

Predicting Oncological and Functional Outcomes by Nephrectomy Type for T1 Renal Tumors Using Machine Learning Models

Dongrul Shin, Maisy Song, Jungyo Suh, Cheryn Song

Department of Urology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Received January 8, 2025

Revised March 13, 2025

Accepted March 17, 2025

Corresponding author:

Jungyo Suh

Department of Urology, Asan Medical Center, Ulsan University College of Medicine 88, Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea

Email: uro_jun@amc.seoul.kr

<https://orcid.org/0000-0002-3867-4778>

Purpose: Determining the optimal surgical approach for patients with T1 renal tumors requires balancing long-term oncological and renal functional outcomes. Using machine learning algorithms, we aimed to develop a model to predict both outcomes simultaneously, according to each radical (RN) and partial nephrectomy (PN).

Materials and Methods: Using demographic and preoperative variables of 823 patients with clinical T1N0M0 renal tumors who underwent PN or RN between 2007 and 2019, we employed 5 different machine learning algorithms—general linear model (GLM), extreme gradient boosting (XgBoost), gradient boosting machine, distributed random forest, deep learning—and compared to predict recurrence probability and estimated glomerular filtration rate (eGFR) at 5-year after surgery. Model performance for recurrence prediction was evaluated with area under the curve receiver operating characteristic, area under the precision-recall curve, and log-loss, while eGFR prediction was assessed using root mean square error (RMSE) and R^2 .

Results: Of the 823 patients, 463 (56.3%) had T1a tumors and 487 (59.2%) underwent PN. The median preoperative eGFR was 99.1 mL/min/1.73 m², and at 5 years postoperative it was 70.4 after RN and 92.0 after PN. Recurrence within 5 years was observed in 1.1% and 4.2% of T1a and T1b cohorts, respectively. We developed models based on clinically significant preoperative variables. Across the models, the XGBoost demonstrated the highest accuracy for predicting 5-year recurrence, with superior recall (0.0252) and precision (0.0465) compared to other algorithms. For 5-year eGFR prediction, the GLM outperformed other models, achieving RMSE of 12.700 and R^2 of 0.694 on the test set. The 2 models were integrated into a single online interface.

Conclusion: We developed a tool to reliably predict 5-year oncological and renal functional outcomes following each nephrectomy type in patients with T1 renal tumors. Further multi-institutional validation is needed to confirm its generalizability and applicability across diverse clinical settings.

Key Words: Machine learning, Nephrectomy, Renal cell carcinoma, Recurrence

- **Grant/Fund Support:** This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
- **Research Ethics:** The study was approved by the Institutional Review Board (IRB) of Asan Medical Center (IRB No. 2024-1217).
- **Conflicts of Interest:** The authors have nothing to disclose.

INTRODUCTION

Surgical intervention is the treatment of choice for localized renal cell carcinoma (RCC). In current guidelines, the European Association of Urology and the American Urological Association recommend partial nephrectomy (PN) as the preferred treatment for T1a tumors (≤ 4 cm) and as an alternative for T1b tumors (4–7 cm) whenever feasible [1-3].

Despite these guidelines, the decision between PN and radical nephrectomy (RN) remains complex, often requiring a balance between oncologic control and renal function preservation. While PN has preferred for its nephron-sparing benefits, RN is often chosen in patients with chronic kidney disease, high surgical risk, anticipated prolonged operation time, potential PN-associated morbidities, and most importantly when prioritizing oncological outcomes or survival benefits [4,5].

Meanwhile, machine learning (ML) introduces a novel approach to guiding clinical decisions regarding PN and RN by considering oncological and renal functional outcomes [6-8]. While current prediction models, even including traditional approaches, can individually predict oncological outcomes such as recurrence or renal function outcomes, none have simultaneously addressed both recurrence and renal function [1,4,6-10]. Moreover, few studies focus specifically on T1 renal tumors, which is the largest population in the contemporary cohort of RCC patients, presenting a significant clinical challenge.

In the current study, we aimed to present a prognostic model that leverages ML to simultaneously stratify 5-year recurrence probability and estimated glomerular filtration rate (eGFR) by surgical approach in patients with T1 renal tumors, using only preoperative variables.

