Title

Artificial Intelligence Improves the Ability of Physicians to Identify Prostate Cancer Extent

Journal

Journal of Urology, 212(1)

ISSN

0021-0005

Authors

Mota, Sakina Mohammed
Priester, Alan
Shubert, Joshua
et al.

Publication Date

2024-07-01

DOI

10.1097/ju.0000000000003960

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

OPEN

Prostate Cancer

JU Insight

Artificial Intelligence Improves the Ability of Physicians to Identify Prostate Cancer Extent

Sakina Mohammed Mota, Alan Priester, Joshua Shubert, et al.

Correspondence: Sakina Mohammed Mota (sakina@avendahealth.com).

Full-length article available at https://doi.org/10.1097/JU.0000000000003960.

Study Need and Importance: In targeted therapy of prostate cancer, clinicians aim to treat clinically significant disease while minimizing damage to healthy tissue. However, conventional MRI underestimates the extent of disease, confounding treatment decisions. Artificial intelligence (AI) can help overcome this limitation, leveraging multimodal data to map cancer risk in 3D. This approach can reveal disease invisible to MRI, potentially improving patient selection, treatment planning, and oncologic efficacy. A multireader multicase study was conducted to compare physicians' delineations of tumor extent using AI vs standard-of-care (SOC; ie, clinical judgement) and to evaluate AI's impact on treatment decision-making.

What We Found: Ten physicians (7 urologists and 3 radiologists) trained in urologic oncology each evaluated 50 cases using both SOC and AI methods (1000 total evaluations). All patients were diagnosed with intermediate-risk prostate cancer via MRI-targeted biopsy then received radical prostatectomy. Whole mount pathology slides derived from surgical specimens were registered to MRI and used as ground truth (Figure). AI-assisted contours had significantly greater balanced accuracy (84.7% vs 67.2%) and sensitivity (97.4% vs 38.2%) than SOC. AI also achieved a substantially higher negative margin rate than SOC (72.8% vs 1.6%) and caused urologists to alter treatment decisions for 28% of cases.

Limitations: Study data were derived from radical prostatectomy cases from a single institution. Ground truth was defined by a single expert pathologist, with some potential subjectivity in tumor boundary annotation and Gleason grading.

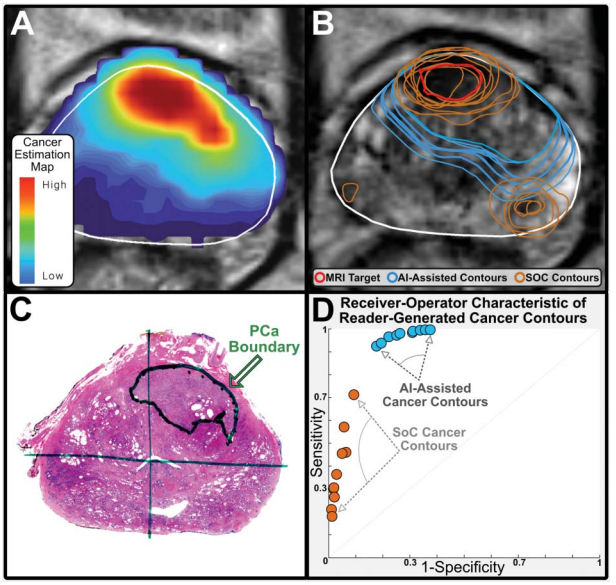

Figure. Exemplary case from reader study showing artificial intelligence (AI) output (A), standard-of-care (SOC; orange) and AI-assisted cancer contours (blue; B), with corresponding ground truth from surgical pathology (C), and reader performance across all cases demonstrated improvement over SOC in all readers when using AI (D). PCa indicates prostate cancer.

Interpretation for Patient Care: AI cancer estimation is a promising improvement upon current practice. The enhanced accuracy of AI-derived tumor delineation could improve therapy planning, targeting, and oncologic efficacy. AI may even influence biopsy strategy, with perilesion (penumbra) cores sampled from higher-risk regions to improve tumor boundary sampling.

https://doi.org/10.1097/JU.0000000000003960
Vol. 212, 52-62, July 2024
Printed in U.S.A.

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

OPEN

Artificial Intelligence Improves the Ability of Physicians to Identify Prostate Cancer Extent

Sakina Mohammed Mota,[1]* Alan Priester,[1,2]* Joshua Shubert,[1] Jeremy Bong,[1] James Sayre,[3] Brittany Berry-Pusey,[1] Wayne G. Brisbane,[2] and Shyam Natarajan[1,2]

[1]Avenda Health, Inc
[2]Department of Urology, David Geffen School of Medicine, Los Angeles, California
[3]Department of Radiological Sciences and Biostatistics, University of California, Los Angeles, California

Purpose: Defining prostate cancer contours is a complex task, undermining the efficacy of interventions such as focal therapy. A multireader multicase study compared physicians' performance using artificial intelligence (AI) vs standard-of-care methods for tumor delineation.

Materials and Methods: Cases were interpreted by 7 urologists and 3 radiologists from 5 institutions with 2 to 23 years of experience. Each reader evaluated 50 prostatectomy cases retrospectively eligible for focal therapy. Each case included a T2-weighted MRI, contours of the prostate and region(s) of interest suspicious for cancer, and a biopsy report. First, readers defined cancer contours cognitively, manually delineating tumor boundaries to encapsulate all clinically significant disease. Then, after $\geq$ 4 weeks, readers contoured the same cases using AI software. Using tumor boundaries on whole-mount histopathology slides as ground truth, AI-assisted, cognitively-defined, and hemigland cancer contours were evaluated. Primary outcome measures were the accuracy and negative margin rate of cancer contours. All statistical analyses were performed using generalized estimating equations.

Results: The balanced accuracy (mean of voxel-wise sensitivity and specificity) of AI-assisted cancer contours (84.7%) was superior to cognitively-defined (67.2%) and hemigland contours (75.9%; $P < .0001$). Cognitively-defined cancer contours systematically underestimated cancer extent, with a negative margin rate of 1.6% compared to 72.8% for AI-assisted cancer contours ($P < .0001$).

Conclusions: AI-assisted cancer contours reduce underestimation of prostate cancer extent, significantly improving contouring accuracy and negative margin rate achieved by physicians. This technology can potentially improve outcomes, as accurate contouring informs patient management strategy and underpins the oncologic efficacy of treatment.

Key Words: artificial intelligence, comparative study, magnetic resonance imaging, prostatic neoplasms, surgical margins

For patients with intermediate-risk prostate cancer (PCa), focal therapy (FT) is gaining acceptance as an alternative to prostatectomy and radiation therapy.[1] Evidence is growing that FT can eradicate PCa while preserving patient quality of life.[2-7] However, successful FT requires accurate contouring of cancer extent. Unfortunately, conventional cancer contouring requires surgeons to cognitively merge imaging, pathology, and clinical factors. In particular, the interpretation of MRI systematically underestimates cancer extent—sometimes severely.[8-10] Patient-specific, reproducible assessment of cancer extent is a pressing and unaddressed need.

Submitted September 8, 2023; accepted March 28, 2024; published June 11, 2024.

Funding/Support: This study was funded in part by the National Cancer Institute at the National Institutes of Health, Grant R01CA218547.

Conflict of Interest Disclosures: Dr Mota, Dr Priester, Mr Shubert, and Mr Bong are employees of Avenda Health. Dr Berry-Pusey and Dr Natarajan are cofounders of Avenda Health. Dr Sayre consults for Avenda Health.

