**Deep-learning models for detection and localization of clinically significant prostate cancer on multi-parametric MRI**

**Abstract**

**Background:** Deep learning for diagnosing clinically significant prostate cancer (csPCa) is feasible but needs further evaluation in patients with prostate-specific antigen (PSA) levels of 4-10 ng/mL.

Purpose: To explore diffusion-weighted imaging (DWI), alone and in combination with T2-weighted imaging (T2WI), for deep-learning-based models to detect and localize csPCa.

**Study Type:** Retrospective.

**Population:** 1628 patients with systematic and cognitive-targeted biopsy-confirmation (1007 csPCa, 621 non-csPCa) were divided into model development (n = 1428) and hold-out test (n = 200) datasets.

**Field Strength/Sequence:** DWI with a diffusion-weighted single-shot gradient echo planar imaging sequence and T2WI with a T2-weighted fast spin echo sequence at 3.0-T and 1.5-T.

**Assessment:** The ground truth of csPCa was annotated by two radiologists in consensus. A diffusion model, DWI and apparent diffusion coefficient (ADC) as input, and a biparametric model (DWI, ADC, and T2WI as input) were trained based on U-Net. Three radiologists provided the PI-RADS (version 2.1) assessment. The performances were determined at the lesion, location, and the patient level.

**Statistical Tests:** The performance was evaluated using the areas under the ROC curves (AUCs), sensitivity, specificity, and accuracy. A P value < 0.05 was considered statistically significant.

Results: The lesion-level sensitivities of the diffusion model, the biparametric model, and the PI-RADS assessment were 89.0%, 85.3%, and 90.8% (*P*=0.289-0.754). At the patient level, the diffusion model had significantly higher sensitivity than the biparametric model (96.0% vs. 90.0%), while there was no significant difference in specificity (77.0%. vs. 85.0%, *P*=0.096). For location analysis, there were no significant differences in AUCs between the models (sextant-level, 0.895 vs. 0.893, *P*=0.777; zone-level, 0.931 vs. 0.917, *P*=0.282), and both models had significantly higher AUCs than the PI-RADS assessment (sextant-level, 0.734; zone-level, 0.863).

Data Conclusion: The diffusion model achieved the best performance in detecting and localizing csPCa in patients with PSA levels of 4-10 ng/mL.

**Key words:**

Deep learning; Prostate cancer; Segmentation

**Introduction**

Technical refinements, standardization, and popularization of multi-parametric MRI (mpMRI) have contributed to a paradigm shift in the detection, localization, and risk stratification of clinically significant prostate cancer (csPCa), as well as guidance for biopsy (1,2). The European Association of Urology guidelines (3) have confirmed the central role of mpMRI in the diagnosis and MRI-guided biopsy of prostate cancer. Among mpMRI sequences, diffusion-weighted imaging (DWI) and the associated apparent diffusion coefficient (ADC) map have been shown to be helpful components for csPCa lesion detection, while T2-weighted imaging (T2WI) has been used for anatomic zone and csPCa lesion detection in the transition zone (TZ) (2). When these are of insufficient diagnostic quality, due to motion, noise, or rectal gas, dynamic contrast enhancement (DCE) imaging may be useful, as stated in the Prostate Imaging Reporting and Data System (PI-RADS) (2).

A moderate level of agreement between observers interpreting prostate mpMRI is observed, influenced by the reader's experience (4-7). In addition, some benign prostate hyperplasia or prostatitis may sometimes show nontypical and heterogeneous signals and mimic or obscure cancers, especially in patients with prostate-specific antigen (PSA) serum levels of 4∼10 ng/mL, which is a diagnostic gray zone (8-10). One problem with early detection of csPCa is the low specificity of PSA between 4 and 10 ng/ml, with only a 25% incidence of prostate cancer (8-10). Patients with serum PSA levels of 4∼10 ng/mL are also very common in daily practice, and there is controversy in determining whether a biopsy should be performed. As prostate cancer incidence continues to increase(11), more accurate and automated methods to detect and delineate within-gland target lesions to direct MRI-guided biopsies are needed.

