IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Detection Of Chronic Kidney Disease (CKD) Using ML Algorithms

Mr.K.Shankar [#1], Ms. P.Bhargavi [#2], Mr.M.Manikanta [#3], Ms.T.Udayasri [#4], Mr.S.Sujith [#5],

[#1]Associate Professor, [2, 3,4,5 B.Tech] Students

Department of Computer Science & Engineering,

NS Raju Institute of Technology (NSRIT), Sontyam,Pendurthi- Anandapuram Highway, Visakhapatnam-531173

ABSTRACT

A major global health issue with a high morbidity and mortality rate, chronic kidney disease (CKD) also causes other diseases. Patients frequently overlook the disease in the early stages of CKD since there are no evident symptoms. Early diagnosis of CKD enables patients to receive effective treatment in time to slow the disease's progression. Due of their rapid and precise recognition performance, machine learning models can help physicians attain this goal in an efficient manner. In this paper, we suggest a machine learning approach for CKD diagnosis. The Irvine(UCI) machine learning repository provided the CKD data set, which contains a significant amount of missing values.. Although patients may overlook particular measurements for a variety of reasons, missing data are frequently observed in real-world medical settings. Six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbour naive bayes classifier, and feed forward neural network) were used to create models after the incomplete data set was successfully filled in. Of these machine learning models, random forest had the highest accuracy. We created an integrated model that combines logistic regression and random forest by employing perceptron, which has the best accuracy; as a result, we hypothesised that this methodology may be used to more complex clinical data for disease diagnosis.

KEYWORDS:

Logistic Regression, Random Forest Classifier, Decision Tree Classifier, Chronic Kidney Disease.

1. INTRODUCTION

The adverse effects of chronic kidney disease (CKD), which include renal failure, cardiovascular disease, and early death, are a serious public health concern worldwide [1]. Chronic kidney disease (CKD), which rose from 27th place in 1990 to 18th place in 2010 according to the Global Burden of Disease Study (GBDS), is one of the top causes of death worldwide [2]. Over 500 million individuals globally suffer from chronic renal disease [3, 4], with South Asia and sub-Saharan Africa bearing a disproportionately high burden [5]. In high-income countries, there were 110 million people with CKD (men 48.3 million, women 61.7 million), but in low- and middle-income nations, there were 387.5 million [6].

In Bangladesh, a heavily populated developing nation in Southeast Asia, chronic kidney disease rates continue to climb. In a global survey of six countries, including Bangladesh, the prevalence of CKD was estimated to be 14% [7]. A 26% prevalence of chronic renal disease was found in a different study among urban Dhaka residents over the age of 30 [8], and a 13% prevalence was found in a different study among urban Dhaka inhabitants over the age of 15 [9]. One-third of rural individuals in Bangladesh were at risk of developing CKD, which was frequently misdiagnosed at the time, according to a community-based prevalence survey conducted there in 2013 [10].

When a person has a specific illness, they must make an expensive and time-consuming trip to the doctor. Also, if the user is far from a doctor or hospital, it could be challenging for them to diagnose their illness. So, it might be simpler for the patient and the process if the aforementioned procedure could be carried out utilising an automated programme that can save time and money. There are other systems that use data mining techniques to examine the patient's risk status and predict heart-related diseases. A web-based tool called Disease Predictor makes health predictions for users based on their reported symptoms.

Data sets for the Illness Prediction System were gathered from several websites that deal with health. The user will be able to determine the likelihood of the disease based on the listed symptoms with the aid of the Disease Predictor. People are always interested in learning new things as internet usage increases daily. When an issue happens, people always try to turn to the internet for assistance. Compared to hospitals and doctors, people have access to the internet. When someone has a certain ailment, they do not immediately have a choice. As a result, this method can benefit the populace as they have constant access to the internet.

2. LITERATURE SURVEY

To conduct a literature survey on the detection of chronic kidney disease (CKD), we can start by identifying key terms and phrases related to your research question, such as "chronic kidney disease detection," "early diagnosis of CKD," "novel CKD biomarkers," etc. You can then search for these terms in online databases such as PubMed, Google Scholar, and Scopus to find relevant articles. Chronic kidney disease (CKD) is a condition in which the kidneys are damaged and are unable to filter waste and excess fluids from the body efficiently. This can lead to a range of health problems, including anaemia, high blood pressure, and nerve damage. There are several causes of CKD, including diabetes, high blood pressure, and glomerulonephritis (inflammation of the kidneys).

It is also more common in older adults and those with a family history of kidney disease. Treatment for CKD usually involves lifestyle changes, such as eating a healthy diet and exercising regularly, and may also include medications to control blood pressure and manage symptoms. In advanced stages of the disease, dialysis or a kidney transplant may be necessary. There is ongoing research into the causes and treatment of CKD. Some studies have suggested that certain dietary interventions, such as reducing the intake of protein or increasing the intake of certain minerals, may be beneficial for those with the condition. Other research has focused on the use of medications to slow the progression of CKD and prevent complications. It is important for individuals with CKD to closely monitor their condition and follow their treatment plan to help manage the disease and reduce the risk of complications.

