Prostate Age Gap: An MRI Surrogate Marker of Aging for Prostate Cancer Detection

Alvaro Fernandez-Quilez, PhD,[1,2,3]* ● Tobias Nordström, MD, PhD,[3,4]

Fredrik Jäderling, MD, PhD,[5,6] Svein Reidar Kjosavik, MD, PhD,[7] and Martin Eklund, PhD[3]

Background: Aging is the most important risk factor for prostate cancer (PC). Imaging techniques can be useful to measure age-related changes associated with the transition to diverse pathological states. However, biomarkers of aging from prostate magnetic resonance imaging (MRI) remain to be explored.

Purpose: To develop an aging biomarker from prostate MRI and to examine its relationship with clinically significant PC (csPC, Gleason score ≥7) risk occurrence.

Study Type: Retrospective.

Population: Four hundred and sixty-eight (65.97 ± 6.91 years) biopsied males, contributing 7243 prostate MRI slices. A deep learning (DL) model was trained on 3223 MRI slices from 81 low-grade PC (Gleason score ≤6) and 131 negative patients, defined as non-csPC. The model was tested on 90 negative, 52 low-grade (142 non-csPC), and 114 csPC patients.

Field Strength/Sequence: 3-T, axial T2-weighted spin sequence.

Assessment: Chronological age was defined as the age of the participant at the time of the visit. Prostate-specific antigen (PSA), prostate volume, Gleason, and Prostate Imaging-Reporting and Data System (PI-RADS) scores were also obtained. Manually annotated prostate masks were used to crop the MRI slices, and a DL model was trained with those from non-csPC patients to estimate the age of the patients. Following, we obtained the prostate age gap (PAG) on previously unseen csPC and non-csPC cropped MRI exams. PAG was defined as the estimated model age minus the patient's age. Finally, the relationship between PAG and csPC risk occurrence was assessed through an adjusted multivariate logistic regression by PSA levels, age, prostate volume, and PI-RADS ≥ 3 score.

Statistical Tests: T-test, Mann–Whitney U test, permutation test, receiver operating characteristics (ROC), area under the curve (AUC), and odds ratio (OR). A P value <0.05 was considered statistically significant.

Results: After adjusting, there was a significant difference in the odds of csPC (OR = 3.78, 95% confidence interval [CI]: 2.32–6.16). Further, PAG showed a significantly larger bootstrapped AUC to discriminate between csPC and non-csPC than that of adjusted PI-RADS ≥ 3 (AUC = 0.981, 95% CI: 0.975–0.987).

Data Conclusion: PAG may be associated with the risk of csPC and could outperform other PC risk factors.

Level of Evidence: 3

Technical Efficacy: Stage 3

Population-based unorganized screening for prostate cancer (PC) has commonly relied on prostate-specific antigen (PSA) levels in serum for an initial risk assessment of PC.[1,2]

Typically, elevated PSA levels are used to refer patients to a systematic or transrectal ultrasound-guided biopsy and have been associated with a decrease in PC mortality.[3] However,
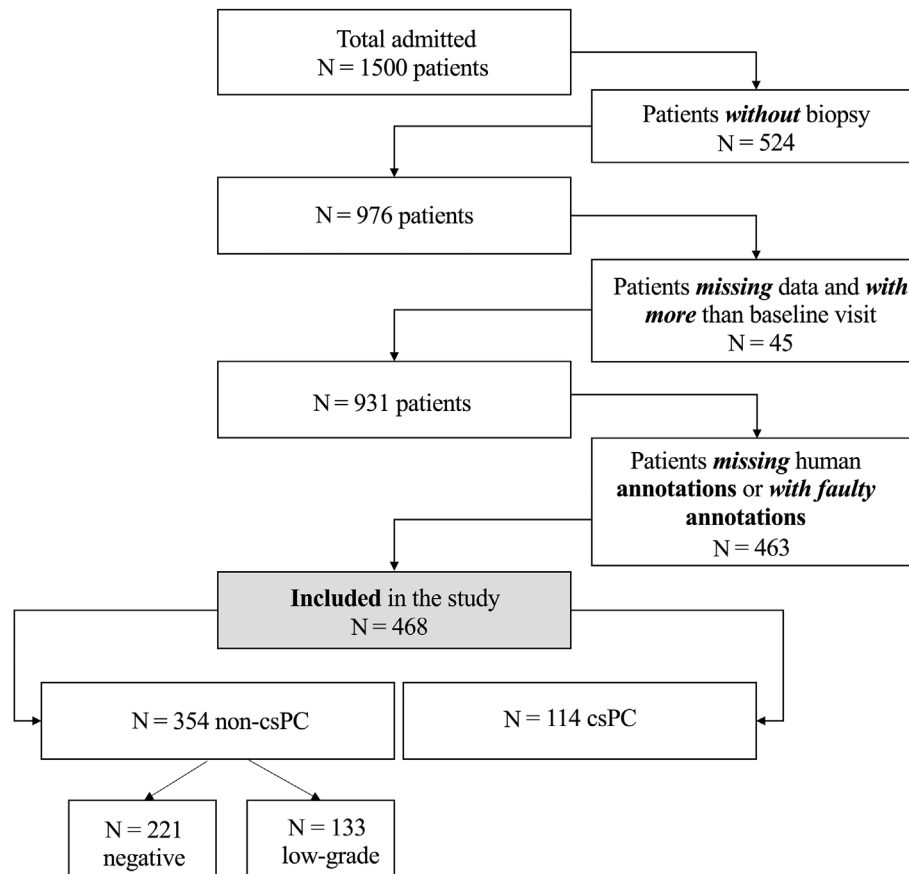
FIGURE 1: Patient inclusion criteria.

no general consensus exists regarding the threshold to define abnormal PSA levels.[3,4] Further, the associations with over-detection of low-grade PC (Gleason score ≤6) and subsequent unnecessary testing practices, treatment, and related complications have resulted in the absence of governmental bodies recommending organized PC screening solely relying on PSA testing.[4,5]

