Abstract Type : Poster
Abstract Submission No. : 1361

Machine Learning Based 2-year Risk Prediction Tool in IgA Nephropathy

Jong Hyun Jhee[1], Yujeong Kim[2], Chan Min Park[2], Donghwan Oh[1], Eunji Yang[1], Beom Jin Lim[3], Hoon Young Choi[1], Dukyong Yoon[2], Hyeong Cheon Park[1]
[1]Department of Internal Medicine-Nephrology, Gangnam Severance Hospital, Korea, Republic of
[2]Department of Biomedical Systems Informatics, Yongin Severance Hospital, Korea, Republic of
[3]Department of Pathology, Gangnam Severance Hospital, Korea, Republic of

Objectives: This study aimed to develop a machine learning-based 2-year risk prediction model for the early identification of patients with rapid progressive immunoglobulin A nephropathy (IgAN). We also assessed its performance to predict the long-term prognosis of kidney outcomes.

Methods: A retrospective cohort of 1,864 patients with biopsy-proven IgAN from two tertiary hospitals was used. A random forest-based prediction model was developed to predict the composite kidney outcome and improvement of proteinuria within 2 years of follow-up after kidney biopsy and the model was externally validated for the generalizability. Prediction performances were evaluated and the correlation between the predicted outcomes and disease-free survival period was assessed.

Results: During a median follow-up of 285 [168 - 453] days for composite kidney outcome, 159 events (training set, 89.9%, test set, 10.1%) occurred. Of the improvement of proteinuria, median follow-up duration was 85 [25 - 194] days and 485 events (training set, 86.0%, test set, 14.0%) occurred. For the prediction of composite kidney outcomes, precision, recall, AUROC, and AUPRC were 0.259, 0.875, 0.771, and 0.242, respectively. The values for the improvement of proteinuria were 0.899, 0.912, 0.664, and 0.894, respectively. From Shapley Additive exPlanations (SHAP) analysis, the most informative feature identifying the composite kidney outcome were urine protein/creatinine ratio (UPCR), serum albumin, and systolic blood pressure. For the improvement of proteinuria, UPCR, urine red blood cell, and serum albumin were informative. The group with a higher probability value of the composite kidney outcome and the group with a lower probability value of the improvement of proteinuria showed significantly increased risks; hazard ratios of 3.33 and 2.17, respectively, for the long-term kidney outcome.

Conclusions: Our 2-year risk prediction models for composite kidney outcomes and improvement of proteinuria showed reliable performance and significant association with long-term kidney outcomes.

Table 1. Prediction model performance metrics for training and test set