MATERIALS AND METHODS

1. Patient Characteristics and Study Variables

After obtaining approval from the institutional ethics committee (IRB No. 2024-1217), data of 892 consecutive patients with T1N0M0 RCC who underwent RN or PN at Asan Medical Center between 2007 and 2019 was retrieved. Patients with single kidney, bilateral tumors, hereditary renal

cancers involving multiple nephron-sparing procedures in the same kidney, histology other than primary adenocarcinoma of the kidney were excluded. Patients without preoperative ^{99m}Tc -diethylenetriaminepentaacetic acid renal scans or follow-up until 5 years after surgery were also excluded, leaving a total of 823 patients in the analytical cohort.

In the cohort, we reviewed patient demographics and clinical characteristics, including age, sex, body mass index, and comorbidities, along with tumor-specific parameters such as size, radiologic assessment of histologic subtype, and R.E.N.A.L. nephrometry score [11], as well as renal function-related data. Recurrence was defined as radiographic evidence of loco-regional recurrence or distant metastasis identified during the 5-year postoperative follow-up period. Renal functional outcome was assessed with eGFR calculated using CKD-EPI formula at postoperative 5 years [12].

2. Model Development

Model development incorporated preoperative variables regarded clinically significant based on established renal tumor surgery literature. Preprocessing was applied, and these variables were directly included in model training without univariate analysis or feature selection, given their established clinical relevance. To ensure the proportionality of RN and PN groups, the dataset was initially separated into RN and PN subgroups before stratification by recurrence. Each subgroup was then splitted into 80% training and 20% testing sets to maintain the recurrence rate of 2.4% be consistent across both sets to address class imbalance. The stratified RN and PN subsets were subsequently merged. K-fold cross-validation ($k=5$) was implemented during training to enhance evaluation robustness and model stability.

For performance assessment, 5 ML models—gradient boosting machine (GBM), extreme gradient boosting (XGBoost), general linear model (GLM), distributed random forest (DRF), and deep learning (DL)—were employed to predict eGFR and recurrence probabilities in clinical T1 renal tumor patients 5 years postoperatively, considering the 2 different surgical methods. GBM, XGBoost, and DRF are based on decision tree models. GBM combines many small decision trees that learn in stages, often achieving strong performance albeit risking overfitting if parameters are not

carefully tuned. XGBoost is a faster, more advanced form of GBM that reduces overfitting with built-in regularization and speeds up training by using parallel processing to run different parts of the model at the same time. DRF trains many decision trees simultaneously on different data subsets, providing robust performance with simpler tuning. GLM links inputs to outputs through linear mathematical relationships, making it easy to interpret but often less capable of handling complex data. DL uses neural networks with multiple layers to automatically learn features, with this study specifically using a multilayer perceptron.

Model evaluation metrics included the area under the receiver operating characteristic curve (AUROC) to assess

classification quality, the area under the precision-recall curve (AUPRC) to emphasize performance on the imbalanced dataset, and log-loss to measure overall predictive accuracy. Precision and recall were also analyzed to evaluate the models' ability to identify true positives and minimize false negatives, as accuracy alone can be misleading in imbalanced datasets. Data preprocessing, model training, and performance evaluation were conducted via Python version 3.11.8, with statistical significance determined at $p < 0.05$.

RESULTS

The analytical cohort included 463 (56.3%) T1a cases,

Table 1. Baseline demographic and clinical characteristics of patients who underwent nephrectomy for clinically suspected T1 renal tumors

