Data Analysis on Chronic Kidney Disease Prognosis

Divya Jain, Parshavi Bolya, Aaditya Maheshwari, Yogendra Singh Solanki

Abstract: ***We have taken our dataset from UCI Machine Learning Repository. Our study is about Chronic Kidney Diseases based on 24 input attributes to produce one output attribute i.e. a patient is suffering from chronic kidney disease or not. We have used three major attributes in our study i.e. PCV, RBCC and Hemoglobin with respect to Age for optimum result. These attributes play major role in our study.***

Keywords: -CKD, Analytics, Weka, PCV, RBCC, Hemoglobin.

I. INTRODUCTION

Chronic Kidney Disease (CKD) defines gradual loss of kidney function over a period of time [1]. Risk of chronic kidney disease increases due to the following factors i.e. diabetes, high blood pressure, age and genetics [2]. The probability of patients suffering from CKD is due to anemia which depends on PCV, concentration of hemoglobin and TC of RBCC. It has been observed that concentration of hemoglobin, PCV and TC of RBCC is significantly lower in three stages of CKD [3]. The factors also includesrecurrent urinary tract infections, hypertension, aillness or urinary obstruction that affects the kidneys [4].The most prevalent symptoms include inability to urinate, fatigue, malaise, headaches, nausea and vomiting, bone pain, frequent epistaxis and swelling [5]. It has been contemplated that by growing age we tend to lose kidney function usually in men than in women. So, it is advised especially over age 50 to get screened for CKD. Also as we age, we're more likely to develop Type 2 diabetes and high blood pressure [6].

II. METHODOLOGY

The dataset for the project is taken from the MachineLearningRepository UCI. http://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease.The dataset comprises of 400 instances and 24 input attributes plus 1 class attribute. The output is based on clinical results on whether a person suffers from chronic kidney disease or not.For our study we have used Weka 3.8.4 tool to classify data for chronic kidney disease patients and vice- versa. On the basis of age we have mapped the values for PVC, RBCC and Hemoglobin. Also Microsoft Excel has been used for graphical representation of the same.

Revised Manuscript Received on March 5, 2020.

Divya Jain*, Techno India NJR Institute of Technology, Udaipur (Raj)-313003. ***jaind2710@gmail.com**

Parshavi Bolya, Techno India NJR Institute of Technology, Udaipur (Raj)-313003

Aaditya Maheshwari, Techno India NJR Institute of Technology, Udaipur (Raj)-313003

Yogendra Singh Solanki, Techno India NJR Institute of Technology, Udaipur (Raj)-313003

III. EXPERIMENTAL RESULT AND ANALYSIS

To classify chronic kidney disease patient's various data processing algorithms have been used in this study.
Thedataset comprises of 400 instances of CKD. The following
table was generated to determine the accuracy of the dataset.

Table 1: Classification results using Weka 3.8.4

Classifier	CC	MAE	RMSE	RRSE	KS	CCI
bayes.BayesNet	0.985	0.0153	0.0998	20.614	0.968	394
fnctions.SimpleLogistic	0.995	0.0237	0.079	16.2936	0.9894	398
functions.Logistic	0.9775	0.0196	0.1289	26.596	0.9523	391
functions.SGD	0.9875	0.0125	0.1118	23.0627	0.9736	395
lazy.Kstar	0.975	0.0521	0.1769	36.548	0.9411	389
meta.AdaBoostM1	0.995	0.01	0.0717	14.8125	0.9893	398
meta.AttributeSelectedClassifier	0.9825	0.0303	0.0875	18.0817	0.9623	393
meta.Bagging	0.9975	0.0483	0.1006	20.7854	0.9947	399
meta.ClassificationViaRegression	0.99	0.0475	0.1209	24.9691	0.9787	396
meta.FilteredClassifier	0.99	0.0244	0.0996	20.5752	0.9788	396
meta.LogitBoost	0.9925	0.0156	0.0719	14.8507	0.984	397
meta.MultiClassClassifier	0.99	0.012	0.1037	21.4286	0.9788	396
meta.RandomSubSpace	0.9975	0.0657	0.1125	23.2476	0.9947	399
trees.RandomForest	0.9975	0.0234	0.0592	12.2238	0.9947	399
tress.J48	0.9825	0.0326	0.0925	19.1041	0.9623	393

where,
CC : Correlation Coefficient
MAE: Mean Absolute Error
RMSE: Root Mean Squared Error
RRSE:Root Relative Square Error
KS:KappaStatitics
CCI:Correctly Classified Instances

