American Urological Association

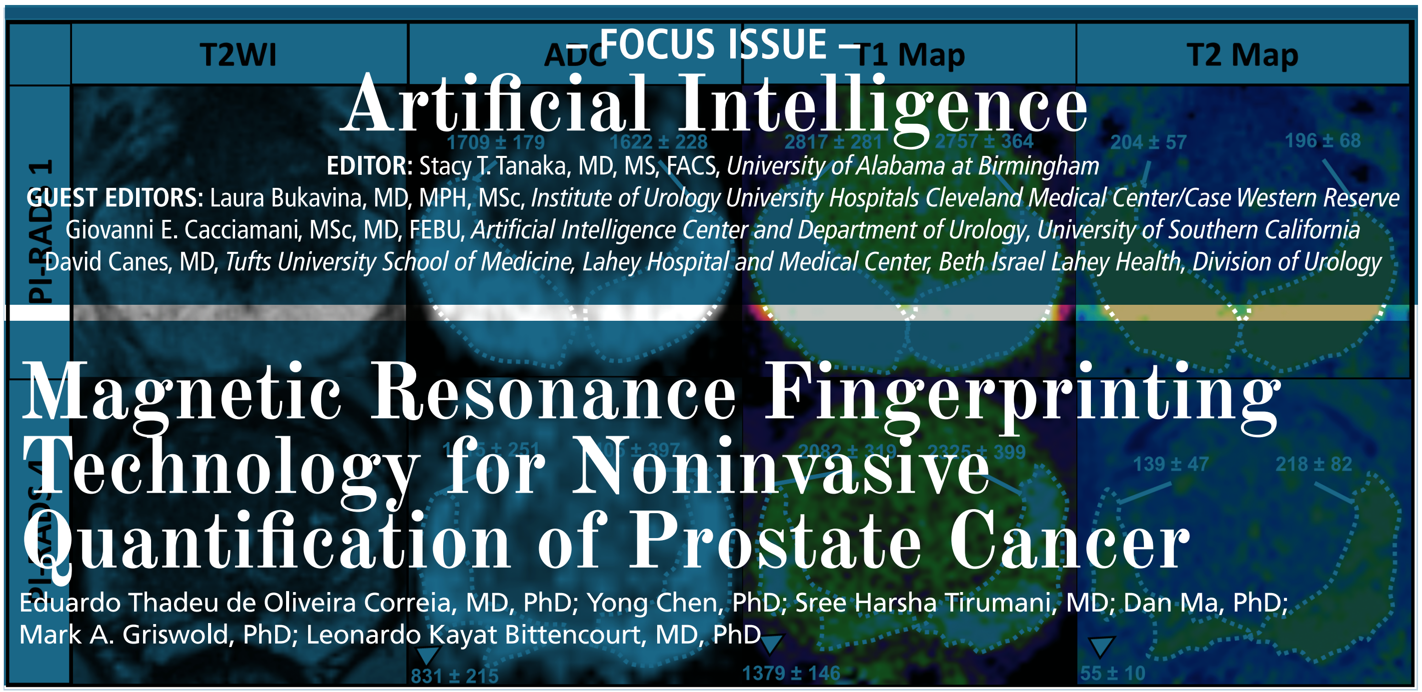
AUANews

JANUARY 2024

VOLUME 29 | ISSUE 1

AUANews.net

THE OFFICIAL NEWSMAGAZINE OF THE AMERICAN UROLOGICAL ASSOCIATION



– FOCUS ISSUE –

Artificial Intelligence

EDITOR: Stacy T. Tanaka, MD, MS, FACS, *University of Alabama at Birmingham*

GUEST EDITORS: Laura Bukavina, MD, MPH, MSc, *Institute of Urology University Hospitals Cleveland Medical Center/Case Western Reserve*

Giovanni E. Cacciamani, MSc, MD, FEBU, *Artificial Intelligence Center and Department of Urology, University of Southern California*

David Canes, MD, *Tufts University School of Medicine, Lahey Hospital and Medical Center, Beth Israel Lahey Health, Division of Urology*

Magnetic Resonance Fingerprinting Technology for Noninvasive Quantification of Prostate Cancer

Eduardo Thadeu de Oliveira Correia, MD, PhD; Yong Chen, PhD; Sree Harsha Tirumani, MD; Dan Ma, PhD; Mark A. Griswold, PhD; Leonardo Kayat Bittencourt, MD, PhD

Leveraging Artificial Intelligence for Prehabilitation Interventions to Improve Perioperative Outcomes

Meghana Noonavath, BS
Chris W. Lewis, MD
Hanna Hunter, MD
Sarah P. Psutka, MD, MS

Artificial Intelligence Chatbots: How Accurate Is the Information?

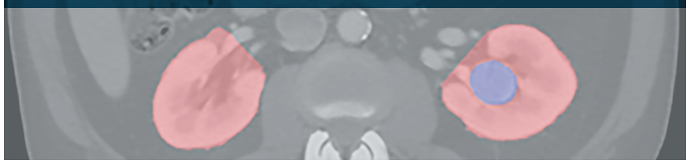
Bristol B. Whiles, MD
Russell S. Terry, MD

European Perspective on Importance of Artificial Intelligence Lectures and Sessions in Urology Meetings

Kari A. O. Tikkinen, MD, PhD
Jochen Walz, MD

Artificial Intelligence–Powered Imaging for Renal Cancer and Nephrometric Scoring Automation

Andrew M. Wood, MD
Nour Abdallah, MD
Rebecca A. Campbell, MD
Christopher J. Weight, MD



INSIDE THIS ISSUE

COMING SOON in January AUANews Extra

NON-PROFIT ORG.
U.S. POSTAGE
PAID
PERMIT NO. 797
RICHMOND, VA

American Urological Association
1000 Corporate Boulevard
Linthicum, Maryland 21090

Xtandi®
(enzalutamide)
40 mg tablets | 80 mg tablets

NOW APPROVED FOR A NEW INDICATION

LEARN ABOUT THE RESULTS OF THE PIVOTAL TRIAL

Scan the QR code
to view the publication of the pivotal trial

astellas ONCOLOGY

Pfizer Oncology

ARTIFICIAL INTELLIGENCE

Magnetic Resonance Fingerprinting Technology for Noninvasive Quantification of Prostate Cancer

Eduardo Thadeu de Oliveira Correia, MD, PhD
University Hospitals Cleveland Medical Center, Ohio

Yong Chen, PhD
University Hospitals Cleveland Medical Center, Ohio
Case Western Reserve University, Cleveland, Ohio

Sree Harsha Tirumani, MD
University Hospitals Cleveland Medical Center, Ohio
Case Western Reserve University, Cleveland, Ohio

Dan Ma, PhD
Case Western Reserve University, Cleveland, Ohio
University Hospitals Cleveland Medical Center, Ohio

Mark A. Griswold, PhD
University Hospitals Cleveland Medical Center, Ohio
Case Western Reserve University, Cleveland, Ohio

Leonardo Kayat Bittencourt, MD, PhD
University Hospitals Cleveland Medical Center, Ohio
Case Western Reserve University, Cleveland, Ohio

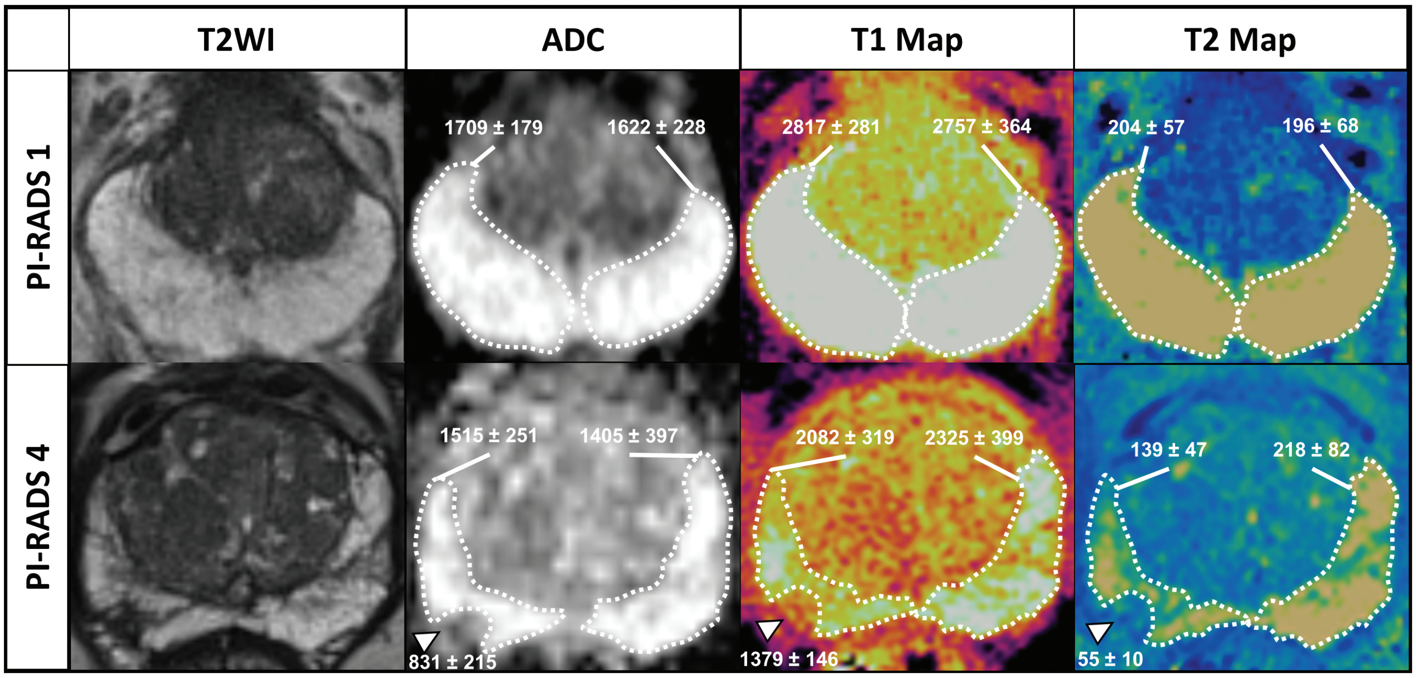


Figure. Magnetic resonance fingerprinting (MRF) mapping of the prostate in patients with and without a focal suspicious lesion. The top row illustrates a case of a Prostate Imaging Reporting and Data System (PI-RADS) 1 prostate MRI, with no focal suspicious lesions in the peripheral zone. The bottom row illustrates a case of a PI-RADS 4 MRI, with a focal suspicious lesion in the right peripheral zone (indicated by a white arrow). As demonstrated in previous studies,[5,13] MRF-derived T1 and T2 values of the suspicious lesions are significantly lower than T1 and T2 values of the normal-appearing peripheral zone (contoured by a dotted line). ADC indicates apparent diffusion coefficient; T2WI, T2-weighted imaging.

