Radiology: Imaging Cancer

ORIGINAL RESEARCH

Development and Validation of a Deep Learning Model Based on MRI and Clinical Characteristics to Predict Risk of Prostate Cancer Progression

Christian Roest, MSc[1] • *Thomas C. Kwee, MD, PhD*[1] • *Igle J. de Jong, MD, PhD*[2] • *Ivo G. Schoots, MD, PhD*[3,4] • *Pim van Leeuwen, MD, PhD*[5] • *Stijn W. T. P. J. Heijmink, MD, PhD*[4] • *Henk G. van der Poel, MD, PhD*[5,6] • *Stefan J. Fransen, MSc*[1] • *Anindo Saha, MSc*[7] • *Henkjan Huisman, PhD*[7] • *Derya Yakar, MD, PhD*[1,4]

Author affiliations, funding, and conflicts of interest are listed at the end of this article.

Radiology: Imaging Cancer 2025; 7(1):e240078 • https://doi.org/10.1148/rycan.240078 • Content codes: GU OI AI MR

Purpose: To validate a deep learning (DL) model for predicting the risk of prostate cancer (PCa) progression based on MRI and clinical parameters and compare it with established models.

Materials and Methods: This retrospective study included 1607 MRI scans of 1143 male patients (median age, 64 years; IQR, 59–68 years) undergoing MRI for suspicion of clinically significant PCa (csPCa) (International Society of Urological Pathology grade > 1) between January 2012 and May 2022 who were negative for csPCa at baseline MRI. A DL model was developed using baseline MRI and clinical parameters (age, prostate-specific antigen [PSA] level, PSA density, and prostate volume) to predict the time to PCa progression (defined as csPCa diagnosis at follow-up). Internal and external testing was performed. The model's ability to predict progression to csPCa was assessed by Cox regression analyses. Predictive performance of the DL model up to 5 years after baseline MRI in comparison with the European Randomized Study of Screening for Prostate Cancer (ERSPC) future-risk calculator, Prostate Cancer Prevention Trial (PCPT) risk calculator, and Prostate Imaging Reporting and Data System (PI-RADS) was assessed using the Harrell C-index. Optimized follow-up intervals were derived from Kaplan–Meier curves.

Results: DL scores predicted csPCa progression (internal cohort: hazard ratio [HR], 1.97 [95% CI: 1.61, 2.41; $P < .001$]; external cohort: HR, 1.32 [95% CI: 1.14, 1.55; $P < .001$]). The model identified a subgroup of patients (approximately 20%) with risks for csPCa of 3% or less, 8% or less, and 18% or less after 1-, 2-, and 4-year follow-up, respectively. DL scores had a C-index of 0.68 (95% CI: 0.63, 0.74) at internal testing and 0.56 (95% CI: 0.51, 0.61) at external testing, outperforming ERSPC and PCPT (both $P < .001$) at internal testing.

Conclusion: The DL model accurately predicted PCa progression and provided improved risk estimations, demonstrating its ability to aid in personalized follow-up for low-risk PCa.

Supplemental material is available for this article.

Prostate cancer (PCa) is the most common cancer among men, and the incidence of low-risk disease (ie, International Society of Urological Pathology [ISUP] grade < 2) is high (1). Men with low-risk PCa have a substantial risk of overdiagnosis and overtreatment, making active surveillance an increasingly adopted approach for this group (2,3). In addition to proper inclusion criteria, effective follow-up is equally vital. The ideal follow-up strategy aims to detect high-risk disease early while minimizing unnecessary examinations (4). Current strategies to reduce unnecessary biopsies include the use of an MRI-first approach, with selective biopsies for patients with MRI examinations showing findings positive for PCa (5). A combination of MRI and prostate-specific antigen (PSA) density has shown potential to further reduce the need for biopsies (6).

Risk calculators could play an important role in management of PCa by estimating personalized future risk levels that guide the timing and intensity of subsequent follow-up examinations (7). Although several established clinical tools exist for the primary diagnosis of PCa, such as the Prostate Cancer Prevention Trial (PCPT) version 2 risk calculator (8), only a few studies have explored tools for future risk of progression. One promising tool for estimating PCa risk is the European Randomized Study of Screening for Prostate Cancer (ERSPC) future-risk calculator, which assesses PCa risk after 4 years (9). Unfortunately, currently available risk calculators have limited reliability because of their dependence on clinical variables with high variability.

Although including MRI information into risk models could improve the accuracy and reliability of risk assessment tools, this is challenging because of the lack of image processing capabilities in clinical tools (10,11). Prostate Imaging Reporting and Data System (PI-RADS) (12) scores could serve as an indirect way of including MRI information, but these scores also demonstrate considerable interreader variation (13,14), which may hinder the reliability of risk estimates.

Deep learning (DL) models based on MRI features have the potential to enhance follow-up strategies by training them to capture patterns relevant to the risk of progression, and these models may offer more precise, patient-specific risk estimates for disease progression. Recent studies have demonstrated that DL imaging predictors in other cancer types, including breast cancer (15) and pancreatic cancer (16), are predictive of disease progression. However, to our knowledge, no previous study has explored this approach in the setting of PCa.

Therefore, the present study aimed to develop and validate a DL-based risk model that predicts the time to progression to clinically significant PCa (csPCa, defined as ISUP grade > 1) in

Abbreviations

csPCa = clinically significant prostate cancer, DL = deep learning, ERSPC = European Randomized Study of Screening for Prostate Cancer, HR = hazard ratio, ISUP = International Society of Urological Pathology, NKI = Netherlands Cancer Institute, PCa = prostate cancer, PCPT = Prostate Cancer Prevention Trial, PI-RADS = Prostate Imaging Reporting and Data System, PSA = prostate-specific antigen, RUMC = Radboud University Medical Center

Summary

The proposed deep learning model based on MRI and clinical characteristics showed good performance in predicting progression from low-risk prostate cancer to clinically significant prostate cancer.

Key Points

- A deep learning (DL) model developed using baseline MRI and clinical data predicted progression of low-risk prostate cancer (PCa) to clinically significant PCa on both internal (hazard ratio [HR], 1.97 [95% CI: 1.61, 2.41]; $P < .001$) and external (HR, 1.32 [95% CI: 1.14, 1.55]; $P < .001$) testing datasets.
- The DL model identified a subgroup (approximately 20% of patients) with progression risks for clinically significant PCa of 3% or less, 8% or less, and 18% or less after 1-, 2-, and 4-year follow-up, respectively.
- DL risk scores outperformed established clinical tools in predicting risk of progression to clinically significant PCa on the internal testing dataset, with the highest C-index (0.68 [95% CI: 0.63, 0.74]).

