Chronic Kidney Disease Prediction Using Machine Learning

Sarthak Arora[1]

Student[1]

Department Of Information Technology

Maharaja Agrasen Institute Of Technology, Rohini, Delhi India

Abstract

Chronic Kidney Disease prediction is one of the most important issues in healthcare. The most challenging task in day to day life is prediction in the medical field. Chronic Kidney Disease is a disease that doesn't show any symptoms at all or it doesn't show any specific symptoms of a disease. It is hard to detect, predict or prevent such a disease and could lead to health damage permanently, machine learning can be hope in this problem it is best in prediction and analysis. In this research paper, we employ machine learning techniques to predict the chronic kidney disease using clinical data.

We used Logistic Regression algorithm to predict the chronic kidney disease. Technological development which includes machine learning has a great impact on health by effective analysis of many chronic diseases for a more accurate diagnosis. Kidney disease is a main chronic disease associated with hypertension, aging, and diabetes, affecting people of 60 and over.

Its major cause is malfunctioning of the kidney in disposing toxins from the blood. This study analyzes chronic kidney disease using machine learning techniques based on a chronic kidney disease (CKD) data set from the UCI machine learning data warehouse. The data set is reprocessed by completing and normalizing missing data. The most relevant features are selected from the data set for improved accuracy and reduced training time.

Keywords:-Chronic Kidney Disease, Logistic Regression

I.   INTRODUCTION

Computer vision has been one of the most remarkable breakthroughs for the machine learning and in particular for active healthcare applications. Machine learning allows to build the models to quickly analyze data and deliver results for the given data. Healthcare service providers can make better decisions on patient's disease diagnosis and treatment for the particular disease with the help of machine learning. The massive quantities of data are analyzed using machine learning. It delivers faster and more accurate results in order to identify the risks, it may also require additional time and resources to train it proper manner. Supervised machine learning algorithms can be applied to predict the future events with the help of what has been learned in the past to new data using labeled examples. Firstly, known training data set is analyzed with the learning algorithm that produces an inferred function to make predictions about the output values. After sufficient training the system is able to provide targets for any new inputs. Supervised learning algorithms use patterns to predict label values on additional unlabeled data. Machine learning algorithms are classified in two types; they are supervised machine learning algorithms and unsupervised machine learning algorithms. Supervised machine learning algorithms are based on input-output pair's patterns. These algorithms aim to predict output values based on given input values. Supervised machine learning algorithms mainly focuses on classification and regression.

We all know that Kidney is essential organ in human body. Which has main functionalities like excretion and regulations? In simple words we can say that all the toxic and unnecessary material from the body is collected and thrown out by kidney and excretion system. There are approximately 1 million cases of Chronic Kidney Disease (CKD) per year in India. Chronic kidney disease is also called renal failure. It is a dangerous disease of the kidney which produces gradual loss in kidney functionality. CKD is a slow and periodical loss of kidney function over a period of several years. A person will develop permanent kidney failure. If CKD is not detected and cured in early stage then patient can show following Symptoms: Blood Pressure, anemia, week bones, poor nutrition health and nerve damage, Decreased immune response because at advanced stages dangerous levels of fluids, electrolytes, and wastes can build up in your blood and body. Hence it is essential to detect CKD at its early stage but it is unpredictable as its symptoms develop slowly and aren't specific to the disease. Some people have no symptoms at all so machine learning can be helpful in this problem to predict that the patient has CKD or not. Machine learning does it by using old CKD patient data to train predicting model. The earlier disease is detected the better chance of stopping its progression.

II. LITERATURE REVIEW

There are many researchers who work on prediction of CKD with the help of many different classification algorithm. And those researchers get expected output of their model.

Gunarathne W.H.S.D. has compared results of different models. And finally they concluded that the Multiclass Decision forest algorithm gives more accuracy than other algorithms which is around 99% for the reduced dataset of 14 attributes.

S.Ramya and Dr.N.Radha worked on diagnosis time and improvement of diagnosis accuracy using different classification algorithms of machine learning. The proposed work deals with classification of different stages of CKD according to its gravity. By analysing different algorithms like Neural Network, RBF and RF. The analysis results indicates that RBF algorithm gives better results than the other classifiers and produces 85.3% accuracy.

S.Dilli Arasu and Dr. R. Thirumalaiselvi has worked on missing values in a dataset of chronic Kidney Disease. Missing values in dataset will reduce the accuracy of our model as well as prediction results. They find solution over this problem that they performed a recalculation process on CKD stages and by doing so they got up with unknown values. They replaced missing values with recalculated values.

