Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning

[1]Akash Maurya,[2]Rahul Wable,[3]Rasika Shinde ,[4]Sebin John ,[5]Rahul Jadhav, [6]Dakshayani.R
[1234]Student, [56]Assistant Professor
Department of Computer Engineering
Fr. Conceicao Rodrigues Institute of Technology, Vashi, Navi Mumbai
[1] akash.am343@gmail.com[2]racul803@gmail.com, [3]rasikashinde976@gmail.com,
[4]johnsebin97@gmail.com,[5]rahulj212@gmail.com,[6]gdakshayani@gmail.com
Department of Computer Engineering
Fr. C. Rodrigues Institute of Technology, Vashi, India

***Abstract*— Chronic kidney disease (CKD) is a type of kidney disease in which there is gradual loss of kidney function over a period of months or years. Prediction of this disease is one of the most important problems in medical fields. So automated tool which will use machine learning techniques to determine the patient's kidney condition that will be helpful to the doctors in prediction of chronic kidney disease and hence better treatment. The proposed system extracts the features which are responsible for CKD, then machine learning process can automate the classification of the chronic kidney disease in different stages according to its severity. The objective is to use machine learning algorithm and suggest suitable diet plan for CKD patient using classification algorithm on medical test records. Diet recommendation for patient will be given according to potassium zone which is calculated using blood potassium level to slow down the progression of CKD.**

Keywords—Chronic kidney disease, prediction, diet, machine learning

I. INTRODUCTION

The healthcare industry is producing copious amounts of data which need to be mined in order to discover hidden information for effective prediction, diagnosis and decision making. Currently, kidney disease has been a crucial problem. It is one of the leading causes of death in India. Chronic kidney disease(CKD), is delineated by the gradual loss of kidney function. Kidneys filter wastes and excess fluids from your blood, which are then excreted in your urine. If this disease gets worse, wastes can accumulate in the blood and can cause difficulties like high blood pressure, anemia, weakening of bones, poor nutritional health and nerve damage. Also, kidney disease increases the risk of having heart and blood vessel disease.

The harmful outcomes can be avoided and prevented by early detections, according to researchers conducted. Awareness of CKD among patients is gradually increasing, but still low. The Global Burden of Disease (GBD) 2015 ranks chronic kidney disease as the eighth leading cause of death in India. All over the world, the highest count of patient with diabetes is in India with the projection figure of 57.2 million cases in 2025and also the count of patient with hypertension is expected to double from 2000 to 2025,hence these will make India the reservoir of CKD [1].The burden of CKD management thus falls largely on primary care providers (PCPs). Hence an accurate, convenient, and automated CKD detection method is important for clinical practice

Undiagnosed CKD can be identified, predicting the likelihood that patients will develop chronic disease, and present patient-specific prevention interventions with Machine learning techniques. Accurate predictive models can be created by health systems, which lower risks and eventually improve standards. The data mining techniques of classification, clustering and association helps in extracting knowledge from large amount of data. Machine learning and data mining techniques together have been the prime factors in determining and diagnosis of various critical diseases.

Management of diet depends on the current Glomerular Filtration Rate (GFR rate) and the severity of the disease. We will be classifying the disease in five stages- Stage 1, stage 2 and stage 3, Stage 4, Stage 5. Stage 1 is safe and requires a lenient diet plan to be followed. Whereas stage 2, a potential CKD patient will be given a restricted and strict diet. Keeping the balance of minerals, electrolytes, and liquids inside body will be difficult for stage 3 to 5 patient . Therefore, they have to be under proper dietary guidance. An important diet for a renal improvement and prevent further harm is essential, which also helps in keeping balance of electrolytes and water in the body.

Other than stages of severity, many other factors will contribute in shaping the diet. The blood potassium level, urea level, calcium level, phosphorous level and so on. In this study, to identify suitable diet plan for a CKD patient the main focus will be on blood potassium level.

II. LITERATURE SURVEY

Anusorn Charleonnan et al [2] projected that revolves around four classification algorithms which make predictive models for chronic kidney disease. The goal here was to find the best classifier amongst the four: logistic regression, Support Vector

978-1-5386-9166-3/19/$31.00 ©2019 IEEE

Machines, Decision trees classifier and K-nearest neighbors. Chronic kidney disease dataset was used to construct the predictive model and later comparison between their performances was done to find the best classifier amongst these to predict chronic kidney disease.

M.P.N.M. Wickramasinghe et al [3] presented a research study, by fetching data from patient's medical records and then applying classification algorithms on these records, which would in turn give a suitable diet plan to the patients of CKD..

