Prediction of Chronic Kidney Disease Using Data Mining

Preethi K Sanil1, Anusha2, Prajna Kotian3, K Shravya4

# Abstract : Machine learning has earned a good position in healthcare center because of its capability to enhancing the disease prediction in healthcare center. Machine Learning and Data Mining techniques are used in healthcare center. Nowadays most dangerous health related problem is kidney disease. It is increasing day by day because of not having proper food habits, drinking less amount of water and lack of health consciousness. So we need some technique to decrease Chronic Kidney Disease. Chronic Kidney disease is a serious problem that is introduced by either kidney pathology or reducing the kidney function. In India according to health station 63538 kidney disease cases has been registered. Usually the kidney disease occurs between the age of 48 to 70. According to survey male having more Chronic Kidney disease than female . In 2015 , India ranks 17th position in CKD. The main focus of work is to analyze the dataset of chronic kidney failure and perform the classification of CKD and non CKD cases. Dataset is preserved to the learning techniques. The training dataset to the selected features is fed into various classifiers to determine which classifier play a accurate role in detection of chronic kidney disease. The proper dataset is classified using various algorithm like Linear Regression, K-NN, Naïve Bayes, Support Vector Machine, Random Forest, ANN,C4.5.

***Index Terms*- Machine Learning, Data Mining, CKD - Chronic Kidney Disease, GFR - Glomerular Filtration Rate ( GFR ) , Decision tree, ANN - Artificial Neural Network, Naïve Bayes Classifier, SVM - Support Vector Machine, Logistic Regression Algorithm, Random Forest , C4.5, KNN - K- Nearest Neighbor, Gradient Classifier.**

1. INTRODUCTION

T

he chronic kidney disease is one of the long-lasting disease of kidney which leads to the failure of renal . Now

a days there are many major health problems in which the chronic kidney disease is one among them. According to survey, many doctors are trying to spend their time just to find whether the patients having kidney disease or not. To overcome this scenario the best machine learning algorithm is used. The algorithm helps to predict kidney disease. Algorithm will give the high accuracy and recall rate. There are many popular algorithms some of them are “K-nearest neighbor,Support Vector Machine, Random Forest Algorithm,

Naive Bayes Algorithm, Decision Tree , Artificial Neural Network , Logistical Regression Classifier, Gradient Classifier” . Kidney disease is the major health problem in many states and it is also identified as the dangerous health problem by government . User expects everything that means better things will be recommended by the system. To make system to be recommendation capable so it must have a capability to take a decision by itself. To take the decision by itself it must need to have classified data. All these made us interested to do such kind of research based work. Our work is mainly dependent on machine learning techniques and data mining procedures .

# 1.1 Chronic Kidney Disease

Unfortunately kidney disease most often doesn’t have any symptoms unless there is a significant decline in kidney function . Initally there may be some weight symptoms like unexplained anemia , feeling weak, tiredness . If such symptoms are found then kidney must be tested whether it is functioning proper or not . Apart from these when kidney functions decline beyond 90% then only symptoms like swelling of feet, swelling of face , nausea which may lead to Chronic kidney disease can be seen . Chronic kidney disease is irreversible problem or damage in the renal function which is developed over a period of years. At initial stage it just seems to be a biochemical abnormality but later on symptoms such as loss of excretory, metabolic functions of kidney leads to the symptoms and signs of renal failure. These conditions are also referred as uraemia. If chronic kidney disease is found in the early stage it can be cured with the medicines else transplantation of kidney must be done or it may lead to death. Chronic kidney disease usually has five stages: that is stage 1 to stage 5.

# Glomerular Filtration Rate(GFR)

Doctors, nurses and pharmacists usually uses the equations to estimate the function of the kidney from the blood creatinine. Creatinine is the unwanted product developed from muscle metabolism and it is excreted from the body by kidneys. If the creatinine level in the blood is high, then efficiency of the kidneys become low[20]. Glomerular Filtration Rate(GFR) is the term used to identify the level of kidney function. Glomerular Filtration Rate shows the rate at which the kidneys filter toxins and unwanted wastes from blood. The stage of kidney disease is based on GFR[5].

The below shows the formula for calculation of GFR:

# 186\*(Creatinine/88.4)-1.154 \*(Age)-0.203 \*(0.742 if female)

**\* (1.210 if black)**

* 1. **Five stage of CKD and GFR for each stage**

Stage 1 : In stage 1 , GFR is ratherly preserved that is >=90. So this can be considered as normal or close to normal . Here the prevalence is about 3.5%[9] . In this stage the evaluation plan is to diagnose and treat cause of slow progression and also evaluate the risk for heart disease.

