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

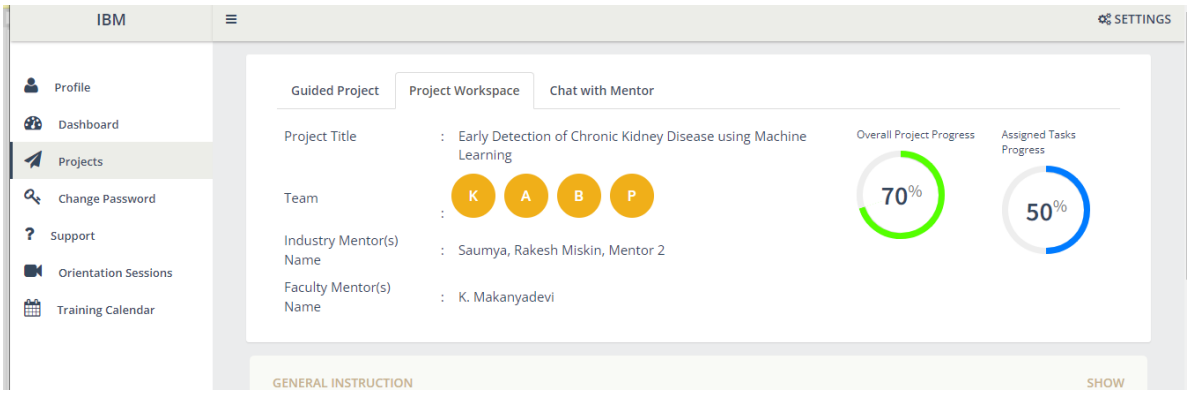
IV YEAR / VII SEMESTER (ODD)

BATCH: 2019-2023

ACADEMIC YEAR 2022-2023

NAALAIYA THIRAN

TEAM ID	:	PNT2022TMID15455
TITLE OF THE PROJECT	:	Early Detection of Chronic Kidney Disease using Machine Learning
DOMAIN	:	Applied Data Science
TEAM MEMBERS	:	Abirami V Bhava Dharani G Kaviya N Pradeep M M
INDUSTRY MENTOR(S) NAME	:	Saumya, Rakesh Miskin
FACULTY MENTOR(S) NAME	:	K. Makanyadevi



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] TITLE: Prior Stage Kidney Disease Prediction Using AI & Machine Learning.

AUTHOR: Barot mitishal

DESCRIPTION: 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.

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

AUTHOR: Jing Xiao and Ruifeng Ding

DESCRIPTION: 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.

[3] TITLE: Chronic Kidney Disease Prediction using Machine Learning.

AUTHOR: Reshma S

DESCRIPTION: 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.

[4] TITLE: Predict Chronic Kidney Disease using Data Mining in Hadoop.

AUTHOR: Guneet Kaur

DESCRIPTION: This proposed a system for predicting the CKD using Data Mining Algorithms in Hadoop. They use two data mining classifiers like KNN and SVM. Here the

predictive analysis is performed based upon the manually selected data columns. SVM classifier gives the best accuracy than KNN in this system.

[5] TITLE: Chronic Kidney Disease Prediction Using Machine Learning Methods.

AUTHOR: Imesh Udara Ekanayake

DESCRIPTION: Imesh Udara Ekanayake and Damayanthi Herath used the following methods to forecast chronic kidney disease in their paper. The features with more than 20% missing values were removed from the analysis. The KNN Imputer technique is utilized to fill in the missing values. In the model training, 11 categorization models were explored. They are KNN regression, logistic regression, SVC with a linear kernel, decision tree classifier, SVC with an RBF kernel, XGB classifier, random forest classifier, extra trees classifier, Gaussian NB, an AdaBoost classifier, and a conventional neural network. As a result of their research, six algorithms outperformed 11 algorithms in terms of training accuracy, cross-validation accuracy and testing accuracy. Random forest classifiers, Decision tree classifiers, XGB classifiers, AdaBoost classifiers, extra trees classifiers, and traditional neural network classifiers are among them. Logistic regression, k-Nearest Neighbours (KNN) regression, SVC, Gaussian NB, decision tree classifier, random forest classifier, XGB classifier, extra trees classifier, an ad a boost classifier and a classical neural network. Highest accuracy of 96% by using Decision Tree Classifier, Random Forest Classifier.

