Web-Based Solution for Effective Management ofChronic Kidney Disease

Tejas Parandekar[1], Shubham Kature[2]

[1]*Student, Computer Engineering, PVG's COET and GKPIM, Pune*
[2] *Student, Computer Engineering, PVG's COET and GKPIM, Pune*

Abstract - *Chronic kidney disease is a perennial condition in medical field that has no cure. It is a progressive disease that occurs in the general adult population, especially in people with diabetes and hypertension. Based on the preliminary available data, chronic kidney disease seems to be associated with enhanced risk of COVID-19 infection. It is important to identify the disease at the early stage to prevent or cure the disease. The main objective of this research work is to create a web-based system that can be used by doctor as well as patient to detect the disease in early stages. The focus of research is concentrated on building a system for effective management of the disease. Research also shows the practical aspects of data collection and attribute selection in machine learning algorithms. Random Forest algorithm is used for the classification.*

Key Words: Chronic kidney disease, Machine Learning, Random Forest, Prediction, Web.

1.INTRODUCTION

Chronic kidney disease (CKD) is a critical health condition in medical field that involves a gradual loss of kidney function. The term "chronic" explains slow degradation of the kidney cells which continues over a long period of time. Main function of kidney act as filtration system for blood and to remove toxins from body. Many people in early stages shows no symptoms and thus most of the cases are diagnosed in the advanced stage. With help of machine learning it is possible to detect the disease in its early stage. Each stage of a chronic kidney disease refers to the performance of the kidneys. The stages of chronic kidney disease ranges from stage 1 to stage 5. In the earlystages of 1 to 3, kidneys can still filter out the waste of ourblood but in the later stages of 4 and 5, kidneys need to work harder for the filtration process. Chronic Kidney Disease gets worse with time and in the worst case the kidneys may stop working. Thus, it becomes very important to detect the chronic kidney disease in its earlystage as possible and provide an efficient and effective way for the disease management.

The main objective of this paper is to create a web-based system for early detection of chronic kidney disease according to stages using classification algorithm. Section2 shows the proposed system, CKD dataset, Feature selection and algorithm analysis. Section3 shows conclusions.

2. Proposed System

In our proposed system we try to provide a system where the effective management of the disease can be done.

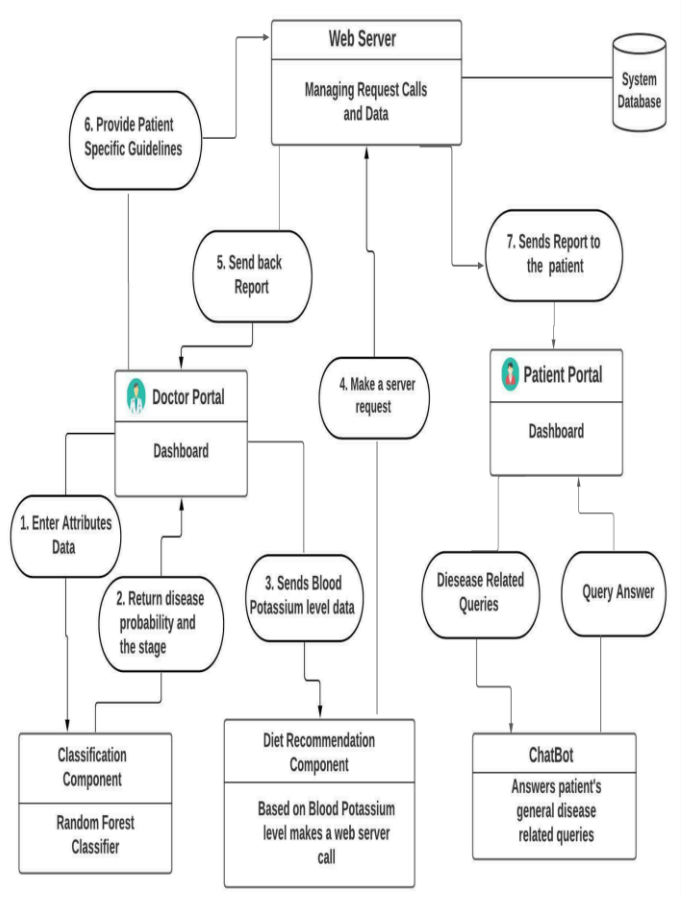

Fig -1: System Architecture

2.1 Dataset and Feature Selection

The dataset required for the chronic kidney disease prediction was collected from UCI repository of machine learning. The metadata for the dataset is:

Data Set Characteristics: Multivariate
Attribute Characteristics: Real
Number of Instances: 400
Number of Attributes: 25
Attribute Information:

Total 24 attributes + class = 25 (11 numeric ,14 nominal)

