Machine learning in renal cell carcinoma research: the promise and pitfalls of 'renal-izing' the potential of artificial intelligence

Zine-Eddine Khene, Alexander Kutikov, Riccardo Campi

Khene Zine-Eddine (Orcid ID: 0000-0002-4124-789X)
Campi Riccardo (Orcid ID: 0000-0001-5237-0888)

Machine Learning in Renal Cell Carcinoma Research: The Promise and Pitfalls of Renal-izing the Potential of Artificial Intelligence

Zine-Eddine KHENE[1,2]*, Alexander KUTIKOV [3], Riccardo CAMPI [4,5],

On behalf of the EAU-YAU Renal Cancer working group

1: Department of urology, Rennes University Hospital, Rennes, France
2: Signal and Image Processing Laboratory (LTSI), University of Rennes INSERM, U1099, Rennes, France
3: Department of Surgical Oncology, Division of Urologic Oncology, Fox Chase Cancer Center, Philadelphia, PA, USA.
4: Unit of Urological Minimally Invasive, Robotic Surgery and Kidney Transplantation, Careggi Hospital, University of Florence, Florence, Italy; Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.
5: European Association of Urology (EAU) Young Academic Urologists (YAU) Renal Cancer Working Group, Arnhem, the Netherlands.

*** Corresponding author:**
Zine-Eddine KHENE
Department of urology, Rennes University Hospital
2, rue Henri Le Guilloux, 35033 Rennes Cedex, France
E-mail : zineddine.khene@gmail.com

Keywords: Renal cell carcinoma, Kidney cancer, Artificial intelligence, Machine learning

This letter received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest: The authors have nothing to disclose.

Artificial intelligence and Machine learning (ML) are increasingly applied for study of patients with renal cell carcinoma (RCC) [1]. Advanced techniques, such as neural networks or random forests can analyze vast amount of clinical data to uncover specific prognostic features that may not be detectable with traditional statistical methods. However, despite some promising results, recent studies showcase that benefits of ML are not ubiquitous, especially when deployed on imperfect and non-granular datasets[2].

In this issue of BJUI, Boulenger de Hauteclocque employ data from the French multi-institutional kidney cancer database UroCCR (ClinicalTrials.gov Identifier: NCT03293563), to investigate the ability of seven ML algorithms at predicting pT3a upstaging in a cohort of patients who underwent surgery for cT1/cT2a RCC. Using supervised ML algorithms, they reported a prediction accuracy (measured by the area under the receiver operating curve) of 0.77 for their best model [3]. While the results are intriguing, the design of the study raises questions that should be considered by clinicians when interpreting the results.

First, the accuracy of the model is heavily dependent on the fidelity and granularity of the dataset used for training. For instance, missing data have a significant noxious impact on the performance of the model. For their study, Boulenger et al. worked with a database with a high rate (205) of missing observations in the variables used to construct the prognostic model, such as the R.E.N.A.L. nephrometry score. Thus, limitations of the presented models stemmed, at least in part, from the nature and handicap of the dataset. Indeed, deep neural network approach is particularly effective when deployed for analyzing large, complex, and high-dimensional datasets, such as cross-sectional imaging, that are highly granular and are of high fidelity [4].

Second, ML algorithms have several adjustable hyperparameters. Hyperparameters can be compared to knobs of an amplifier that allow to fine-tune the bass and treble of an audio track. They are usually set before the training process begins. A search for optimal hyperparameters must be done using the training set to achieve a high performance (there are usually two strategies for hyperparameter tuning: GridSearchCV or RandomizedSearchCV). A detailed description of the hyperparameters tuning process is lacking and would strengthen this report [5]. Similarly, decision curve analysis is also missing. While the authors do report receiver operating characteristic curves and calibration measurements of the models, details of decision curve analysis are important since they provide a probability of certainty for the decisions that can be made in daily practice based on the results of the algorithm [6].

Third, ML approaches have a “shadow zone”, referred to as a “black box”, which results in challenges of understanding the model estimates. Boulenger de Hauteclocque and colleagues acknowledge this limitation and use the shapley additive explanations (SHAP) values to explain each patient’s probability of being upstaged. SHAP is a

popular method used to understand how ML models make predictions, but it has limitations. One issue is that SHAP assumes the included variables are independent from one another, while in reality, there may be interdependencies among the variables. In that case, the approximations made by SHAP can be skewed (for instance tumor size and RENAL score, both of which are used in the ML model, are intimately linked) [7].

The Standardized Reporting of Machine Learning Applications in Urology (STREAM-URO) framework was developed to provide a set of recommendations to standardize the way ML studies are reported [8]. It includes providing a detailed report of the model, along with the code, especially in case of complex ML systems. The authors of the study largely adhered to the STREAM-URO guidelines. However, the manuscript omits the models themselves, which prohibits the readers from utilizing them to predict outcomes on new populations and to validate the model with external data.