M. Dr. S. Vijayarani et al [4] discussed about a comparison made between two classification algorithms namely Support Vector Machines and Artificial Neural Networks. Based on their respective accuracies and timings, the goal to predict CKD was achieved. The one with higher accuracy and good timing was chosen.

Ms.Astha Ameta et al [5] concentrated mainly on data mining techniques and ways by which it could predict chronic kidney diseases. Thus they made it clear that data mining was a more efficient tool to predict chronic kidney diseases.

S.Ramya et al [6] ,here the objective was to overcome leading time in diagnosis and also improve its accuracy. The patient's medical reports were classified using algorithms: Back-propagation neural network, Random Forest, Radial Basis Function. According to experimental result, Radial Basic Function was founded to be more accurate method than the others. It also showed that the classification of various stages was according to the severity the disease was in.

III. SYSTEM DESIGN

A. Data Collection

The dataset collected is real time data obtained from UCI machine learning repository [7] and from various hospital in Mumbai .The dataset has 1000 instances and 25 attributes .It has two types of attributes which are 11 numeric and 14 nominal attributes. As we are using machine learning techniques the dataset will be divided into two set one for training and testing data.

B. Data Pre-processing

Data pre-processing is a way to convert the noisy and huge data into relevant and clean data, as the data available is Real world data, so it contains inaccurate data, missing values and other noisy data, for removing this inconsistent data from the dataset, the proposed system has to clean the raw data. This is an important part to complete the prediction model. It reduces the dimensionality and helps the machine to achieve better results.

Following data pre-processing steps are followed:

1) Finding Missing Values

When the data collected is real world data, and then it will contain missing values. This brings more change in the prediction accuracy. An efficient way to handle missing values is to use mean, average of the observed attribute or value. This way we lead to more genuine data and better prediction results.

2) Data Transformation

In this step we transform the given real data into required format. The data downloaded consist of Nominal, Real and Decimal values. In this step we convert the Nominal data into numerical data of the form 0 and 1(yes/1 or No/0). Now the resultant csv file comprises of all the integer and decimal values for different CKD related attributes.

C. Adding new attributes

We are adding GFR attribute (using MDRD equation) to determine which stage of CKD patient has and also deriving a new attribute called ZONE on the basis of potassium level which will be helpful for diet recommendation module. we define 3 zones on basis of blood potassium level they are as follows

- SAFE ZONE: - Blood potassium level should be between 3.5 -5.0.
- CAUTION ZONE: - Blood potassium level should be between 5.1-6.0.
- DANGER ZONE: - Blood potassium level should be above 6.1.

MDRD equation for GFR calculation are as follows:

$GFR = 175 \times (S_{cr})^{-1.154} \times (Age)^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if African American})$

Where,

GFR (glomerular filtration rate) = mL/min/1.73 m^2

S_{cr} (standardized serum creatinine) = mg/dL

Age = years

D. Prediction module

1) Feature selection

In this step we select subset of relevant attributes from the total given attributes. This stage helps in reducing the dimensionality and making the model simpler and easy to use, thus leading to short training time and high accuracy. To obtain highly dependent features for CKD prediction we will either use correlation or Regression to form.

