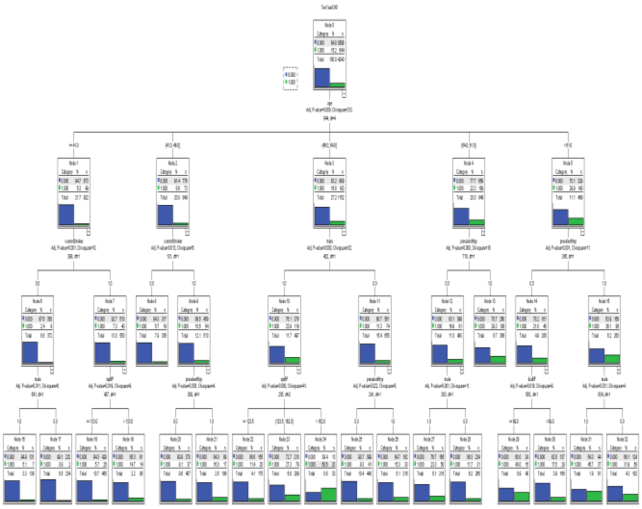


Fig. 6 - Outcome of DT Analysis Carried Out on Relatively Important Variables Determined from RF Analysis of CVD Dataset

Table 6 shows the accuracy level of Decision Tree analysis. The Decision tree analysis carried out on relatively important variables determined from RF Analysis of CVD dataset predicts the accuracy of 85.0%. The variables found to be relatively significant in decision tree analysis are age followed by currentSmoker, male, sysBP, prevalentHyp and diaBP.

Table 6. Prediction Accuracy of DT Analysis Carried Out on Relatively Important Variables Determined from RF Analysis of CVD Dataset

Classification			
Observed	**Predicted**		
	0	**1**	**Percent Correct**
0	3583	13	99.60 %
1	624	20	3.10 %
Overall Percentage	99.20 %	0.80 %	85.00 %
Growing Method: CHAID			
Dependent Variable: TenYearCHD			

4C. Logistic Regression (LR) Analysis

1. For CKD Dataset (LR Analysis of CKD Dataset Considering All Variables)

No result was produced for this analysis in this dataset as there are no cases for the estimation to be performed.

2. LR Analysis of CKD Dataset Considering All Variables Except Contradictory Variables

No result was produced for this analysis in this dataset as there are no cases for the estimation to be performed.

3. LR Analysis of CKD Dataset Considering Relatively Important Variables Determined from DT Analysis

The Logistics regression on variable found significant using DT (Figure 7) shows that no variables are significant along with the constant that predicts the presence of CKD. But LR on the data set with the variables selected from DT showed much better accuracy (92.0% in DT) compared to DT performs with all variables.

4. LR Analysis of CKD Dataset Considering Relatively Important Variables Determined from RF Analysis

The Logistics regression on variable found significant using RF shows that no variables are significant along with the constant that predicts the presence of CKD. But LR on the data set with the variables selected from RF (Figure 7) showed much better accuracy (92.0% in DT) compared to DT performs with all variables.

5. LR Analysis of CVD Dataset Considering All Variables

The Logistics regression on all variables age, cigsPerDay, male, prevalentStroke, sysBP and diabetes along with the constant, are statistically significant and predicts the presence of CVD. Equation 1 enables the prediction of the disease with these variables.

log(p/1-p) = –8.328 + 0.555*male + 0.064*age + 0.018*cigsPerDay + 0.002*totChol + 0.015*sysBP + 0.001*glucose ……………………………………………… (1)

The results showed an overall accuracy of 85.6% (Figure 7) which is higher than DT analysis with all variables. Along with that in this method classification accuracy of the presence of CVD is higher than DT analysis with all variables (4.0%). Thus, LR on the data set with the variables selected by DT showed better accuracy in predicting the presence of CVD.