Over the last decade, multi-parametric magnetic resonance imaging (mp-MRI) has emerged as an important component for PC detection, staging, treatment planning, and intervention.[6,7] Some studies have highlighted its benefits in PC screening practices by reducing the number of low-grade PC patients referred to biopsies.[8,9] In most cases, positive prostate MRI findings are defined on the basis of the Prostate Imaging-Reporting and Data System (PI-RADS) v2.1 scoring system.[10] PI-RADS ≥ 3 is commonly considered the cut-off to refer a patient to a biopsy, based on a probable presence of a clinically significant PC (csPC, Gleason score ≥7).[10] Nevertheless, PI-RADS v.2.1 has been reported to suffer from low levels of interobserver and intraobserver agreements due to the subjective definition of the different PC lesion categories.[10,11]

To overcome PI-RADS v.2.1 limitations in estimating csPC risk occurrence, different models have been proposed combining the MRI score with known PC risk factors such as prostate volume, PSA, PSA density (PSAd), and patient chronological age.[10–13] The latest is commonly included under the assumption that it might be representative of the aging process of the patient.[14,15] However, it has been argued that chronological age alone is not sufficient to capture the process, and that more reliable biomarkers of aging might be needed.[16,17] In that regard, medical imaging can display phenotypic changes at the organ level that might be useful to measure age-related changes associated with transitions into pathological states.[17,18] Although some research has explored the use of images such as chest radiographies to estimate the age of the patient with the help of deep learning (DL),[19,20] the relationship between aging and prostate MRI has yet to be explored.

Against this background, we aimed to obtain a surrogate biomarker of aging from prostate MRI of "healthy" participants (low-grade and negative biopsy-confirmed) using DL. Moreover, we aimed to investigate the relationship between the proposed aging biomarker and the csPC risk occurrence in an unseen cohort of patients with biopsy-confirmed PC diagnosis (csPC, negative, and low-grade). We hypothesized that our proposed prostate MRI surrogate biomarker of aging might be associated with an increased risk of csPC occurrence independently of other PC risk factors.

Materials and Methods

Institutional review boards (IRBs) of all the centers contributing to the data used for the purpose of the study waived the need for informed patient consent, conditioned to a retrospective and scientific use.[21]

Study Design

The prostate imaging: cancer artificial intelligence (PI-CAI) is a collection of retrospectively collected prostate MRI exams to validate modern AI algorithms and to estimate radiologists' csPC detection and diagnosis performance.[21] Patient exams were collected on the basis of suspected csPC due to elevated PSA levels and, in some cases, abnormal digital rectal examination (DRE) findings. Patients included in the study did not have a history of previous treatment or prior biopsy-confirmed csPC findings.

The main study of PI-CAI consists of 1500 T2-weighted (T2w) and diffusion-weighted (DW) MRI exams collected from four different centers and two MRI vendors. Out of the publicly available sequences, 425 cases have biopsy-confirmed csPC, from which 220 were annotated at the pixel level by one trained investigator or one radiology resident under the supervision of one expert radiologist, for the purpose of the publicly available dataset. Annotations were obtained with the open-source ITK-SNAP v3.80 software (http://www.itksnap.org/).[21] Clinical and demographic variables including PSA levels (ng/mL), prostate volume (mL), PSAd (ng/mL^2), Gleason scores, PI-RADS scores, and patient

chronological age (years) could be obtained from the available clinical information and patient enrollment process or were provided together with the sequences when a value was reported during clinical routine.[21]

Data and Preprocessing

We used axial T2w spin sequences acquired from 2012 to 2021,[21] together with their paired prostate whole gland masks. Images were acquired with either Siemens (Siemens Health Engineers, Erlangen, Germany) or Philips (Philips & Co, Eindhoven, the Netherlands) scanners with time to echo (TE) ≤90 msec, repetition time (TR) ≥3000 msec, and in-plane resolution of 0.5 mm × 0.5 mm and 3 mm slice thickness and with a surface coil.[10,21] The T2w MRI exams were included in the study on the basis of biopsy-confirmed (systematic, MRI-guided or a combination of both) csPC, low-grade, and negative cases. For the remainder of the article and unless otherwise stated, we consider non-clinically significant PC (non-csPC) to include both biopsy-confirmed low-risk PC and negative biopsy-confirmed MRI exams.

Patients with missing clinical, demographic data, human annotations, or faulty annotations at the time the study was performed were discarded. Further, only baseline visits were considered, where baseline is defined as the first visit of the patient when multiple visits for the same patient were available. Figure 1 shows the inclusion criteria of the study. After following the inclusion criteria,
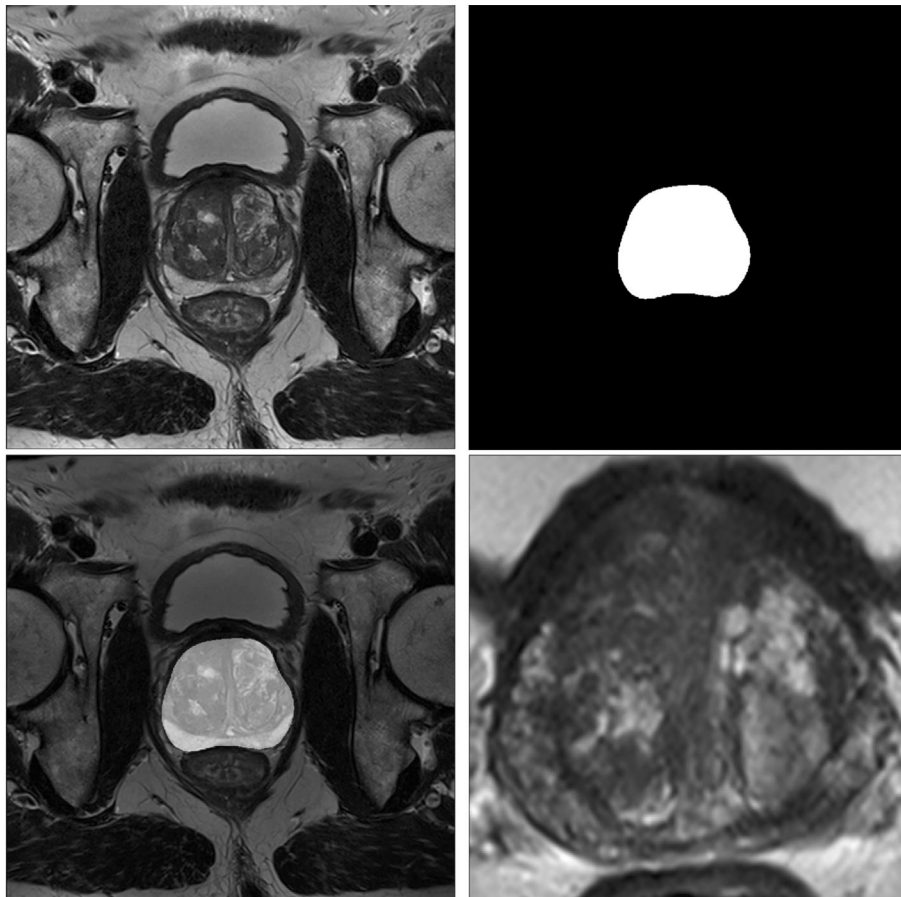