Stage 2 : In stage 2, GFR rate is slightly reduced between 60 to 89. In this stage it seems to be mild CKD. Here the prevalence is about 3.9%. In this stage 2 evaluation plan is to estimate progression.

Stage 3: Stage 3 is the middle round kidney disease. In stage 3, GFR rate is mildly reduced between 30 to 59 not too low not too high. In this stage you may begin to see the complications of CKD . Here the prevalence is about 7.6%. In this stage 3 evaluation plan is to estimate and treat the complications. Fatigue, urination change , swelling of extremities are some of the symptoms of Stage 3

Stage 4 : In Stage 4 chronic kidney disease , GFR is between 15-29 . In this stage kidney functions and features may halt . Here the prevalence is about 0.4% . In this stage 4 evaluation plan is either preparing for dialysis or creation of access or may be transplantation. Symptoms of stage 4 is similar to that which may come across in stage 3 but here loss of appetite takes place and some more symptoms are:

* + - Urea build up in the blood leads to bad breath.
    - Loss of appetite
    - Loss of focus or concentration
    - Nerve problems.

Stage 5 : This is the most severe stage in chronic kidney disease. In stage 5 CKD, GFR is less than 15. Here the prevalence is about 0.1% . In this stage the evaluation plan is to take dialysis[9] . In this stage there are many symptoms than any other stage. Some of the symptoms such as itching, trouble sleeping , difficult in concentrating, nausea and vomiting also takes place.

1. LITERATURE SURVEY

In [1] **Sahana B J** The goal of this paper is to anticipate Chronic Kidney Disease (CKD) utilizing arrangement methods like Naive Bayes and to foresee the phases of endless kidney illness utilizing the Artificial Neural Network (ANN) . Interminable kidney illness is a condition where kidneys end up plainly harmed and can't channel poisonous materials in our body. This work dominatingly concentrates on identifying life debilitating ailments like Chronic Kidney Disease (CKD) utilizing Classification calculations like Naive Bayes and Artificial Neural Network(ANN) to predicts phases of Chronic kidney disease(CKD). This project is a

medical sector application which helps the medical practitioners in predicting the disease types based on the symptoms. Patients can also predict diseases by entering symptoms in the form of sentences. It is automation for disease prediction and it identifies the disease, its types and complications from the clinical database in an efficient and an economically faster manner.It is successfully accomplished by applying the Naïve Bayes algorithm for classification. This classification technique comes under data mining technology. This algorithm takes symptoms as input and predicts the disease based on old patients data.

In [2] **S.Dilli Arasu,Dr. R.Thirumalaiselvi** This work aims to analyze the various data mining techniques in medical domain and some of the algorithms used to predict kidney diseases eventually. it is proven that results may vary for different stages of kidney disease diagnosis based on the tools and techniques used. Data mining provides better results in disease diagnosis when appropriate techniques used. Thus, data mining is the significant field for healthcare predictions

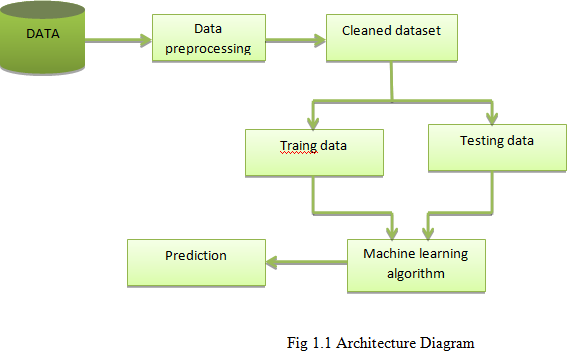
In [3] **El-Houssainy A. Radya, Ayman S. Anwarb** Data Mining was utilized in this study because it is a process of identifying novel, potentially useful, valid and ultimately understandable patterns in data. Supervised and unsupervised learning.Three different types of the most commonly used artificial neural network algorithms and support vector machine algorithms have been used for this study, to determine which algorithm will give the best classification results, so as to identify the stage of chronic kidney disease, based on patient clinical and laboratory data. Finally, as a result the Probabilistic Neural Networks algorithm gives the highest overall classification accuracy percentage of 96.7%, compared to other algorithms in classifying the stages of CKD patients. On the other hand, the Multilayer Perceptron requires a minimum execution time whereas the Probabilistic Neural Network requires less time to finalize the analysis. These algorithms have been compared with classification accuracy based on correctly classified stages of CKD patients, time taken to construct the model, and time taken to test the model. The Probabilistic Neural Networks algorithm yields a better classification accuracy and prediction performance to predict the stages of chronic kidney disease patients.