MpMRI combined with deep learning has the potential to serve as a triage test to determine which patient should undergo prostate biopsy. Recently, many studies have reported the feasibility of diffusion or biparametric MRI models using deep learning techniques to automatically diagnose csPCa (12-21). Deep learning has excellent detection ability for advanced cancer and obvious cancer, but only a few studies (14,18,19) have focused on cancer detection in patients with PSA in the range of 4-10 ng/mL. ADC only, T2WI only, or their combination (biparametric) have been studied, and the diagnostic performance shows high variability among centers(12). Investigations of the generalization ability of models for images from different MRI protocols or scanners are limited(14,22), and few studies have focused on optimal sequence exploration(23).

Thus, the aim of this study was to develop and compare two deep-learning-based models trained using diffusion images, one with and one without T2WI, to explore which is best for fully automated detection and localization of csPCa in patients with PSA in the range of 4-10 ng/mL.

**Materials and Methods**

This retrospective study was approved by the institutional review board of our institution [IRB number: 2021(342)], with a waiver of informed consent.

**Patient enrollment**

Figure 1 shows the patients’ enrollment and distribution. The inclusion criteria were mpMRI examination performed on 1 of 7 scanners at our institution from June 2017 to October 2021; consecutive patients who underwent mpMRI before biopsy, with a clinical suspicion of prostate cancer due to elevated serum PSA level, abnormal digital rectal examination and/or abnormal transrectal ultrasound (TRUS) results regardless of serum PSA levels. Exclusion criteria were (1) patients without biopsy performed within 1 month after mpMRI examination (progression of serum PSA level, abnormal digital rectal examination, or MR findings); (2) biopsy negative for prostate cancer without clinical follow-up ≥1 year or showing potential prostate cancer during the clinical follow-up; (3) images with severe artifacts; (4) mismatch between the mpMRI image and the pathology result, such as a discrepancy in tumor location, MRI-invisible csPCa, and images that show csPCa but are pathologically negative.

A total of 1628 patients were retrospectively included in this study. All of the patients underwent combined biopsy and clinical follow-up, and 21.3% (346/1628) of patients underwent further radical prostatectomy. The detection rate of the combined biopsy was 61.9% (1007/1628). Of the 1628 patients, 748 patients had PSA levels of 4∼10 ng/mL, and 880 patients had PSA levels of <4 or >10 ng/ml. We preselected 200 patients (100 PCa, 100 non-PCa) from the patients with PSA serum levels of 4∼10 ng/mL to build a hold-out test dataset to evaluate and compare the models developed and PI-RADS assessment. The remaining 1428 patients were used to develop models and were randomly allocated into a training dataset (total 1142, 719 PCa, 423 non-PCa), validation dataset (total 143, 88 PCa, 55 non-PCa), and test dataset (total 143, 100 PCa, 43 non-PCa) at a ratio of 8:1:1. All statistics are based on post-processed data in the hold-out test dataset. The demographic and clinical characteristics of the included patients are shown in Table 1.

**mpMRI data**

All examinations were obtained using 1 of 5 3.0 T MR scanners (i.e., Scanner A: Magnetom Prisma, Siemens Healthcare, Erlangen, Germany, Scanner B: Achieva TX, Philips Healthcare, Best, the Netherlands, Scanner C: Discovery HD 750, GE Healthcare, Milwaukee, WI, USA, Scanner D: uMR790, UIH, Shanghai, China, Scanner E: Ingenia, Philips Healthcare, Best, The Netherlands) and 1 of 2 1.5 T MR scanners (i.e., Scanner F: Magnetom Aera, Siemens Healthcare, Erlangen, Germany, and Scanner G: Multiva, Philips Healthcare, Best, The Netherlands). The phased array coil was used as the receiving coil. DWI exploited a diffusion-weighted single-shot gradient echo planar imaging sequence, and T2WI exploited a T2-weighted fast spin echo sequence at 3.0 T and 1.5 T. The scanner type and vendor as well as the mpMRI protocols are summarized in Table 2. The ADC maps were calculated (b values 0, 800 - 2500 s/mm2) using the corresponding algorithm.

