Kidney Res Clin Pract 2024;43(6):697-699 pISSN: 2211-9132 • eISSN: 2211-9140 https://doi.org/10.23876/j.krcp.24.998



Brief review for "Machine learning-based 2-year risk prediction tool in immunoglobulin A nephropathy"

Hae-Ryong Yun¹, Tae-Hyun Yoo²

¹Department of Internal Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea ²Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

See Article on Pages 739-752

Immunoglobulin A nephropathy (IgAN) is one of the most common kidney diseases worldwide, which exhibits various clinical courses [1]. Patients can experience a range of disease severity from mild to severe and may progress to kidney failure with renal replacement therapy or kidney transplantation due to the challenges in treatment. These serious complications can significantly impair patient survival and quality of life, highlighting the importance of accurately predicting IgAN prognosis to initiate timely and appropriate treatment and management.

In this context, a machine learning model for predicting the prognosis of IgAN patients can be a highly useful tool in the medical field [2–4]. Through this model, nephrologists can identify individual patient risks and can intervene early to improve prognosis. Therefore, this study emphasizes that the development and evaluation of machine learning models for predicting IgAN prognosis could be a crucial turning point in the treatment and management of kidney diseases. Such algorithms can assist healthcare providers in properly managing patients and formulating treatment plans based on accurate prognosis prediction.

The study utilized a retrospective cohort of 1,301 patients with IgAN to derive and externally validate a machine learning-based random forest model [5]. This model was employed to predict both primary outcomes (a 30% decline in estimated glomerular filtration rate from baseline or the need for renal replacement therapy within 2 years after renal biopsy) and secondary outcomes (improvement in proteinuria). As a result, for the 2-year prediction of primary outcomes, metrics such as accuracy, recall, area under the curve, precision-recall curve, F1 score, and Brier score were found to be 0.259, 0.875, 0.771, 0.242, 0.400, and 0.309, respectively (Table 1). In contrast, results for secondary outcomes were observed to be 0.904, 0.971, 0.694, 0.903, 0.955, and 0.113, respectively. Shapley Additive exPlanations (SHAP) analysis revealed that baseline proteinuria was the most informative feature for identifying primary and secondary outcomes. Furthermore, using probabilities derived from the 2-year primary outcome prediction model to forecast 10-year renal outcomes in Kaplan-Meier analysis, the high (hazard ratio [HR], 13.00; 95% confidence interval [CI], 9.52-17.77) and moderate (HR, 12.90; 95% CI, 9.92-16.76) groups exhibited higher risks compared to the low-risk group. In the 2-year secondary outcome prediction model, the low (HR, 1.66; 95% CI, 1.42-1.95) and

Received: May 22, 2024; Accepted: May 23, 2024

Correspondence: Hae-Ryong Yun

Department of Internal Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Republic of Korea. E-mail: siberian82@yuhs.ac ORCID: https://orcid.org/0000-0002-7038-0251

^{© 2024} by The Korean Society of Nephrology

[©] This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial and No Derivatives License (http:// creativecommons.org/licenses/by-nc-nd/4.0/) which permits unrestricted non-commercial use, distribution of the material without any modifications, and reproduction in any medium, provided the original works properly cited.

Table 1. Performance metrics for the 2-year prediction model

Metric	Primary outcome	Secondary outcome
Accuracy ^a	0.259	0.904
Recall ^b	0.875	0.971
Area under the $\mbox{curve}^{\mbox{\tiny c}}$	0.771	0.694
Precision-recall curve ^d	0.242	0.903
F1 score ^e	0.400	0.955
Brier score ^f	0.309	0.113

^aThe ratio of correctly predicted instances to the total instances. ^bThe ability of the model to correctly identify positive instances. ^cMeasures the ability of the model to distinguish between classes. ^dA curve that shows the tradeoff between precision and recall for different threshold values. ^cThe harmonic mean of precision and recall, providing a balance between the two. ^fMeasures the mean squared difference between predicted probabilities and actual outcomes, with lower values indicating better accuracy.

moderate (HR, 1.42; 95% CI, 0.99–2.03) groups were found to have greater risks in 10-year prognosis compared to the high-risk group. Therefore, the machine learning-based 2-year risk prediction model for predicting IgAN progression demonstrated reliable performance and effectively predicted long-term renal outcomes.