Keywords

MRI, Prostate Cancer, Deep Learning

patients with low-risk PCa, based on MRI and clinical parameters. We also compared this model with the aforementioned ERSPC future-risk and PCPT risk calculators and PI-RADS classifications.

Materials and Methods

Study Design and Sample

This retrospective study included patients from two health care institutions in the Netherlands: Radboud University Medical Center (RUMC), Nijmegen, and Netherlands Cancer Institute (NKI), Amsterdam. The need for informed consent was waived by the respective institutional review board (RUMC: IRB 2016–3045; NKI: IRB 22–159).

Patients meeting the following criteria were included: *(a)* underwent prostate MRI (referred to as "baseline MRI"); *(b)* were suspected of having PCa because of elevated PSA levels, abnormal digital rectal examination findings, or lower urinary tract symptoms; and *(c)* underwent at least one subsequent biopsy or follow-up MRI 6 months after the baseline MRI. The exclusion criteria were as follows: *(a)* biopsy-confirmed csPCa before or within 6 months after the baseline MRI; *(b)* previous treatment for prostate cancer; *(c)* poor quality of the baseline MRI scan; *(d)* missing axial T2-weighted or diffusion-weighted imaging; and *(e)* positive (defined as PI-RADS score ≥ 3) follow-up MRI findings without histopathologic confirmation. Relevant clinical parameters, including PSA, PSA density, prostate volume, patient age, and baseline ISUP grade group, were extracted from patient files.

Multiple MRI examinations for a patient meeting inclusion and exclusion criteria were treated as distinct observations, with each additional scan serving as a new baseline for risk prediction. Figure 1 presents flow diagrams showing patient selection in both the internal and external cohorts, in accordance with Standards for Reporting of Diagnostic Accuracy Studies (17).

The study used the ISUP grade group derived from biopsy, prostatectomy, or transurethral resection of the prostate as the reference standard (1). PCa progression was defined as the detection of csPCa at follow-up histopathologic assessment. For patients without documented progression, data points were censored on the date of the last negative radiologic or histopathologic assessment before the cutoff date. Follow-up was truncated at 5 years after MRI.

Imaging Protocol

All patients underwent biparametric or multiparametric MRI of the prostate. For each MRI examination, axial T2-weighted imaging, monoexponential apparent diffusion coefficient maps, and diffusion-weighted images with b value of 1400 sec/mm^2 or greater were extracted. Dynamic contrast-enhanced imaging was excluded from this study because of its limited role in current prostate MRI guidelines (18,19). Detailed information regarding imaging parameters can be found in Table S1.

Time-to-Progression Prediction Models

We developed a novel DL model to predict PCa progression risk using baseline MRI and clinical parameters (ie, age, PSA, PSA density, and prostate volume). The model consisted of two main components. First, a U-Net model was used to automatically identify regions suspicious for PCa in the baseline MRI scans, generating heatmaps representing the likelihood of pathology-confirmed PCa lesions (ISUP grade ≥ 1) at a voxel level (20,21). These heatmaps, together with clinical variables and the time interval between MRI and follow-up, were passed to a DL classifier model, which was trained to predict, at the patient level, the likelihood of detecting csPCa following each patient's respective progression or censoring time (9). The model was optimized by minimizing the binary cross-entropy between the predicted and target label. During inference, personalized risk scores were computed for a fixed 4-year interval, facilitating long-term risk estimation. Because of the overall shorter follow-up period across the cohort, we chose to predict at 4 years; the sample size at this time period was sufficient for accurate predictions.

On the basis of their predicted risk scores, patients were assigned to one of five risk categories (referred to as "DL risk groups"), aligning with established Likert-based scoring systems, such as PI-RADS and ISUP. To address variations in cohort constitution affecting risk score distribution (eg, scanner parameters, institutional practices) and to optimize clinical utility (6), we used a dedicated 5% calibration set to establish institution-specific thresholds, grouping the remaining patients into five risk categories of similar size. Fivefold cross-validation was performed using the RUMC (internal) cohort to obtain predictions for each observation, enabling a well-powered internal test. External testing was conducted using the NKI (external) cohort by averaging predictions from the five cross-validated models trained on internal data. Further details on data preprocessing, DL architecture, and training parameters are provided in Appendix S1.

