



Machine learning-based 2-year risk prediction tool in immunoglobulin A nephropathy

Yujeong Kim^{1,*}, Jong Hyun Jhee^{2,*}, Chan Min Park¹, Donghwan Oh², Beom Jin Lim³, Hoon Young Choi^{2,4},
Dukyong Yoon^{1,5,†}, Hyeong Cheon Park^{2,4,†}

¹Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

²Division of Nephrology, Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

³Department of Pathology, Yonsei University College of Medicine, Seoul, Republic of Korea

⁴Severance Institute for Vascular and Metabolic Research, Yonsei University College of Medicine, Seoul, Republic of Korea

⁵Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea

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

Methods: A retrospective cohort of 1,301 patients with biopsy-proven IgAN from two tertiary hospitals was used to derive and externally validate a random forest-based prediction model predicting primary outcome (30% decline in estimated glomerular filtration rate from baseline or end-stage kidney disease requiring renal replacement therapy) and secondary outcome (improvement of proteinuria) within 2 years after kidney biopsy.

Results: For the 2-year prediction of primary outcomes, precision, recall, area-under-the-curve, precision-recall-curve, F1, and Brier score were 0.259, 0.875, 0.771, 0.242, 0.400, and 0.309, respectively. The values for the secondary outcome were 0.904, 0.971, 0.694, 0.903, 0.955, and 0.113, respectively. From Shapley Additive exPlanations analysis, the most informative feature identifying both outcomes was baseline proteinuria. When Kaplan-Meier analysis for 10-year kidney outcome risk was performed with three groups by predicting probabilities derived from the 2-year primary outcome prediction model (low, moderate, and high), 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 showed higher risks compared with the low group. From the 2-year secondary outcome prediction model, low (HR, 1.66; 95% CI, 1.42–1.95) and moderate (HR, 1.42; 95% CI, 0.99–2.03) groups were at greater risk for 10-year prognosis than the high group.

Conclusion: Our machine learning-based 2-year risk prediction models for the progression of IgAN showed reliable performance and effectively predicted long-term kidney outcome.

Keywords: Chronic kidney failure, Immunoglobulin A nephropathy, Machine learning, Proteinuria

Received: March 22, 2023; **Revised:** June 22, 2023; **Accepted:** July 17, 2023

Correspondence: Dukyong Yoon

Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu Seoul 03722, Republic of Korea. E-mail: dukyong.yoon@yonsei.ac.kr

ORCID: <https://orcid.org/0000-0003-1635-8376>

Hyeong Cheon Park

Division of Nephrology, Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, Seoul 06273, Republic of Korea. E-mail: amp97@yuhs.ac

ORCID: <https://orcid.org/0000-0002-1550-0812>

*Yujeong Kim and Jong Hyun Jhee contributed equally to this work study as co-first authors.

†Dukyong Yoon and Hyeong Cheon Park contributed equally to this work as co-corresponding authors.

© 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.

Introduction

Immunoglobulin A nephropathy (IgAN) is a prevalent primary glomerulonephritis, especially in Asian countries where it comprises up to 50% of cases [1,2]. IgAN is common in young adults, aged 20 to 30 years [3]. After IgAN diagnosis, kidney function deteriorates, and >30% of IgAN patients progress to end-stage kidney disease (ESKD), requiring dialysis in 10 to 25 years [4,5]. Moreover, the incidence of cardiovascular complications and mortality are significantly increased in patients with IgAN compared with individuals of the same age and sex [6,7]. Hence, research has focused on early risk stratification of IgAN disease progression to prevent further development of adverse outcomes. Various factors, including baseline kidney function, proteinuria, blood pressure (BP), and histologic findings, contribute to the pathophysiology of IgAN, disrupting the accurate prediction of the disease prognosis [8]. Growing evidence suggests that modifying these factors early in the disease course may prevent the long-term decline in kidney function [9–11]. However, creating precise long-term prediction tools for IgAN outcomes, especially during the early stages, is difficult due to the disease's rarity, time lag between diagnosis and outcome development, and infrequent hard outcomes (such as a 50% reduction in estimated glomerular filtration rate (eGFR) or ESKD) [10]. Moreover, based on the disease complexity that multifactorial risks affect the progression of IgAN, an accurate establishment of prediction models for IgAN progression is challenging [12].

Recently, various risk-scoring systems to predict IgAN progression were established using multiple clinical and pathological findings [13,14]. However, these systems have limitations such as insufficient sample sizes, different pathological scoring criteria, and relatively few variables. To overcome the shortcomings of prediction models from conventional statistical methods, machine learning approaches have been applied in models for predicting the progression of IgAN [12,15–17]. Machine learning-based prediction algorithms show better predictive performance, cover larger data sets, and interpret more complex interactions than conventional tools. Nevertheless, the accuracy and practical applicability of these prediction models for determining long-term outcomes in patients with IgAN remain uncertain. In addition, the potential impact of short-

term outcomes, especially those occurring within 1 to 2 years after biopsy, on the long-term prognosis of IgAN is unclear due to the lack of tools to predict the short-term outcome.

This study therefore aimed to develop a machine learning-based 2-year risk prediction model for IgAN progression, based on kidney function decline and proteinuria improvement using a database of kidney biopsy-proven IgAN patients. We also validated the usefulness of this prediction model for predicting 10-year long-term kidney outcome.

Methods

Study population and design

This was a retrospective study using databases from two tertiary hospitals (Severance Hospital and Gangnam Severance Hospital at Yonsei University College of Medicine). The overall research workflow is summarized in Fig. 1. A total of 1,864 patients with biopsy-proven IgAN from May 2005 to January 2021 were initially screened. The patients were excluded on the basis of the following criteria: patients aged <18 years, baseline eGFR of <15 mL/min/1.73 m², underwent dialysis or kidney transplantation, and missing laboratory test results. Finally, a total of 1,301 patients were included for primary outcome analysis. For secondary outcome analysis, as the secondary outcome was defined as improvement in urine protein to creatinine ratio (UPCR) <1.0 g/g Cr accompanied by a 30% decline from baseline, a total of 597 patients were included after excluding individuals with a baseline UPCR of <1.0 g/g Cr or those with missing follow-up UPCR data (Fig. 2).

This study was approved by the Institutional Review Board of Yonsei University Severance Hospital (No. 3-2021-0059). In addition, the study was conducted following the guiding principles of the Declaration of Helsinki. Furthermore, the requirement for written informed consent was waived because of the study's retrospective nature. Overall, the methods and results followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines [18].

