



OPEN

## Development and preliminary validation of a predictive model for IgA nephropathy progression

Jinjin Wang<sup>1,9</sup>, Tianmu Chen<sup>2,9</sup>, Yan Fu<sup>3,9</sup>, Tianqi Zhu<sup>4</sup>, Yaning Feng<sup>5</sup>, Yuxi Feng<sup>1</sup>, Xuan Zhang<sup>1</sup>, Yi Cai<sup>6</sup>, Li Gao<sup>7</sup>, Yongqiang Lin<sup>2</sup>✉, Qinyang Jin<sup>8</sup>✉ & Qin Zhu<sup>1</sup>✉

IgA nephropathy (IgAN), the most common primary glomerular disease worldwide, poses challenges in predicting progression at diagnosis—particularly in primary care. This study developed and validated a model to estimate 5-year renal survival and support early risk stratification for personalized management. A total of 1135 patients with biopsy-confirmed IgAN from Hangzhou Hospital of Traditional Chinese Medicine (2014–2017) were retrospectively enrolled and randomly assigned to training and internal validation cohorts in a 7:3 ratio. An external validation cohort comprising 352 patients was obtained from three independent centers (2015–2020). To identify robust prognostic variables, 1000 bootstrap samples were generated from the training set, each subjected to tenfold cross-validation to determine the optimal regularization parameter ( $\lambda$ ) for the LASSO-Cox model. Variables with non-zero coefficients were recorded across iterations, and the five most frequently selected were used to construct both the XGBoost survival model and a corresponding nomogram. Model performance was evaluated through discrimination, calibration, and clinical utility using ROC curves, Brier scores, calibration curve and decision curve analysis. The area under the curve (AUC) of the nomogram was 0.951 (95% CI, 0.914–0.988) in the training cohort and 0.927 (95% CI, 0.877–0.978) in the internal validation cohort. In the external validation cohort, the AUC was 0.913 (95% CI, 0.870–0.955). The Brier scores were 0.029 and 0.045 for the internal and external validation cohorts, respectively. DCA further demonstrated the favorable clinical utility of the nomogram. A clinically practical prognostic model incorporating routine clinical and pathological features was developed to estimate 5-year renal survival in patients with IgAN. Specifically designed for primary care settings, the model leverages easily accessible data to enable early identification of high-risk individuals and support personalized long-term management. Its simplicity and applicability in resource-limited environments make it a valuable tool for improving outcomes beyond specialist centers.

**Keywords** IgA nephropathy, Machine learning, Initial diagnosis, Follow-up care

IgA nephropathy (IgAN) is the most prevalent primary glomerulonephritis (GN) globally. According to a global risk map derived from genome-wide association studies, the eastern region of China is identified as a high-prevalence area for IgAN<sup>1</sup>. The clinical progression of IgAN is highly variable: while some patients maintain stable renal function for years, others progress rapidly to end-stage renal disease (ESRD). Reported 10-year renal survival rates range from 57 to 91%<sup>2</sup>. Due to the recurrent nature of renal inflammation and the risk of progressive kidney function deterioration, early identification of high-risk patients and the implementation of personalized management strategies are crucial for improving long-term outcomes.

<sup>1</sup>Department of Nephrology, Hangzhou TCM Hospital Affiliated to Zhejiang Chinese Medical University, Hangzhou 310003, Zhejiang, People's Republic of China. <sup>2</sup>Department of Nephrology, Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine, Wenzhou 325088, Zhejiang, People's Republic of China. <sup>3</sup>The Second School of Clinical Medicine, Hangzhou Normal University, Hangzhou, China. <sup>4</sup>Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China. <sup>5</sup>School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, China. <sup>6</sup>The First Clinical College, Zhejiang Chinese Medical University, Hangzhou, China. <sup>7</sup>Zhejiang Kangjing Hospital, Hangzhou, China. <sup>8</sup>Heart Center, Department of Geriatrics, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou 310003, Zhejiang, People's Republic of China. <sup>9</sup>Jinjin Wang, Tianmu Chen and Yan Fu: contributed equally to this work. <sup>10</sup>Tianmu Chen and Yan Fu are recognized as co-first authors. ✉email: wzzxylinyongqiang@126.com; jqy119@163.com; zhuqinfeifei@126.com

Currently, prognostic risk assessment in IgAN primarily relies on traditional models such as the Oxford classification and the International IgAN Prognostic Score. Previous studies have often utilized conventional logistic regression models, which typically suffer from limited sample sizes and fail to adequately address issues like missing or imbalanced clinical data, thus restricting the development of effective predictive tools. In recent years, machine learning (ML) has emerged as a powerful tool in data mining and has rapidly gained traction in the medical field due to its strong computational capabilities, allowing for the exploration of complex relationships between multiple variables and disease outcomes. For instance, Ryunosuke Noda et al. developed a noninvasive diagnostic prediction model for IgAN using ML techniques<sup>3</sup>, and Ying Zhang et al. constructed a short-term risk prediction model for chronic kidney disease (CKD) progression in patients with IgAN based on clinical data from 443 individuals<sup>4</sup>. While these studies highlight the potential of ML in IgAN diagnosis and short-term risk prediction, an integrated long-term prognostic model incorporating clinical data, laboratory parameters, and renal histopathologic features has yet to be developed. This gap limits the clinical applicability of ML approaches in the long-term management and personalized treatment of patients with IgAN.

This study aims to develop and validate a 5-year survival prediction model for IgAN using ML approaches, integrating clinical characteristics, laboratory findings, and renal histopathologic features. The objective is to provide accurate long-term outcome predictions for patients with IgAN, thereby improving risk stratification, optimizing individualized treatment strategies, and enhancing the clinical translation of these models into practice.

## Methods

### Study participants

This retrospective study utilized a follow-up dataset of 1,135 patients diagnosed with IgAN who underwent long-term follow-up at Hangzhou Hospital of Traditional Chinese Medicine from January 2014 to December 2017. An additional dataset of 352 patients diagnosed with IgAN, who underwent long-term follow-up at Hangzhou Hospital of Traditional Chinese Medicine, Zhejiang Provincial People's Hospital and Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine between February 2015 and December 2020, was also included. This study received approval from the ethics committee (approval number: 2021KY045). All IgAN diagnoses were confirmed based on histological and immunofluorescence studies of renal biopsies. The inclusion criteria were as follows: 1) age  $\geq 18$  years; 2) eGFR  $\geq 15$  mL/min/1.73 m<sup>2</sup> at baseline; 3) follow-up duration  $\geq 6$  months; and 4) at least 8 glomeruli. Exclusion criteria were as follows: 1) patients with comorbid chronic kidney diseases; 2) patients with secondary causes of mesangial IgA deposits; 3) patients with acute renal failure; and 4) patients with incomplete or ambiguous data. Ultimately, 723 patients were included in the study. The study flowchart is presented in Fig. 1.