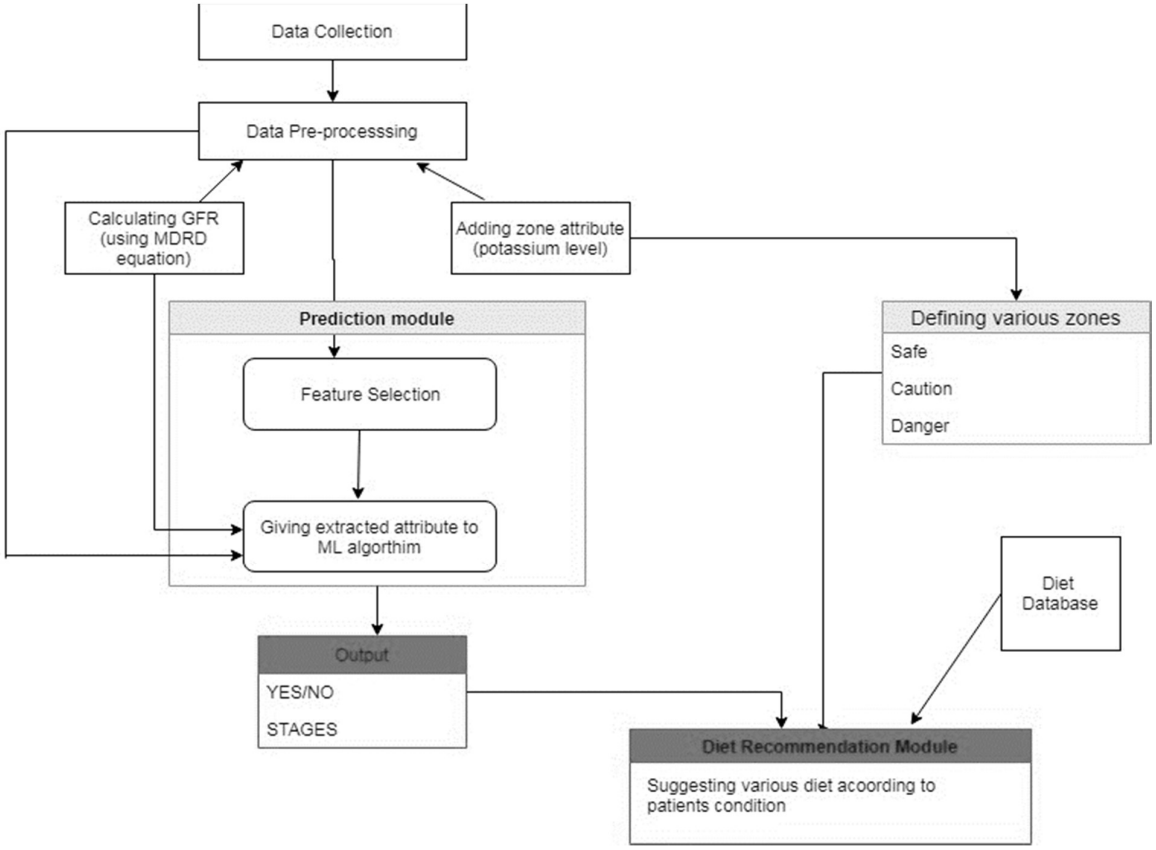

Fig. 1. System design

2) Prediction algorithm

Chronic kidney disease patient usually does not have much symptoms or health issues and even most of CKD patients didn't even know they have CKD or not during stage 1&2. So early prediction of CKD will help patient and doctors to deal with proper medication at early stage from getting condition worse.

First the selected subset of attributes which was extracted from 25 attributes along with GFR attribute will given input to machine learning algorithm for training purpose. Once the training of predictive model is done, we will test the model using another set of data to check whether the same result is obtained as in training phase. After achieving expected accuracy, we will test predictive model but this time we will eliminate the GFR attribute in order to that the algorithm will predict CKD or non CKD along with the CKD stage patient is in.

E. Diet recommendation module

As dietary management plays a important role during chronic kidney disease to slow down the progression of CKD. Mostly patients with diabetes and high blood pressure conditions should a very strict diet to prevent from kidney failure.

So in this module, based on zone detected (using potassium level) and output from prediction module a patient will be given suitable diet which will fetched from diet database. Also CKD patient with high blood pressure condition will be given an alert to lower down sodium intake.

IV. CONCLUSION

This paper elaborates the proposed system consisting of 4 main modules, which are data preprocessing, feature extraction, defining zones based on blood potassium level, diet recommendation module. Dataset has 25 main features, appropriate statistical method like regression extracts highly effective parameters in the decision of CKD detection. Main aim of machine learning module is to label the patient's CKD status and categories them into various stages. Diet recommendation module is purely based on blood potassium level. Overall this system detects and suggest diet which will be useful to the doctors as well as patients

V. REFERENCES

[1] M. R. a. M. Dabhi, "Burden of disease – prevalence and incidence," *Clinical Nephrology,* vol. 74, pp. 9-12, /2010.

[2] T. F. T. N. W. C. Anusorn Charleonnan, "Predictive Analytics for Chronic Kidney Disease," in *The 2016 Management and Innovation Technology International Conference (MITiCON-2016)*, 2016.

[3] D. P. a. K. K. M.P.N.M. Wickramasinghe, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," in *2017 IEEE Life Sciences Conference (LSC)*, Sydney, NSW, Australia, 2017.

[4] M. Dr. S. Vijayarani, "data mining classification algorithma for kidney disease prediction," *International Journal on Cybernetics & Informatics (IJCI),* vol. 4, no. 4, p. 13, August 2015.

[5] M. K. J. Ms. Astha Ameta, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review," *International Journal Of Engineering And Computer Science ,* vol. 4, no. 2, p. 20376, february 2017.

[6] D. N. S.Ramya, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 4, no. 1, pp. 812-813, january 2016.

[7] Dr.P.Soundarapandian, "Chronic_Kidney_Disease Data Set," Apollo Hospitals, 03 july 2015. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease. [Accessed 215].