Asif salekin and john stankovic they use novel approach to detect CKD using machine learning algorithm. They get result on dataset which having 400 records and 25 attributes which gives result of patient having CKD or not CKD. They use k-nearest neighbours, random forest and neural network to get results. For feature reduction they use wrapper method which detect CKD with high accuracy.
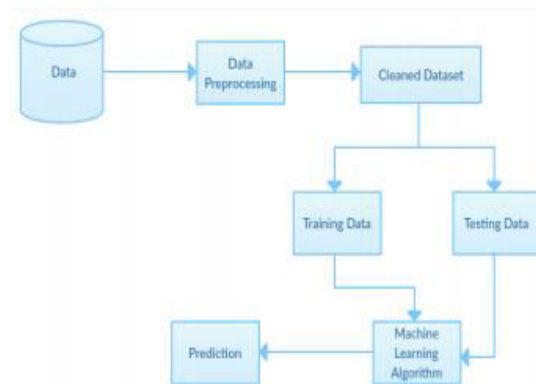
III. METHODOLOGY

Fig1: Working

i) Data set - A Data set of prediction of chronic kidney disease using machine learning algorithm is installed from the UCI repository. The data set presents 400 patient records. Also they include 25 attributes but we take only 14 attributes for building model. Age, Blood pressure, Albumin, Red blood cells, Pus cell, Pus cells clumps, Serum creatinine, Haemoglobin, White blood cell count, Red blood cell count, Anemia, Classification, Appetite, Packed cell volume all this 14 attributes are used to build model.

ii) In the dataset we have a lot of missing data for the columns rbc, pcv,wbcc, rbcc, pot, sod .So , for example, if we want to calculate the missing values of sodium , we can apply Multi-Variate Linear Regression on the rest of the data without considering the other 5 columns.

iii) Training and Testing Dataset: The dataset is divided into two sub datasets both containing 24 attributes. Training data: training dataset is derived from main dataset and it contains 300 out of 400 records in main dataset of CKD. Testing data: testing dataset is of 100 out of 400 records from main CKD dataset.

iv) Classifiers: Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function is also called the sigmoid function. It was invented by statisticians to describe the properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those   limits

1 / (1 + e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function) and value is the actual numerical value that you want to   transform.

v) Evaluating the model using Logistic Regression Technique

IV. RESULT

```
warning: called from
    csvread at line 39 column 5
    logisticKidneyMain at line 4 column 6
Chronic Disease Probability 19.0%
Accuracy of Model : 99.0%
>>
```

Fig2:Output Screen

On the basis of above output screen we can say that the accuracy of model comes out to be 99% and the chance of chronic kidney disease for this particular patient is only 19%. Hence we can use this model for early detection and prevention of chronic kidney disease before it gets worse.

V.CONCLUSION

In this paper we have studied Logistic Regression machine learning algorithms. We have analyzed 24 different attributes related to CKD patients and predicted accuracy for Logistic Regression machine learning algorithms. From the results analysis, it is observed that the Logistic Regression algorithms gives the accuracy of 99% .The advantage of this system is that, it will help in early prediction of the disease. It will help the doctors to start the treatments early for the CKD patients. Limitations of this study are the strength of the data is not higher because of the size of the data

set and the missing attribute values. To improve the machine learning model targeting chronic kidney disease we will need a large number of records with zero missing values. Also we can use other classification algorithms such as Support Vector Machines ,Decision Trees, Neural Networks, Random Forest etc.

VI. REFERENCES

[1] Abhinandan Dubey(2015), A Classification of CKD Cases Using MultiVariate K-Means Clustering. International Journal of Scientific and Research Publications, 5(8), (pp.1-5).

[2] Agarwal, S.K and Srivastava,(2009), Chronic kidney disease in India: challenges and solutions. Nephron Clin Pract,111(3), (pp.197-203).

[3] Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq(2016),Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease.International Journal of Database Management Systems,8(3),.

[4] Firman G(2009), Definition and Stages of Chronic Kidney Disease (CKD),(pp.2-5).

[5] Kirubha, V, and S. Manju Priya(2016).Survey on Data Mining Algorithms in Disease Prediction International Journal of Computer Trends and Technology, 38(3), (pp.124-128)

[6] W.G. Couser, G. Remuzzi, S. Mendis, M. Tonelli(2011), The contribution of chronic kidney disease to the global burden of major noncommunicable diseases Kidney Int, 80 (pp. 1258-1270).

[7] A.S. Levey, J. Coresh, E. Balk, A.T. Kausz, A. Levin, M.W. Steffes, et al.(2003) National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification Ann Intern Med, 139 (2),(pp. 137-147).