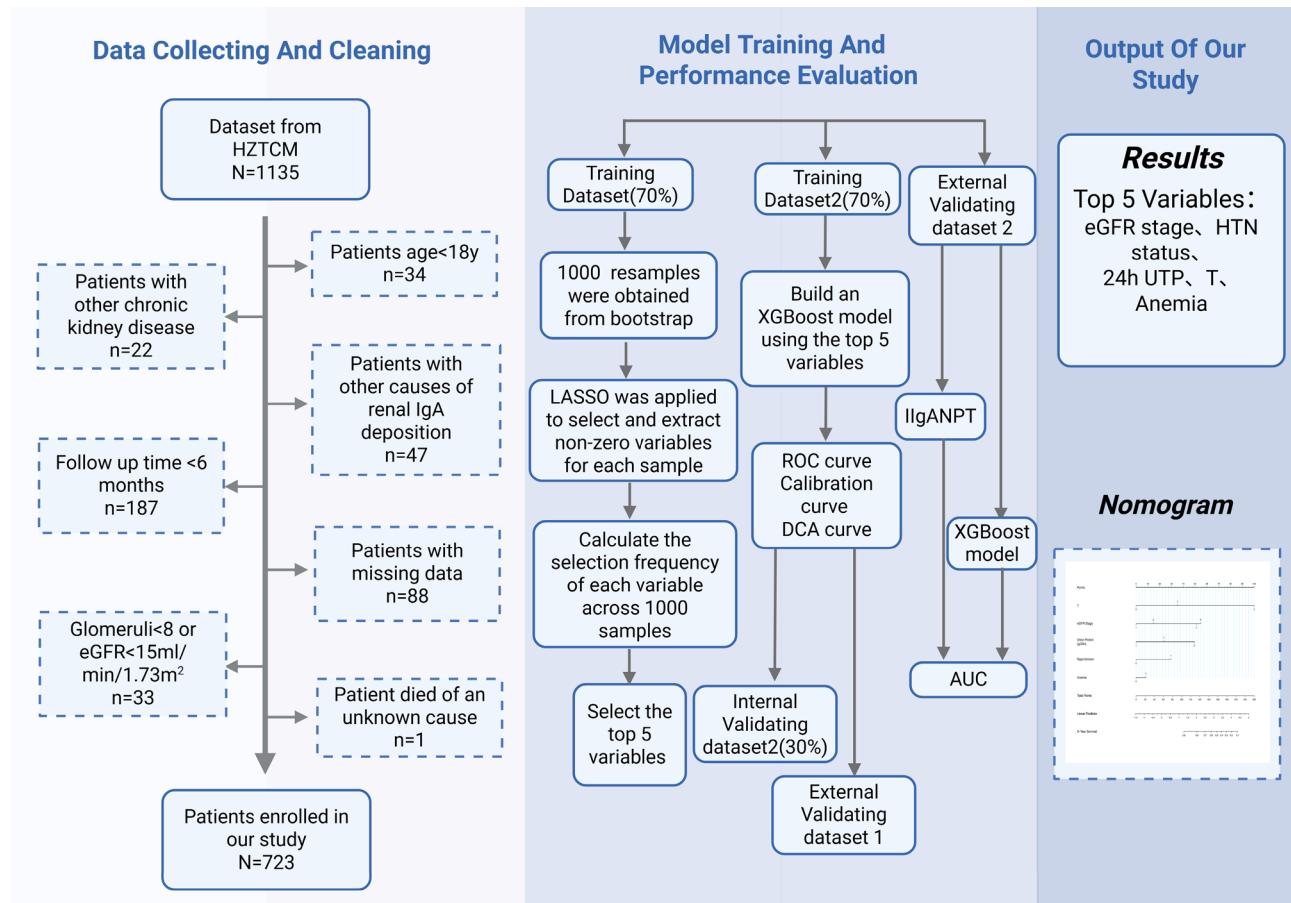
### Clinical and laboratory data

Clinical characteristics were collected at the time of renal biopsy (baseline). Data were retrieved from the medical record system of Hangzhou Hospital of Traditional Chinese Medicine and included variables such as renal biopsy time, gender, age, body mass index (BMI), hemoglobin (Hb), serum albumin (SA), total cholesterol (TC), low-density lipoprotein (LDL), high-density lipoprotein (HDL), blood uric acid (BUA), 24-h urinary total protein (24 h UTP), microscopic hematuria, estimated glomerular filtration rate (eGFR), serum complement C3 (C3), serum complement C4 (C4), and serum immunoglobulins IgA, IgG, and IgM. The baseline clinical characteristics are summarized in Table 1. Anemia was defined according to standard clinical criteria as Hb  $< 120$  g/L in males and  $< 110$  g/L in females<sup>5</sup>. The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula, and eGFR stages were classified as follows: G5 ( $< 15$  mL/min/1.73 m<sup>2</sup>), G4 (15–29 mL/min/1.73 m<sup>2</sup>), G3 (30–59 mL/min/1.73 m<sup>2</sup>), G2 (60–89 mL/min/1.73 m<sup>2</sup>), and G1 ( $\geq 90$  mL/min/1.73 m<sup>2</sup>). All indicators were defined according to traditional cutoff values, following the Kidney Disease Improving Global Outcomes (KDIGO) guidelines<sup>6</sup>. Additional details of each variable's definition are presented in Table 1.

### Tissue pathology data

Histopathological data were collected by the Nephrology Department of Hangzhou TCM Hospital, Zhejiang Chinese Medical University. Routine staining of renal biopsy tissue included hematoxylin and eosin, periodic acid-Schiff, hexamine silver, and Masson staining. Direct immunofluorescence was performed to assess the expression intensities and renal deposition sites of IgA, IgG, IgM, C3, C4, and C1q (Fig. 2).

For each patient, pathological indicators were scored using the MEST-C classification. Electron microscopy images were used to evaluate foot process effacement (FPE). The MEST-C score characterizes tissue lesions as follows: M for mesangial, E for endocapillary, S for glomerular sclerosis, T for tubulointerstitial damage, and C for crescents. Mesangial lesions were classified as M0 (absence) or M1 (presence) based on the proportion of mesangial cells. Endocapillary and glomerular sclerosis lesions were categorized as E0/E1 and S0/S1, respectively, depending on their presence or absence. Crescents were defined as extravascular proliferation of capillaries of any size exceeding two cell layers. Renal tubular atrophy/interstitial fibrosis and crescent lesions were graded as T0 (0–25%), T1 (26–50%), or T2 ( $> 50\%$ ) and C0 (none), C1 (0–25%), or C2 ( $\geq 25\%$ ), respectively<sup>7</sup>. FPE was defined by the disappearance of the normal fissure diaphragm between adjacent foot processes or complete effacement. The extent of FPE was expressed as the percentage of the length of the FPE relative to the total length of each capillary loop; FPE  $\geq 50\%$  was classified as widespread FPE, while FPE  $< 50\%$  was considered segmental FPE<sup>8,9</sup>. Renal pathology inflammatory cell infiltration (RPICI) primarily consisted of lymphocytes, monocytes, and plasma cells, with neutrophils and eosinophils being less frequent. The degree of inflammatory cell infiltration was classified into four categories: no infiltration,  $< 25\%$ ,  $> 25\%$  but  $< 50\%$ , and  $\geq 50\%$  inflammatory



**Fig. 1.** The flowchart of our study.

cell infiltration<sup>7</sup>. All biopsy specimens were reviewed by two renal pathologists who were blinded to the clinical data.

#### Endpoint and and follow-up

A composite endpoint was defined as the development of ESRD (eGFR < 15 mL/min/1.73 m<sup>2</sup>, dialysis, or kidney transplantation), a doubling of serum creatinine (Scr) levels, or a 40% decline in eGFR within five years after the diagnostic renal biopsy. Information on Scr, eGFR, total follow-up duration, and whether the patient reached the renal composite endpoint within five years was recorded during each follow-up. In cases where information was unavailable, follow-up data were obtained via phone inquiry. To account for the potential competing risk of death occurring prior to renal endpoints, the single patient who died of an unknown cause was excluded from the analysis. This approach was undertaken to minimize bias arising from competing events during model development and evaluation, thereby preserving the integrity of the study cohort and ensuring the robustness of the predictive model.

#### Statistical analysis

The overall proportion of missing values per variable ranged from 0.2 to 10.6%, with BMI missing in 3.1% of cases, Hb, LDL, and renal C1q in 0.2%, SA in 0.4%, TC in 0.6%, 24 h UTP in 0.8%, MH and renal C4 in 1.2%, serum IgG, IgA, and IgM in 2.3%, serum C3 and C4 in 2.9%, and FPE in 10.6%; overall, no variable exceeded 15% missingness. Missing data were imputed using multiple imputation by chained equations (MICE) implemented via the mi package in R (version 4.4.1), ensuring sample representativeness and improving result accuracy. Each variable was imputed using a method appropriate to its type: 1) continuous variables: predictive mean matching (PMM); 2) categorical variables: multinomial logistic regression; 3) ordinal variables: proportional odds model. We generated five imputed datasets, each subjected to model fitting, and pooled estimates were obtained according to Rubin's rules to ensure robust inference. Given the single-center origin of the dataset, stringent internal validation protocols were implemented to reduce overfitting and enhance model generalizability. The cohort was randomly split into training (70%) and independent internal validation (30%) sets.

In the training set, bootstrap resampling generated 1000 replicates, each undergoing tenfold cross-validation to identify the optimal regularization parameter ( $\lambda$ ) that minimized cross-validation error. LASSO Cox proportional hazards models were then applied to each bootstrap sample at the selected  $\lambda$  threshold, recording variables with non-zero coefficients. Variables were ranked based on their selection frequency across iterations.