Data collection

Baseline demographic data, including age, sex, BP, pulse

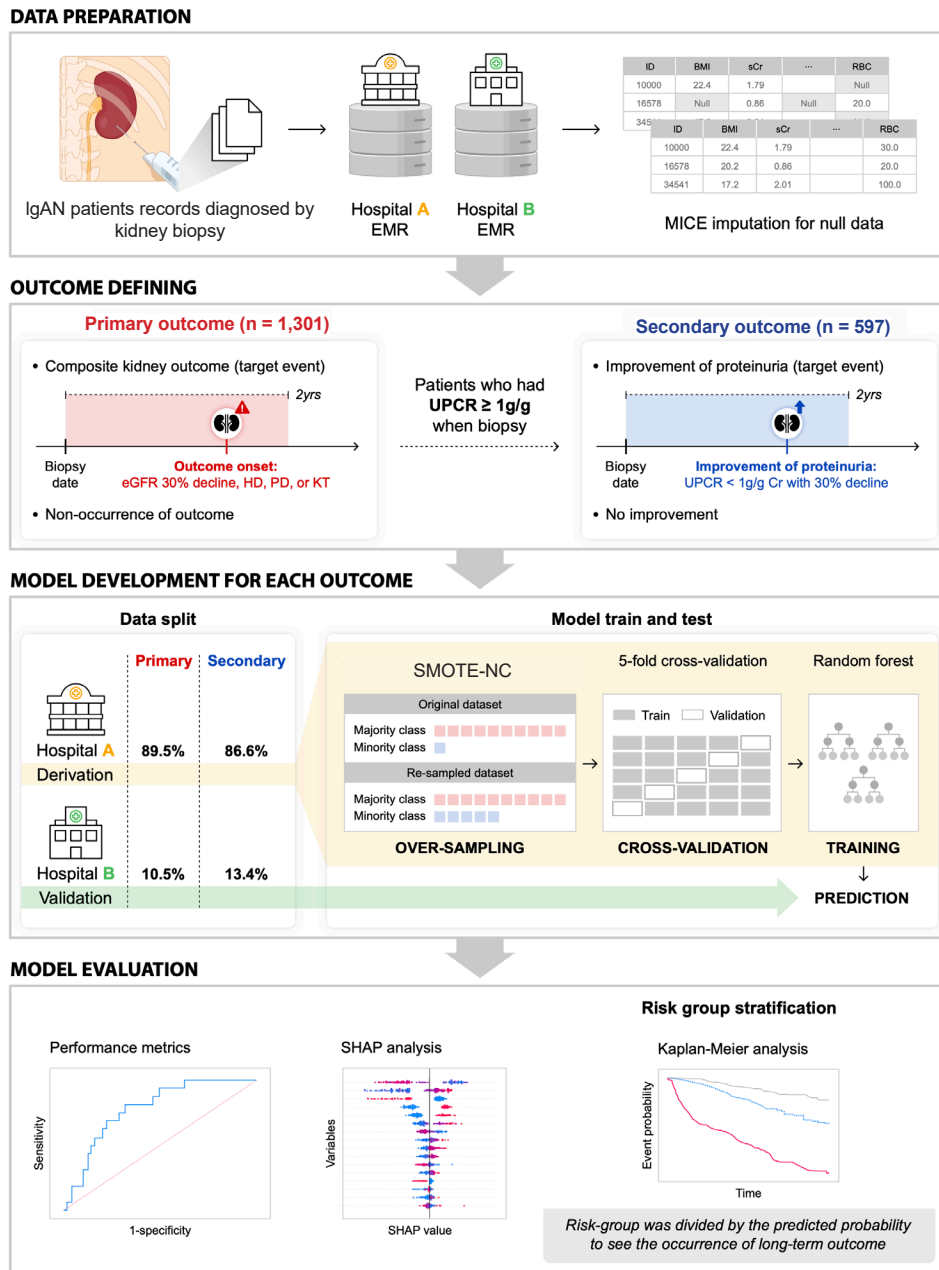


Figure 1. Model development overview. Electronic medical record (EMR) data of patients with immunoglobulin A nephropathy (IgAN), including demographic information, laboratory tests, drugs, and medical history. MEST classification, and features from computed tomography and sonography, were derived from hospitals A and B. After arranging data according to the biopsy date, multiple imputation by chained equations (MICE) imputation was adopted, and both outcomes were defined. The primary outcome is the occurrence of the composite kidney outcome, and the secondary outcome is an improvement of proteinuria. For model development, each hospital's data were used as the derivation and validation cohorts. Synthetic minority over-sampling technique for nominal and continuous features (SMOTE-NC) and five-fold cross-validation were used during the model development. A random forest algorithm was adopted, and the model was first evaluated through performance metrics and Shapley Additive exPlanations (SHAP) analysis. Additionally, three risk groups were stratified to evaluate the occurrence of composite kidney outcomes. Kaplan-Meier analysis for 10 years and incidence ratio analysis was performed.

BMI, body mass index; eGFR, estimated glomerular filtration rate; HD, hemodialysis; KT, kidney transplant; PD, peritoneal dialysis; RBC, red blood cell; sCr, serum creatinine; UPCR, urine protein/creatinine ratio.

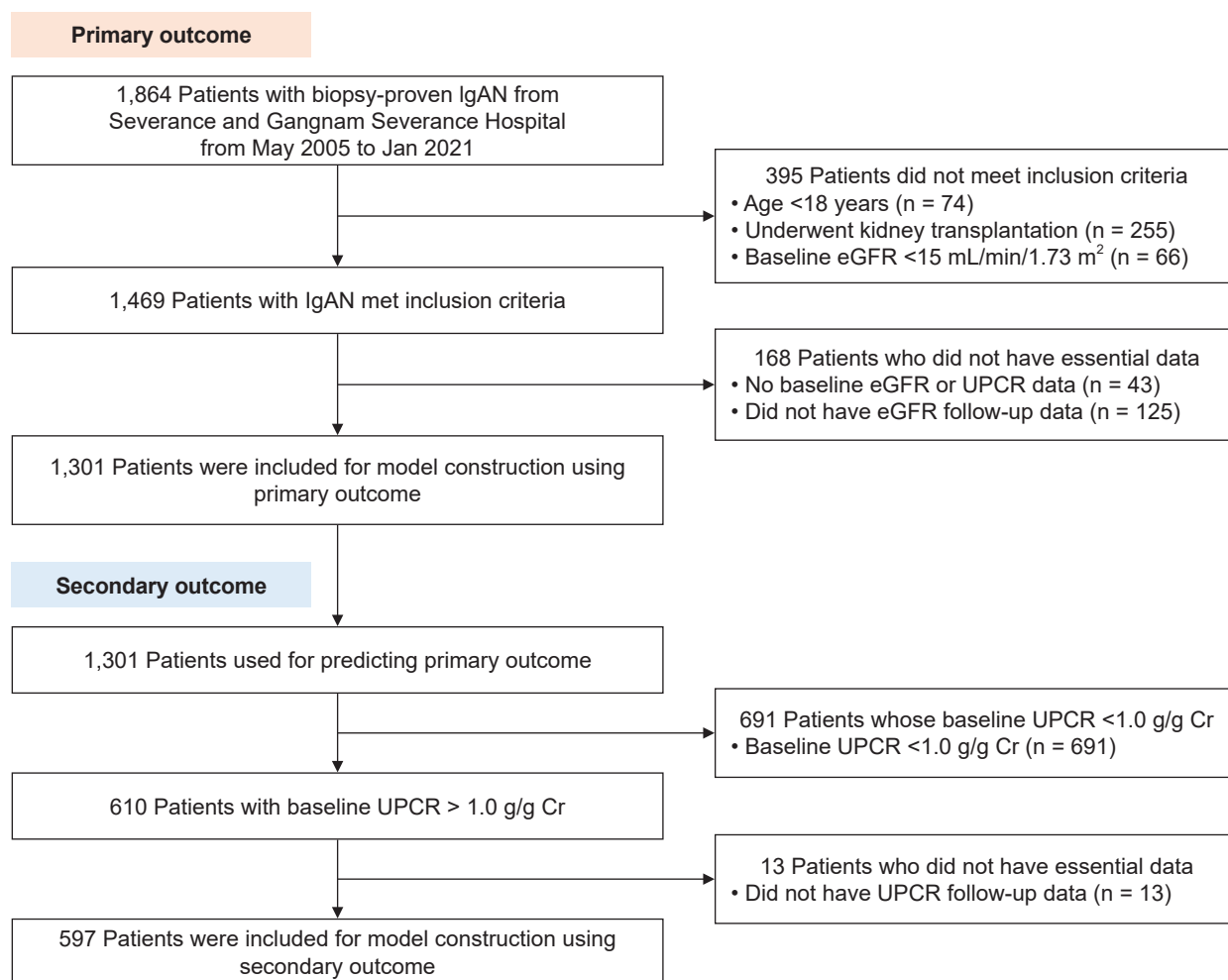


Figure 2. Patient flow chart. Patients diagnosed with immunoglobulin A nephropathy (IgAN) were selected as the whole study population. After the exclusion criteria were applied, 1,301 patients remained to predict the primary outcome and IgAN. eGFR, estimated glomerular filtration rate; UPCr, urine protein/creatinine ratio.

