IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING APPROACH TO CHRONIC KIDNEY DISEASE PREDICTION AND DIET PLAN RECOMMENDATION

Mahesh G. Huddar[1], Ms. Sneha Hirekodi[2], Ms. Sneha Ganachari[2], Ms. Bandavva Satigoudar[2], Ms. Ashwini Maled[2]

Associate Professor[1], Student[2]

Department of Computer Science & Engineering,

Hirasugar Institute of Technology, Nidasoshi, Belagavi, India.

Abstract:

Chronic kidney disease (CKD) is a type of kidney disease in which there is gradual loss of kidney function over a period of months or years. Prediction of this disease is one of the most important problems in medical fields. So automated tool which will use machine learning techniques to determine the patient's kidney condition that will be helpful to the doctors in prediction of chronic kidney disease and hence better treatment. The proposed system extracts the features which are responsible for CKD, then machine learning process can automate the classification of the chronic kidney disease in different stages according to its severity. The objective is to use machine learning algorithm and suggest suitable diet plan for CKD patient using classification algorithm on medical test records. Diet recommendation for patient will be given according to potassium zone which is calculated using blood potassium level to slow down the progression of CKD.

Index Terms **– Machine Learning, Chronic Kidney Disease, Ensemble Approach**

I INTRODUCTION

Chronic kidney disease (CKD) is a major burden on the healthcare system because of its increasing prevalence, high risk of progression to end-stage renal disease, and poor morbidity and mortality prognosis. It is rapidly becoming a global health crisis. Unhealthy dietary habits and insufficient water consumption are significant contributors to this disease. Without kidneys, a person can only live for 18 days on average, requiring kidney transplantation and dialysis. It is critical to have reliable techniques at predicting CKD in its early stages. Machine learning (ML) techniques are excellent in predicting CKD. The current study offers a methodology for predicting CKD status using clinical data, which incorporates data preprocessing, a technique for managing missing values, data aggregation, and feature extraction. A number of physiological variables, as well as ML techniques are implemented to train models for reliable prediction. Depending on the predictions the diet suggestion is done automatically to slow the progression of chronic kidney disease, Figure below depicts the block diagram of the proposed system. The framework utilizes the CKD prediction dataset. After preprocessing and feature selection, the DT, KNN, and logistic regression algorithms have been used.

The main objective of the develop a machine learning model for prediction of Chronic kidney diseases(CKD). The objectives of the project are to:

- Collect the dataset and pre-process the data.
- Use different machine learning algorithms like Random Forest Classifier, Gradient boosting classifier etc to perform the classification.
- Compare the performance of the classification algorithm and then Develop an smart diet suggestion system depending on the predictions.
- Develop a web application using flask which can be used to predict the CKD and view diet suggestions depending on the parameters entered.

II PROPOSED WORK AND METHOLOGY

These section different functionalities of the project or the functional requirements of the project. This section illustrates various resources and approaches that are used in this work. Primarily, the description of dataset is provided to understand how to work on it, followed by the preprocessing steps involved. Finally, the internal working and understanding of the analytical models used are explained.