Characteristics	HZSZZY				External Validation1	External Validation2		
	Overall	Outcome						
		Non-endpoint event group	Endpoint event group	p				
Patients (n)	723	679	44		352	352		
Age (mean±SD)	35 [29,44]	35 [29,44]	42.5 [30,49]	0.024*	37 [30,47]	37 [30,47]		
Gender = Male (%)	304 (42)	284 (41.8)	20 (45.5)	0.637	156 (44.3)	156 (44.3)		
BMI (mean±SD) Kg/m <sup>2</sup>	22.66 [20.40,24.81]	22.68 [20.45,24.81]	21.53 [18.92,25.45]	0.086	23.45 [21.23,25.62]	23.45 [21.23,25.62]		
Hb (mean±SD) g/L	120 [109,133]	121 [110,133]	110 [100,122.75]	<0.001*	124 [111,136.75]	124 [111,136.75]		
SA (mean±SD) g/L	37.9 [35.3,40.5]	38 [35.5,40.7]	34.45 [30.43,38.38]	<0.001*	37.6 [34.73,40.6]	37.6 [34.73,40.6]		
TC (mean±SD) mmol/L	4.41 [3.79,5.10]	4.4 [3.77,5.07]	4.63 [4.01,5.71]	0.053	4.44 [3.97,5.18]	4.44 [3.97,5.18]		
LDL (mean±SD) mmol/L	2.59 [2.07,3.14]	2.57 [2.06,3.10]	2.92 [2.39,3.60]	0.006*	2.68 [2.29,3.22]	2.68 [2.29,3.22]		
HDL (mean±SD) mmol/L	1.08 [0.91,1.29]	1.08 [0.91,1.30]	1.09 [0.93,1.25]	0.815	0.96 [0.85,1.14]	0.96 [0.85,1.14]		
BUA (mean±SD) µmol/L	306 [250,367]	304 [248,364]	365 [308.5,410]	<0.001*	305 [244,365]	305 [244,365]		
HTN=1 (%)	272 (37.6)	234 (34.5)	38 (86.4)	<0.001*	147 (41.8)	147 (41.8)		
24 h UTP (mean±SD) g/24 h	0.9 [0.46,1.74]	0.85 [0.45,1.57]	2.40 [1.25,3.77]	<0.001*	0.88 [0.46,1.8]	0.88 [0.46,1.8]		
Microscopic Hematuria (%)				0.141				
0	180 (24.9)	165 (24.3)	15 (34.1)		83 (23.6)	83 (23.6)		
1	106 (14.7)	99 (14.6)	7 (15.9)		46 (13.1)	46 (13.1)		
2	134 (18.5)	129 (19)	5 (11.4)		68 (19.3)	68 (19.3)		
3	248 (34.3)	231 (34)	17 (38.6)		117 (33.2)	117 (33.2)		
4	55 (7.6)	55 (8.1)	0 (0)		38 (10.8)	38 (10.8)		
eGFR Stage				<0.001*				
G1	363 (50.2)	359 (52.9)	4 (9.1)		196 (55.7)	196 (55.7)		
G2	214 (29.6)	208 (30.6)	6 (13.6)		83 (23.6)	83 (23.6)		
G3	121 (16.7)	100 (14.7)	21 (47.7)		59 (16.8)	59 (16.8)		
G4	25 (3.5)	12 (1.8)	13 (29.5)		14 (4)	14 (4)		
Serum IgG (mean±SD) mg/dl	1060 [905,1250]	1060 [905,1250]	1033.5 [864.25,1215]	0.372	1110 [950.25,1250]	1110 [950.25,1250]		
Serum IgA (mean±SD) mg/dl	295 [235,358]	295 [235,359]	289 [243.5,350.75]	0.866	298 [227.25,375]	298 [227.25,375]		
Serum IgM (mean±SD) mg/dl	108 [80,144]	107 [79,144]	116 [84.25,145.75]	0.501	101 [72,140.75]	101 [72,140.75]		
Serum C3 (mean±SD) mg/dl	97 [86,110]	97 [86,111]	90 [78,107.5]	0.052	97 [85,110]	97 [85,110]		
Serum C4 (mean±SD) mg/dl	23 [19, 27]	23 [19, 27]	24 [19,26.75]	0.854	23 [18, 28]	23 [18, 28]		
Renal IgA deposition (%)				0.722				
0	0 (0)	0 (0)	0 (0)		0 (0)	0 (0)		
1	12 (1.7)	11 (1.6)	1 (2.3)		3 (0.9)	3 (0.9)		
2	135 (18.7)	130 (19.1)	5 (11.4)		54 (15.3)	54 (15.3)		
3	497 (68.7)	462 (68)	35 (79.5)		266 (75.6)	266 (75.6)		
4	79 (10.9)	76 (11.2)	3 (6.8)		29 (8.2)	29 (8.2)		
Renal IgG deposition (%)				0.094				
0	590 (81.6)	558 (82.2)	32 (72.7)		246 (69.9)	246 (69.9)		
1	105 (14.5)	97 (14.3)	8 (18.2)		91 (25.9)	91 (25.9)		
2	23 (3.2)	20 (2.9)	3 (6.8)		14 (4)	14 (4)		
3	5 (0.7)	4 (0.6)	1 (2.3)		1 (0.3)	1 (0.3)		
Renal IgM deposition (%)				0.459				
0	96 (13.3)	89 (13.1)	7 (15.9)		15 (4.3)	15 (4.3)		
1	444 (61.4)	422 (62.2)	22 (50)		217 (61.6)	217 (61.6)		
2	166 (23)	152 (22.4)	14 (31.8)		117 (33.2)	117 (33.2)		
3	17 (2.4)	16 (2.4)	1 (2.3)		3 (0.9)	3 (0.9)		
Renal C3 deposition (%)				0.392				
0	30 (4.1)	29 (4.3)	1 (2.3)		16 (4.5)	16 (4.5)		

Continued

Characteristics	HZSZZY				External Validation1	External Validation2		
	Overall	Outcome						
		Non-endpoint event group	Endpoint event group	p				
1	83 (11.5)	75 (11)	8 (18.2)		40 (11.4)	40 (11.4)		
2	294 (40.7)	276 (40.6)	18 (40.9)		130 (36.9)	130 (36.9)		
3	313 (43.3)	296 (43.6)	17 (38.6)		165 (46.9)	165 (46.9)		
4	3 (0.4)	3 (0.4)	0 (0)		1 (0.3)	1 (0.3)		
Renal C4 deposition (%)				0.699				
0	665 (92)	625 (92)	40 (90.9)		330 (93.8)	330 (93.8)		
1	48 (6.6)	47 (6.9)	1 (2.3)		19 (5.4)	19 (5.4)		
2	8 (1.1)	7 (1)	1 (2.3)		3 (0.9)	3 (0.9)		
3	2 (0.3)	0 (0)	2 (4.5)		0 (0)	0 (0)		
Renal C1q deposition (%)				0.003*				
0	549 (75.9)	523 (77)	26 (59.1)		317 (90.1)	317 (90.1)		
1	159 (22)	146 (21.5)	13 (29.5)		32 (9.1)	32 (9.1)		
2	14 (1.9)	9 (1.3)	5 (11.4)		3 (0.9)	3 (0.9)		
3	1 (0.1)	1 (0.1)	0 (0)		0 (0)	0 (0)		
RPICI (%)				<0.001*				
0	1 (0.1)	1 (0.1)	0 (0)		2 (0.6)	2 (0.6)		
1	444 (61.4)	438 (64.5)	6 (13.6)		217 (61.6)	217 (61.6)		
2	216 (29.9)	202 (29.7)	14 (31.8)		93 (26.4)	93 (26.4)		
3	62 (8.6)	38 (5.6)	24 (54.5)		40 (11.4)	40 (11.4)		
FPE (%)				<0.001*				
0	109 (15.1)	108 (15.9)	1 (2.3)		35 (9.9)	35 (9.9)		
1	369 (51)	354 (52.1)	15 (34.1)		195 (55.4)	195 (55.4)		
2	170 (23.5)	154 (22.7)	16 (36.4)		72 (20.5)	72 (20.5)		
3	67 (9.3)	56 (8.2)	11 (25)		45 (12.8)	45 (12.8)		
4	8 (1.1)	7 (1)	1 (2.3)		5 (1.4)	5 (1.4)		
M=1 (%)				1.000				
0	3 (0.4)	3 (0.4)	0 (0)		23 (6.5)	23 (6.5)		
1	720 (99.6)	676 (99.6)	44 (100)		329 (93.5)	329 (93.5)		
E=1 (%)				0.081				
0	545 (75.4)	507 (74.7)	38 (86.4)		217 (61.6)	217 (61.6)		
1	178 (24.6)	172 (25.3)	6 (13.6)		135 (38.4)	135 (38.4)		
S=1 (%)				0.185				
0	97 (13.4)	94 (13.8)	3 (6.8)		56 (15.9)	56 (15.9)		
1	626 (86.6)	585 (86.2)	41 (93.2)		296 (84.1)	296 (84.1)		
T (%)				<0.001*				
0	510 (70.5)	503 (74.1)	7 (15.9)		249 (70.7)	249 (70.7)		
1	173 (23.9)	159 (23.4)	14 (31.8)		79 (22.4)	79 (22.4)		
2	40 (5.5)	17 (2.5)	23 (52.3)		24 (6.8)	24 (6.8)		

Continued

Characteristics	HZSZZY				External Validation1	External Validation2		
	Overall	Outcome						
		Non-endpoint event group	Endpoint event group	p				
C (%)				0.077				
0	258 (35.7)	235 (34.6)	23 (52.3)		114 (32.4)	114 (32.4)		
1	406 (56.2)	390 (57.4)	16 (36.4)		209 (59.4)	209 (59.4)		
2	59 (8.2)	54 (8)	5 (11.4)		29 (8.2)	29 (8.2)		
EndEvent=1 (%)	44 (6.1)	0 (0)	44 (100)		20 (5.7)	18 (5.1)		