rate, body mass index (BMI), and alcohol or smoking status, were collected. Medical histories such as hypertension or diabetes, use of medications such as angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, statins, steroids, or other immunosuppressants, and initial presenting symptoms such as edema or hematuria were collected. Pathological findings from kidney biopsy were reported using Oxford classification and represented with M, E, S, and T scores (MEST scores) which means mesangial hypercellularity (M), endocapillary hypercellularity (E), segmental glomerulosclerosis (S), and tubular atrophy/interstitial fibrosis (T) [19]. Laboratory data were collected from fasting blood samples. Serum creatinine levels were measured using the rate-blanked compensated

Jaffe kinetic method with the Roche reagent Creatinine and the Roche Calibrator for Automated Systems, traceable to an isotope dilution mass spectrometry (IDMS) reference method in the Hitachi Automatic Analyzer (Hitachi). Since April 2011, the Severance Hospital adopted an IDMS traceable method to measure serum creatinine (the value of serum creatinine measured before this date was modified using the following equation: new serum creatinine = 1.049 × previous value - 0.129) [20]. The eGFR was calculated using the CKD-EPI (Chronic Kidney Disease-Epidemiology Collaboration) equation [21]. Urine samples were collected in the morning after the first voiding. Fresh urine samples were analyzed using URISCAN Pro II (YD Diagnostics Corp.). The presence of proteinuria was assessed

by the UPCR. Features from the computed tomography scan and sonography, such as kidney size, echogenicity, and the presence of renal stone or cyst, were obtained. For the missing values of each variable, those that had >50% null were excluded. For the remaining variables, multiple imputation by chained equations (MICE) were applied to impute the missing data [22]. Finally, 43 variables were included to predict the study outcome.

Study outcome

The primary outcome of this study was the deterioration of kidney function within 2 years after the kidney biopsy. The deterioration of kidney function was defined as a composite of a 30% decline in eGFR from baseline, the development of ESKD requiring hemodialysis or peritoneal dialysis, or kidney transplantation. The proportion of each outcome definition for the primary outcome is depicted in [Supplementary Fig. 1](#) (available online). The secondary outcome was an improvement of proteinuria among patients whose baseline UPCR was ≥ 1.0 g/g Cr. The improvement of proteinuria was defined as an improvement in UPCR to < 1.0 g/g Cr and a 30% decline in UPCR from baseline within 2 years of follow-up after kidney biopsy.

Two-year prediction model development and performance test

This study developed a two-step 2-year outcome prediction model. Firstly, we developed the prediction model evaluating the risk of a 30% decline in eGFR or development of ESKD (primary outcome), involving a total of 1,301 subjects including those with a baseline UPCR of < 1.0 g/g Cr. We then developed a second prediction model for evaluating the improvement of UPCR to < 1.0 g/g Cr with a 30% decrease from the baseline (secondary outcome), involving 597 subjects after excluding 691 subjects with a baseline UPCR of < 1.0 g/g Cr.

To construct the prediction model, multicenter data were divided into the model derivation and validation cohort according to the patient's hospital; Severance Hospital data was the model derivation cohort, and Gangnam Severance Hospital data was the validation cohort. Owing to the imbalanced ratio between the positive (event) and negative (no event) records in the derivation cohort that can cause

bias during the model development, the synthetic minority over-sampling technique (SMOTE) [23] was applied to balance the derivation cohort for the primary outcome. Among various SMOTE methods, we selected a method that can handle numerical and categorical data, called SMOTE-nominal and continuous (SMOTE-NC) [23]. The final derivation cohort, used as input for the prediction model, has a ratio of 1:2 between cases with composite outcomes and cases without composite outcomes for the primary outcome. We used a random forest-based machine learning algorithm to develop the prediction models. Random forest is an ensemble machine learning algorithm that can handle binary outcomes [24]. The grid search was performed using the derivation cohort with internal five-fold cross-validation for criterion after the random search; the method used to measure the split quality, `max_depth`; the maximum depth of the tree, `max_features`; the number of features to split the tree well, and `n_estimator`; the number of trees to adapt the best combination of parameters for a random forest. The range of each parameter is listed: `max_depth` from 2 to 10, `max_features` every 5 from 5 to 30, `n_estimators` every 10 from 10 to 100, and criterion either gini or entropy. For the primary outcome, 10 for `max_depth`, 5 for `max_features`, 90 for `n_estimators`, and entropy criterion were chosen. For the secondary outcome, 8 for `max_depth`, 10 for `max_features`, 10 for `n_estimators`, and entropy criterion were chosen as a combination of best parameters.

The 2-year prediction model's performances were evaluated based on the following performance indexes from both cohorts: accuracy, precision, recall, the area under the precision-recall curve (AUPRC), the area under the receiver operating characteristic curve (AUROC), Brier score, and F1-score. The classification threshold for the primary outcome was defined as the point where the recall reached approximately 0.8.

Feature importance analysis

To produce an explainable prediction model, feature importance was calculated using the Shapley Additive exPlanations (SHAP) method. SHAP analysis provides information about which features have a high contribution to predicting the outcome, with each feature value's contribution to deciding whether the sample is positive or negative by calculating Shapley value [25,26].

Long-term risk stratification

To test the predictive value of our 2-year prediction models on long-term (10-year follow-up) risk of kidney disease progression, the association between the predicted probability of primary and secondary outcomes from the 2-year prediction models and long-term risk of kidney outcome was evaluated. Each predictive probability derived from the 2-year risk prediction tools was categorized into three groups based on predictive probability: less than 0.5 (low), between 0.5 to 0.75 (moderate), and greater than 0.75 (high). Ten-year risk of kidney disease progression was defined as a composite of a 30% decline in eGFR from baseline and the development of ESKD during the follow-up period. Kaplan-Meier analysis was performed to compare the 10-year kidney survival among the three groups based on predictive probability. Statistical comparisons for each group were made by log-rank test, and Cox proportional hazards regression analysis was used to calculate the relative hazard ratio (HR) among the three groups.

Software

Python version 3.7 and R studio version 4.0.3 were used for the data preprocessing and model development. For data imputation, MICE package version 3.13.0 was used.

SMOTE-NC package version 0.8 was used for data balancing, and scikit-learn package version 0.43 was used for model development. Lifelines30 package version 0.26 and SHAP package version 0.39 were used to evaluate the model.

