INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1351)
Available online at: https://www.ijariit.com

Prediction of chronic kidney disease and diet recommendation

Annapoorna B. A.
annapooranaba.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Nisarga Y. N.
nisargayn.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Rachana R. Shastry
rachanarshastry.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Sreelatha P. K.
sreelatha.pk@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

ABSTRACT

Chronic renal disorder is that the sort of disease within which there's a decrease in kidney function over a period of months or years. Early prediction of CKD is one in all the main problem in medical fields. So automated tools which use machine learning techniques determine the patient's kidney condition which will be helpful to the doctors in prediction of disease. Our system retrieves the features which are significantly affects the human with CKD, and so the ML technique which automates the classification of the disease into different stages. Our main goal is to predict the disease stage and suggest suitable diet for CKD patients using classification algorithms on medical test records. Diet recommendations for patients are going to be given per the potassium zone which is calculated using blood potassium level to weigh down the progression of CKD.

Keywords: *Chronic Kidney Disease, Glomerular Filtration Rate, Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbour Classifier*

1. INTRODUCTION

Chronic Kidney Disease (CKD) is a dangerous health issue due to its expensive treatment there is a possibility of death rate is high. CKD is a type of kidney disease caused due to the damage to both the kidneys and it is being revealed by the abnormal excretion of albumin or decrease in the kidney function. It is a long-term disorder. There is no cause and the damage caused to the kidneys is permanent which can lead to ill health. In few cases, dialysis or transplantation may be helpful and essential. Chronic Kidney Disease is basically found frequently in old people and it seems to increase in the population in a large volume. CKD is basically defined as illness or the presence of kidney damage, which is revealed by the excretion of abnormal albumin or decrease in the kidney function.

The Chronic Kidney Disease is being divided into 5 stages on the basis of defined range of Glomerular Filtration Rate. (GFR): - CKD 1, CKD 2, CD 3a, CKD 3b, CKD 4, CKD 5. In order to avoid the deaths, patients having CKD at 5th stage have to undergo kidney transplantation or dialysis. In particular, the treatment of kidney transplantation provides a great possibility to survive. The issue is that only few will have a chance to undergo this treatment due to a huge waiting list. Due to which many patients have taken the alternative of undergoing dialysis treatment. There are two types of dialysis treatment – Haemodialysis(HD) and Peritoneal dialysis(PD). Both these types are created on the basis of same principles: solute diffusion and fluid ultra - filtration. HD treatment is performed in the clinic via machine and in PD treatment blood inside the body is cleaned which is done at patient's home on the basis of the natural tendency of progression from the stages of kidney disease 1 to 5, where patients must frequently consult the doctor for various suggestions in order to maintain kidney health.

Since the number of patients, and the total information about each patient is large and also it keeps increasing, the doctors and the medical staff face difficulty in handling the personalized data and treatment plans. The disease trend, especially the progression patterns are very useful as decision making support tool. The current study uses machine learning technique which develops a classification model capable of predicting chronic kidney disease stages 1 to 5 and also suggests a suitable diet on the basis of the patient's condition.

2.LITERATURE SURVEY

By the following survey, the classification techniques such as Naïve Bayes, KNN, Random Forest and Decision Tree algorithms are used to predict the stages of the disease. Apart from the selection of classifiers, several components which concentrates on influencing factors related to the kidney

diseases are Hypertension, Diabetes, Smoking, Obesity, Heart diseases, alcohol intake, drug overdose, family history of kidney disease, age, gender and ethnicity / race. During the training of the prediction model all the parameters related are being considered to classify the different stages. Below given are the previously used procedures and conclusions drawn from them.

1.Imesh Udara Ekanayake and Damayanthi Herath [4], proposed an approach using Machine Learning techniques in the year 2020. The tree structure algorithms are unstable and small change in the data can lead to a large change in the result.

Procedures used - Machine Learning algorithm techniques
Algorithms – Logistic Regression, KNN, SVC Gaussian, Decision Tree, Random Forest, XGB
Conclusion – Filling missing values based on the distribution of them along with the collocation of other attributes by KNN imputer instead of replacing with a constant directly leads to work done using some dataset. This suggests new workflow including data pre-processing, missing values handling and feature selection to predict CKD status as positive or negative

2.Devika R, Sai Vaishnavi and V Subramaniyaswamy [3] proposed the idea of using different algorithms and comparing them in the year 2019. Random Forest algorithm is complex and it consumes time.
Procedures used - Machine Learning Techniques
Algorithms - Naïve Bayes, KNN, Random Forest
Conclusion – In this paper, we can compare the overall performances of the used classifiers with the other current classifiers. New classifiers can be used and their performances can be evaluated to locate the higher solutions of the objective feature in destiny paintings.

