

Early Detection of Chronic Kidney Disease using Machine Learning

Introduction

In terms of the current state of society's health, chronic kidney disease (CKD) is seen as a serious concern. With routine testing, chronic renal disease can be identified. There are numerous treatments and laboratory tests that can halt the onset of disease, slow its course, and problems caused by a lower glomerular filtration rate (GFR) improved survival rates, decreased risk of cardiovascular disease, and lifestyle quality[1]. It is one of the main reasons behind excessive medical expenses. 2% to 3% of the annual medical budget in high-income countries is spent on the cost of transplantation and dialysis [2]. The majority of renal failure patients in low- and middle-income nations do not have adequate access to life-saving dialysis and kidney transplants [3]. In developing nations like China and India, the prevalence of renal failure is predicted to climb unexpectedly [4]. In 2005, there were almost 58 million fatalities globally. The World Health Organization (WHO) estimates that chronic diseases affect 35 million people. According to current estimates, CKD will affect one in five men and one in four women worldwide between the ages of 65 and 74. Clinical information, blood testing, imaging scans, and finally a biopsy are typically used to diagnose CKD. In this paper, we propose low-cost, straightforward, and non-invasive tests that are simple to carry out. We do this by employing machine learning approaches. The data was taken from a dataset of CKD patients that was taken from the UCI machine learning repository[5]. This approach aims to "down-stage" the disease to a point where it is more receptive to curative treatment by increasing the percentage of CKD that is discovered at an early stage.

Literature Review

[1] Proposed a system that uses various data mining techniques like Random Forest algorithm and Back propagation neural Network. Here they compare both of the algorithm and found that Back Propagation algorithm gives the best result as it uses the supervised learning network called feedforward neural network.

Advantages: A wrapper method used here for feature selection is ACO. ACO is a meta-heuristic optimization algorithm. Out of the 24 attributes present 12 best attributes are taken for prediction.

Disadvantages: It contains only 400 samples of two different classes. Out of 25 attributes, eleven are numeric and 13 are nominal and one is class attribute.

[2] A deep learning model for the early diagnosis of chronic disease is presented in this work. In this research, the authors looked at the Recursive Feature Elimination approach to identify which features are the most important for prediction. The most essential CKD features are packed red blood cell count, albumin, cell volume, serum creatinine, specific gravity, hemoglobin, and hypertension.

Advantages: Classification algorithms are fed with a set of features. Different metrics, including classification accuracy, recall, precision, and f-measure, are used to estimate the comparative analysis. The proposed deep neural model outperformed the other five classifiers (Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic regression, Random Forest, and Naive Bayes classifier) by achieving 100% accuracy. The accuracy of KNN, SVM, Naive Bayes, Decision tree, Random Forest, logistic regression is 92%, 92%, 95%, 97%, and 99%, respectively.

Disadvantages: As object values cannot be used for the analysis. So, we have to convert the numeric values with type as object to float64 type. It will take extra time.

[3] Kunwar, etal. entitled “Chronic Kidney Disease Analysis is Using Data Mining Classification Techniques” published in 2016. Data mining is the process of extracting hidden information from massive dataset, categorizing valid and unique patterns in data. There are many data mining techniques like clustering, classification, association analysis, regression.

Advantages: It is automation for CKD disease prediction and it identifies the disease, its stages in an efficient and economically manner. It is successfully accomplished by applying the KNN and Naive Bayes algorithms for classification.

Disadvantages: There are 158 completed cases in the data collection, with the remainder missing. Ignoring records is the simplest technique to deal with missing values; however, this is not practical for small data sets. The data set is examined during the data preparation process to see whether any attribute values are missing.

[4] The kidneys are positioned in the abdominal cavity, on each side of the spine. They generally weigh around 5 times their body weight yet receives only 20% of the blood flow from the heart. The urine generated by each kidney drains into the urinary bladder, which is positioned in the pelvic area, via a distinct urethra.

