JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Utilizing Hybrid Machine Learning Techniques for Chronic Kidney Disease Diagnosis

Harshada Krishnapuram,
Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana.

Dr. A Satyanarayana,
Professor, Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana.

Y.Sowjanya,
Assistant Professor, Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana.

K.Uma,
Assistant Professor, Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana.

Abstract - Chronic Kidney Disease (CKD) has a widespread influence on the world's population and causes substantial numbers of fatalities every year due to kidney malfunction. The objective of this venture is to conduct a comprehensive analysis of the CKD data collection with the purpose of extracting meaningful information. By visualizing and evaluating the data, our aim is to bring to light the symptoms of CKD and determine the impact that different elements have on the ailment.

This project details the steps taken for the formation of a data set, the preparation of a model, and the prediction of disease. The process begins with the procurement of structured and unstructured data from diverse sources, referred to as data collection. Following the collection, the data undergoes preprocessing and is partitioned into a cleaning data set and a test data set. The training data set is utilized to train machine learning algorithms such as the Random Forest, Gaussian Naive Bayes, and Gradient Boosting Classifier for multiple cycles, with the goal of enhancing the accuracy of prediction outcomes. Upon reaching the desired level of accuracy, the model is ready for testing.

1. INTRODUCTION

Chronic Kidney Disease, or CKD for short, is a term utilized to describe a condition where the kidneys lose the capability to fulfill their responsibilities in an adequate manner. These important organs are accountable for purifying blood of waste materials and extra fluids, which are then expelled from the body through the process of urination.

At the earliest stages of CKD, one may not experience any noticeable symptoms. However, as the disease advances and the kidney function deteriorates, there may be a buildup of hazardous levels of fluid, electrolytes, and waste products in the body.

It is of utmost importance to identify and treat CKD in its initial stages in order to prevent its progression. The objective of treatment is to manage the underlying cause and impede further damage to the kidneys. This can lead to better results and a slower progression of the disease.

Should CKD be left untreated, it may advance to a stage referred to as end-stage kidney failure, which is lethal unless treated through artificial filtering procedures such as dialysis or a kidney transplant.

Individuals who have risk factors for CKD, such as a family history of kidney disease, diabetes, or high blood pressure, must prioritize seeking medical attention and managing these conditions effectively in order to slow or prevent the progression of CKD.

The Chronic Kidney Disease (CKD) is a state in which the kidneys, over the course of several months or years, lose the proficiency to operate in a proper manner due to a pre-existing condition or illness. The following are some frequent origins of CKD:

Diabetes, which may be either Type 1 or Type 2

High Blood Pressure

Inflammation of the Kidney's Filtering Units, referred to as Glomerulonephritis

Inflammation of the Kidney's Tubules and Surrounding Structures, known as Interstitial Nephritis

Polycystic Kidney Disease

It is significant to be cognizant of the factors that increase the risk of developing CKD and to seek medical attention if symptoms present themselves. Early recognition and treatment of the underlying cause can assist in slowing the progression of the disease and bettering the outcome.

The diagnosis of Chronic Kidney Disease (CKD) is of utmost importance for the successful administration of the disease and to minimize the likelihood of developing complications. It must be noted that CKD is a persistent and irreversible condition that can ultimately culminate in end-stage renal disease (ESRD) and necessitate dialysis or a kidney transplant.

To diagnose CKD, hybrid machine learning methods may be implemented. These methods employ a combination of multiple algorithms, including decision tree algorithms, support vector machines (SVM), and artificial neural networks (ANN), to augment the accuracy and stability of the diagnosis. Demographic information, laboratory test results, and medical history are used as input features in these models.

To train the hybrid machine learning models, a substantial dataset comprising of CKD patients is utilized. The efficacy of these models can then be evaluated based on their ability to diagnose the disease, such as their sensitivity, specificity, and accuracy.

In conclusion, hybrid machine learning techniques offer a valuable tool for the diagnosis of CKD. The combination of multiple algorithms and the utilization of relevant input features improve the accuracy and stability of the diagnosis, thus contributing to reducing the likelihood of complications.

PURPOSE

The Chronic Kidney Disease (CKD) describes a condition where there is a slow degradation of the functionality of the kidneys. The kidneys play a vital role in the human body as they remove waste and surplus liquids from the blood. The CKD implies that the kidney's performance is weakened and cannot effectively filter the blood. If left unattended, CKD can cause dangerous accumulations of waste, liquids and electrolytes in the body.

