World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:16, No:11, 2022

A Machine Learning Model for Dynamic Prediction of Chronic Kidney Disease Risk Using Laboratory Data, Non-Laboratory Data, and Metabolic Indices

Authors : Amadou Wurry Jallow, Adama N. S. Bah, Karamo Bah, Shih-Ye Wang, Kuo-Chung Chu, Chien-Yeh Hsu

Abstract : Chronic kidney disease (CKD) is a major public health challenge with high prevalence, rising incidence, and serious adverse consequences. Developing effective risk prediction models is a cost-effective approach to predicting and preventing complications of chronic kidney disease (CKD). This study aimed to develop an accurate machine learning model that can dynamically identify individuals at risk of CKD using various kinds of diagnostic data, with or without laboratory data, at different follow-up points. Creatinine is a key component used to predict CKD. These models will enable affordable and effective screening for CKD even with incomplete patient data, such as the absence of creatinine testing. This retrospective cohort study included data on 19,429 adults provided by a private research institute and screening laboratory in Taiwan, gathered between 2001 and 2015. Univariate Cox proportional hazard regression analyses were performed to determine the variables with high prognostic values for predicting CKD. We then identified interacting variables and grouped them according to diagnostic data categories. Our models used three types of data gathered at three points in time: non-laboratory, laboratory, and metabolic indices data. Next, we used subgroups of variables within each category to train two machine learning models (Random Forest and XGBoost). Our machine learning models can dynamically discriminate individuals at risk for developing CKD. All the models performed well using all three kinds of data, with or without laboratory data. Using only non-laboratory-based data (such as age, sex, body mass index (BMI), and waist circumference), both models predict chronic kidney disease as accurately as models using laboratory and metabolic indices data. Our machine learning models have demonstrated the use of different categories of diagnostic data for CKD prediction, with or without laboratory data. The machine learning models are simple to use and flexible because they work even with incomplete data and can be applied in any clinical setting, including settings where laboratory data is difficult to obtain.

Keywords : chronic kidney disease, glomerular filtration rate, creatinine, novel metabolic indices, machine learning, risk prediction
Conference Title : ICRMRT 2022 : International Conference on Reproductive Medicine and Reproductive Technology
Conference Location : London, United Kingdom
Conference Dates : November 18-19, 2022