Introduction

Over the last decade, there has been a growing interest in the investigation of quantitative MRI techniques in the field of prostate imaging. The primary goal of quantitative imaging is to improve both intra- and inter-reader reproducibility by using objective tissue property values to diagnose and stage suspicious prostate cancer (PCa) lesions. Nonetheless, early conventional mapping techniques were often time-consuming and reproducibility across different MRI scanners is often a concern. Our team at Case Western Reserve University, in collaboration with Siemens Healthineers, developed a native quantitative MRI technique known as magnetic resonance fingerprinting (MRF).[1] MRF allows for the simultaneous acquisition of T1 and T2 maps,[1] as illustrated in the Figure, and more recently, diffusion maps,[2] all within a clinically feasible time. Importantly, the multiple tissue properties maps acquired by MRF are incoherently coregistered, enabling direct multiparametric tissue analysis.

To this date, within the field of abdominal radiology, particularly in genitourinary imaging, the primary focus of MRF has been to improve the noninvasive detection and characterization of PCa in both the transition zone (TZ) and peripheral zone (PZ).[3-5] This paper provides an overview of prior developments in prostate MRF while highlighting emerging applications that hold potential for reshaping PCa management.

Current Applications of Prostate MRF

One pioneering application of prostate MRF was introduced by Yu et al, who proposed a novel prostate MRI examination that combines standard apparent diffusion coefficient (ADC) maps with MRF-derived T1 and T2 maps.[3] In this paper, Yu et al demonstrated that T1 and T2 relaxation times of areas of known PCa were markedly lower than those of the normal-appearing PZ. Furthermore, integrating ADC, T1, and T2 maps yielded an area under the curve (AUC) of 0.83 for distinguishing low-grade from intermediate- and high-grade PCa lesions.[3] Similar findings with the same methodology were shown by Panda et al, albeit focusing on the TZ.[4] In this investigation, Panda et al showed that the combination of MRF-derived T1 maps and standard ADC mapping could allow for the differentiation of PCa lesions from the normal TZ, with an AUC of 0.94.[4] Subsequently, Shiradkar et al assessed the likely biological basis behind variations in T1 and T2 relaxation times using histopathology specimens from radical prostatectomy.[6] Their study showed that parameters of tissue composition ratio (percentage of epithelium, stroma, and lumen) differed between areas of normal prostatic tissue, prostatitis, and PCa.[6] Remarkably, T1 and T2 relaxation times were correlated with parameters of tissue composition ratio.[6] Beyond localized PCa, Choi et al also demonstrated that prostate MRF–derived T1 and T2 maps could differentiate PCa bone metastasis from normal bone, expanding the horizon of potential prostate MRF applications.[7]

Following these groundbreaking applications, several investigations were published looking at more technical aspects of prostate MRF. For instance, Sushentsev et al explored the feasibility of performing prostate MRF examinations at 1.5 T (tesla).[8] Though promising, multicenter studies with higher sample sizes are still required to validate the acquisition of prostate MRF maps using lower field scanners. Another study by Sushentsev et al raised questions about the impact of contrast administration on T1 relaxation times, which could potentially affect the ability of T1 maps to distinguish TZ lesions from normal TZ.[9] Nevertheless, a subsequent study from Lee et al, involving a larger cohort, showed that T1 and T2 relaxation times remained significantly lower for PCa compared with normal areas of the PZ and TZ after contrast administration.[10] This finding may pave the way for the development of truly quantitative delayed contrast-enhanced sequences for prostate imaging using MRF. Han et al also contributed to the advancement of prostate MRF by showing the feasibility of a 3D acquisition of prostate MRF maps.[11]

→ Continued on page 5

AUANews

January 2024, Volume 29 | Issue 1

PUBLICATIONS DEPARTMENT
American Urological Association
publications@AUAnet.org

PRODUCTION
KnowledgeWorks Global Ltd.
2905 Byrdhill Road, Richmond, VA 23228

ADVERTISING SALES
The Walchli Tauber Group
(443) 512-8899
stephen.tauber@wt-group.com

MAGNETIC RESONANCE FINGERPRINTING TECHNOLOGY FOR NONINVASIVE QUANTIFICATION OF PROSTATE CANCER

→ Continued from page 3

Table. Key Studies on Prostate Magnetic Resonance Fingerprinting

Study, y	Study design	Patient/volunteer population	N	Aim	Main conclusions
Correia et al 2023[13]	Retrospective, single center	Patients with suspected PCa that had PI-RADS 1-5 MRIs and MRF maps available	124 Patients	Comprehensively assess the distribution of MRF-derived T1 and T2 relaxation times of the whole normal-appearing PZ	The mean T1 relaxation time was 1941 ms, while the mean T2 relaxation time was 88 ms
Lo et al 2022[12]	Prospective, multicenter (5 different scanners across 3 institutions)	Patients with suspected PCa that underwent a prostate MRF	24 Patients	Investigate the multicenter reproducibility and repeatability of T1 and T2 relaxation times	Intrascanner variation was about 2% for T1 and 4.7% for T2. Interscanner variation between institutions was about 4.9% for T1 and 8.1% for T2
Lee et al 2022[10]	Retrospective, single center	Patients without a previous history of PCa that underwent a TRUS biopsy and had a prebiopsy contrast-enhanced prostate MRI with MRF acquisitions	57 Patients	Assess MRF-derived T1 and T2 relaxation times of noncontrast-enhanced and contrast-enhanced MRF for both the normal PZ and TZ and also for PCa	Median nonenhanced and contrast-enhanced T1 values were 1906 and 880 ms for the normal PZ, 1624 and 542 ms for the normal TZ, and 1510 and 605 ms for PCa. Median nonenhanced and contrast-enhanced T2 values were 180 and 186 ms for the normal PZ, 101 and 91 ms for the normal TZ, and 81 and 73 ms for PCa
Choi et al 2021[7]	Retrospective, single center	Patients with suspected PCa that had pelvic bone metastasis on MRI	30 Patients	Assess the feasibility of using MRF to evaluate PCa bone metastasis	ROIs of bone metastasis had significantly higher nonenhanced and contrast-enhanced T1 relaxation times and significantly lower nonenhanced and contrast-enhanced T2 relaxation times compared with benign bone
Han et al 2021[11]	Retrospective, single center	Patients with suspected PCa that underwent prostate MRF	90 Patients	Assess the feasibility of a 3D prostate MRF acquisition	T1 and T2 relaxation times obtained from a 3D prostate MRF acquisition had an excellent correlation with relaxation times obtained in the phantom study
Sushentsev et al 2021[8]	Prospective, single center	Volunteers without previous diagnosis or treatment for PCa	10 Healthy volunteers	Assess the reproducibility of MRF-derived T1 relaxation times from phantoms and healthy volunteers at both 1.5 T and 3 T field strengths	Mean T1 relaxation time was significantly higher at 3 T than at 1.5 T. There was a strong linear correlation between 1.5 T and 3 T T1 relaxation times. Interscanner agreement was acceptable for both 1.5 T and 3 T T1 mapping
Shiradkar et al 2021[6]	Retrospective, single center	Patients with suspected PCa that underwent a prostate MRF and were subsequently submitted to radical prostatectomy	14 Patients	Investigate the histopathological basis that justifies prostate MRF measurements for characterizing prostatitis and PCa	Areas of normal PZ, prostatitis, and PCa had different tissue composition ratios. There were significant correlations between different parameters of tissue composition ratio and T1 and T2 relaxation times
Sushentsev et al 2020[9]	Prospective, single center	Patients with MRI-visible biopsy-proven PCa on active surveillance	14 Patients	Evaluate the variation of T1 relaxation time after contrast administration	Mean T1 relaxation times before and after contrast administration were 2521 and 1270 ms for the normal PZ, 1753 and 724 ms for the normal TZ, and 1666 and 718 ms for PCa lesions. Contrast administration impaired the ability of T1 to differentiate TZ lesions from areas of normal TZ
Panda et al 2019[4]	Retrospective, single center	Patients with suspected PCa submitted to targeted biopsy that had a prebiopsy MRF	67 Patients	Investigate the role of MRF combined with clinical ADC mapping to characterize TZ lesions	The combination of MRF-derived T1 and standard ADC maps could differentiate TZ lesions from the normal TZ with an AUC of 0.94
Panda et al 2019[5]	Retrospective, single center	Patients with suspected PCa that were submitted to a targeted biopsy	85 Patients	Investigate if MRF-based T1 and T2 relaxation times in addition to conventional ADC mapping are able to distinguish PZ lesions from areas of benign prostatic tissue	MRF-derived T1 and conventional ADC maps could differentiate between PCa and negative biopsies with an AUC of 0.83
Yu et al 2017[3]	Retrospective, single center	Patients with suspected PCa that underwent systematic or targeted biopsies and had a prebiopsy MRF	131 Patients	Assess the role of MRF-derived T1 and T2 relaxation times combined with clinically available ADC maps to characterize prostatic diseases	Standard ADC combined with MRF-derived T2 maps had the highest AUC (0.83) to differentiate high- or intermediate-grade tumors from low-grade PCa. Mean T1, T2, and ADC values of PCa were significantly lower than those from the normal PZ

Abbreviations: ADC, apparent diffusion coefficient; AUC, area under the curve; MRF, magnetic resonance fingerprinting; MRI, magnetic resonance imaging; N, number; PCa, prostate cancer; PI-RADS, Prostate Imaging Reporting and Data System; PZ, peripheral zone; ROI, region of interest; T, tesla; TRUS, transrectal ultrasound; TZ, transition zone.

In preparation for larger-scale implementation of prostate MRF examinations, Lo et al conducted a study across 5 different scanners from 3 institutions in the United States and Brazil. Their findings demonstrated minimal intrascanner and interscanner variations in MRF-derived T1 and T2 relaxation times.[12] This underscores the repeatability and reproducibility of prostate MRF acquisitions across various scanners and centers. More recently, Correia et al

→ Continued on page 6

MAGNETIC RESONANCE FINGERPRINTING TECHNOLOGY FOR NONINVASIVE QUANTIFICATION OF PROSTATE CANCER

→ Continued from page 5

provided the first reference values for T1 and T2 relaxation times of the whole normal-appearing PZ across patients in different Prostate Imaging Reporting and Data System (PI-RADS) categories.[13] A comprehensive overview of key studies on prostate MRF is available in the Table.

