DATE	13 OCTOBER 2022
TEAM ID	PNT2022TMID08848
PROJECT NAME	Early Detection of Chronic Kidney Disease using Machine Learning
MAXIMUM MARKS	4 MARKS

LITERATURE SURVEY

EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

[1] Author Name: Tomas E

The small dataset of 400 records have been collected from Apollo Hospital, India in 2015 taken over a two-month period. ANOVA test, the Pearson's correlation, and the Cramer's V test are applied and removed the redundant features in dataset. By using filter feature selection method, Three features (hemoglobin, albumin, and specific gravity) are selected and trained using Logistic regression, support vector machines, random forest, and gradient boosting algorithm and reached an accuracy of 99.1% accuracy. Data used in this research is small. So, in future, need to validate the results by using big dataset and for reducing the prevalence of CKD, planned to predict if a person with CKD risk factors such as diabetes, hypertension, and family history of kidney failure will have CKD in the future or not by using appropriate dataset.

[2] Author Name: Ebrahime Mohammed Senan, Nizar Alsharif

The dataset was collected from 400 patients containing 24 features. The dataset was divided into 75% training and 25% testing and validation. The dataset was processed to remove outliers and replace missing numerical and nominal values using mean and mode statistical measures respectively. The RFE algorithm was applied to select the most strongly representative features of CKD. Then Random forest algorithm is applied. Data used in this research is small. So, in future, need to validate the results by using big dataset.

[3] Author Name: Bidri Deepik, Vasudeva Rao KR, Dharmaj N Rampure, Prajwal P and Devanand Gowda G

Early Prediction of Chronic Kidney Disease by using Naive bayes, K-Nearest neighbor. KNN algorithm takes CKD parameters as input and predicts the disease based on old CKD patient's data. To develop user interface for create easier communication between doctors and patients.

[4] Author Name: Sreeji S, Balamurugan Balusamy

In Design System For Early Detection And Prediction Of Chronic Kidney Disease Using Machine Learning Techniques, The SVM, and Naive Bayes classification algorithms were applied on the processed data.The information assortment contains 400 patient records, for certain qualities missing. It comprises of 24 clinical highlights that show up in the anticipation of ongoing kidney illness, with one class quality demonstrating the event of persistent renal disappointment in the patient. Data used in this research is small. In future by using better algorithms the accuracy can be increased and also need to validate the results by using big dataset.

[5] Author Name: Reshma S , Salma Shaji , S R Ajina, Vishnu Priya S R , Janisha A

In this research, The dataset contains 400 samples of two different classes. Out of 25 attributes, 11 are numeric and 13 are nominal and one is class attribute. The data set contains number of missing values. Here the information of dataset uses the patient's data like age, blood pressure, specific gravity, albumin, sugar, red blood cells etc. The sample Data used in this research is small. So, in future by using by using better deep learning concepts the accuracy can be increased.Random forest algorithm outperformed all other applied algorithms

[6] Author Name: Hamida Ilyas, Sajid Ali, Mahvish Ponum

In this study for predicting the various stages of CKD using machine learning classification algorithms on the dataset obtained from the medical records of affected people. Specifically, they 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. Comparative analysis of the results revealed that J48 predicted CKD in all stages better than random forest with an accuracy of 85.5%. The study also showed that J48 shows improved performance over Random Forest

[7] Author Name: Zixian Wang, Jae Won Chung

In this research, CKD is detected using the Apriori association technique for 400 instances of chronic kidney patients with 10-fold-cross-validation testing, and the results are compared across a number of classification algorithms including ZeroR, OneR, naive Bayes, J48, and IBk (k-nearest-neighbor). The dataset is preprocessed by completing and normalizing missing data. The most relevant features are selected from the dataset for improved accuracy and reduced training time. The results for selected features of the dataset indicate 99% detection accuracy for CKD based on Apriori. The identified technique is further tested using four patient data samples to predict their CKD. Analysed with different supervised and unsupervised machine learning

techniques and feature selection techniques with additional performance metrics need to be used for better CKD prediction.

[8] Author Name: Suman Bala, Krishan Kumar

In the health care industry the data mining is mainly used for predicting the diseases from the datasets. The Data mining classification techniques, namely Decision trees, ANN, Naive Bayes are analyzed on Kidney disease data set. Decision Trees, ANN and Naïve Bayes, Logistic Regression, Genetic Algorithms are applied on processed datasets and detects the kidney disease.

[9] Author Name: Vishal Dineshkumar Soni

This research first propose an efficient hybrid dimensionality reduction method consisting of ReliefF and PCA method. The key aspect is the selection of an appropriate threshold for the elimination of irrelevant and redundant features from the dataset. The presented work is suitable for both text and micro-array datasets that shows remarkable results with different chronic disease datasets. A novel adaptive classification system using Support Vector Machine (SVM) Classifier is proposed for the diagnosis of chronic diseases. The generalization performance of SVM classifier highly depends on the appropriate setting of its hyperparameters.

[10] Author Name: Junaid Rashid, Saba Batool

In this paper, we propose a novel augmented artificial intelligence using an artificial neural network (ANN) with particle swarm optimization (PSO) to predict five prevalent chronic disease including breast cancer, diabetes, heart attack and kidney disease. Seven classification algorithms are compared to evaluate the proposed model's prediction performance. Our proposed approach gave the highest accuracy of 99.67%. A limitation of this study is that all classification methods are applied to selected diseases, while a dataset with different patient may lead to suboptimal predictions.