FIGURE 2: Pre-processing including normalization and prostate gland cropping.
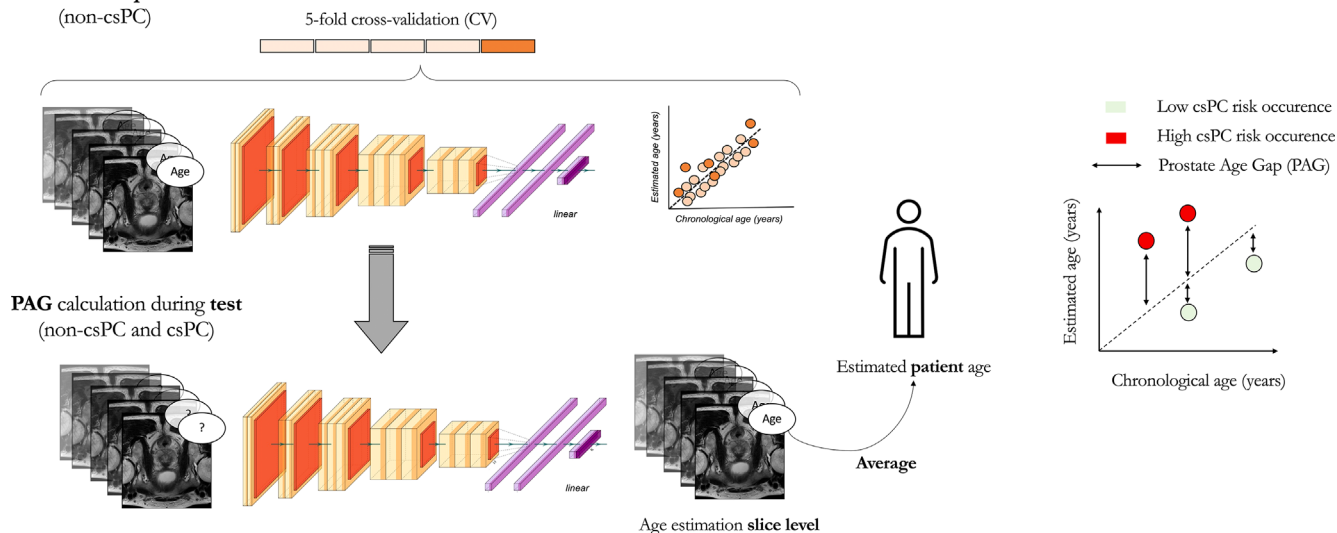
FIGURE 3: Technical approach to the project.

468 participants (221 negative, 133 low-grade, and 114 csPC) were included in the study.

The multi-center and multi-vendor nature of the data used in the study led to noticeable visual differences in the sequences. To palliate them, we normalized the pixel intensity range and applied a cropping around the prostate gland using the available prostate masks. To avoid edge cases where parts of the prostate gland could be omitted due to restrictive mask ranges, we extended the cropping area 40 pixels in the horizontal and vertical directions.[22] Figure 2 depicts an example of an original sequence and the result following the pre-processing steps.

Deep Learning for Chronological Age Prediction

With the aim of capturing patterns that are manifested as part of a normal aging process,[17,18] we trained a DL model to predict the chronological age of non-csPC patients. For that purpose, we trained the DL model with 354 non-csPC (221 negative and 133 low-grade) patient T2w axial MRI exams out of the original 468 (354 ncsPC and 114 csPC) exams included in the study. Figure 3 depicts the development and testing process of the DL model, including data splitting.

To train the DL model, we started by splitting the 354 non-csPC (221 negative and 133 low-grade) patient T2w MRI exams stratifying by PSA levels and non-csPC sub-group (i.e., negative or low-grade). The splitting was performed at the patient level to avoid cross-contamination and in a 60%/40%, to obtain the non-csPC training dataset and non-csPC test dataset. The process resulted in 212 non-csPC (131 negative and 81 low-grade) patient T2w MRI exams included in the training set and 142 non-csPC in the test set (90 negatives and 52 low-grade), respectively. Finally, we combined the 142 non-csPC test patient exams with the 114 csPC patient exams that were reserved for testing purposes to obtain the final test set (Fig. 3). Table 1 depicts the baseline characteristics of the non-csPC training set whilst Table 2 depicts the characteristics of the testing set including non-csPC and csPC patients.

TABLE 1. Baseline Characteristics of Non-csPC (Negative and Low-Grade) Biopsy-Confirmed Participants Used to Train the DL Model

Characteristic	Non-csPC (N = 212)
Age (years)	64.78 ± 6.67
PAG (years)	-2.98 ± 0.03
PSA (ng/mL)	11.90 ± 17.70
PSA > 3 ng/mL	-
Yes	204 (96.22)
No	8 (3.78)
Prostate volume (mL)	60.29 ± 30.05
PSAd (ng/mL^2)	0.228 ± 0.442
PI-RADS ≥ 3	-
Yes	146 (68.37)
No	66 (31.13)
Biopsy type	-
Systematic	66 (31.13)
MRI guided	108 (50.94)
MRI (+systematic)	38 (17.93)

Non-csPC = Gleason score ≤6 and negative cases; PAG = prostate age gap; PSA = prostate-specific antigen; PSAd = prostate-specific antigen density; PI-RADS = Prostate Imaging-Reporting and Data System.