6. LR Analysis of CVD Dataset Considering Relatively Important Variables Determined from DT Analysis

The Logistics regression on variable found significant using DT shows that age, cigsPerDay, male, prevalentStroke, sysBP and diabetes along with the constant, are statistically significant and predicts the presence of CVD. Equation 2 enables the prediction of the disease with these variables.

log(p/1-p) = –7.540 + 0.065*age + 0.020*cigsPerDay + 0.476*male + 1.021*prevalentStroke + 0.014*sysBP + 0.794*diabetes …………………………………………... (2)

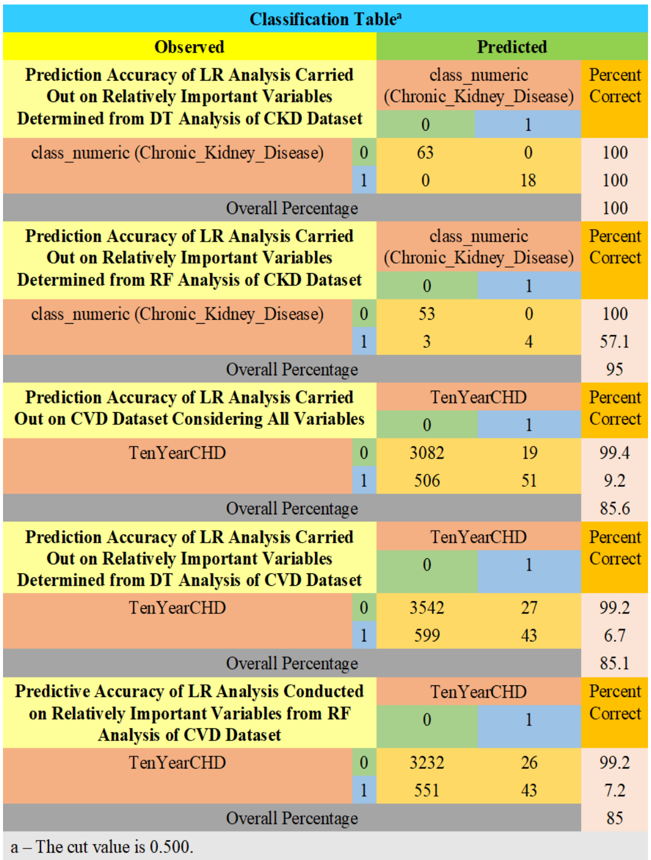
The results showed an overall accuracy of 85.1% (Figure 7) which is same as DT analysis with all variables. But in this method classification accuracy of the presence of CVD is higher than DT analysis with all variables (4.0%). Thus, LR on the data set with the variables selected by DT showed better accuracy in predicting the presence of CVD.

7. LR Analysis of CVD Dataset Considering Relatively Important Variables Determined from RF Analysis

The Logistics regression on variable found significant using RF shows that glucose, sysBP, age, male and currentSmoker along with the constant, are statistically significant and predicts the presence of CVD. Equation 3 enables the prediction of the disease with these variables.

log(p/1-p) = –8.061 + 0.007*glucose + 0.016*sysBP + 0.64*age + 0.594*male + 0.382*currentSmoker + 0.262*prevalentHyp …………………………………….. (3)

The results showed an overall accuracy of 85.0% (Figure 7). Thus, LR showed better accuracy in predicting the existence of CVD in the data set with all variables.



Classification Table[a]				
Observed		Predicted		
Prediction Accuracy of LR Analysis Carried Out on Relatively Important Variables Determined from DT Analysis of CKD Dataset		class_numeric (Chronic_Kidney_Disease)		Percent Correct
		0	1	
class_numeric (Chronic_Kidney_Disease)	0	63	0	100
	1	0	18	100
Overall Percentage				100
Prediction Accuracy of LR Analysis Carried Out on Relatively Important Variables Determined from RF Analysis of CKD Dataset		class_numeric (Chronic_Kidney_Disease)		Percent Correct
		0	1	
class_numeric (Chronic_Kidney_Disease)	0	53	0	100
	1	3	4	57.1
Overall Percentage				95
Prediction Accuracy of LR Analysis Carried Out on CVD Dataset Considering All Variables		TenYearCHD		Percent Correct
		0	1	
TenYearCHD	0	3082	19	99.4
	1	506	51	9.2
Overall Percentage				85.6
Prediction Accuracy of LR Analysis Carried Out on Relatively Important Variables Determined from DT Analysis of CVD Dataset		TenYearCHD		Percent Correct
		0	1	
TenYearCHD	0	3542	27	99.2
	1	599	43	6.7
Overall Percentage				85.1
Predictive Accuracy of LR Analysis Conducted on Relatively Important Variables from RF Analysis of CVD Dataset		TenYearCHD		Percent Correct
		0	1	
TenYearCHD	0	3232	26	99.2
	1	551	43	7.2
Overall Percentage				85