The CKD is caused by an ailment or a condition that weakens the kidney's performance and can progress to the End-stage Renal Disease (ESRD), which is life-threatening if not treated. During the early stages of

CKD, there may not be any noticeable signs or symptoms. However, as the condition progresses, the dangerous accumulations in the body increase. The CKD is divided into five stages, with the fifth stage being the most critical as the kidneys are unable to perform most of their functions.

The Glomerular Filtration Rate (GFR) test is the best way to determine the stage of CKD and measure the kidney's functionality. The test is based on factors such as blood creatinine, age, gender, race and others.

The prevalence of CKD is increasing globally, particularly in low and medium-income countries. Approximately 10% of the world's population is affected by the disease, and the number of deaths caused by CKD is on the rise. In the last two decades, various health organizations have conducted research that highlights the devastating impact of CKD on individuals and communities.

The number of people suffering from the End-stage Renal Disease is also increasing, which is the final stage of CKD. To survive this stage, patients require either regular dialysis or a kidney transplant.

In conclusion, CKD is a severe and progressive condition that requires early diagnosis and treatment. The GFR test helps in determining the stage of CKD and guides the appropriate treatment plan. The global impact of CKD on individuals and communities highlights the ongoing need for research and efforts to prevent and manage the disease.

SIGNIFICANCE OF THE STUDY

Chronic kidney disease (CKD), a condition where the kidney loses its ability to properly filter waste and excess fluids from the blood, is causing a surge in alarm in India due to the increasing number of cases being reported each year. If CKD is not detected and treated early, it can lead to dangerous accumulations of fluids, electrolytes, and waste in the body, ultimately causing permanent damage to the kidneys and increasing the risk of other dangerous conditions such as hypertension and anemia.

It is imperative that CKD be diagnosed at an early stage in order to prevent its progression to end-stage renal disease (ESRD). However, CKD is difficult to detect in its early stages as it may not exhibit any signs or symptoms. The stage of CKD can be determined through a test called the Glomerular Filtration Rate (GFR), which measures kidney function based on various factors such as blood creatinine, age, race, and gender. People who have a family history of CKD, hypertension, or diabetes are more susceptible to the disease.

The global impact of CKD on individuals and communities, as well as the growing number of end-stage renal disease cases, is a cause for concern. It is imperative that ongoing research and efforts be made to prevent and manage the disease. In this context, a machine learning methodology to diagnose CKD at an early stage and at a low cost could greatly benefit those affected by the disease. The goal is to enable early treatment and prevent the progression of CKD to ESRD

2. LITERATURE SURVEY

The occurrence of chronic kidney disease (CKD) affects a significant portion of the global population, with yearly fatalities resulting from untended kidney failure being recorded at around one million. Recognizing CKD at its early stages is imperative in the prevention of its progression and improving patient results. Nonetheless, the absence of early symptoms makes diagnosing the disease a challenge without medical tests.

To confront this problem, the project endeavors to study the relationship between health attributes and chronic kidney disease. The project involves identifying the important precursors to CKD that can be utilized for machine learning in the future. This includes comprehending the impact of various health attributes on the results of CKD tests. The study comprises graphical representations and additional processing for a clearer understanding of the connections. The aim is to determine the dependence of each attribute on the disease and ascertain the significant factors causing CKD.

Furthermore, the project examines how health issues such as anemia, diabetes mellitus, coronary artery disease, pedal edema, and hypertension can affect the probability of having CKD. The objective is to identify the most impactful factor in the cause of the disease and provide insight into the root causes of CKD. With this analysis, prompt medical interventions can be carried out and early detections can be made.

Prior research [2] has shown the numerous factors that lead to Chronic Kidney Disease (CKD) and the interconnections between its various attributes. This study presents crucial medical insights, but it lacks statistical analysis. Another study [3] performed by Gulvahid Shaikh and team conducted descriptive statistics and ANOVA, but it only looks into the impact of osmol gap on CKD.

Our analysis, in contrast, strives to take a more all-encompassing approach by considering all aspects that contribute to CKD, rather than just one aspect. Thus, there is a requirement for a more thorough examination that encompasses all factors that lead to the disease, beyond just one aspect.

The study described in reference [4] examines the connection between urinary osmolality and the presence of Chronic Kidney Disease (CKD). The analysis employs techniques like ANOVA, Kruskal-Wallistest, $\chi 2$ test, linear regression analysis and multivariable hazard models to evaluate the relationship between the variables. However, this research is limited in its scope, focusing only on urinary osmolality as a factor for CKD. In contrast, our project seeks to expand this examination by including additional health attributes like diabetes, coronary artery disease, blood sugar levels, blood pressure, and others, with the goal of providing a more comprehensive understanding of the disease and its symptoms.