To maximize the amount of data available for training, we trained the model at the slice level. That is, we used two-dimensional (2D) slices from each patient's T2w exams as inputs of

TABLE 2. Baseline Characteristics of Non-csPC (Low-Grade and Negative) and csPC Biopsy-Confirmed Participants Used to Evaluate PAG

| Characteristic | Group | | P Value |
	Non-csPC (N = 142)	csPC (N = 114)	
Age (years)	66.86 ± 7.39	67.08 ± 6.42	0.802
PAG (years)	-4.11 ± 0.32	4.53 ± 2.22	<0.001*
PSA (ng/mL)	11.91 ± 12.79	13.96 ± 10.11	0.163
PSA > 3 ng/mL	-		0.998
Yes	137 (96.47)	111 (97.36)	
No	5 (3.53)	3 (2.64)	
Prostate volume (mL)	65.78 ± 38.26	54.39 ± 26.60	0.007*
PSAd (ng/mL^2)	0.213 ± 0.244	0.303 ± 0.239	0.003*
PI-RADS ≥ 3	-		0.003*
Yes	96 (67.60)	96 (84.21)	
No	46 (32.39)	18 (15.78)	
Biopsy type	-		0.740
Systematic	46 (32.39)	18 (15.78)	
MRI guided	74 (52.12)	94 (82.46)	
MRI (+systematic)	22 (15.49)	2 (1.76)	

Non-csPC = low-grade (Gleason score ≤ 6) and negative cases; csPC = clinically significant (Gleason score ≥ 7); PAG = prostate age gap; PSA = prostate-specific antigen; PSAd = prostate-specific antigen density; PI-RADS = Prostate Imaging-Reporting and Data System.
*P values <0.05 were considered statistically significant.

the DL model. The approach resulted in 3223 non-csPC 2D inputs available during training. We chose a ResNet34[23] as the DL model based on its simplicity and performance in similar tasks.[24] Rotation and flipping data augmentations were used in an online fashion with probability $P = 0.5$ at training time.[22,25]

During testing, we defined prostate age gap (PAG) as the residual obtained from subtracting the chronological age from the estimated DL of the patient. We considered PAG to be a MRI surrogate marker of normal aging of the patients. We work under the assumption that the model is able to successfully extract patterns from the non-csPC MRI exams, and that they are representative of a normal aging process. Hence, during testing, a big residual between the estimated model age and the chronological patient age (PAG) can be indicative of the presence of processes captured by the patient's MRI examination that are not associated with a normal aging process in the PC context.

The model was trained with a 5-fold cross validation (CV) for 120 epochs, where we aimed to minimize the mean absolute error (MAE) between the DL model estimated age and the chronological patient age. In every fold, the validation loss was closely monitored and the one reaching the minimum in the 120 epochs was saved and used to report the MAE at the slice level for the fold under

consideration. Model training and validation was carried out on an NVIDIA A100-80G GPU (NVIDIA Corporation, Santa Clara, California, USA).

Evaluation

We evaluated PAG using the independent test dataset including 142 non-csPC (90 negative and 52 low-grade) and 114 csPC patient MRI examinations. Due to the MRI slice-level design of the DL model, PAG was obtained for each patient slice. We calculated the PAG at the patient level by averaging the PAG slice values belonging to each patient's MRI exam, resulting in a unique PAG per patient. Figure 3 depicts the calculation process followed for every patient MRI exam.

Statistical Analysis

All analyses were performed in Python 3 (https://www.python.org/downloads/) with the open-sourced statsmodels 0.14.0 module (https://www.statsmodels.org/stable/index.html). We reported continuous variables as mean and standard deviation (mean $\pm$ SD), whilst categorical variables were reported as number of occurrences and percentage (N [%]). We performed unpaired t-test, Mann–Whitney U tests, and permutation tests where appropriate to assess the statistical significance of the differences between csPC and non-
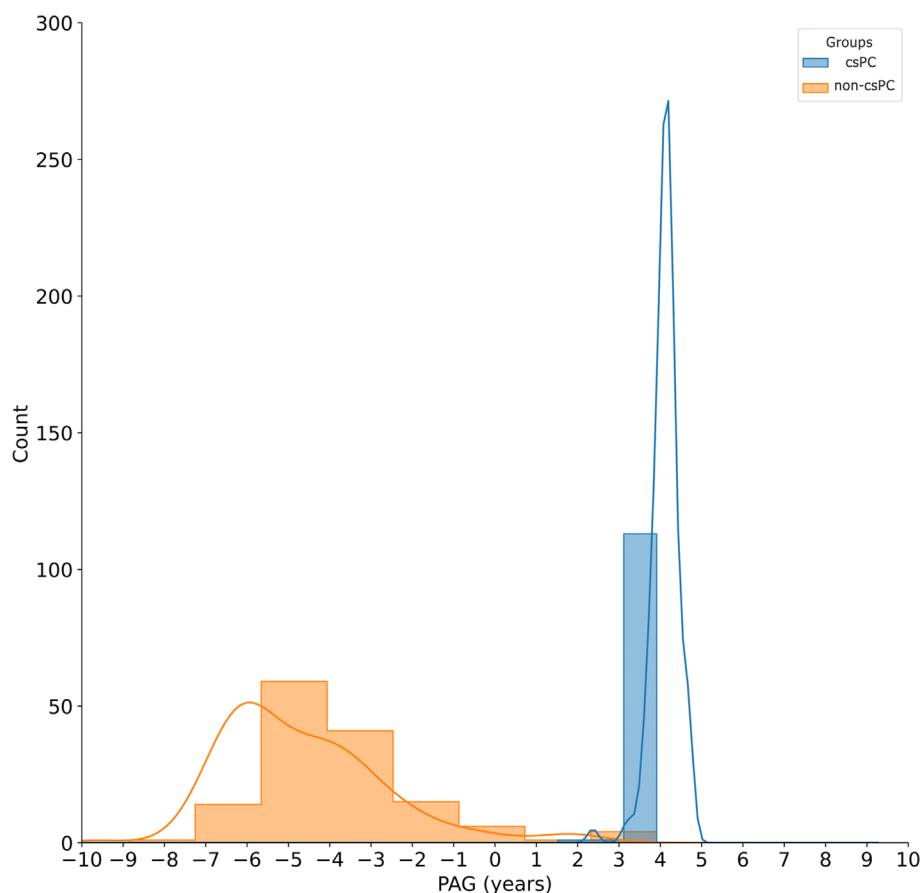
FIGURE 4: Prostate age gap (PAG) distributions for the non-csPC (negative and low-grade) and csPC patients in the test set.