Ethics Statement: This study was retrospectively performed on extant anonymized data from an Independent Review Board–approved study (IRB No. 11-001580).

Author Contributions:

Conception and design: Mota, Priester, Shubert, Sayre, Berry-Pusey, Natarajan.

Data analysis and interpretation: Mota, Priester, Sayre, Berry-Pusey, Brisbane.

Data acquisition: Mota, Priester, Berry-Pusey, Bong.

Drafting the manuscript: Mota, Priester, Sayre, Brisbane, Natarajan.

Critical revision of the manuscript for scientific and factual content: Mota, Priester, Shubert, Berry-Pusey, Brisbane, Natarajan, Bong.

Statistical analysis: Mota, Priester, Sayre.

Supervision: Berry-Pusey, Brisbane, Natarajan.

Data Availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Corresponding Author: Sakina Mohammed Mota, PhD, Avenda Health, Data Science, 4130 Overland Ave, Culver City, CA 90230 (sakina@avendahealth.com).

* Co-first authors.

https://doi.org/10.1097/JU.0000000000003960
Vol. 212, 52-62, July 2024
Printed in U.S.A.

Frequently, instead of precise tumor contouring, the entire cancer-bearing hemisphere (hemigland) is treated.[2,11-13] Hemigland treatment is rarely optimal, frequently undertreating cases with bilateral cancer[14] while encompassing large volumes of benign tissue. Artificial intelligence (AI) overcomes these limitations, improving the identification and characterization of PCa compared to visual MRI interpretation alone.[15-19] Furthermore, tracked biopsy data derived from MRI-ultrasound fusion devices[20] are an underutilized but powerful resource for cancer contouring.

Software recently cleared by the Food and Drug Administration ("Unfold AI," K221624, Avenda Health, Culver City, California) uses an AI algorithm to visualize cancer probability based on multimodal clinical data. In a retrospective study, the AI software outperformed conventional contouring methods, including the original radiology-defined region of interest (ROI), a 10-mm margin around the ROI, and hemigland margins.[21] Furthermore, AI software accurately assessed the likelihood of residual clinically significant prostate cancer (csPCa; defined as Gleason Grade Group [GG] $\geq$2) outside a cancer contour.[21] This study demonstrated the potential of AI software to improve FT treatment planning. However, no prior work has characterized the performance of this AI tool in the hands of physicians.

To that end, we conducted a multireader multicase study to evaluate the cancer contouring and clinical decision-making of physicians with and without AI software. We compared cognitive and hemigland contouring methodologies to AI-assisted cancer contours. Whole mount (WM) pathology data were used as ground truth, enabling retrospective comparison of predictions to the true extent of csPCa.

MATERIALS AND METHODS

Whole Mount Dataset Description and Case Selection

WM-derived data from an Independent Review Board–approved study at the University of California, Los Angeles served as ground truth for the location of csPCa.[8] Out of 124 cases accrued from 2011 to 2016, 50 plausibly eligible FT cases were selected according to the following inclusion criteria:

- GG 2 or 3
- At least one magnetic resonance–visible ROI with Prostate Imaging Reporting and Data System version 2 (PI-RADS)[22] $\geq$ 3
- Unilateral csPCa, or csPCa confined to the anterior zone, based on biopsy and imaging
- No prior prostate cancer treatment

All patients in the study dataset had received presurgical multiparametric MRI. The prostate capsule was segmented using T2-weighted MRI, and cancer-suspicious ROIs were defined using PI-RADS22 or a Likert-like precursor with similar performance. Imaging was followed by radical prostatectomy and WM slide preparation. Patient-specific 3D-printed molds based on the prostate capsule segmentation were used to align the excised prostate with preoperative image slices. The prostate specimen was sliced, processed into WM slides, and then annotated by an expert pathologist (>10 years of expertise) to demarcate tumor boundaries and determine tumor-level GG. A validated algorithm was used to register WM tumor contours to corresponding MR images and interpolate them into 3D surfaces.[8]

The pathologic stages for these 50 WM cases were pT2 (34%), pT2a (4%), pT2b (6%), pT2c (14%), pT3a (38%), and pT3b (4%). Extraprostatic extension (EPE) was observed on WM pathology in 21/50 (42%) cases. Out of these 21 cases, EPE was noted on preoperative imaging in 12 cases and missed on preoperative imaging in 9 cases. Seminal vesicle invasion was also observed on 2 out of 21 of EPE cases. Regardless of EPE presence, assessment of cancer contours was limited to intraprostatic tissue for this study.

Study Design

The multireader multicase study design was fully crossed, with each reader reviewing each case. Ten physicians (urologists and radiologists) were recruited from 5 institutions to represent a broad range of clinical oncology experience (mean 9 years, range 2-23 years). First, after a chart review of conventional diagnostic data, each reader produced cognitive standard-of-care (SOC) cancer contours via manual delineation. Then, a 4-week washout period was imposed to minimize reading order bias (ie, reader recall of the prior cancer contour). Finally, readers repeated cancer contouring with the assistance of AI software. All cases were anonymized, with case identifiers and read order randomized between contouring sessions. Following the completion of contouring, task duration was recorded in a logbook.

The first reader was trained for the first study session (SOC) on March 22, 2022, and the last reader completed the second study session (AI) on October 11, 2022. During both sessions (SOC and AI), readers were instructed primarily to encapsulate all csPCa and secondarily to exclude non-csPCa tissue. In order to mirror clinical practice, readers were provided a text report including pathology diagnoses for each biopsy core and serum PSA (Figure 1, A). Readers were also provided with ROI contours and PI-RADS scores produced by an expert radiologist (>10 years of experience).

SOC and AI-assisted contouring methodologies are described in the following sections. Additionally, a follow-up SOC session was performed to investigate the influence of exposure to AI software, ie, a potential "learning effect." Details specific to hemigland hemisphere selection and follow-up SOC contouring are presented in the supplementary materials (https://www.jurology.com).

SOC Cancer Contouring

T2-weighted MRI and contours of the prostate capsule and ROI(s) were preloaded to open-source medical imaging software (3D Slicer version 5.1.0[23]). Readers were trained to manually delineate cancer contours with the "Segmentation Editor" module of 3D Slicer. They could

A

Patient is 52 years old, presenting with PSA 3.1 ng/mL. Patient had 2 MRI-visible regions of interest suspicious for cancer in the LPPM and RPPA, with PIRADS scores of 3 and 3 respectively. Patient received tracked biopsy followed by microscopic analysis of tissue samples. Cores bearing clinically significant cancer (Gleason Score = 3+4) were found in the LPPM ROI and elsewhere in the right hemisphere. Additional cores bearing cancer with Gleason Score = 3+3 were found in the LPPM ROI and elsewhere in the left and right hemisphere. Per-Core biopsy results are presented below.

