iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52281

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Chronic Kidney Disease Prediction Using Machine Learning

Rohit More[1], Subodh Shinde[2], Pratik Shah[3], Riya Raut[4]
Zeal College of Engineering and Research

Abstract: ***Chronic kidney disease (CKD) is a life-threatening condition that can be difficult to diagnose early because there are no symptoms. The purpose of the proposed study is to develop and validate a predictive model for the prediction of chronic kidney disease. Machine learning algorithms are often used in medicine to predict and classify diseases. Medical records are often skewed. Chronic Kidney Disease (CKD) or chronic renal disease has become a major issue with a steady growth rate. A person can only survive without kidneys for an average of 18 days, which makes a huge demand for a kidney transplants and Dialysis. It is important to have effective methods for the early prediction of CKD. Machine learning methods are effective in CKD prediction. This work proposes a workflow to predict CKD status based on clinical data, incorporating data prepossessing, a missing value handling method with collaborative filtering and attribute selection. The extra tree classifier and random forest classifier are shown to result in the highest accuracy and minimal bias to the attributes. The project also considers the practical aspects of data collection and highlights the importance of incorporating domain knowledge when using machine learning for CKD status prediction.***

Keywords: CKD, Random Forest Classifier, Extra Tree Classifier

I. INTRODUCTION

Engineers and medical researchers are trying to develop machine-learning algorithms and models that can identify chronic kidney disease at an early stage. The problem is that the data generated in the health industry is large and complex, making data analysis difficult. However, we can process this data into a data format using data mining technology, and then this data can be translated into machine learning algorithms. A combination of estimated glomerular filtration rate (GFR), age, diet, existing medical conditions, and albuminuria can be used to assess the severity of kidney disease but requires more accurate information about the risk to the kidney is required to make clinical decisions about diagnosis, treatment, and referral. This model aims to develop and validate predictive models for chronic kidney disease. The main goal will be to evaluate kidney failure, which means the need for kidney dialysis or kidney transplant first. These models also teach the patient how to live a healthy life, help the doctor see the risk and severity of the disease, and how to proceed with the treatment in the future. It may be possible to identify patterns of data collection using ANN, and mining methods and the future occurrence of certain diseases that may cause harm can be predicted in advance [1]

II. RELATED WORKS

A review of literature pertaining to Chronic Kidney Disease prediction using Machine Learning is presented here in this section.
Machine learning techniques can be used to ascertain the existence of chronic kidney disease by imposing various classification algorithms on the patient's medical record. Empirical work is performed on different algorithms like Support Vector Machine, Random Forest, XGBoost, Logistic Regression, Neural Networks, and Naive Bayes Classifier. The work is primarily concentrated on finding the best suitable classification algorithm which can be used for the diagnosis of CKD based on the classification report and performance factors. The experimental results show that Random Forest and XGBoost give better results when compared to other classification algorithms and generate 99.29% accuracy. The classification techniques, that is, tree-based decision trees, random forests, and logistic regression have been analyzed. The different measures has been used for comparison between algorithms for the dataset collected from the standard UCI repository.
Another study proposes and evaluates a Kernel-based Extreme Learning Machine to predict Chronic Kidney Disease. Subsequently, various kernel-based ELM were evaluated. The performance of four kernels-based ELM, namely RBF-ELM, Linear-ELM, Polynomial-ELM, Wavelet-ELM, and of standard ELM were compared. The result showed that the radial basis function extreme learning machine (RBF -ELM) was higher than those from the other tested and give the best prediction sensitivity and specificity of 99.38% and 100% respectively.

©IJRASET: All Rights are Reserved

Another study demonstrates the early prediction model of kidney diseases using an adaptive neuro-fuzzy logic system (ANFIS). This model diagnoses the stages of kidney diseases so that treatment can be provided according to the disease condition. Mat lab-based ANFIS CKD stage prediction model is presented with an accuracy of 94 percent in terms of actual output to estimated output.
Another study extracts the features which are responsible for CKD, then the machine learning process can automate the classification of chronic kidney disease in different stages according to its severity. The objective is to use a machine learning algorithm and suggest suitable diet plans for CKD patients using a classification algorithm on medical test records. Diet recommendations for the patient will be given according to potassium zone which is calculated using blood potassium level to slow down the progression of CKD.