3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing system there was no proper method to identify the chronic kidney disease prediction using data mining algorithms. The following are the main limitations in the existing system.

LIMITATION OF PRIMITIVE SYSTEM

1. More Time Delay in finding the route cause of kidney diseases
2. There is no prevention technique due to late prediction.
3. There is no early prediction of chronic kidney disease.

4. There is no method to identify the kidney diseases using ML algorithms

4. PROPOSED SYSTEM AND ITS ADVANTAGES

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables.

It performs better results for classification problems. Random Forest works in two-phase first is to create the random forest by combining N decision tree and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree and assign the new data points to the category that wins the majority votes.

5. IMPLEMENTATION PHASE

Here in this section we are going to implement several algorithms which are used to implement the current application. They are discussed in detail:

1) SVM: "Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well .

The Working process can be explained in the below steps and diagram:

Step 1: Load the important libraries.

Step 2: Import dataset and extract the X variables and Y separately.

Step 3: Divide the dataset into train and test.

Step 4: Initializing the SVM classifier model.

Step 5: Fitting the SVM classifier model.

Step 6: Coming up with predictions.

Step 7: Evaluating model's performance

2. Decision Tree:

The decision tree algorithm is a supervised machine learning algorithm. It can be used for both a classification problem as well as for regression problem. The general objective of using Decision tree is to create a model that predicts classes or values of target variables by generating decision rules derived from training datasets. Decision tree algorithm follows a tree structure with roots, branches and leaves. The attributes of decision making are the internal nodes and class labels are represented as leaf nodes. Decision tree algorithm is easy to understand compared with other classification algorithms.

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm.

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3.

Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

3. Random Forest:

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems. Random Forest works in two-phase first is to create the random forest by combining N decision tree and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

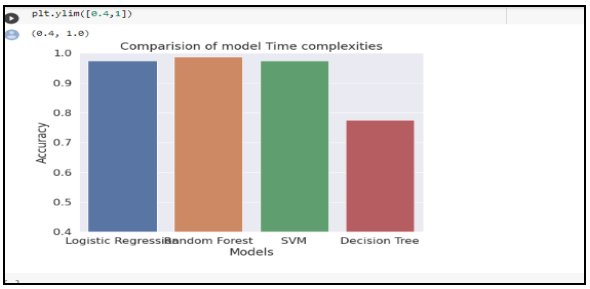
Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree and assign the new data points to the category that wins the majority votes.

6. EXPERIMENTAL RESULTS

In this section we try to design our current model using Python as programming language and we used Google Collab as working environment for executing the application.

COMPARISION OF ML-ALGORITHMS



From the above window we can clearly the comparison of several ML algorithms and we can see Random Forest has good accuracy.

7. CONCLUSION

This system presented the best prediction algorithm to predict CKD at an early stage. The dataset shows input parameters collected from the CKD patients and the models are trained and validated for the given input parameters. Random forest tree Classifier, Decision Tree Classifier, Support Vector Machine (SVM) learning models are constructed to carry out the diagnosis of CKD. The performance of the models is evaluated based on a variety of comparison metrics are being used, namely Accuracy. The results of the research showed that Random Forest Classifier model better predicts CKD in comparison to the other models taking all the metrics under consideration.

8. REFERENCES

[1] M. P. N. M. Wickrama singhe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, 2017, pp. 300-303

[2] R. Devika, S. V. Avilala and V. Subramaniya swamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naïve Bayes, KNN and Random Forest," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 679-684. 42

[3] G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 973-979.

[4] Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using

Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar,Bangladesh, 2019, pp. 1-6.

[5] S. Vijayarani, S. Dhayanand, "KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS", International Journal of Computing and Business Research (IJCBR), vol.6, no. 2, 2015.

[6] Deepti Sisodiaa , Dilip Singh Sisodia, Prediction of Diabetes using three classification algorithms, Computational International Conference on Data Intelligence and Science (ICCIDS 2018)

[7] Maryam Soltanpour Gharibdousti Kamran Azimi, Saraswathi Hathikal, Dae H Won, Prediction of Chronic Kidney Disease using Data mining techniques, 2017 Industrial and Systems Engineering Conference.

[8] Krupa Joel Chabathula, Jaidhar C.D, Ajay Kumara M.A Comparative Study of Principal Component Analysis Based Intrusion Detection Approach Using Machine Learning Algorithms , 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN).

[9] Guneet Kaur , Ajay Sharma , Predict Chronic Kidney Disease Using
Data Mining Algorithms in Hadoop,International Journal of Advanced
Computational Engineering and Networking, ISSN: 2320-2106,
Volume-5, Issue- 6, Jun-2017

[10]Chronic_Kidney_DiseaseDataSet",Archive.ics.uci.edu,2015.Available:http s://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

[11]www.kidneyfund.org/kidneydisease/chronic-kidney-diseaseckd/