International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 402-404 www.ijaem.net ISSN: 2395-5252

Machine Learning Algorithms in CKD Prediction-A Review

Snehal Jejurkar,Komal Jadhav,Pragati Thorat,Gauri Kanade,Varsharani Jape

Submitted: 25-05-2021 Revised: 01-06-2021 Accepted: 05-06-2021

ABSTRACT: Chronic kidney disease(CKD) can cause various difficulties for kidney function.It may cause kidney failure and lead to death of the person. It needs to be cured in early stages to avoid further complications. Prediction of Future Possibilities of CKD can be done effectively by using Machine Learning predictive models. The process of the whole workflow includes data collection,data processing and attribute handling. This paper discusses various classification algorithms and describes the performance of each algorithm. The main contribution of these works lies in comparative study of various classifiers.This Study Shall help to find the best model on the basis of accuracy.

Keywords: CKD, Machine Learning, Classifiers, Prediction.

I. INTRODUCTION

The average size of a human kidney ranges from 10 to 13 cm. They lie below the rib cage, one on each side of the spine. 120 to 150 quarts of blood get filtered every day to produce about 1 to 2 quarts of urine. The main function of the kidneys is to remove the waste material and excess fluid from the body through the urine. The production of urine involves highly complex steps of excretion and reabsorption. Exertion through the kidneys maintains the balance of body chemicals. If kidneys lose their ability to filter waste material it results in kidney disease. CKD progression can be considered as a function of various parameters including underlying renal diseases, blood pressure, hypertension, proteinuria, and age. Early diagnosis of CKD requires great attention among physicians, especially in determining the appropriate time to apply medical treatments and to control identified risk factors that reflect on the disease progression to End-Stage Renal Disease. Chronic kidney disease (CKD) is becoming a major issue worldwide, with nearly 14% of the world population suffering from CKD. There are five stages of kidney functionality. The function is normal in stage 1 and minimally reduced in stage 2 but the majority of cases are at stage 3. The application of machine learning methods to predict CKD has been explored based on multiple data sets. Datasets are collected from the UCI repository. This work also highlights the importance of statistical analysis as well as the domain knowledge of the features when making a prediction based on clinical data related to CKD. The paper “Chronic kidney disease: global dimension and perspectives” discusses the burden, risk factors, and causes of chronic kidney disease in relation to levels of socio-economic development and healthcare systems.

II. LITERATURE REVIEW

Ahmed J. Aljaaf et al., proposed a system for early prediction of chronic kidney disease using machine learning algorithms. The proposed methodology was supported by the use of predictive analytics. Predictive analytics helped the authors to select the optimum subset of parameters to feed to the machine learning. In this study a dataset of 24 parameters was taken and 30% of them as an ideal sub set to predict Chronic Kidney Disease. In this work a total of 4 machine learning based classifiers were simulated. Classification and regression models(RPART) are showing good result and here two blackbox models used for early prediction of CKD, i.e. SVM and MLP models. Results show that the highest AUC and TPR have been achieved by MLP model, while the highest TNR of 1.00 has been obtained by RPART model.[1]

DOI: 10.35629/5252-0306402404

I. U. Ekanayake and D. HerathThis paper uses Random forest and ANN algorithms to predict early disclosure of chronic kidney diseases. Their work shows that the accuracy of the Random forest algorithm is 97.12% and ANN is 94.5%. The main objective is to predict early disclosure of chronic kidney diseases.[2]

El-Houssainy A. Radya proposed a system for Prediction of kidney disease using data mining algorithms. In this work Data Mining Algorithms has been implemented. They are Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) algorithms. The main objective is to find the best algorithm based on its performance of prediction. It shows that the Probabilistic Neural Networks algorithm gives the highest overall classification accuracy percentage of 96.7%.[3]

Vivekanand Jha, Guillermo Garcia-Garcia published a paper. In this paper discusses CKD, its causes, preventions and global perspectives.[4]

S. Vijayarani, S. Dhayanand et al. proposed a system, Data mining classification algorithms for kidney disease prediction. This work discusses scope of data mining in the field of medical industry. various data mining techniques are available for the prediction of disease.This techniques are clustering,classification, association rule and many more.Main aim is to find best algorithm among various data mining algorithms based on classification accuracy and execution time performance.[5]

Joseph A. Vassalotti, et al. published a paper, A Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician. This paper discusses the Outcomes Quality Initiative by primary care clinicians guide about assessment and care of chronic kidney disease. The objective of chronic kidney disease management is to prevent disease progression and to reduce complications.[6]