**PI-RADS Assessment**

All mpMRI images in the hold-out test dataset were retrospectively interpreted according to PI-RADS version 2.1(2) by two independent urogenital radiologists (G.G. and H.W. with 10 and 12 years of experience). If there was a difference, the third senior radiologist (X.W. with 30 years of experience) was invited to provide a final decision on the PI-RADS score. DCE imaging (if available) was not used in the assessment. The radiologists were blinded to the clinical information of the patients, such as age, biopsy history, PSA, pathology, and any previous MRI reports. The lesions detected were manually delineated on a prostate sector map (24) embedded instructed report (Figure 2).

**Ground truth and image annotation**

In this study, all patients underwent TRUS-guided systematic (12- or 6-core needles) and cognitive-targeted biopsy (combined biopsy). Based on structured reports prepared by 1 of 5 dedicated urogenital radiologists with varying seniority (12, 16, 20, 22, and 28 years, respectively), lesions suspected of malignancy were marked on a prostate sector map (24) for cognitive-targeted biopsy during the clinical routine. At least 1 urologist and 1 urogenital radiologist reviewed MR images before biopsy in a multidisciplinary meeting to ensure accurate localization of suspicious lesions. When performing cognitive-targeted biopsies, the urologists examined each suspicious lesion with an additional needle core (2- to 5-core needles). A urogenital pathologist with 11 years of experience conducted the histopathology analysis on each specimen. The pathology of systematic lesions and targeted lesions were fused into a combination reference (25). Those sextants were given the highest International Society of Urological Pathology (ISUP) grade as determined by either systematic biopsy or cognitive-targeted biopsy, whereas all other sextants were given a systematic ISUP grade. csPCa was defined as ISUP grade 2 or higher based on histopathology findings and scored as Gleason score 3 + 4 or higher (26).

Two urogenital radiologists (Z.S. with 4 years of experience and X.W.) retrospectively reviewed all cases and mapped the pathology results of each focus to the MR images in consensus. If there was whole-mount step section pathology, then that was the reference standard; if not, the combined biopsy was the reference standard. Each lesion was manually delineated by the urogenital radiologist (Z.S.) and then modified by the other urogenital radiologist (X.W.) on DWI using the software ITK-SNAP (27). The modified annotations were used as ground truth.

The prostate segmentation model (28) was used to presegment the prostate gland into sextant areas (Supplementary Material 1) and anatomic zones (Supplementary Material 2) (29). The areas were classified into cancerous and non-cancerous zones according to the presence or absence of cancer in each area. For example, if the ground truth segmentation overlapped with an anatomic zone, then the zone was considered to be a cancerous zone.

**Preprocessing**

The DWI, ADC maps, and T2WI were registered by rigid transformation using coordinate information stored in the DICOM image headers. The prostate gland region was automatically presegmented using the model (28) previously developed in our institution. All the prostate areas were unified and cropped to 64×64×64 (x, y, z), and pixel intensities were normalized to an interval of [0,1]. We augmented the data in the training set by random rotation (rotation angle within 10°), adding random noise, and parallel translation at a range of [(-0.1; 0.1); (-0.1; 0.1)] pixels.

**Deep learning**

The AI system utilizes MR images as inputs and automatically segments the whole prostate gland, the anatomic zones, and the csPCa region in a step-by-step manner. A cascade 3D U-Net (30) segmentation framework was used for prostate cancer segmentation. The network architecture and training setup are detailed in Supplementary Material 3. In the same model development dataset, the diffusion model took the combination of DWI and ADC maps as input, and the biparametric model took the combination of DWI, ADC maps, and T2WI images as input. The training and testing process is shown in Figure 2.

