CHRONIC KIDNEY DISEASE PREDICTION USING PYTHON AND MACHINE LEARNING

Naveya Bhutani

DEPARTMENT OF INFORMATION TECHNOLOGY

MAHARAJA AGRASEN INSTITUTE OF TECHNOLOGY

Abstract

One of the most serious illnesses nowadays is chronic kidney disease, for which a correct diagnosis must be made as soon as possible. Machine learning techniques are currently used in medicine. The clinician can identify the ailment early with the use of a machine learning system. In this paper , chronic kidney disease prediction has been covered. Disease prediction and detection is a vital and difficult subject since it aids in early disease diagnosis by assisting pathologists and medical professionals in their decision-making.In recent studies and research, "The Artificial Neural Network" provides a useful method to address a variety of everyday issues in life or in many different fields, such as the medical field, where it can be used to anticipate a specific disease based on some provided data. In this article, we use the ANN approach to describe a system for predicting chronic kidney disease.This neural network output lets us know whether a patient has chronic renal illness or not. Artificial neural networks produce outcomes that are more accurate than those of other machine learning algorithms in the aforementioned topic of kidney disease prediction after extensive investigation.

Keywords : Chronic Kidney Disease , Machine Learning , Artificial Neural Network(ANN)

Introduction

Nowadays, people's changing eating habits and work cultures are leading to a number of health problems. Even if people are increasingly concerned about their health, they tend to pay attention more when sickness symptoms start to show. However, because it doesn't present with any distinctive symptoms, chronic kidney disease is difficult to diagnose. It is really challenging to predict otherwise. The signs of this illness are not readily apparent in the early stages. It is possible that the symptoms might not be noticable until a considerable amount of damage has been done to kidneys. The primary objective of treating CKD is frequently to halt the underlying cause in order to stop the progression of kidney damage.Without artificial filtration (dialysis) or a kidney transplant, end-stage renal failure, which is deadly, can develop from chronic kidney disease. Machine learning achieves this by creating prediction models utilising past CKD patient data. This will enable us to detect CKD before it has catastrophic consequences.

The kidney is an essential organ in the human body with key roles in excretion and osmoregulation. The kidneys are part of the excretory system, which collects and eliminates all waste and hazardous substances the body produces. The prevalence of morbidity, death, and hospital admission rates among patients globally have all been dramatically impacted by chronic kidney disease (CKD), a non-communicable illness. As it grows swiftly, it will soon rank among the leading killers globally. In India, there are 1 million cases of chronic kidney disease (CKD) annually. Kidney function gradually deteriorates as a result of this dangerous disorder. CKD is a condition in which the kidneys gradually lose their ability to function over a number of years. A person will acquire renal failure that is irreversible. The patient may exhibit the following symptoms if CKD is not identified and treated at an early stage: Hypertension, anaemia, brittle bones, inadequate nutrition, poor health, and

nerve damage reduced immune response because harmful levels of fluids, electrolytes, and wastes can accumulate in your blood and body at advanced stages. Therefore, it's critical to spot CKD in its early stages. Different models and algorithms have been developed for the prediction. Here, I've experimented with using an alternative Artificial Neural Networks technique to try and predict CKD.

Literature Survey

Numerous studies use various categorization methods to predict CKD, and these researchers receive the predicted results from their model.

S. Ramya and Dr. N. Radha [2] used various machine learning classification techniques to increase diagnosis speed and accuracy. The classification of several stages of CKD according to their severity is the focus of the proposed investigation. by examining various techniques, including RBF, RF, and Basic Propagation Neural Network. The analysis's findings show that the RBF algorithm produces 85.3% accuracy, outperforming the other classifiers.

Results from several models have been compared by Gunarathne W.H.S.D et al. Finally, they came to the conclusion that the Multiclass Decision forest algorithm provides greater accuracy than other algorithms, with a 99% accuracy rate for the dataset that has been condensed to just 14 attributes.

An innovative method developed by Asif Salekin and John Stankovic [5] makes use of a machine learning algorithm to identify CKD. They receive results based on a dataset with 400 records and 25 attributes that indicate whether a patient has CKD or not. To obtain results, they employ neural networks, random forests, and k-nearest neighbours. They employ a wrapper approach, which accurately detects CKD, for feature reduction.