a – The cut value is 0.500.

Fig. 7 - Prediction Accuracy of LR Analysis

5. Results and Discussions

This research stresses that the addition and exclusion of conflicting variables do not boost results reliability. For example, prediction accuracy, presence of CKD as well as an absence of CKD, using Logistics Regression remains the same on the inclusion of significant variables and exclusion of conflicting variables. Authors also found from the analysis of the CKD dataset that hypertension is the factor with the highest score followed by age. Equation 4 gives this finding expression.

RFi ∩ DTi ∩ LRi = { htn, age } ……………………….. (4)

Where RFi defines RF analysis on dataset; DTi defines DT analysis on dataset; LRi defines LR analysis on dataset; i represents original dataset.

From the analysis of CVD dataset, authors found the Systolic Blood Pressure, Age in years and male(gender) as the factors with the highest score. Equation 5 and 6 give the expression of this finding.

RFi ∩ DTi ∩ LRi = { sysBP, age, male } ……………... (5)

DTii ∩ LRii = { sysBP, age, male } ……………………. (6)

Where ii represents a revised set of data containing only relatively significant variables.

Table 7. Comparison of Accuracy Levels and Identification of Significant Variables for CKD Dataset

Approaches	age	bp	su	sc	pot	htn	Accuracy Predicting 1	Accuracy Predicting 0
RF	1	1	1	-	-	1	50.1	-
DT – I	1	-	-	1	1	1	95.6	86.2
DT – II	1	-	-	-	-	1	94.8	86.2
DT – III	-	-	-	-	-	-	95.6	86.2
LR – I	-	-	-	-	-	-	-	-
LR – II	-	-	-	-	-	-	100	100
LR – III	-	-	-	-	-	-	57.1	100
DA – I	-	-	-	-	-	-	-	-
DA – II	-	-	-	-	-	-	-	-
DA – III	-	-	-	-	-	-	-	-
TOTAL	3	1	1	1	1	3	-	-

Table 8. Comparison of Accuracy Levels and Identification of Significant Variables for CVD Dataset

Approaches	Male	Age	currentSmoker	cigsPerDay	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	glucose	Accuracy Predicting 1	Accuracy Predicting 0
RF	1	1	1	-	-	1	-	-	1	1	1	13.7	-
DT – I	1	1	1	1	1	1	1	-	1	-	-	4	99.6
DT – II	1	1	1	-	-	1	-	-	1	1	-	3.1	99.6
DT – III	-	-	-	-	-	-	-	-	-	-	-	-	-
LR – I	1	1	-	1	-	-	-	1	1	-	1	9.2	99.4
LR – II	1	1	-	1	1	-	1	-	1	-	-	6.7	99.2
LR – III	1	1	1	-	-	-	-	-	1	-	1	7.2	99.2
DA – I	-	-	-	-	-	-	-	-	-	-	-	-	-
DA – II	-	-	-	-	-	-	-	-	-	-	-	-	-
DA – III	-	-	-	-	-	-	-	-	-	-	-	-	-
TOTAL	6	6	4	3	2	3	2	1	6	2	3		