Advantages: The dataset contains input parameters obtained from CKD patients, and the models are trained and validated using the valid parameters. It makes more efficient to find the disease.

Disadvantages: Outliers are extreme values located far away from the feature central tendency. Invalid outliers occur due to data entry errors, which are referred to as a noise in the data.

[5] Today’s real-world datasets are susceptible to missing, noisy, redundant, and inconsistent data, especially clinical datasets. Working with low-quality data leads to low-quality results. Therefore, the first step in every machine learning application is to explore the dataset and understand its characteristics in order to make it ready for the modeling stage.

Advantages : This aim by applying four machine learning classifiers: logistic regression, SVM, random forest, and gradient boosting on a small dataset of 400 records.

Disadvantages: Drawback of decision tree is that it suffers from two major problems overfitting and it is based on greedy method. overfitting happened due to decision tree split dataset aligned to axis it means it need a lot of nodes to split data.

[6] The proposed work deals with classification of different stages of CKD according to its gravity. By analysing different algorithms like Basic Propagation Neural Network, RBF and RF. The analysis results indicates that RBF algorithm gives better results than the other classifiers and produces 85.3% accuracy.

Advantages : The prediction process is less time consuming. It will help the doctors to start the treatments early for the CKD patients and also it will help to diagnose more patients within a less time period. 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.

Disadvantages: The first was Inverse Document Frequency (IDF), which grants higher impact to less frequent codes than frequent ones.

[7] As ESRD demands kidney dialysis and involves severe comorbidities, accurate prediction of patients who are likely to deteriorate to ESRD at high likelihood mortality is critical. A variety of methods have been proposed to predict ESRD.

Advantages : Health maintenance organization and hospitals, when a patient is approaching the threshold ESRD risk, a warning message can be sent electronically to the physician, to initiate a referral to for a nephrology consultation.

Disadvantages: presented performance comparison of Artificial Neural Networks, Decision Tree and Logical Regression are used for Kidney dialysis survivability. The data mining techniques were evaluated based on the accuracy measures such as classification accuracy, sensitivity and specificity. They achieved results using 10 fold cross-validations and confusion matrix for each technique.

References:

[1] Guneet Kaur, “Predict Chronic Kidney Disease using Data Mining in Hadoop, International Conference on Inventive Computing and Informatics, 2017.

[2] Vanholder, R.; Annemans, L.; Brown, E.; Gansevoort, R.; Gout-Zwart, J.J.; Lameire, N.; Morton, R.L.; Oberbauer, R.; Postma, M.J.; Tonelli, M.; et al. Reducing the costs of chronic kidney disease while delivering quality health care: A call to action. *Nat. Rev. Nephrol.* 2017, 13, 393–409. [CrossRef] [PubMed]

[3] Bidri Deepika*, Vasudeva Rao KR, Dharmaj N Rampure, Prajwal P and Devanand Gowda G, Early Prediction of Chronic Kidney Disease by using Machine Learning Techniques Vol. 8 No.2:7 (2020).

[4] A.Vaishnovi Anuhya¹, Ayyala Ganesh², Nallabathini Poojitha³, Amandeep Singh⁴, Amitha S K⁵, Dr. DHANANJAYA. Early Detection of Chronic Kidney Disease Using Advanced Machine Learning Models, 5 | May 2022.

[5] Marwa Almasoud¹, Tomas E Ward² Information System Department College of Computer and Information Science Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia¹

Insight Center for Data Analytics, Dublin City University, Dublin, Ireland. Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors,(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 8, 2019.

[6]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.

[7] Zvi Segal¹, Dan Kalifa¹, Kira Radinsky¹, Bar Ehrenberg¹, Guy Elad¹, Gal Maor¹, Maor Lewis¹, Muhammad Tibi¹, Liat Korn² and Gideon Koren², Machine learning algorithm for early detection of end-stage renal disease <https://doi.org/10.1186/s12882-020-02093-0> (2020).