Emerging Applications

Our institution has developed an MRF technique for the simultaneous acquisition of relaxometry and diffusion maps of the brain.[2] We are currently exploring the adaptation of this sequence for use in prostate imaging to generate repeatable, reproducible, and motion-robust ADC maps in addition to the already validated T1 and T2 maps. Our group is also examining the use of rigid coregistration between prostatectomy specimens and optimized prostate MRF acquisitions to improve the correlation between whole-mount histopathology and MRF maps, striving to get nearer to the concept of virtual pathology. Another promising application under investigation by our group involves the integration of prostate MRF to optimize existing biopsy decision-making workflows. Moreover, MRF holds the potential to improve the characterization of PI-RADS 3 lesions, which represent uncertainty and equivocal findings in which the current standard multiparametric MRI does not add value to biopsy decisions. MRF-derived T1 and T2 assessment can provide quantitative information to aid biopsy decision-making and potentially improve the management of the PI-RADS 3 subgroup. Additionally, leveraging radiomics to assess multiparametric voxel-wise quantitative data obtained with prostate MRF may unveil new promising applications in clinical practice.

Conclusions

Prostate MRF stands as a valuable quantitative MRI technique with proven clinical applications for noninvasive PCa detection and characterization. However, robust evidence supporting its widespread adoption in clinical practice is still needed. This gap may be attributed to the absence of large-scale, multi-institutional clinical studies supporting the benefit of integrating prostate MRF into clinical workflows. Future research should focus on the optimization and prospective large-scale validation of prostate MRF to facilitate its broader implementation in clinical practice. ■

1. Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature.* 2013;495(7440): 187-192.
2. Afzali M, Mueller L, Sakaie K, et al. MR fingerprinting with b-tensor encoding for simultaneous quantification of relaxation and diffusion in a single scan. *Magn Reson Med.* 2022;88(5):2043-2057.
3. Yu AC, Badve C, Ponsky LE, et al. Development of a combined MR fingerprinting and diffusion examination for prostate cancer. *Radiology.* 2017;283(3):729-738.
4. Panda A, Obmann VC, Lo WC, et al. MR fingerprinting and ADC mapping for characterization of lesions in the transition zone of the prostate gland. *Radiology.* 2019;292(3):685-694.
5. Panda A, O'Connor G, Lo WC, et al. Targeted biopsy validation of peripheral zone prostate cancer characterization with magnetic resonance fingerprinting and diffusion mapping. *Invest Radiol.* 2019;54(8):485-493.
6. Shiradkar R, Panda A, Leo P, et al. T1 and T2 MR fingerprinting measurements of prostate cancer and prostatitis correlate with deep learning-derived estimates of epithelium, lumen, and stromal composition on corresponding whole mount histopathology. *Eur Radiol.* 2021;31(3):1336-1346.
7. Choi MH, Lee S-W, Kim HG, et al. 3D MR fingerprinting (MRF) for simultaneous T1 and T2 quantification of the bone metastasis: initial validation in prostate cancer patients. *Eur J Radiol.* 2021;144:109990.
8. Sushentsev N, Kaggie JD, Slough RA, Carmo B, Barrett T. Reproducibility of magnetic resonance fingerprinting-based T1 mapping of the healthy prostate at 1.5 and 3.0 T: a proof-of-concept study. *PLoS One.* 2021;16(1):e0245970.
9. Sushentsev N, Kaggie JD, Buonincontri G, et al. The effect of gadolinium-based contrast agent administration on magnetic resonance fingerprinting-based T_1 relaxometry in patients with prostate cancer. *Sci Rep.* 2020;10(1):20475.
10. Lee YS, Choi MH, Lee YJ, Han D, Kim DH. Magnetic resonance fingerprinting in prostate cancer before and after contrast enhancement. *Br J Radiol.* 2022;95(1131):20210479.
11. Han D, Choi MH, Lee YJ, Kim DH. Feasibility of novel three-dimensional magnetic resonance fingerprinting of the prostate gland: phantom and clinical studies. *Korean J Radiol.* 2021;22(8):1332-1340.
12. Lo W, Bittencourt LK, Panda A, et al. Multicenter repeatability and reproducibility of MR fingerprinting in phantoms and in prostatic tissue. *Magn Reson Med.* 2022;88(4):1818-1827.
13. de Oliveira Correia ET, Qiao PL, Griswold MA, Chen Y, Bittencourt LK. Magnetic resonance fingerprinting based comprehensive quantification of T1 and T2 values of the background prostatic peripheral zone: correlation with clinical and demographic features. *Eur J Radiol.* 2023;164:110883.

ARTIFICIAL INTELLIGENCE

Leveraging Artificial Intelligence for Prehabilitation Interventions to Improve Perioperative Outcomes

Meghana Noonavath, BS
University of Washington, Seattle

Chris W. Lewis, MD
University of Washington, Seattle

Hanna Hunter, MD
University of Washington, Seattle
Fred Hutchinson Cancer Center, Seattle, Washington

Sarah P. Psutka, MD, MS
University of Washington, Seattle
Fred Hutchinson Cancer Center, Seattle, Washington

Prehabilitation—interventions developed to increase a patient's preoperative physiological reserve to help them cope with the stresses of surgery and recovery, improve postoperative outcomes, and facilitate return to function—is a relatively new field of study and a rapidly growing research area.[1-5] As prehabilitation interventions evolve, exciting new developments are being driven by the integration of artificial intelligence (AI)–powered technology.[6] With the advent of cutting-edge AI solutions such as wearable trackers, chatbots, and predictive modeling, patients and health care providers are witnessing a paradigm shift in the delivery of personalized health care (Figure). In this article, we will review how these technological advances can be leveraged for the delivery of prehabilitation-focused interventions and highlight anticipated challenges of AI integration, with particular attention to the field of urology.

Perhaps one of the most pressing challenges in prehabilitation is to ensure the acceptability and feasibility of exercise intervention completion, even in the most vulnerable patients. The lack of standardized

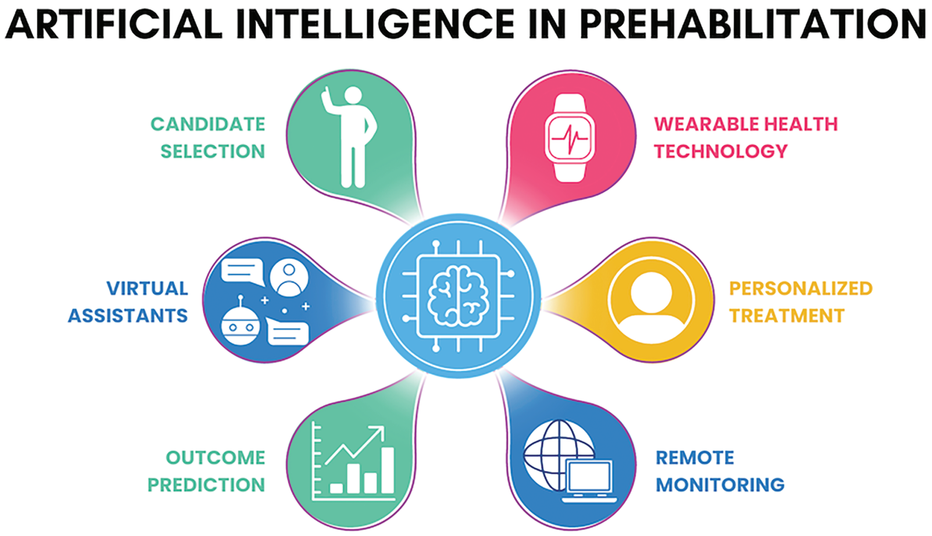


Figure. A visual map of common applications of artificial intelligence in prehabilitation, including selection of appropriate candidates, virtual assistants for patient/provider education, prediction of potential outcomes, remote monitoring of patient adherence and health metrics, personalized treatments for each individual patient, and wearable health technology for convenient patient monitoring and prompting.

→ Continued on page 7

LEVERAGING ARTIFICIAL INTELLIGENCE FOR PREHABILITATION INTERVENTIONS TO IMPROVE PERIOPERATIVE OUTCOMES

➔ Continued from page 6

prehabilitation protocols and the heterogeneity of patients and their disease course further complicate this issue,[7] requiring a robust solution based on large, diverse datasets. Modern AI technology utilizes large-language models and natural language processing to analyze vast amounts of disparate patient data—including but not limited to radiographic images, medical history, and surgical history—and capture complex, nonlinear relationships that may otherwise have been missed in such highly dimensional data. Using patterns obtained from this analysis, AI can personalize interventions according to patient characteristics and preferences[8] and characterize personalized risk profiles to anticipate patient outcomes.[9] For example, researchers utilized recurrent neural networks to predict eligibility for deep inspiration breath-hold radiotherapy treatment of patients with left breast cancer—these candidates were selected based on their functional capacity for breath holding, as well as anatomic and other clinical features.[10] The model produced a binary result, easily allowing providers to discern which patients would likely be able to successfully receive the treatment. Similarly, when it comes to selecting patients for prehabilitation, diverse data incorporating functional capacity and clinicopathologic features can be combined to tailor the most suitable adjuvant intervention for each patient's unique needs, to maximize successful treatment.

While candidate selection is essential, appropriate selection of the prehabilitation intervention is equally important. Traditional models of prehabilitation focus on multiple supervised exercise sessions, which demand resources (eg, physical facilities, equipment, and supervising trainers), are geographically limiting, and offer little flexibility.[11] Unsupervised sessions conducted at home, popularized during the COVID-19 pandemic, avert many of these issues but cannot ensure adequate adherence or intensity. Poor patient adherence, ineffective resource allocation and disparities in equity can all impede a successful prehabilitation intervention.[7] There is clearly a need for technology that monitors, reports, and responds to patients' physical activity. Some examples are already ubiquitous—wearable fitness trackers and smartphone applications aid the general public in tracking health metrics and achieving fitness goals. This technology has the power to respond dynamically to human behavior—for example, prompting one to stand up and walk around after a period of inactivity. Waller et al found that a prehabilitation program using a smartwatch and smartphone application for exercise and nutrition counseling resulted in increased exercise adherence and improved functional outcomes.[12] Wearables can go beyond simple tracking—some, such as a smart brace for joint issues, can provide dynamic feedback and appropriately adjust the type of activity being performed.[13] Others can automatically send tracker-obtained data to a health care team, allowing for remote patient monitoring without the need for in-person visits.[14,15] Tracker data such as patterns of active vs sedentary time or duration of tolerable exercise can be used to train algorithms on patterns of patient behavior to optimize adherence or develop interventions that are more likely to be sustainably adopted.

Following prehabilitation and surgery, AI-powered technology can continue to aid patients and physicians. The aforementioned preoperative predictive analytics can be applied postoperatively as well—AI-based remote monitoring can assist in the early detection of cancer recurrence or complications following treatment.[16] When patients undergoing radical cystectomy were given wearable activity monitors, lower postoperative step counts were associated with longer length of stay and higher rate of postdischarge readmissions.[17] This information can help medical teams prepare for potential challenges and help surgeons convey this personalized set of possible complications when obtaining informed consent. In urology, AI-powered technology has demonstrated benefit in patient education. Cakir et al found that ChatGPT, a natural language-processing platform that generates responses to prompts, was able to accurately and sufficiently answer more than 95% of potential questions about urolithiasis.[18] In this same way, one might apply generative AI under the supervision of urologists to help patients understand more about their medical condition, goals of prehabilitation, and care pathway.