In the chronic kidney disease dataset , S.Dilli Arasu and Dr. R. Thirumalaiselvi [3] have worked on missing values. Missing values in the dataset will make our model less accurate and provide less accurate prediction outcomes. They solved this issue by recalculating the stages of CKD, but in the process, they came up with unknown values. They recalculated the values in place of the missing ones.

Dataset and Attributes : The CKD dataset [4] is retrieved from the UCI repository for this paper. 400 patient records with 25 attributes are part of this collection. These 25 characteristics are all essential characteristics associated with CKD illness. out of 25 characteristics' worth of patient records. These 25 characteristics are all essential characteristics associated with CKD illness.

Problem Definition

Determine a person's prognosis for chronic kidney disease based on the person's health data, such as haemoglobin, albumin, packed cell volume, hypertension, etc. Create a custom Python program utilising artificial neural networks(ANN) to forecast and categorise individuals as having chronic kidney disease (CKD) or not .

Objective

Everyone in the modern period strives to be health conscious, but because of their job and hectic schedules, they only pay attention to their health when certain symptoms arise. However, because CKD is a disease with no symptoms at all, or in some cases, no symptoms at all, it is difficult to predict, detect, and prevent such a disease, which could result in long-term health damage. However, machine learning offers hope in this situation because it excels at prediction and analysis.

The goal of this project is to predict CKD using machine learning. By applying various classification algorithms, medical test results of CKD patients can be used to suggest diet programmes. We create a model using the Artificial Neural Network, compile it, store it, and then use it to forecast chronic kidney disease.Two machine learning libraries are using Keras and sklearn Chronic Kidney Disease dataset has been taken from the UCI repository. Then the dataset is cleaned and manipulated. All the null or empty values are replaced with mean. Then we use Split and Scale. Then the data is divided into training data and testing data. Then the model is built. After building the model , the model is compiled and then saved. Plot the loss and accuracy of the model to visualize how well the model did on the training dataset. Get the training and test datashape. Loop through the saved model. Then determine the accuracy, loss, prediction, and original values of the model on the test data.

Methodology
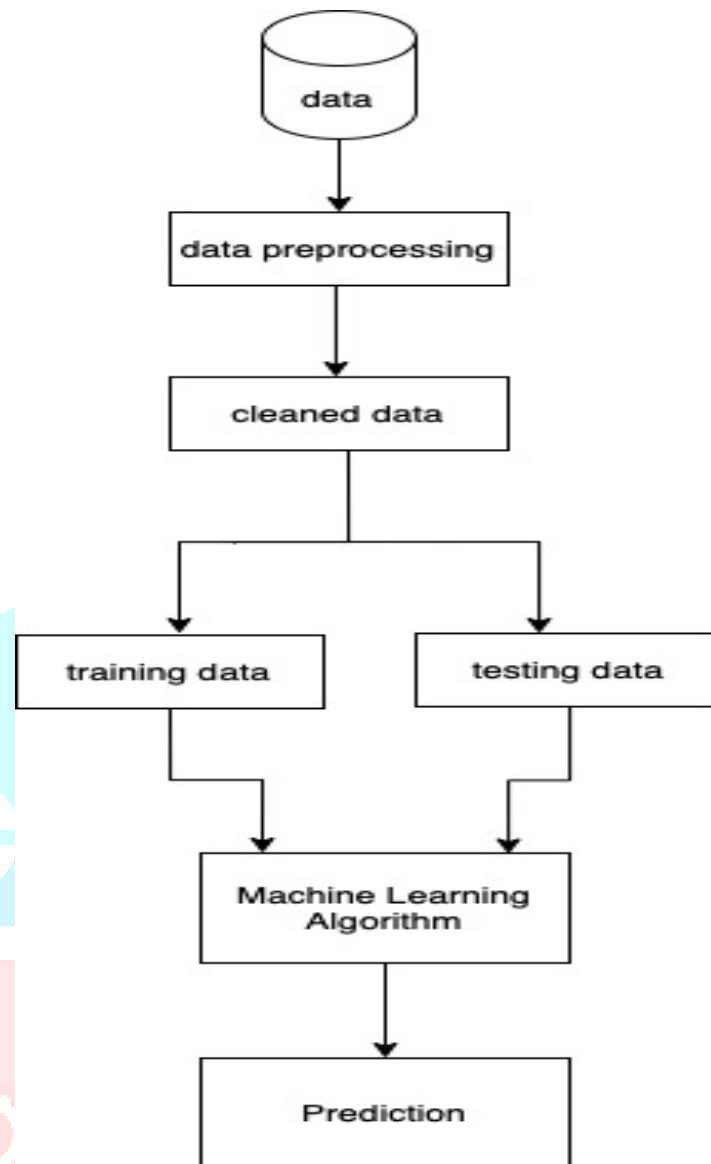
Import libraries needed like glob , numpy , pandas , keras and sklearn.