**Postprocessing**

Based on the current capabilities of mpMRI, csPCa greater than or equal to 0.5 cc in volume may be detected (24). Considering the current prevalence of high false positives in artificial intelligence algorithms for prostate cancer(12,31,32), to minimize the influence of very small predicted tumor foci, the outputs were filtered with a threshold value of 0.5 cc. In the hold-out test dataset, we also removed tumor foci less than 0.5 cc from the ground truth of PCa lesions. This postprocessing was not applied to the training and validation sets. The final models were selected based on post-processing performance.

**Statistical analysis**

The Dice similarity coefficient (DSC) was calculated to evaluate segmentation performance (Supplementary Material 4). Statistical analysis was performed using SPSS 24.0 (SPSS Inc., Chicago, IL, USA) and MedCalc 15.8 (MedCalc Software, Ostend, Belgium). Pathology was regarded as the ground truth. At the lesion level, the predicted areas of the models overlapping the manual csPCa segmentation lesions with at least one voxel were considered a true positive; otherwise, the predicted lesions were considered false positive. At the patient level, the prediction of the models was defined as a true positive when at least one index lesion, i.e., ISUP Grade >=2, was detected. For location analysis, when one area overlapped with an index lesion, it was considered a true positive area; otherwise, it was considered a true negative area (33). For the PI-RADS assessment, PI-RADS ≥ 3 was considered positive for csPCa. A receiver operating characteristic (ROC) curve analysis was employed, and the area under the curve (AUC) values of both models and the PI-RADS assessment were calculated. The AUCs were compared using the DeLong test. The sensitivity, specificity, and accuracy were calculated and compared using the chi-square test. The free response receiver characteristic curve (FROC) was plotted using the R (version 4.2.0) package ‘RJafroc’ (<https://dpc10ster.github.io/RJafroc/>). A two-sided p < 0.05 was considered statistically significant.

**Results**

*At the lesion level*

In the hold-out test dataset, 100 patients with 109 csPCa lesions were included as the reference standard, including 14 lesions with PI-RADS scores of 2, 8 lesions with PI-RADS scores of 3, and 91 lesions with PI-RADS scores of 4–5. The mean volume was 1.56 ± 1.35 cm3. There were 43 lesions located in the peripheral zone (PZ), 41 lesions in the transition zone (TZ), and 25 lesions in multiple zones. Table 3 shows the number of lesions before and after post-processing. After the post-processing, the positive calls per patient decreased from 1.185 to 0.135 for the diffusion model and from 1.270 to 0.105 for the biparametric model. The lesion-level sensitivity of the diffusion model, biparametric model, and PI-RADS assessment was 90.0% (97/109), 85.3% (93/109), and 90.8% (99/109), respectively (*P* = 0.289, 0.625, and 0.754, respectively). Table 4 shows the subgroup sensitivity at the lesion level of the two deep-learning-based models and the PI-RADS assessment. In addition, DSC based on csPCa lesions in the hold-out test dataset was 0.69 ± 0.28 and 0.67 ± 0.30 for the diffusion model and biparametric model, respectively (*P = 0.184*) (Figure 3). Figure 4a-d shows examples of csPCa segmentation with good performance for typical csPCa. Figure 4e-h demonstrates a false-negative example of csPCa segmentation. In this case, the csPCa lesion was successfully detected by the diffusion model and missed by the biparametric model. Figure 4i-p demonstrates false-positive examples of csPCa segmentation. In the post-processing, some small true positive lesions with volumes less than 0.5 cc may be filtered out (Figure 5). Note that patients with coexisting lesions greater than 0.5 cc are still predicted to be cancerous without changing the prediction at the patient level. FROC analysis showed that the curves of the two models were very close and partially crossed, suggesting that the detection efficacy of the two models was similar at the lesion level (Figure 6).

