Short-term predictive models for post-kidney transplant diabetes mellitus using machine learning approach: preliminary data

Seoyoung Choi[1], Jieun Hong[3], Myeong Ju Kim[3], Jong Cheol Jeong[2], Sejoong Kim[2]
[1]Department of Preliminary Medicine, Seoul National University College of Medicine, Korea, Republic of
[2]Department of Internal Medicine-Nephrology, Seoul National University Bundang Hospital, Korea, Republic of
[3]Department of Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Korea, Republic of

Objectives: Post-transplant diabetes mellitus (PTDM) increases morbidity and mortality of transplant recipients. This study aims to develop a predictive model for PTDM from Korean Organ Transplant Registry cohort data and further build a platform preventing PTDM.

Methods: A prospective cohort study was conducted on KOTRY data of 6455 kidney transplant patients. The data consisted of a total of 110 variables (23 binary, 9 categorical, 78 continuous) including demographic and clinical data. Patients were classified by the occurrence of diabetes 6 months post transplantation. The total study population was divided into training set (n=5809, 90%) and test set (n=646, 10%). Machine learning was applied in finding the model with optimized accuracy. Four machine learning methods (Logistic Regression, XGBoost, CatBoost, Lightgbm) were constructed to analyze the data. Performances of algorithm were calculated by AUC (area under the receiver operating characteristic curve) score. Significance of each feature was determined using SHAP method.

Results: Out of 6455 patients, 2569 (40%) showed the incidence of PTDM within 6 months. Nine percent were patients with newly developed diabetes, and 31% were patients who had pre-transplant diabetes. Four machine learning methods (Logistic Regression, XGBoost, CatBoost, Lightgbm) developed prediction models using the training set. The 6 months prediction AUC scores in the test set were 0.92, 0.91, 0.92, and 0.91, respectively. Categorical features were especially more associated with PTDM. The significance of modifiable features was in the order of pretransplant HbA1c, fasting serum glucose, and body mass index.

Conclusions: We developed a prediction model for PTDM within 6 months after kidney transplantation based on biological cohort data of patients. We found that PTDM was most successfully predicted by using CatBoost model among 4 machine learning models. HbA1c, fasting serum glucose, and body mass index were proved to be the most significant modifiable features in predicting PTDM.