**Table 1.** Baseline characteristics of clinical variables for patients with IgAN enrolled in our study at the time of renal biopsy. 1. Outcome was defined as: combined event of ESRD (eGFR<15 mL/min/1.73m<sup>2</sup>, dialysis, or kidney transplantation) or Scr level doubled or a 40% decline in eGFR within 5 years after diagnostic renal biopsy. 2. p-value refers to the p-values from statistical tests comparing patients with and without outcome (Chi-square test for categorical variables, and t-test for continuous variables).\* Abbreviations: BMI, Body Mass Index; Hb, Hemoglobin; SA, Serum Albumin; TC, Total Cholesterol; LDL, Low-Density Lipoprotein; HDL, High-Density Lipoprotein; BUA, Blood Uric Acid; 5y Scr, 5-year Serum creatinine; 24-h UTP, 24-h Urinary Total Protein; IgA, Immunoglobulin A; IgG, Immunoglobulin G; IgM, Immunoglobulin M; Serum C3:, Serum complement C3; Serum C4, Serum complement C4; MH, Microscopic Hematuria, MH0(without MH), MH1(+), MH2(++)+, MH3(++)+, MH4(++)+, Renal IgA deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition, 4(++) deposition; Renal IgG deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition; Renal IgM deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition; Renal C3 deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition, 4(++) deposition; Renal C4 deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition; Renal C1q deposition, 0(without deposition), 1(+ deposition), 2(++) deposition, 3(++) deposition; eGFR, estimated Glomerular Filtration Rate (caculated by EPI formula), G4(<30 mL/min/1.73 m<sup>2</sup>), G3 (30-59 mL/min/1.73 m<sup>2</sup>), G2 (60-89 mL/min/1.73 m<sup>2</sup>), G1 (≥90 mL/min/1.73 m<sup>2</sup>); HTN, Hypertension; RPICI, Renal Pathology Inflammatory Cell Infiltration, RPICI0(no inflammatory cell infiltration), RPICI1(<25%), RPICI2(25-50%), RPICI3(>50%); FPE, Foot Process Effacement, FPE0(<15%), FPE1(15-50%), FPE2(50-75%), FPE3(75-90%), FPE4(>90%); M, Mesangial, M0(≤50% of glomeruli have mesangial cell proliferation) M1(>50% of glomeruli have mesangial cell proliferation); E, Endocapillary, E0(No capillary endothelial cell proliferation), E1(Capillary endothelial cell proliferation is present); S, Glomerular sclerosis, S0(No segmental glomerulosclerosis), S1(Presence of segmental glomerulosclerosis); T, Tubulointerstitialfibrosis, T0 (Tubular atrophy and interstitial fibrosis≤25%), T1 (Tubular atrophy and interstitial fibrosis26-50%), T2 (Tubular atrophy and interstitial fibrosis>50%); C, Crescents, C0 (none), C1 (0-25% of glomeruli with crescent formation), C2 (≥25% of glomeruli with crescent formation).

The top five features were analyzed for correlation and collinearity, with results visualized in Fig. 3 using SHAP plots. These features informed the XGBoost model for predicting 5-year IgAN survival. Model performance was evaluated based on discrimination (AUC), calibration (Brier scores and calibration curve), and clinical utility (decision curve analysis [DCA]). External validation was performed using datasets from Zhejiang Provincial People's Hospital and Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine to assess clinical applicability, with AUC values confirming model promising predictive ability. Two validation analyses were performed: the first used an endpoint defined as the development of ESRD (eGFR<15 mL/min/1.73 m<sup>2</sup>, initiation of dialysis, or kidney transplantation), a doubling of serum creatinine (Scr) levels, or a 40% decline in eGFR within five years after the diagnostic renal biopsy; the second used an endpoint defined as the development of ESRD (eGFR<15 mL/min/1.73 m<sup>2</sup>, initiation of dialysis, or kidney transplantation), a doubling of Scr levels, or a 50% decline in eGFR within five years after the diagnostic renal biopsy to allow comparison with the International IgA Nephropathy Prediction Tool (IgANPT). Model discrimination was assessed using the AUC.

Continuous variables are presented as mean ± SD and compared using t-tests (for normally distributed data) or Kruskal-Wallis tests (for non-normally distributed data). Categorical variables are expressed as frequencies and compared using chi-square tests. Statistical significance was set at  $P<0.05$ . Analyses were performed using SPSS (v29.0, Mac) and R (v4.4.1).

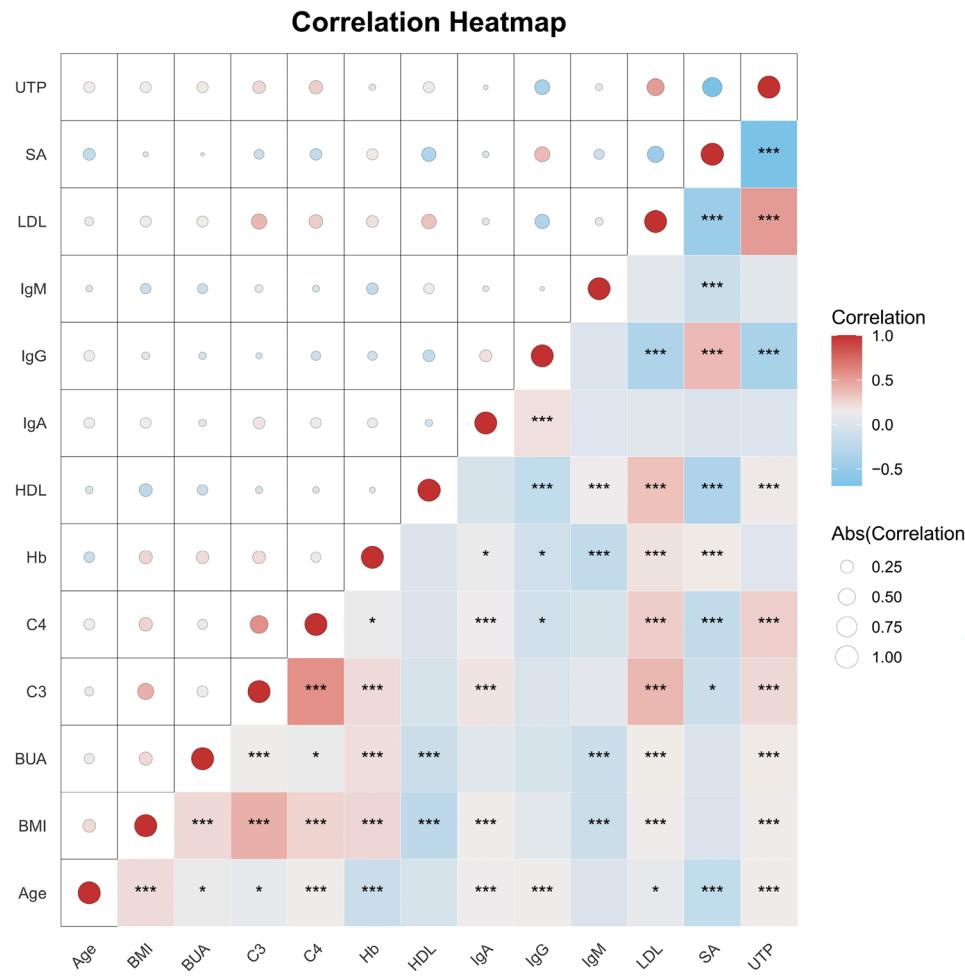
## Methods statements

All methods were performed in accordance with the relevant guidelines and regulations.

## Results

### Characteristics of the clinical variables of studied participants

Each patient's baseline was defined as the time of renal biopsy, which ranged from January 2014 to December 2017 in the dataset. The total follow-up period was 5 years following the renal biopsy. Within this cohort, 44



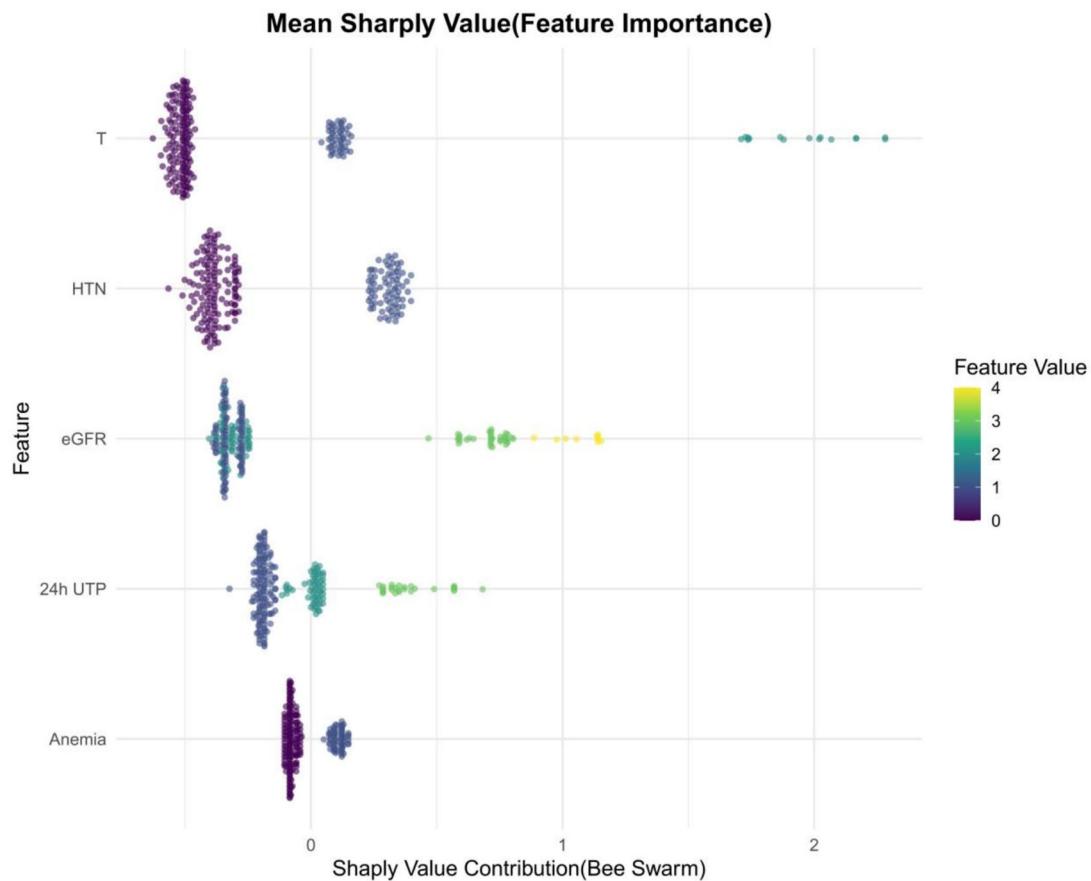
**Fig. 2.** Correlation heat map of continuous variables, color depth represents correlation coefficient (r value), \*  $< 0.05$ ; \*\*\*  $< 0.001$ .