A. Proposed System

The dataset required for the project is Chronic Kidney Disease Dataset from Kaggle. The proposed system aims to develop a Chronic Kidney Disease Prediction system using machine learning algorithms. The system plans to use Random Forest Classifier and XGBoost Classifier to predict the CKD. The system will evaluate the performance of the machine learning model using various evaluation metrics such as the confusion matrix.

B. System Architecture

The system architecture consists of a backend where the machine learning model is trained and tested using the CKD dataset obtained from Kaggle. Random Forest Classifier and XGBoost Classifier algorithm are used for model training.
The front end is a simple web app created using the Python Flask framework.

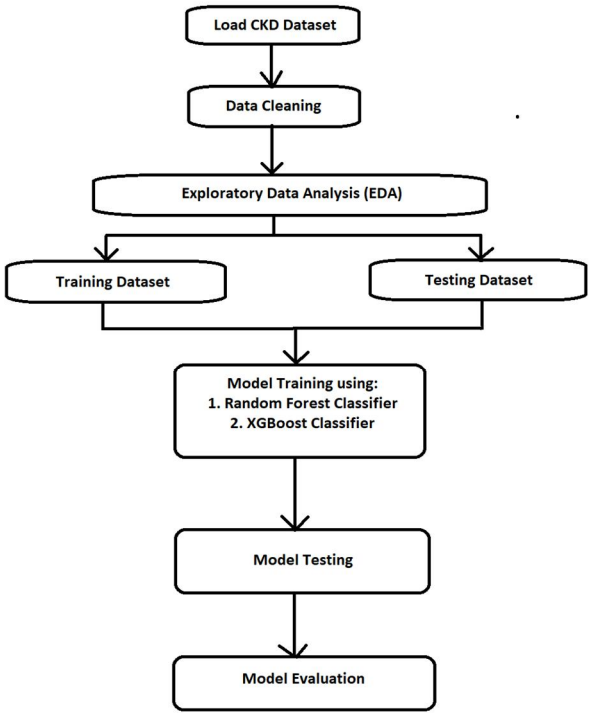


Fig 1 - System Architecture

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the model's performance. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

XGBoost (eXtreme Gradient Boosting) is a popular supervised-learning algorithm used for regression and classification on large datasets. It uses sequentially-built shallow decision trees to provide accurate results and a highly-scalable training method that avoids overfitting. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements Machine Learning algorithms under the Gradient Boosting framework. It provides a parallel tree boosting to solve many data science problems in a fast and accurate way.