A unique approach was developed with the use of machine learning methods to predict CKD and CVD variables. The CKD-CVD relationship has been extensively documented in the literature. Both CKD and CVD may contribute to some common traditional risk factors such as age, smoking, obesity, hypertension, diabetes mellitus, and dyslipidemia. Hypertension is generally defined as systolic blood pressure is greater than or equal to 140 mm Hg or diastolic blood pressure is greater than or equal to 90 mm Hg or antihypertensive drug use [37]. Nonetheless, in patients with CKD, CVD remains under suspicion or undertreated. From the study of CKD and CVD datasets, authors found age and hypertension (calculated field, added to CKD dataset using literature [37]) as the highest score factors in both datasets (Table 7 and 8). Predictive accuracies of CKD and CVD using age and hypertension as predicting factor in RF analysis are 82.5% (highest is 100%) and 17.7% (highest) respectively (Figure 8 and 9 shown below).

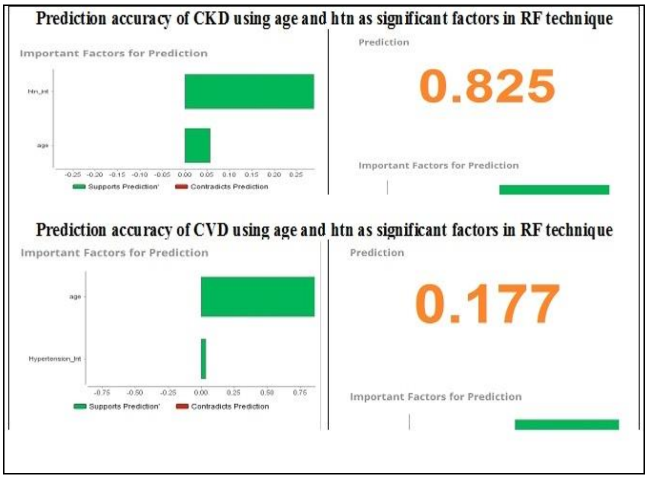

Prediction accuracy of CKD using age and htn as significant factors in RF technique
Important Factors for Prediction
Htn_int
age
-0.25 -0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20 0.25
Supports Prediction'
Contradicts Prediction
Prediction
0.825
Important Factors for Prediction
Prediction accuracy of CVD using age and htn as significant factors in RF technique
Important Factors for Prediction
age
Hypertension_Int
-0.75 -0.50 -0.25 0.00 0.25 0.50 0.75
Supports Prediction'
Contradicts Prediction
Prediction
0.177
Important Factors for Prediction

Fig. 8 - RF Analysis on CKD and CVD Dataset Taking age and hypertension as Independent Variables

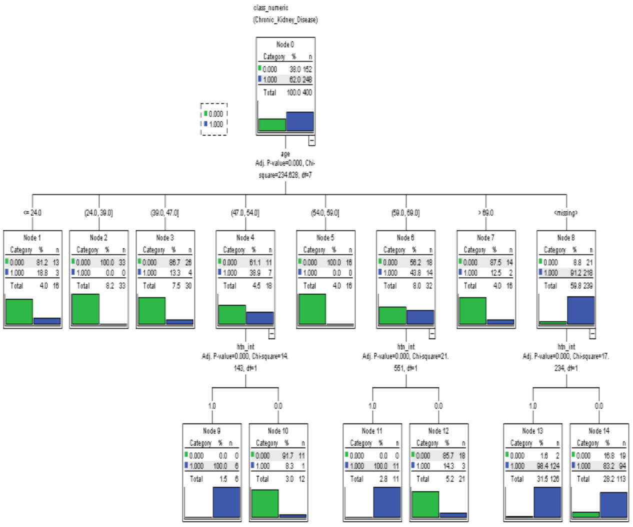


Fig. 9 - Analysis on CKD Dataset Taking age and hypertension as Independent Variables