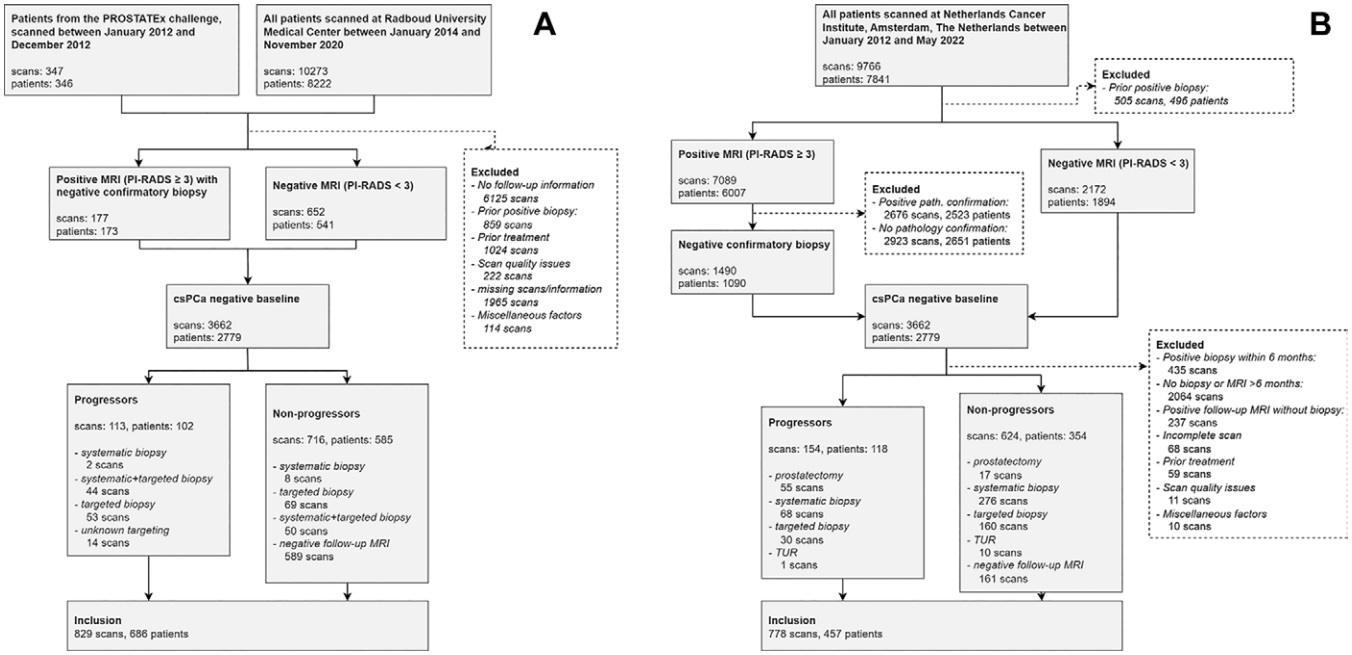

Figure 1: Flow diagrams show patient selection in the **(A)** Radboud University Medical Center and **(B)** Netherlands Cancer Institute cohorts. csPCa = clinically significant prostate cancer, PI-RADS = Prostate Imaging Reporting and Data System, TUR = transurethral resection of the prostate.

Our DL model was compared with established risk tools: ERSPC future-risk calculator (9) and PCPT risk calculator version 2 (9,22). The PCPT risk calculator lacked suggested cutoff values, so we stratified PCPT risk scores into five equally sized bins to allow Kaplan–Meier analysis, with "PCPT risk 1" and "PCPT risk 5" referring to the 20% lowest risk predictions and 20% highest risk scores, respectively. We also compared our model to PI-RADS scores, following contemporary PI-RADS guidelines (18). The code is available at the following repository: *https://github.com/0xC4/mri-risk*.

Outcomes

The primary outcome was the detection of progression to PCa with an ISUP grade of 2 or higher at follow-up histopathologic assessment. Optimized follow-up schedules were derived from Kaplan–Meier curves for each predictor's risk groups. Inspired by established guidelines for follow-up after CT screening for lung cancer, we extracted follow-up times based on a 10% risk threshold for missing csPCa (7). This threshold was considered conservative in comparison to previously reported detection rates for csPCa at repeat biopsy of 13% (23). Secondary outcomes included time-to-progression analyses for the detection of PCa with ISUP grade of 3 or higher.

Statistical Analysis

Baseline characteristics (including clinical parameters, PI-RADS classification, and ISUP grades) were compared between patients with and without disease progression. The Harrell C-index was used to assess the predictive value for the risk of progression to csPCa of each predictor (DL risk scores, ERSPC future-risk scores, PCPT scores, and PI-RADS scores) in internal and external data (24). Differences in C-index between predictors were assessed for significance using the method of Kang et al (25). Cox proportional hazard models were fit to assess the predictor significance using a Wald test (26). ERSPC, PCPT, and DL scores were modeled as continuous parameters to preserve the granularity of the data. PI-RADS scores, in contrast, were modeled as categorical predictors to account for the potentially distinct effect that each category may have on the outcome. Continuous predictors were standardized before Cox model fitting to allow for meaningful effect size comparisons because the resulting exponentiated coefficient is the hazard ratio (HR) corresponding to a 1-SD increase in the predictor. Cox models were adjusted for within-subject correlation using a marginal model with clustering by patient identification number (27). Proportional hazards were assessed using the Schoenfeld test, and linearity was evaluated via a model fitted with restricted cubic splines (28). Analyses were repeated for the internal and external cohorts. All statistical analyses were performed using R software, 4.2.2 (R Foundation for Statistical Computing). *P* values less than .05 indicated a statistically significant difference.

Results

Study Sample

After exclusion criteria were applied, 1607 MRI examinations from 1143 male patients (median age, 64 years; IQR, 59–68 years) were included as baseline scans for risk assessment. Progression to csPCa was observed in 113 examinations (13.6%) in the internal cohort and 154 examinations (19.8%) in the external cohort. Median follow-up was 2.1 years (IQR, 1.2–3.2 years) and 2.5 years (IQR, 1.4–4.2 years), respectively. Baseline characteristics are presented in Table 1.

Time to Progression to PCa with ISUP Grade 2 or Higher

On the internal testing dataset, DL-based risk scores significantly predicted PCa progression (HR, 1.97 [95% CI: 1.61,

Table 1: Overview of the Patient Characteristics for the Internal and External Cohorts

	RUMC			NKI		
Baseline Characteristic	Progression	No Progression/Censored	*P* Value	Progression	No Progression/Censored	*P* Value
No. of patients	113 (13.6)	716 (86.4)		154 (19.8)	624 (80.2)	
Age (y)	67 (62–70)	63 (59–67)	<.001	66 (59–69)	64 (59–69)	.48
PSA level (μg/L)	7.5 (5.5–10.6)	7.5 (5.5–11)	.41	7.7 (5.4–10.2)	7.4 (4.9–11.1)	.78
Prostate volume (mL)	57 (42–77)	74 (54–98)	<.001	41 (33–55)	55 (40–80)	<.001
PSA density (μg/mL^2)	0.14 (0.09–0.19)	0.11 (0.08–0.15)	<.001	0.18 (0.13–0.26)	0.13 (0.09–0.2)	<.001
Time until event/censor (y)	1.7 (1.1–2.7)	2.2 (1.3–3.3)		2.1 (1.2–3.1)	2.8 (1.6–4.7)	
PI-RADS score			<.001			<.001
1–2	58 (51.3)	594 (84.7)		11 (7.1)	136 (21.8)	
3	8 (7.1)	26 (3.8)		10 (6.5)	120 (19.2)	
4	29 (25.7)	69 (10.1)		73 (47.4)	211 (33.8)	
5	18 (15.9)	10 (1.5)		60 (39)	157 (25.2)	
Baseline pathologic result			<.001			.04
ISUP grade < 1	23 (20.4)	83 (12.1)		23 (14.9)	148 (23.7)	
ISUP grade 1	33 (29.2)	35 (5.1)		123 (79.9)	423 (67.8)	
Not performed (MRI-negative)	57 (50.4)	566 (82.8)		8 (5.2)	53 (8.5)	
Follow-up pathologic result						
ISUP grade 2	67 (59.2)	...		111 (72.1)	...	
ISUP grade 3	22 (19.5)	...		31 (20.1)	...	
ISUP grade 4	15 (13.3)	...		5 (3.3)	...	
ISUP grade 5	9 (8)	...		7 (4.6)	...	

