M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 & ISO 14001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NALAIYA THIRAN PROJECT

TEAM ID : PNT2022TMID15455

TITLE OF THE PROJECT : Early Detection of Chronic Kidney Disease using Machine Learning

DOMAIN : Applied Data Science

SUBJECT CODE : 18CSE040L

SUBJECT TITLE : Professional Readiness for Innovation, Employability and Entrepreneurship

INDUSTRY MENTOR(S) NAME : Lalitha Gayathri

FACULTY MENTOR(S) NAME : K. Makanyadevi

PROJECT MEMBERS :

REGISTER NO	NAME	ROLE
927619BCS4057	KAVIYA N	LEADER
927619BCS4003	ABIRAMI V	MEMBER
927619BCS4014	BHAVA DHARRANI G	MEMBER
927619BCS4092	PRADEEP M M	MEMBER

Early Detection of Chronic Kidney Disease using Machine Learning

ABSTRACT

Early detection and cure of Chronic Kidney Disease (CKD) is extremely desirable as it can lead to the prevention of unwanted consequences. Machine learning methods are being extensively advocated for early detection of symptoms and diagnosis of several diseases recently. With the same motivation, the aim of this study is to predict the various stages of CKD using machine learning classification algorithms on the dataset obtained from the medical records of affected people. Specifically, we have used the Random Forest and J48 algorithms to obtain a sustainable and practicable model to detect various stages of CKD with comprehensive medical accuracy. CKD, i.e., gradual decrease in the renal function spanning over a duration of several months to years without any major symptoms, is a life-threatening disease. It progresses in six stages according to the severity level. It is categorized into various stages based on the Glomerular Filtration Rate (GFR), which in turn utilizes several attributes, like age, sex, race and Serum Creatinine. Among multiple available models for estimating GFR value, Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), which is a linear model, has been found to be quite efficient because it allows detecting all CKD stages. CKD is often diagnosed in later stages when dialysis or kidney transplant are the only options left to save the patient's life. Whereas an early diagnosis can lead to the prevention of kidney failure. The best way to measure the kidney function or to predict the stages of kidney disease is to monitor the Glomerular Filtration Rate (GFR) on regular basis.

LITERATURE REVIEW

[1] Chronic Kidney Disease Prediction Using Machine Learning Techniques.

This paper deals with the prediction of CKD in people. 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. Prediction is done using the machine learning technique, SVM. In this classification problem SVM classifies the output into two class with CKD and without CKD. The main objective of this study was to predict patients with CKD using less number attributes while maintaining a higher accuracy. Here we obtain an accuracy of about 96 percentage.

[2] Comparative Analysis for Prediction of Kidney Diseases using ML models.

The proposed system utilizes the CKD prediction dataset. After pre-processing and feature selection, the DT, KNN, and logistic regression algorithms have been used. J. Hussain and the team have obtained an accuracy of 0.995 in predicting CKD in early stages using multilayer perception while including pre-processing of data set with neural networks to fill the missing values. The workflow includes discarding the outliers, selecting the optimal seven attributes with statistical analysis, and discarding the attributes which have a higher inter co-relation by principal component analysis (PCA).

[3] Prior Stage Kidney Disease Prediction Using AI & Supervised Machine Learning Techniques.

In this a system is proposed that uses various data mining techniques like KNN, DT, NB, and SB classifiers. The metrics provided below gives us information on the quality of the outcomes that we get in this study. A confusion matrix helps us with this by describing the performance of the classifier. This paper deals with the prediction of CKD in people. A wrapper method used here. SBC is a meta-heuristic algorithm. Out of the 24 attributes present 12 best attributes are taken for prediction. Prediction is done using the machine learning technique, SBC. In this classification problem SBC classifies the output into two class with CKD and without CKD.

[4] Comparison and development of machine learning tools in the prediction of chronic kidney disease progression.