Characteristic	Total (n=823)	T1a stage (n=463)	T1b stage (n=360)
Age (yr)	54 (46–62)	53 (45–62)	55 (47–62)
Sex			
Male	567 (68.9)	328 (70.8)	239 (66.4)
Female	256 (31.1)	135 (29.1)	121 (33.6)
BMI (kg/m ²)	25.1 (23.3–27.2)	24.9 (23.1–26.8)	25.7 (23.6–27.7)
Comorbidities			
Diabetes mellitus	123 (14.9)	63 (13.6)	60 (16.7)
Hypertension	301 (36.6)	164 (35.4)	137 (38.1)
History of stroke	22 (2.7)	13 (2.8)	9 (2.5)
Nephrectomy type			
Radical nephrectomy	336 (40.8)	113 (24.4)	223 (61.9)
Partial nephrectomy	487 (59.2)	350 (75.6)	137 (38.1)
Tumor size, median (cm)	3.5 (2.4–4.7)	2.5 (2.0–3.0)	4.9 (4.5–5.5)
Presumed histologic subtype on CT			
Clear cell	681 (82.7)	383 (82.7)	298 (82.8)
Non-clear cell	142 (17.3)	80 (17.3)	62 (17.2)
Nephrometry score			
Radius			
1	462 (56.1)	463 (100)	0 (0)
2	350 (42.6)	0 (0)	349 (96.9)
3	11 (1.3)	0 (0)	11 (3.1)
Exophytic/endophytic			
1	343 (41.7)	164 (35.4)	179 (49.7)
2	344 (41.8)	202 (43.6)	142 (39.4)
3	136 (16.5)	97 (21.0)	39 (10.8)
Nearness to collecting system			
1	222 (27.0)	182 (39.3)	40 (11.1)
2	125 (15.2)	83 (17.9)	42 (11.7)
3	476 (57.8)	198 (42.8)	278 (77.2)
Location relative to polar lines			
1	391 (47.5)	270 (58.3)	121 (33.6)
2	174 (21.1)	82 (17.7)	92 (25.6)
3	258 (31.4)	111 (24.0)	147 (40.8)
Preoperative serum creatinine (mg/dL)	0.85 (0.72–0.97)	0.84 (0.73–0.97)	0.86 (0.72–0.97)
Preoperative eGFR (mL/min/1.73 m ²)	99.1 (89.3–107.2)	100.0 (89.8–107.4)	97.6 (87.9–105.8)
5-year postoperative eGFR (mL/min/1.73 m ²)	84.0 (68.5–96.9)	89.9 (73.8–100.6)	76.0 (64.6–90.2)
Relative renal function (%)	49.0 (46.0–52.0)	49.7 (47.3–52.1)	47.9 (44.8–51.9)
Recurrence	20 (2.4)	5 (1.1)	15 (4.2)

Values are presented as median (interquartile range) or number (%).

BMI, body mass index; CT, computed tomography; eGFR, estimated glomerular filtration rate.

75.6% of which underwent PN, and 360 (43.7%) T1b cases, 38.1% of which underwent PN (Table 1). Median age was 54 years (interquartile range, 46–62), with 68.9% being male. Histologic subtype was presumed to be of clear cell type in 82.7% of patients. Recurrence was observed in 1.1% of T1a and 4.2% of T1b patients at 5 years. The training set included 657 patients (79.8%) and the test set 166 (20.2%), maintaining similar distributions of surgery method and clinical stage (Table 2).

We developed a model including only the clinically significant preoperative parameters (Fig. 1). Among the models, the XGBoost demonstrated the highest performance in

predicting 5-year recurrence, achieving an AUROC of 0.8850 and an AUPRC of 0.3790, indicating strong predictive accuracy (Table 3). Conversely, DL had the poorest performance, with the lowest AUROC (0.5679). GLM, GBM, and DRF produced intermediate outcomes, with AUROC values ranging from 0.6474 to 0.8488 (Fig. 2).

With respect to 5-year eGFR prediction, the GLM demonstrated the lowest RMSE in the test set (12.700), with respective R² values of 0.688 and 0.694, indicating strong predictive accuracy (Table 4). The GBM and XGBoost models exhibited comparable RMSE and R² values; however, they showed larger differences between training and test set performance. Overall, GLM was the most accurate model for predicting 5-year eGFR (Fig. 2).

Table 2. Distribution of T1 renal tumor patients by nephrectomy type and clinical stage for model training and validation

Variable	Training set (n=657)	Test set (n=166)
Nephrectomy type		
Radical nephrectomy	268 (40.8)	68 (41.0)
Partial nephrectomy	389 (59.2)	98 (59.0)
Clinical stage		
T1a	373 (56.8)	90 (54.2)
T1b	284 (43.2)	76 (45.8)
Recurrence after 5 years	16 (2.4)	4 (2.4)

Values are presented as number (%).

DISCUSSION

Our study demonstrates a novel application of ML to predict both oncological (recurrence) and renal function (eGFR) outcomes in patients with clinical T1 renal tumors undergoing nephrectomy. This dual-prediction framework bridges a significant gap in the literature, which has typi-