Note.—Continuous parameters are reported as medians, with IQRs in parentheses, and were tested for significance using univariate Cox proportional hazards models; categorical parameters are reported as frequencies, with percentages in parentheses, and were tested for significance using log-rank tests. ISUP = International Society of Urological Pathology, NKI = Netherlands Cancer Institute, PI-RADS = Prostate Imaging Reporting and Data System, PSA = prostate-specific antigen, RUMC = Radboud University Medical Center.

2.41]; $P < .001$). The C-index for DL was 0.68 (95% CI: 0.63, 0.74) (Table 2), which was higher than that for ERSPC (C-index, 0.59 [95% CI: 0.52, 0.65]; $P < .001$) and PCPT (C-index, 0.59 [95% CI: 0.52, 0.65]; $P < .001$). Kaplan–Meier curves for the internal testing show the time to progression over 5 years of follow-up (Fig 2).

During external testing, DL scores significantly predicted PCa progression (HR, 1.32 [95% CI: 1.14, 1.55]; $P < .001$). The C-index was 0.56 (95% CI: 0.51, 0.61) (Table 2). No evidence of differences was found when we compared the predictive performance of DL risk scores to ERSPC (C-index, 0.61 [95% CI: 0.56, 0.66]; $P = .11$) and PCPT (C-index, 0.56 [95% CI: 0.51, 0.61]; $P = .95$). PI-RADS significantly outperformed DL scores (C-index, 0.63 [95% CI: 0.59, 0.68]; $P = .01$). Kaplan–Meier curves for the external testing show the time to progression over 5 years of follow-up (Fig 3).

In the multivariable Cox regression analysis, both DL risk scores (HR, 1.66 [95% CI: 1.32, 2.09]; $P < .001$) and PI-RADS scores were significant predictors of PCa progression. Higher PI-RADS scores were associated with an increased risk: PI-RADS score of 3 to 5 had HRs ranging from 2.18 to 7.96 (all $P \leq .048$). In the external cohort, all models except PCPT were significantly associated with progression, with DL risk scores (HR, 1.28 [95% CI: 1.07, 1.53]; $P = .006$) and ERSPC (HR, 1.60 [95% CI: 1.09, 2.36]; $P = .02$) showing significant associations, along with PI-RADS scores of 4 and 5 (HRs, 2.44 and 3.14; $P \leq .007$).

Optimized Follow-up Schedules

Figure 4 illustrates the optimized follow-up times for the internal evaluation. DL scores identified a patient stratum with the 48% lowest DL risk scores (DL risk group ≤ 2), who could delay follow-up for 3.5 years with less than 10% risk of missing progression to csPCa. Conversely, patients assigned to DL risk group of 5 or higher reached a greater than 10% risk of missed csPCa within 1 year.

On external testing, in the 15% of patients with the lowest DL-predicted risk (DL risk group ≤ 1), follow-up could be delayed for 2 years with less than 10% risk of missing progression to csPCa, and for 3 years at a risk of 13% (Fig 4). Patients assigned to DL risk group 2 or higher reached 10% risk after 2 years.

Across both cohorts, the lowest DL risk group had progression risks for csPCa of 3% or less, 8% or less, and 18% or less after 1-, 2-, and 4-year follow-up, respectively. Conversely, patients in DL risk group 5 reached a 29% or greater risk of csPCa progression within 3 years.

Risk of Progression to PCa with ISUP Grade 3 or Higher

DL risk scores predicted the risk of progression to PCa with ISUP grade of 3 or higher in both the internal (HR, 1.87 [95% CI: 1.37, 2.55]; $P < .001$) and external (HR, 1.44 [95% CI: 1.08, 1.93]; $P = .01$) cohorts, with C-index values of 0.70 (95% CI: 0.63, 0.78) and 0.57 (95% CI: 0.5, 0.65), respectively. Ka-

Table 2: Univariable and Multivariable Cox Regression Results and C-Indexes for All Predictors

	RUMC			NKI		
Variable	C-Index	HR	*P* Value	C-Index	HR	*P* Value
Univariable						
DL scores	0.68 (0.63, 0.74)	1.97 (1.61, 2.41)	<.001*	0.56 (0.51, 0.61)	1.32 (1.14, 1.55)†	<.001*
ERSPC future-risk calculator	0.59 (0.52, 0.65)	0.81 (0.65, 1.01)	.06	0.61 (0.56, 0.66)	1.24 (1.07, 1.43)†‡	.004§
PCPT risk calculator	0.59 (0.52, 0.65)	1.09 (0.92, 1.3)	.25	0.56 (0.51, 0.61)	1.09 (0.91, 1.27)	.27
PI-RADS score	0.72 (0.67, 0.77)			0.63 (0.59, 0.68)		
1/2	...	Reference	...	...	Reference	...
3	...	2.69 (1.25, 5.79)	<.01‖	...	0.86 (0.34, 2.15)	.75
4	...	4.32 (2.7, 6.9)	<.001*	...	2.68 (1.4, 5.16)	.003§
5	...	10.55 (5.92, 18.8)	<.001*	...	3.67 (1.87, 7.19)	<.001*
Multivariable						
DL scores	...	1.66 (1.32, 2.09)	<.001*	...	1.28 (1.07, 1.53)	.006§
ERSPC future-risk calculator	...	0.96 (0.54, 1.7)	.88	...	1.6 (1.09, 2.36)	.02‖
PCPT risk calculator	...	1.26 (0.76, 2.11)	.37	...	0.67 (0.44, 1.03)	.07
PI-RADS score					...	
1/2	...	Reference	...	...	Reference	...
3	...	2.18 (1, 4.74)	.048‖	...	0.81 (0.33, 2.032)	.67
4	...	3.35 (1.8, 6.24)	<.001*	...	2.44 (1.27, 4.7)	.007§
5	...	7.96 (3.96, 15.97)	<.001*	...	3.14 (1.59, 6.19)	<.001*

Note.—Hazard ratios represent the change in risk associated with a 1-SD increase in the predictors. Data in parentheses are 95% CIs. DL = deep learning, ERSPC = European Randomized Study of Screening for Prostate Cancer, HR = hazard ratio, PI-RADS = Prostate Imaging Reporting and Data System, PCPT = Prostate Cancer Prevention Trial.
* $P < .001$.
† Predictors for which proportional hazards assumptions were not met.
‡ Predictors for which linearity assumptions were not met.
§ $P < .01$.
‖ $P < .05$.

plan–Meier curves for time to progression for ISUP grade of 3 or higher are included in Figure S2.