*At the patient level*

For the 200 patients (csPCa, 100; non-csPCa, 100) in the hold-out test dataset, the performance and comparisons of the models and the PI-RADS assessment at the patient level are shown in Table 5 and Figure 7a. The diffusion model, biparametric model, and PI-RADS assessment had similar patient-level AUCs of 0.865, 0.875, and 0.920, respectively (*P* = 0.678, 0.058, and 0.094, respectively). The diffusion model had significantly higher sensitivity than the biparametric model (96.0% vs. 90.0%) and had lower specificity without statistical significance (77.0%). vs. 85.0%, *P* = 0.096).

*At the sextant level*

A total of 1200 sextants from the hold-out test dataset were analyzed, including 301 sextants of csPCa and 899 sextants of non-csPCa. The diagnostic efficacy and comparisons of the models and the PI-RADS assessment based on sextants are summarized in Table 5 and Figure 7b. The diffusion model and the biparametric model (AUC, 0.895 vs. 0.893, *P* = 0.777) had significantly higher AUCs than the PI-RADS assessment (AUC, 0.734). There was no significant difference in the sextant level sensitivity between the diffusion model and the biparametric model (87.0% vs. 85.4%, *P* = 0.332), with both being significantly higher than that of PI-RADS assessment (53.5%). There were no significant differences in the sextant level specificity between the diffusion model, biparametric model, and PI-RADS assessment (91.8%, 93.1%, and 93.4%, *P* = 0.118, 0.327, and 1.000, respectively).

*At the anatomic zone level*

A total of 800 zones from the hold-out test dataset were analyzed, including 125 zones of csPCa and 675 zones of non-csPCa. The diagnostic efficacy and comparisons of the models and the PI-RADS assessment based on zones are summarized in Table 6 and Figure 7c. There were no significant differences in diagnostic efficacy between the diffusion model and biparametric model both at the level of total zones and each zone (*P* = 0.282-0.812). For total zone analysis, both the diffusion model and the biparametric model had significantly higher AUCs (0.931 vs. 0.917, *P* = 0.282) than that of the PI-RADS assessment (AUC, 0.863). For the PZ analysis, there were no significant differences in the AUCs of the diffusion model, biparametric model, and PI-RADS assessment (0.926, 0.907, and 0.910, *P* = 0.282, 0.535, and 0.936, respectively). For the TZ analysis, both the diffusion model and biparametric model had significantly higher AUCs (0.898 vs. 0.903, *P* = 0.812) than the PI-RADS assessment (AUC, 0.810).

**Discussion**

This study improved the training segmentation model by incorporating data heterogeneity from multiple vendors, and the performance of detection and location of csPCa in patients with PSA in the range of 4-10 ng/mL approached or exceeded that of PI-RADS (version 2.1) assessment.

The elevated PSA levels in prostate cancer and benign prostatic situations overlap, to a large extent, at a range of 4–10 ng/ml(9,10). The probability of prostate cancer occurrence has been reported to be nearly 20%, and approximately 70% of men with gray-zone PSA levels (4–10 ng/mL) undergo unnecessary biopsies (8,34). MpMRI has the potential to serve as a triage test to identify candidates for biopsy. Prostate mpMRI combined with PI-RADS (v2 or v2.1) at a cutoff value of 3 has been reported to detect csPCa in patients with gray-zone PSA, yielding AUCs of 0.732-0.932 (35-37). The efficacy of our two deep-learning models falls within this interval and is close to similar AI studies (14,15,17,18,38). However, only a few of these studies have focused on performance in men with PSA levels of 4–10 ng/mL(14,18,19). Thus, our models have the potential to accomplish the clinical task of biopsy candidate identification.