Dataset is taken from the UCI repository. It is the dataset for prediction of chronic kidney disease using machine learning algorithm . It has 25 attributes age , bp , sg , al , su , rbc , pc , pcc , ba , bgr , bu , sc , sod , pot , hemo , pcv , wc , rc , htn , dm , cad , appet , pe , ane , classification . We only take important attributes required to build the model.

Here , bp stands for blood pressure , sg stands for specific gravity , al stands for albumin , su is the glucose level , rbc stands for red blood cell , pc stands for pus cells , psc stands for pus cell clumps , bu stands for blood urea , sc stands for serum creatinine , sod stands for sodium , pot stands for potassium , hemo stands for haemoglobin , pcv stands for packed cell volume , wc stands for white blood cells , rc stands for red blood cells , htn stands for hypertension , ane stands for anaemia , appet stands for appetite .

After loading the "kidney_disease.csv" dataset , it is time for data preprocessing.

Data preprocessing :
Data cleaning : When we gather open source raw data of CKD patients it does not contain the name of all attributes or has missing values or blank spaces in the dataset. So first we need to handle such values.
Here we will replace the missing values with the mean of the other values. We could also use the approach where we remove the missing values , but it might lead to losing important information.

Data Reduction : The dataset we have loaded has 25 attributes . All the attributes in the dataset are not needed. We only need some attributes out of all these to successfully carry out the prediction. Here , we keep the columns 'sg' , 'al' , 'sc' , 'hemo' , 'pcv' , 'wbcc' , 'rbcc' , 'htn' , 'classification' . These are the attributes we'll be using to build the predictive model.

Split and Scale the data :
To transform non numeric data into columns we use Label encoder.
We split the data into independent(X) dataset that features and dependent(y) dataset the target.
Here we use the feature scaling. Feature scaling is used to normalize the range of independent variables or features of data.
Min-Max scaler method is used . It scales the dataset so that all input features lie between 0 and 1.

$$x' = \frac{x - min(x)}{max(x) - min(x)}$$

Where x is the original value and x' is normalised value.

Once we are done with this much we divide the data into **training and testing data**. We split the dataset into 80% training (X_train and y_train) and 20% testing (X_test and y_test) data sets , and shuffle the data before training.

Build the Model (ANN)
For the model's architecture we'll add 2 layers. We have the first layer with 256 and the 'ReLu' activation function with a normal distribution initializer for the weights.
We must also specify the number of features/columns in the data set len(X.columns).
In the second layer we will have one neuron which uses 'hard_sigmoid' activation function. The second layer is also the last layer here.

Now , **compile the model** and give "Binary_crossentropy" which is the loss function used for binary classification. It helps us evaluate how well the model did on the training and then using an optimizer it tries to improve on it.

We have used the 'adam' optimizer . To see how well the model does we will use some metrics on the model's accuracy.

Using the training datasets (X_train and y_train) we **train the model** . 200 epochs are given and a batch size equal to number of patients in the dataset.

Next step is to **save the model** .