Discussion

We introduced a DL model based on clinical and MRI characteristics to predict the risk of csPCa progression and demonstrated potentially useful follow-up schedules for patients with low-risk PCa based on their predicted risk scores. The DL model achieved a promising performance on both internal (C-index, 0.68 [95% CI: 0.63, 0.74]; HR, 1.97 [95% CI: 1.61, 2.41; $P < .001$]) and external (C-index, 0.56 [95% CI: 0.51, 0.61]; HR, 1.32 [95% CI: 1.14, 1.55; $P < .001$]) testing datasets in predicting disease progression in comparison to established clinical risk calculators and PI-RADS. In our analysis, follow-up schedules based on DL risk scores would have allowed a substantial proportion of patients to forego follow-up for 3 years with 13% or lower risk of missing disease progression. External testing further demonstrated the model's prognostic value and showed its generalizability across institutions. Our proposed method could help to reduce the overuse of biopsies and MRI by guiding follow-up frequency in low-risk PCa surveillance. To encourage further research, we provide public access to the code and trained DL models.

Our DL model showed robust prognostic value across internal and external cohorts. In contrast, existing clinical calculators (ERSPC and PCPT) predicted PCa progression only in the external cohort. Despite similar C-indexes among the predictors in external data, the DL risk scores exhibited stronger effect sizes for predicting PCa progression compared with the clinical calculators. This suggests that DL-based risk estimates could enhance the robustness of risk assessments across health care centers compared with existing tools. Although PI-RADS scores showed larger effect sizes for PCa progression than DL scores, their performance may have been inflated because of their inclusion in the reference standard for negative follow-up, known as verification bias.

DL scores also predicted the risk of progression to higher ISUP grading (ie, PCa with ISUP grade ≥ 3). This result is particularly relevant because some active surveillance protocols may allow patients who have PCa with an ISUP grade of 2 to enroll under specific conditions (eg, the absence of cribriform growth [4]) or use less strict criteria (eg, ISUP grade ≥ 3) to trigger active treatment (29,30). Furthermore, surveillance inclusion and termination criteria are still subject to change. The model's ability to estimate the risk of progression to PCa with ISUP grade 3 or higher suggests the utility of DL risk scores across diverse surveillance protocols.

The proposed DL model may also have value for optimizing patient selection by identifying patients at higher risk of developing aggressive disease. For example, our DL model identified patients at 29% or greater risk of csPCa progression within 3 years. Whether patients with such a high risk of progression can

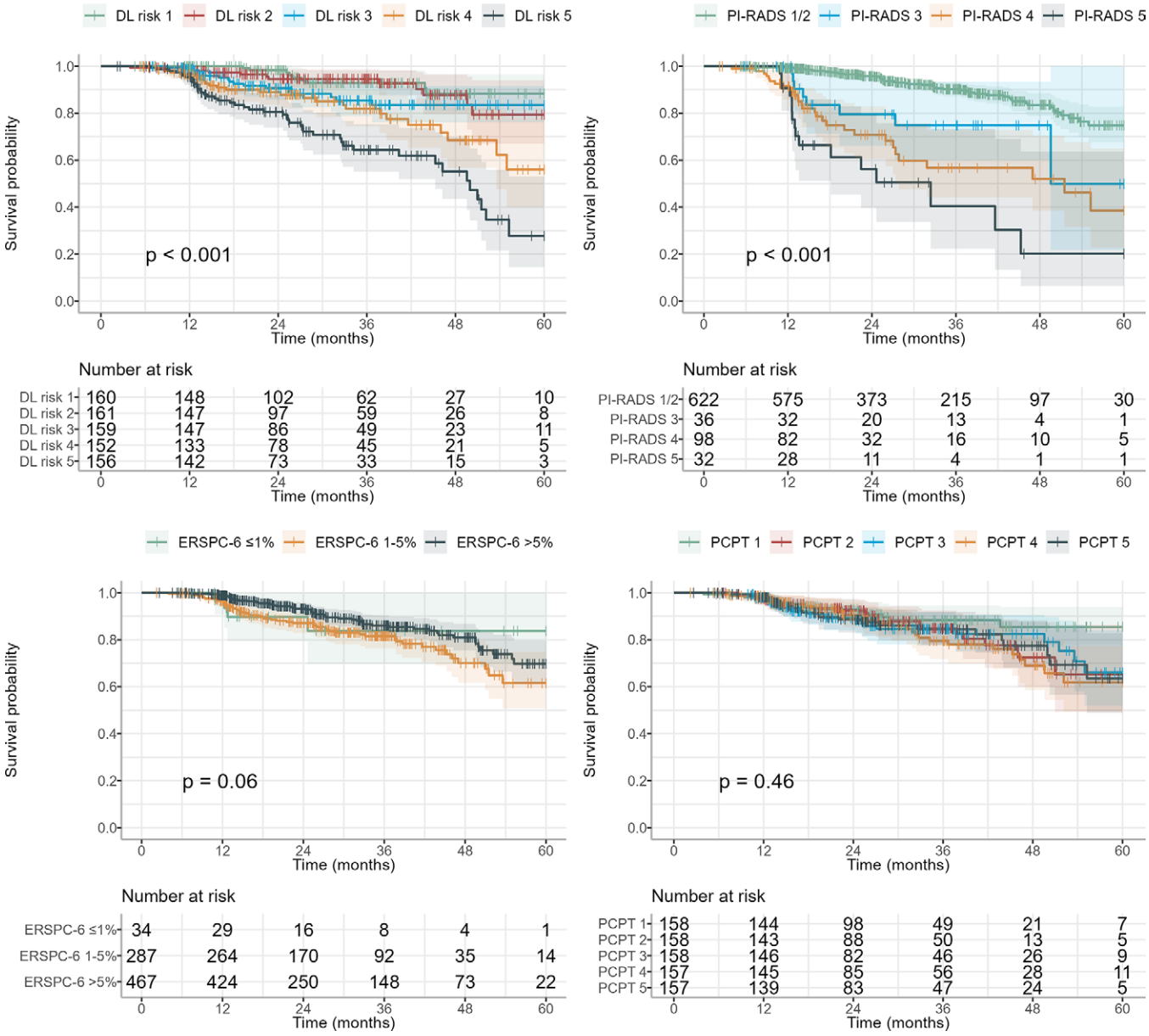

Figure 2: Kaplan–Meier curves for time to prostate cancer progression in the internal (Radboud University Medical Center) cohort, stratified by **(A)** deep learning (DL) predictions, **(B)** Prostate Imaging Reporting and Data System (PI-RADS), **(C)** European Randomized Study of Screening for Prostate Cancer (ERSPC) future-risk calculator scores, and **(D)** Prostate Cancer Prevention Trial (PCPT) risk calculator scores.

be considered suitable candidates for active surveillance should be decided on an individual basis. Similarly, patients with PI-RADS grade 5 lesions were at considerably higher risk of csPCa detection at follow-up in both cohorts (34% and 16% after 1.5 years, respectively). Currently, none of the published surveillance guidelines incorporate MRI information in their eligibility criteria (4,30,31). Thus, there may be an opportunity to refine inclusion criteria and enhance overall patient outcomes.