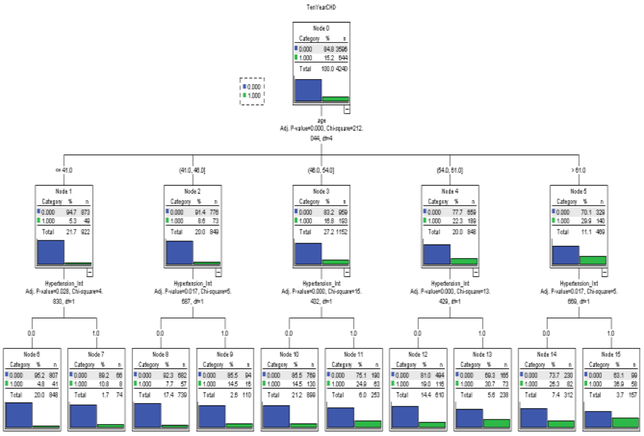


Fig. 10 - Study of age and hypertension of CVD Dataset as Independent Variables

The collective association rule related with DT analysis technique from Figure 9 and 10 above on CKD and CVD datasets is:

- If age > 47.0 and age <= 54.0, then htn will be the next best predictor.
    - If htn = 1.0, then patients may be attacked with CKD and CVD.
    - This is referred to as a terminal node since it has no child nodes.
- If age > 59.0 and age <= 69.0, then htn will be the next best prediction factor.
    - If htn = 1.0, then patients may be attacked with CKD and CVD.
    - This is referred to as a terminal node since it has no child nodes.

From the literature review, we have again found that ([25 - 36]) high blood potassium levels are one of the most significant heart failure factors. For CKD patients, if the potassium level is high, they may be suggested to take additional tests such as taking the number of cigarettes per day, whether presently smoker or not, prevalent hypertension present or not, attacked with stroke or not, and diabetes for CVD prevention. Clinicians must be able to recognize that CKD patients are a high-risk group for developing CVD and cardiovascular events. Additional studies to determine the dangerous causes of CVD in CKD patients are needed to develop and implement methods of prevention and treatment to minimize high morbidity and mortality in CKD patients. CKD is an independent risk factor for CVD and most patients die as a result of CVD than progression to End-Stage Renal Disease (ESRD). This risk increases as the severity of kidney dysfunction deteriorate. Distinguishing evidence of early CKD patients is important because prevention works better than cure.

6. Decreased Kidney Function and CVD

The higher prevalence of typical risk factors is only partially responsible for the increased occurrence of CVD in CKD patients. This has turned our attention to finding a link between CVD and CKD and to locate the common ‘novel’ risk factors that cause CKD as well as CVD at the same time. The current findings have significant consequences for public health and clinical effects. Epidemiologic studies have documented an extraordinarily high incidence of death from CVD among CKD patients. In this research article analysis highlights a poor risk factor profile for CVD in patients suffering from CKD. The predictive value of such a profile for CVD has not been well studied in patients with CKD. However, increased morbidity and mortality from CVD have been associated with these risk factors in the general population [40 - 41, 43]. Our findings provide additional encouragement for close screening and management of patients with CKD to lower their risk of CVD sufficiently.

The destruction of the blood vessels within it significantly decreases kidney functions. This destruction is caused due to high blood pressure. High blood pressure (Hypertension) spreads blood vessels and eventually weakens them. Once the blood pressure has increased, the vessels spread faster, so the blood flows faster inside the kidney. The kidney’s damaged blood vessels can prevent waste and excess fluid from being filtered out of the body. If the blood vessels have damaged the blood vessel’s additional fluid may increase blood pressure, further leading to a hazardous cycle [42]. CKD is a different coronary artery disease risk factor (CAD). It is the primary cause of death and mortality for patients with CKD [38]. Coresh et al. [39], in their research, found that 70% of people with an elevated level of serum creatinine have high blood pressure. This high blood pressure causes damage to the blood vessels, which will increase the blood pressure further. Elevated blood pressure, CAD, and hypertension are strong diagnostic features for CKD.

The two major predictors of CKD are sodium and potassium. Balancing of potassium and sodium is of utmost importance for the human body. The excess level of sodium and potassium in the body increases the fatality rate among CKD patients. A person attacked with CKD will neither be able to remove potassium, sodium, and fluid from the body nor be able to accumulate into the bloodstream and body cells gradually. An elevated sodium level causes high blood pressure [42]. Simultaneously, patients with an advanced CKD level suffer from high potassium levels in the bloodstream called hyperkalemia. Hyperkalemia results in numbness, fatigue, nausea, lower limb swelling, foot ulcer, chest pain, less self-confidence, anxiety, or slow pulse rate among CKD patients.