AI also presents inherent challenges. Research studies often do not disclose their code, rendering algorithm assessment difficult. The opaqueness of neural networks, commonly referred to as the "black box" effect, hinders comprehension of their decision-making processes and raises concerns about potential biases. Additionally, questions arise concerning patient data ownership and confidentiality. Effective oversight is essential to ensure the use of high-quality data and equitable representation of diverse patient groups in algorithmic inputs. When integrating AI into health care, rigorous validation is imperative to mitigate the risks of harm and biased outcomes stemming from inadequately trained data. Upholding patient safety and data privacy is of paramount importance. Furthermore, it is vital to be mindful of the potential financial implications of AI technology—a comprehensive cost-benefit analysis is necessary for prudent resource allocation.[19]

AI-powered technology has the power to transform the field of prehabilitation for those undergoing major surgical or oncological interventions. It can help identify appropriate patients for prehabilitation, improve adherence via prompting, and serve as an informative tool for patients and physicians via digital health applications and chatbots. While the possibilities are exciting, there is much work to be done, from ascertaining validity and accuracy to ensuring lack of bias, before such technology can be widely utilized. For now, the push for AI technology in prehabilitation is a promising new frontier that is ready for exploration. ■

1. Durrand J, Singh SJ, Danjoux G. Prehabilitation. *Clin Med.* 2019;19(6):458-464.
2. Kaye DR, Schafer C, Thelen-Perry S, et al. The feasibility and impact of a presurgical exercise intervention program (prehabilitation) for patients undergoing cystectomy for bladder cancer. *Urology.* 2020;145:106-112.
3. Waterland JL, McCourt O, Edbrooke L, et al. Efficacy of prehabilitation including exercise on postoperative outcomes following abdominal cancer surgery: a systematic review and meta-analysis. *Front Surg.* 2021;8:628848.
4. Briggs LG, Reitblat C, Bain PA, et al. Prehabilitation exercise before urologic cancer surgery: a systematic and interdisciplinary review. *Eur Urol.* 2022;81(2):157-167.
5. Stout NL, Fu JB, Silver JK. Prehabilitation is the gateway to better functional outcomes for individuals with cancer. *J Cancer Rehabil.* 2021;4:283-286.
6. Yoon HK, Yang HL, Jung CW, Lee HC. Artificial intelligence in perioperative medicine: a narrative review. *Korean J Anesthesiol.* 2022;75(3):202-215.
7. Zhu H, Moffa ZJ, Gui X, Carroll JM. Prehabilitation: care challenges and technological opportunities. Paper presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 23, 2020; Honolulu, HI.
8. Piette JD, Newman S, Krein SL, et al. Patient-centered pain care using artificial intelligence and mobile health tools: a randomized comparative effectiveness trial. *JAMA Intern Med.* 2022;182(9):975-983.
9. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2(4):230-243.
10. Vendrame A, Cappelletto C, Chiovati P, et al. Artificial intelligence-based patient selection for deep inspiration breath-hold breast radiotherapy from respiratory signals. *Appl Sci.* 2023;13(8):4962.
11. Al-Naime K, Al-Anbuky A, Mawston G. Remote monitoring model for the preoperative prehabilitation program of patients requiring abdominal surgery. *Future Internet.* 2021;13(5):104.
12. Waller E, Sutton P, Rahman S, Allen J, Saxton J, Aziz O. Prehabilitation with wearables versus standard of care before major abdominal cancer surgery: a randomised controlled pilot study (trial registration: NCT04047524). *Surg Endosc.* 2022;36(2):1008-1017.
13. Riffitts M, Cook H, McClincy M, Bell K. Evaluation of a smart knee brace for range of motion and velocity monitoring during rehabilitation exercises and an exergame. *Sensors.* 2022;22(24):9965.
14. Al-Naime K, Al-Anbuky A, Mawston G. IoT based pre-operative prehabilitation program monitoring model: implementation and preliminary evaluation. Paper presented at: 4th International Conference on Biomedical Engineering (iBioMed); October 18-19, 2022; Yogyakarta, Indonesia.
15. Al-Naime K, Al-Anbuky A, Mawston GA. IoT for movement recognition and prehabilitation support system. Paper presented at: 27th Asia Pacific Conference on Communications (APCC); October 19-21, 2022; Jeju Island, Republic of Korea.
16. Guo C, Wang J, Wang Y, et al. Novel artificial intelligence machine learning approaches to precisely predict survival and site-specific recurrence in cervical cancer: a multi-institutional study. *Transl Oncol.* 2021;14(5):101032.
17. Khetrapal P, Bains PS, Jubber I, et al. Digital tracking of patients undergoing radical cystectomy for bladder cancer: daily step counts before and after surgery within the iROC randomised controlled trial. *Eur Urol Oncol.* 2023;10.1016/j.euo.2023.09.021.
18. Cakir H, Caglar U, Yildiz O, Meric A, Ayranci A, Ozgor F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *Int Urol Nephrol.* 2023;10.1007/s11255-023-03773-0.
19. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg.* 2020;155(2):148-158.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence Chatbots: How Accurate Is the Information?

Bristol B. Whiles, MD
University of Kansas Medical Center, Kansas City

Russell S. Terry, MD
University of Florida College of Medicine, Gainesville

Within a week of its release in November 2022, OpenAI's ChatGPT large language model (LLM) had already provided millions of internet users with unprecedented access to a transformative artificial intelligence (AI) tool.[1] While prior advanced LLMs had existed for some time, ChatGPT's novelty was its user-focused interface and conversational communication style—one no longer needed to have a background in computer programming to meaningfully use the technology. The ChatGPT chatbot platform is remarkable for its ability to provide confident-sounding answers to almost any query about almost any topic, including health care and urology. ChatGPT met the needs of our interconnected world, providing us with information that aligned with the persistent desire for instantaneous access.

Health care providers quickly began using this technology and testing its ability to perform rote tasks such as creating call schedules, drafting prior authorization letters to insurance companies, and formatting replies to messages in the electronic medical record. When a global cohort of 456 urologists was surveyed in April 2023 about their LLM use and experiences, 48% of them reported using LLMs in academic practice for tasks such as idea generation.[2] Nearly 20% of respondents also reported using ChatGPT in clinical practice, primarily for patient education applications. Interestingly, while equal numbers of respondents (29.6%) reported either trusting or not trusting LLMs to provide accurate information, 78% and 56% of respondents believed that LLMs could play an important role in academic and clinical practice, respectively. Considering the increased uptake among urologists, our patients are also likely to begin utilizing these AI tools. It is therefore important for us to understand the accuracy and limitations of current AI chatbot outputs, especially for urologic health care advice.

> "Considering the increased uptake among urologists, our patients are also likely to begin utilizing these AI tools."

ChatGPT (version 3.5) is the most studied AI chatbot model in our field (Table). Common response attributes evaluated by research

→ Continued on page 9

Table. Summary Table Containing Many of the Published Studies That Have Examined Large Language Models in Urology

Article	Large language model	Urologic topic evaluated	Scoring system	Accuracy, appropriateness, or correctness evaluation	Other info
Caution! AI Bot Has Entered the Patient Chat: ChatGPT Has Limitations in Providing Accurate Urologic Healthcare Advice[3]	ChatGPT (3.5) February 13, 2023	Nononcology	Brief DISCERN Yes or no for appropriateness	60% Appropriate	92.3% Had ≥1 incorrect, misinterpreted, or nonfunctional citation Quality content in 54% of responses
Evaluating the Effectiveness of Artificial Intelligence Powered Large Language Models Application in Disseminating Appropriate and Readable Health Information in Urology[4]	ChatGPT (3.5) February 28, 2023	Oncology Nononcology Emergency	5-Point Likert scale for appropriateness Utilized scoring aspects from DISCERN and QUEST tools	78% Appropriate 100% Accurate 61% Compressive	College reading level (Flesch Reading Ease of 35.5 ± 10.2 and Flesh-Kincaid Reading Grade Level 13.5 [SD = 1.74])
Quality of Information and Appropriateness of ChatGPT Outputs for Urology Patients[5]	ChatGPT (3.5) April-May 2023	Oncology Nononcology Emergency	5-Point Likert scale for accuracy, comprehensiveness, or appropriateness Section 2 of the DISCERN tool for information quality	52% Appropriate Nononcology > oncology or emergency questions (59% vs 53% vs 11%; *P* = .03) Poor quality via DISCERN assessment	College graduate reading level (Flesch Reading Ease of 18 [IQR = 21] and Flesch-Kincaid Reading Grade Level 15.8 [IQR = 3])
How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies?[6]	ChatGPT (3.5) Perplexity Chat Sonic Microsoft Bing AI April 10, 2023 (all)	Oncology	DISCERN	No appropriateness assessment 4 Out of 5 for quality of information	Moderate understandability (PEMAT-P of 66.7%), and actionability was moderate to poor (40%)
Evaluating the Performance of ChatGPT in Answering Questions Related to Urolithiasis[7]	ChatGPT (3.5) August 3, 2023	Nephrolithiasis	4-Point Likert scale	95% Completely correct	
Development of a Personalized Chat Model Based on the European Association of Urology Oncology Guidelines: Harnessing the Power of Generative Artificial Intelligence in Clinical Practice[8]	Uro_Chat ChatGPT (3.5 and 4.0)	Oncology	Yes or no for adequate response	100% Adequate	Custom model developed based on EAU guidelines; freely available online
ChatGPT and Most Frequent Urological Diseases: Analysing the Quality of Information and Potential Risks for Patients[9]	ChatGPT (4.0) March 18, 2023	Oncology Nononcology	DISCERN	Not assessed	

Abbreviations: EAU, European Association of Urology.

ARTIFICIAL INTELLIGENCE CHATBOTS: HOW ACCURATE IS THE INFORMATION?

→ Continued from page 8

groups have included appropriateness, comprehensiveness, clarity, and reproducibility. These assessments were performed using a variety of scoring instruments across the different studies. Appropriateness, although defined differently in different studies, is used by the researchers as a surrogate end point which combines an assessment of response accuracy, comprehensiveness, and clarity. Across studies, 52% to 78% of ChatGPT responses were deemed appropriate.[3-5] The most commonly reported reason for judging responses to be inappropriate was the absence of vital information, and response clarity was the least common reason.[4] Concerningly, the one study which examined response reproducibility within a single chatbot (ChatGPT version 3.5) by posing the same question in 3 independent chat instances found substantial response inconsistency—25% of question sets produced responses with dissimilar appropriateness ratings.[3] The LLM was also noted by multiple groups to provide responses at a difficult reading level, rated as college level or higher.[4-6] This body of research highlights that despite the hype and excitement about this new tool, the information currently provided by chatbots to urologic health care queries is generally of low quality and should not be considered actual, reliable medical advice.