Another role of prostate mpMRI is to provide precise localization for the selected candidates for biopsy(29,33). Based on sextants, the diffusion model and biparametric model had higher sensitivity to csPCa localization than the PI-RADS (V 2.1) assessment, and there was no significant difference in specificity. Based on zones, the diffusion model and biparametric model had higher AUCs than the PI-RADS assessment. Radiologists may tend to focus on the main lesion, while the models developed in this study were designed to detect all lesions, potentially helping in detecting lesions other than the main lesion. In addition, our models could also delineate the boundary of csPCa and provide the visualization of suspicious lesions and the adjacent vital structures, which could be helpful for MRI-TRUS fusion targeted biopsy and preoperative preparation, as well as for patient education (39). In the next step, more precise automated divisions, such as 39 sectors (24) or 41 sectors (2), could be investigated to guide prostate biopsy with greater precision.

In this study, the lesion-level sensitivity for detecting csPCa lesions and the DSC for segmenting csPCa lesions were similar to those in previous investigations (19,20,33,40). Compared with these investigations, the data in our study were collected from seven MRI vendors. The results of subgroup analyses (1.5 T vs 3.0 T and Discovery HD 750 scanner vs other scanners) suggest that the models may achieve at least hospital-wide generalization. From a clinical practice point of view, a model with a lower segmentation performance (DSC ~0.6) in a heterogeneous dataset may be more desirable than one with a higher segmentation performance (DSC > 0.8) (21,41) in a homogeneous dataset.

At present, the most popular sequences for csPCa segmentation using deep learning are T2-weighted images or diffusion images. Few studies have compared the performance of diffusion and biparametric deep learning models or directly compared them with clinical assessment (33). Based on lesions, sextants, and zones, the two deep-learning-based models in this study had a comparable diagnostic performance. Based on patients, the sensitivity of the diffusion model was higher than that of the biparametric model. The addition of T2WI increased the patient-level specificity of the biparametric model, but this increase was not statistically significant.

A high false-positive rate of prostate cancer AI systems has been found by some investigators (12,31,32), and some post-processing methods have been tried to reduce this. There is a possibility that our study methodology could falsely increase the false negative rate. Hiremath et al (20) removed all the regions with the largest diameter of < 5 mm to remove some false positives before evaluating the sensitivity and positive predictive value of lesion detection. According to the current detection capabilities of mpMRI in guidelines (24), we took the post-processing step of removing connected domains less than 0.5 cc to reduce false positives. Although some small true positive lesions will also be removed by this approach, it improves the overall effectiveness of the models. There may be a false increase in false positive rates due to the imperfect match between the reference standard annotation and the real pathology. When compared to radical prostatectomy specimens, combined biopsy pathology may miss some lesions (25). It is possible that our radiologists underestimate the extent of lesions or miss them when outlining csPCa foci on MRI images based on cognitive biopsy results (42). If missed lesions were detected by AI, the false positive rates increased.

**Limitations**

First, there is a possible discrepancy introduced by the definition of lesion detection as at least 1 voxel overlap. With this definition, it is possible to have an approximate Dice score of 0 on the lesion level, indicating complete segmentation failure while at the same time reporting 100% lesion detection. Second, as discussed above, postprocessing filters out the smaller ground truth lesion segmentations. Third, this was a single-center retrospective study. AI systems should be tested with multi-center prospective randomized studies to assess performance. Fourth, there is a lack of analysis of data that mismatches between MR images and pathology. Reference standards rely most heavily on image-guided biopsies in this study. Despite the high sensitivity to csPCa using TRUS-guided systematic and targeted biopsy, there is still a false negative rate compared to radical prostatectomy. A training set with spatially well-correlated histopathology is desirable in the next phases. Last, there was a lack of radiologist interaction. In the future, we will invite multiple radiologists with different levels of experience to interpret mpMRI aided with and without AI and compare the reports to the final clinical outcome to determine if AI adds value in real clinical practice.

**Conclusion**

This study shows that the diffusion model trained using ADC and DWI achieved the best performance in detection and localization of csPCa in patients with PSA in the range of 4-10 ng/mL and had a comparable performance with radiologists using PI-RADS (Version 2.1).