So, CKD is one of the strongest factors for CVD but at present clinically used measures of kidney disease (urine albumin to creatinine ratio, uACR and estimated glomerular filtration rate, eGFR) mainly focused on glomerular health. The kidney tubules perform a multitude of functions that are important for homeostasis maintenance and contain important CVD prognostic information. The authors demonstrated the importance of the hypertension factor in the diagnosis of CVD risk factors from CKD. This study will improve the identification of people who suffered from CVD, a common complication of CKD.

7. Conclusion

To strengthen and implement preventive and remedial approaches to reduce high morbidity and mortality in CKD patients, further studies are needed to further identify risk factors for CVD in patients with CKD. The incidence of CKD is rapidly increasing worldwide, which also speaks to a significant burden for patients and society due to the smaller scale and macro vascular complexities which individuals with this condition may face, and hence CVDs that are the most prevalent causes of CKD patients’ morbidity and mortality.

This study shows that data mining-based methodologies can be used to determine the factors that affect the risk of CKD and thus estimate the likelihood of adult patients having heart disease. Instead of traditional illustrative statistical analysis approaches and methodologies that only involve expert variables, the use of decision trees, random forests, and logistics regression models give a useful list of risk factors that some have been incorporated into the existing studies; meanwhile, some others have been missing from the related literature. For the overall model, hypertension is the factor with the highest score followed by age is the most important factor.

The after-effect of the test also shows that the system of forecasts is designed to accurately, most critically, and timely predict the risk of heart disease. It means that the application software is fit to assist a doctor in making decisions regarding health risks to the patient. This produces outcomes that bring this closer to the real circumstances, making data mining in the healthcare segment increasingly supportive, which means that it is necessary for knowledge discovery in the healthcare sector.

References

[1] World Health Organization, “Preventing Chronic Disease: A Vital Investment,” Geneva, WHO, 2005.

[2] Grassmann, A., Gioberge, S., Moeller, S., & Brown, G. (2005). ESRD patients in 2004: global overview of patient numbers, treatment modalities and associated trends. Nephrol. Dial. Transplant., 20(12), 2587-2593.

[3] Abdelhmid, S. M. S., & Ajith, A. (2014). Novel Ensemble Decision Support and Health Care Monitoring System. Journal of Network and Innovative Computing, 2, 041-051.

[4] Han, J., & Kamber, M. (2003). Data mining: concepts and techniques, 3rd ed. Burlington, MA, Elsevier.

[5] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A Review on Ensembles for the Class Imbalance Problem: Bagging, Boosting and Hybrid-Based Approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4), 463-484.

[6] Sen, A. K., Patel, S. B., & Shukla, D. D. (2013). A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level. International Journal of Engineering & Computer Science, 2(9), 2663-2671.

[7] Herland, M., Khoshgoftaar, T. M., & Wald, R. (2014). A review of data mining using big data in health informatics. Journal of Big Data, 1(1), 1.

[8] Masethe, H. D., & Masethe, M. A. (2014). Prediction of Heart Disease using Classification Algorithms. in Proc. of the World Congress on Engineering and Computer Science, 2, 22-24.

[9] Sarvestani, A. S., Safavi, A. A., Parandeh, N. M., & Salehi, M. (2010, October). Predicting breast cancer survivability using data mining techniques. In 2010 2nd International Conference on Software Technology and Engineering (Vol. 2, pp. V2-227). IEEE.

[10] Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung Journal of Medical Sciences, 29(2), 93-99.

[11] Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E., & Hsu, C. Y. (2004). Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. New Engl J Med, 351(13), 1296-1305.

[12] Bright, R. (1836). Cases and observations illustrative of renal disease accompanied with the secretion of albuminous urine. Med. Chir. Rev., 25(49), 23-35.