TABLE 3. Odds Ratio in Univariate and Multivariate Analysis of PAG and csPC

Model	Adjustment	PAG (1-Year Increase)		
		OR	95% CI	P Value
I (unadjusted)	-	3.07	[2.15, 4.38]	<0.001*
II (adjusted)	Age, volume, and PSA	3.77	[2.33, 6.11]	<0.001*
III (adjusted)	Age, volume, PSA, and PI-RADS	3.78	[2.32, 6.16]	<0.001*
IV (adjusted)	Age and PSAd	3.61	[2.31, 5.64]	<0.001*
V (adjusted)	Age, PSAd, and PI-RADS	13.18	[2.88, 60.28]	<0.001*
VI (adjusted)	Age and PI-RADS	3.15	[2.17, 4.58]	<0.001*

PAG = prostate age gap; PSA = prostate-specific antigen; PSAd = prostate-specific antigen density; PI-RADS = Prostate Imaging-Reporting and Data System.
*P values <0.05 were considered statistically significant.

csPC groups. Multivariate logistic regressions were used to estimate the odds ratio (OR) and association between PAG and risk of csPC, and in a discriminative modeling approach to assess the performance differentiating between csPC and non-csPC groups of PAG.

We started by presenting a descriptive analysis in the test dataset of the demographic, clinical, and PAG for csPC and non-csPC groups. Following, we introduced PAG as a continuous variable (years) and considered different scenarios with a univariate analysis, partially adjusted, and fully adjusted models. In particular, we presented Model I unadjusted and Model II adjusted by age (years), volume (mL), and PSA (ng/mL). Model III was adjusted by age (years), volume (mL), PSA (ng/mL), and PI-RADS ≥ 3. Both Model
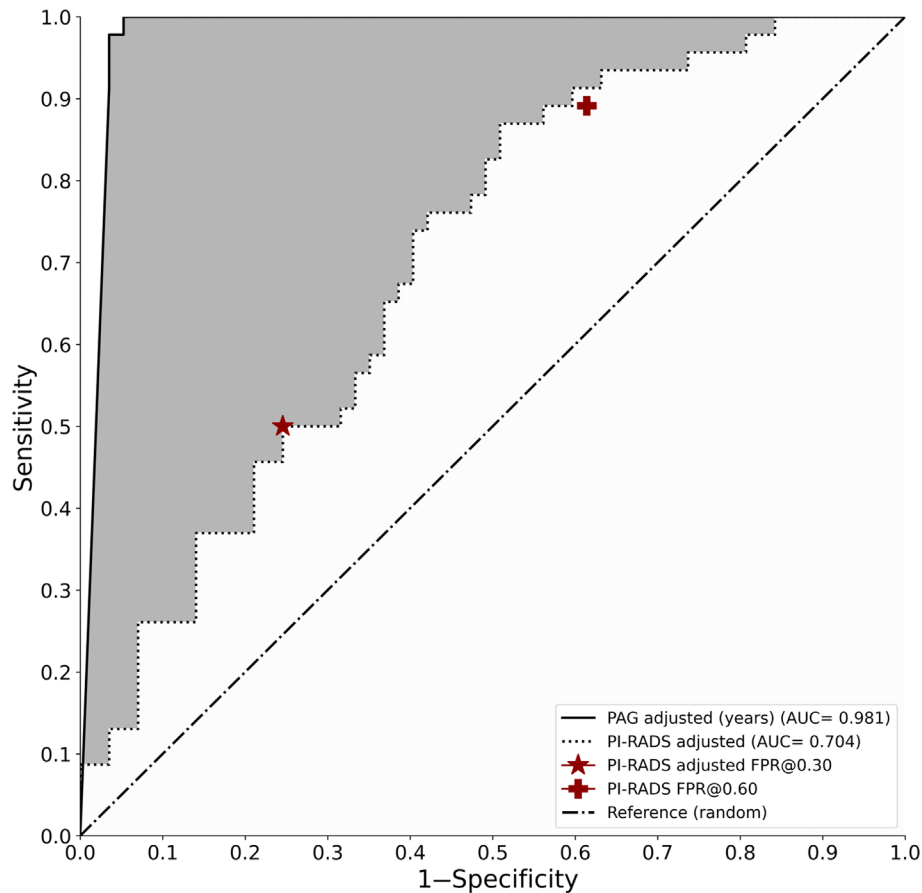
FIGURE 5: Multivariate logistic regression and discriminative ability of prostate age gap (PAG) and PI-RADS ≥ 3 when adjusted by PSA, prostate volume, and chronological age.

II and Model III were intended to provide a hypothetical standard PC screening scenario accounting for PSA and PI-RADS ≥ 3 along with the chronological age of the patient. Model IV and Model V considered a case scenario where PSA (ng/mL) and volume (mL) in Model II and Model III were replaced with PSAd (ng/mL^2). Finally, Model VI considered a PC-screening scenario based on age and PI-RADS ≥ 3 as adjusting factors.

We used a multivariate logistic regression to estimate the discriminative value of PAG adjusted by PSA (ng/mL), prostate volume (mL), and chronological patient age (years). Following, we compared its discriminative ability with PI-RADS ≥ 3 adjusted by PSA (ng/mL), prostate volume (mL), and chronological patient age (years), simulating common screening PC pathways. Predictive performance was measured in terms of area under the curve (AUC) for the risk of occurrence of csPC. We obtained an estimate of AUC through bootstrapping of the test set with N = 1000 replicates without repetition. Confusion matrices are presented together with AUC for specific false positive rates (FPRs) chosen to emulate commonly reported PC FPRs.[21] A *P* value <0.05 was considered statistically significant.

