Prediction of chronic kidney disease and diet recommendation

Annapoornna B. A.
annapoornab17cs@saividya.ac.in
Sai Vidy Institute of Technology,
Bangalore, Karnataka

Nisarga Y. N.
nisargayn17cs@saividya.ac.in
Sai Vidy Institute of Technology,
Bangalore, Karnataka

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

Sreelatha P. K.
sreelatha.pk@saividya.ac.in
Sai Vidy 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, Naïve 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.

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
International Journal of Advance Research, Ideas and Innovations in Technology

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


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.

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.

5. 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.


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%.


8. Dilli Arasu [6] proposed an approach using Machine Learning algorithms where proper diet is recommended for the patients having CKD.

Conclusion – The diet recommendation model is purely based on blood potassium level. The system predicts and suggests diet to the patients.

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:
  \[
  \text{GFR (female)} = 175 \times (\text{SCR}) - 1.154 \times (\text{Age}) - 0.203 \times (0.742) \\
  \text{GFR (African American)} = 175 \times \text{SCR} - 1.154 \times (\text{Age}) - 0.203 \times (1.212)
  \]
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.

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).

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.
4.2 Accuracy

Table 1: Accuracy table

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>93</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>KNN</td>
<td>84</td>
<td>0.87</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Decision tree</td>
<td>96</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Random Forest</td>
<td>98</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
</tr>
</tbody>
</table>

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 Naive 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