In [4] **N. Afhami** In this paper ,we predicate the disease progression in patients with diabetic chronic kidney. It has been done based on properties of 249 patients. First stage is input dataset pre-process and in this stage missing value replacing,normalizing and balancing are done. Then in second stage, different techniques applied on dataset. J48, Naïve Bayes, Bayesian Network, SVM, SMO, Random Forest, Bagging, Multilayer perceptron were employed on input dataset. Efficiency of different methods studied and compared using precision, Recall, F-measure,Random Forest had better performance.Also we extract rule for prediction of the disease progression to the last stage of renal disease or mortality.In future works, We can study patient condition and the disease progression, over a period of time.

In [5] **Sai Prasad Potharaju** In this paper, used a rebalancing algorithm called SMOTE to enhance the accuracy of different induction and decision tree models in order to predict kidney disease of patients. the proposed framework improves the accuracy of models by balancing the imbalanced dataset. For this, a technique for sampling the minority class called SMOTE is applied on existing dataset and percentage of variation between classes is minimized. this method can attain the average accuracy of 98.73%.This method can be applied in other areas to improve the accuracy in case of imbalanced dataset. In case of Big Data also SMOTE can be applied using Hadoop framework and Mapreduce programming model with new algorithmic approach. In this research, rebalance algorithm i.e. SMOTE is applied on imbalanced dataset to make it balanced dataset. This research article presented on the improvement of analytical framework for the prediction of kidney function.data- set has been gathered and integrated to form the target dataset. In the data preparation phase, traditional tech- niques for removing missing values are employed and normalization technique is applied to remove bias and bugs identified in the data. At the end, various rule- based induction models and Decision tree models are applied for the classification, such that the generating hidden knowledge is interpretable.Further, SMOTE can be applied for Big Data analysis using Hadoop framework with the help of mapreduce programming model with new algorithmic approach, which is our future work.

1. METHODOLOGY

Disease Prediction has been already implemented using different techniques like ANN, Random Forest, ID3 and Naïve Byes algorithm. From the analysis it was found that Naïve Bayes is more accurate than other techniques. So, Disease Predictor also uses Naïve Bayes for the prediction of Chronic Kidney Diseases.



The above figure 1.1 shows the Architecture Diagram of prediction of chronic kidney disease. The steps can be illustrated as follows

1. **Dataset** - Dataset for prediction of chronic kidney disease using machine learning algorithm is collected from different resources. In that dataset there are 400 patient records are included. Also include 25 attributes

# Data prepossessing -

* 1. **Data Cleaning:** Data obtained from resources does not contain the name of the attribute so first we have to

assign the names to the attribute. Missing values in the dataset like NA’s or blank values are removed by using WEKA function “ReplaceMissingValues” used, which replaces NA’s with the mean values of that attribute.

* 1. **Training and Testing Dataset:** The dataset is divided into two sub datasets both containing attributes. Training data training dataset is derived from main dataset and it contains 300 out of 400 records in main dataset of CKD.Testing data testing dataset is of 100 out of 400 records from main CKD dataset.
  2. **Machine learning algorithm:** the training and testing datasets are applied to machine learning algorithm. In our project we are using naive Bayes algorithm and applying data sets into that.
  3. **Prediction:** In this section the system will predict that the person is having chronic kidney disease or not .

# MODULES IN APPLICATION :

**User:**

User need to first register themselves providing the required details. And then user can login using the details which he/she provided.

* View Profile:Which is the section where user can view his/her profile details.
* Prediction: Here there will be nearly 14 questions, where user need to answer all the questions based on the earlier patient datasets. Then it will predict whether the person is having the CKD or not. If the result is positive then it will go for prediction stage, there a person can know in which stage he is.
* Appointment: Even if the prediction is positive or negative the user can get a appointment from the doctor.
* Feedback: Here user can give the feedback about the doctor.

# Admin:

* View User details: In this section Admin can view the details about the user who has logged in for this app.
* Add Doctor: which is the section where admin can edit like add doctor, remove doctor, and add the information about the doctor etc.
* Change Password: Here only admin can change the password.
* View FeedBack: Here admin can view the feedback of the users.