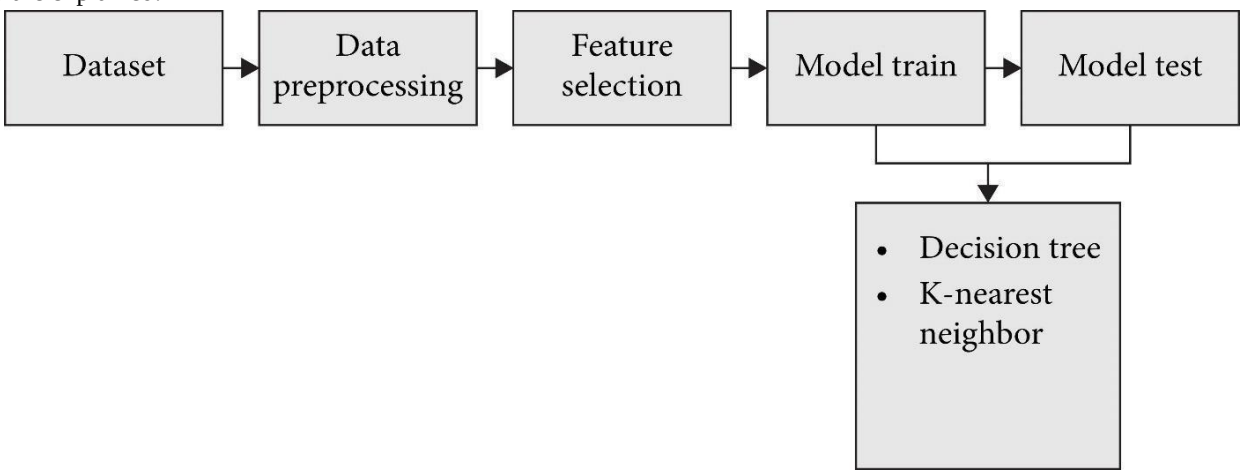

Fig 1. System Architecture

- The dataset is available on the Kaggle website, and it has 450 instances and 25 attributes including numeric and nominal. The selected dataset will be split into train and test.

https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

Tabel 1. Dataset

Attribute No.	Representation	Information Attribute	Description
1	Age	Age	Numerical
2	Blood Pressure	Bp	Numerical
3	Specific Gravity	Sg	Nominal
4	Albumin	Al	Nominal
5	Sugar	Su	Nominal
6	Red Blood Cell	Rbc	Nominal
7	Pus Cell	Pc	Nominal
8	Pus Cell Clumps	Pcc	Nominal
9	Bacteria	Ba	Nominal
10	Blood Glucose Random	Bgr	Numerical
11	Blood Urea	Bu	Numerical
12	Serum Creationism	Sc	Numerical
13	Sodium	So	Numerical
14	Potassium	Pot	Numerical
15	Hemoglobin	Hemo	Numerical
16	Packed Cell Volume	Pcv	Numerical
17	White Blood Cel Count	Wc	Numerical
18	Red Blood Cell Count	Rc	Numerical
19	Hypertension	Htn	Nominal
20	Diabetes Mellitus	Dm	Nominal
21	Coronary Artery Disease	Cad	Nominal
22	Appetite	Appet	Nominal
23	Pedal Edema	Pe	Nominal
24	Anemia	Ane	Nominal
25	Class	Class	Nominal

- Preprocessing is a method to obtain complete, consistent, interpretable data. The data quality affects the mining results that are obtained using machine learning algorithms. Quality data results in a quality decision.
- Feature Selection: In this step we select subset of relevant attributes from the total given attributes. This stage helps in reducing the dimensionality and making the model simpler and easy to use, thus leading to short training time and high accuracy. To obtain Highly dependent features for CKD prediction we will either use correlation or Regression to form.
- Analysis involves using different machine learning algorithms for solve the problem discussed above.
- Diet suggestion module: As dietary management plays an important role during chronic kidney disease to slow down the progression of CKD. Mostly patients with diabetes and high blood pressure conditions should a very strict diet to prevent from kidney failure. So, in this module, based on zone detected (using potassium level) and output from prediction module a patient will be given suitable diet which will fetched from diet database. Also, CKD patient with high blood pressure condition will be given an alert to lower down sodium intake.

The following machine learning algorithms are used for the purpose of the project :

- Decision Tree (DT) is one of the simplest algorithms, yet most effective and useful. It is a tree which comprises of three nodes, first is the chance node, second is decision node and at the last end node. The chance node shows the probable outcomes of a particular node whereas the decision node is a node on which a decision is to be made based on the outcome. The end node is the last node of the tree which gives the final result of a path.
- K-Nearest Neighbor (KNN) is also a supervised classification algorithm. It predicts the target class on the basis of how similar that particular data is from other provided training data labels to the model. This can be understood as, the characteristic(features) of that data, whose target label needs to be predicted, is compared with features of existing data (except the target class).
- Naive Bayes Classifier Algorithm Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- SVM Classifier: SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.
- Random Forest is a commonly used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.
- Gradient Boosting is a functional gradient algorithm that repeatedly selects a function that leads in the direction of a weak hypothesis or negative gradient so that it can minimize a loss function. Gradient boosting classifier combines several weak learning models to produce a powerful predicting model.