J.N.Kale, et al. Proposed iot based framework for protection of fruit trees. So a similar kind of framework,using suitable sensors,can be developed for data gathering steps in the proposed methodology.[7]

III. M to confirm the randomness of missing
value**ETHODOLOGY**

Data Preprocessing:

IV. In the pre-processing step, missing values have to be handled based on their distributions to achieve reasonable accuracy. In this work, to confirm the randomness of missing values, Little's MCAR test was performed. The potential bias due to missing data depends on the mechanism causing the data to be missing. This work uses the MCAR tests, the missingness are tested using the chi-square test of MCAR for multivariate quantitative data. It tests whether there exists a significant difference between the means of different missing-value patterns.

Testing and Training Data:

After preprocessing the data , the main dataset is split into test and training dataset. The Training and Testing data are subsets of the main dataset. Training data used to train models and the test dataset is used to provide an unbiased evaluation of a final model fit on the training dataset.

Model Training:

In the proposed system by I. U. Ekanayake et al., there are a total of 11 training models included in the training. They are logistic regression, k-Nearest Neighbors (KNN) regression, SVC with a linear kernel, SVC with RBF kernel, Gaussian NB, decision tree classifier, random forest classifier, XGB classifier, extra trees classifier, an ADA boost classifier, and a classical neural network. Here the dataset was divided into three parts, 70 percent training data, 15 percent for testing, and the rest for validation randomly.

Decision tree classifier, random forest classifier, XGB classifier, extra trees classifier, ADA boost classifier, and classical neural network these algorithms outperformed in training accuracy, testing accuracy, and cross-validation accuracy. Those are the decision tree classifier, random forest classifier.

Model Evolution and Selection

In proposed system by I. U. Ekanayake et al., performance of all the mentioned algorithms are checked for all the datasets and Based on these results, the algorithms who have highest accuracy were selected. Those algorithms are decision tree classifier, random forest classifier, XGB classifier, extra trees classifier in ada boost classifier. As it is important to find which attributes have the highest impact on this prediction, so the standard deviation of the feature importance of each algorithm was calculated. This study shows that the extra trees classifier has the lowest bias towards features next to the random forest classifier. The decision tree classifier has the highest bias out of all.

Accuracies of each Algorithm:

Algorithm	**Training accuracy**	**Cross validation accuracy**	**Testing accuracy**
Decision Tree Classifier	100.00%	100.00%	100.00%
Random Forest Classifier	100.00%	100.00%	100.00%
XGB Classifier	99.28%	100.00%	100.00%
Extra Trees Classifier	100.00%	100.00%	100.00%
AdaBoost Classifier	100.00%	100.00%	100.00%
KNN	97.85%	98.33%	98.33%
Classical Neural Network	97.81%	97.50%	97.50%
SVC Linear	97.14%	96.66%	96.66%
Logistic Regression	96.07%	96.66%	95.00%
SVC RBF	94.64%	95.00%	95.99%
Gaussian NB	95.35%	95.00%	93.33%

V. CONCLUSION

Stage 3 of CKD can cause kidney failure and lead patients to go through expensive treatment. In some poor countries people can't afford expensive treatment of CKD. So, it is better to detect CKD at an early stage. To predict such possibilities, a Machine Learning based prediction system will help people. This review is taken for the design of a predictive model to choose a best predictive model among various classification algorithms.

REFERENCES

[1]. A. J. Aljaaf, D. Jumeily, H. M.Haglan, "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics", IEEE, 2018.

[2]. I. U. Ekanayake, D.Herath, "Chronic Kidney Disease Prediction Using Machine Learning Methods",IEEE,2020.

[3]. El-Houssainy A. Radya , Ayman S. Anwarb,"Prediction of kidney disease stages using data mining algorithms",ScienceDirect,2019.

[4]. Vivekanand Jha, Guillermo Garcia-Garcia, Kunitoshi Iseki, Zuo Li,"Chronic kidney disease: global dimension and perspectives",thelancet,2013.

[5]. Dr. S. Vijayarani1 , Mr.S.Dayananda,"Data mining classification Algorithm for kidney disease prediction",IJCI,2015.

[6]. Joseph A. Vassalotti, MD, Robert Centor, MD, Barbara J. Turner, MD, MSED,"A Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician ",AJM,2015.

[7]. J.N.Kale,R..Nikam,N.G.Pardeshi,"Fruit trees protection using IoT ", IJARIIE,vol-6 Issue-3,2020.