3. Akash Maurya, Rahul Wable, Rasika Shinde, Sebin John, Rahul Jadhav [1] proposed an approach using Machine learning techniques where proper diet is recommended for the patients having CKD.
Procedures used – Machine Learning Techniques
Algorithm used – Prediction algorithm
Conclusion – The diet recommendation model is purely based on blood potassium level. The system predicts and suggests diet to the patients.

4. M. Dogruyol Basar, A. Akan [5] proposed an approach where classification techniques considered in this paper can be used and evaluated to find rapid solutions for the patient. The main aim of this study is to reduce the number of classifiers used so that CKD can be diagnosed efficiently and rapidly. And Rep Tree and subspaces classifier and Naïve Bayes algorithm is used for the best results.
Procedures used - Reduced Individual Classifier
Algorithms - Random Tree, REP Tree, Naïve Bayes
Conclusion - The best results are obtained from REPTree and Random Subspaces classifiers as 99.17%. It was shown that Random Subspace technique has the highest accuracy and kappa values in every reduced type of features. Classification techniques considered can be used and evaluated to find the rapid solutions for patients. The main objective of this work is to reduce the number of classifiers being used so that CKD can be diagnosed rapidly and efficiently.

5.Dr. R.Thirumalaiselvi,.S. Dilli Arasu [6] proposed an approach whose goal is to analyze the different data mining techniques in medical domain and some of the algorithms used to predict kidney disease. They use Data Mining techniques to predict the kidney diseases and SVM is used as a classifier and also C4.5 algorithm is used. Procedures used - Data Mining Techniques
Algorithms - SVM
Conclusion - This paper proves that results may vary for different stages of kidney disease diagnosis based on the techniques and the tools being used. Data mining provides better results in disease diagnosis when appropriate techniques are used. Thus, data mining is the significant field for healthcare predictions.

6.S. Ramya, Dr. N. Radha, [7] proposed an approach using Machine Learning algorithms in the year 2016. Attribute used is GFR for prediction of kidney diseases.
Procedures used - Machine Learning Techniques
Algorithms - BPN, RBF, RF
Conclusion - The models are evaluated with four different measures like Kappa, Accuracy, Sensitivity and Specificity. From the experimental results, the Radial Basis Function (RBF) yields a better accuracy for predicting CKD and it attains the accuracy of 85.3%.

7.Ashfaq Ahmed, K Aljahadali, S Hussain, S.N [2] proposed an approach using support vector machine and random forest classification techniques in the year 2016.
Procedures used - Classification Techniques
Algorithms - SVM, Random Forest
Conclusion - It is concluded that the varying results are observed with SVM classification technique with different kernel functions. The Random Forest also yields results comparable with parameter tuned SVM results. The results can be better analyzed with confusion matrix. This can be further extended with other new kernel functions and other classification techniques

8.P Swathi Baby and Panduranga Vital, [8] proposed an approach using Machine Learning Algorithms like AD Trees, J48, KStar, Naïve Bayes, Random Forest for prediction of kidney disease. It shows that Naïve Bayes Algorithm obtained the highest 100 percent accuracy.
Procedures used - Machine Learning Techniques
Algorithms - AD Trees, J48, KStar, Naïve Bayes, Random Forest
Conclusion – The datasets are collected from various hospitals are being processed through data mining techniques tool such as Weka and Orange. The machine learning algorithms are used for the performance study of each algorithm which gives the statistical analysis and predicting kidney diseases using these algorithms.

3. DESIGN

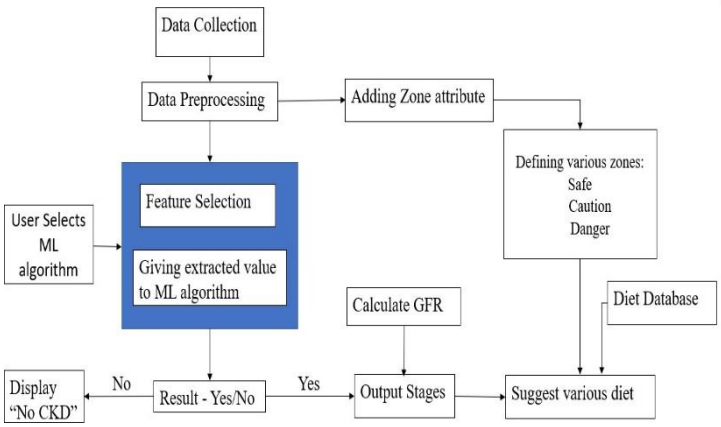

Fig 1: System architecture

3.1 Data Collection

Dataset is obtained from UCI machine learning repository and is real time data. Dataset has 25 attributes and 400 instances which includes nominal and numeric data. Since machine learning techniques are used, dataset will be divided into two sets (training data-67%, testing data-33%).

