IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GUI BASED IMPROVED MACHINE LEARNING MODEL FOR CHRONIC KIDNEY DISEASE PREDICTION

Naveya Bhutani
DEPARTMENT OF INFORMATION TECHNOLOGY
MAHARAJA AGRASEN INSTITUTE OF TECHNOLOGY

Abstract

Chronic kidney disease is a significant health concern in modern times, and it is critical to diagnose it correctly and promptly. Medical professionals currently employ machine learning techniques to aid in disease identification at an early stage. The focus of this is on the prediction of chronic kidney disease. Detecting and predicting diseases is a complex and vital aspect of medicine that supports pathologists and clinicians in decision-making by allowing early diagnosis. Artificial Neural Network (ANN) has been demonstrated to be a valuable tool in various fields, including medicine, by predicting specific diseases based on provided data. This study employs the ANN approach to develop a chronic kidney disease prediction system. The neural network output provides information on whether a patient has chronic renal disease. After extensive research, it has been shown that artificial neural networks are more accurate than other machine learning algorithms in predicting kidney disease. A comparative analysis has also been shown . A GUI to predict CKD on the basis of input values is present.

Keywords : GUI , Machine Learning Model , Kidney Disease

Introduction

In modern times, the changing eating habits and work culture of people have led to numerous health issues. Despite growing concerns about their health, individuals tend to prioritise paying attention to sickness symptoms when they arise. However, diagnosing chronic kidney disease is challenging since it lacks distinctive symptoms, making it difficult to predict. Signs of this disease are not noticeable in the early stages, and sometimes symptoms only manifest after significant damage to the kidneys has occurred. The primary goal of CKD treatment is often to stop the progression of kidney damage by addressing the underlying cause. In the absence of artificial filtration, such as dialysis, or a kidney transplant, end-stage renal failure, which is fatal, can develop from chronic kidney disease. To detect CKD before it reaches catastrophic consequences, machine learning utilises past CKD patient data to create prediction models.

The kidney is a crucial organ in the human body responsible for excretion and osmoregulation. As part of the excretory system, it eliminates all waste and harmful substances produced by the body. Chronic kidney disease (CKD) is a non-communicable disease that has had a significant impact on morbidity, mortality, and hospitalisation rates globally. It is rapidly becoming one of the leading causes of death worldwide. In India, there are approximately one million cases of CKD each year. CKD is a dangerous disorder that gradually deteriorates kidney function over several years, leading to irreversible renal failure. If not detected and treated early, the patient may exhibit symptoms such as hypertension, anaemia, brittle bones, poor nutrition, nerve damage, reduced immune response, and other health problems. In advanced stages, harmful levels of fluids, electrolytes, and wastes can accumulate in the body and bloodstream. Thus, early detection of CKD is critical. Several models and algorithms have been developed to predict CKD, including Artificial Neural Networks. In this study, an alternative ANN approach was utilised to 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.

Asif Salekin and John Stankovic [5] use a novel approach to detect CKD using a machine learning algorithm. They get results on a dataset which has 400 records and 25 attributes which gives the result of a patient having CKD or not CKD. They use k-nearest neighbours, random forest and neural networks to get results. For feature reduction they use a wrapper method which detects CKD with high accuracy.

Pinar Yildirim [7] searches for the effect of class imbalance when we train the data by using the development of neural network algorithms for making medical decisions on chronic kidney disease. In this proposed work, a comparative study was performed using a sampling algorithm. This study reveals that the performance of classification algorithms can be improved by using the sampling algorithms.

Problem Definition

Objective

In the contemporary era, individuals strive to prioritize their health, but their busy schedules and work commitments often make them neglect it until they experience specific symptoms. Unfortunately, CKD is a disease that may not present any symptoms or only mild ones, making it challenging to predict, detect, and prevent the condition, which can result in long-term health consequences. However, machine learning techniques offer a glimmer of hope in this scenario as they are adept at prediction and analysis.

The objective of this project is to utilize machine learning for the prediction of CKD. By employing various classification algorithms, the medical test results of CKD patients can be utilized to recommend specific diet programs. To accomplish this, an Artificial Neural Network model is developed, compiled, stored, and used for predicting chronic kidney disease. Two machine learning libraries, Keras and sklearn, are employed, and the Chronic Kidney Disease dataset is obtained from the UCI repository. We have used Pearson Correlation Coefficient.The Pearson correlation coefficient is a statistical metric that gauges the degree of linear association between two variables. Then a heatmap is generated for better understanding of the correlation. Correlation heatmaps refer to a graphical representation that displays the intensity of connections among numeric variables.The dataset is then cleansed and manipulated by substituting all null or empty values with the mean. The data is split and scaled, followed by division into training and testing datasets, after which the model is constructed, compiled, and saved. The loss and accuracy of the model are plotted to visualize its performance on the training dataset. The shape of the training and test data is obtained, and the saved model is looped through to determine its accuracy, loss, prediction, and original values on the test data. We have also build a UI for the prediction of presence of CKD using the model we have made on the basis of inputs provided by the user.

Methodology

First step would be to create the UI in anvil using the different form components available there which are very simple to use .

