As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 9(04): 31-34, 2023
Copyright © 2023International Journal for Modern Trends in Science and Technology
ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0904005
Available online at: http://www.ijmtst.com/vol9issue04.html

Random Forest Based Chornic Kidney Disease Prediction and Analysis

Dr. Y. Jaya Babu | Peketi Durga Pavani | Battula Naga Chakra Satya Sai Hima Madhuri | Seekoti Sudheer Babu | Maddipudi Sai Ravi Teja | Rao Venkata Rama Gopala Krishna Rohith

Department of Computer Science and Engineering, Pragati Engineering College (A), Surampalem (East Godavari) A.P,India.

To Cite this Article
Dr. Y. Jaya Babu, Peketi Durga Pavani, Battula Naga Chakra Satya Sai Hima Madhuri, Seekoti Sudheer Babu, Maddipudi Sai Ravi Teja and Rao Venkata Rama Gopala Krishna Rohith. Random Forest Based Chornic Kidney Disease Prediction and Analysis. International Journal for Modern Trends in Science and Technology 2023, 9(04), pp. 31-34. https://doi.org/10.46501/IJMTST0903005

Article Info
Received: 02 March 2023; Accepted: 25 March 2023; Published: 30 March 2023.

ABSTRACT

The basic purpose of the project is to detect whether the person is having chronic kidney disease or not. Chronic Kidney Disease (CKD) is a global health issue that causes a high incidence of morbidity and death, as well as the onset of additional illnesses. Because there are no clear symptoms in the early stages of CKD, people frequently miss it. Early identification of CKD allows patients to obtain prompt therapy to slow the disease's development. Due of their rapid and precise identification capabilities, machine learning models can successfully assist doctors in achieving this aim. The prediction models used in our project helps to detect. whether the person is suffering from chronic kidney disease or not.

1. INTRODUCTION

Chronic Kidney Disease (CKD) is considered as an important threat for the society with respect to the health in the present era. CKD can be caused due to lack of water consumption, smoking, improper diet, loss of sleep and many other factors. This disease affected 753 million people globally in 2020 in which 417 million are females and 336 million are males. Majority of the time the disease is detected in its final stage and which sometimes leads to kidney failure. This disease is characterized by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious symptoms in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. Hence, the prediction and of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease.

2. LITERATURESURVEY

This chapter describes the research literature relevant to the primary aspects of this thesis. The core aspects of this thesis are machine learning applications to detect whether the is suffering with chronic kidney disease. Data mining techniques can be classified into supervised learning technique and unsupervised learning technique. This section consists of the reviews of various technical and review articles on data mining techniques applied to predict Kidney Disease.

1. chronic kidney disease prediction using machine learning techniques:

Goal three of the UN's Sustainable Development Goal is good health and well-being where it clearly emphasized that non-communicable diseases is emerging challenge. One of the objectives is to reduce premature mortality from non-communicable disease by third in 2030. Chronic kidney disease (CKD) is among the significant contributor to morbidity and mortality from non-communicable diseases that can affected 10–15% of the global population. Early and accurate detection of the stages of CKD is believed to be vital to minimize impacts of patient's health complications such as hypertension, anemia (low blood count), mineral bone disorder, poor nutritional health, acid base abnormalities, and neurological complications with timely intervention through appropriate medications. Various researches have been carried out using machine learning techniques on the detection of CKD at the premature stage. Their focus was not mainly on the specific stage's prediction. In this study, both binary and multi classification for stage prediction have been carried out. The prediction models used include Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). Analysis of variance and recursive feature elimination using cross validation have been applied for feature selection. Evaluation of the models was done using tenfold cross-validation. The results from the experiments indicated that RF based on recursive feature elimination with cross validation has better performance than SVM and DT.