Age	PSA (ng/mL)	Core Label	Core Location	Primary Gleason Score	Secondary Gleason Score	Cancer Length (mm)	% Cancer in Core	Core Length (mm)	Prostate Volume (CC)	PIRADS Score
52	3.1	1-1	LEFT LATERAL BASE					16	33.42	
52	3.1	2-1	LEFT LATERAL MID	3	3	6	55%	11	33.42	
52	3.1	3-1	LEFT LATERAL APEX					14	33.42	
52	3.1	4-1	LEFT BASE					15	33.42	
52	3.1	5-1	LEFT MID					6	33.42	
52	3.1	6-1	LEFT APEX					17	33.42	
52	3.1	7-1	RIGHT BASE					12	33.42	
52	3.1	8-1	RIGHT MID	3	4	3	19%	16	33.42	
52	3.1	9-1	RIGHT APEX	3	3	2	15%	13	33.42	
52	3.1	10-1	RIGHT LATERAL BASE					15	33.42	
52	3.1	11-1	RIGHT LATERAL MID					17	33.42	
52	3.1	12-1	RIGHT LATERAL APEX					12	33.42	
52	3.1	114-2	LPPM	3	3	7	37%	19	33.42	3
52	3.1	114-2	LPPM	3	3	7	100%	6	33.42	3
52	3.1	114-2	LPPM	3	3	3	21%	14	33.42	3
52	3.1	114-2	RPPA	3	4	3	17%	18	33.42	3
52	3.1	114-2	RPPA					9	33.42	3
52	3.1	114-2	RPPA					15	33.42	3

Site Acronyms:

LPPM = Left Posteriolateral Peripheral Midgland

RPPA = Right Posteriolateral Peripheral Apex

B

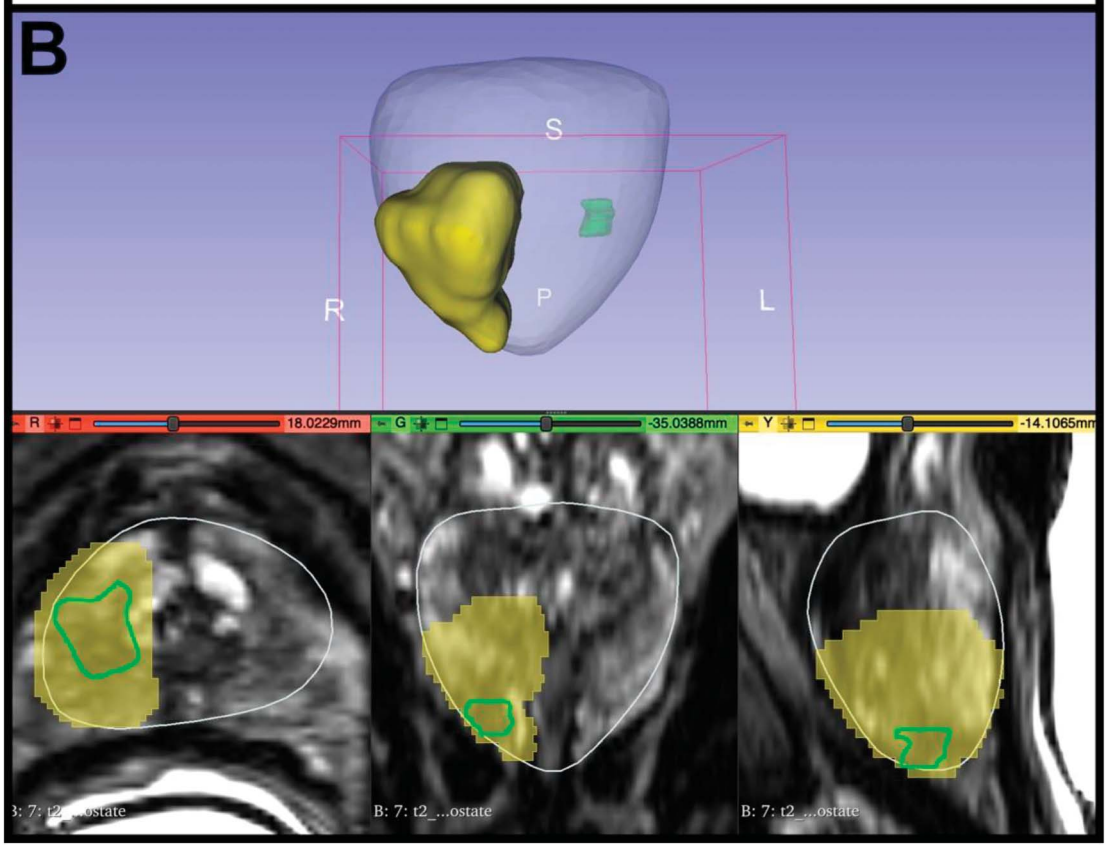

Figure 1. Exemplary image showing the prediction of cancer extent using standard-of-care (SOC) methodology. A, Biopsy pathology summary document. First, a text summary of the case is presented including serum PSA, followed by a table with 1 row per biopsy core. B, 3D Slicer software showing 3D surfaces (top) and scrollable T2-weighted MRI in axial, coronal, and sagittal views (bottom). The prostate surface model is shown in white, Prostate Imaging Reporting and Data System (PIRADS) regions of interest suspicious for prostate cancer are shown in green, and an exemplary manually-drawn SOC cancer contour is shown in yellow.

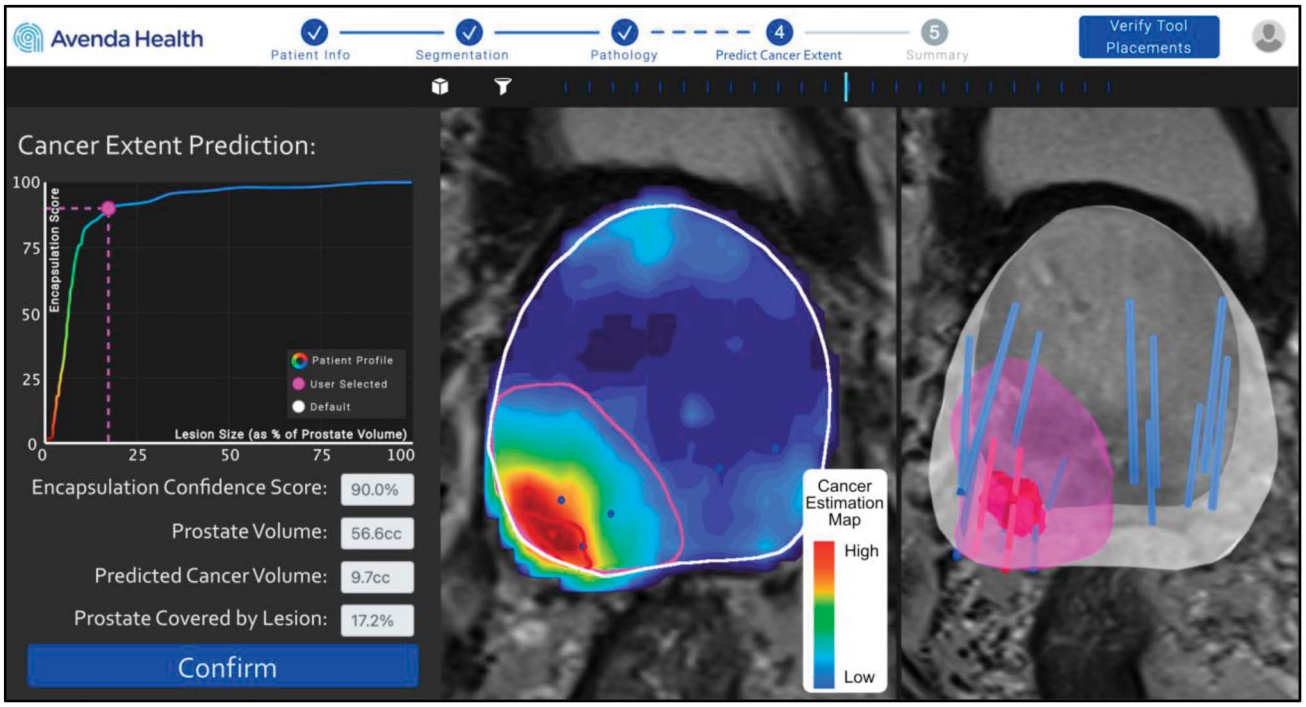

Figure 2. Exemplary image showing the prediction of cancer extent using artificial intelligence software. The central panel shows T2-weighted MRI overlaid with the cancer estimation map, biopsy cores bearing clinically significant prostate cancer (red circles), biopsy cores bearing benign tissue (blue circles), Prostate Imaging Reporting and Data System region of interest (red contour), the prostate capsule (white), and the currently selected cancer contour (pink). The right-hand panel shows the same elements in 3D. The left-hand panel shows the curve for encapsulation confidence vs cancer contour volume (as a percentage of the prostate), as well as other cancer contour statistics.