The best classification result is observed from Random Forest classification algorithm with CCI equal to 399, also the value of Kappa Statistics > 0.8(i.e. 0.9947) and the Mean Absolute Error is 0.0234 which is less than that of RandomSubSpace(0.0657).

IV. GRAPHICAL REPRESENTATION OF ANALYSIS

A. PCV

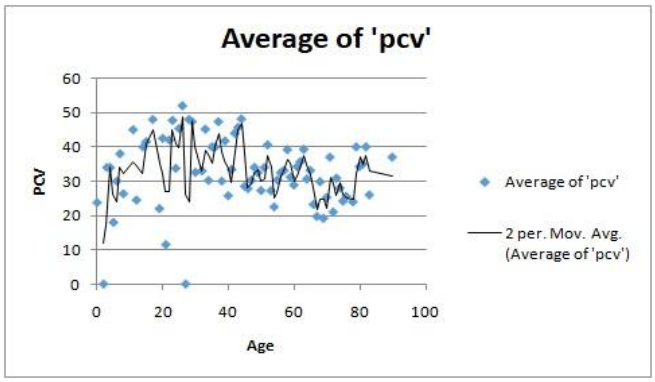

Figure 1: Concentration of PCV w.r.t. Age groups.

From the above graph it is observed that in age group 61-80 the concentration of PCV is lowest and in age group 21-40 it is maximum.

Table 2: Statistical values of PCV.

age	avg	max	min
0-20	31.17222	42.5	0
21-40	36.49791	52	0
41-60	33.5599	48.16667	22.5
61-80	29.21336	40	19.16667
81-100	34.5	40	26

The above table represents the average, maximum and minimum value of PCV w.r.t. Age.

B. RBCC

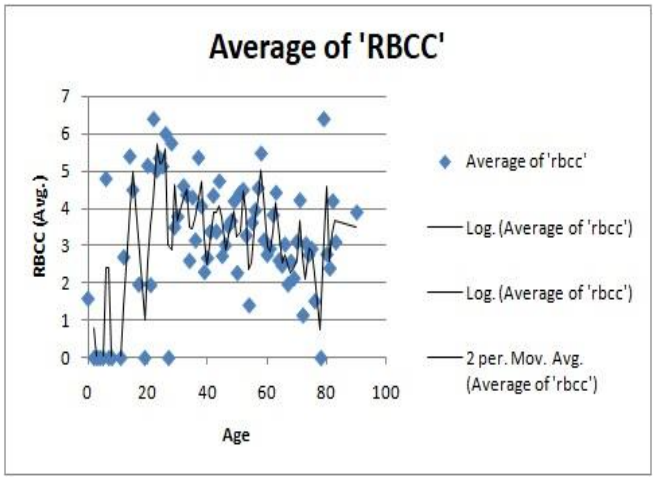

Figure 2: Concentration of RBCC w.r.t. Age groups.

From the above graph it is observed that in age group 21-40 and 61-80 the concentration of PCV is maximum and in age group 0-40 and 61-80 it is minimum.

Table 3: Statistical values of RBCC.

age	avg	max	min
0-20	1.740278	5.4	0
21-40	4.016792	6.4	0
41-60	3.619799	5.48	1.41
61-80	2.836851	6.4	0
81-100	3.4	4.2	2.4

The above table represents the average, maximum and minimum value of PCV w.r.t. Age.