2.Prediction of chronic kidney disease using RF algorithm and XG boost

Data from 98,432 participants were included, with a median age of 43 years (25th–75th percentiles: 36.6 to 51.4). Fifty-four (55.1%) studies examined participants thought be at high risk of CKD (people with hypertension, diabetes, HIV, or sickle cell disease), while 33 (33.7%) were conducted in the general population or subjects not known to be at risk of CKD, and 11 (11.2%) studies in both. The included studies applied various estimators of GFR. Fifty-eight applied a single equation; the most frequently used being the Modification of Diet in Renal Disease (MDRD) eq. (32 studies), followed by the Cockcroft-Gault formula (17 studies), the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula (7 studies), and the Cystatin C equation (two studies). Of the 22 studies that compared eGFR using two or more equations, Cockcroft-Gault and MDRD equations were used in 11 studies, CKD-EPI and MDRD used in 3 studies, and the three formulas (Cockcroft-Gault, CKD-EPI, and MDRD) used in 8 studies. The most common methods used to assess proteinuria were the urine dipstick test (46 studies), followed by the spot urine albumin to creatinine ratio (20 studies), and the 24-h urine collection (three studies).

3.Prevalence of Chronic Kidney Disease and Associate Factors among patients with diabetes in north west Ethiopia:

Chronic kidney disease (CKD) is increasingly recognized as a global health issue and it affects 10% to 15% of the world population. Diabetes mellitus is the leading cause of end-stage renal disease. More than 422 million adults in the world populations are living with diabetes mellitus, 40% of whom will develop CKD. CKD in diabetes increases the risk of early death and cardiovascular morbidity and mortality..

3. SYSTEM ANALYSIS

A. Existing System

Chronic Kidney Disease is a long persisting renal disease which is known at advanced stages. Until now, in majority of cases full features have been taken to consideration. Hence, it took more time to detect the disease. Also, in existing system datasets with small size are used so that accuracy to detect the disease is not so high. In existing system some machine learning algorithms like Logistic Regression and Support Vector Machine (SVM) are used to detect whether the person is having kidney disease or not.

DISADVANTAGES OF EXISTING SYSTEM

As this system deals with small datasets the accuracy changes when dataset size changes. The algorithms used here gives the detection with low accuracy. It takes more time for training and testing of the data. Also, it doesn't deal with missing and redundant values. It takes all the features into consideration where the time to process the data is very high. As the disease is life threatening it should be detected with high accuracy to save lives of people.

B. Proposed System

In our proposed system we detect the kidney disease using Random Forest (RF) Algorithm and along with XGBoost model. In this system we will be going through the chronic kidney disease dataset and doing the complete analysis. Here we used some data pre-processing techniques to improve the quality of dataset. This study focuses on detection of kidney disease using machine learning models based on dataset with big size. The Random Forest algorithm is a supervised learning model. The output of this algorithm selects from majority of votes so that gives output with high accuracy. RF algorithm takes less training time as compared to other algorithms. The other model we used is XGBoost. It is an optimized distributed gradient boosting library designed for efficient and scalable training of machine. XGBoost stands for Extreme Gradient Boosting. It has an ability to handle large datasets and handle the missing values.

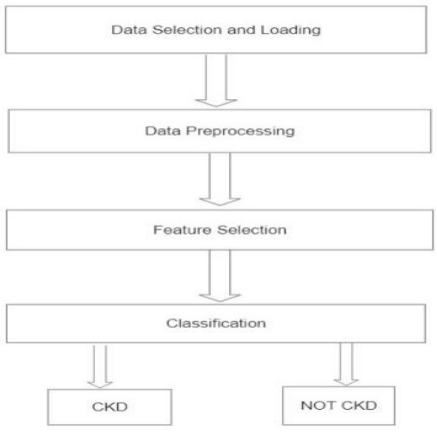
ADVANTAGES OF PROPOSED SYSTEM:

The Random Forest algorithm can handle large datasets efficiently. It provides an effective way of handling missing data. The random forest algorithm provides a higher level of accuracy in predicting outcomes. Also the XGBoost model is a scalable and highly accurate implementation of gradient boosting that pushes the limits of computing power for boosted tree algorithms. Hence in our proposed system we used random forest algorithm and XGBoost model gives the output with high accuracy.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture of Detection of Chronic Kidney Disease using RF Algorithm



5. SYSTEM IMPLEMENTATION

MODULES

1.Dataset Selection and Loading.
2.Data Pre-processing
3.Feature Selection
4.Classification

1.Dataset Selection and Loading

The Dataset here we use is publicly available CKD Dataset. Here the information of dataset uses the patient's data like age, blood pressure, specific gravity, albumin, sugar, red blood cells etc. To detect the CKD, we need the data set with suitable attributes. CKD is caused due to diabetes and high blood pressure. Due to Diabetes our many organs get affected and it will be followed by high blood sugar. So, it is important to predict the disease as early as possible. This study improvises some of the machine learning techniques to predict the disease.