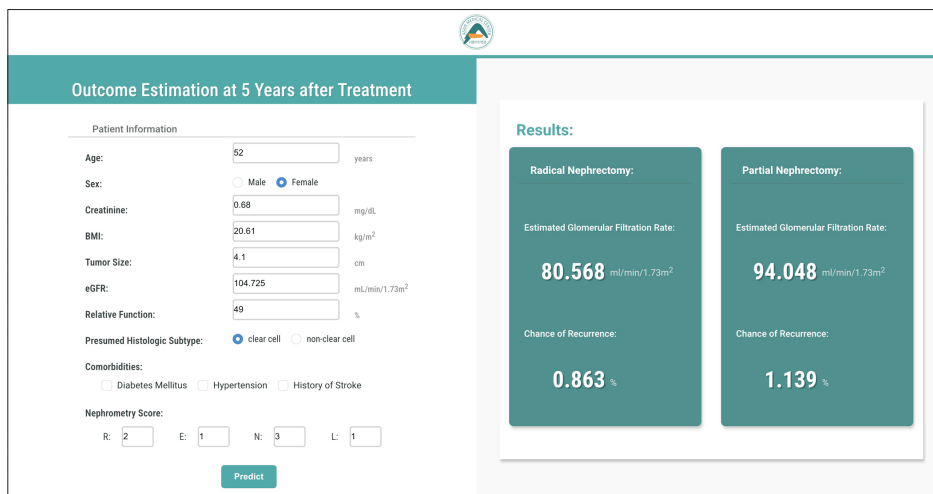


Fig. 1. An online-based predictive model interface for estimating 5-year estimated glomerular filtration rate and recurrence outcomes in T1 renal tumor patients based on nephrectomy type.

Table 3. Performance comparison of machine learning models for predicting 5-year recurrence in patients with T1 renal tumors after nephrectomy

Model	AUROC	AUPRC	Log-loss	Recall	Precision
General linear model	0.8426	0.0769	0.0992	0.0288	0.0999
Extreme gradient boosting (XGBoost)	0.8850	0.3790	0.1027	0.0252	0.0465
Gradient boosting machine	0.8488	0.1294	0.0989	0.0223	0.3175
Distributed random forest	0.6474	0.2817	0.2842	0.0000	0.4565
Deep learning	0.5679	0.0259	0.1539	0.0014	0.0014

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

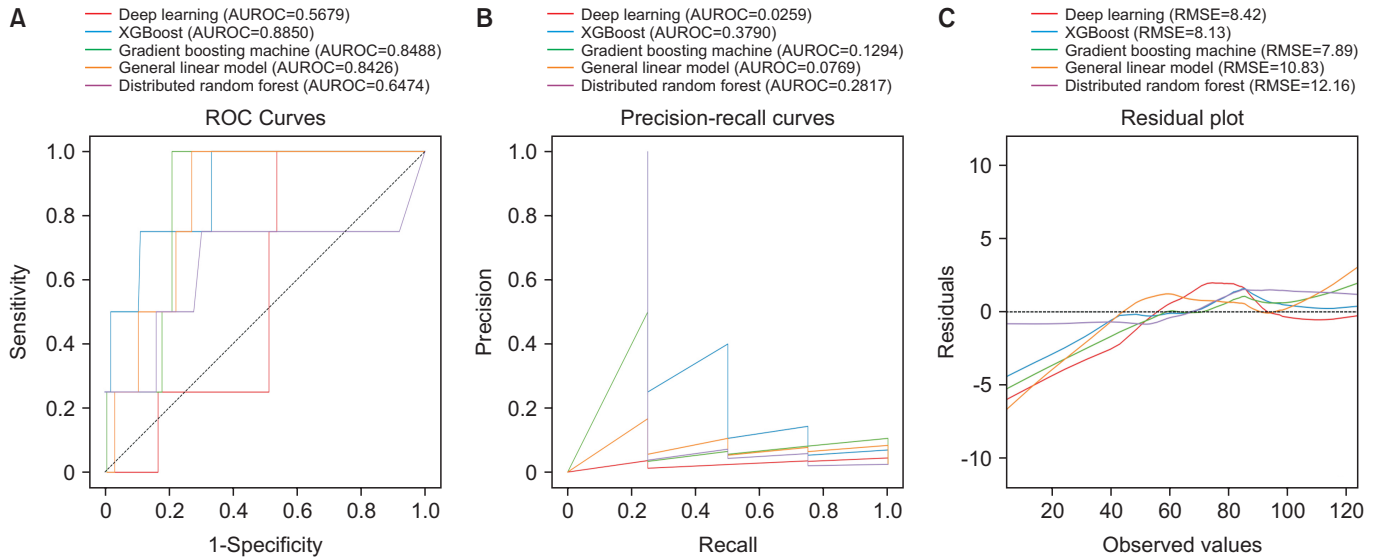


Fig. 2. Area under the receiver operating characteristic (AUROC) curves (A) and precision-recall curves (B) of machine learning models for predicting tumor recurrence in T1 renal tumor patients: models evaluated include deep learning, gradient boosting machine, general linear model, extreme gradient boosting (XgBoost), and distributed random forest (C). Residual plots of machine learning models for predicting 5-year estimated glomerular filtration rate in T1 renal tumor patients: models evaluated include deep learning, gradient boosting machine, general linear model, extreme gradient boosting, and distributed random forest. RMSE, root mean square error.