A key strength of this paper is its significant contribution to conditions like IgAN, where accurate prognosis is challenging. The validation of the model through external validation and reproducibility studies highlights its generalizability. Additionally, the enhancement of interpretability through SHAP analysis and confirmation of the model's ability to predict long-term outcomes via Kaplan-Meier analysis are also noteworthy strengths of this research.

However, the development and evaluation of such models entail several key considerations. First, additional external validation and reproducibility studies are needed to enhance the reliability of the model's performance and generalizability. With many variables being utilized, there's a risk of the model overfitting to the training data, potentially leading to issues of overfitting. Indeed, while performance metrics within the internal dataset of this study appear promising, there's a notable decrease in performance when applied to external datasets. Therefore, reinforcing the model's generalizability using results from external datasets and ensuring consistent predictive performance across diverse environments are crucial. This would help elevate the model's reliability and maintain the consistency of predictions. Second, efforts to enhance the interpretability and transparency of the model are vital. Interpretability refers to the ability to clearly understand and explain which variables or features the model bases its predictions on. This aids healthcare professionals and patients in comprehending and trusting the model's predictions. Transparency pertains to the clear disclosure of the model's internal workings and decision-making processes. In other words, understanding how the model analyzes data and makes predictions should be made evident. High transparency models not only enhance confidence in medical decisions but also provide evidence for why certain predictions are made, supporting the decision-making process. Lastly, empirical studies are necessary to confirm whether the model can be effectively utilized in clinical settings. Evaluating whether doctors or clinical teams can easily apply and comprehend the model is paramount. Assessing the impact of the model on actual patients and evaluating its usefulness and effectiveness in real-world scenarios should follow suit. Empirical studies in clinical settings should be conducted iteratively from the model's development stages. Evaluating how the model operates in real-world settings allows for the verification of its performance and validity. This process enables the assessment of whether the model can provide tangible value in medical settings and, if necessary, allows for model supplementation and improvement. Therefore, empirical studies play a crucial role in confirming how the model can be applied in real-world settings, assessing its practical utility, and determining its clinical validity.

In conclusion, this research has made significant strides by developing and evaluating a machine learning-based predictive model for forecasting the prognosis of IgAN patients. However, further research and validation are required, and additional efforts are needed to evaluate the medical utility and clinical applicability of the model.

Conflicts of interest

Tae-Hyun Yoo is the Editor-in-Chief of *Kidney Research and Clinical Practice* and was not involved in the review process of this article. All authors have no other conflicts of interest to declare.

Data sharing statement

The data presented in this study are available from the cor-

responding author upon reasonable request.

ORCID

Hae-Ryong Yun, https://orcid.org/0000-0002-7038-0251 Tae-Hyun Yoo, https://orcid.org/0000-0002-9183-4507

References

- 1. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med* 2013;368: 2402–2414.
- 2. Schena FP, Anelli VW, Trotta J, et al. Development and testing of an artificial intelligence tool for predicting end-stage kidney dis-

ease in patients with immunoglobulin A nephropathy. *Kidney Int* 2021;99:1179–1188.

- **3.** Haaskjold YL, Lura NG, Bjørneklett R, Bostad L, Bostad LS, Knoop T. Validation of two IgA nephropathy risk-prediction tools using a cohort with a long follow-up. *Nephrol Dial Transplant* 2023;38:1183–1191.
- 4. Schena FP, Anelli VW, Abbrescia DI, Di Noia T. Prediction of chronic kidney disease and its progression by artificial intelligence algorithms. *J Nephrol* 2022;35:1953–1971.
- 5. Kim Y, Jhee JH, Park CM, et al. Machine learning-based 2-year risk prediction tool in immunoglobulin A nephropathy. *Kidney Res Clin Pract* 2024;43:739-752.