2.Data Pre-processing

Data Pre-Processing is that stage where the data that is distorted, or encoded is brought to such a state that the machine can easily analyze it. A dataset can be observed as a group of data objects. Data objects are labelled by a number of features, that ensures the basic features of an object, such as the mass of a physical object or the time at which an event ensured. In the dataset there may be missing values, they can either eliminated or estimated. The most common method of dealing with missing values is filling them in with mean, median or mode value of respective feature.

3. Feature Selection

Feature Selection is the method where we computationally select the features which contribute most to our prediction variable or output. In our data set we have some attributes which are important to detect the disease. So, we select that attributes and by using our machine learning algorithm we detect the output.

4. Classification

For classification we use Random Forest (RF) algorithm to predict the disease. we have imported the libraries like NumPy, pandas etc. for classification. This is the final step of our project in this we give input data like age, blood pressure, sugar, bacteria etc. and

then compare it with data in data set and detect if the person has CKD or not.

6.CONCLUSION AND FUTURE WORK

Early prediction is very crucial for both experts and patients to prevent and slow down the progress of chronic kidney disease to kidney failure. proposed. The model supervised imputation of missing values with a selection of the most promising data imputation methods and could be used to diagnose CKD Gradient Boosting imputation. Hence this project was to predict patients with CKD using less number attributes while maintaining a higher accuracy. Here we obtain an accuracy of about 98 percentage. In the future, once our analysis of the data is complete, we will train a statistical model in hopes of improving the generalization performance of the model. We will collect a more representative, and more complex, set of data to improve the model.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Joseph A Akinyele, Christina Garman, Ian Miers, Matthew W Pagano, Michael Rushanan, Matthew Green, and Aviel D Rubin. Charm: a framework for rapidly prototyping cryptosystems. Jour YUK0nal of Cryptographic Engineering, 3(2):111–128, 2013.

[2] Ittai Anati, Shay Gueron, Simon Johnson, and Vincent Scarlata. Innovative technology for cpu based attestation and sealing. In Workshop on hardware and architectural support for security and privacy (HASP), volume 13, page 7. ACM New York, NY, USA, 2013.

[3] Alexandros Bakas and Antonis Michalas. Modern family: A revocable hybrid encryption scheme based on attribute-based encryption, symmetric searchable encryption and SGX. In SecureComm 2019, pages 472–486, 2019.

[4] Amos Beimel. Secure schemes for secret sharing and key distribution. PhD thesis, PhD thesis, Israel Institute of Technology, Technion, Haifa, Israel, 1996.

[5] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In S&P 2007, pages 321–334. IEEE, 2007.

[6] Victor Costan and Srinivas Devadas. Intel sgx explained. IACR Cryptology ePrint Archive, 2016(086):1–118, 2016.

[7] Ben Fisch, Dhinakaran Vinayagamurthy, Dan Boneh, and Sergey Gorbunov. IRON: functional encryption using intel SGX. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, pages 765–782, 2017.

[8] Manjula Devarakonda Venkata1, Sumalatha Lingamgunta& K Murali, Health Care Automation in Compliance to Industry 4.0 Standards: A Case Study of Liver Disease Prediction ,Journal of Scientific & Industrial Research , Vol. 82, February 2023, pp. 263-268, DOI: 10.56042/jsir.v82i2.70215

[9] Manjula Devarakonda Venkata1, Sumalatha Lingamgunta & K Murali, Health Care Automation in Compliance to Industry 4.0 Standards: A Case Study of Liver Disease Prediction ,Journal of Scientific & Industrial Research , Vol. 82, February 2023, pp. 263-268, DOI: 10.56042/jsir.v82i2.70215

[10] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In Advances in Cryptology-CRYPTO 1999, pages 537–554. Springer, 1999.

[11] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In ACM CCS 2006, pages 89–98. ACM, 2006.

[12] Jinguang Han, Willy Susilo, Yi Mu, Jianying Zhou, and Man Ho Allen Au. Improving privacy and security in decentralized ciphertext-policy attribute-based encryption. IEEE transactions on information forensics and security, 10(3):665–678, 2015.