# Doctor:

* Here doctor can upload doctor name, doctor email , qualification, specification etc which is viewed by the user.
* Doctor can also view the appointment of the user.

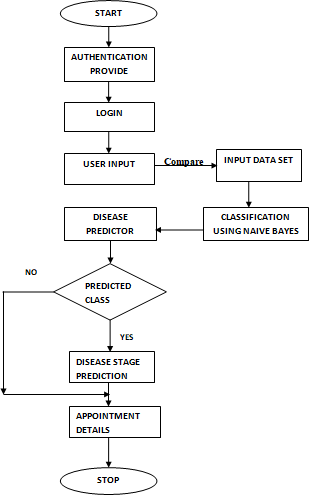
# Menu:

* Information about the chronic kidney disease and its stages are displayed.
* It also gives the information about symptoms and dietary recommendations .
* It also searches the nephrologists available near the user

.

# Dataflow Diagram :

At first user have to make a registration by entering all the necessary details.Once he completes the registration process he can login to this app.Based on the 12 attributes of the dataset there will be several questions asked. After answering all the question it uses the naïve bayes classifer algorithm to predict the Chronic Kidney Disease. If many symptoms of a person matches the symptoms in the dataset then it results as positive indicating that he is suffering from CKD , otherwise negative. If the result is positive it shows the stage of the kidney disease . It also shows the details about the Chronic Kidney Disease, its symptoms, precautions , dietary recommendations.



1. CONCLUSION

The prediction of chronic kidney disease is the project which is an application for medical field. It will be helpful for the medical practitioner to predict whether chronic kidney disease is present or not based on the previous data sets that is collected from the different patient, who suffered through the CKD . This project is also an automation of disease prediction. This paper uses different data mining techniques and machine learning classification algorithm for the correct prediction of the chronic kidney disease. In these papers it was also shown the comparison of different algorithms based on the measures

like accuracy, sensitivity, specificity, precision etc. By comparing different machine learning algorithm it was proven that Naïve Bayes and Decision tree algorithm gives the highest accuracy compared to other algorithm.

REFERENCES

1. Prediction of Chronic Kidney Disease using Data Mining Classification Techniques and ANN,Sahana B J International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 NCETEIT - 2017 Conference Proceedings
2. Review of Chronic Kidney Disease based on Data Mining Techniques S.Dilli Arasu,Dr. R.Thirumalaiselvi International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13498-13505
3. Prediction of kidney disease stages using data mining algorithms El- Houssainy A. Radya, Ayman S. Anwarb,Informatics in Medicine Unlocked 15 (2019) 100178
4. N Afhami”Prediction of Diabetic Chronic Kidney Disease Progression using Data Mining Technique”,International Journal of Computer Science Engineering(IJCSE)ISSN:2319-7323 Vol.7 No.02 Mar-Apr 2018
5. Sai Prasad Potharaju ,M.Sreedevi “An Improved Prediction of Kidney Disease using SMOTE”,Indian journal of Science and Technology Vol 9(31),DOI:10.17485/ijst/2016/v9i31/95634,Aug 2016
6. Champa M S,Rekha B S”Chronic Kidney Disease Detection using Microservice Architecture”,International Jouranl of Innovative Research in Computer Science and Technology (IJIRCST) ISSN:23475552,Vol 7,Issue 3,May 2019 DO1 : 10.21276/ijircst .2019.7.3.10
7. Sirage Zeynu and Shruthi Patil”Survey on Prediction of Chronic Kidney Disease Using Data Mining Classification Technique and Feature Selection”,International Journal of Pure and Applied Mathematics Vol 118 No.8 2018,149-156
8. Tabassum S,Mamatha Bai B G”Analysis and Prediction of Chronic Kidney Disease using Data Mining Technique”International Journal of Engineering Research in Computer Science and Engineering(IJERCSE)Vol 4,Issue 9,September 2017
9. S Gopika,Dr M Vanita”Survey on Prediction of Kidney Disease by using Data Mining Techniques” ,International Journal of Advanced Research in Computer and Communication Engineering IS03297:2007 Certified Vol 6,Issue 1,January 2017
10. Njoud Abudullah Almansour,Hajra Fahim Syed, Nuha Radwan Khayat, Rawan Kanaan Altheeb, Renad Emad Juri, Jamal Alhiyaﬁ∗, Saleh Alrashed, Sunday O. Olatunji”Neural network and support vector machine for prediction of chronic kidney disease:A comparative study,Computer in Biology and Medicine 109(2019) 101-111