C. Hemoglobin

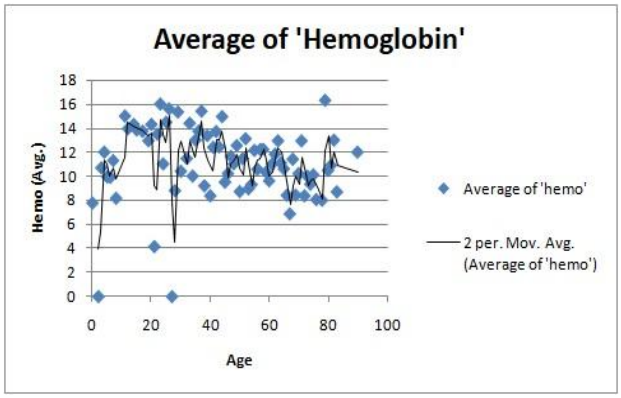

Figure 3: Concentration of Hemoglobin w.r.t. Age groups.

From the above graph it is observed that in age group 61-80 the concentration of Hemoglobin is minimum and approximately equal values in other age groups.

Table 4: Statistical values of Hemoglobin.

age	avg	max	min
0-20	11.19167	15	0
21-40	11.49037	16	0
41-60	11.35979	14.95	8.725
61-80	10.33991	16.3	6.871429
81-100	11.15	13	8.7

The above table represents the average, maximum and minimum value of Hemoglobin w.r.t. Age.

V. ABBREVIATIONS

Abbreviation used in this paper are as follows:

age	-	age
bp	-	blood pressure
sg	-	specific gravity
al	-	albumin
su	-	sugar
rbc	-	red blood cells
pc	-	pus cell
pcc	-	pus cell clumps
ba	-	bacteria
bgr	-	blood glucose random
bu	-	blood urea
sc	-	serum creatinine
sod	-	sodium
pot	-	potassium
hemo	-	hemoglobin
pcv	-	packed cell volume
wc	-	white blood cell count
rc	-	red blood cell count
htn	-	hypertension
dm	-	diabetes mellitus
cad	-	coronary artery disease
appet	-	appetite
pe	-	pedal edema
ane	-	anemia
class	-	class

VI. CONCLUSION

The analysis suggests that the best accuracy is obtained using Random Forest algorithm that is 99.75%, also it portrays that PVC, RBCC and Hemoglobin are the major factors for the study of CKD and its treatment. It is found out that the values of PVC, RBCC and Hemoglobin are lowest in age group between 61-80 therefore the chances of renal failure is maximum in this age group.

REFERENCES

1. G "What Is Chronic Kidney Disease?". National Institute of Diabetes and Digestive and Kidney Diseases. June 2017. Retrieved 19 December 2017.
2. The National Kidney Foundation (NKF)
3. Hsu CY, Mcculloch CE, Curhan GC. Epidemiology of anemia associated with chronic renal insufficiency among adults in the United States: results from the Third National Health and Nutrition Examination Survey. J Am SocNephrol. 2002;13:504–10
4. National Kidney Foundation. K/DOQI, clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002;39 (2 suppl 1):S1–266
5. Carissa Stephens, RN, CCRN, CPN on July 10, 2017 — Written by Christine DiMaria, Marijane Leonard, and Tim Jewell
6. Rashid HU. Haemodialysis: how to reduce cost. Bangladesh Renal J. 1989;8(2):43–4

AUTHORS PROFILE

Divya Jain, Department of Computer Science &Engineering, Techno India NJR Institute of Technology, Udaipur 313003

ParshaviBolya,Department of Computer Science & Engineering, Techno India NJR Institute of Technology, Udaipur 313003

Aditya Maheshwari is working in the capacity of Project Lead- New Initiative & Research at Techno India NJR Institute of Technology Udaipur. He is an Open Source Advocate, leading technical speaker and Expert of User Research (UI/UX). Leading students by creating and developing projects applying cloud technologies at an engineering institute. Achieved specialization and certification in IBM Watson and BlockChain, and applied in developing projects.

Yogendra Singh Solanki, is working as Asst. Professor, Department of Electronics and Communications, Techno India NJR Institute of Technology, Udaipur. He is practicing the IoT Devices, AI/ML and VLSI in day to day life.