> "Across studies, 52% to 78% of ChatGPT responses were deemed appropriate.[3-5] The most commonly reported reason for judging responses to be inappropriate was the absence of vital information, and response clarity was the least common reason.[4]"

Figure. The Chatbot will see you now. This image was created by the authors using ChatGPT 4 with DALL-E plugin, October 16, 2023 version.

One of the major limitations of the current body of research on LLMs in health care is that there does not currently exist a validated assessment instrument by which the models' responses can be evaluated in a standard and reproducible fashion. Researchers so far have designed their own Likert scales or adapted existing evaluation tools which were originally designed for other purposes (eg, DISCERN, Brief DISCERN, QUEST). Given the rapid adoption and the degree to which LLMs have already become entrenched within society, as well as the potential for real harm from low-quality responses to potentially serious medical questions, there is an urgent need for the development of a standardized evaluation tool that is specifically designed to assess chatbot responses to medical questions. Such a tool will be critical for the performance and interpretation of high-quality future research in this emerging field.

In summary, chatbot responses to urologic health care queries are sometimes appropriate. While they are often written in clear language, they frequently provide information which is either not factual or not comprehensive. Responses are also limited in their practicality by being written at a very high reading level and often lacking actionable instructions for their end users, who may be patients. Nonetheless, it is increasingly apparent that AI and chatbots are here to stay (Figure). We as a field can choose to ignore the proverbial elephant in the room, or we can choose to participate meaningfully in their evaluation and evolution to ensure that future iterations of these tools wield their considerable computational power for healing and not for harm. ■

> "Responses are also limited in their practicality by being written at a very high reading level and often lacking actionable instructions for their end users, who may be patients."

1. Roose K. The brilliance and weirdness of ChatGPT. *New York Times.* December 5, 2022:B1.
2. Eppler M, Ganjavi C, Ramacciotti LS, et al. Awareness and use of ChatGPT and large language models: a prospective cross-sectional global survey in urology. *Eur Urol.* 2023;10.1016/j.eururo.2023.10.014.
3. Whiles BB, Bird VG, Canales BK, DiBianco JM, Terry RS. Caution! AI bot has entered the patient chat: chatGPT has limitations in providing accurate urologic healthcare advice. *Urology.* 2023;180:278-284
4. Davis R, Eppler M, Ayo-Ajibola O, et al. Evaluating the effectiveness of artificial intelligence-powered large language models application in disseminating appropriate and readable health information in urology. *J Urol.* 2023;210(4):688-694.
5. Cocci A, Pezzoli M, Lo Re M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate Cancer Prostatic Dis.* 2023;10.1038/s41391-023-00754-3.
6. Musheyev D, Pan A, Loeb S, Kabarriti AE. How well do artificial intelligence chatbots respond to the top search queries about urological malignancies?. *Eur Urol.* 2023;10.1016/j.eururo.2023.07.004.
7. Cakir H, Caglar U, Yildiz O, Meric A, Ayranci A, Ozgor F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *Int Urol Nephrol.* 2023;10.1007/s11255-023-03773-0.
8. Khene Z-E, Bigot P, Mathieu R, Rouprêt M, Bensalah K. Development of a personalized chat model based on the European Association of Urology oncology guidelines: harnessing the power of generative artificial intelligence in clinical practice. *Eur Urol Oncol.* 2023;10.1016/j.euo.2023.06.009.
9. Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, Diaz Goizueta FJ, García Tello A, Llanes González L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. *World J Urol.* 2023;41(11):3149-3153.

ARTIFICIAL INTELLIGENCE

European Perspective on Importance of Artificial Intelligence Lectures and Sessions in Urology Meetings

Kari A. O. Tikkinen, MD, PhD
University of Helsinki and Helsinki University Hospital, Finland
South Karelian Central Hospital, Lappeenranta, Finland

Jochen Walz, MD
Institut Paoli-Calmettes Cancer Centre, Marseille, France

Artificial intelligence (AI) encompasses a broad spectrum of technologies, including machine learning, natural language processing, and computer vision, which equip machines with the ability to learn, reason, and make decisions in a manner that mimics human cognition. AI is dedicated to developing systems and technologies that can execute tasks traditionally requiring human intellect.

In the world of urology, AI-focused lectures and workshops at professional gatherings are essential for disseminating knowledge, fostering collaboration, and advancing education. At the upcoming Annual Congress of the European

→ Continued on page 10

EUROPEAN PERSPECTIVE ON IMPORTANCE OF ARTIFICIAL INTELLIGENCE LECTURES AND SESSIONS IN UROLOGY MEETINGS

→ Continued from page 9

Association of Urology (EAU) in Paris, France, next spring, the AI session will once again be a cornerstone of our congress program. For the fourth consecutive year, our thematic AI session is designed to empower urologists and other professionals to leverage AI's capabilities to enhance patient care and drive innovation within the field.

To date, EAU Annual Congress AI sessions have covered several critical topics, such as:

1. AI foundations in urology: introducing the core principles of AI and its specific applications in urology, providing a solid knowledge foundation for participants
2. Imaging and diagnostics: discussing a breadth of subjects from computational imaging to renal mass imaging and the support AI offers in imaging and pathology
3. AI in prostate cancer management: presenting AI's utilization in diagnosing and treating prostate cancer, including screening, imaging, and Gleason grading
4. Ethical and patient considerations: addressing ethical implications in AI use and considering patient viewpoints on AI, underscoring the necessity for responsible and patient-focused AI applications
5. Standardization and reporting: emphasizing the need for standardizing AI/machine learning applications and the critical nature of transparent and accurate reporting of AI findings
6. Surgical planning and guidance: exploring how AI, augmented reality, and robotics can assist in surgical planning, guidance, and implementation

> "In the world of urology, AI-focused lectures and workshops at professional gatherings are essential for disseminating knowledge, fostering collaboration, and advancing education."

Figure 1. Illustration generated by ChatGPT 4.0 with DALL·E 3 in response to a request for a figure to accompany the original title of this article ("Enhancing Urologists' Knowledge: A European Perspective on the Importance of Artificial Intelligence Lectures and Sessions in Urology Meetings").

Figure 2. Revised illustration (of Figure 1) generated by ChatGPT 4.0 with DALL·E 3, modified at the authors' request to remove the uterus and ovaries, and instead, to emphasize the male urinary tract in the context of an artificial intelligence lecture at a urology conference.

7. Challenges and limitations of AI: discussing the importance of recognizing the limitations of observational AI data and the instances when randomized trials are necessary for trustworthy evidence

In the forthcoming EAU24 congress, participants in the AI session will gain insights into:

1. The role of AI in decision support, data analytics, and predictive tools
2. The application of AI in diagnostic support, with a focus on imaging and pathology
3. The integration of AI into surgical planning via augmented reality and robotics for improved surgical accuracy
4. The specific functions of AI, particularly ChatGPT, in patient education and assistance
5. The involvement of AI and ChatGPT in grant proposals, scientific publications, and conference presentations

Beyond the EAU meetings, we are delighted to observe an increasing presence of AI sessions also at numerous other urological conferences worldwide (Figures 1 and 2), such as the AUA Annual Meeting. AI's significance is on the rise within urology and beyond, indicating that AI sessions are set to be a permanent feature at future EAU congresses.

> "For the fourth consecutive year, our thematic AI session is designed to empower urologists and other professionals to leverage AI's capabilities to enhance patient care and drive innovation within the field."

Acknowledgments

ChatGPT with DALL·E 3 contributed to the drafting and revising of this manuscript. ■

ARTIFICIAL INTELLIGENCE

Artificial Intelligence–Powered Imaging for Renal Cancer and Nephrometric Scoring Automation

Andrew M. Wood, MD
Glickman Urological and Kidney Institute, Cleveland Clinic, Ohio

Nour Abdallah, MD
Glickman Urological and Kidney Institute, Cleveland Clinic, Ohio

Rebecca A. Campbell, MD
Glickman Urological and Kidney Institute, Cleveland Clinic, Ohio

Christopher J. Weight, MD
Glickman Urological and Kidney Institute, Cleveland Clinic, Ohio

Since the launch of ChatGPT in November 2022, public awareness of the capabilities and potential uses of artificial intelligence (AI) and machine learning (ML) has increased exponentially. In tandem with this public explosion, a similarly meteoric rise in biomedical research involving AI or ML components has occurred. In the 5 years from 2017 to 2022, the number of PubMed articles including the terms "artificial intelligence" or "machine learning" increased from 10,156 to 51,995.[1] Within the medical field, AI can be used to interpret radiologic, endoscopic, and histologic images; analyze a patient's disease, comorbidities, and other components of medical care such as previous treatments, number of visits, potential side effects, and costs; recommend management strategies; and predict outcomes.[2] Within the scope of urologic oncology, we have found some of the most impressive applications lie within the realm of kidney tumor imaging.

The first step in any fully automated process involving imaging of kidney tumors is to create an algorithm that can reliably and accurately differentiate a malignant tumor from the surrounding renal parenchyma and hilar structures. Segmentation, as this process is often referred to, is the bedrock of any future automated quantitative examinations of the radiologic features of kidney tumors. In order to mobilize international interest and effort toward this lofty goal, we launched segmentation challenges, including the 2019, 2021, and 2023 Kidney Tumor Segmentation Challenges (KiTS). In KiTS 2019, 106 international teams used a public training set of 210 cross-sectional CT images with kidney tumors and corresponding hand-drawn semantic segmentation masks (generated by human annotators) to develop automated systems predicting the segmentation masks of 90 test CT images (Figures 1 and 2). The winning model achieved a Sørensen-Dice coefficient of 0.974 for the segmentation of the kidney and 0.851 for the tumor, nearing the interannotator agreement for both the kidney (0.983) and the tumor (0.923).[3] KiTS 2021 was a sequel to this, with an innovative challenge design and a larger dataset. The highest-ranked teams performed better than those of the KiTS 2019 challenge, with respective scores achieving even closer to the human-level performance.[4] KiTS 2023 is ongoing and is the first KiTS to incorporate nonarterial contrast phases into both the training and test sets, hopefully resulting in broader applicability.

As AI-generated segmentation has garnered more interest and supporting evidence, we've found ourselves running into one of the biggest obstacles to the broader adoption of AI in the clinical setting: the so-called "black box" issue.[5] Physicians are reluctant to trust AI in their practice as they lack an understanding of the processes underpinning the ML algorithms. Similarly, patients are skeptical of AI-based technologies, and while they may tolerate human errors, AI errors are often more difficult to accept. Because of these issues, increasing comprehension and trust in AI algorithms are just as important as the development of the algorithms themselves. To combat this, our team elected to adopt a stepwise approach to opening the AI black box in kidney tumor imaging. Small steps were taken to explain and familiarize the use of AI in kidney cancer to build trust and acceptance.