1.Age(numerical) age in years
2.Blood Pressure(numerical) bp in mm/Hg
3. Specific Gravity(nominal) sg - (1.005,1.010,1.015,1.020,1.025)
4.Albumin(nominal) al - (0,1,2,3,4,5)
5. Sugar(nominal) su - (0,1,2,3,4,5)
6. Red Blood Cells(nominal) rbc - (normal, abnormal)
7. Pus Cell (nominal) pc - (normal, abnormal)
8. Pus Cell clumps(nominal) pcc - (present,notpresent)
9. Bacteria(nominal) ba - (present,notpresent)
10. Blood Glucose Random(numerical) bgr in mgs/dl
11.Blood Urea(numerical) bu in mgs/dl
12.Serum Creatinine(numerical) sc in mgs/dl
13.Sodium(numerical) sod in mEq/L
14.Potassium(numerical) pot in mEq/L
15.Hemoglobin(numerical) hemo in gms
16.Packed Cell Volume(numerical)
17.White Blood Cell Count(numerical) wc in cells/cumm
18.Red Blood Cell Count(numerical) rc in millions/cmm
19. Hypertension(nominal) htn - (yes, no)
20. Diabetes Mellitus(nominal) dm - (yes, no)
21.Coronary Artery Disease(nominal) cad - (yes, no)
22.Appetite(nominal) appet - (good, poor)
23.Pedal Edema(nominal) pe - (yes, no)
24.Anemia(nominal) ane - (yes, no)
25.Class (nominal) class - (ckd,notckd)

The dataset used for the classification of chronic kidney disease contains 24 features along with the class object, a process of feature selection was applied on the dataset to extract the most important features with which we can successfully classify if the patient is suffering from chronic kidney disease or not. The process used for feature selection is the wrapping method which follows brute force approach. After applying feature selection, the best predictive subset of features needed for the classification were specific gravity, albumin, hypertension, serum creatinine, hemoglobin, packed cell volume, red blood cell count and thus the number of features required for the classification were reduced to 7.

2.2 Algorithm Analysis

The algorithm we trained to predict the probability of patient suffering from chronic kidney disease is Random Forest Classifier. Random Forest contains various decision trees on various dataset subset and it takes average of these decision trees to improve the predictive accuracy of the dataset. The dataset used for prediction contains some amount of missing data and random forest can maintain accuracy even where the dataset contains missing values. The algorithm is also robust to the overfitting issue. Therefore, the features data of specific gravity, albumin, hypertension, serum creatinine, hemoglobin, packed cell volume, red blood cell count, which were selected from feature selection process were provided as a training set to the Random Forest Classifier. The algorithm analysis is done on the metrics of Sensitivity, Specificity and Accuracy.

$$\text{Sensitivity: } \frac{TP}{TP+FN}$$

$$\text{Specificity: } \frac{TN}{TN+FP}$$

$$\text{Accuracy: } \frac{TP+TN}{TP+TN+FP+FN}$$

Here, TP: True Positive, TN: True Negative FP: False Positive, FN: False Negative

Based on the accuracy metric our selected algorithm provided accuracy of 98%.

2.3 System Architecture

1. **Doctor Portal**: The doctor portal consists of a dashboard so that the doctor can enter attributes data of the patient and based on the input data our trained model can predict whether the patient is suffering from chronic kidney disease or not

    a. **Classification Component**: As mentioned in the section of algorithm analysis, the

algorithm we trained to predict the probability of patient suffering from chronic kidney disease is Random Forest Classifier. According to the feature selection process the best predictive subset of features are specific gravity, albumin, hypertension, serum creatinine, hemoglobin, packed cell volume and red blood cell count. Based on these values our trained classifier is able to predict the probability of patient suffering from chronic kidney disease and also the stage of the disease.

b. **Diet Recommendation Component**: When the doctor gets the stage of chronic kidney disease from the classifier component then that stage is sent to the diet recommendation component. Doctor also provides patients potassium level to this component. This component makes a specific server call to the web server based on the Blood Potassium level. Therefore, based on this Blood Potassium level of the patient a zone is specified to the patient.

- If the blood potassium level is in between 3.5 - 5.0 the patient is inthe SAFE zone.
- If the blood potassium level is in between 5.1 - 6.0 the patient is inthe CAUTION zone.
- If the blood potassium level is higher than 6.1 the patient is in the DANGER zone.

The data required for diet recommendation was obtained from research done by Anonnya Banerjee [1] and others.

c. **Report Generation Component**: Based on the request call made from the Diet Recommendation Module, the web server extracts the recommended diet stored in the database for that zone. Then a generalized report with the recommended diet is send to the doctor. The doctor will go through the report and add his or her important guidelines for the patient. Then the doctor will send the report to thatspecific patient.

2. **Patient Portal**: It contains a dashboard that displays the report. The patient can see the reportand doctor's suggested guidelines here.

    a. **Chatbot Component**: It may happen that doctor might not be available for patient queries so we provided a chatbot system which can assist patient and provide general information about the chronic kidney disease and its stages, what factors may lead to the disease and some general queries related assistance.

3. CONCLUSIONS

In our proposed system we proposed a way where the management of chronic kidney disease can be done in a systematic and in an effective way. The trained random forest classifier was able to classify the probability of patient suffering from chronic kidney disease or not with an accuracy of 98 percent. The process of feature selection wasable to identify the best predictive subset of features which includes specific gravity, albumin, hypertension, serum creatinine, hemoglobin, packed cell volume and red blood cell count.

REFERENCES

[1] Anonnya Banerjee, Alaa Noor, Nasrin Siddiqua, Mohammed Nazim Uddin "Food Recommendation usingMachine Learning for Chronic Kidney Disease Patients" in 2019 International Conference on Computer Communication and Informatics.