Now, we'll **visualise** how well the model did on the training dataset . Plot accuracy and loss versus epoch for the same .
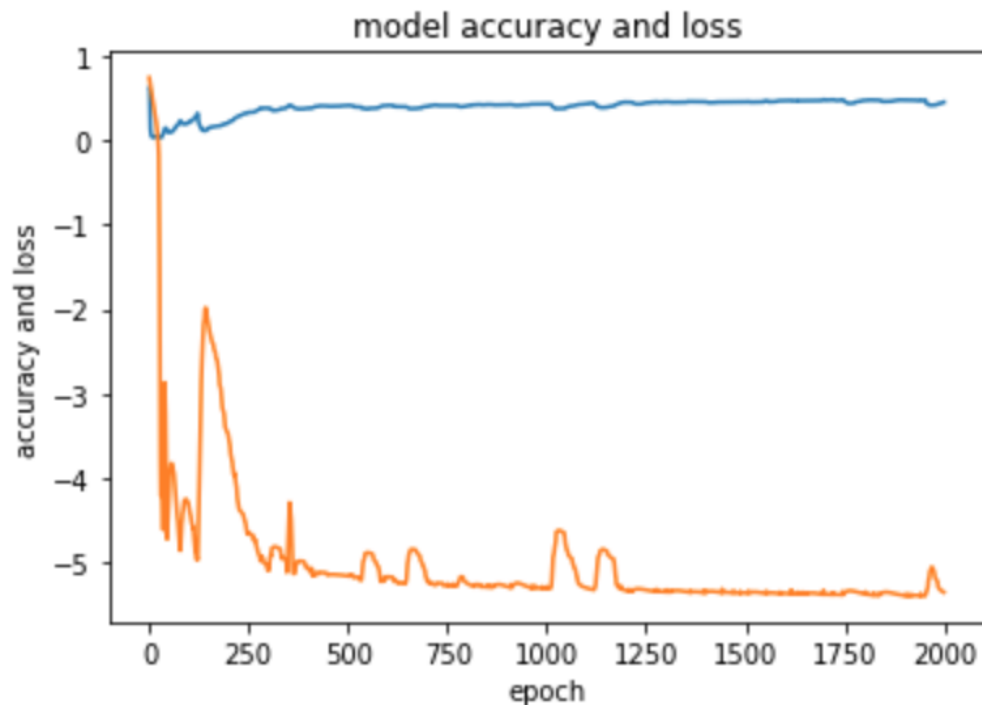
Now we get the **shape of training data and testing data** .

Now it is time for the **Prediction** . We get the model's accuracy, loss, prediction and original values on the test data .

Conclusion :
We have built a program that predicts the presence of Chronic Kidney Disease . It is a good model as our experimental results show the loss and accuracy versus epochs have a good graph. This can basically act as the foundation of CKD prediction. It is a model made using the concept of artificial neural network.

Future scope:
This can be viewed as the foundation of the CKD patient healthcare system. It will aid in CKD prediction so that the appropriate precautions can be taken to help stop it from getting worse. It will benefit our quality of life and help us avoid serious health hazards.
Artificial neural network (ANN) is created to function similarly to how the human brain operates. In fact, the core of deep learning is an ANN. Deep learning can therefore be added to this in the future to improve the outcomes of this research. This could be used in the future in conjunction with deep learning.

References:

[1]. Gunarathne W.H.S.D,Perera K.D.M, Kahandawaarachchi K.A.D.C.P, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and
Forecasting through Data Analytics for Chronic Kidney Disease (CKD)",2017 IEEE 17th International Conference on Bioinformatics and Bioengineering.
[2]. S.Ramya, Dr. N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Proc. International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 1, January 2016.
[3]. S.Dilli Arasu and Dr. R.Thirumalaiselvi, "Review of Chronic Kidney Disease based on Data Mining Techniques",International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13498-13505
[4]. L. Rubini, "Early stage of chronic kidney disease UCI machine learning repository,"2015. [Online].Available:http://archive.ics.uci.edu/ml/datasets/Chronic Kidney Disease.
[5]. Asif Salekin, John Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," Proc. IEEE International Conference on Healthcare Informatics (ICHI), IEEE, Oct. 2016, doi:10.1109/ICHI.2016.36.