freely define lesion geometry in order to best accomplish the contouring task. An example of preloaded SOC data, the corresponding biopsy pathology document, and an SOC cancer contour are shown in Figure 1.

AI-Assisted Cancer Contouring

The same data used for SOC interpretation were preloaded to the AI software, and readers had access to the same text-based biopsy report. AI software inputs included T2-weighted MRI, simulated 3D biopsy core locations, PSA, and biopsy GG. Detailed information regarding the AI algorithm, its development, and model parameters are presented in Priester et al.[21] AI software was used to generate a 3D cancer estimation map (CEM), color-coding each prostate voxel with the estimated probability of csPCa. The software also displayed an "encapsulation confidence score," the estimated probability that a cancer contour would completely encapsulate all csPCa.

A default cancer contour, maximizing encapsulation confidence while minimizing contour volume, was selected automatically by the software and initially displayed to readers. Then, for each case, readers selected a threshold using the encapsulation confidence curve to apply to the CEM (Figure 2). Using this reader-selected threshold, the AI software produced a 3D cancer contour, converting the CEM into cancer contours comparable to SOC-defined or hemigland contours.

Table 1. *Cancer Contour Performance*

	Mean performance			Mean differences		95% CIs		*P* values		Mean coefficients of variation	
Study measures	SOC	AI	Hemi	AI-SOC	AI-hemi	AI-SOC	AI-hemi	AI vs SOC	AI vs hemi	SOC	AI
Sensitivity, %	38.2	97.4	98.4	59.2	−1.0	+33.4, +46.4	−1.2, +2.8	**< .0001**	.42	0.77	0.04
Specificity, %	96.2	72.1	53.4	−24.2	18.7	−29.8, −23.9	+8.8, +17.3	**< .0001**	**< .0001**	0.04	0.13
Balanced accuracy, %	67.2	84.7	75.9	17.5	8.8	+2.9, +10.2	+4.8, +9.0	**< .0001**	**< .0001**	0.14	0.05
Negative margin rate, %	1.6	72.8	86.0	71.2	−13.2	+75.1, +92.4	−9.4, +14.0	**< .0001**	.70	—	—
Prediction time, min	3.5	2.0	—	−1.5	—	−1.0, −3.2	—	**< .0001**	—	0.53	0.56

Abbreviations: AI, artificial intelligence; hemi, hemigland; SOC, standard-of-care.
All *P* values were derived from generalized estimation equations analysis. The application of a Bonferroni correction for the 5 measures (2 primary and 3 secondary) establishes a statistical significance threshold of $\alpha = .01$.
Bolded text indicates significance.

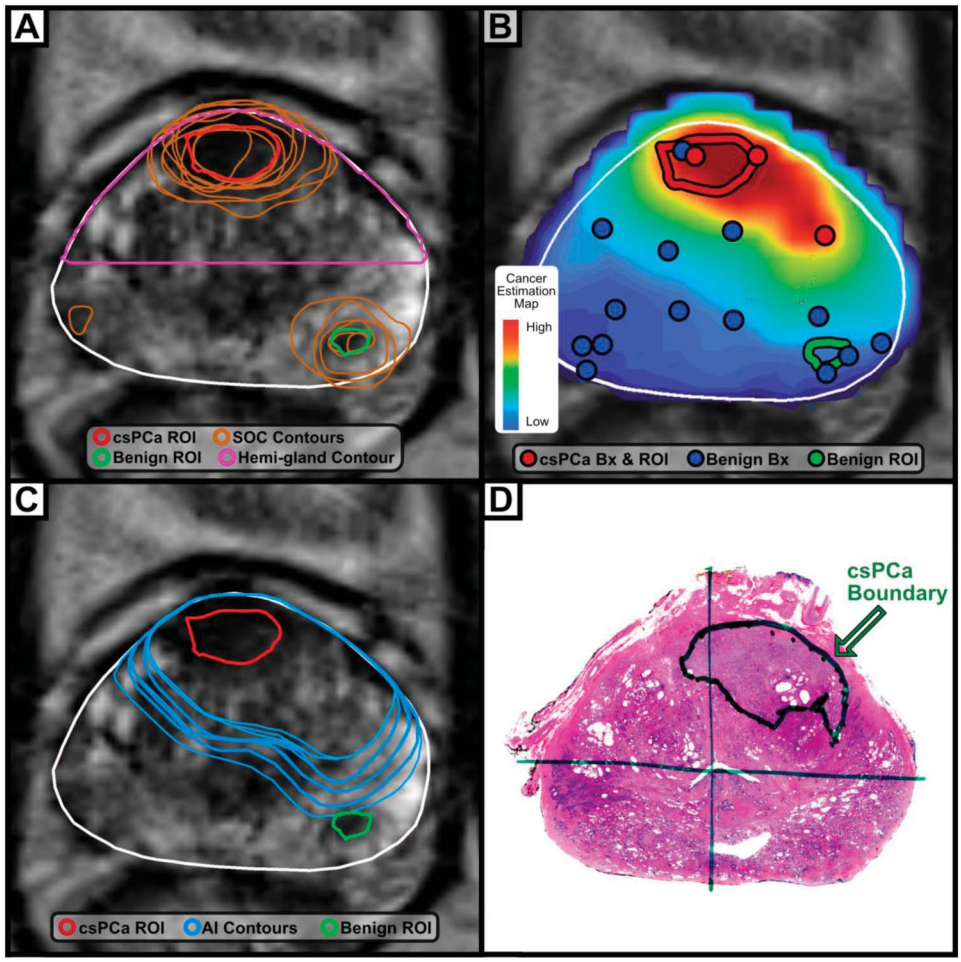

Figure 3. Exemplary images from cases wherein the standard-of-care (SOC) cancer contours had below-average mean sensitivity. Figure elements include SOC and hemigland cancer contours (A), case displayed using artificial intelligence (AI) software (B), AI-assisted cancer contours (C), and whole mount ground truth (D). The prostate is always contoured in white; legends and labels within each figure panel denote other color correspondences. The mean balanced accuracy of SOC, AI, and hemigland contours for this case (#44) was 55.2%, 82.6%, and 77.4%, respectively. Negative margin rates were 0% for SOC, 100% for AI, and 100% for hemigland contours. Bx indicates biopsy; csPCa, clinically significant prostate cancer; ROI, region of interest.