Next , write the anvil code to connect the button with the desired click function , so that whatever desired function we call works from the code of ML model .

After this we have to create the uplink and link with the google colab.

Now open google colab code and following steps are performed.

Install the Anvil uplink with the command : !pip install anvil-uplink

Import the anvil server with the command along with the unique server id from uplink :

```
import anvil.server
```

Then we will import the needed libraries like pandas , keras , sklearn , numpy , etc.

Next step would be to load the dataset taken from the UCI repository . Chronic Kidney Disease Prediction Dataset .

Preprocessing of Data will take place after loading the dataset.

Preprocessing is done by cleaning the data and reduction in data . There are several missing attributes or values or blank spaces in the dataset. So we need to handle them carefully .
Here we will remove the missing values to obtain the accuracy we need .
For the data reduction we remove the unnecessary attributes and only keep the needed attributes for correct prediction. Columns we have retained are : specific gravity , albumin , serum creatinine , haemoglobin , packed cell volume , white blood cell count , red blood cell count , hypertension , classification.

Next step will be to Split and Scale the data , using Label encoder we transform non numeric data to numeric . We will use feature scaling to normalise a range of independent variables or features of the data .
We will use the Min-Max scaler method . It will scale the data input to lie between 0 and 1.

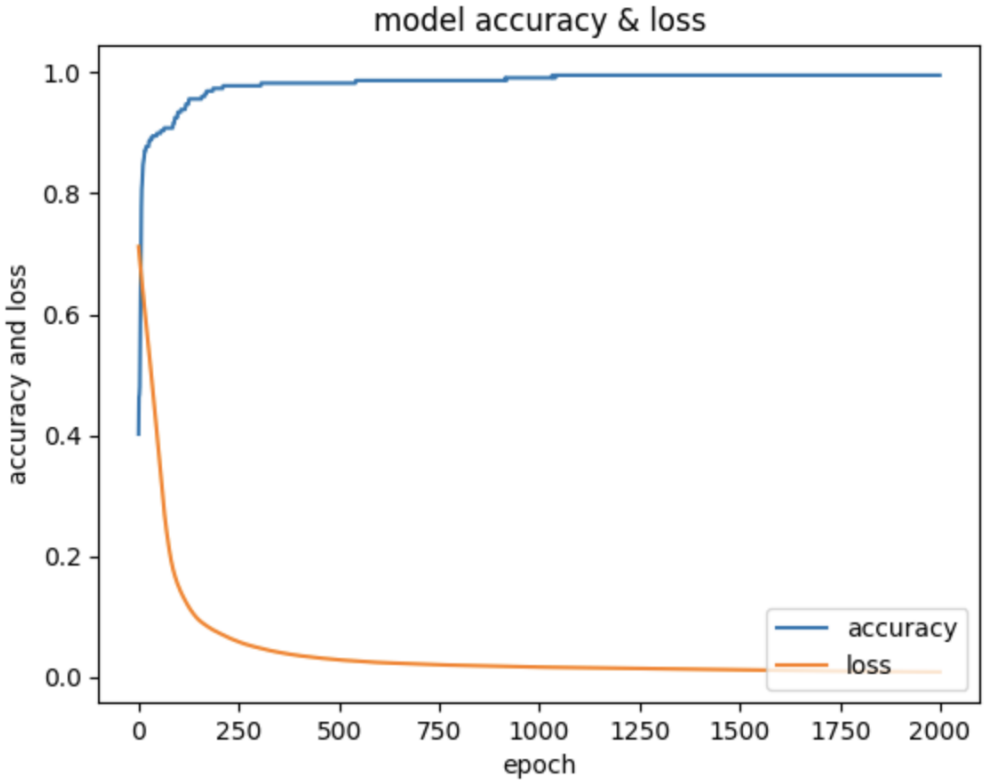
Now let's divide the data into testing and training data.

Next step is to build the model. "ReLu" Function is used here to build the model.
In the other layer we use "Hard sigmoid" Function .

Next will be the compilation of model. Optimizer used will be the 'adam optimizer'

Next we will train the model . 2000 epochs will be used for the same.

After training we will save the model .
Next is to visualise the accuracy and loss of the model.



Now after visualising the model we will evaluate the model . Here we get:

Test Loss : 0.010090748779475689

Test Accuracy : 1.0

Now it's time to compare it with other models. This is our comparison result :

Decision Tree Model:

Precision: 0.9833743842364532

Recall: 0.9827586206896551

F1-Score: 0.9827740377203351

Neural Network Model:

Precision: 1.0

Recall: 1.0

F1-Score: 1.0

KNN Model:

Precision: 1.0

Recall: 1.0

F1-Score: 1.0

Random Forest Model:

Precision: 1.0

Recall: 1.0

F1-Score: 1.0

Next we need to create a prediction function that will help us get predictions when the appropriate button is clicked in the GUI . We will link these with our GUI to make predictions.

So the next step would be to link the predict button with the anvil server.

Next we will display the prediction result .

Conclusion

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.
[6] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
[7] Pinar Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," Proc. 41st IEEE International Conference on Computer Software and Applications (COMPSAC), IEEE, Jul. 2017, doi: 10.1109/COMPSAC.2017.84