Table 4. Performance comparison of machine learning models for predicting 5-year eGFR in patients with T1 renal tumors after nephrectomy

Model	RMSE train	RMSE test	R ² train	R ² test	Within ±5% of actual values	Within ±10% of actual values
General linear model	10.831	12.700	0.688	0.694	26.51%	51.81%
Extreme gradient boosting (XGBoost)	8.125	12.636	0.825	0.697	22.89%	53.01%
Gradient boosting machine	7.886	12.892	0.835	0.684	26.51%	50.60%
Distributed random forest	12.158	12.780	0.607	0.690	24.10%	48.80%
Deep learning	8.420	13.697	0.812	0.644	23.49%	51.2%

eGFR, estimated glomerular filtration rate; RMSE, root mean square error.

cally focused on either oncological or renal outcomes separately. Our findings emphasize the importance of pre-operative factors such as tumor size, nephrometry score, and patient characteristics in prediction, which aligns with previous studies that identified these variables as critical in surgical outcomes [6,7,9,10]. Unlike prior models that often incorporated postoperative data, our approach relied exclusively on preoperative variables, thereby enhancing its relevance and applicability to real-world surgical decision-making.

For example, in our study, radiographic assessment of histologic subtype (clear cell vs. non-clear cell) was incorporated as a variable, which contributed to improving the accuracy of the recurrence prediction model. Despite potential limitations in the accuracy of radiographic histologic

subtype assessment, we included this parameter given its clinical relevance; it is an essential factor that clinicians consider when determining the appropriate surgical approach [13].

The XGBoost model excelled in predicting 5-year recurrence, achieving the highest AUROC (0.8850) and AUPRC (0.3790) compared to other models. Precision-recall curves were employed to better assess model performance in this imbalanced setting, which provided more nuanced insights compared to traditional ROC curves [14,15]. High AUROC and AUPRC scores reinforce the reliability of the model in detecting true recurrence cases. The XGBoost model’s particularly high recall for recurrence prediction (0.0252) is notable given the low incidence rate (2.4%) in the cohort, with high recall being especially valuable to ensure true

cases are not missed. Its strong recall shows its robustness in handling imbalanced data, especially when hyperparameters are fine-tuned to avoid overfitting [16].

On the other hand, the GLM exhibited the highest predictive accuracy for 5-year eGFR, as evidenced by the low RMSE in both training (10.831) and test (12.700) sets, and relatively strong R^2 values with 0.694 in the test set. This aligns with the nature of eGFR, which can often be estimated accurately through relatively linear associations with preoperative factors such as age, baseline renal function, and tumor size. Although some prior studies have employed binary classification for postoperative eGFR assessment [4,17], predicting eGFR as a continuous variable provides more precise information by offering exact estimates of renal function at 5 years, which is one of the goals of this study. The effectiveness of GLM in modeling straightforward linear correlations with continuous preoperative variables contributes to its predictive reliability in estimating renal function [18].

Several limitations need to be acknowledged in this study. First, its retrospective design may introduce biases related to patient selection and data availability. Additionally, the relatively small, single-institution sample limits cohort diversity, particularly regarding race, ethnicity, and regional healthcare practices. Another limitation is the lack of external validation; although our models performed well internally, their applicability to other clinical settings remains untested. Finally, the model does not account for intraoperative factors or postoperative complications that may affect long-term outcomes. Future research should include multi-institutional validation to better assess generalizability across varied populations and healthcare environments.

CONCLUSIONS

In this study, we developed and internally validated an online interface using ML to predict 5-year oncological and renal functional outcomes for patients with T1 renal tumors. Further multi-institutional validation is needed to confirm its generalizability and applicability across diverse clinical settings.

NOTES

• Author Contribution:

Conceptualization: CS; Data curation: DS, MS, JS, CS; Formal analysis: MS, JS; Methodology: DS, MS, JS, CS; Project administration: CS; Visualization: DS, MS; Writing - original draft: DS; Writing - review & editing: DS, MS, JS, CS.