Clinical Decision-Making Survey

To investigate changes in clinical decision-making, a subset of readers (N = 7) recorded their recommended treatment strategy after SOC and AI-assisted cancer contouring. Following SOC contouring, readers selected a recommended treatment: active surveillance, radical prostatectomy, radiation therapy, or focal therapy. After the 1-month washout period and AI-assisted contouring, readers again selected a recommended treatment.

Postprocessing and Cancer Contour Assessment

The results consisted of N = 10 sets of 50 matched cancer contours (SOC, AI-assisted, and hemigland) and contouring durations (SOC and AI-assisted). Reader performance was assessed using WM radical prostatectomy as ground truth. Evaluation of reader performance was purely objective and carried out through automated test scripts written in MATLAB 2020b (MathWorks, Natick, Massachusetts). The overlap between ground-truth csPCa regions and each cancer contour was used to compute sensitivity and specificity. Voxel-wise sensitivity was defined as the proportion of csPCa-bearing voxels correctly included within the cancer contour and voxel-wise specificity as the proportion of non-csPCa voxels correctly excluded from the cancer contour.

The primary outcome measures were voxel-wise balanced accuracy, defined as (sensitivity + specificity)/2, and the negative margin rate, ie, the proportion of cases wherein cancer contours enclosed all csPCa-bearing regions. These measures were selected to represent clinical priorities, where cancer control is prioritized over sparing damage to benign prostatic tissue. Secondary measures consisted of sensitivity, specificity, and time spent contouring. Note that all sensitivity, specificity, and balanced accuracy values reported in this article represent voxel-wise measures. Changes in the recommended treatment strategy were computed and compared for significance. All statistical comparisons were conducted via generalized estimating equations analysis, and a Bonferroni correction was applied to adjust for multiple comparisons.

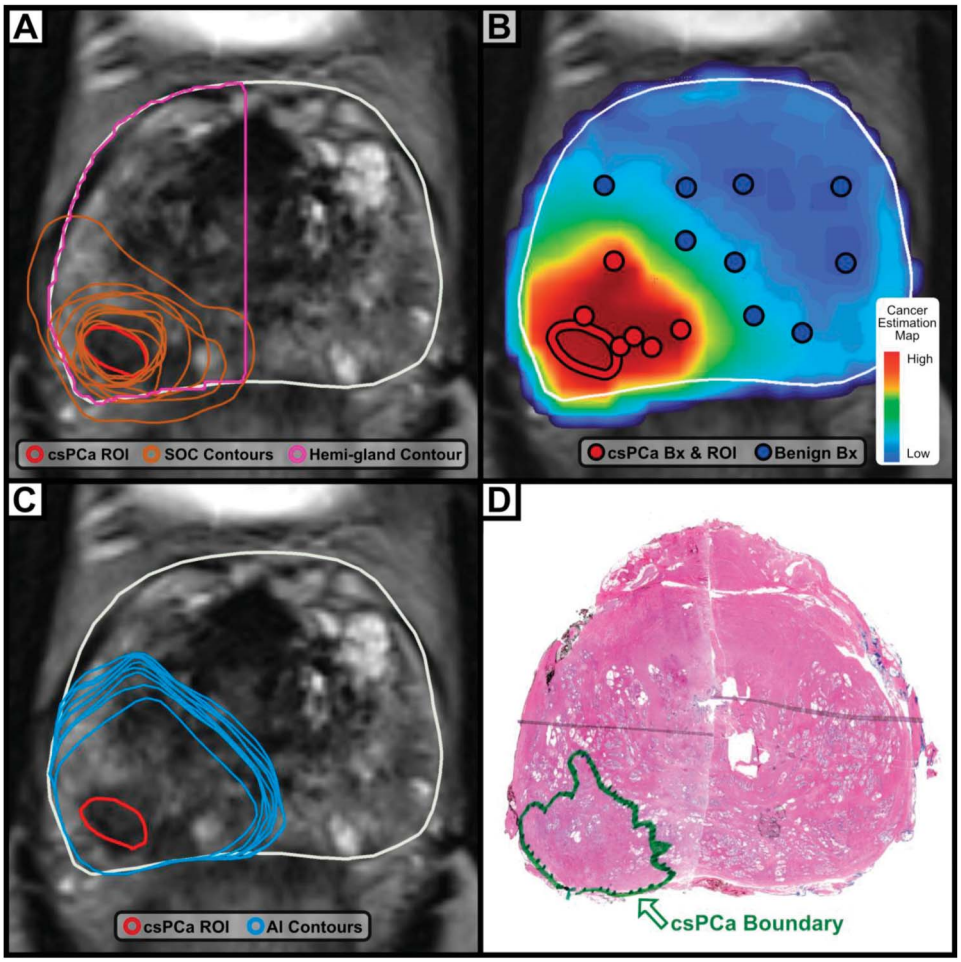

Figure 4. Exemplary images from cases wherein the standard-of-care (SOC) cancer contours had approximately average mean sensitivity. Figure elements include SOC and hemigland cancer contours (A), case displayed using artificial intelligence (AI) software (B), AI-assisted cancer contours (C), and whole mount ground truth (D). The prostate is always contoured in white; legends and labels within each figure panel denote other color correspondences. The mean balanced accuracy of SOC, AI, and hemigland contours for this case (#22) was 64.0%, 87.5%, and 77.7%, respectively. Negative margin rates were 0% for SOC, 70% for AI, and 0% for hemigland contours. Bx indicates biopsy; csPCa, clinically significant prostate cancer; ROI, region of interest.

RESULTS

Predictive Performance of Cancer Contours

Results are summarized in Table 1 and illustrated in Figures 3-6. Performance comparison between mean SOC (N = 10 readers), mean AI (N = 10 readers), and hemigland cancer contours is outlined for each of the 50 cases in supplementary materials (Table S1, https://www.jurology.com). Cancer contours produced using AI software had balanced accuracy superior to both SOC and hemigland methods (84.7% vs 67.2% and 75.9%, respectively, $P < .001$). AI-assisted cancer contours had greater sensitivity than SOC (mean 97.4% vs 38.2%, $P < .001$) and greater specificity than hemigland (mean 72.1% vs 53.4%, $P < .001$). The negative margin rate achieved using AI software was substantially higher than SOC (mean 72.8% vs 1.6%, $P < .001$), and the time required to produce cancer contours was reduced (mean 2.0 vs 3.5 minutes, $P < .001$). The negative margin rate for hemigland contours was greater than AI software, but the difference was statistically insignificant (86.0% vs 72.8%, $P = .7$). The mean performance of urologists (N = 7) and radiologists (N = 3) was similar; balanced accuracy was 68.0% vs 65.4% for SOC contours and 84.6% vs 85.0% for AI-assisted contours.

Figure 6 shows a receiver-operator characteristic plotting AI-assisted, SOC, and hemigland cancer contours. AI-assisted cancer contours had higher volume and more sensitivity than SOC cancer contours, with a greater negative margin rate. AI-assisted cancer contours were smaller and more specific than hemigland contours, with a negative margin rate that differed insignificantly. Furthermore, readers using AI software had less inter-reader variability than SOC as seen in Figure 6, and coefficient of variation measures in Table 1.

