

Early Detection of Chronic Kidney Disease Using Machine Learning:

Project Name	Early Detection of Chronic Kidney Disease Using Machine Learning
Team ID	PNT2022TMID22216

PROJECT OBJECTIVE :

Chronic kidney disease (CKD) is a worldwide public health problem, usually diagnosed in the late stages of the disease. To alleviate such issue, investment in early prediction is necessary. The purpose of this study is to assist the early prediction of CKD, addressing problems related to imbalanced and limited-size datasets. We used data from medical records of Brazilians with or without a diagnosis of CKD, containing the following attributes: hypertension, diabetes mellitus, creatinine, urea, albuminuria, age, gender, and glomerular filtration rate.

We present an oversampling approach based on manual and automated augmentation. We experimented with the synthetic minority oversampling technique (SMOTE), Borderline-SMOTE, and Borderline-SMOTE SVM. We implemented models based on the algorithms: decision tree (DT), random forest, and multi-class AdaBoosted DTs.

We also applied the overall local accuracy and local class accuracy methods for dynamic classifier selection; and the k-nearest oracles-union, k-nearest oracles-eliminate, and META-DES for dynamic ensemble selection. We analyzed the models' performances using the hold-out validation, multiple stratified cross-validation (CV), and nested CV. The DT model presented the highest accuracy score (98.99%) using the manual augmentation and SMOTE. Our approach can assist in designing systems for the early prediction of CKD using imbalanced and limited-size datasets