We presented optimized follow-up times at a 10% risk level as an illustrative example, inspired by established guidelines for the follow-up of pulmonary nodules detected in lung cancer screening (7). Currently, no such agreed-upon cutoff exists for low-risk PCa management. By providing validated risk levels for different risk calculators, this research promotes discussions on the determination of achievable risk thresholds in low-risk PCa follow-up.

Our results are supported by previous works. In 2020, Wang et al (11) evaluated the PI-RADS score as a marker for PCa progression in 344 patients enrolled in active surveillance in a single center and detected csPCa in 30% of patients classified as low-risk 4 years after baseline MRI. Likewise, a systematic review by Grivas et al in 2022 (32) revealed that higher reported MRI suspicion scores (ie, PI-RADS or Likert scores) were associated with an elevated csPCa detection rate at repeat biopsy following an initial negative biopsy result. We identified only one previous study that used artificial intelligence for risk estimation (Jia et al, 2022 [10]), which showed that MRI radiomics predicted progression-free survival in 191 patients with localized PCa over 3 years. Although these previous works provided evidence for the use of MRI for PCa risk estimation, they were limited by small cohort sizes and lacked comparisons with existing clinical tools. The present

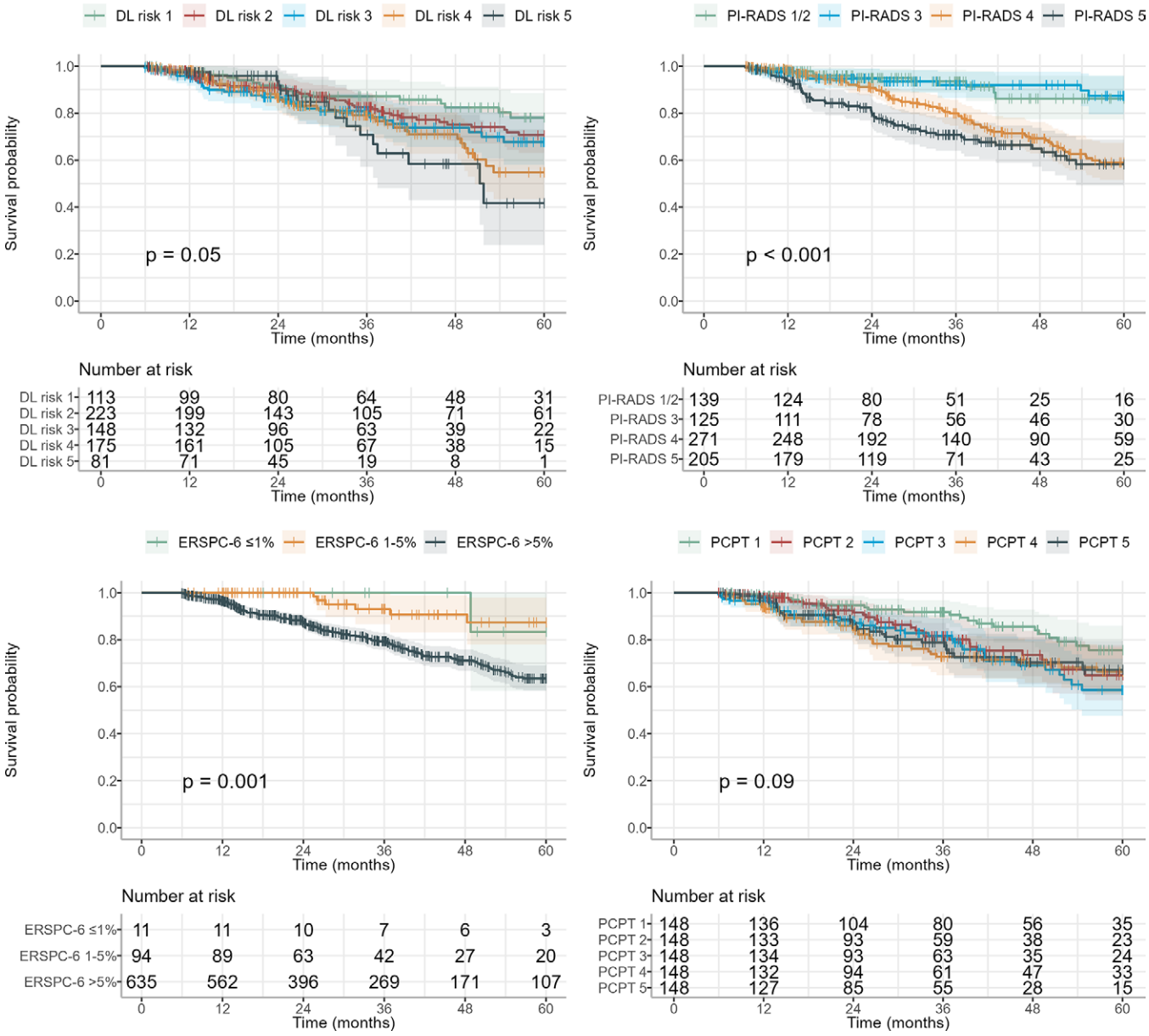

Figure 3: Kaplan–Meier curves for time to prostate cancer progression in the external (Netherlands Cancer Institute) cohort, stratified by **(A)** deep learning (DL) predictions, **(B)** Prostate Imaging Reporting and Data System (PI-RADS), **(C)** European Randomized Study of Screening for Prostate Cancer (ERSPC) future-risk calculator scores, and **(D)** Prostate Cancer Prevention Trial (PCPT) risk calculator scores.

study used a substantially larger patient sample and validated the results in a multicenter setting.