3.2 Data Pre-processing

As data collected is real-time data, it contains noisy and inaccurate data. Role of data pre-processing is to clean these raw data. This process is used to convert huge and noisy data into clean and relevant data. This procedure is important to complete prediction model. This process includes 2 steps:

- Removing null values
- Data transformation

3.3 Prediction

This module has 4 sub modules

(a) Selection of algorithm: We have implemented four different types of algorithms which include Naïve Bayes, KNN, Decision tree and Random Forest. User can select any of these algorithms to predict the stage.

(b) Feature selection: From whole set of attributes, relevant attributes are selected. From 24 attributes, 21 attributes are extracted. Feature selection helps to make model simpler and easy to use by reducing the dimensionality. It gives high accuracy in short training time.

(c) Prediction algorithm: During early stages (1 and 2), most of the patients do not have many symptoms, So, the doctors can deal with proper medication if CKD is predicted early. Subset of attributes obtained from feature selection will be given as input to the algorithm for training. After the process of training, model is tested to check whether same result is obtained as in training phase. Finally, result is displayed either as yes (CKD detected) or no (no CKD).

(d) Adding new attributes: If the predicted result is yes, a new attribute called GFR (Glomerular Filtration Rate) is added to determine the stage. Formula to calculate GFR is as follows:

$$GFR\ (female) = 175 * (SCR) - 1.154 * (Age) - 0.203 * (0.742)$$

$$GFR\ (African\ American) = 175 * SCR - 1.154 * (Age) - 0.203 * (1.212)$$

Where: SCR stands for Standardized Serum Creatinine.

A new attribute called ZONE is derived on basis of blood potassium level. ZONE attribute helps in recommending diet. There are 3 levels of zone as follows:

Safe zone: 3.5 – 5.0
Caution zone: 5.1 – 6.0
Danger zone: > 6.1

3.4 Diet Recommendation

Diet recommendation plays very important role for slowing down progression of CKD. Patients with critical conditions such as high BP, diabetes must follow strict diet to prevent kidney failure. Based on the ZONE detected and output from prediction model, patient will be recommended suitable diet. Food items for the diet recommendation is fetched from diet database. Diet database consists of 4 attributes and 198 instances. KNN algorithm is used in this module.

Attributes used: Food items, potassium, quantity, category (high, medium, low)

4. RESULTS

Final results are categorized into 2. Stage of the CKD patient and suitable diet recommendation. Total number of algorithms used in this project are 4. Each algorithm produces different accuracy.

4.1 Output

Output is categorized into 3.

- Prediction of CKD
- Stage of CKD patient
- Diet recommendation

4.1.1 Prediction of CKD: Output of this will be either yes or no. If yes is predicted, then it will display stage. If no is predicted, then it is terminated.

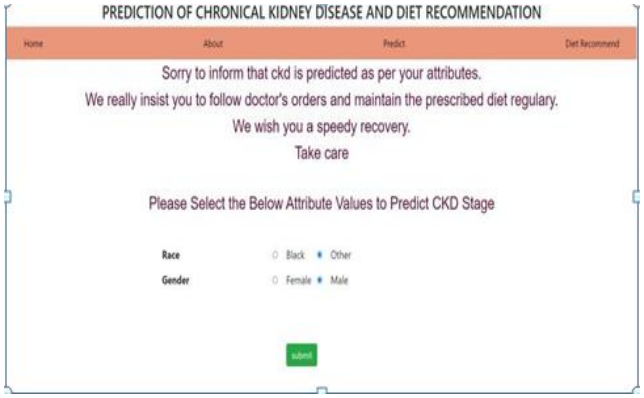

Fig 2: Prediction of CKD

4.1.2 Stage of CKD patient: Once CKD is predicted, stage is displayed after addition of two more attributes (race, gender). Total number of stages are 6: (1, 2, 3a, 3b, 4, 5).