patients (6.1%) reached the endpoint event within 5 years after biopsy. Based on the occurrence of endpoint events, the 723 patients were divided into two groups: the endpoint events group and the non-endpoint events group. Thirty-one candidate predictors were considered, including demographics, clinical characteristics, and pathological features. Overall, patients undergoing renal biopsy were middle-aged, with a median age of 35 years, and predominantly female (58.0%). The median BMI was 22.66 kg/m<sup>2</sup>. Specifically, patients in the endpoint events group had higher values for age (mean 42.5 vs. 35 years), LDL (mean 2.92 mmol/L vs. 2.57 mmol/L), BUA (mean 365  $\mu$ mol/L vs. 304  $\mu$ mol/L), 24 h UTP (mean 2.40 g/24 h vs. 0.85 g/24 h), and HTN (86.4% vs. 34.5%) and lower Hb (mean 110 g/L vs. 121 g/L) and SA (mean 34.45 g/L vs. 38 g/L) compared to those without endpoint events. These differences were statistically significant ( $P < 0.05$ ). Notably, patients with lower baseline eGFR stages were more likely to progress to end-stage kidney disease (ESKD). Specifically, 1.8% of patients in Group 1 (eGFR stage G4) progressed to ESKD, compared to 29.5% in Group 2. Similar trends were observed for renal C1q deposition, RPICI, FPE, and T. In addition, several predictors, including gender, BMI, TC, HDL, MH, serum immunoglobulins (IgA, IgG, IgM), serum C3, serum C4, and renal IgA, IgG, IgM, C3, and C4 deposition, as well as the M, E, S, and C scores, showed no statistically significant differences between the two groups at baseline. The clinical characteristics of the endpoint event group and non-endpoint event group from the Hangzhou TCM Hospital of Zhejiang Chinese Medical University are summarized in Table 1.

The external validation cohort, comprising patients from Hangzhou Hospital of Traditional Chinese Medicine, Zhejiang Provincial People's Hospital and Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine, had follow-up durations and variables consistent with the training and internal validation cohorts. The external cohort was predominantly composed of middle-aged individuals, with a median age of 37 years, and a majority of females (55.7%). The median BMI was 23.45 kg/m<sup>2</sup>, and 20 patients (5.7%) reached the endpoint event. The characteristics of the external validation cohort are presented in Table 1.

### Model performance

A correlation analysis was performed on the included variables, revealing strong correlations between TC and LDL, SA and 24 h UTP, as well as RPICI and T. Based on prior research findings, TC, SA, and RPICI were excluded from further analysis. A correlation heat map of continuous variables (Fig. 2) showed significant



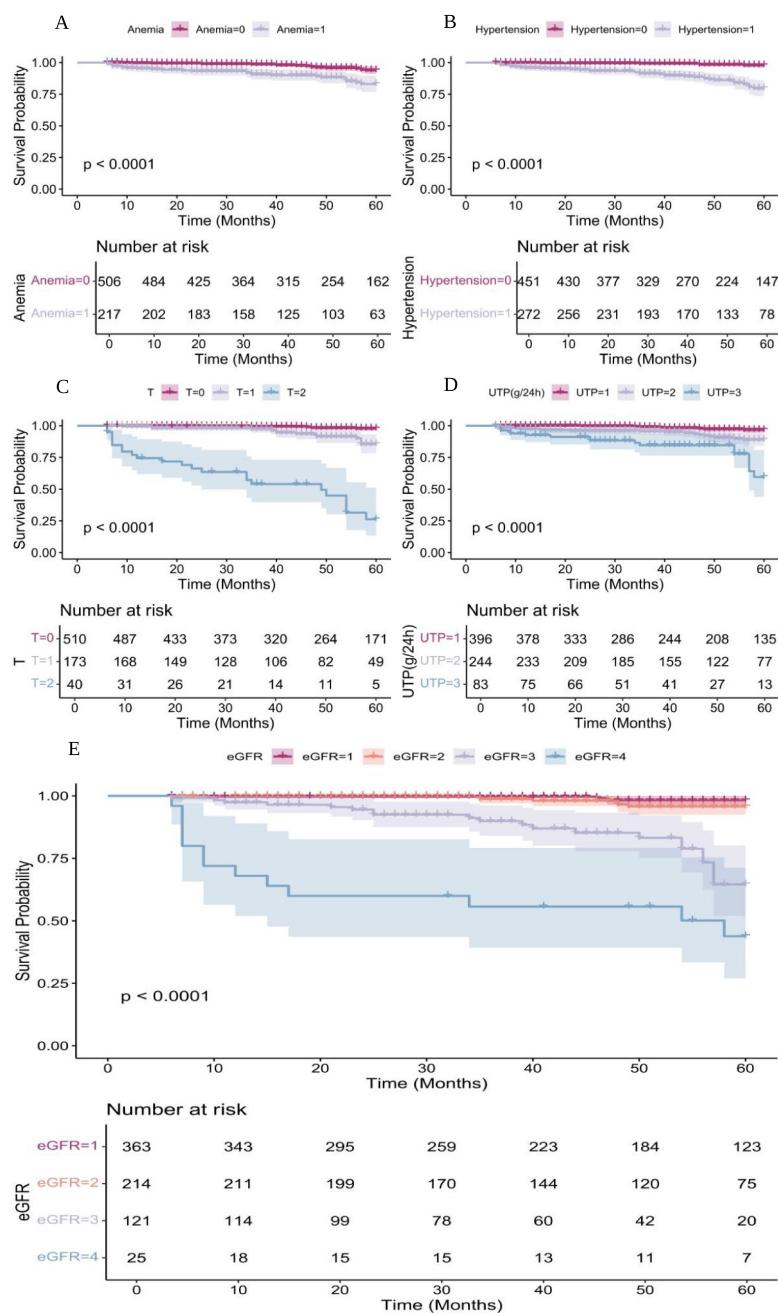
**Fig. 3.** SHAP figure of variables selected by LASSO. Abbreviations: SHAP, SHapley Additive exPlanations; LASSO, Least Absolute Shrinkage and Selection Operator; HTN, hypertension; 24 h UTP, 24-h Urinary Total Protein; eGFR, estimated Glomerular Filtration Rate (calculated by EPI formula).

correlations among several variables, with color depth representing the correlation coefficient. To address the potential impact of multicollinearity on the model, a LASSO Cox proportional hazards model was employed for variable selection. The top five selected variables were further analyzed for correlation and collinearity, and the results indicated no significant multicollinearity among them. These selected variables—Hb, 24 h UTP, HTN, eGFR, and Oxford classification T—were used to construct the XGBoost model for predicting the 5-year survival rate of patients with IgAN. Figure 4 illustrates the Kaplan–Meier curve showing the association of these variables with endpoint events. In the training set, the AUC of the XGBoost model was 0.951 (95% CI: 0.914–0.988), and in the internal validation set, it was 0.927 (95% CI: 0.877–0.978), indicating that the model did not exhibit overfitting. The Brier scores for the internal and external validation cohorts were 0.029 and 0.045, respectively. Both the Brier score and DCA confirmed the model's good calibration and clinical applicability. External validation using data from patients with IgAN at Hangzhou Hospital of Traditional Chinese Medicine, Zhejiang Provincial People's Hospital and Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine showed an AUC of 0.913 (95% CI: 0.870–0.955), suggesting that the model demonstrates promising predictive performance and potential clinical utility. A more detailed presentation of the model's performance is provided in Table 2, with the ROC curve and DCA displayed in Fig. 5.

Using the second endpoint definition—which included the development of ESRD, a doubling of serum creatinine levels, or a  $\geq 50\%$  decline in eGFR within five years—the external validation demonstrated that our XGBoost model achieved an AUC of 0.915 (95% CI: 0.870–0.961), whereas the IgANPT showed an AUC of 0.715 (95% CI: 0.559–0.872). These findings indicate that our model exhibited superior discriminative performance compared with the IgANPT in the external cohort.