Results

Baseline characteristics

The baseline characteristics at the time of biopsy of 1,301 patients included for the development of the primary outcome model are shown in [Table 1](#). In the derivation cohort (n = 1,165), the mean age of patients was 39.8 years and 54.8% were female. The mean age of patients in the validation cohort (n = 136) was 35.8 years and 50.7% were female. For both cohorts, the drug use proportion including for angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, statin, steroids, and other immunosuppressants was similar. In particular, the use of angiotensin receptor blockers was 40.1% and 36.0% in the derivation and validation cohorts, respectively. In the laboratory tests, no significant difference was observed between the derivation and validation cohorts' mean serum creatinine levels; 1.03 ± 0.53 and 1.07 ± 0.55 mg/dL, respectively (p = 0.447). However, there was a significant difference in UPCR levels between the two cohorts: median of 0.88 g/g Cr (interquar-

Table 1. Baseline characteristics from the primary outcome (n = 1,301)

Characteristic	Derivation cohort	Validation cohort	p-value
Demographics			
No. of patients	1,165 (89.5)	136 (10.5)	
Age (yr)	39.82 ± 13.41	35.78 ± 13.58	<0.001
Female sex	639 (54.8)	69 (50.7)	0.41
SBP (mmHg)	130.90 ± 17.58	125.39 ± 16.53	0.001
DBP (mmHg)	84.16 ± 11.86	80.70 ± 11.84	0.001
Pulse pressure (mmHg)	77.15 ± 12.32	77.65 ± 10.30	0.65
Body mass index (kg/m ²)	23.58 ± 4.13	23.67 ± 3.67	0.81
Alcohol			
Current	437 (37.5)	48 (35.3)	
Past	85 (7.3)	9 (6.6)	
Never	607 (52.1)	78 (57.4)	
Smoking			
Current	151 (13.0)	9 (6.6)	0.07
Past	104 (8.9)	16 (11.8)	
Never	897 (76.7)	110 (80.9)	

(Continued to the next page)

Table 1. Continued

Characteristic	Derivation cohort	Validation cohort	p-value
Medical history			
Hypertension	387 (33.2)	41 (30.1)	0.29
Diabetes mellitus	40 (3.4)	3 (2.2)	0.62
Symptoms and signs at biopsy			
Edema	23 (2.0)	2 (1.5)	>0.99
Gross hematuria	134 (11.5)	29 (21.3)	0.003
Oxford classification			
Mesangial (M) 1	205 (17.6)	20 (14.7)	0.43
Endocapillary (E) 1	219 (18.8)	31 (22.8)	0.31
Glomerular sclerosis (S) 1	796 (68.3)	96 (70.6)	0.88
Tubulointerstitial damage (T) 1	116 (10.0)	18 (13.2)	0.004
Tubulointerstitial damage (T) 2	19 (1.6)	8 (5.9)	
Drugs			
ACEi	20 (1.7)	3 (2.2)	0.73
ARB	467 (40.1)	49 (36.0)	0.35
Statin	169 (14.5)	22 (16.2)	0.74
Steroid	23 (2.0)	4 (2.9)	0.52
Other immunosuppressants	17 (1.5)	2 (1.5)	>0.99
Laboratory tests			
Serum creatinine (mg/dL)	1.03 ± 0.53	1.07 ± 0.55	0.447
Serum albumin (g/dL)	4.04 ± 0.51	4.02 ± 0.54	0.81
Fasting plasma glucose (mg/dL)	99.00 ± 19.48	108.54 ± 35.65	<0.001
Serum hemoglobin (g/dL)	12.79 ± 1.68	13.22 ± 1.87	0.005
Serum uric acid (mg/dL)	5.81 ± 1.72	6.45 ± 1.80	<0.001
Serum triglyceride (mg/dL)	141.76 ± 113.55	148.77 ± 108.32	0.51
Serum LDL-C (mg/dL)	115.09 ± 38.17	115.02 ± 30.86	0.99
Serum HDL-C (mg/dL)	54.25±15.52	53.44 ± 18.12	0.59
C reactive protein (mg/L)	0.8 (0.3–2.6)	0.7 (0.3–1.7)	0.16
C3 (mg/dL)	111.32 ± 20.18	112.82 ± 24.17	0.55
C4 (mg/dL)	27.98 ± 9.18	33.15 ± 12.58	<0.001
IgA (mg/dL)	326.77 ± 114.06	301.98 ± 84.10	0.08
IgG (mg/dL)	1,150.07 ± 351.40	1,057.42 ± 302.58	0.04
IgM (mg/dL)	116.44 ± 61.56	118.98 ± 49.29	0.74
UPCR (g/g Cr)	0.88 (0.42–1.72)	1.19 (0.56–2.16)	0.005
Urine RBC (/HPF)	30 (10–100)	20 (5–100)	0.001
RBC dysmorphism (%)	22.04 ± 27.17	16.04 ± 18.89	0.08
Radiologic findings			
Right kidney size (cm)	10.19 ± 2.99	10.56 ± 0.99	0.21
Left kidney size (cm)	10.29 ± 1.02	10.47 ± 1.31	0.08
Chronic change	278 (23.9)	29 (21.3)	0.61
Renal cyst	121 (10.4)	16 (11.8)	0.66
Renal stone	31 (2.7)	6 (4.4)	0.26

Data are expressed as mean ± standard deviation, median (interquartile range), or number (%).

ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blockers; C3, complement component 3; C4, complement component 4; DBP, diastolic blood pressure; HDL, high-density lipoprotein cholesterol; HPF, high power field; Ig, immunoglobulin; LDL-C, low-density lipoprotein cholesterol; RBC, red blood cell; SBP, systolic blood pressure; UPCR, urine protein/creatinine ratio.

tile range [IQR], 0.42–1.72 g/g Cr) and 1.19 g/g Cr (IQR, 0.56–2.16 g/g Cr), respectively ($p = 0.005$). In the Oxford classification, over 70% of patients had glomerular sclerosis in two cohorts. There were no significant differences in radiologic features between the two cohorts. The baseline characteristics of 597 patients included for the secondary outcome (improvement of proteinuria) analysis are shown in [Supplementary Table 1](#) (available online).

Performance of 2-year risk prediction tools

During a follow-up for the primary outcome (median, 285 days; IQR, 168–453 days), 159 events (derivation set, 12.27%; validation set 11.76%) occurred. For the secondary outcome, 474 events (derivation set, 78.53%; validation set, 85.00%) occurred during the follow-up (median, 85 days; IQR, 25–194 days). Over all performances of the 2-year risk prediction models for both primary and secondary outcomes are summarized in [Table 2](#) and [Fig. 3](#). In the model derivation cohort for the primary outcome, AUROC and AUPRC were 0.993 and 0.991 with precision of 0.999 and recall of 0.986, and the Brier and F1 scores were 0.005 and 0.993. In the validation cohort for the primary outcome, the AUROC and AUPRC were 0.771 and 0.242 with precision of 0.259 and recall of 0.875, and the Brier and F1 scores were 0.309 and 0.400. Regarding the secondary outcome in the derivation cohort, the AUROC and AUPRC were 0.829 and 0.914 with precision of 0.914 and recall of 1.000, and the Brier and F1 scores were 0.074 and 0.955, respectively. In the validation cohort for the secondary outcome, the AUROC and AUPRC were 0.694 and 0.903 with precision of 0.904 and recall of 0.971, and the Brier and F1 scores were 0.113 and 0.955, respectively.