Chronic Kidney Disease (CKD) is a persistent issue afflicting the kidneys' structure and performance, diagnosed by a glomerular filtration rate (GFR) less than 60 mL/min/1.73 m2 or albuminuria of 30 mg over 24 hours, continuing for more than three months. This ailment impacts a segment of the world's population, ranging from 8% to 16%, with diabetes and hypertension being the most prevalent causes in advanced nations. Nonetheless, only a negligible proportion of individuals with early-stage CKD are cognizant of their affliction.

3. SYSTEM ANALYSIS:

Chronic Kidney Disease (CKD) is a grave matter in healthcare and can be remedied if identified in its early phases. It is often not realized that the medical evaluations conducted for various reasons possess information that is significant regarding kidney diseases. Thus, the characteristics of the different medical evaluations are scrutinized to ascertain which characteristics might offer useful information concerning the sickness. The objective is to gauge the magnitude of the issue and employ this knowledge to construct a machine learning model that foretells Chronic Kidney Disease.

3.1 EXISTING SYSTEM

Chronic Kidney Disease, referred to as chronic renal malady, is a rising issue that has come to occupy a prominent place of concern. The fact that a person can only endure without kidneys for a limited time span of 18 days on average, accentuates the critical requirement for kidney transplantation and dialysis. It is imperative to have efficient techniques for the prompt prediction of Chronic Kidney Disease.

3.2 PROPOSED SYSTEM

In this study, the use of machine learning techniques for predicting Chronic Kidney Disease (CKD) was examined. A comprehensive workflow was developed, which includes steps such as data preprocessing, addressing missing values through collaborative filtering, and attribute selection. Of the

11 machine learning methods evaluated, the Random Forest Classifier and the Naive Bayes Classifier were determined to be the most accurate and exhibit the least attribute bias. The study also considered the practical aspects of data collection, highlighting the need for incorporating domain expertise in the application of machine learning for CKD prediction.

4. SYSTEM DESIGN:

4.1 SYSTEM ARCHITECTURE

Chronic Kidney Disease (CKD) is a prevalent ailment that, if detected in its early stages, can be effectively treated. However, individuals are frequently oblivious to the fact that medical tests performed for different reasons can comprise significant information regarding kidney ailments. Thus, an examination was carried out on the attributes of various medical tests to determine which attributes contain beneficial information regarding the illness. The results of the examination demonstrate that this information can be used to assess the gravity of the ailment and construct a machine learning model for predicting Chronic Kidney Disease.