#### Constructing a nomogram

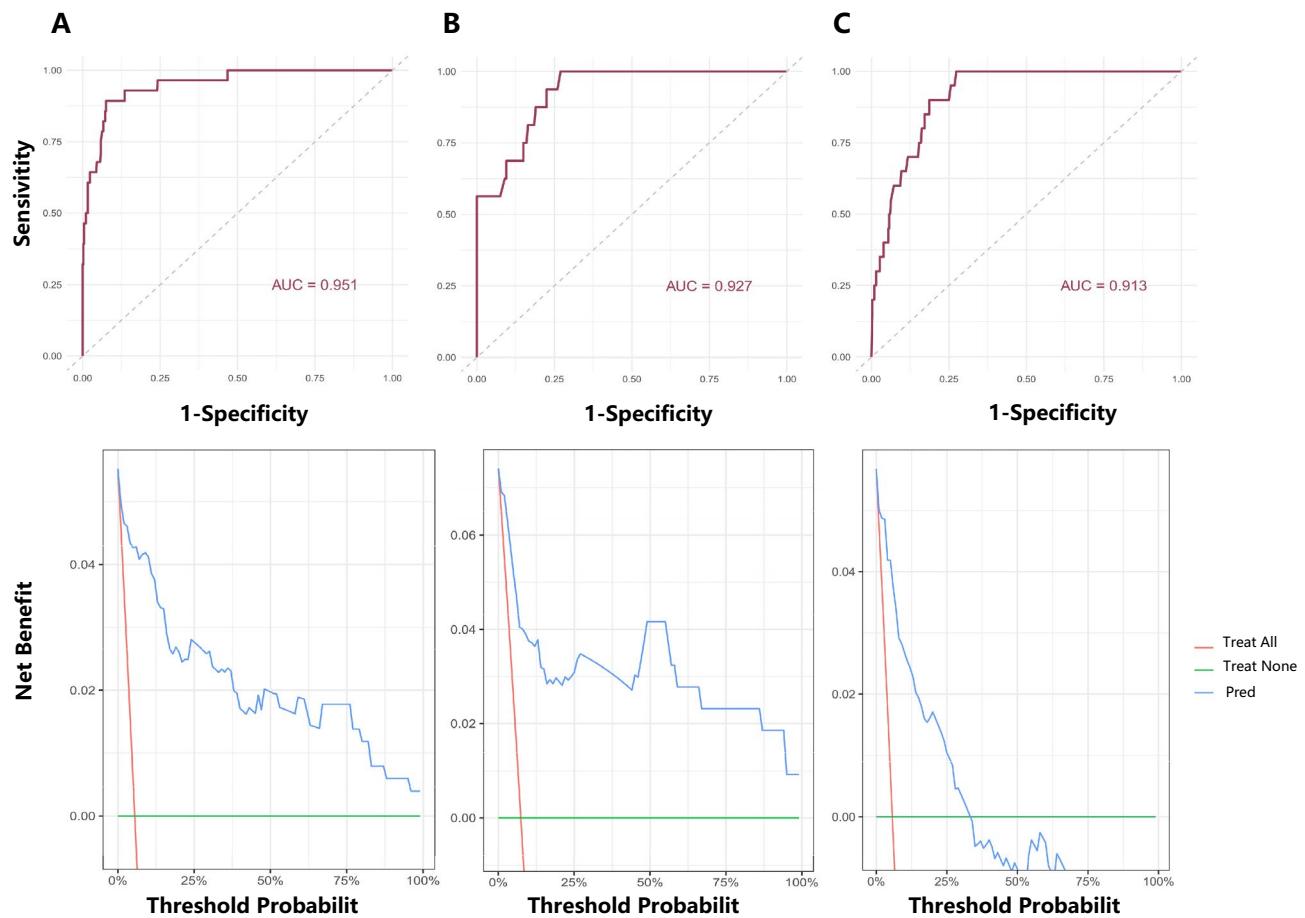
Based on the variables selected by LASSO—Hb, 24 h UTP, HTN, eGFR, and Oxford classification T—a nomogram was developed to predict the 5-year renal survival probability for patients with IgAN. The nomogram calculates the total score by projecting the values of each variable onto the points axis and summing the scores. The total score is then mapped to the risk axis to estimate the patient's 5-year survival probability (Fig. 6). For example, for a patient with IgAN exhibiting 24 h UTP of 4.51 g/24 h, Oxford classification T=2, HTN=1, eGFR=G4, and no anemia, the respective scores are approximately 43.5, 100, 52.6, 47.5, and 0. The total score is 243.6, which corresponds to a 5-year survival rate of about 13%, indicating a high risk of progression to ESRD.



**Fig.4.** KM survival curve of clinical variables selected by LASSO. (A) KM survival curve of anemia status. (B) KM survival curve of hypertension. (C) KM survival curve of T. (D) KM survival curve of 24 h UTP. (E) KM survival curve of eGFR stage. Abbreviations: KM, Kaplan-Meier Curve; LASSO, Least Absolute Shrinkage and Selection Operator; T, T in Oxford classification; eGFR stage, estimated Glomerular Filtration Rate, according to the KDIGO CKD staging, 1-G1, 2-G2, 3-G3, 4-G4; Urine protein, 1-urine protein < 1 g/24 h, 2-1 g/24 h < = urine protein < 3 g/24 h, 3-urine protein > = 3 g/24 h

Dataset	AUC	95%CI		Brier score	p value
		Lower	Upper		
Train set	0.951	0.914	0.988	0.029	0.729
Internal validation	0.927	0.877	0.978	0.039	0.753
External validation	0.913	0.870	0.955	0.045	0.295

**Table 2.** Nomogram performance in training, internal and external validation cohorts.



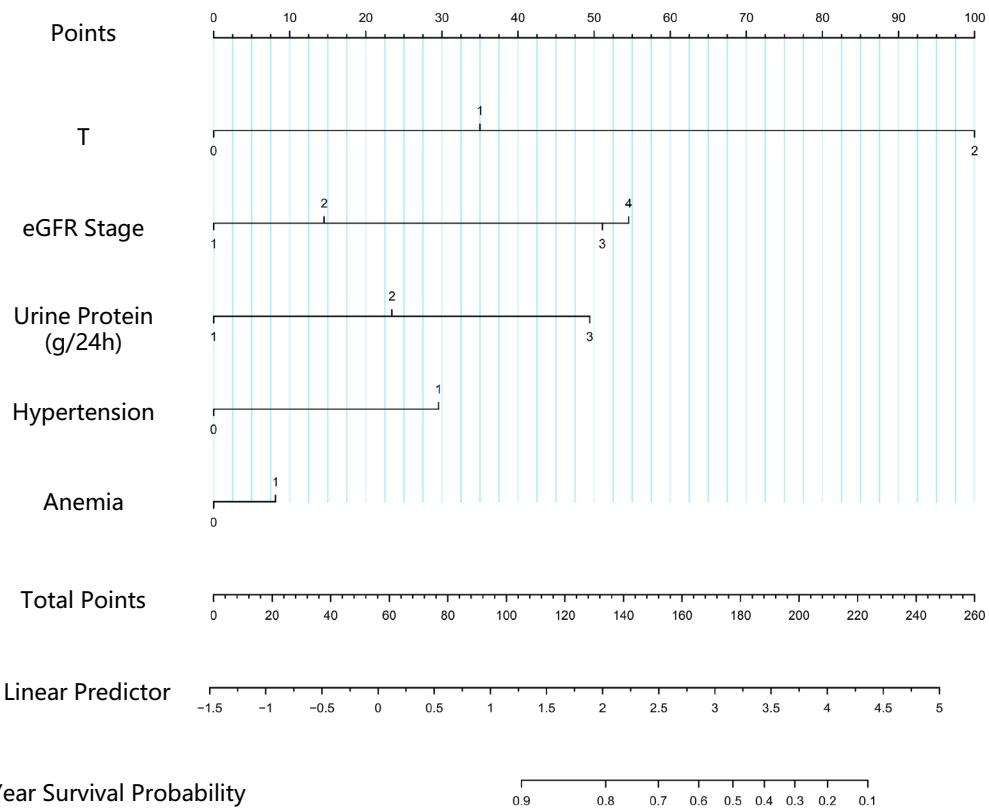
**Fig.5.** ROC curve and DCA curve of XGBoost model. (A) ROC curve and DCA curve of XGBoost model on the training set. (B) ROC curve and DCA curve of XGBoost model on the internal validation set. (C) ROC curve and DCA curve of XGBoost model on the external validation set. Abbreviations: DCA, decision curve analysis; XGBoost, Extreme Gradient Boosting.

This suggests that the patient warrants close clinical monitoring. Clinicians should comprehensively assess the patient's condition, develop an individualized treatment plan, and regularly monitor renal function and 24 h UTP.

## Discussion

This study developed an XGBoost model and a corresponding nomogram to predict the 5-year renal survival probability in patients with IgAN. The nomogram incorporated five readily available variables (anemia status, eGFR stage, HTN status, Oxford classification T score, and 24 h UTP) and demonstrated strong discriminatory performance and calibration in both internal and external validation cohorts. As a user-friendly tool, the nomogram enables primary care physicians to easily interpret the contribution and weight of each risk factor, thus facilitating more accurate and personalized risk assessment for patients with IgAN.