A distinct learning effect was observed during follow-up SOC contouring, with exposure to AI

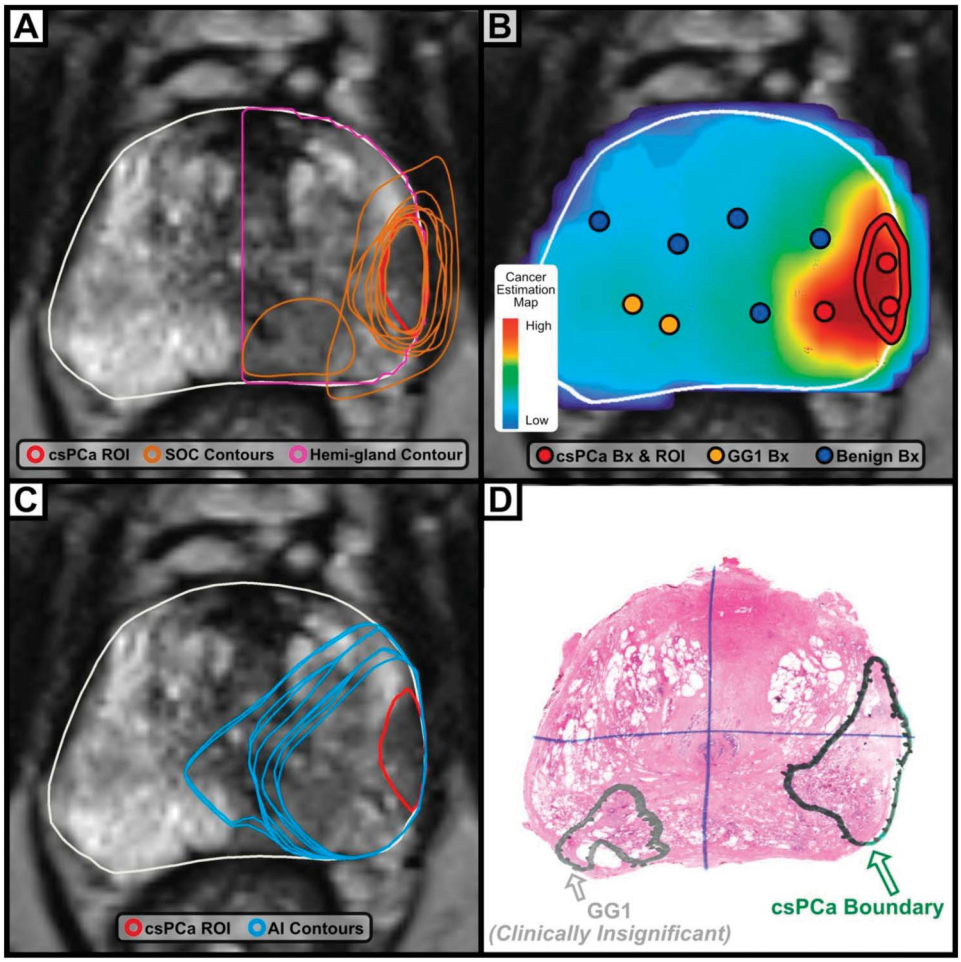

Figure 5. Exemplary images from cases wherein the standard-of-care (SOC) cancer contours had above-average mean sensitivity. Figure elements include SOC and hemigland cancer contours (A), case displayed using artificial intelligence (AI) software (B), AI-assisted cancer contours (C), and whole mount ground truth (D). The prostate is always contoured in white; legends and labels within each figure panel denote other color correspondences. The mean balanced accuracy of SOC, AI, and hemigland contours for this case (#25) was 75.1%, 88.8%, and 75.7%, respectively. Negative margin rates were 10% for SOC, 90% for AI, and 100% for hemigland contours. Bx indicates biopsy; csPCa, clinically significant prostate cancer; GG, Gleason Grade Group; ROI, region of interest.

markedly improving physician performance. Data pertaining to follow-up SOC contouring are presented in Supplementary Table 2 and Supplementary Figure 1 (https://www.jurology.com).

Impact of AI Software on Clinician Decision-Making for PCa Management

Survey results, completed by N = 7 readers, were used to analyze trends in PCa management decision-making (Table 2). AI software usage led to altered treatment recommendations in 42% (28% urologists, 61% radiologists) of cases ($P < .001$). After exposure to AI software, urologists tended to recommend FT more frequently (12.5% SOC vs 21.5% AI-assisted) and radical prostatectomy less frequently (58.0% vs 52.0%); radiologists tended to recommend Radiation Therapy (RT) less frequently (38.0% vs 33.3%) and radical prostatectomy more frequently (11.3% vs 16.7%). These trends are graphically represented using a grouped bar chart in Figure 7.

DISCUSSION

Generating cancer contours is a complex cognitive process requiring a surgeon to consider imaging, pathology, and clinical features simultaneously. In the presented study, SOC contours systematically underestimated csPCa extent. Given access to standard clinical information, physicians tended to err toward small "high-specificity" contours that fail to encapsulate the entire tumor. This likely occurred as physicians relied on a tumor's MRI appearance, but the true extent of prostate cancer is often MRI invisible. The underestimation of tumor size is consistent with prior work,[8-10,21] and plausibly accounts for the variable rates of residual disease reported during focal treatment trials to date.[24]

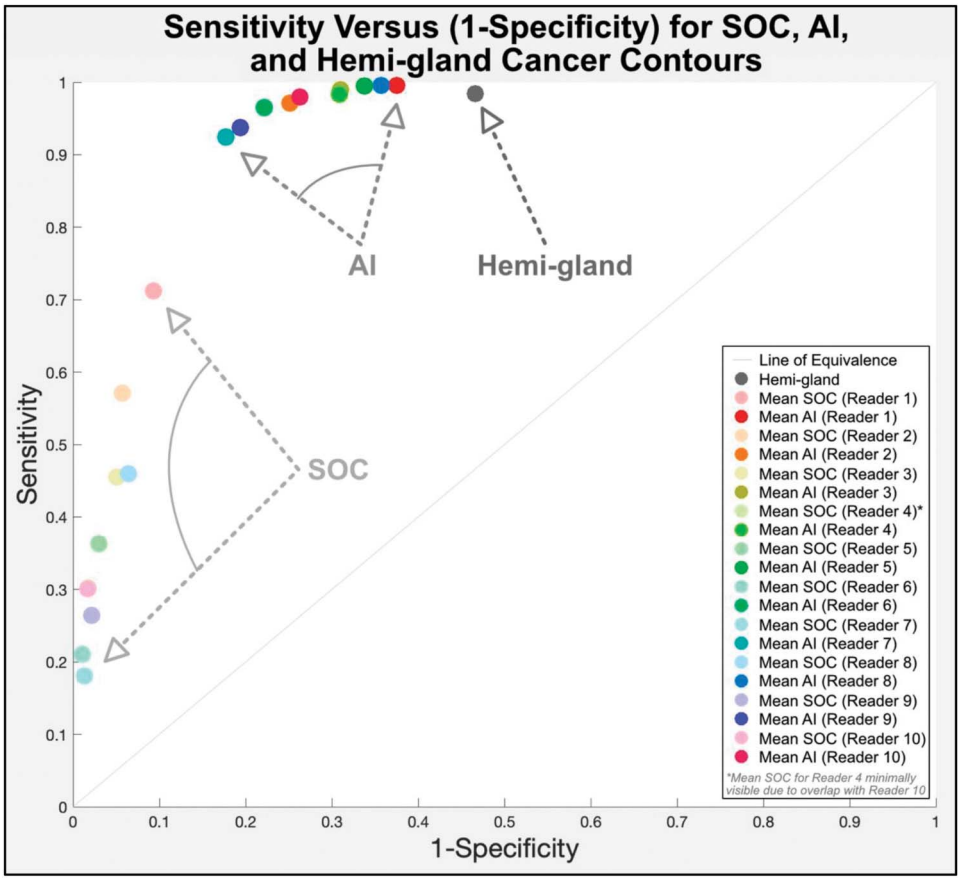

Figure 6. Receiver-operator characteristic showing mean sensitivity vs 1-specificity measures derived from all (N = 50) cancer contour predictions for each reader (N = 10) and each prediction method: artificial intelligence (AI)–assisted, standard-of-care (SOC), and hemigland. Note that AI predictions had moderately decreased specificity but substantially improved sensitivity compared to SOC, leading to greater balanced accuracy and substantial improvements in negative margin rates. Likewise, note that AI predictions were more closely clustered, indicating lower inter-reader variability than SOC.

