

## LITERATURE SURVEY

# EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

### INTRODUCTION:

Chronic kidney Disease (CKD) means your kidneys are damaged and not filtering your blood the way it should. The primary role of kidneys is to filter extra water and waste from your blood to produce urine and if the person has suffered from CKD, it means that wastes are collected in the body. This disease is chronic because of the damage gradually over a long period. It is flatterer a common disease worldwide. Due to CKD may have some health troubles.

There are many causes for CKD like diabetes, high blood pressure, heart disease. Along with these critical diseases, CKD also depends on age and gender. If your kidney is not working, then you may notice one or more symptoms like abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, vomiting.

There are two main diseases of CKD:

- (i) Diabetes
- (ii) High blood pressure. So that controlling of these two diseases is the prevention of CKD. Usually, CKD does not give any sign till kidney is damaged badly. CKD is being increased rapidly as per the studies hospitalization cases increase 6.23 per cent per year but the global mortality rate remains fixed.

There are few diagnostic tests to check the condition of CKD:

- (i) Estimated Glomerular filtration rate
- (ii) Urine test
- (iii) Blood pressure.

### AUTHORS AND THEIR PROPOSAL :

#### 1. Author Name: Tomas E

The small dataset of 400 records have been collected from Apollo Hospital, India in 2015 taken over a two-month period. ANOVA test, the Pearson's correlation, and the Cramer's V test are applied and removed the redundant features in dataset. By using filter feature selection method, Three features (hemoglobin, albumin, and specific gravity) are selected and trained using Logistic regression, support vector machines, random forest, and gradient boosting algorithm and reached an accuracy of 99.1% accuracy. Data used in this research is small. So, in future, need to validate the results by using big dataset and for reducing the prevalence of CKD, planned to predict if a person with CKD risk factors such as

diabetes, hypertension, and family history of kidney failure will have CKD in the future or not by using appropriate dataset.

## **2.Author Name: Ebrahime Mohammed Senan, Nizar Alsharif**

The dataset was collected from 400 patients containing 24 features. The dataset was divided into 75% training and 25% testing and validation. The dataset was processed to remove outliers and replace missing numerical and nominal values using mean and mode statistical measures respectively. The RFE algorithm was applied to select the most strongly representative features of CKD. Then Random forest algorithm is applied. Data used in this research is small. So, in future, need to validate the results by using big dataset.

## **3.Author Name: Bidri Deepik, Vasudeva Rao KR, Dharmaj N Rampure, Prajwal P and Devanand Gowda G**

Early Prediction of Chronic Kidney Disease by using Naive bayes, K-Nearest neighbor. KNN algorithm takes CKD parameters as input and predicts the disease based on old CKD patient's data. To develop user interface for create easier communication between doctors and patients.

## **4.Author Name: Sreeji S, Balamurugan Balusamy**

In Design System For Early Detection And Prediction Of Chronic Kidney Disease Using Machine Learning Techniques, The SVM, and Naive Bayes classification algorithms were applied on the processed data. The information assortment contains 400 patient records, for certain qualities missing. It comprises of 24 clinical highlights that show up in the anticipation of ongoing kidney illness, with one class quality demonstrating the event of persistent renal disappointment in the patient. Data used in this research is small. In future by using better algorithms the accuracy can be increased and also need to validate the results by using big dataset.

## **5.Author Name: Reshma S , Salma Shaji , S R Ajina, Vishnu Priya S R , Janisha A**

In this research, The dataset contains 400 samples of two different classes. Out of 25 attributes, 11 are numeric and 13 are nominal and one is class attribute. The data set contains number of missing values. Here the information of dataset uses the patient's data like age, blood pressure, specific gravity, albumin, sugar, red blood cells etc. The sample Data used in this research is small. So, in future by using by using better deep learning concepts the accuracy can be increased. Random forest algorithm outperformed all other applied algorithms.

## **6.Author Name: Hamida Ilyas, Sajid Ali, Mahvish Ponum**

In this study for predicting the various stages of CKD using machine learning classification algorithms on the dataset obtained from the medical records of affected people. Specifically, they have used the Random Forest and J48 algorithms to obtain a sustainable

and practicable model to detect various stages of CKD with comprehensive medical accuracy. Comparative analysis of the results revealed that J48 predicted CKD in all stages better than random forest with an accuracy of 85.5%. The study also showed that J48 shows improved performance over Random Forest

#### **7.Author Name: Zixian Wang, Jae Won Chung**

In this research, CKD is detected using the Apriori association technique for 400 instances of chronic kidney patients with 10-fold-cross-validation testing, and the results are compared across a number of classification algorithms including ZeroR, OneR, naive Bayes, J48, and IBk (k-nearest-neighbor). The dataset is preprocessed by completing and normalizing missing data. The most relevant features are selected from the dataset for improved accuracy and reduced training time. The results for selected features of the dataset indicate 99% detection accuracy for CKD based on Apriori. The identified technique is further tested using four patient data samples to predict their CKD. Analysed with different supervised and unsupervised machine learning techniques and feature selection techniques with additional performance metrics need to be used for better CKD prediction.

#### **8.Author Name: Suman Bala, Krishan Kumar**

In the health care industry the data mining is mainly used for predicting the diseases from the datasets. The Data mining classification techniques, namely Decision trees, ANN, Naive Bayes are analyzed on Kidney disease data set. Decision Trees, ANN and Naïve Bayes, Logistic Regression, Genetic Algorithms are applied on processed datasets and detects the kidney disease.