The IIgANPT is a widely used model that integrates clinical parameters and the Oxford MEST score to estimate the risk of a 50% decline in eGFR or progression to ESRD within 60 months. It includes two versions: one with and one without a race parameter<sup>10</sup>. In our study, we conducted two external validation analyses to compare the performance of our XGBoost model with the IIgANPT. Using the endpoint consistent with model development (ESRD, doubling of serum creatinine, or a  $\geq 40\%$  decline in eGFR), our model demonstrated robust predictive performance. To enable direct comparison with the IIgANPT, we redefined the external cohort endpoint as a  $\geq 50\%$  decline in eGFR, under which our XGBoost model achieved a substantially higher AUC of 0.915, indicating superior discriminative ability. Moreover, adopting a  $\geq 40\%$  eGFR decline as the primary endpoint allows our model to identify high-risk patients earlier than the IIgANPT, providing a longer window for clinical intervention and increasing sensitivity to early disease progression. Several other models have also been proposed. For instance, Zhang-Yu Tian et al. developed a prediction model based on five variables ( $n=519$ )-LDL, Oxford S and T scores, cardiovascular disease, and 24 h UTP-to estimate the risk of initiating renal replacement therapy or reaching an  $eGFR < 15 \text{ mL} \cdot \text{min}^{-1} \cdot (1.73 \text{ m}^2)^{-1}$  at 5, 8, and 10 years<sup>11</sup>. Notably, this model includes only the M, E, S, and T scores of the Oxford classification, excluding the C score—a pathological feature that is closely associated with poor prognosis in Asian patients with IgAN—thereby potentially limiting the comprehensiveness and accuracy of the model's prognostic assessment for this population.



**Fig.6.** Nomogram based on XGBoost model. The nomogram was constructed to estimate the 5-year renal survival probability in patients with IgAN. To utilize the nomogram, each predictor is located on its corresponding axis, and a vertical line is drawn upward to determine the number of points assigned. The total score is calculated by summing the individual points from all variables. This total score is then projected downward onto the probability scale to estimate the 5-year renal survival likelihood. Abbreviations: XGBoost, Extreme Gradient Boosting; T, T in Oxford classification; eGFR stage, estimated Glomerular Filtration Rate, according to the KDIGO CKD staging, 1-G1, 2-G2, 3-G3, 4-G4; Urine protein, 1-urine protein < 1 g/24 h, 2-1 g/24 h < = urine protein < 3 g/24 h, 3-urine protein > = 3 g/24 h; anemia, 0-without anemia, 1-with anemia; hypertension, 0-without hypertension, 1-with hypertension.

With the advancement of ML, Yngvar Lunde Haaskjold et al. applied a random forest approach to data from Norwegian patients ( $n = 232$ ) to identify predictive features and develop a risk prediction model. The selected predictors included T%, segmental glomerular sclerosis, normal glomeruli, global sclerotic glomeruli, segmental adherence, and perihilar glomerular sclerosis. While the model showed potential, it has not undergone rigorous external validation, and its clinical utility remains uncertain<sup>12</sup>.

To address these limitations, this study contributes novel insights to the existing literature by developing a prognostic model using multicenter data and conducting independent external validation, which confirmed the model's strong discriminatory performance. The final model incorporated anemia status, eGFR stage, HTN status, the T score from the Oxford classification, and 24 h UTP. Among these variables, the Oxford T score, eGFR stage, and 24 h UTP align with KDIGO guidelines and are widely recognized as key prognostic indicators in IgAN. Early identification of CKD stages is clinically significant for risk stratification and the development of personalized intervention strategies, potentially delaying renal function deterioration and improving long-term outcomes and quality of life. Proteinuria is the most extensively studied risk factor for IgAN progression to ESRD, and recent reports have identified proteinuria reduction as a surrogate endpoint in IgAN trials<sup>13</sup>. Our study found that proteinuria levels > 1 g/24 h are a risk factor for poor prognosis, which is consistent with KDIGO recommendations that suggest targeting proteinuria < 1 g/24 h as a reasonable therapeutic goal<sup>13</sup>. A retrospective study by Heather N. Reich et al., which included follow-up data from 542 patients with IgAN, reported that those with persistent proteinuria > 3 g/day experienced a renal function decline rate 25 times faster than patients with proteinuria < 1 g/day<sup>14</sup>. However, after adjusting for time-dependent confounding factors, Chen Tang et al. observed that patients with proteinuria  $\geq 0.5$  g/24 h were still at an increased risk of kidney failure, highlighting the importance of controlling proteinuria in IgAN management, even at lower levels<sup>15</sup>. The prognostic significance of tubulointerstitial lesions has also been well established. A retrospective cohort study and meta-analysis by Bingxin Yu et al. confirmed that the presence of T lesions is a strong predictor of poor renal outcomes in IgAN, independent of all clinical and laboratory parameters<sup>16</sup>. Our findings support

this conclusion. In the Oxford classification, the T score reflects the extent of tubulointerstitial injury. During the pathogenesis and progression of IgAN, the deposition of immune complexes—particularly IgA—in the tubulointerstitium activates local immune responses and triggers sustained inflammation, ultimately leading to interstitial fibrosis<sup>17</sup>. Therefore, a higher T score is associated with an increased risk of progression to ESRD and necessitates close clinical monitoring and targeted interventions.

In addition to the previously discussed indicators, anemia was identified as a strong and independent predictor of poor renal prognosis in patients with IgAN. Chronic kidney injury can contribute to anemia through several mechanisms, including reduced red blood cell survival, erythropoietin (EPO) deficiency, suppressed erythropoiesis, and iron dysregulation<sup>18–23</sup>. Anemia may further aggravate renal damage by inducing tissue hypoxia, potentially via activation of the hypoxia-inducible factor (HIF) signaling pathway, which plays a role in the progression of chronic kidney disease<sup>24,25</sup>. This mechanistic pathway helps explain the observed association between anemia and adverse renal outcomes in our cohort. A retrospective study by Tae Ryom Oh et al., involving 4,326 patients with biopsy-confirmed IgAN, found that for each 1.0 g/dL increase in Hb, the risk of disease progression decreased by 13%, aligning with our findings<sup>26</sup>. Renal anemia is driven by various mechanisms, including reduced red blood cell survival, EPO deficiency, suppressed erythropoiesis, and iron imbalance<sup>18–23</sup>. Among these, EPO deficiency and functional iron deficiency—both related to the HIF pathway—have attracted increasing attention for their roles in renal anemia and disease progression<sup>27</sup>. Furthermore, prior studies have documented a correlation between anemia and tubulointerstitial injury in IgAN, further underscoring its prognostic significance<sup>28,29</sup>. Our study further validates anemia as a risk factor for poor renal prognosis. Additionally, the importance of HTN in IgAN is reinforced. HTN at the time of renal biopsy has been linked to increased urinary protein excretion during follow-up, suggesting its ongoing impact on glomerular injury. Histopathological findings indicate that hypertensive patients typically exhibit more severe vascular lesions, which, compared to mesangial cell proliferation, play a more dominant role in glomerulosclerosis<sup>30</sup>. Therefore, in IgAN individuals with comorbid HTN, regular blood pressure monitoring and the use of antihypertensive medications during treatment are crucial for improving patient outcomes.

This study has several limitations. Firstly, event-per-variable (EPV) ratio of our study is relatively low, with 31 candidate predictors and 44 outcome events (EPV = 1.19), which may introduce some risk of overfitting and affect the stability of parameter estimates. In future studies, larger cohorts with more outcome events are needed to increase the EPV and improve the robustness and generalizability of the predictive model. Secondly, as a retrospective study, while multiple imputation using the “mi” package was applied to address missing data, residual bias cannot be entirely ruled out. Prospective studies with standardized and comprehensive data collection are necessary to validate these findings and minimize potential bias. Finally, this study focused solely on clinical and histopathological data, excluding genetic information. With the increasing accessibility of genomic sequencing technologies, integrating clinical and genetic data may provide additional value in improving risk prediction, diagnosis, and prognostication in IgAN, and should be explored in future research.

## Conclusion

Using real-world data and a data-driven approach, this study developed a 5-year survival prediction model for IgAN utilizing the XGBoost ML algorithm. The model incorporates five routinely available clinical variables at the time of renal biopsy and demonstrated promising predictive performance in both internal and external validation cohorts. Moreover, when compared with the widely used IIgANPT, our model achieved higher discriminative efficiency and, by adopting a ≥ 40% decline in eGFR as the primary endpoint, was able to identify high-risk patients at an earlier stage, providing a longer window for clinical intervention and greater sensitivity to early disease progression. An intuitive nomogram derived from the model enables individualized risk assessment at diagnosis, helping primary care physicians recognize high-risk individuals, optimize follow-up strategies, and guide early, personalized management. By facilitating timely risk stratification and supporting evidence-based decision-making, this model may contribute to more precise, proactive, and resource-efficient care in IgAN, ultimately improving long-term outcomes. Further multicenter, prospective studies are warranted to validate its generalizability and support its integration into routine clinical practice.

## Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Received: 5 July 2025; Accepted: 9 December 2025

Published online: 14 December 2025

## References

1. Floege, J. Primary glomerulonephritis: A review of important recent discoveries. *Kidney Res. Clin. Pract.* **32**(3), 103–110. <https://doi.org/10.1016/j.krcp.2013.06.004> (2013).
2. D’Amico, G. Natural history of idiopathic IgA nephropathy: Role of clinical and histological prognostic factors. *Am. J. Kidney Dis.* **36**(2), 227–237. <https://doi.org/10.1053/ajkd.2000.8966> (2000).
3. Noda, R., Ichikawa, D. & Shibagaki, Y. Machine learning-based diagnostic prediction of IgA nephropathy: Model development and validation study. *Sci. Rep.* **14**(1), 12426. <https://doi.org/10.1038/s41598-024-63339-7> (2024).
4. Zhang, Y., Wang, Z., Tang, W., Yuan, X. & Xie, X. Development and internal and external validation of a nomogram model for predicting the risk of chronic kidney disease progression in IgA nephropathy patients. *PeerJ* **30**(12), e18416. <https://doi.org/10.7717/peerj.18416> (2024).
5. Luo, H. L., Zhang, H. N., Wei, W., Chen, C., Tang, K. N., Cui, X. N. Application of low-dose chest CT in predicting the grading of anemia severity. *Lin Chuang Fang She Xue Za Zhi*. **44**(06):1015–1020. <https://doi.org/10.13437/cnki.jcr.2025.06.004> (2025).

6. Levey, A. S., de Jong, P. E., Coresh, J., et al. The definition, classification, and prognosis of chronic kidney disease: A KDIGO Controversies Conference report [published correction appears in *Kidney Int.* 2011 Nov;80(9):1000]. *Kidney Int.* 80(1):17–28. <https://doi.org/10.1038/ki.2010.483> (2011).
7. Trimarchi, H., Barratt, J., Cattran, D. C., et al. IgAN classification working group of the international IgA nephropathy network and the renal pathology society; conference participants. Oxford classification of IgA nephropathy 2016: an update from the IgA Nephropathy Classification Working Group. *Kidney Int.* 91(5):1014–1021. <https://doi.org/10.1016/j.kint.2017.02.003> (2017).
8. Peng, Y. C. et al. Long-term prognosis of primary focal segmental glomerulosclerosis in children. *Zhonghua Er Ke Za Zhi* 60(9), 894–900. <https://doi.org/10.3760/cma.j.cn112140-2022028-00161> (2022).
9. Zhao, Y. et al. Plasma soluble urokinase receptor level is correlated with podocytes damage in patients with IgA nephropathy. *PLoS ONE* 10(7), e0132869. <https://doi.org/10.1371/journal.pone.0132869> (2015).
10. Hu, L. et al. External Validation of the International Prognosis Prediction Model of IgA Nephropathy. *Ren. Fail.* 46(1), 2313174. <https://doi.org/10.1080/0886022X.2024.2313174> (2024).
11. Tian, Z. Y. et al. Prognostic value of low-density lipoprotein cholesterol in IgA nephropathy and establishment of nomogram model. *Front. Endocrinol. (Lausanne)* 10(14), 1037773. <https://doi.org/10.3389/fendo.2023.1037773> (2023).
12. Haaskjold, Y. L. et al. Long-term follow-up of IgA nephropathy: Clinicopathological features and predictors of outcomes. *Clin. Kidney J.* 16(12), 2514–2522. <https://doi.org/10.1093/ckj/sfad154> (2023).
13. Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO 2021 Clinical Practice Guideline for the Management of Blood Pressure in Chronic Kidney Disease. *Kidney Int.* 99(3S):S1–S87. <https://doi.org/10.1016/j.kint.2020.11.003> (2021).
14. Reich, H. N.\*; Troyanov, Ste[Combining Acute Accent]phan†; Scholey, J. W.\*; Cattran, D. C.\* for the Toronto Glomerulonephritis Registry. Remission of Proteinuria Improves Prognosis in IgA Nephropathy. *J. Am. Soc. Nephrol.* 18(12):3177–3183. <https://doi.org/10.1681/ASN.2007050526>.
15. Tang, C. et al. Time-varying proteinuria and progression of IgA nephropathy: A cohort study. *Am. J. Kidney Dis.* 84(2), 170–178. e1. <https://doi.org/10.1053/j.ajkd.2023.12.016> (2024).
16. Bingxin, Yu. et al. Evaluation of the Oxford classification in immunoglobulin A vasculitis with nephritis: A cohort study and meta-analysis. *Clin. Kidney J.* 14(2), 516–525. <https://doi.org/10.1093/ckj/sfaa129> (2021).
17. Suzuki, H. & Novak, J. IgA nephropathy: Significance of IgA1-containing immune complexes in clinical settings. *J. Clin. Med.* 13, 4495. <https://doi.org/10.3390/jcm13154495> (2024).
18. Eschbach, J. W. et al. Erythropoiesis in patients with renal failure undergoing chronic dialysis. *N. Engl. J. Med.* 276, 653–658. <https://doi.org/10.1056/NEJM196703232761202> (1967).
19. Vos, E. E. et al. Red blood cell survival in long-term dialysis patients. *Am. J. Kidney Dis.* 58(4), 591–598. <https://doi.org/10.1053/j.ajkd.2011.03.031> (2011).
20. Sahay, M. et al. Diabetes and Anemia: International diabetes federation (IDF) - Southeast Asian Region (SEAR) position statement. *Diabetes Metab. Syndr.* 11(Suppl 2), S685–S695. <https://doi.org/10.1016/j.dsx.2017.04.026> (2017).
21. Eschbach, J. W. The anemia of chronic renal failure: Pathophysiology and the effects of recombinant erythropoietin. *Kidney Int.* 35, 134–148. <https://doi.org/10.1038/ki.1989.18> (1989).
22. MacDougall, I. C. et al. A randomized controlled study of iron supplementation in patients treated with erythropoietin. *Kidney Int.* 50, 1694–1699. <https://doi.org/10.1038/ki.1996.487> (1996).
23. Babitt, J. L. & Lin, H. Y. Molecular mechanisms of hepcidin regulation: Implications for the Anemia of CKD. *Am. J. Kidney Dis.* 55, 726–741. <https://doi.org/10.1053/j.ajkd.2009.12.030> (2010).
24. Schödel, J. & Ratcliffe, P. J. Mechanisms of hypoxia signalling: New implications for nephrology. *Nat. Rev. Nephrol.* 15, 641–659. <https://doi.org/10.1038/s41581-019-0182-z> (2019).
25. Ito, M. et al. Prolyl hydroxylase inhibition protects the kidneys from ischemia via upregulation of glycogen storage. *Kidney Int.* 97, 687–701. <https://doi.org/10.1016/j.kint.2019.10.020> (2020).
26. Oh, T. R., Song, S. H., Choi, H. S., Kim, C. S., Han, S. H., Kang, K. P., Kwon, Y. J., Kim, S. W., Ma, S. K., Bae, E. H., on behalf of the Korean GlomeruloNEphritis sTudy (KoGNET) Group. The Association between Serum Hemoglobin and Renal Prognosis of IgA Nephropathy. *J. Clin. Med.* 10:363. <https://doi.org/10.3390/jcm10020363> (2021).
27. Gupta, N. & Wish, J. B. Hypoxia-inducible factor prolyl hydroxylase inhibitors: A potential new treatment for anemia in patients with CKD. *Am. J. Kidney Dis.* 69, 815–826. <https://doi.org/10.1053/j.ajkd.2016.12.011> (2017).
28. Wang, Y. et al. Clinical and pathological factors of renal anaemia in patients with IgA nephropathy in Chinese adults: A cross-sectional study. *BMJ Open* 9, e023479. <https://doi.org/10.1136/bmjopen-2018-023479> (2019).
29. Zhu, B. et al. The association of low hemoglobin levels with IgA nephropathy progression: A two-center cohort study of 1828 cases. *Am. J. Nephrol.* 51, 624–634. <https://doi.org/10.1159/000508770> (2020).
30. Ikee, R. et al. Impact of hypertension and hypertension-related vascular lesions in IgA nephropathy. *Hypertens Res.* 29(1), 15–22. <https://doi.org/10.1291/hypres.29.15> (2006).

## Author contributions

Jinjin Wang:Data Curation,Methodology,Visualization,Writing-Original Draft; Tianmu Chen:Methodology,Software,Formal analysis,Supervision; Yan Fu:Data Curation,Investigation; Tianqi Zhu:Data Curation,Investigation; Yaning Feng:Data Curation,Resources; Yuxi Feng:Investigation;Junwei Huang:Investigation; Xuan Zhang:Investigation;Junwei Huang:Investigation; Yi Cai:Data Curation; Li Gao:Data Curation; Yongqiang Lin:Conceptualization,Writing-Review&Editing; Qinyang Jin:Data Curation,Writing-Review&Editing; Qin Zhu:Writing-Review&Editing,Supervision,Funding acquisition.

## Funding

This study was supported by the National Natural Science Foundation of China (No. 82205008), the Zhejiang Provincial Administration of Traditional Chinese Medicine Project (No. 2023ZF137), the Zhejiang Provincial Department of Health Project (No. 2023RC242),the Research Project of the Zhejiang Province Natural Science Foundation (No. LQ24H180007),and the Joint TCM Science & Technology Projects of National Demonstration Zones for Comprehensive TCM Reform(No.GZY-KJS-ZJ-2026–090) .

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethical approval

Our study was approved by the Ethics Committee of Hangzhou Hospital of Traditional Chinese Medicine (IRB number 2021KY045). Written informed consent was provided by all participants.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-32280-8>.

**Correspondence** and requests for materials should be addressed to Y.L., Q.J. or Q.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025