[6] TITLE: Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease.

AUTHOR: Mohamed Elhoseny

DESCRIPTION: Density-based Feature Selection (DFS) with Ant Colony-based Optimization (D-ACO). D-ACO shows the better output with 95.00 % accuracy.

[7] TITLE: Comparative Analysis for Prediction of Kidney Diseases using ML models.

AUTHOR: Gazi Mohammed Ifraz

DESCRIPTION: 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).

[8] TITLE: Prediciton of Chronic Kidney Disease using ML algorithms.

AUTHOR: Kübra Eroğlu

DESCRIPTION: 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.

[9] TITLE: Comparative Analysis of Prior Kidney Disease using Intelligent Algorithms.

AUTHOR: Naganna Chetty

DESCRIPTION: Chronic Kidney Disease (CKD) has become a common problem to the public in recent days. CKD refers to kidney damage and is measured in terms of GFR (Glomerular Filtration Rate). Researchers from healthcare as well as the academicians are working on the CKD problem and design an efficient model in order to predict and classify the CKD patient in the initial stage of CKD. This can help health care persons to take the necessary treatment to prevent or cure CKD. There are several classification models built using different classification algorithms. Some of the methods are best first search method and Wrapper subset attribute evaluator to predict and classify the CKD and non-CKD patients. These models are implemented on the CKD dataset that was downloaded from the UCI repository. These models show better performance in classifying CKD and non-CKD cases.

[10] TITLE: Prediction of Chronic Kidney Disease Using Machine Learning Algorithm.

AUTHOR: Siddheshwar Tekale

DESCRIPTION: In this paper they have studied different machine learning algorithms. They have analysed 14 different attributes related to CKD patients and predicted accuracy for different machine learning algorithms like Decision tree and Support Vector Machine. From the results analysis, it is observed that the decision tree algorithms give the accuracy of 91.75% and SVM gives accuracy of 96.75%. When considering the decision tree algorithm, it builds the tree based on the entire dataset by using all the features of the dataset. The advantage of this system is that, 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. To build a machine learning model targeting chronic kidney disease with overall accuracy of 99.99%, will need millions of records with zero missing values.

S.No	Year	Researcher	Title	Algorithm	Remarks
01	2021	Barot mitisha1, prof. Barkha bhavsa	Prior Stage Kidney Disease Prediction Using AI & Supervised Machine Learning Techniques	KNN, DT, NB, and SB classifiers	Highest accuracy of 99% by using SBC Classifier
02	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, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XGBoost, neural network and k-nearest neighbor	The model with the highest sensitivity was Elastic Net (85%)
03	2021	Gazi Mohammed Ifraz et.al	Comparative Analysis for prediction of kidney diseases using Intelligent Machine Learning Models	K-Nearest Neighbor, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine and Multi-Layer Perceptron Algorithm	Individual F1 scores are 95% for non -CKD and 97% for CKD
04	2020	Imesh Udara Ekanayake	Chronic Kidney Disease Prediction Using Machine Learning Methods	logistic regression, k-Nearest Neighbors (KNN) regression, SVC,Gaussian NB, decision tree classifier, random forest classifier, XGB classifier, extra trees classifier, an ada boost classifier and a classical neural network.	Highest accuracy of 96% by using Decision Tree Classifier, Random Forest Classifier
05	2018	Siddheshwar Tekale	Prediction of Chronic Kidney Disease Using Machine Learning Algorithm	Decision Tree, GFR, SVM, Machine Learning	Highest accuracy of 96.75% by using SVM
06	2017	Reshma S, et.al.,	Chronic Kidney Disease Prediction using Machine Learning	ACO, Support Vector Machine (SVM)	Highest accuracy of 96% by using SVM

REFERENCES

1. Barot mitishal(2021) - Prior Stage Kidney Disease Prediction Using AI & Supervised Machine Learning.2
2. Gazi Mohammed Ifraz et.al(2021) - Comparative Analysis for prediction of kidney Diseases using Intelligent Machine Learning Models.
3. Mohamed Elhoseny,Ket.al., Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease.
4. Jing Xiao and Ruifeng Ding et.al., (2019) - Comparison and development of machine learning tools in the prediction of chronic kidney disease progression.
5. Imesh Udara Ekanayake et.al - Chronic Kidney Disease Prediction Using Machine Learning Methods.