[13] Gansevoort, R. T., Correa-Rotter, R., Hemmelgarn, B. R., Jafar, T. H., Heerspink, H. J. L., Mann, J. F., ... & Wen, C. P. (2013). Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. Lancet, 382(9889), 339-352.

[14] Eckardt, K. U., Coresh, J., Devuyst, O., Johnson, R. J., Köttgen, A., Levey, A. S., & Levin, A. (2013). Evolving importance of kidney disease: from subspecialty to global health burden. Lancet, 382(9887), 158-169.

[15] Eknoyan, G., Lameire, N., Eckardt, K., Kasiske, B., Wheeler, D., Levin, A., ... & Levey, A. S. (2013). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int, 3(1), 5-14.

[16] Matsushita, K., van der Velde, M., Astor, B. C., Woodward, M., Levey, A. S., de Jong, P. E., ... & Chronic Kidney Disease Prognosis Consortium. (2010). Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet, 375(9731), 2073-2081.

[17] Gansevoort, R. T., Matsushita, K., Van Der Velde, M., Astor, B. C., Woodward, M., Levey, A. S., ... & Coresh, J. (2011). Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. Kidney Int, 80(1), 93-104.

[18] Wen, C. P., Cheng, T. Y. D., Tsai, M. K., Chang, Y. C., Chan, H. T., Tsai, S. P., ... & Wen, S. F. (2008). All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan. Lancet, 371(9631), 2173-2182.

[19] Hemmelgarn, B. R., Clement, F., Manns, B. J., Klarenbach, S., James, M. T., Ravani, P., ... & Jindal, K. (2009). Overview of the Alberta kidney disease network. BMC Nephrol, 10(1), 30.

[20] Franco, O. H., Steyerberg, E. W., Hu, F. B., Mackenbach, J., & Nusselder, W. (2007). Associations of diabetes mellitus with total life expectancy and life expectancy with and without cardiovascular disease. Arch Intern Med, 167(11), 1145-1151.

[21] Franco, O. H., Peeters, A., Bonneux, L., & De Laet, C. (2005). Blood pressure in adulthood and life expectancy with cardiovascular disease in men and women: life course analysis. Hypertension, 46(2), 280-286.

[22] Upadhyay, A., Earley, A., Haynes, S. M., & Uhlig, K. (2011). Systematic review: blood pressure target in chronic kidney disease and proteinuria as an effect modifier. Ann Intern Med, 154(8), 541-548.

[23] Lv, J., Ehteshami, P., Sarnak, M. J., Tighiouart, H., Jun, M., Ninomiya, T., ... & Strippoli, G. F. (2013). Effects of intensive blood pressure lowering on the progression of chronic kidney disease: a systematic review and meta-analysis. Cmaj, 185(11), 949-957.

[24] Kokubo, Y., Nakamura, S., Okamura, T., Yoshimasa, Y., Makino, H., Watanabe, M., ... & Kawano, Y. (2009). Relationship between blood pressure category and incidence of stroke and myocardial infarction in an urban Japanese population with and without chronic kidney disease: the Suita Study. Stroke, 40(8), 2674-2679.

[25] Conway, R., Creagh, D., Byrne, D. G., O'Riordan, D., & Silke, B. (2015). Serum potassium levels as an outcome determinant in acute medical admissions. Clin Med, 15(3), 239.

[26] Kovesdy, C. P. (2015). Management of hyperkalemia: an update for the internist. Am J Med, 128(12), 1281-1287.

[27] Ingelfinger, J. R. (2015). A new era for the treatment of hyperkalemia?. New Engl J Med, 372(3), 275.

[28] Krogager, M. L., Eggers-Kaas, L., Aasbjerg, K., Mortensen, R. N., Køber, L., Gislason, G., ... & Søgaard, P. (2015). Short-term mortality risk of serum potassium levels in acute heart failure following myocardial infarction. Eur Heart J Cardiovasc Pharmacother, 1(4), 245-251.