Feature importance in the 2-year risk prediction models

The SHAP method was used to identify the feature importance of our machine learning-based 2-year risk prediction models ([Fig. 4](#)). The top features from the SHAP analysis can interpret the outcome risk discrimination from each model. For the 2-year risk of the primary outcome, the most important feature of the prediction model was UPCR. The next most important features were serum albumin, systolic BP (SBP), urine red blood cell (RBC) dysmorphism, and serum immunoglobulin G (IgG). The SHAP force plots in [Supplementary Fig. 2](#) (available online) show the explanation for the prediction of each patient with one of the highest or lowest probabilities of outcome risk determined by our prediction models. For the 2-year risk of the primary outcome prediction model, a large amount of UPCR and proportion of urine dysmorphic RBC and low serum albumin and IgG levels indicated the increased risk of a 2-year kidney outcome. For the 2-year risk of the secondary outcome prediction model, UPCR was the most informative feature of the model. The next most important features were serum C4, serum creatinine, serum C3, IgG, and BMI. Less UPCR and low serum creatinine levels were indicative of an increased potential for proteinuria improvement in the 2-year period.

Long-term risk stratification by 2-year risk prediction tools

Subsequently, we evaluated whether our 2-year risk prediction tools have a predictive value on the 10-year long-term risk of kidney disease progression. Firstly, we evaluated three risk probability groups derived from 2-year primary outcome risk prediction model ($n = 1,301$) for the long-term risk ([Fig. 5A](#)). During the median of approximately 3.5

Table 2. Prediction model performance metrics for the derivation and validation cohorts

Outcome	Accuracy	Precision	Recall	AUPRC	AUROC	Brier score	F1-score
Primary outcome							
Derivation cohort	0.995	0.999	0.986	0.991	0.993	0.005	0.993
Validation cohort	0.691	0.259	0.875	0.242	0.771	0.309	0.400
Secondary outcome							
Derivation cohort	0.998	0.914	1.000	0.914	0.829	0.074	0.955
Validation cohort	0.850	0.904	0.971	0.903	0.694	0.113	0.955

AUPRC, area under the precision and recall curve; AUROC, area under the receiver operating characteristic curve.

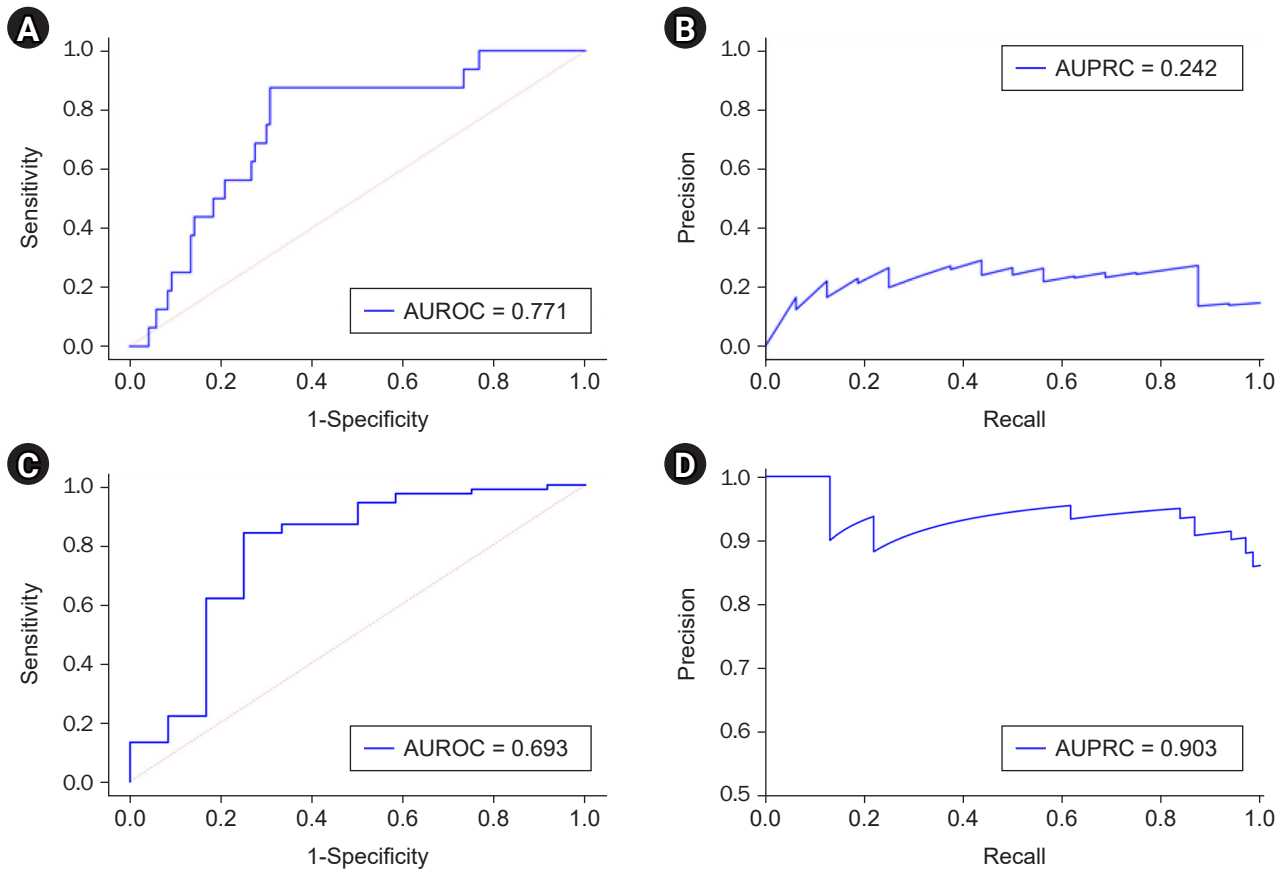


Figure 3. Performance of 2-year risk prediction models. (A) and (B) are the receiver operating characteristic (ROC) and precision-recall (PRC) curves for the primary outcome. (C) and (D) are the ROC and PRC curves for the secondary outcome. The solid line presents the ROC and PRC, and the dotted line for the ROC curve presents the situation when samples were classified randomly. AUROC, area under the ROC curve; AUPRC, area under the PRC curve.

years (IQR, 1.3–6.5 years) of follow-up, 357 events (27.4%) occurred. In the Kaplan-Meier analysis, the higher-risk groups showed an increased long-term risk of kidney disease progression compared with the lowest group (HR, 13.00; 95% confidence interval [CI], 9.52–17.77 in the high group and HR, 12.90; 95% CI, 9.92–16.76 in the moderate group).

Next, we evaluated three risk probability groups derived from the 2-year secondary outcome risk prediction model ($n = 597$) for the long-term risk (Fig. 5B). During the follow-up (median, about 2.9 years; IQR, 1.0–5.8 years), 244 events (41.0%) occurred. Among three risk probability groups, groups representing a lower probability to improve proteinuria within 2 years after biopsy showed an increased long-term risk of kidney disease progression compared with the highest group (HR, 1.66; 95% CI, 1.42–1.95

in the low group and HR, 1.42; 95% CI, 0.99–2.03 in the moderate group).