Results

A total of 212 non-csPC participants at baseline (64.78 ± 6.67 years) were considered for the DL model

training. Table 1 depicts the baseline characteristics of the participants. During testing, we considered 114 participants with biopsy-confirmed csPC (67.08 ± 6.42 years) and 142 non-csPC (66.86 ± 7.39 years) participants who were not included in the DL model training. As depicted by the baseline characteristics in Table 2, both groups of participants were equally distributed in terms of age. However, csPC patients were more likely to have a higher PSA level, less likely to have a larger prostate, and more likely to have a larger PSAd level. As it can also be seen in Table 2, PAG was more likely to be substantially larger and positive for csPC patients. Figure 4 depicts the PAG distributions for csPC and non-csPC patients in the final test dataset.

Table 3 shows the different univariate and multivariate logistic regression models adjusted for chronological age (years), PSA (ng/mL), prostate volume (mL), PSAd (ng/mL^2), and PI-RADS score. As observed in the univariate model (model I), there was a significant difference in the odds of csPC (OR = 3.07, 95% confidence interval [CI]: [2.15, 4.38]).

When partially adjusting Model I by PSA and volume of the prostate (Model II), there was a significant difference in the adjusted OR per 1-year increase of PAG of csPC
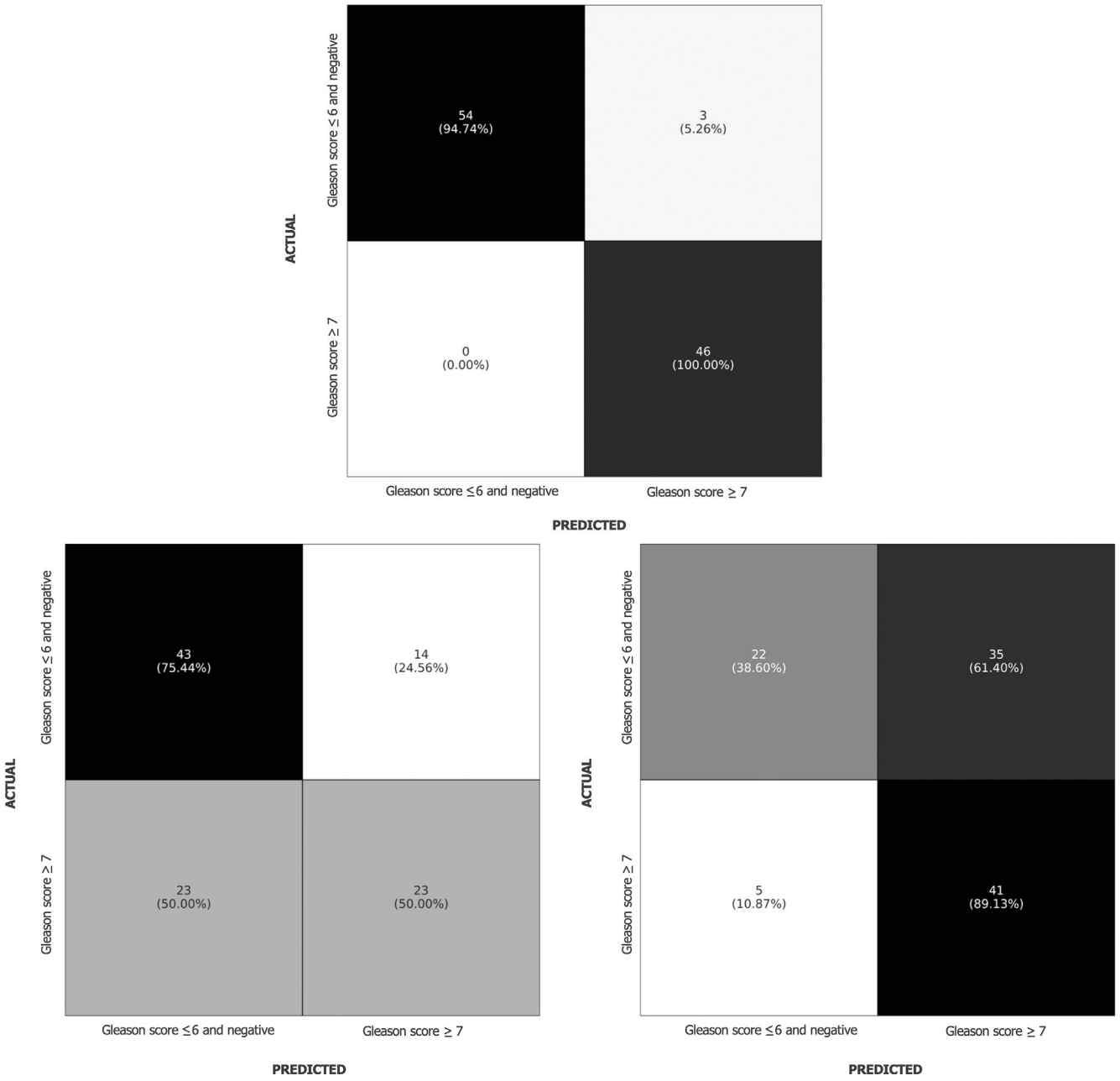
FIGURE 6: Confusion matrices for multivariate logistic regression and prostate age gap (PAG) when adjusted (FPR @ 0.10), PI-RADS adjusted (FPR @ 0.30), and PI-RADS unadjusted (FPR @ 0.60).

(OR = 3.77, 95% CI: [2.33, 6.11]). Therefore, we considered a fully adjusted model (Model III) accounting for PI-RADS ≥ 3 and found that the significant difference in the adjusted OR of csPC was kept (OR = 3.78, 95% CI: [2.32, 6.16]).

In Model IV, we found that the adjusted OR still remained associated with csPC risk (OR = 3.61, 95% CI: [2.31, 5.64]) when adjusting by PSAd (ng/mL^2) and chronological age (years). Finally, in Model V, we found an association between the adjusted OR per 1-year PAG increase and csPC risk (OR = 13.18, 95% CI: [2.88, 60.28]) when adjusting by PI-RADS ≥ 3 and chronological patient age (years).

We compared the discriminative ability of PAG and PI-RADS ≥ 3 adjusted. Figure 5 depicts the bootstrapped AUC of both models, with a significant improvement for the PAG-adjusted model when compared to the PI-RADS model (AUC = 0.981, 95% CI: [0.975, 0.987]; AUC = 0.704, 95% CI: [0.652, 0.756]).