Finally, from a clinical and surgical standpoint, the ability to predict pT3a upstaging among patients with localized renal masses who are candidates for elective surgery is not always actionable. On one hand, decision-making in these patients is complex and nuanced; as such, beyond tumor-related characteristics, the treatment decisions should also rely on careful assessment of patient-, kidney- and provider-related factors [9]. On the other hand, PN and RN appear to achieve similar cancer control even if the final histology shows unexpected pT3a RCC upstaging [10]. In addition, in a recent systematic review, age, tumor size and RENAL score were the three predictors of upstaging [11]. In this regard, the added value of ML methods as compared to conventional logistic regression models for prediction of pathological features of RCC is unclear. Indeed, several reports indicate that logistic regression is at least as good as ML techniques in predict such outcomes [12]. Of note, an AUC of 0.77 (as reported in the study by Boulenger de Hauteclocque et al) is close to halfway between a random coin flip (AUC of 0.5) and a perfect prediction (AUC of 1.0), being comparable to previously reported “standard” modes [13].

In conclusion, the report by the authors is novel and noteworthy; however, the available evidence does not entirely support the claim that prediction models obtained using ML algorithms on standard clinical datasets are superior and more clinically useful when compared to conventional predictive tools. While ML techniques harbor new opportunities in the field of RCC prognostication, it is imperative to standardize and improve the quality of reporting of these novel AI algorithms, as well as to thoroughly validate them before integration into shared decision-making in daily clinical practice.

References

[1] Lee M, Wei S, Anaokar J, Uzzo R, Kutikov A. Kidney cancer management 3.0: can artificial intelligence make us better? Curr Opin Urol 2021;31:409–15. https://doi.org/10.1097/MOU.0000000000000881.

[2] Khene Z-E, Bigot P, Doumerc N, Ouzaid I, Boissier R, Nouhaud F-X, et al. Application of Machine Learning Models to Predict Recurrence After Surgical Resection of Nonmetastatic Renal Cell Carcinoma. Eur Urol Oncol 2022:S2588-9311(22)00137-7. https://doi.org/10.1016/j.euo.2022.07.007.

[3] Boulenger de Hauteclocque A, Ferrer L, Ambrosetti D, Ricard S, Bigot P, Bensalah K, et al. Machine learning approach for prediction of pT3a upstaging and outcomes of localized RCC (UroCCR-15). BJU Int 2023. https://doi.org/10.1111/bju.15959.

[4] Heller N, Tejpaul R, Isensee F, Benidir T, Hofmann M, Blake P, et al. Computer-Generated R.E.N.A.L. Nephrometry Scores Yield Comparable Predictive Results to Those of Human-Expert Scores in Predicting Oncologic and Perioperative Outcomes. J Urol 2022;207:1105–15. https://doi.org/10.1097/JU.0000000000002390.

[5] Liu Y, Chen P-HC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA 2019;322:1806. https://doi.org/10.1001/jama.2019.16489.

[6] Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Med Decis Making 2006;26:565–74. https://doi.org/10.1177/0272989X06295361.

[7] Fryer D, Strümke I, Nguyen H. Shapley Values for Feature Selection: The Good, the Bad, and the Axioms. IEEE Access 2021;9:144352–60. https://doi.org/10.1109/ACCESS.2021.3119110.

[8] Kwong JCC, McLoughlin LC, Haider M, Goldenberg MG, Erdman L, Rickard M, et al. Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework. Eur Urol Focus 2021;7:672–82. https://doi.org/10.1016/j.euf.2021.07.004.

[9] Chandrasekar T, Boorjian SA, Capitanio U, Gershman B, Mir MC, Kutikov A. Collaborative Review: Factors Influencing Treatment Decisions for Patients with a Localized Solid Renal Mass. Eur Urol 2021;80:575–88. https://doi.org/10.1016/j.eururo.2021.01.021.

[10] Capitanio U, Stewart GD, Klatte T, Akdogan B, Roscigno M, Marszalek M, et al. Does the Unexpected Presence of Non-organ-confined Disease at Final Pathology Undermine Cancer Control in Patients with Clinical T1N0M0 Renal Cell Carcinoma Who Underwent Partial Nephrectomy? European Urology Focus 2018;4:972–7. https://doi.org/10.1016/j.euf.2017.02.020.

[11] Veccia A, Falagario U, Martini A, Marchioni M, Antonelli A, Simeone C, et al. Upstaging to pT3a in Patients Undergoing Partial or Radical Nephrectomy for cT1 Renal Tumors: A Systematic Review and Meta-analysis of Outcomes and Predictive Factors. Eur Urol Focus 2021;7:574–81. https://doi.org/10.1016/j.euf.2020.05.013.

[12] A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models | Elsevier Enhanced Reader n.d. https://doi.org/10.1016/j.jclinepi.2019.02.004.

[13] Kutikov A, Smaldone MC, Egleston BL, Manley BJ, Canter DJ, Simhan J, et al. Anatomic features of enhancing renal masses predict malignant and high-grade pathology: a preoperative nomogram using the RENAL Nephrometry score. Eur Urol 2011;60:241–8. https://doi.org/10.1016/j.eururo.2011.03.029.