Discussion

In this study, we developed a machine learning-based 2-year risk prediction tool for primary (30% decline in eGFR or ESKD) and secondary (improvement of proteinuria) outcomes within 2 years of follow-up after kidney biopsy in patients with IgAN. The highest feature importance by SHAP analysis for the 2-year risk prediction tools, for both primary and secondary outcomes, was the amount of proteinuria at the time of biopsy. However, serum creatinine level at biopsy was ranked as a less contributing variable for feature importance in the 2-year risk prediction tools. Finally, the 2-year risk prediction tools effectively dis-

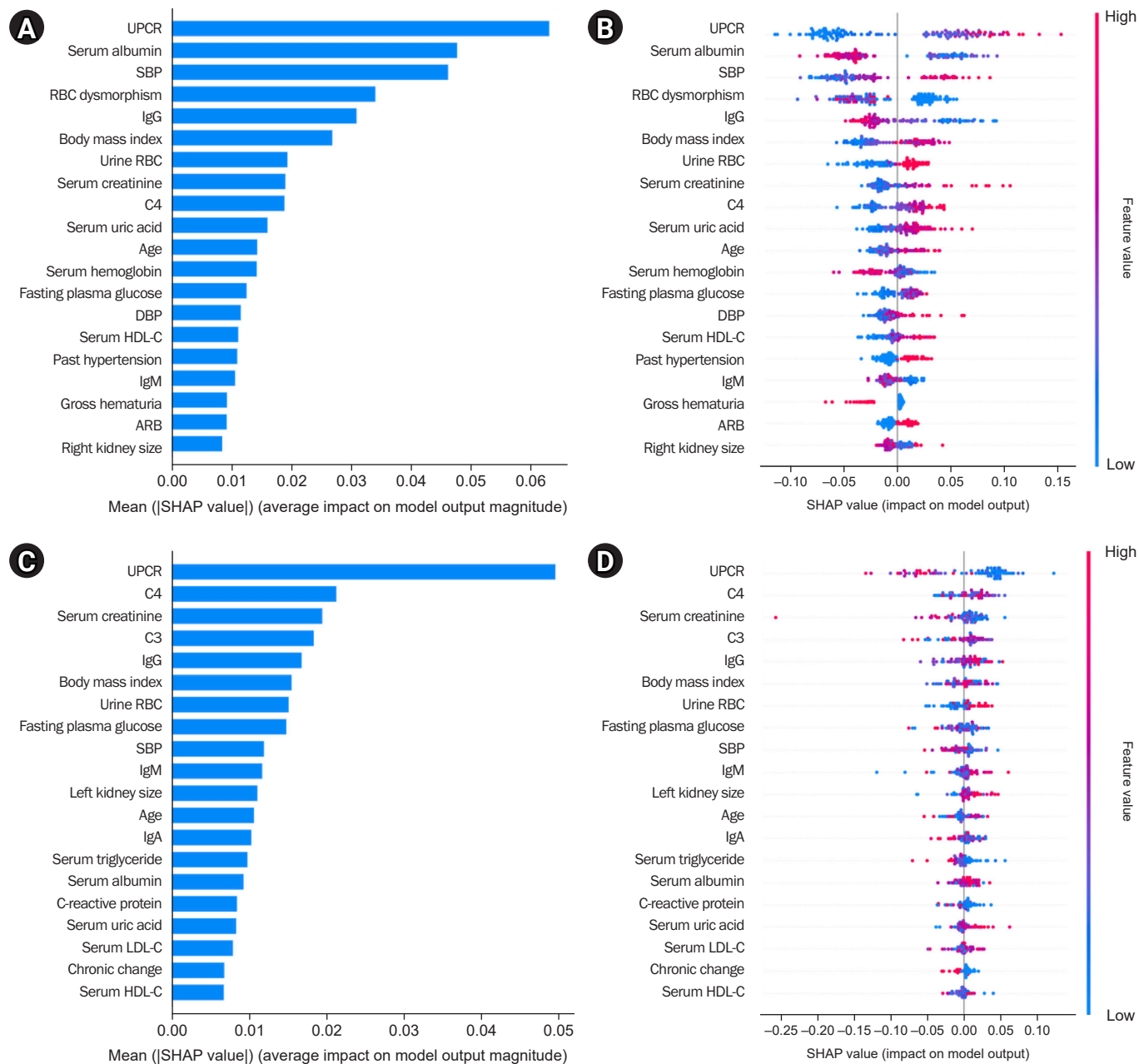


Figure 4. Summary and force plots to interpret the model using SHAP analysis. (A) and (B) for the primary. (C) and (D) for the secondary outcome prediction. (A) and (C) show feature importance in the order, high SHAP value to low SHAP value. (B) and (D) shows more detailed importance than (A) and (C). (B) and (D) depict the relationship between each feature's value and their impact on predicting the event. Red represents a high feature value, and blue represents a low feature value.

ARB, angiotensin receptor blocker; C3, complement component 3; C4, complement component 4; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; Ig, immunoglobulin; RBC, red blood cell; SBP, systolic blood pressure; SHAP, Shapley Additive exPlanations; UPCR, urine protein/creatinine ratio.

