LITERATURE SURVEY

[1] proposed an algorithm for the diagnostic procedure to identify patients with CKD requiring further nephrological care. Screening procedures included a microalbuminuria dipstick test accompanied by blood pressure measurement and medical questionnaire. In further diagnosis of CKD, estimated glomerular filtration rate (eGFR), albumin concentration in urine, urinalysis and ultrasound examination were used according to the algorithm. Multivariate logistic regression was performed to identify associations between participants' characteristics and albuminuria. Out of 9,700 invited subjects, 2,471 individuals participated in the PolNef study. Albuminuria was detected in 15.6% of the investigated population using the dipstick test and thereafter confirmed in 11.9% by the turbidimetric method. The modeling of multivariate logistic regression indicated the following independent predictors of albuminuria: male sex, diabetes, nocturia and hypertension. The proposed diagnostic algorithm seems to be a powerful tool to identify subjects at risk of CKD.

In [2], in this paper the dataset used is samples collected from Apollo Hospital present in UCI data repository. They have removed redundant features such as ANOVA test, the Pearson's correlation and Cramer's V test. They made use of 10-fold cross-validation to train and test using algorithms like support vector machines,Logistic regression, random forest, and gradient boosting. For replacing missing values, multiple imputations (MI) is used which is based on linear regression for predicting continuous variables and logistic regression for categorical variables. For feature association, Pearson's correlation, Cramer's V and ANOVA tests have been used to find relationships between variables. It is found that after applying filter feature selection it is found that haemoglobin, albumin, and specific gravity have the most impact to predict the CKD.

In [3], they have used various machine learning models to predict the possibility of chronic kidney disease CKD using the data available in UCI repository. The clinical data is preprocessed using chi-square testing of MCAR for multivariate quantitative data which identifies whether a noticeable difference is between means of missing value pattern. Along with this, the MCAR test includes finding p-value, degrees of freedom and missing patterns. After pre-processing, feature selection is carried out. Then , the model is trained using 11 classification models which are logistic regression, k-Nearest Neighbors (KNN) regression, SVC with a linear kernel, SVC with RBF kernel, Gaussian NB, decision tree classifier, random forest classifier, XGB classifier, extra trees classifier, an adaboost classifier and a classical neural network. The dataset was divided into 3 parts - 70% for training the data, 15% for cross validation of the data, 15% for testing the data. It is followed by hyperparameter tuning from a genetic algorithm and grid search for the training dataset. Based on the result, 6 algorithms outperformed in training accuracy, testing accuracy and in crossvalidation accuracy. Those are the extra trees classifier,ada boost classifier, random forest classifier,decision tree classifier, XGB classifier and classical neural network. It can be concluded that filling missing values using KNN-imputer based approach instead of a constant gives high accuracy.

[4] have performed an evaluation on a dataset of 400 patients, 250 among them have early stage of CKD. To handle missing and noisy values a classification algorithm was needed. Hence, they evaluated three classifiers: k-nearest neighbours, random forest, and neural networks to find a good solution for this application. To reduce over-fitting as well as to identify the most important predictive attributes for CKD, feature reduction is done using two methods: the wrapper method and LASSO regularization. Through their evaluation they find that, the random forest algorithm with a reduced attribute set of 12 members can detect CKD with highest accuracy of .998 and with a 0.107 root mean square error, which is a 57% RMSE reduction compared to the state of the art solutions. Through their evaluation they find hemoglobin, along with previously explored serum creatinine, and albumin are highly predictive attributes for CKD.

[5] Applies various ML algorithms in order to assess and compare the accuracy of each,along with other performance parameters for the detection of chronic kidney disease. Eight different algorithms (LR, KNN, SVM, DT, NB, MLP, RF, QDA) were used for the same dataset and analyzed. performance parameters like accuracy, precision, sensitivity, F1 score and ROC-AUC. With and without the tuning of hyperparameters for each algorithm was checked. This tuning was done using RandomizedSearchCV and accuracies were significantly higher after applying the same. Among the models, Random Forest displayed the highest accuracy of 99.75%.

[6] suggests a deep learning algorithm that analyzes retinal images in order to detect chronic kidney disease. This algorithm can be introduced in the existing screening methods used to enhance the effectiveness of detecting CKD. The retina shares a close biological relationship with the kidney, and retinal microvascular abnormalities like retinopathy and other vascular features have been shown to be associated with chronic kidney disease, even in people without diabetes suggesting that the retina could provide clues for chronic kidney disease. A conventional development, validation, and external testing study on three DLAs (retinal images only, RF only, and combined retinal and RF hybrid) using retinal images and clinical data collected from three population-based studies. The DLA takes as inputs two standardized macula-centered images (1 image per eye per participant) with resolution of 512 × 512. The output for the DLA was a binary classifier with two nodes classifying the presence or absence of chronic kidney disease status. The deep learning model is based on cCondenseNet20 with 5 blocks.

Ultimately this method was found to have potential to identify chronic kidney disease in community populations. Since access to digital retinal photography is increasing at the community and primary care level, a retinal image-based DLA has the potential to be adopted for first stage chronic kidney disease screening before confirmatory tests.

REFERENCES

1. Król E, Rutkowski B, Czarniak P, Kraszewska E, Lizakowski S, Szubert R, Czekalski S, Sułowicz W, Więcek A: Early Detection of Chronic Kidney Disease: Results of the PolNef Study. Am J Nephrol 2009;29:264-273.
2. Marwa Almasoud,, Tomas E Ward : Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors, 2019.
3. Imesh Udara Ekanayake, Damayanthi Herath : Chronic Kidney Disease Prediction Using Machine Learning Methods , 2020
   DOI: http://dx.doi.org/10.1109/MERCon50084.2020.9185249.
4. Salekin, Asif & Stankovic, John. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. 262-270.
5. Muntasir Nishat M, Faisal F, Rahman Dip R, Nasrullah SM, Ahsan R, Shikder F, Ar-Raihan Asif MA-, Hoque MA. A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms.
6. Sabanayagam, Charumathi, et al. “A Deep Learning Algorithm to Detect Chronic Kidney Disease from Retinal Photographs in Community-Based Populations.” *The Lancet Digital Health*, vol. 2, no. 6, June 2020, pp. e295–e302, 10.1016/s2589-7500(20)30063-7. Accessed 25 Mar. 2022.