Our DL risk estimation may be adapted to clinical purposes beyond PCa progression. Most risk models are based on statistical regression and thus cannot incorporate imaging information directly without manually reducing the image data to features, such as through radiomics extraction or by using PI-RADS scores. We show that imaging data and clinical values can be combined in a multimodal DL model to generate useful risk predictions.

Our study had limitations. First, this study was retrospective and was thus subject to inherent biases associated with the collection of historical data. Second, there was considerable heterogeneity between the baseline pathology grades of the internal and external cohorts, which can be attributed to differences in institutional standards. In general, RUMC avoided biopsies in patients without any visible lesions at MRI, resulting in a considerable proportion of patients included on the basis of negative MRI results. In contrast, NKI typically conducted biopsies before the initial MRI being performed. However, the model's consistent performance across both cohorts demonstrates robustness to diverse clinical settings. Third, the broad time span for patient inclusion may have led to variability in MRI scan quality, reflecting technologic advancements. Fourth, although some assumptions for proportional hazards models were not met, these deviations can be considered minor, given that the primary aim was to assess predictor significance. Despite a suboptimal fit, all affected predictors showed statistical significance, and their *P* values would likely have been more noteworthy with correct modeling (28). Nonetheless, caution is advised in interpreting HRs from these models because they may produce erroneous predictions in certain intervals. Finally, our study sample and outcome differed from the intended population and outcome of the PCPT risk calculator. A lack of publicly available future risk models motivated our decision to

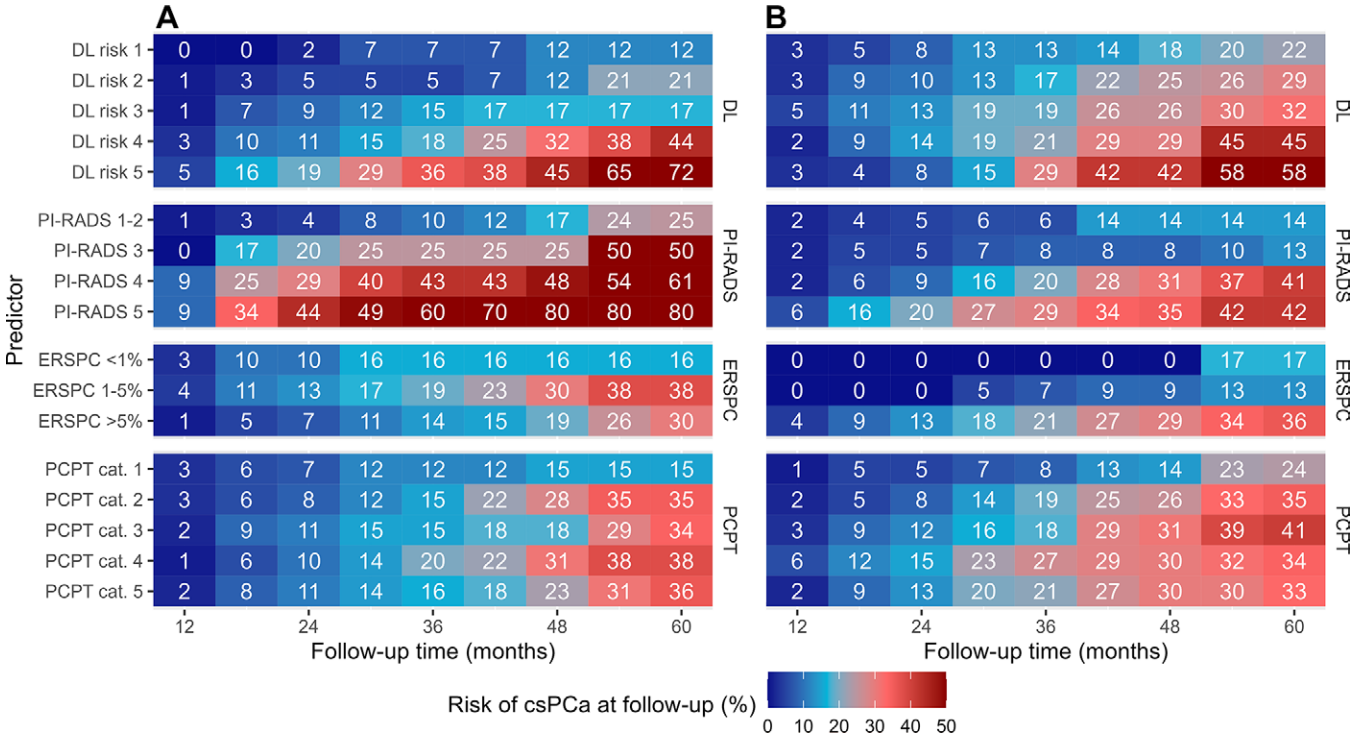

Figure 4: Heatmaps show risk of clinically significant prostate cancer (csPCa) progression (International Society of Urological Pathology grade ≥ 2) at follow-up in **(A)** internal cross-validation and **(B)** external testing cohorts, stratified by deep learning (DL) risk scores and other prediction models. Models include the European Randomized Study of Screening for Prostate Cancer (ERSPC) future-risk calculator, Prostate Cancer Prevention Trial (PCPT) risk calculator, and Prostate Imaging Reporting and Data System (PI-RADS).

include this widely used model as a benchmark. In addition, its inclusion provides insights into the applicability of primary diagnostic tools for predicting the time to csPCa progression.

In conclusion, a DL-based prediction model was developed using MRI and clinical parameters and accurately predicted progression to csPCa up to 5 years after the MRI. DL-based risk estimates may help to individualize follow-up times to the risk level of the patient. Prospective studies with longer follow-up times are needed to confirm these results.

Author affiliations:
[1] Department of Radiology, University Medical Center Groningen, Groningen, the Netherlands

[2] Department of Urology, University Medical Center Groningen, Groningen, the Netherlands

[3] Department of Radiology and Nuclear Medicine, Erasmus MC Cancer Institute, University Medical Centre, Rotterdam, the Netherlands

[4] Department of Radiology, Netherlands Cancer Institute, Amsterdam, the Netherlands

[5] Department of Urology, Netherlands Cancer Institute, Amsterdam, the Netherlands

[6] Department of Urology, Amsterdam University Medical Centers, Amsterdam, the Netherlands

[7] Department of Medical Imaging, Radboudumc, Nijmegen, the Netherlands

Received March 19, 2024; revision requested April 25; revision received September 10; accepted October 31.

Address correspondence to C.R. (email: *c.roest@umcg.nl*).

Funding: Supported by Siemens Healthineers.