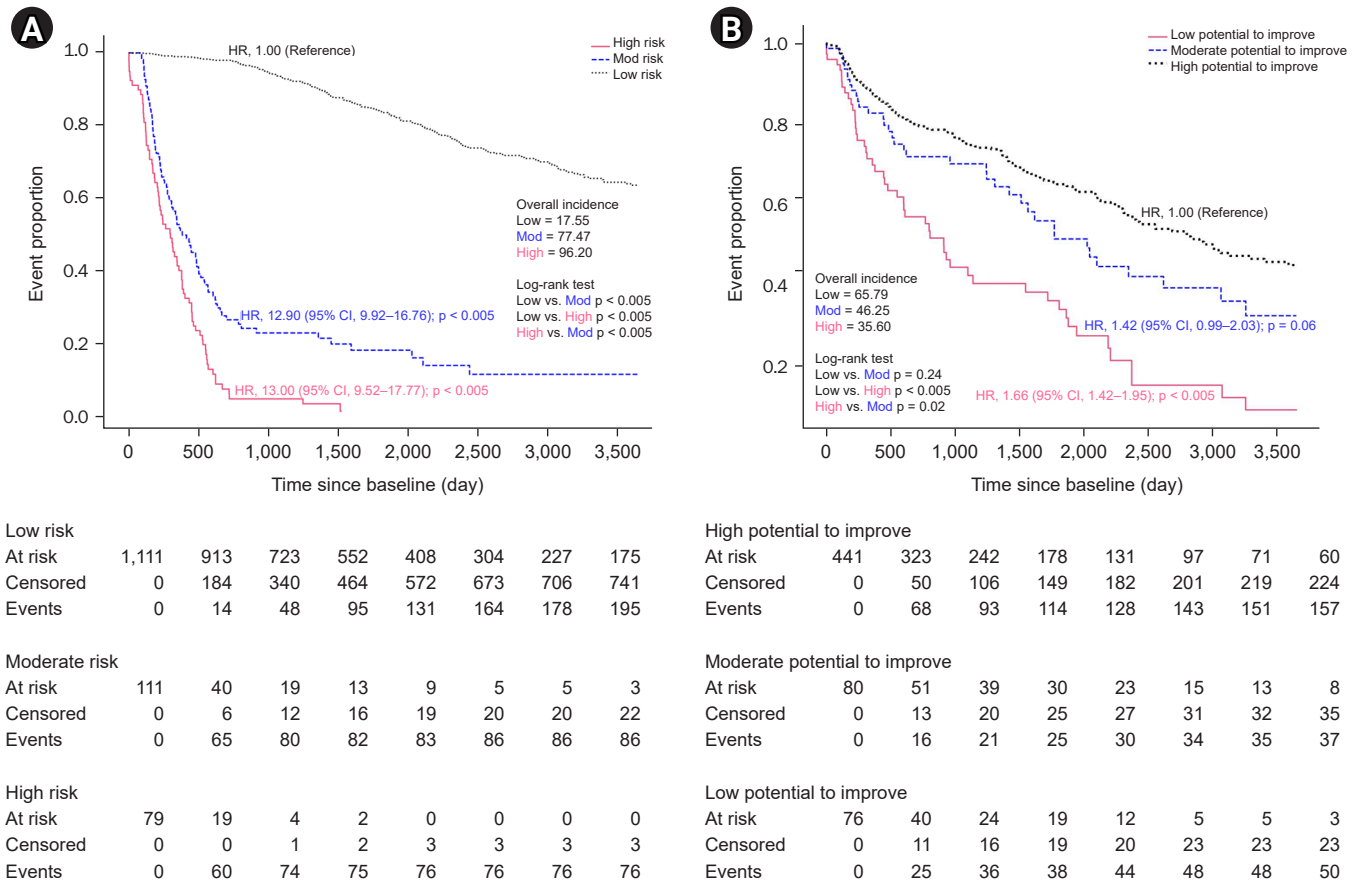


Figure 5. Long-term risk stratification according to the predicted 2-year risk probability. Kaplan-Meier analysis for composite kidney outcome according to the predicted 2-year risk probability based on (A) primary outcome analysis and (B) secondary outcome analysis. CI, confidence interval; HR, hazard ratio.

criminated the risk for the 10-year long-term progression of IgAN using risk groups based on the predictive probability of 2-year risk prediction tools.

Although IgAN shows a slowly progressive nature compared with other glomerular diseases, there is a substantial heterogeneous clinical course in IgAN [27,28]. Furthermore, multiple risk factors are involved in the progression of IgAN, and an individualized estimate of a patient's risk of disease progression is essential. Risk prediction may help collaborative decision-making regarding treatment strategies. Recently, various risk prediction tools were developed for IgAN [12,16,17,29]. The KDIGO (Kidney Disease Improving Global Outcomes) 2021 guidelines for managing glomerular disease suggested the International IgAN Prediction Tool (IIgAN-PT) as a useful resource to assess the risk of progression and facilitate shared decision-making with patients [30]. The IIgAN-PT was externally validated

using models with multiple ethnic cohorts, including over 4,000 participants [31]. However, Barbour et al. [32] showed that the original IIgAN-PT did not predict outcomes as accurately as when used 1 year after biopsy. Although IgAN has a slow progressive nature, treatment guidelines recommend immunosuppressant treatment if there is no improvement in supportive care within a short period after diagnosis [30]. Hence, it is essential to determine how clinical changes within 1 to 2 years after diagnosis affect the long-term risk. Therefore, the IIgAN-PT should be re-evaluated to predict the risk of disease progression after observation and supportive care. Moreover, the model incorporated established risk factors for the progression of IgAN such as age, BP, eGFR, proteinuria, renin-angiotensin system blockade use, immunosuppressant use, and MEST-C score. In addition, this tool provides data on the long-term risk, especially at least 5-year predicted risks without data on

short-term clinical courses following the diagnosis of IgAN. On the other hand, in various risk prediction studies, the long-term risk was predicted by accumulating follow-up data for at least 2 years after diagnosis [33,34]. However, the need for prolonged periods of follow-up data is one of the limitations of previous IgAN prediction models, which reduced their clinical utility. Therefore, the present study is novel for examining the effect of predicted probability for event occurrence based on a machine learning approach within 2 years on predicting the long-term risk for IgAN progression. No previous study has evaluated the usefulness of the machine learning-based 2-year risk probability on long-term risk prediction in IgAN.

Early change in proteinuria is a reliable surrogate outcome in IgAN, and a basis for treatment decision-making [8,10]. An individual participant-level meta-analysis performed by Inker et al. [11] provided evidence that an early reduction in proteinuria of 30% from baseline would confer the probabilities of at least 90% treatment benefits on the long-term risk of disease progression in IgAN. In this study, a 50% reduction in proteinuria after 9 months of treatment was associated with a 60% decrease in the risk of composite kidney outcome. Furthermore, reduction in proteinuria accounted for 11% and 29% of the treatment effect from RASB and steroids, respectively. Canney et al. [10] also supported these findings and concluded that a shorter duration needed to achieve proteinuria remission was associated with significant reductions in the risk of disease progression in IgAN. In accordance with previous studies, our 2-year risk prediction tool for the improvement of proteinuria successfully predicted the probability of 10-year long-term kidney outcome risk. This study is the first to develop a machine learning-based short-term risk prediction model for changes in proteinuria.

In the present study, the highest feature importance evaluated by SHAP analysis for the 2-year risk prediction models for both primary and secondary outcomes was UPCR at the time of biopsy. However, serum creatinine level at biopsy was ranked as a less contributing variable for feature importance in both prediction models. Historically, eGFR and proteinuria during diagnosis are the most significant factors in determining long-term prognosis in IgAN. Nevertheless, our findings showed that proteinuria is the most useful for predicting short-term (within 2 years) clinical outcomes. In contrast, serum creatinine level may

have a less significant effect. However, since most patients in this study showed normal kidney function at biopsy, the results should be interpreted cautiously for patients with advanced kidney disease during diagnosis. RBC dysmorphism emerged as an important variable in the prediction model for primary outcome. The accompanying dysmorphic RBC may be closely associated with glomerular structural damage and a decrease in eGFR [35–37]. Additionally, previously known risk factors for the progression of IgAN such as SBP and BMI showed high feature importance in the primary outcome prediction model. Furthermore, serum albumin emerged as a high-importance feature in both prediction models in accordance with previous studies that hypoalbuminemia is a significant risk factor for poor kidney outcome [38,39].

Our study has some limitations. First, the prediction model was derived from the retrospective cohort data of patients with IgAN. Furthermore, the data consisted of a single Korean ethnic group; hence, the findings should be generalized with caution. Thus, prospective cohort studies including multi-ethnic participants are needed to evaluate the validity of this 2-year risk prediction model in patients with IgAN. Second, it is unclear whether therapeutic interventions during the follow-up period of IgAN may change the disease course from our prediction model. Third, the importance of pathologic data contributing to the prediction model for IgAN is limited in this study. Only the Oxford classification was considered as pathologic data in this study. Therefore, incorporating various other features such as digital image analysis, immune cell infiltration, vascular features, and other microscopic findings from biopsy specimens may provide further insights for enhancing the prediction model [2]. Future studies are necessary to investigate this further. Lastly, the limited number of outcome events that occurred within 2-years after kidney biopsy hindered the development of prediction models using specific kidney events such as a 30% decline in eGFR or ESKD requiring dialysis. To validate this, further studies with larger cohorts will need to be performed.

In conclusion, our 2-year prediction models for the progression of IgAN based on eGFR decline, development of ESKD, or improvement of proteinuria successfully predicted the 10-year long-term risk of kidney outcome in patients with IgAN. Our 2-year prediction models can be implemented in clinical practice during biopsy and may provide

individualized benefits for treatment decision-making in the early course of IgAN.

Conflicts of interest

All authors have no conflicts of interest to declare.

Funding

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI22C0452) and supported by a faculty research grant of Yonsei University College of Medicine (6-2022-0118). This study was also supported by a new faculty research seed money grant of Yonsei University College of Medicine for 2021 (2021-32-0051).

Data sharing statement

The clinical data used to develop the model cannot be shared publicly because of the Personal Information Protection Act enforced by the government.