As depicted in Fig. 6, we observed a trade-off of a low FPR at the expense of missing 50% of csPC cases in the case of adjusted PI-RADS ≥ 3 and when operating under a FPR of 0.30. In the case of adjusted PAG, we observed that the model was able to detect all the csPC cases when operating at a 0.05 FPR. Figure 7 shows the consistently positive values
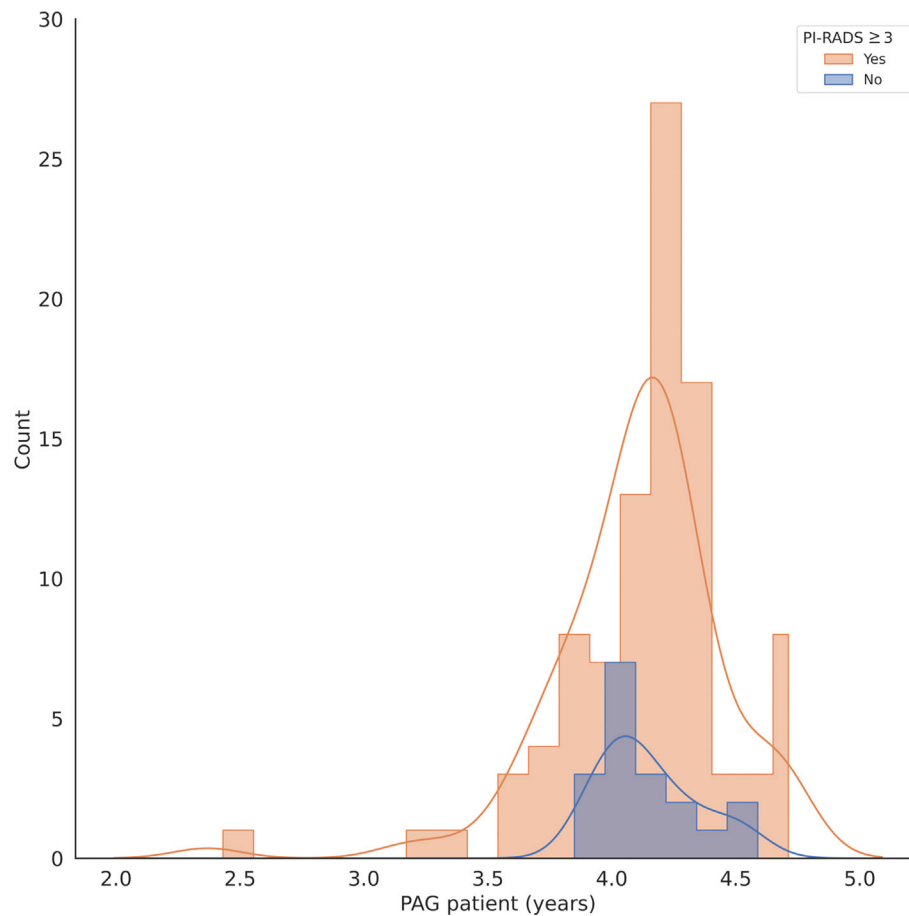
FIGURE 7: Prostate age gap (PAG) distribution for PI-RADS ≤ 2 and PI-RADS ≥ 3 biopsy-confirmed csPC patients.

in the distribution of PAG for the csPC patients with PI-RADS ≤ 2 and PI-RADS ≥ 3.

Discussion

In this retrospective study, we proposed PAG as a surrogate biomarker of patient aging from prostate MRI. Our retrospective study provides some degree of evidence of the independent association of 1-year increase in PAG and the odds of csPC when adjusted by PSA (ng/mL), prostate volume (mL), and chronological patient age (years). Further, we evaluated the discriminative ability of PAG to distinguish between csPC and non-csPC (negative and low-grade) biopsy-confirmed exams in a multivariate logistic regression when adjusting by the same risk factors. Our results support the hypothesis that PAG is independently associated with csPC risk even in the presence of PSA and PI-RADS ≥ 3, two of the most used markers for PC unorganized population-screening practices.[26–28]

Previous studies in the literature have tackled the csPC detection from MRI leveraging DL techniques, yielding AUCs in the range of 0.82–0.89.[22,29–31] Despite not being directly comparable, our results suggest that the efficacy of PAG to detect csPC occurrence might fall within the interval or even surpass that of other DL studies.[22,29,30] In contrast to those studies, our approach might benefit from being less

opaque.[17,32] In particular, the concept of aging might be more relatable and interpretable whilst DL approaches might suffer from an inherent lack of interpretability and transparency.[22,29,30,32] In that perspective, the hypothetical clinical deployment and routinely use of a non-interpretable and non-transparent approach can pose more ethical and practical challenges than a more transparent and interpretable alternative.[32,33] Further, as reflected by our internal test results and based on the multi-center and multi-vendor data used to develop the study, PAG might be a robust marker in the presence of heterogenous data sources.[21]

With the increase in interest in MRI as a central component of the PC diagnostic pathway,[9,27,28] PAG might offer a reader-independent MRI risk occurrence marker of csPC.[10,11] In that regard, we show that PAG may help to detect csPC presence even in the event of PI-RADS ≤ 2.[11] As depicted by our multivariate logistic regression analysis, PAG might be complementary to commonly used risk factors of csPC such as prostate volume (mL), PI-RADS ≥ 3, PSA (ng/mL), and PSAd (ng/mL^2).[10,12,29–31] Our results indicate that PAG might be suggestive of phenotypical changes in the prostate MRI that are not associated with a normal patient aging process and that other PC risk factors might not be able to capture.[16,17]

Limitations

The study had several limitations. First, our study was limited by the small sample size and its retrospective nature. Future studies including external evaluations of the DL algorithm are required to verify our findings and assess the clinical value. Additionally, further validation using prospective data would be desirable to confirm causality. Second, the design of the best DL algorithm for chronological age estimation was considered to be out of the scope of this study. Hence, future studies will tone to improve and validate the algorithm. Finally, we did not compare our DL model with other biological age markers (eg, epigenetic clocks) or with DL models trained on the same data to directly predict Gleason score from MRI images. Future comparisons might provide additional insights about our DL model and approach.