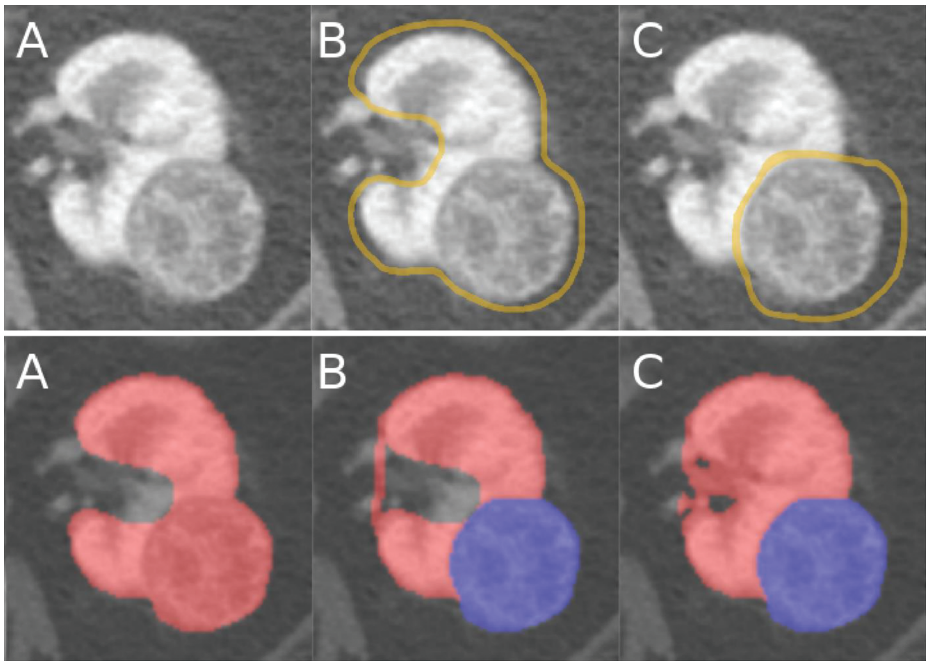


Figure 1. Illustration of hand-drawn ground truth kidney + tumor segmentation (top, B and C) and artificial intelligence-generated segmentation masks (bottom, B and C) for a single patient with a renal mass.[3]

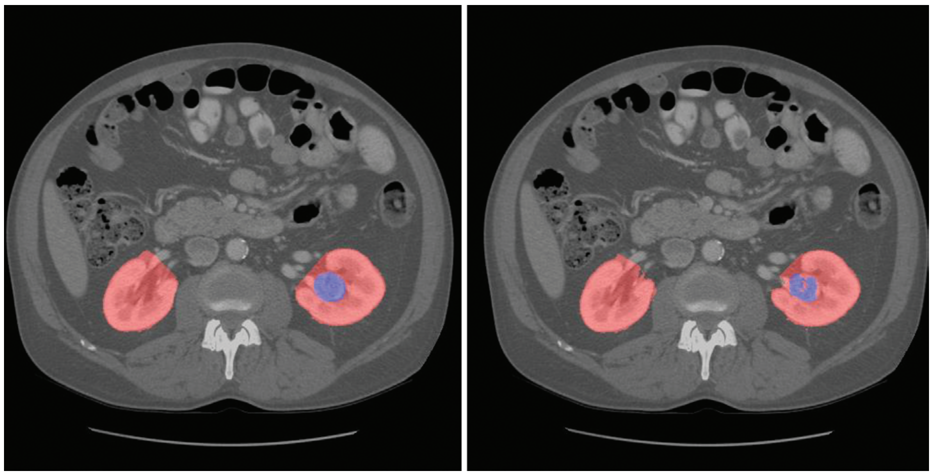


Figure 2. Comparison of human-generated (left) and artificial intelligence–generated (right) segmentation mask of kidneys and left central kidney tumor (Sørensen-Dice score 0.92).

Our first goal in establishing trust in AI-based kidney tumor segmentation was to replicate known and trusted clinical tumor scoring systems using the publicly available KiTS 2019 data and the winning segmentation model. Thus, we automated the R.E.N.A.L. (for radius, exophytic/endophytic, nearness of tumor to collecting system, anterior/posterior, location relative to polar line) nephrometry score on 300 preoperative CT scans and created an AI-generated score. The fully automated AI R.E.N.A.L. score was able to predict meaningful patient-centered and oncologic outcomes with similar predictive utility to human-generated scores, including presence of malignancy, presence of necrosis, high-grade disease, and high-stage disease.[6]

Building on these results, our next goal was to demonstrate how AI/computer-aided systems might enhance, rather than simply replicate, existing models. One major limitation of the R.E.N.A.L. score is the categorical, unweighted nature of its components. In other words, the R.E.N.A.L. score categorizes its variables, which would otherwise be

→ Continued on page 12

ARTIFICIAL INTELLIGENCE–POWERED IMAGING FOR RENAL CANCER AND NEPHROMETRIC SCORING AUTOMATION

➔ Continued from page 11

continuous in nature (radius, endophycity, nearness to the collecting system, and the longitudinal location) in order to simplify the calculation. Thus, we also used the KiTS data to examine a computer-aided transition to continuous instead of categorical variables.

In this study, we explored the ability of continuous versions of R.E.N.A.L. components generated from fully automated AI-based segmentation to predict oncologic outcomes of patients with renal mass. The oncologic predictions made by our AI+ score (AI segmentations, continuous variables) surpassed those of the AI-generated (AI segmentations, categorical variables) and the human expert–generated R.E.N.A.L. scores.[7] Notably, calculations with the AI+ score can be performed without human intervention at any step, given the complete automation of the generation of segmentation masks, measurement of continuous R.E.N.A.L. components, and weighting and combination of components into a multivariate model, each of which is a complex and time-consuming step for humans.

Our hope is that our work provides reliable and relatable evidence for trusting fully automated versions of previously understood scoring systems, while at the same time promising benefits to nephrometry score utilization in both clinical practice and research settings. Our study is also an essential intermediary step in developing and implementing more complex ML-based radiomic scoring systems that cannot be realistically calculated by humans.

In more broadly looking to the future of AI within clinical care, we hope that AI can help to resolve inherent biases and noise within clinician decision-making. Although AI does not follow common sense and may be biased, it can be reprogrammed. AI is faster, more reproducible, continuously adapting, and with less noise. Of course, how AI tools are developed and implemented will remain crucial to their safety and reliability. It is up to urologists and urologic researchers to maintain an active involvement in the development and validation of AI algorithms to ensure their appropriate use within the realm of urology.[8] ■

1. PubMed, National Library of Medicine. *"(Artificial Intelligence) OR (Machine Learning)"* (search term). 2023. https://pubmed.ncbi.nlm.nih.gov/
2. Rasmussen R, Sanford T, Parwani AV, Pedrosa I. Artificial intelligence in kidney cancer. *Am Soc Clin Oncol Educ Book.* 2022;42:300-310.
3. Heller N, Isensee F, Maier-Hein KH, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med Image Anal.* 2021;67:101821.
4. Heller N, Isensee F, Trofimova D, et al. The KiTS21 challenge: automatic segmentation of kidneys, renal tumors, and renal cysts. *arXiv.* 2023;10.48550/arXiv.2307.01984.
5. Poon AIF, Sung JJY. Opening the black box of AI-medicine. *J Gastroenterol Hepatol.* 2021;36(3):581-584.
6. Heller N, Tejpaul R, Isensee F, 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(5):1105-1115.
7. Abdallah N, Wood A, Benidir T, et al. AI-generated R.E.N.A.L.+ score surpasses human-generated score in predicting renal oncologic outcomes. *Urology.* 2023;180:160-167.
8. Heller N, Weight C. "The algorithm will see you now": the role of artificial (and real) intelligence in the future of urology. *Eur Urol Focus.* 2021;7(4):669-671.

ARTIFICIAL INTELLIGENCE

ChatGPT in Medical Education: Teaching Optimization Not Conceptualization

Michael Tradewell, MD
Desai Sethi Urology Institute, University of Miami, Florida

Justin Dubin, MD
Memorial Healthcare System, Aventura, Florida

ChatGPT (Generative Pretrained Transformer) is an artificial intelligence (AI) large language model developed to mimic human conversation. The AI tool is able to complete a wide variety of complex tasks (eg, writing original music, language translation, or debugging computer code). While ChatGPT was not specifically designed for medicine, it made headlines when it passed the USMLE (United States Medical Licensing Examination) Step 1, 2, and 3 licensing exams.[1] Unsurprisingly, urologists have been eager to assess how ChatGPT can impact medical education in our field.

ChatGPT did not perform as well on the AUA Self-Assessment Study Program as it did on the USMLE exams. It answered only 42.3% and 30% of the 2021 and 2022 questions correctly, respectively. The AI performed very well on first-order information recall questions and worse on questions with multiple reasoning steps. This result is unsurprising as the AI chatbot was trained on massive amounts of writing on the internet and questions requiring high-order reasoning introduce more variables and increased chances for the AI to incorporate relevant but incorrect information. The authors found that ChatGPT provided a "highly logical and coherent rationale for its answer choice, regardless of whether it chose the correct answer."[2] Nonetheless, this calls into question the readiness of ChatGPT as a primary teaching tool in urology.[3]

More interestingly, ChatGPT is not static. It can be trained. Researchers have built a ChatGPT trained on the 2023 European Association of Urology guidelines. This custom AI responded with more concise and precise answers compared to the untrained ChatGPT.[4] Future custom GPTs may be built and validated as education and clinical decision support tools.

While we are happy to report that urologists won't yet be replaced by AI in clinical decision-making, for now, like any new technology, the true potential for ChatGPT has yet to be realized. In terms of medical education, ChatGPT poses a huge upside for medical students, residents, fellows, and attending physicians. There are multiple tasks ChatGPT can do to elevate the medical education experience including teaching assistance, personalized learning with materials and study plans, research assistance, content creation, documentation, and patient interactions.[5] Just like any complex tool, proper usage must be taught. While research on ChatGPT is still in its adolescence, one of the most important aspects of optimization of its use comes from understanding how to appropriately prompt and direct the AI. Studies show the accuracy and efficiency of the desired outcome depends on the user. With so many potential ways to utilize this technology, medical schools and residency programs need to focus on teaching trainees how to elicit the appropriate responses from ChatGPT to properly achieve their goal.[3,4] Similar to how medical students learn to properly perform a PubMed search or statistics, ChatGPT optimization should be taught in the classroom. By providing students with the right tools, they can best take full advantage of ChatGPT to improve their knowledge and, at the end of the day, improve patient care. ■

1. ChatGPT passed the USMLE. What does it mean for med ed?. American Medical Association. March 3, 2023. Accessed October 31, 2023. https://www.ama-assn.org/practice-management/digital/chatgpt-passed-usmle-what-does-it-mean-med-ed
2. Deebel NA, Terlecki R. ChatGPT performance on the American urological association self-assessment study program and the potential influence of artificial intelligence in urologic training. *Urology.* 2023;177:29-33.
3. Gupta R, Pedraza AM, Gorin MA, Tewari AK. Defining the role of large language models in urologic care and research. *Eur Urol Oncol.* 2023;S2588-9311(23)00162-1.
4. Manolitsis I, Feretzakis G, Tzelves L, et al. Training ChatGPT models in assisting urologists in daily practice. *Stud Health Technol Inform.* 2023;305:576-579.
5. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci.* 2023;39(2):605-607.