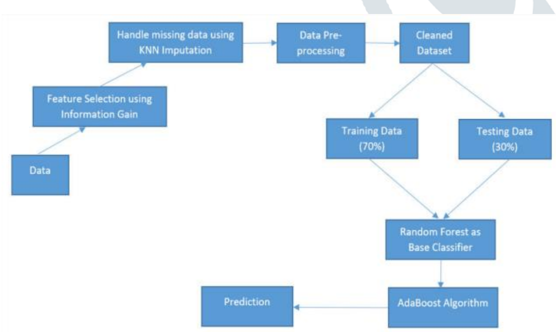


Figure 4.1: Architecture diagram of CKD

4.2 UML DIAGRAMS

4.2.1 USE CASE DIAGRAM

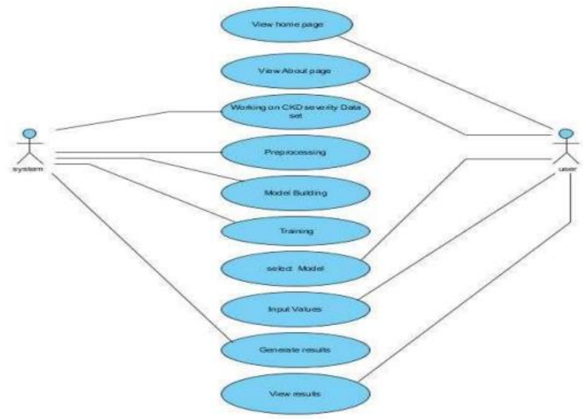


Figure 4.2.1 CKD Use Case Diagram

4.2.2 SEQUENCEDIAGRAM

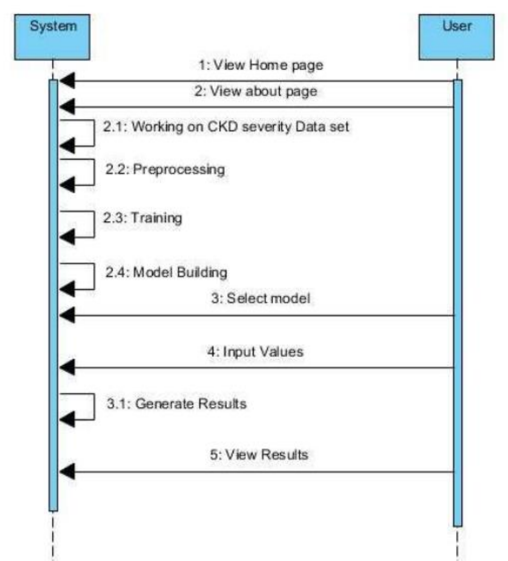


Figure 4.2.2: CKD Sequence diagram

4.2.3 ACTIVITY DIAGRAM

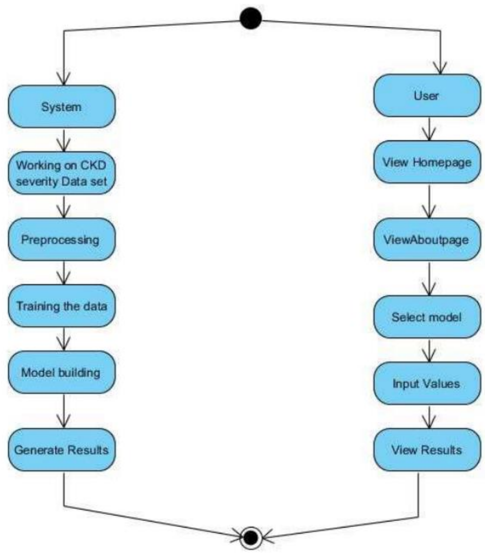


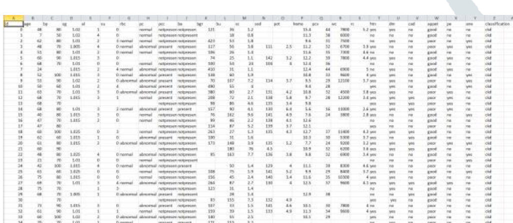
Figure4.2.3: CKD Activity Diagram

CLASS DIAGRAM:

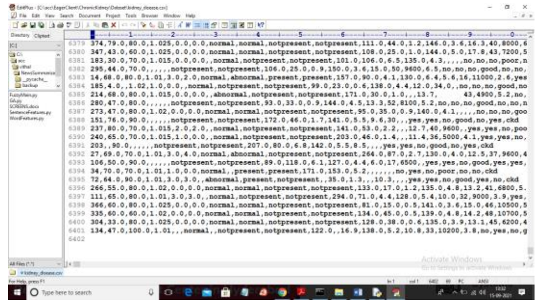
Figure 4.2.4: CKD Class Diagram

5. SCREEN SHOTS

Dataset:



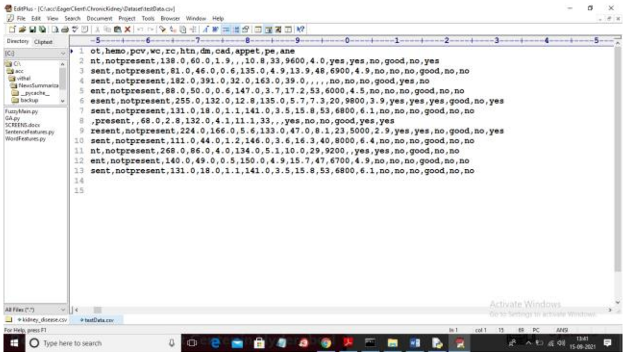
In this effort, the diagnosis of Chronic Kidney Disease (CKD) is realized by employing a hybrid machine learning algorithm that employs Min_Max random values. This approach has led to a substantial augmentation of the dataset, now boasting over 6000 records. A visual representation of the altered dataset is appended herein.



The detection of Chronic Kidney Disease was attained through the implementation of a hybrid machine learning algorithm that consisted of a conglomeration of Support Vector Machine, Random Forest, Decision Tree, Bagging Classifier, and AdaBoost Classifier techniques. For the optimization of features, the usage of the Principal Component Analysis algorithm was executed for the purpose of feature selection.