The clinical and blood biochemical results from 551 patients with proteinuria were collected. Thirteen blood-derived tests and 5 demographic features were used as non-urinary clinical variables to predict the 24-h urinary protein outcome response. Nine predictive models were established and compared, including logistic regression, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XG Boost, neural network and k-nearest neighbour. The AU-ROC, sensitivity (recall), specificity, accuracy, log-loss and precision of each of the models were evaluated. The effect sizes of each variable were analysed and ranked. The linear models including Elastic Net, lasso regression, ridge regression and logistic regression showed the highest overall predictive power, with an average AUC and a precision above 0.87 and 0.8, respectively. Logistic regression ranked first, reaching an AUC of 0.873, with a sensitivity and specificity of 0.83 and 0.82, respectively.

[5] Prediction of Chronic Kidney Disease Using Machine Learning Algorithm.

The proposed system makes a comparison of the performance of the different classification methods and ensemble algorithms that are used for detection of chronic kidney disease. This study involves six different basic classifiers namely: k nearest neighbour (KNN), naive bayes, support vector machines (SVM), random trees, J48, decision tables and three different ensemble algorithms namely: bagging, AdaBoost, random subspace are used in the study. Classification results were derived using three different performance evaluation criteria (kappa, accuracy and the area under the ROC curve (AUC)). The result says that J48 basis algorithm for use with random subspace and bagging ensemble algorithms and random tree basis algorithm for use with bagging ensemble algorithm has provided 100% classification success. Data mining is vastly being investigated in diagnostic results. Huge amount of un-mined data derived from healthcare industry is used to discover sensible information in order to diagnose and help in decision making. Data mining helps in extracting hidden information from huge dataset. It also helps in categorizing data, validate them and derive unique patterns in them. Several datamining techniques like classification, clustering, regression, association analysis, etc are used in healthcare data mining. The objective of this paper is for the prediction of Chronic Kidney Disease (CKD)through classification techniques like Artificial Neural Network (ANN) and Naive Bayes.

TABLE OF ARTICLES

S.No	YEAR	AUTHOR NAME	TITLE	ALGORITHM	DRAWBACKS
01	2022	Saurabh P al	Chronic Kidney Disease Prediction Using Machine Learning Methods	Decision Tree, GFR, SVM, Machine Learning	It has been a challenging task to find out the correct set of attributes.
02	2021	Vineeta Gulati and Neeraj Raheja	Comparative Analysis for prediction of kidney diseases using Intelligent Machine Learning Models	K-Nearest Neighbors, Logistic Regression, Decision Tree method Random Forest and, Naïve Bayes, Support Vector Machine and Multi-Layer Perceptron Algorithm	Sometimes setting a biasing values and activation function may consume more time for processing the inputs.
03	2021	Barot mitisha1, prof. Barkha bhavsa	Prior Stage Kidney Disease Prediction Using AI & Supervised Machine Learning Techniques	KNN, DT, NB, and SB classifiers	In this, there lies a problem in the selection of best suitable attributes.
04	2019	Jing Xiao and Ruifeng Ding et.al	Comparison and development of machine learning tools in the prediction of chronic kidney disease progression.	Logistic regression, k-Nearest Neighbors (KNN) regression, SVC, Gaussian NB, decision tree classifier, Random Forest classifier.	This requires the huge amount of dataset collection. Those data should be pre processed by an efficient method.
05	2018	Siddheshw ar Tekale	Prediction of Chronic Kidney Disease Using Machine Learning Algorithm	Logistic regression, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XGBoost, neural network and k-nearest neighbor.	It has the problem of conducting the performance analysis criteria as there is need for specifying a optimum criteria value.

REFERENCES

1. Saurabh Pal - Chronic Kidney Disease Prediction Using Machine Learning Methods. Issue:16,August 2022
2. Vineeta Gulati and Neeraj Raheja - Comparative Analysis for prediction of kidney Diseases using Intelligent Machine Learning Models.Issue:2021
3. Barot mitishal - Prior Stage Kidney Disease Prediction Using AI & Supervised Machine Learning.Issue:12,Dec 2021
4. Jing Xiao and Ruifeng Ding et.al - Comparison and development of machine learning tools in the prediction of chronic kidney disease progression.Issue:2019
5. Siddheshwar Tekale,Pranjal Shingavi et.al - Prediction of Chronic Kidney Disease Using Machine Learning Algorithm.Issue:10.Oct 2018