III. APPLICATIONS

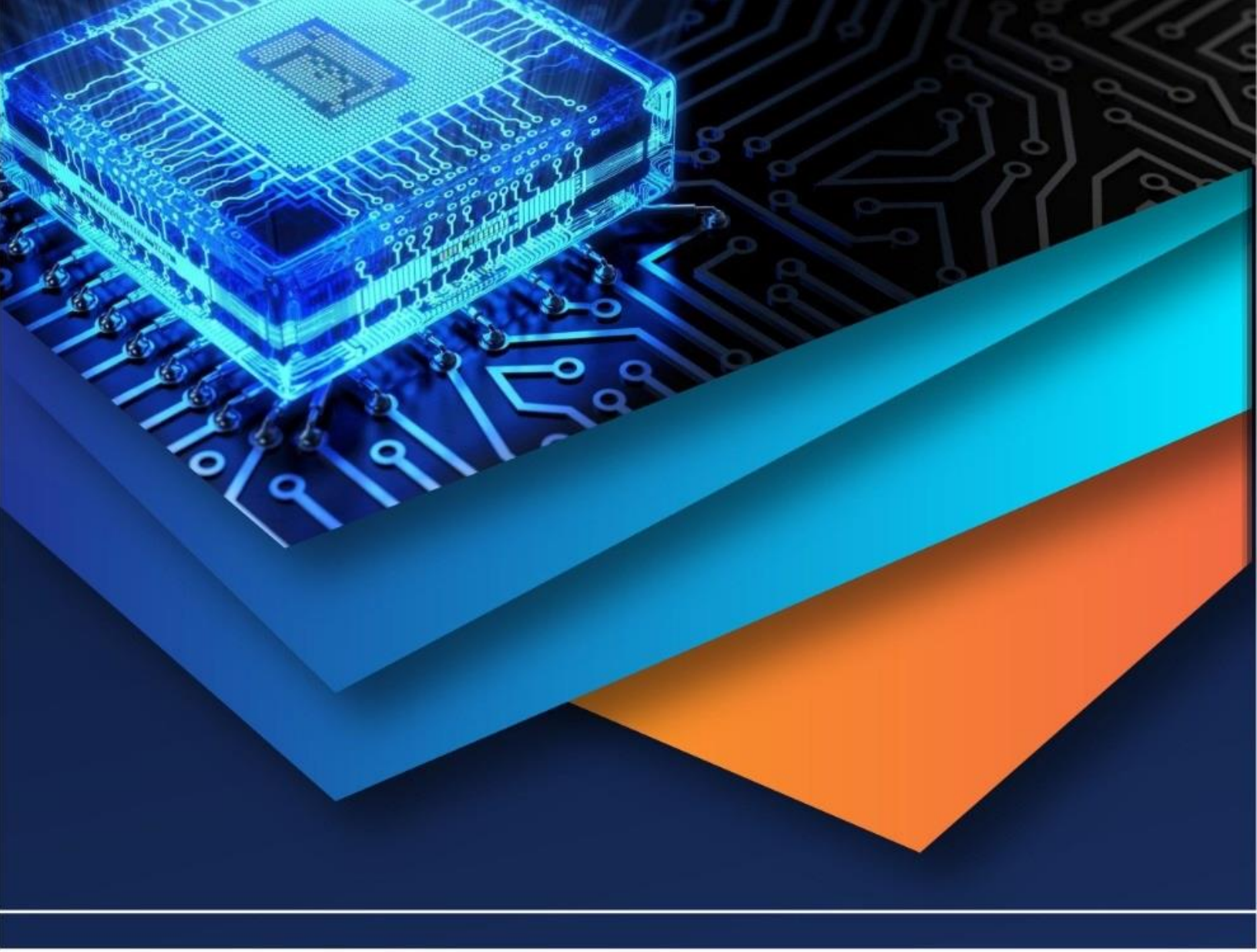
1) Web-based platform for complete detection and Prediction of Chronic Kidney Disease
2) A base system for building the system for detection and prediction of different Chronic Kidney Diseases such as Atypical hemolytic uremic syndrome (aHUS), Alport syndrome, Amyloidosis, Cystinosis, Fabry disease, Focal segmental glomerulosclerosis (FSGS), Glomerulonephritis (glomerular disease), Goodpasture syndrome.
3) A system for understanding and studying the effects of different features on Chronic Kidney Disease.

IV. CONCLUSION

In this project, we have used the chronic kidney disease dataset collected from the Kaggle repository. We have developed a chronic kidney disease prediction model using two machine learning classifiers Random Forest and XGBoost Classifier to measure the performance of the prediction model. The performance of the model depends upon the confusion matrix. The developed chronic kidney disease prediction model has been trained by categorical and non_categorical chronic kidney disease dataset attributes. After applying the base classifiers we find that the Random Forest classifier got an accuracy of 95.83% and XGBoost Classifier got an accuracy of 98.33%. The Random Forest classifier performed better than XGBoost Classifier. This can help medical practitioners and patients in the early prediction of chronic kidney disease to save a life. In the future, the model can be further tuned by applying feature selection methods to increase the performance of the prediction.

REFERENCES

[1] H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018, pp. 1-4, DOI: 10.1109/ICICOS.2018.8621762.
[2] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav, and R. Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 2019, pp. 1-4, DOI: 10.1109/ICNTE44896.2019.8946029.
[3] N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha and C. Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 642-647, DOI: 10.1109/ICCS45141.2019.9065309.
[4] R. Gupta, N. Koli, N. Mahor and N. Tejashri, "Performance Analysis of Machine Learning Classifier for predicting Chronic Kidney Disease 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, DOI: 10.1109/INCET49848.2020.9154147
[5] K. Damodara and A. Thakur, "Adaptive Neuro Fuzzy Inference System based Prediction of Chronic Kidney Disease," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 973-976, DOI: 10.1109/ICACCS51430.2021.9441989.
[6] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav and R. Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 2019, pp. 1-4, doi: 10.1109/ICNTE44896.2019.8946029.
[7] P. Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 2017, pp. 193-198, doi: 10.1109/COMPSAC.2017.84.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)