1) The following steps have been devised to bring forth the implementation of the project in question:
2)
3) The dataset pertaining to Chronic Kidney Disease is to be uploaded onto the application.
4) The preprocessing of the dataset shall be executed by replacing all missing values with 0, and transforming non-numeric labels into numeric format via the utilization of the Label Encoder. The processed dataset shall then be divided into two separate portions, with 80% allocated for training and the remaining 20% reserved for testing the accuracy of the trained algorithm.
5) The Hybrid Machine Learning Algorithm shall then be initiated on the processed data to construct a model that has been trained.
6) The generation of a Confusion Matrix Graph shall take place through the prediction of class labels for the test data and calculation of the corresponding confusion matrix.
7) The Accuracy Comparison Graph shall be plotted to exhibit the precision, accuracy, recall and FScore of the Hybrid Algorithm.
8) Finally, the prediction of the Disease from the Test Data shall be performed through uploading the test data and utilizing the Hybrid Algorithm to determine the labels of CKD or NO-CKD.

Below is the test data screen shots



The steps involved in the combination of multiple algorithms into a hybrid algorithm are as follows:

Identification of algorithms to be incorporated into the hybrid system.

Integration of outputs of the algorithms and utilization of the integrated data to make predictions.

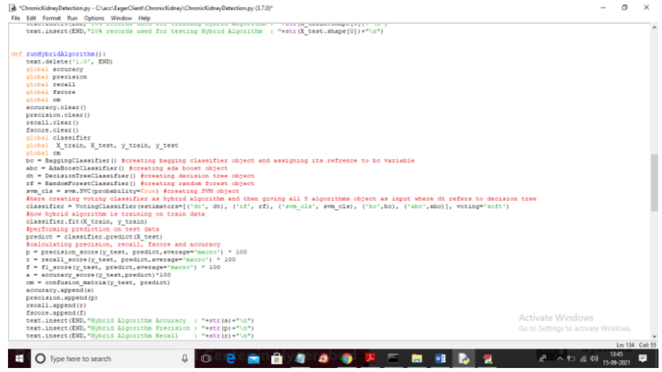
Assessment of the performance of the hybrid algorithm by carrying out tests on test data, and comparison of the results to predictions made by each individual algorithm.

Refinement of the hybrid algorithm, if necessary, by adjusting the weight of each individual algorithm to enhance performance.

Further verification of the proper functioning of the hybrid algorithm and improvement of performance, if necessary, through tests on additional test data.

Continual repetition of testing and refinement processes until the desired results are achieved by the hybrid algorithm.

It can be concluded that the combination of multiple algorithms into a single hybrid system allows for leveraging the strengths of each algorithm, leading to an overall improvement in accuracy and performance of the model.



In above screen read red colour comments to know about hybrid algorithm implementation

SCREEN SHOTS

In order to initiate the project, undertake the procedure as follows:

Search for the file labeled "run.bat".

Execute a double click on the file designated as "run.bat".

Upon executing the above steps, the project shall commence and the anticipated screen shall be presented.

6. CONCLUSION

This investigation seeks to determine the feasibility of identifying Chronic Kidney Disease (CKD) by utilizing machine learning algorithms with a reduced number of tests or features. A machine learning classifier, specifically Logistic Regression, was employed to carry out this examination. The correlation between variables was scrutinized in order to decrease the quantity of features and remove duplication.

The guided study furnished significant knowledge of the application of Logistic Regression to real-life problems and the differentiation between regression and classification problems. Moreover, it deepened comprehension of the treatment of missing values and the significance of Label encoding. The Pickle library, which was previously unfamiliar, was also introduced through this project.

Further advancement of this research can be achieved by incorporating more categories and enhancing its efficiency. Employing various classifiers on the data set will offer a clearer comprehension of the ideal classifier for this study.

Future Scope:

This software enables individuals to forecast if they are affected by Chronic Kidney Disease (CKD). With this prediction, necessary steps can be undertaken to cure CKD before the kidneys completely fail.

REFERENCES

https://www.kaggle.com/code/kianwee/exploratory-data-analysis-chro nic-kidney-disease/notebook

Risk Factors for Development and Progression of Chronic Kidney Disease - PMC (nih.gov)

Shaikh G, Sehgal R, Sandhu S, Vaddineni S, Fogel J, Rubinstein S. Changes in osmol gap in chronic kidney disease: an exploratory study. Ren Fail. 2014 Mar;36(2):198-201. doi: 10.3109/0886022X.2013.838052. Epub 2013 Oct 11. PMID: 24111718.

Lee M, J, Chang T, I, Lee J, Kim Y, H, Oh K, -H, Lee S, W, Kim S, W, Park J, T, Yoo T, -H, Kang S, -W, Choi K, H, Ahn C, Han S, H: Urine Osmolality and Renal Outcome in Patients with Chronic Kidney Disease: Results from the KNOW-CKD. Kidney Blood Press Res 2019;44:1089-1100. doi: 10.1159/000502291.