Author contributions: Guarantors of integrity of entire study, **C.R., T.C.K., D.Y.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **C.R., D.Y.**; clinical studies, **C.R., I.J.d.J., S.W.T.P.J.H., H.G.v.d.P.**; experimental studies, **C.R., H.G.v.d.P., A.S., H.H., D.Y.**; statistical analysis, **C.R., H.H.;** and manuscript editing, **C.R., T.C.K., I.J.d.J., I.G.S., P.v.L., S.W.T.P.J.H., H.G.v.d.P., S.J.F., H.H., D.Y.**

Disclosures of conflicts of interest: C.R. Support from Siemens Healthineers. **T.C.K.** Support from Siemens Healthineers. **I.J.d.J.** Payment or honoraria from Bayer for postgrade course; on Bayer advisory board. **I.G.S.** No relevant relationships. **P.v.L.** No relevant relationships. **S.W.T.P.J.H.** No relevant relationships. **H.G.v.d.P.** No relevant relationships. **S.J.F.** Funding was provided by Siemens Healthineers and Health-Holland (grant number: LSHM20103) (payment to institution). **A.S.** No relevant relationships. **H.H.** Research support from Siemens Healthineers. **D.Y.** Siemens Healthineers research grant (payment made to institution); NWO grant, Hanarth fund grant, and Health Holland grant (payment made to institution); payment or honoraria to author from Astellas; speaker fee from Bayer; MDPI travel grant paid to author; advisory board EIBIR and advisory board ICAI lab (no payments to anyone).

References

1. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: definition of grading patterns and proposal for a new grading system. Am J Surg Pathol 2016;40(2):244–252.
2. Cooperberg MR, Meeks W, Fang R, Gaylis FD, Catalona WJ, Makarov DV. Time trends and variation in the use of active surveillance for management of low-risk prostate cancer in the US. JAMA Netw Open 2023;6(3):e231439.
3. Chen RC, Rumble RB, Loblaw DA, et al. Active surveillance for the management of localized prostate cancer (Cancer Care Ontario guideline): American Society of Clinical Oncology clinical practice guideline endorsement. J Clin Oncol 2016;34(18):2182–2190.
4. van den Bergh RC, Roemeling S, Roobol MJ, Roobol W, Schröder FH, Bangma CH. Prospective validation of active surveillance in prostate cancer: the PRIAS study. Eur Urol 2007;52(6):1560–1563.
5. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389(10071):815–822.
6. Deniffel D, Healy GM, Dong X, et al. Avoiding unnecessary biopsy: MRI-based risk models versus a PI-RADS and PSA density strategy for clinically significant prostate cancer. Radiology 2021;300(2):369–379.
7. Callister ME, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. Thorax 2015;70(Suppl 2):ii1–ii54 [Published correction appears in Thorax 2015;70(12):1188.].
8. Grill S, Fallah M, Leach RJ, et al. Incorporation of detailed family history from the Swedish Family Cancer Database into the PCPT risk calculator. J Urol 2015;193(2):460–465.

9. Roobol MJ, Zhu X, Schröder FH, et al. A calculator for prostate cancer risk 4 years after an initially negative screen: findings from ERSPC Rotterdam. Eur Urol 2013;63(4):627–633.
10. Jia Y, Quan S, Ren J, et al. MRI radiomics predicts progression-free survival in prostate cancer. Front Oncol 2022;12:974257 [Published correction appears in Front Oncol 2023;12:1125641.].
11. Wang AZ, O'Conno LP, Yerram NK, et al. PI-RADS® Category as a predictor of progression to unfavorable risk prostate cancer in men on active surveillance. J Urol 2020;204(6):1229–1235 [Published correction appears in J Urol 2021;205(2):638.].
12. Rosenkrantz AB, Kim S, Lim RP, et al. Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales. Radiology 2013;269(2):482–492.
13. Westphalen AC, McCulloch CE, Anaokar JM, et al. Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel. Radiology 2020;296(1):76–84.
14. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS Version 2 Lexicon: a multicenter study of six experienced prostate radiologists. Radiology 2016;280(3):793–804.
15. Dammu H, Ren T, Duong TQ. Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. PLoS One 2023;18(1):e0280148.
16. Yao J, Cao K, Hou Y, et al. Deep learning for fully automated prediction of overall survival in patients undergoing resection for pancreatic cancer: a retrospective multicenter study. Ann Surg 2023;278(1):e68–e79.
17. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6(11):e012799.
18. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 update of Prostate Imaging Reporting and Data System version 2. Eur Urol 2019;76(3):340–351.
19. Belue MJ, Yilmaz EC, Daryanani A, Turkbey B. Current status of biparametric MRI in prostate cancer diagnosis: literature analysis. Life (Basel) 2022;12(6):804.
20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, 2015; 234–241.
21. Roest C, Kwee TC, Saha A, Fütterer JJ, Yakar D, Huisman H. AI-assisted biparametric MRI surveillance of prostate cancer: feasibility study. Eur Radiol 2023;33(1):89–96.
22. Ankerst DP, Hoefler J, Bock S, et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. Urology 2014;83(6):1362–1367.
23. Bokhorst LP, Valdagni R, Rannikko A, et al. A decade of active surveillance in the PRIAS Study: an update and evaluation of the criteria used to recommend a switch to active treatment. Eur Urol 2016;70(6):954–960.
24. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361–387.
25. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. Stat Med 2015;34(4):685–703.
26. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 1943;54(3):426–482.
27. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. In: Statistics for Biology and Health. Springer, 2000.
28. Harrell FE Jr. Regression modeling strategies. In: Springer Series in Statistics. Springer, 2015.
29. Selvadurai ED, Singhera M, Thomas K, et al. Medium-term outcomes of active surveillance for localised prostate cancer. Eur Urol 2013;64(6):981–987.
30. de Vos II, Luiting HB, Roobol MJ. Active surveillance for prostate cancer: past, current, and future trends. J Pers Med 2023;13(4):629.
31. Komisarenko M, Martin LJ, Finelli A. Active surveillance review: contemporary selection criteria, follow-up, compliance and outcomes. Transl Androl Urol 2018;7(2):243–255.
32. Grivas N, Lardas M, Espinós EL, et al. Prostate cancer detection percentages of repeat biopsy in patients with positive multiparametric magnetic resonance Imaging (Prostate Imaging Reporting and Data System/Likert 3-5) and negative initial biopsy. A mini systematic review. Eur Urol 2022;82(5):452–457.