DEPARTMENT OF INFORMATION TECHNOLOGY

LITERATURE REVIEW OF NALAYATHIRAN PROJECT

TITLE   :   EARLY DETECTION OF CHRONIC KIDNEY USING MACHINE LEARNING

DOMAIN   :   APPLIED DATA SCIENCE

FACULTY MENTOR   :   Mr.Rajkumar/Assistant Professor/IT

INDUSTRY MENTOR   :   Prof.LALITHA GAYATHRI

PROJECT MEMBERS   :

REGISTER NO	NAME	ROLE
1919106118	YESVANTINI S V	LEADER
1919106085	SANJAY S T	MEMBER
1919106102	SRI CHANDRIKA S	MEMBER
1919106097	SNEHA S	MEMBER
1919106116	YAKESH G	MEMBER

LITERATURE SURVEY

EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

[1] Author Name: Delpino, Felipe & Costa, Â.K. & Farias, S.R. & Filho, A.D.P. & Arcêncio, Ricardo & Nunes, Bruno. (2022)

Machine learning for predicting chronic diseases: a systematic review. Public Health. 205. 1425.10.1016/j.puhe.2022.01.007. Objectives We aimed to review the literature regarding the use of machine learning to predict chronic diseases.Study design-This was a systematic review.Methods- The searches included five databases. We included studies that evaluated the prediction of chronic diseases using machine learning models and reported the area under the receiver operating characteristic curve values. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis scale was used to assess the quality of studies.Results- In total, 42 studies were selected. The best reported area under the receiver operating characteristic curve value was 1, whereas the worst was 0.74. K-nearest neighbors, Naive Bayes, deep neural networks, and random forest were the machine learning models most frequently used for achieving the best performance. Conclusion- We found that machine learning can predict the occurrence of individual chronic diseases, progression, and their determinants and in many contexts. The findings are original and relevant to improve clinical decisions and the organization of health care facilities.

[2]Author Name: J. Siddegowda & Devi, A. (2022)

A Literature Review on Prediction of Chronic Diseases using Machine Learning Techniques. International Journal of Management, Technology, and Social Sciences. 28-49. 10.47992/IJMTS.2581.6012.0209. Background/ Purpose: Reviewing of the various work and literature in the proposed areas will help in developing a strong foundation of the domain on which the research is planned. The reason forth for the literature review is to become familiar in the health care domain. Since the area selected is the health care domain, the recent literature review is carried out as it is very important. Objective: A strong background on health care domain is developed and a new problem which is not addressed is discussed. The gaps in the research area are identified. A new solution for solving the problem is designed and developed. Design/Methodology: This work has adapted secondary source of data which is mainly journals, articles and review comments. The relevant literature is selected and a detailed study is conducted. This has helped in drafting the problem statement. Findings/Results: The finding and drawbacks of all the recent work are well studied. The reason for the gap is also well studied and the results of each work are also well analyzed. Research Limitations: A detailed study done on the chronic diseases and its impact has helped to open up the importance of studying about comorbid diseases. The limitations of various machine learning algorithms are also studied. Originality/Value: This paper aims at studying the relevant existing literature that includes research journals, conference papers, technical book chapter and few web sources. All the papers selected were relevant to the proposed work and all papers are recent and from well reputed publisher. The papers are cited by many authors. Paper Type: Literature review paper is carried out on scientific papers, especially from well indexed services.

[3]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.

[4]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.

[5]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.

[6]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.

[7]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

[8]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

[9]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.

[10]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.