AI can help resolve this challenging problem. AI was significantly more accurate than hemigland and SOC contours, indicating a better balance between sensitivity and specificity. Furthermore, exposure to AI software encouraged physicians to define larger and more patient-specific contours, improving the negative margin rate. When AI software was used in our study population, the negative margin rate rose from 1.6% to 72.8%—a 45-fold increase. Though the sensitivity of AI and hemigland contours were comparable, AI contours were smaller and more specific. It is therefore plausible that AI may facilitate faster treatment with fewer side effects, particularly because hemigland ablation entails lethal treatment in close proximity to the urethra, external sphincter, and ipsilateral nerve bundle. Avoidance of these structures, though common in clinical implementation, would reduce hemigland contour size and likely compromise their efficacy for a subset of cases.

Understandably, we observed that AI assistance impacted clinical decision-making, with recommended treatment altered in over a quarter of cases. The cancer estimation map improved confidence in recommending focal therapy over whole gland

Table 2. *Recommended Treatment Survey Statistics*

	Urologist (N = 4)			Radiologist (N = 3)		
Management decision	SOC mean	AI mean	AI-SOC	SOC mean	AI mean	AI-SOC
Active surveillance, %	19.0	16.5	−2.5	9.3	10.0	0.7
Focal therapy, %	12.5	21.5	9.0	38.7	40.0	1.3
Radiation therapy, %	10.5	8.0	−2.5	38.0	33.3	−4.7
Radical prostatectomy, %	58.0	52.0	−6.0	11.3	16.7	5.3
Other, %	0.0	2.0	2.0	2.7	0.0	−2.7

Abbreviations: AI, artificial intelligence; SOC, standard-of-care.

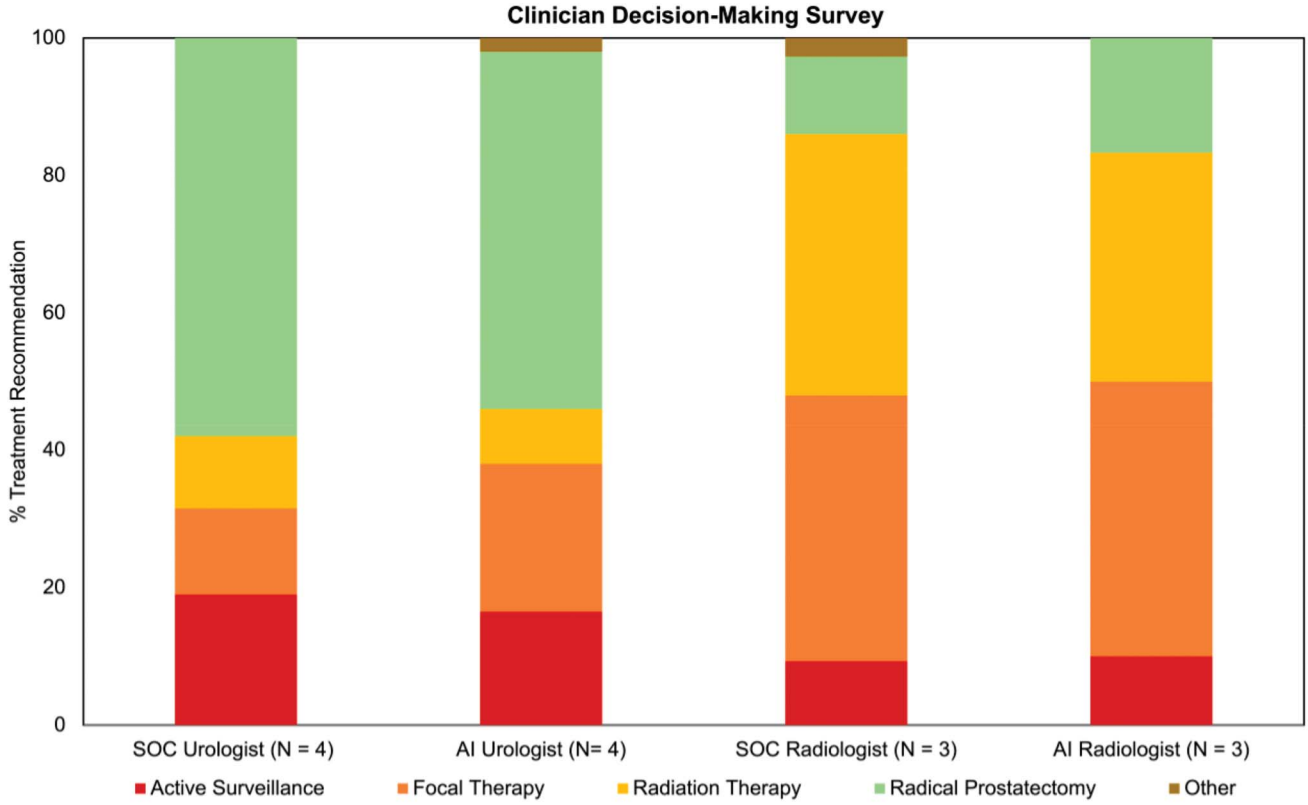

Figure 7. Bar chart showing the survey trends in prostate cancer treatment decision-making after standard-of-care (SOC) vs artificial intelligence (AI)–predicted cancer contouring, grouped by urologists and radiologists. AI usage led to statistically significant changes in decision-making ($P < .001$).

therapy, with urologists trending towards a more targeted approach. Clinician reluctance to recommend focal therapy under SOC conditions mirrors documented shortcomings in patient selection based on MRI and systematic biopsy alone.[14,21,25]

This study was subject to several limitations. First, though the readers were multi-institutional, the study cases were derived from a single institution and may not represent performance in broader populations. Additionally, the ground truth was set up for this study by a single expert pathologist and lacks a method to address observer dependency in tumor boundary annotation and Gleason grading. The required elements for robust algorithm validation (WM data registered to preoperative MRI) are rare, and future work will entail acquiring and testing data from a more diverse patient population while using WM slide annotations verified by more than 1 experienced pathologist as the truth. Second, the dataset was derived from radical prostatectomy cases, representing relatively advanced disease. This limitation was mitigated through study inclusion criteria, intended to select only plausible FT candidates. However, the inclusion of only patients with apparently unilateral cancer in the cohort likely inflated the performance of hemigland contours relative to AI-assisted contours. Prior work on broader populations suggests a substantial incidence of undetected bilateral disease.[14,21] Third, the study design lacked a crossed-over approach, wherein the reading order of AI vs SOC interpretation would be reversed for half the dataset. A noncrossed-over study design was selected to prevent reading order bias, which results from exposure to AI software and would not have been representative of true SOC performance. Follow-up SOC data analysis (supplementary materials, https://www.jurology.com) confirmed and characterized this effect.