[29] Ahmed, M. I., Ekundayo, O. J., Mujib, M., Campbell, R. C., Sanders, P. W., Pitt, B., ... & Aronow, W. S. (2010). Mild hyperkalemia and outcomes in chronic heart failure: a propensity matched study. Int J Cardiol, 144(3), 383-388.

[30] Michel, A., Martín-Pérez, M., Ruigómez, A., & García Rodríguez, L. A. (2015). Risk factors for hyperkalaemia in a cohort of patients with newly diagnosed heart failure: a nested case–control study in UK general practice. Eur J Heart Fail, 17(2), 205-213.

[31] Abbas, S., Ihle, P., Harder, S., & Schubert, I. (2015). Risk of hyperkalemia and combined use of spironolactone and long-term ACE inhibitor/angiotensin receptor blocker therapy in heart failure using real-life data: a population-and insurance-based cohort. Pharmacoepidemiology and Drug Safety, 24(4), 406-413.

[32] Ronco, C., McCullough, P., Anker, S. D., Anand, I., Aspromonte, N., Bagshaw, S. M., ... & Daliento, L. (2010). Cardio-renal syndromes: report from the consensus conference of the acute dialysis quality initiative. Eur Heart J, 31(6), 703-711.

[33] Ahmed, A., Rich, M. W., Sanders, P. W., Perry, G. J., Bakris, G. L., Zile, M. R., ... & Shlipak, M. G. (2007). Chronic kidney disease associated mortality in diastolic versus systolic heart failure: a propensity matched study. Am J Cardiol, 99(3), 393-398.

[34] Yancy, C. W., Jessup, M., Bozkurt, B., Butler, J., Casey, D. E., Colvin, M. M., ... & Hollenberg, S. M. (2017). 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Am Coll Cardiol, 70(6), 776-803.

[35] Palmer, B. F. (2004). Managing hyperkalemia caused by inhibitors of the renin–angiotensin–aldosterone system. New Engl J Med, 351(6), 585-592.

[36] Packham, D. K., Rasmussen, H. S., Lavin, P. T., El-Shahawy, M. A., Roger, S. D., Block, G., ... & Singh, B. (2015). Sodium zirconium cyclosilicate in hyperkalemia. New Engl J Med, 372(3), 222-231.

[37] Mertens, I. L., & Van Gaal, L. F. (2000). Overweight, obesity, and blood pressure: the effects of modest weight reduction. Obes Res, 8(3), 270-278.

[38] Cai, Q., K Mukku, V., & Ahmad, M. (2013). Coronary artery disease in patients with chronic kidney disease: a clinical update. Current cardiology reviews, 9(4), 331-339.

[39] Coresh, J., Wei, G. L., McQuillan, G., Brancati, F. L., Levey, A. S., Jones, C., & Klag, M. J. (2001). Prevalence of high blood pressure and elevated serum creatinine level in the United States: findings from the third National Health and Nutrition Examination Survey (1988-1994). Archives of internal medicine, 161(9), 1207-1216.

[40] de Gonzalo-Calvo, D., Martínez-Camblor, P., Bär, C., Duarte, K., Girerd, N., Fellström, B., ... & Rossignol, P. (2020). Improved cardiovascular risk prediction in patients with end-stage renal disease on hemodialysis using machine learning modeling and circulating microribonucleic acids. Theranostics, 10(19), 8665.

[41] Mezzatesta, S., Torino, C., De Meo, P., Fiumara, G., & Vilasi, A. (2019). A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. Computer methods and programs in biomedicine, 177, 9-15.

[42] Salekin, A., & Stankovic, J. (2016, October). Detection of chronic kidney disease and selecting important predictive attributes. In 2016 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 262-270). IEEE.

[43] Yuan, J., Zou, X. R., Han, S. P., Cheng, H., Wang, L., Wang, J. W., ... & C-STRIDE study group. (2017). Prevalence and risk factors for cardiovascular disease among chronic kidney disease patients: results from the Chinese cohort study of chronic kidney disease (C-STRIDE). BMC nephrology, 18(1), 23.