Authors' contributions

Conceptualization: DY, HCP, DO, BJL, HYC

Data curation, Formal analysis: YK, CMP

Funding acquisition: DY, JHJ

Investigation: YK, JHJ

Resources: DO, BJL, HYC, DY, HCP

Visualization: YK

Writing—original draft: YK, JHJ

Writing—review & editing: DO, BJL, HYC, DY, HCP, YK, JHJ

All authors read and approved the final manuscript.

ORCID

Yujeong Kim, <https://orcid.org/0000-0003-3674-993X>

Jong Hyun Jhee, <https://orcid.org/0000-0002-1255-1323>

Chan Min Park, <https://orcid.org/0000-0002-6088-6048>

Donghwan Oh, <https://orcid.org/0000-0001-7570-0288>

Beom Jin Lim, <https://orcid.org/0000-0003-2856-0133>

Hoon Young Choi, <https://orcid.org/0000-0002-4245-0339>

Dukyong Yoon, <https://orcid.org/0000-0003-1635-8376>

Hyeong Cheon Park, <https://orcid.org/0000-0002-1550-0812>

References

1. McGrogan A, Franssen CF, de Vries CS. The incidence of primary glomerulonephritis worldwide: a systematic review of the literature. *Nephrol Dial Transplant* 2011;26:414–430.
2. Chang JH, Kim DK, Kim HW, et al. Changing prevalence of glomerular diseases in Korean adults: a review of 20 years of experience. *Nephrol Dial Transplant* 2009;24:2406–2410.
3. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med* 2013;368:2402–2414.
4. Manno C, Strippoli GF, D'Altri C, Torres D, Rossini M, Schena FP. A novel simpler histological classification for renal survival in IgA nephropathy: a retrospective study. *Am J Kidney Dis* 2007;49:763–775.
5. Magistroni R, D'Agati VD, Appel GB, Kiryluk K. New developments in the genetics, pathogenesis, and therapy of IgA nephropathy. *Kidney Int* 2015;88:974–989.
6. Jarrick S, Lundberg S, Welander A, et al. Mortality in IgA nephropathy: a nationwide population-based cohort study. *J Am Soc Nephrol* 2019;30:866–876.
7. Lee H, Kim DK, Oh KH, et al. Mortality of IgA nephropathy patients: a single center experience over 30 years. *PLoS One* 2012;7:e51225.
8. Le W, Liang S, Hu Y, et al. Long-term renal survival and related risk factors in patients with IgA nephropathy: results from a cohort of 1155 cases in a Chinese adult population. *Nephrol Dial Transplant* 2012;27:1479–1485.
9. Oh TR, Choi HS, Oh SW, et al. Association between the progression of immunoglobulin A nephropathy and a controlled status of hypertension in the first year after diagnosis. *Korean J Intern Med* 2022;37:146–153.
10. Canney M, Barbour SJ, Zheng Y, et al. Quantifying duration of proteinuria remission and association with clinical outcome in IgA nephropathy. *J Am Soc Nephrol* 2021;32:436–447.
11. Inker LA, Heerspink HJ, Tighiouart H, et al. Association of treatment effects on early change in urine protein and treatment effects on GFR slope in IgA nephropathy: an individual participant meta-analysis. *Am J Kidney Dis* 2021;78:340–349.
12. Chen T, Li X, Li Y, et al. Prediction and risk stratification of kidney outcomes in IgA nephropathy. *Am J Kidney Dis* 2019;74:300–309.
13. Berthoux F, Mohey H, Laurent B, Mariat C, Afiani A, Thibaudin L. Predicting the risk for dialysis or death in IgA nephropathy. *J Am*

- Soc Nephrol* 2011;22:752–761.
14. Goto M, Wakai K, Kawamura T, Ando M, Endoh M, Tomino Y. A scoring system to predict renal outcome in IgA nephropathy: a nationwide 10-year prospective cohort study. *Nephrol Dial Transplant* 2009;24:3068–3074.
 15. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;319:1317–1318.
 16. Pesce F, Diciolla M, Binetti G, et al. Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. *Nephrol Dial Transplant* 2016;31:80–86.
 17. Schena FP, Anelli VW, Trotta J, et al. Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. *Kidney Int* 2021;99:1179–1188.
 18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
 19. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society; Cattran DC, Coppo R, et al. The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. *Kidney Int* 2009;76:534–545.
 20. Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. *JAMA* 2012;307:1941–1951.
 21. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* 2012;367:20–29.
 22. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–67.
 23. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321–357.
 24. Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
 25. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv: 1705.07874 [Preprint]. arXiv; 2017 [v1 posted 2017 May 22; v2 revised 2017 Nov 25; cited 2023 Mar 21]. Available from: <https://doi.org/10.48550/arXiv.1705.07874>
 26. Shapley L. A value for n-person games. In: Kuhn H, Tucker A, Contributions to the theory of games (AM-28), vol. II. Princeton University Press; 1953. p. 307–318.
 27. Pattapornpisut P, Avila-Casado C, Reich HN. IgA nephropathy: core curriculum 2021. *Am J Kidney Dis* 2021;78:429–441.
 28. Han SY, Jung CY, Lee SH, et al. A multicenter, randomized, open-label, comparative, phase IV study to evaluate the efficacy and safety of combined treatment with mycophenolate mofetil and corticosteroids in advanced immunoglobulin A nephropathy. *Kidney Res Clin Pract* 2022;41:452–461.
 29. Joo YS, Kim HW, Baek CH, et al. External validation of the international prediction tool in Korean patients with immunoglobulin A nephropathy. *Kidney Res Clin Pract* 2022;41:556–566.
 30. Rovin BH, Adler SG, Barratt J, et al. Executive summary of the KDIGO 2021 Guideline for the Management of Glomerular Diseases. *Kidney Int* 2021;100:753–779.
 31. Barbour SJ, Coppo R, Zhang H, et al. Evaluating a new international risk-prediction tool in IgA nephropathy. *JAMA Intern Med* 2019;179:942–952.
 32. Barbour SJ, Coppo R, Zhang H, et al. Application of the International IgA Nephropathy Prediction Tool one or two years post-biopsy. *Kidney Int* 2022;102:160–172.
 33. Bartosik LP, Lajoie G, Sugar L, Cattran DC. Predicting progression in IgA nephropathy. *Am J Kidney Dis* 2001;38:728–735.
 34. Barbour SJ, Espino-Hernandez G, Reich HN, et al. The MEST score provides earlier risk prediction in IgA nephropathy. *Kidney Int* 2016;89:167–175.
 35. Saha MK, Massicotte-Azarniouch D, Reynolds ML, et al. Glomerular hematuria and the utility of urine microscopy: a review. *Am J Kidney Dis* 2022;80:383–392.
 36. Pollock C, Liu PL, Györy AZ, et al. Dysmorphism of urinary red blood cells: value in diagnosis. *Kidney Int* 1989;36:1045–1049.
 37. Avguštin Rotar N, Jerman A, Škoberne A, Borštnar Š, Kojc N, Lindič J. The predictive value of urinary erythrocyte morphology for proliferative glomerular kidney disease. *Clin Nephrol* 2021;96:49–55.
 38. Walther CP, Gutiérrez OM, Cushman M, et al. Serum albumin concentration and risk of end-stage renal disease: the REGARDS study. *Nephrol Dial Transplant* 2018;33:1770–1777.
 39. Kikuchi H, Kanda E, Mandai S, et al. Combination of low body mass index and serum albumin level is associated with chronic kidney disease progression: the chronic kidney disease-research of outcomes in treatment and epidemiology (CKD-ROUTE) study. *Clin Exp Nephrol* 2017;21:55–62.