ARTIFICIAL INTELLIGENCE

Shifting From Dr Google to Dr GPT: The Potential Impact on Patient Safety of Changing e-Providers

Stacy Loeb, MD, MSc, PhD (Hon)
New York University Langone Health and Manhattan Veterans Affairs, New York

Over 90% of US adults use the internet, and there is a substantial amount of online content about health topics. Unfortunately, substantial limitations have been identified with urological information on the internet. Major problems include insufficient content presented at the recommended 6th grade reading level for consumer health information and a high prevalence of circulating misinformation. For example, in a series of studies evaluating information about prostate cancer on Instagram and TikTok, we reported misinformation in 40% to 41% of the content containing objective information.[1,2] Even the websites of National Cancer Institute–designated cancer centers, on average, provided sufficient information to answer only 19% of key questions for prostate cancer decision-making.[3] These issues are not unique to prostate cancer, with studies showing a substantial amount of poor-quality content about a range of benign and malignant urological conditions across different online platforms.[4-6] This leaves a lot of room for improvement from the "care" that Dr Google has been providing our patients to date.

The key question is whether Dr ChatGPT can improve upon this and provide better advice to our patients. Our group has published several studies on the quality of consumer health information from ChatGPT and other artificial intelligence (AI) chatbots. First, we examined information about the most common urological cancers (prostate, bladder, kidney, and testicular cancer) from ChatGPT, Perplexity, Chat Sonic, and Microsoft Bing AI.[7] Using the top 5 Google search queries about

→ Continued on page 14

> "These preliminary findings suggest that ChatGPT and other AI chatbots may provide higher quality information than many other online sources and appear less likely to spread misinformation."

SHIFTING FROM DR GOOGLE TO DR GPT: THE POTENTIAL IMPACT ON PATIENT SAFETY OF CHANGING E-PROVIDERS

➔ Continued from page 13

each cancer as prompts, we found that AI chatbot responses were generally good quality (median score of 4 out of 5 on the validated DISCERN instrument) and lacked misinformation. However, actionability of the responses was poor (median actionability score of 40% out of 100% on the validated Patient Education Materials Assessment Tool [PEMAT] instrument).

Using similar methods, we compared information from the 4 AI chatbots (ChatGPT, Perplexity, Chat Sonic, and Microsoft Bing AI) related to the most common 5 cancers in the US (skin, lung, breast, colorectal, and prostate cancer).[8] The top 5 Google search queries about each cancer were used as prompts. The quality of text responses generated by the AI chatbots was high (median score of 5 out of 5 on the validated DISCERN instrument); however, actionability was poor (median score of 20% out of 100% on the validated PEMAT instrument) and responses were written at a college reading level.

More recently, we examined information about erectile dysfunction from ChatGPT, Perplexity, Chat Sonic, and Microsoft Bing AI.[9] Using the top 5 Google search queries and headings from the National Institute of Diabetes and Digestive and Kidney Diseases website as inputs, we found that the quality of information was high (median score of 4 out of 5 on the validated DISCERN instrument) but actionability was low (median score of 20% out of 100% on the validated PEMAT instrument) and responses were written at a median Flesch-Kincaid grade level of 14.

Similarly, Davis et al examined the responses of ChatGPT to 18 patient questions about signs/symptoms or treatment for benign, oncologic, and emergency urology topics.[10] Overall, the majority of responses (77.8%) were deemed appropriate; however, the information was presented at a mean grade level of 13.5.

These preliminary findings suggest that ChatGPT and other AI chatbots may provide higher quality information than many other online sources and appear less likely to spread misinformation. However, the information is not readily actionable and is written above the recommended reading level for consumer health information. Therefore, we have yet to identify the optimal "e-provider" with high-quality information that is also actionable and understandable for lay health consumers. In the meantime, it is prudent to provide patients with a list of vetted resources of additional information about their condition. ■

1. Xu AJ, Taylor J, Gao T, Mihalcea R, Perez-Rosas V, Loeb S. TikTok and prostate cancer: misinformation and quality of information using validated questionnaires. *BJU Int.* 2021;128(4):435-437.
2. Xu AJ, Myrie A, Taylor JI, et al. Instagram and prostate cancer: using validated instruments to assess the quality of information on social media. *Prostate Cancer Prostatic Dis.* 2022;25(4):791-793.
3. Dulaney C, Barrett OC, Rais-Bahrami S, Wakefield D, Fiveash J, Dobelbower M. Quality of prostate cancer treatment information on cancer center websites. *Cureus.* 2016;8(4):e580.
4. Loeb S, Taylor J, Borin JF, et al. Fake news: spread of misinformation about urological conditions on social media. *Eur Urol Focus.* 2020;6(3):437-439.
5. Dubin JM, Aguiar JA, Lin JS, et al. The broad reach and inaccuracy of men's health information on social media: analysis of TikTok and Instagram. *Int J Impot Res.* 2022;1-5.
6. Kanner J, Waghmarae S, Nemirovsky A, Wang S, Loeb S, Malik R. TikTok and YouTube videos on overactive bladder exhibit poor quality and diversity. *Urol Pract.* 2023;10(5):493-500.
7. Musheyev D, Pan A, Loeb S, Kabarriti AE. How well do artificial intelligence chatbots respond to the top search queries about urological malignancies?. *Eur Urol.* 2023; S0302-2838(23)02972-X.
8. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. *JAMA Oncol.* 2023;9(10):1437-1440.
9. Pan A, Musheyev D, Loeb S, Kabarriti AE. Quality of erectile dysfunction information from ChatGPT and other artificial intelligence chatbots. *BJU Int.* 2023;10.1111/bju.16209.
10. Davis R, Eppler M, Ayo-Ajibola O, et al. Evaluating the effectiveness of artificial intelligence–powered large language models application in disseminating appropriate and readable health information in urology. *J Urol.* 2023;210(4):688-694.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence and Residency Applications: Is This the End of the Personal Statement?

Adam B. Weiner, MD
David Geffen School of Medicine, University of California, Los Angeles

Jonathan Bergman, MD, MPH
David Geffen School of Medicine, University of California, Los Angeles

The urology application process has evolved rapidly over the past few years. Formerly, United States Medical Licensing Examination Step 1 scores were used by some programs to screen applicants for interviews and were considered among the most important factors for assessing candidates. Amid concerns for the burden of exam preparation and desire for a more holistic application process, Step 1 was made a pass/fail exam in January 2022. During the COVID pandemic, interviews were changed from in person to 100% virtual, primarily to reduce virus transmission. However, the virtual requirement was maintained to reduce the financial burden to applicants and ensure equity among the applicant pool. More recently, preference signaling has received positive feedback from both programs and applicants with the aim of providing an equitable system for students to demonstrate programmatic interest in place of other methods which might drive socioeconomic disparities (eg, away rotations and mentor or home department connections).

> "Personal statements leveraging similar AI language models may start to seem homogeneous—potentially diminishing the worth of personal statements to programs trying to discern applicant fit."

These changes were reactions to improve the application process. Most were intended to help encourage a holistic approach to ranking. With less focus on exam scores and the potential loss of interpersonal connections with virtual interviews, greater attention will be paid towards other application components that can highlight the individual traits of each student, including letters of recommendation and the personal statement. The personal statement, in particular, is poised to grow in importance. In the

Figure. Artificial intelligence (AI)–generated image of AI writing a personal statement.

➔ Continued on page 15

AI AND RESIDENCY APPLICATIONS: IS THIS THE END OF THE PERSONAL STATEMENT?

→ Continued from page 14

past, this part of the application varied from incredibly intimate and telling of an applicant's fit to "cookie cutter" with repetitive platitudes. With the changes noted above, the personal statement stands out as an applicant's opportunity to showcase their journey to medicine and urology in a format that can be honed into their voice.

However, as the application process changed to accommodate the COVID pandemic and augment equity, so too will interpretation of personal statements with the arrival of artificial intelligence (AI) language models such as ChatGPT. ChatGPT has already been shown to produce credible personal statements for residency applications that can be indistinguishable from those written by humans (Figure).[1,2] With this new tool comes the need to forecast potential pros and cons so that our specialty can adapt appropriately.

Pro 1: Access to tools such as ChatGPT is equitable. Thus, resources spent on writing support for personal statements and access to mentors or other contacts who can assist with the writing process will play less of a role in creating disparities in the application process.

Con 1: Personal statements leveraging similar AI language models may start to seem homogeneous—potentially diminishing the worth of personal statements to programs trying to discern applicant fit. This could cause a subsequent need for programs to rely more on objective measures of applicant quality such as medical school reputation, further disadvantaging certain applicants. If students are going to use AI to write their entire personal statements, we might as well just call them statements.

Pro 2: Medical students applying to subspecialties are usually extremely busy at the time of application drafting with finishing clerkships, away rotations, and electives. AI language models can help medical students, and many others, save time initiating the writing process. This time can then be leveraged for balancing the many expectations of medical students transitioning from the final year of school. Importantly, AI should be used in this capacity to generate ideas or narratives that might be useful with the applicant only drafting what is actually personally relevant to them (see **Con 1**).

Con 2: Everyone is prone to mistakes using AI language models. Recently, a peer-reviewed publication was retracted after clear, accidental evidence of copying and pasting from ChatGPT was noted.[3] This mistake was not caught during the peer-review process or copyediting. With so many applicants every year, these mistakes are likely to manifest in awkward ways in personal statements that could greatly jeopardize an applicant's odds of matching—much worse than an honest typographical error.

Pro 3: International medical graduates, or those who otherwise might benefit from English language support, could leverage AI language models to improve writing clarity. This can help level the disadvantage felt by worthy applicants for whom English was not their first learned language, for instance. As noted in **Con 1**, however, these tools should be used only to the extent to prevent misrepresenting oneself in written words. Otherwise, if all students use the most common language models for grammar such as Grammarly, we might encounter issues with homogenizing individual written voices.

Con 3: Essay materials generated by AI might seem appealing to applicants to include, even if the words don't truly represent the applicant. This could lead to inaccurate judgment of fit by programs, which could harm chances of interview offers at good fit programs or increase the likelihood of offers from poor fit programs. Again, personal statements should be kept personal.