IV Results and Analysis

The proposed methodology is tested on chronic kidney disease dataset which is available on the Kaggle website. The following evaluation metrics are used to evaluate the performance of the proposed model.

Accuracy: Accuracy is a performance metric commonly used in machine learning to measure the proportion of correct predictions made by a model over the total number of predictions made. The formula for accuracy is:

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions}$$

Precision: Precision is a performance metric in machine learning that measures the proportion of true positive predictions made by a model over the total number of positive predictions Precision helps to evaluate the accuracy of a model's positive predictions. The formula for precision is:

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Recall: Recall is a performance metric in machine learning that measures the proportion of true Positive predictions made by a model over the total number of actual positives in the dataset. Recall helps to evaluate the ability of a model to identify all positive instances in the dataset. The formula for recall is:

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

F1 Score: F1 score is a performance metric in machine learning that combines both precision and recall to evaluate the overall performance of a binary classification model. The F1 score is the harmonic mean of precision and recall, and it ranges from 0 to 1, where a higher value indicates better performance. The formula for F1 score is:

$$F1\ score = 2 * \left(\frac{precision * recall}{precision + recall}\right)$$

The table shows the results from different machine learning algorithms.

Table 2. Algorithm Results

Algorithm Name	Precision	Recall	F1-Score	Accuracy
Random Forest Classifier	0.98	0.97	0.97	0.975
Gradient Boosting Classifier	0.98	0.97	0.97	0.975
Decision Tree Classifier	0.99	0.99	0.99	**0.991**
Naïve Bayes Classifier	0.95	0.95	0.95	0.95
SVM Classifier	0.90	0.90	0.90	0.9
K-Nearest Neighbor Classifier	0.93	0.93	0.93	0.93

Among above mentioned algorithm Decision Tree algorithm gives high accuracy and best accuracy. That is 99% of accuracy. So we used decision tree algorithm for further process.

IV CONCLUSION

Chronic Kidney Disease prediction and Diet Plan recommendation using Machine Learning is the software where user need to fill the requirement of his/her body parameters like blood sugar, height, weight, blood pressure and so on. Also, the software provide the result with its disease stage and recommend the diet plan for the parser who is facing chronic kidney disease.

V REFERENCES

[1] S. Levey, R. Atkins, J. Coresh et al., "Chronic kidney disease as a global public health problem: approaches and initiatives–a position statement from kidney disease improving global outcomes," Kidney International, vol. 72, no. 3, pp. 247–259, 2007.

[2] V. Jha, G. Garcia-Garcia, K. Iseki et al., "Chronic kidney disease: global dimension and perspectives," The Lancet, vol. 382, no. 9888, pp. 260–272, 2013.

[3] N. R. Hill, S. T. Fatoba, J. L. Oke et al., "Global prevalence of chronic kidney disease – a systematic review and meta-analysis," PLoS One, vol. 11, no. 7, article e0158765, 2016.

[4] 5- H. Nasri, "World kidney day 2014; chronic kidney disease and aging: a global health alert," Iranian Journal of Public Health, vol. 43, no. 1, pp. 126-127, 2014.

[5] G. Abraham, S. Varughese, T. Thandavan et al., "Chronic kidney disease hotspots in developing countries in South Asia," Clinical Kidney Journal, vol. 9, no. 1, pp. 135–141, 2016.

[6] K. T. Mills, T. Xu, W. Zhang et al., "A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010," Kidney International, vol. 88, no. 5, pp. 950–957, 2015.

[7] Ene-Iordache, N. Perico, B. Bikbov et al., "Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study," The Lancet Global Health, vol. 4, no. 5, pp. e307–e319, 2016.

[8] S. Anand, M. A. Khanam, J. Saquib et al., "High prevalence of chronic kidney disease in a community survey of urban Bangladeshis: a cross-sectional study," Glob Health, vol. 10, no. 1, p. 9, 2014.

[9] L. Ali, K. Fatema, Z. Abedin et al., "Screening for chronic kidney diseases among an adult population," Saudi Journal of Kidney Diseases and Transplantation, vol. 24, no. 3, p. 534, 2013.