Conclusion

In conclusion, we developed a DL model that uses prostate T2w MRI exams to estimate the age of the patients. We proposed PAG, defined as the difference between the DL model estimated age and the chronological patient age as a surrogate biomarker of aging. We found that the PAG of patients with known csPC was higher than those without it, reflecting that the DL patient estimated age was higher than the chronological patient age. Our DL model could potentially serve as an indicator of the presence of atypical phenotypical changes in the prostate that are not associated with a normal aging process and could help improve biopsy referral in organized PC screening initiatives.

Acknowledgments

References

1. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366(11):981-990. https://doi.org/10.1056/NEJMoa1113135.

2. Merriel SW, Pocock L, Gilbert E, et al. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. BMC Med 2022;20(1):54. https://doi.org/10.1186/s12916-021-02230-y.

3. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50–69 years (STHLM3): A prospective population-based diagnostic study. Lancet Oncol 2015 Dec 1;16(16):1667-1676. https://doi.org/10.1016/S1470-2045(15)00361-7.

4. Barry MJ. Screening for prostate cancer—The controversy that refuses to die. N Engl J Med 2009;360(13):1351. https://doi.org/10.1056/NEJMe0901166.

5. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and over-treatment of prostate cancer. Eur Urol 2014;65(6):1046-1055. https://doi.org/10.1016/j.eururo.2013.12.062.

6. Drost FJ, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database Syst Rev 2019;4:1465-1858.

7. Ahmed HU, Bosaily AE, 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.

8. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018;378(19):1767-1777.

9. Eklund M, Jäderling F, Discacciati A, et al. MRI-targeted or standard biopsy in prostate cancer screening. N Engl J Med 2021;385(10):908-920. https://doi.org/10.1056/NEJMoa2100852.

10. Turkbey B, Purysko AS. PI-RADS: Where next? Radiology 2023;307(5):223128. https://doi.org/10.1148/radiol.223128.

11. 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. https://doi.org/10.1148/radiol.2016152542.

12. Sedelaar JM, Schalken JA. The need for a personalized approach for prostate cancer management. BMC Med 2015;13:1-3. https://doi.org/10.1186/s12916-015-0344-1.

13. Mehralivand S, Shih JH, Rais-Bahrami S, et al. A magnetic resonance imaging–based prediction model for prostate biopsy risk stratification. JAMA Oncol 2018;4(5):678-685. https://doi.org/10.1001/jamaoncol.2017.5667.

14. Rawla P. Epidemiology of prostate cancer. World J Oncol 2019;10(2):63-89. https://doi.org/10.14740/wjon1191.

15. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC Med 2014;12:1-7. https://doi.org/10.1186/1741-7015-12-26.

16. Chen R, Wang Y, Zhang S, et al. Biomarkers of ageing: Current state-of-art, challenges, and opportunities. MedComm–future. Medicine 2023;2(2):e50.

17. Babyn PS, Adams SJ. AI analysis of chest radiographs as a biomarker of biological age. Lancet Healthy Longev 2023;4(9):e446-e447.

18. Aging Biomarker Consortium, Bao H, Cao J, et al. Biomarkers of aging. Sci China Life Sci 2023;66(5):1-74.

19. Azarfar G, Ko SB, Adams SJ, Babyn PS. Deep learning-based age estimation from chest CT scans. Int J Comput Assist Radiol Surg 2023;1-9.

20. Mitsuyama Y, Matsumoto T, Tatekawa H, et al. Chest radiography as a biomarker of ageing: Artificial intelligence-based, multi-institutional model development and validation in Japan. Lancet Healthy Longev 2023;4(9):e478-e486.

21. Saha A, Bosma J, Twilt J, et al. Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI—The PI-CAI Challenge. In: *Medical Imaging with Deep Learning*, short paper track 2023 Apr 28. https://doi.org/10.5281/zenodo.6667655

22. Rolfsnes ES, Thangngat P, Eftestøl T, et al. Reconsidering evaluation practices in modular systems: On the propagation of errors in MRI prostate cancer detection. arXiv preprint arXiv:2309.08381. 2023 Sep 15.

23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016. p 770–778.

24. Jónsson BA, Bjornsdottir G, Thorgeirsson TE, et al. Brain age prediction using deep learning uncovers associated sequence variants. Nat Commun 2019;10(1):5409.

25. Lindeijer TN, Ytredal TM, Eftestøl T, et al. Leveraging multi-view data without annotations for prostate MRI segmentation: A contrastive approach. arXiv preprint arXiv:2308.06477. 2023 Aug 12.

26. Nordström T, Discacciati A, Bergman M, et al. Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): A prospective, population-based, randomised, open-label, non-inferiority trial. Lancet Oncol 2021;22(9):1240-1249.

27. Washino S, Okochi T, Saito K, et al. Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naïve patients. BJU Int 2017;119(2):225-233.

28. Wallström J, Geterud K, Kohestani K, et al. Prostate cancer screening with magnetic resonance imaging: Results from the second round of the Göteborg prostate cancer screening 2 trial. Eur Urol Oncol 2022; 5(1):54-60.

29. Saha A, Hosseinzadeh M, Huisman H. End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction. Med Image Anal 2021;73:102155. https://doi.org/10.1016/j.media.2021.102155.

30. Fernandez-Quilez A, Eftestøl T, Kjosavik SR, Goodwin M, Oppedal K. Contrasting axial T2w MRI for prostate cancer triage: A self-supervised learning approach. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*; 2022 Mar 28. p 1–5. IEEE.

31. Schelb P, Kohl S, Radtke JP, et al. Classification of cancer at prostate MRI: Deep learning versus clinical PI-RADS assessment. Radiology 2019;293(3):607-617. https://doi.org/10.1148/radiol.2019190938.

32. Fernandez-Quilez A. Deep learning in radiology: Ethics of data and on the value of algorithm transparency, interpretability and explainability. AI Ethics 2023;3(1):257-265. https://doi.org/10.1007/s43681-022-00161-9.

33. Kurbatskaya A, Jaramillo-Jimenez A, Ochoa-Gomez JF, Brønnick K, Fernandez-Quilez A. Assessing gender fairness in EEG-based machine learning detection of Parkinson's disease: A multi-center study. arXiv preprint arXiv:2303.06376. 2023 Mar 11.