Personalised prediction of chronic kidney disease progression in patients with chronic kidney disease stages 3–5: a multicentre study using the machine learning approach

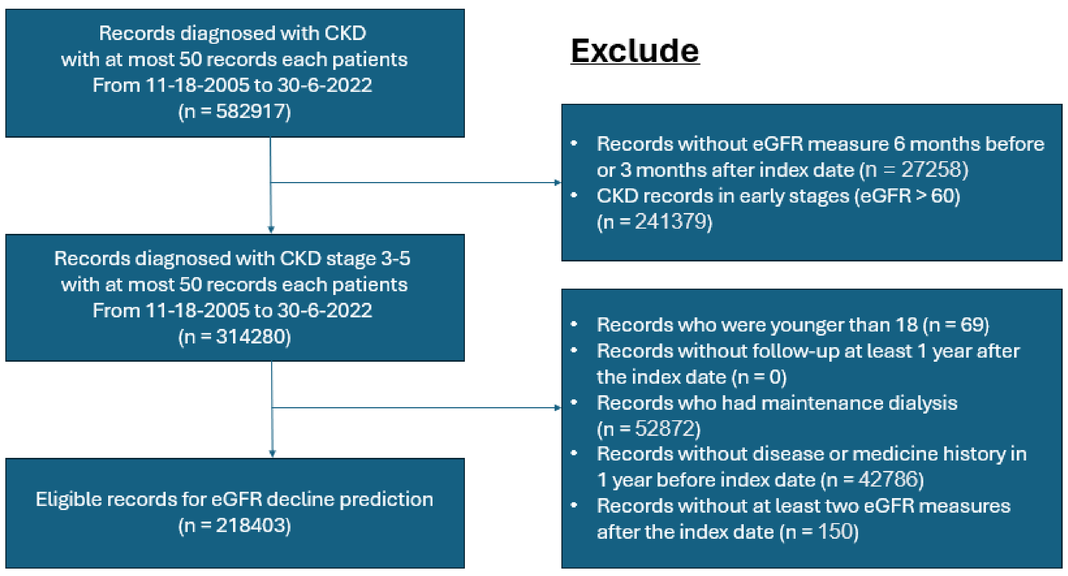
Authors: Minh Tri Nguyen 1, Trung Toan Duong 2 3, Jason C. Hsu 4 5 6 7, Alex PA. Nguyen 5 6 7 8 *
Affiliation: 1. Institute of Data Science, College of Management, Taipei Medical University, New Taipei City, Taiwan; 2. International PhD program of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; 3. Cho Ray Hospital, Ministry of Health, Ho Chi Minh City, Vietnam; 4. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan; 5. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan; 6. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan; 7. Research Center of Data Science on Healthcare Industry, College of Management, Taipei Medical University, Taipei, Taiwan; 8. Graduate Institute of Data Science, College of management, Taipei Medical University, Taipei, Taiwan; * Corresponding author.

Background

Chronic kidney disease (CKD) progression, calculated through estimated glomerular filtration rate (eGFR), is an important measurement to maintain patients' health and prevent other complications such as hypertension. This study aims to train and evaluate machine learning (ML) models for CKD progression prediction within 1-year timeframe among patients with CKD stages 3–5.

Methods

Figure 1: Enrollment process

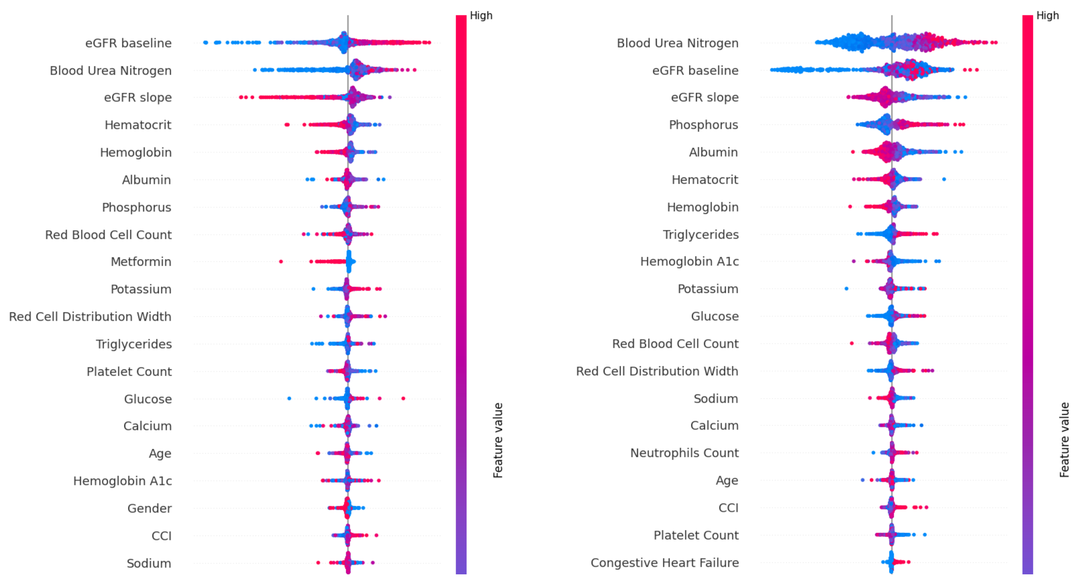


Electronic health record data from the Taipei Medical University clinical research database (TMUCRD) were used for the retrospective dataset. The TMUCRD database has been converted into the OMOP Common Data Model. Our cohort includes patients with CKD stages 3–5 between 2005 and 2021, with a maximum follow-up of 1 year. Patient demographics, comorbidities, medications, and laboratory data were used to develop models. We divide the dataset into training and testing sets and evaluate the model with 5-fold cross-validation to guarantee robust performance. Area under the curve (AUC), sensitivity, specificity, and accuracy were employed as evaluation metrics.

Results

After the enrollment process, 11488 patients were included in model training. The Light Gradient-Boosting Machine model achieved the best results in predicting 5% and 25% eGFR decline, with AUC values of 0.76 and 0.82, respectively. Based on SHAPLEY value calculations, important features that contributed to the prediction's results included baseline eGFR, eGFR slope, and blood urea nitrogen (BUN).

Table 1: Features importances in predicting 5% eGFR decline (left figure) and 25% eGFR decline (right figure)



Conclusions

This study demonstrates the effectiveness of applying an ML approach for predicting CKD progression for patients with CKD stages 3–5. These findings can be used for personalized prevention and treatment strategies and discovering patients at risk for CKD decline. We plan to expand this study into a multicenter study in the future.

Contact: tri.m.nguyen.1999@gmail.com