

EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

Team ID: PNT2022TMID35275

Team Members:

Jeyasri Meenakshi K(2019103530)

Subashini M(2019013530)

Dhamodaran S(2019103515)

Mohammad Farook A(2019103543)

Ishwarya Rani M(2019103527)

LITERATURE SURVEY

1. Qin *et al.* proposed data assertion and sample diagnosis achievable in CKD diagnosis. KNN is used for data assertion. Six classifiers algorithms used for accuracy of diagnosis: logistic regression, random forest, support vector machine, K-nearest neighbor, naive Bayes classifier and feed-forward neural network. In these classifiers random forest gives better accuracy, i.e., 99.75%.
2. Chen *et al.* applied three models on the dataset that is provided by UCI. They used KNN, SVM and soft independent modelling of class analogy (SIMCA) for finding the risk calculation of patient using these classifiers. In which the SVM and KNN model attained, the best accuracy of 99.7% and SVM model has the greatest capability to endure noise disturbance.
3. Padmanaban and Parthiban suggested that the early detection of CKD for diabetic patients with the help of machine learning classifiers algorithms. They collected data from Chennai based diabetes research center and applied Naive Bayes and Decision tree on the dataset. For finding the accuracy they used Weka tool and concluded that Naïve Bayes classifier achieved the highest accuracy of 91%.
4. de Almeida *et al.* in their work applied Decision tree, Random Forest, Support Vector Machine (SVM) and also used SVM with linear, polynomial, sigmoid and RBF functions. For their research, they used the MIMIC-II database. They concluded that random forest and Decision tree got the best result in the form of prediction accuracy of 80% and 87% respectively.

5. Gunarathne *et al.* built a model of various machine learning classifiers algorithm and analysis of which algorithm is best suited to the dataset. They used dataset provided by UCI containing 400 instances and 14 attributes. They concluded that the Multiclass decision forest algorithm was best fitted for the CKD dataset with an accuracy of 99.1%.
6. Polat *et al.* used SVM algorithm for CKD prediction. For the accurate result, they worked on an important feature. For selecting the correct feature, they used two-approach Wrapper and filter with the SVM algorithm. In the Wrapper, there were the greedy stepwise search engine for classifier subset evaluator and best first search engine for Wrapper subset evaluator. In filter, there were the greedy stepwise search engine for correlation feature section subset feature and best first search engine for filtered subset evaluator. The results of all techniques were compared and it was found that SVM gave the highest accuracy with filtered subset evaluator, i.e. 98.5%.
7. Sujata Drall, Gurdeep Singh Drall, Sugandha Singh, Bharat Drall *et al.* worked on CKD dataset given by UCI with 400 instances and 25 attributes. Firstly, data was preprocessed, the missing data was found, filled with 0, then transformed and applied on the dataset. After preprocessing, authors applied algorithm for important attributes and found 5 most important features and then the classification algorithm: Naïve Bayes and K-Nearest Neighbor. The gotten result KNN achieved the highest accuracy.
8. Almasoud and Ward worked with CKD dataset of 400 instances and 25 attributes. They applied the filter feature selection method on attributes and found that haemoglobin, albumin and specific gravity are feature attributes in CKD dataset. After feature selection, they trained the dataset and validated with 10-fold cross-validation. The gradient boosting algorithm achieved the highest accuracy of 99.1%.
9. Shankar *et al.* applied three steps on the same UCI dataset: (i) data preprocessing & feature selection (ii), algorithms' accuracy determination and (iii) diet plan suggestion. In the feature selection method, they applied two approaches: one is the Wrapper and the other is the LASSO method. After the feature selection method, 4 classification algorithms were applied: Logistic Regression, Random forest tree K-Nearest Neighbors, Neural Network and Wide and Deep

Learning. For diet plan suggestion blood potassium level was used. The blood potassium level was divided into three groups based on its value: Safe Zone, Caution Zone and Danger zone.

10. Vijayarani and Dhayanand collected kidney function test (KFT) dataset from medical labs, research centres and hospitals. The dataset contained 584 instances and 6 attributes and two classifier applied algorithms: support vector machine (SVM) and artificial neural network (ANN). It was found that ANN achieved the highest accuracy of 87.7%.

REFERENCES:

1. https://scholar.google.com/scholar?q=A+machine+learning+methodology+for+diagnosing+chronic+kidney+disease&as_occt=title&hl=en&as_sdt=0%2C31
2. https://scholar.google.com/scholar?q=Clinical+risk+assessment+of+patients+with+chronic+kidney+disease+by+using+clinical+data+and+multivariate+models&as_occt=title&hl=en&as_sdt=0%2C31
3. https://scholar.google.com/scholar?q=Applying+machine+learning+techniques+for+predicting+the+risk+of+chronic+kidney+disease&as_occt=title&hl=en&as_sdt=0%2C31
4. https://scholar.google.com/scholar?q=Kidney+failure+detection+using+machine+learning+techniques&as_occt=title&hl=en&as_sdt=0%2C31
5. https://scholar.google.com/scholar?q=Performance+evaluation+on+machine+learning+classification+techniques+for+disease+classification+and+forecasting+through+data+analytics+for+chronic+kidney+disease+%28CKD%29&as_occt=title&hl=en&as_sdt=0%2C31
6. https://scholar.google.com/scholar?q=Huseyin+polat1+%26+homay+danaei+mehr1+%26+aydin+cetin&as_occt=title&hl=en&as_sdt=0%2C31
7. https://scholar.google.com/scholar?q=Chronic+kidney+disease+prediction+using+machine+learning%3A+A+new+approach&as_occt=title&hl=en&as_sdt=0%2C31

8. https://scholar.google.com/scholar?as_q=Detection+of+chronic+kidney+disease+using+machine+learning+algorithms+with+least+number+of+predictors&as_occt=title&hl=en&as_sdt=0%2C31
9. https://scholar.google.com/scholar?as_q=Analysis+and+prediction+of+chronic+kidney+disease&as_occt=title&hl=en&as_sdt=0%2C31
10. https://scholar.google.com/scholar?as_q=Kidney+disease+prediction+using+SVM+and+ANN+algorithms&as_occt=title&hl=en&as_sdt=0%2C31