• ORCID

Dongrul Shin: <https://orcid.org/0000-0002-7758-3342>

Maisy Song: <https://orcid.org/0009-0003-2083-8505>

Jungyo Suh: <https://orcid.org/0000-0002-3867-4778>

Cheryn Song: <https://orcid.org/0000-0002-1823-4281>

REFERENCES

1. Takagi T, Yoshida K, Wada A, Kondo T, Fukuda H, Ishihara H, et al. Predictive factors for recurrence after partial nephrectomy for clinical T1 renal cell carcinoma: a retrospective study of 1227 cases from a single institution. *Int J Clin Oncol* 2020;25:892-8.
2. Ljungberg B, Albiges L, Abu-Ghanem Y, Bedke J, Capitanio U, Dabestani S, et al. European Association of Urology guidelines on renal cell carcinoma: the 2022 update. *Eur Urol* 2022;82:399-410.
3. Aguilar Palacios D, Wilson B, Ascha M, Campbell RA, Song S, DeWitt-Foy ME, et al. New baseline renal function after radical or partial nephrectomy: a simple and accurate predictive model. *J Urol* 2021;205:1310-20.
4. Bhindi B, Lohse CM, Schulte PJ, Mason RJ, Cheville JC, Boorjian SA, et al. Predicting renal function outcomes after partial and radical nephrectomy. *Eur Urol* 2019;75:766-72.
5. Gu L, Ma X, Li H, Chen L, Xie Y, Li X, et al. Comparison of oncologic outcomes between partial and radical nephrectomy for localized renal cell carcinoma: a systematic review and meta-analysis. *Surg Oncol* 2016;25:385-93.
6. Lee Y, Ryu J, Kang MW, Seo KH, Kim J, Suh J, et al. Machine learning-based prediction of acute kidney injury after nephrectomy in patients with renal cell carcinoma. *Sci Rep* 2021;11:15704.
7. Kim HM, Byun SS, Kim JK, Jeong CW, Kwak C, Hwang EC, et al. Machine learning-based prediction model for late recurrence after surgery in patients with renal cell carcinoma. *BMC Med Inform Decis Mak* 2022;22:241.
8. Khene ZE, Bigot P, Doumerc N, Ouzaid I, Boissier R, Nouhaud FX, et al. Application of machine learning models to predict recurrence after surgical resection of nonmetastatic renal cell carcinoma. *Eur Urol Oncol* 2023;6:323-30.

9. Abdel Raheem A, Shin TY, Chang KD, Santok GDR, Alenzi MJ, Yoon YE, et al. Yonsei nomogram: a predictive model of new-onset chronic kidney disease after on-clamp partial nephrectomy in patients with T1 renal tumors. *Int J of Urol* 2018;25:690-7.
10. Wang K, Guo B, Niu Y, Li G. Development and validation of a nomogram to predict recurrence for clinical T1/2 clear cell renal cell carcinoma patients after nephrectomy. *BMC Surg* 2024;24:196.
11. Kutikov A, Uzzo RG. The R.E.N.A.L. nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. *J Urol* 2009;182:844-53.
12. Levey AS, Stevens LA. Estimating GFR using the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD Prevalence estimates, and better risk predictions. *Ame J Kidney Dis* 2010; 55:622-7.
13. Klatte T, Ficarra V, Gratzke C, Kaouk J, Kutikov A, Macchi V, et al. A literature review of renal surgical anatomy and surgical strategies for partial nephrectomy. *Eur Urol* 2015;68:980-92.
14. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevskaya O; written on behalf of AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med* 2019;7:152.
15. Fu G, Yi L, Pan J. Tuning model parameters in class-imbalanced learning with precision-recall curve. *Biometrical J* 2019;61:652-64.
16. Abdu-Aljabar RD, Awad OA. A comparative analysis study of lung cancer detection and relapse prediction using XG-Boost classifier. *IOP Conf Ser: Mater Sci Eng* 2021;1076: 012048.
17. Rathi N, Palacios DA, Abramczyk E, Tanaka H, Ye Y, Li J, et al. Predicting GFR after radical nephrectomy: the importance of split renal function. *World J Urol* 2022;40:1011-8.
18. Yousif JM. A comparative analysis between various machine learning models and generalized linear models [master thesis]. Stockholm (Sweden): Mathematical Statistics Stockholm University; 2023.