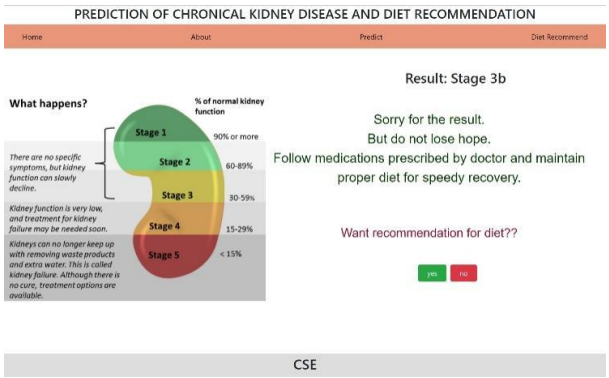

Fig 3: Stage of CKD

4.1.3 Diet Recommendation: Diet recommendation is based on zone attributes. There are 3 zones (safe, caution and danger). Based on these zones, suitable diet is recommended.

Recommended Food Items are

Food	Quantity
Pinto beans	1/2 cup
Lentils beans	1/2 cup
Cheese burger, plain	1
Pomegranate	1 medium size
Complete bran flakes	1 cup
Oat bran muffin	small
Passion fruit juice, purple/yellow	1/2 cup
Avocado	1/2 cup
Condensed milk	1/2 cup
French fries	1/2 cup
Broccoli, cooked	1/2 cup
Chicken vegetable soup	1 cup
BBQ potato chips	1 oz bag
Plain potato chips	1 oz bag
Orange juice	1/2 cup
Prune juice	1/2 cup

Fig 4: Diet recommendation

4.2 Accuracy

Table 1: Accuracy table

Classifier	Accuracy	Precision	Recall	F1-score
Naïve Bayes	93	0.95	0.94	0.94
KNN	84	0.87	0.89	0.88
Decision tree	96	0.99	0.98	0.98
Random Forest	98	1.00	1.00	0.99

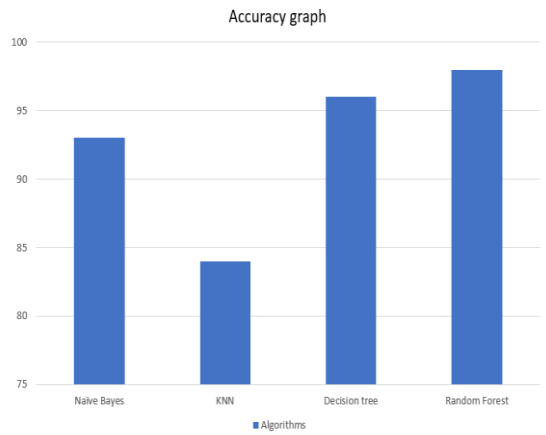

Fig 5: Accuracy Graph

From Table 1 and Fig 5, it is observed that Random Forest gives highest accuracy.

5. CONCLUSION

This system predicts transitional interval of kidney disease from stages 1 to 5 using Machine Learning algorithm and suggests suitable diet according to the patient condition. For classification, user can use Naïve Bayes/Random Forest/KNN/Decision tree classifier which helps to identify the disease and provide guidance for decision makers regarding kidney disease stages for further treatment. From the analysis of all the algorithms, Random Forest gave the highest accuracy.

6. REFERENCES

[1] Akash Maurya Rahul Wable "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning "2019 International Conference on Nascent Technologies in Engineering (ICNTE 2019).

[2] Ashfaq Ahmed, K., Aljahdali, S., Hussain, S.N "Comparative prediction performance with support vector machine and random forest classification techniques" International Journal of Computer Applications, 2016.

[3] Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest" International Conference on Computing Methodologies and Communication (ICCMC 2019)

[4] Imesh Udara Ekanayake and Damayanthi Herath, "Chronic kidney Disease Prediction using Machine Learning Techniques", International journal of Applied, 2020.

[5] M. Dogruyol Basar, A. Akan "Chronic kidney disease prediction with reduced individual classifiers "Electrica 2018, 18 Research article, 2018.

[6] S.Dilli Arasu, Dr. R.Thirumalaiselvi "Review of chronic kidney disease based on data mining techniques" International Journal of Applied Engineering Research, 2017.

[7] S. Ramya, Dr. N. Radha, "Diagnosis of chronic kidney disease using machine learning algorithms", International Journal of Innovative Research in Computer and Communication Engineering, Vol 4, issue 1, January 2016.

[8] Swathi Baby P and Panduranga Vital T, "Statistical analysis and predicting kidney diseases using machine learning algorithms", International Journal of Engineering Research & Technology (IJERT), 2015.

[9] "Potassium and Your CKD Diet", The National Kidney Foundation. [Online]Available: https://www.kidney.org/atoz/content/potassium [Accessed: 01- Jul - 2021].

[10] "UCI Machine Learning Repository: Chronic_Kidney_Disease Data Set", Archive.ics.uci.edu, 2015. [Online]Available: http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease [Accessed: 01- Jul - 2021].