When it comes to personal statements, many are familiar with an 8:1:1 ratio: 8 out of 10 personal statements are "OK" and don't really hurt or help an applicant's appeal; 1 out of 10 essays tend to decrease an applicant's ratings; and 1 out of 10 statements help the applicant's ranking. While this ratio might vary somewhat, we suggest that, going forward, both applicants and programs view the personal statements as a tool to help communicate individual passions. Applicants should take pride in these written works; this alone can prevent many of the issues noted above.

There is little doubt AI will change the landscape of urology residency personal statements. Although AI-driven technologies are improving objectivity, efficiency, and data-driven decision-making, it is unlikely that they will entirely replace the personal statement or the human component in the selection process. Guidance regarding appropriate and ethical use of AI should be taught to medical students. Program directors, applicants, and the medical education community should carefully manage these changes as the landscape evolves to guarantee a fair and comprehensive evaluation of aspiring physicians. Ultimately, we should continue to strive for the personal statement to remain personal. ■

1. Johnstone RE, Neely G, Sizemore DC. Artificial intelligence software can generate residency application personal statements that program directors find acceptable and difficult to distinguish from applicant compositions. *J Clin Anesth.* 2023;89:111185.

2. Patel V, Deleonibus A, Wells MW, Bernard SL, Schwarz GS. The plastic surgery residency application in the era of ChatGPT: a personal statement generated by artificial intelligence to statements from actual applicants. *Ann Plast Surg.* 2023;91(3):324-325.

3. Conroy G. Scientific sleuths spot dishonest ChatGPT use in papers. *Nature.* 2023;10.1038/d41586-023-02477-w.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Renal Cell Carcinoma Histology

Riccardo Bertolo, MD, PhD
Urology Unit, University of Verona, Italy

Anna Caliò, MD, PhD
Section of Pathology, University of Verona, Italy

Matteo Brunelli, MD, PhD
Section of Pathology, University of Verona, Italy

Guido Martignoni, MD
Section of Pathology, University of Verona, Italy

Alessandro Antonelli, MD
Urology Unit, University of Verona, Italy

Morphological analysis, including the determination of renal cell carcinoma (RCC) histotype, tumor grade, presence of lymphovascular invasion, tumor necrosis, and sarcomatoid dedifferentiation, is pivotal for RCC diagnosis. It not only defines prognosis but also predicts the impact of eventual systemic treatments. In contemporary practice, this analysis must be complemented by genetic and cytogenetic assessments.[1] However, RCC histological diagnosis and classification can pose challenges due to its encompassment of a diverse range of histopathological entities, which have recently undergone revisions.[2]

Over the years, the diagnostics in RCC have evolved through the integration of modern counterparts such as electronic health records, digitalized radiology, and virtual pathology. This evolution has generated a huge amount of data, which can be processed using characterization algorithms or artificial intelligence (AI).[3]

The use of AI in RCC histopathology, known as pathomics or

→ Continued on page 16

ARTIFICIAL INTELLIGENCE IN RENAL CELL CARCINOMA HISTOLOGY

➔ Continued from page 15

computational pathology, is relatively new. AI can assist the pathologist in improving efficiency, accessibility, cost-effectiveness, and time consumption. It also enhances accuracy and reproducibility, reducing subjectivity. Additionally, whole slide imaging technology, which refers to scanning of conventional glass slides in order to produce digital slides, enables machine learning in pathology by providing a vast amount of high-quality information for training and testing AI models to identify specific features and patterns that may be challenging for the human eye to discern.[4]

Machine learning, a subfield of AI, utilizes algorithms that enable computers to learn from digital images of tissue samples. In histopathology, it can be employed for the digital analysis of images to identify cell types, different structures, and to segment various regions of a given tissue sample.[5] The capabilities of machine learning have advanced with the development of deep learning: a section of machine learning is now focused on creating virtual neural networks, drawing inspiration from the ways in which the neurons of a human being communicate.[5] Deep learning models are adept at extracting features and learning from data. They can automatically identify complex patterns and relationships within diverse, large datasets, such as those used in cancer diagnostics.

However, choosing the best algorithm for the application of AI in histopathology is challenging. There are 3 primary types of learning: (1) supervised learning, which utilizes labeled data for training; (2) unsupervised learning, which identifies patterns without labels; and (3) weakly supervised learning, which strikes a middle ground by using partially labeled data.[6]

In our daily routines, we are well aware of repetitive and time-consuming tasks, such as the analysis of high-volume biopsy tissue samples and the counting of lymph nodes yielded during surgery. In such cases, AI has the potential to flag suspicious regions for inspection and might enable autonomous assessment. Additionally, AI can assist the pathologist in tasks like classifying different regions of cancer based on varying tumor grades using color-coded highlighting.

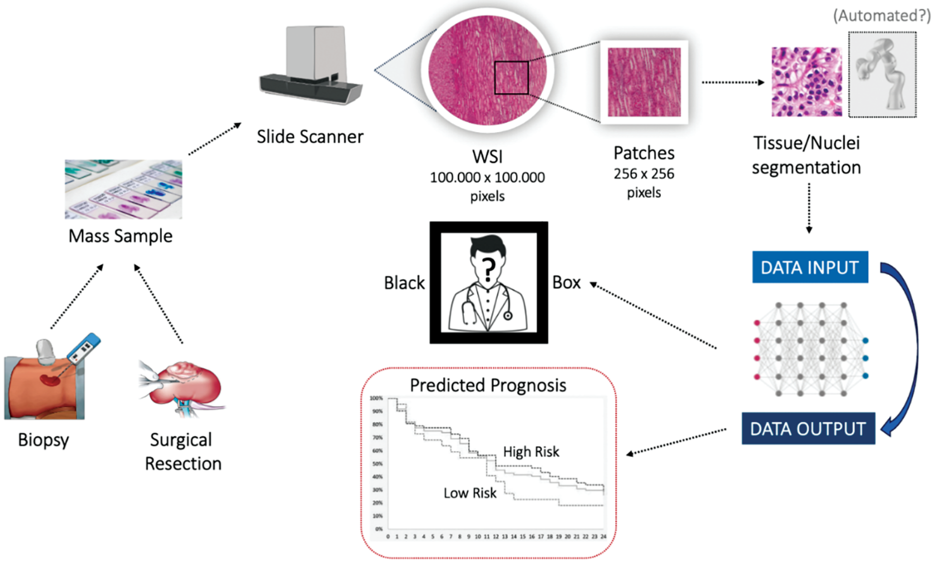


Figure. Pathway for the development of pathomics algorithms. After the sample is obtained from either biopsy or surgical resection, the whole slide imaging (WSI) is created. Derived patches are utilized through a digital scanner to train the algorithm to define diagnostic, prognostic, or predictive models. On the other hand, supervised learning-based algorithms could carry the "black box" issue: the machine generates an answer according to its learned algorithms, which humans cannot survey. Rather, the pathologist must have faith in the findings. Created with BioRender.com.

Moreover, by combining segmentation, detection, and classification techniques, it becomes possible to objectively quantify established biomarkers used in clinical practice. Notably, in the field of RCC pathology, specific instances include the evaluation of tumor-infiltrating lymphocytes and the quantification of PD-L1 (programmed death-ligand 1)–positive cells, which can even be predicted directly from slides.[7,8] Therefore, AI might aid in a wide spectrum of tasks ranging from tumor detection and classification to predictive and prognostic modeling.

Where are we now? Some authors have developed deep learning–based algorithms for RCC diagnosis, subtyping, and grading on biopsy specimens. However, they primarily focused on identifying the main subtypes of RCC without considering benign tumors.[9] Using specimens obtained from surgical resection, other authors have achieved promising results by employing AI in differentiating among RCC subtypes and normal parenchyma.[6] Undoubtedly, the pioneer experiences are witnessing how AI and machine learning in RCC pathology hold promise for the future of RCC diagnosis. They might help us overcome several issues faced by the pathologist with "traditional" histopathology, primarily concerning time consumption and intra-/interobserver variability.

> "AI can assist the pathologist in improving efficiency, accessibility, cost-effectiveness, and time consumption. It also enhances accuracy and reproducibility, reducing subjectivity."

A representation of the ideal pathway we imagine for the development of pathomics algorithms is summarized in the Figure: following either a biopsy or surgical resection, a whole slide image is created and derived patches are used through a digital scanner to train the algorithm in defining diagnostic and prognostic models.

One issue could lie in the fact that supervised learning-based algorithms could lead to the so-called "black box": while these algorithms are efficient at performing the assigned tasks, the generated outputs cannot be visually authenticated (ie, a human cannot oversee them; thus, the pathologist must have faith in the findings). Up to date, the available AI algorithms are either noninferior or even outperform the standard methods, but it is important to note that most technologies are currently unavailable for widespread clinical use, and further evidence is needed.

In our opinion, in the immediate future, AI will assist the community of uropathologists in elevating the average quality of assessments, particularly in recognizing different tumor gradings rather than different histotypes.

We believe this is the major need. From our clinical experience as "second-opinionists," we have observed that the majority of misclassifications with clinical implication occur in assigning tumor grade.

In a more distant future, the perspective of pathomics could lie in aiding the prediction of RCC prognosis. This will be particularly significant for the uro-oncologist, as an "augmented intelligence," relying on extensive big data (including tumor molecular characteristics), could differentiate between apparently similar pT1b grade 3 clear cell RCCs that will have different natural histories. ■

1. Cimadamore A, Caliò A, Marandino L, et al. Hot topics in renal cancer pathology: implications for clinical management. *Expert Rev Anticancer Ther.* 2022;22(12):1275-1287.
2. Caliò A, Marletta S, Brunelli M, Martignoni G. WHO 2022 Classification of Kidney Tumors: what is relevant? An update and future novelties for the pathologist. *Pathologica.* 2022;115(1):23-31.
3. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer.* 2022;3(9):1026-1038.
4. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol.* 2019;20(5):e253-e261.
5. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Comput Struct Biotechnol J.* 2018;16:34-42.
6. Distante A, Marandino L, Bertolo R, et al. Artificial intelligence in renal cell carcinoma histopathology: current applications and future perspectives. *Diagnostics (Basel).* 2023;13(13):2294.
7. Saltz J, Gupta R, Hou L, et al. Spatial organization and molecular correlation of Tumor-Infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* 2018;23(1):181-193.e7.
8. Kapil A, Wiestler T, Lanzmich S, et al. DASGAN–joint domain adaptation and segmentation for the analysis of epithelial regions in histopathology PD-L1 images. *arXiv.* 2019;10.48550/arXiv.1906.11118.
9. Fenstermaker M, Tomlins SA, Singh K, Wiens J, Morgan TM. Development and validation of a deep-learning model to assist with renal cell carcinoma histopathologic interpretation. *Urology.* 2020;144:152-157.