The promising performance of AI-assisted cancer contours suggests several avenues for future research efforts. Targeted therapies may benefit from the prospective usage of AI contours since they dramatically improved negative margin rates relative to cognitive contours. Similarly, AI has the potential to aid in patient selection, with focal treatment only offered to patients when critical structures lie outside the cancer contour. AI could even potentially influence biopsy strategy, with perilesion (penumbra)[26] cores sampled from higher-risk regions in order to reduce core count and improve the sampling of tumor boundaries.

CONCLUSIONS

AI-assisted cancer contouring overcomes many limitations of the current clinical workflow, facilitating

patient-specific therapy selection and treatment planning. The use of AI increased clinician recommendations for focal therapy among patients with unilateral cancer and reduced variation in accurate tumor encapsulation. AI-assisted contours are a promising improvement to the current standard of care, and their impact in a clinical setting will be the subject of further validation efforts.

REFERENCES

1. Connor MJ, Gorin MA, Ahmed HU, Nigam R. Focal therapy for localized prostate cancer in the era of routine multi-parametric MRI. *Prostate Cancer Prostatic Dis.* 2020;23(2):232-243. doi:10.1038/s41391-020-0206-6
2. Ahmed HU, Hindley RG, Dickinson L, et al. Focal therapy for localised unifocal and multifocal prostate cancer: a prospective development study. *Lancet Oncol.* 2012;13(6):622-632. doi:10.1016/S1470-2045(12)70121-3
3. Ward JF, Jones JS. Focal cryotherapy for localized prostate cancer: a report from the national Cryo On-Line Database (COLD) Registry. *BJU Int.* 2012;109(11):1648-1654. doi:10.1111/j.1464-410X.2011.10578.x
4. Natarajan S, Jones TA, Priester AM, et al. Focal laser ablation of prostate cancer: feasibility of magnetic resonance imaging-ultrasound fusion for guidance. *J Urol.* 2017;198(4):839-847. doi:10.1016/j.juro.2017.04.017
5. Ting F, Tran M, Böhm M, et al. Focal irreversible electroporation for prostate cancer: functional outcomes and short-term oncological control. *Prostate Cancer Prostatic Dis.* 2016;19(1):46-52. doi:10.1038/pcan.2015.47
6. Moore CM, Nathan TR, Lees WR, et al. Photodynamic therapy using meso tetra hydroxy phenyl chlorin (mTHPC) in early prostate cancer. *Lasers Surg Med.* 2006;38(5):356-363. doi:10.1002/lsm.20275
7. Langley S, Ahmed HU, Al-Qaisieh B, et al. Report of a consensus meeting on focal low dose rate brachytherapy for prostate cancer. *BJU Int.* 2012;109(suppl 1):7-16. doi:10.1111/j.1464-410X.2011.10825.x
8. Priester A, Natarajan S, Khoshnoodi P, et al. Magnetic resonance imaging underestimation of prostate cancer geometry: use of patient specific molds to correlate images with whole mount pathology. *J Urol.* 2017;197(2):320-326. doi:10.1016/j.juro.2016.07.084
9. Christie DRH, Sharpley CF. How accurately can multiparametric magnetic resonance imaging measure the tumour volume of a prostate cancer? Results of a systematic review. *J Med Imaging Radiat Oncol.* 2020;64(3):398-407. doi:10.1111/1754-9485.13035
10. Pooli A, Johnson DC, Shirk J, et al. Predicting pathological tumor size in prostate cancer based on multiparametric prostate magnetic resonance imaging and preoperative findings. *J Urol.* 2021;205(2):444-451. doi:10.1097/JU.0000000000001389
11. Rischmann P, Gelet A, Riche B, et al. Focal high intensity focused ultrasound of unilateral localized prostate cancer: a prospective multicentric hemiablation study of 111 patients. *Eur Urol.* 2017;71(2):267-273. doi:10.1016/j.eururo.2016.09.039
12. Feijoo ER, Sivaraman A, Barret E, et al. Focal high-intensity focused ultrasound targeted hemiablation for unilateral prostate cancer: a prospective evaluation of oncologic and functional outcomes. *Eur Urol.* 2016;69(2):214-220. doi:10.1016/j.eururo.2015.06.018
13. Ganzer R, Hadaschik B, Pahernik S, et al. Prospective multicenter phase II study on focal therapy (hemiablation) of the prostate with high intensity focused ultrasound. *J Urol.* 2018;199(4):983-989. doi:10.1016/j.juro.2017.10.033
14. Johnson DC, Yang JJ, Kwan L, et al. Do contemporary imaging and biopsy techniques reliably identify unilateral prostate cancer? Implications for hemiablation patient selection. *Cancer.* 2019;125(17):2955-2964. doi:10.1002/cncr.32170
15. Lay N, Tsehay Y, Greer MD, et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. *J Med Imaging.* 2017;4(2):024506. doi:10.1117/1.JMI.4.2.024506
16. McGarry SD, Bukowy JD, Iczkowski KA, et al. Gleason probability maps: a radiomics tool for mapping prostate cancer likelihood in MRI space. *Tomography.* 2019;5(1):127-134. doi:10.18383/j.tom.2018.00033
17. Twilt JJ, van Leeuwen KG, Huisman HJ, Fütterer JJ, de Rooij M. Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. *Diagnostics (Basel).* 2021;11(6):959. doi:10.3390/diagnostics11060959
18. Pellicer-Valero OJ, Marenco Jiménez JL, Gonzalez-Perez V, et al. Deep learning for fully automatic detection, segmentation, and Gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. *Sci Rep.* 2022;12(1):2975. doi:10.1038/s41598-022-06730-6
19. Sushentsev N, Moreira Da Silva N, Yeung M, et al. Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. *Insights Imaging.* 2022;13(1):1-17. doi:10.1186/s13244-022-01199-3
20. Kasivisvanathan V, Stabile A, Neves JB, et al. Magnetic resonance imaging-targeted biopsy versus systematic biopsy in the detection of prostate cancer: a systematic review and meta-analysis. *Eur Urol.* 2019;76(3):284-303. doi:10.1016/j.eururo.2019.04.043
21. Priester A, Fan R, Shubert J, et al. Prediction and mapping of intraprostatic tumor extent with artificial intelligence. *Eur Urol Open Sci.* 2023;54:20-27. doi:10.1016/j.euros.2023.05.018
22. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS prostate imaging—reporting and data system: 2015, version 2. *Eur Urol.* 2016;69(1):16-40. doi:10.1016/j.eururo.2015.08.052
23. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging.* 2012;30(9):1323-1341. doi:10.1016/j.mri.2012.05.001
24. Aker MN, Brisbane WG, Kwan L, et al. Cryotherapy for partial gland ablation of prostate cancer: oncologic and safety outcomes. *Cancer Med.* 2023;12(8):9351-9362. doi:10.1002/cam4.5692
25. Johnson DC, Raman SS, Mirak SA, et al. Detection of individual prostate cancer foci via multiparametric magnetic resonance imaging. *Eur Urol.* 2019;75(5):712-720. doi:10.1016/j.eururo.2018.11.031
26. Brisbane WG, Priester AM, Ballon J, et al. Targeted prostate biopsy: umbra, penumbra, and value of perilesional sampling. *Eur Urol.* 2022;82(3):303-310. doi:10.1016/j.eururo.2022.01.008