**Reference**

1. Heidenreich A, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. European urology 2011;59(1):61-71.

2. 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. European urology 2019;76(3):340-351.

3. Mottet N, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European urology 2021;79(2):243-262.

4. Smith CP, Harmon SA, Barrett T, et al. Intra- and interreader reproducibility of PI-RADSv2: A multireader study. Journal of magnetic resonance imaging : JMRI 2019;49(6):1694-1703.

5. Girometti R, Giannarini G, Greco F, et al. Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. Journal of magnetic resonance imaging : JMRI 2019;49(2):546-555.

6. 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.

7. Byun J, Park KJ, Kim MH, Kim JK. Direct Comparison of PI-RADS Version 2 and 2.1 in Transition Zone Lesions for Detection of Prostate Cancer: Preliminary Experience. Journal of magnetic resonance imaging : JMRI 2020;52(2):577-586.

8. Dwivedi DK, Kumar R, Dwivedi AK, et al. Prebiopsy multiparametric MRI-based risk score for predicting prostate cancer in biopsy-naive men with prostate-specific antigen between 4-10 ng/mL. Journal of magnetic resonance imaging : JMRI 2018;47(5):1227-1236.

9. Liu J, Dong B, Qu W, et al. Using clinical parameters to predict prostate cancer and reduce the unnecessary biopsy among patients with PSA in the gray zone. Scientific reports 2020;10(1):5157.

10. Huang Y, Li ZZ, Huang YL, Song HJ, Wang YJ. Value of free/total prostate-specific antigen (f/t PSA) ratios for prostate cancer detection in patients with total serum prostate-specific antigen between 4 and 10 ng/mL: A meta-analysis. Medicine 2018;97(13):e0249.

11. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022;72(1):7-33.

12. Padhani AR, Turkbey B. Detecting Prostate Cancer with Deep Learning for MRI: A Small Step Forward. Radiology 2019;293(3):618-619.

13. Cipollari S, Pecoraro M, Forookhi A, et al. Biparametric prostate MRI: impact of a deep learning-based software and of quantitative ADC values on the inter-reader agreement of experienced and inexperienced readers. Radiol Med 2022;127(11):1245-1253.

14. Khosravi P, Lysandrou M, Eljalby M, et al. A Deep Learning Approach to Diagnostic Classification of Prostate Cancer Using Pathology-Radiology Fusion. Journal of magnetic resonance imaging : JMRI 2021;54(2):462-471.

15. Hosseinzadeh M, Saha A, Brand P, Slootweg I, de Rooij M, Huisman H. Deep learning-assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge. European radiology 2022;32(4):2224-2234.

16. Mehralivand S, Yang D, Harmon SA, et al. Deep learning-based artificial intelligence for prostate cancer detection at biparametric MRI. Abdom Radiol (NY) 2022;47(4):1425-1434.

17. Hiremath A, Shiradkar R, Fu P, et al. An integrated nomogram combining deep learning, Prostate Imaging-Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. Lancet Digit Health 2021;3(7):e445-e454.

18. Arif M, Schoots IG, Castillo Tovar J, et al. Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI. European radiology 2020;30(12):6582-6592.

19. Reda I, Khalil A, Elmogy M, et al. Deep Learning Role in Early Diagnosis of Prostate Cancer. Technology in cancer research & treatment 2018;17:1533034618775530.

20. Hiremath A, Shiradkar R, Merisaari H, et al. Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps. European radiology 2021;31(1):379-391.

21. Yan K, Wang X, Kim J, Khadra M, Fulham M, Feng D. A propagation-DNN: Deep combination learning of multi-level features for MR prostate segmentation. Computer methods and programs in biomedicine 2019;170:11-21.

22. Thon A, Teichgraber U, Tennstedt-Schenk C, et al. Computer aided detection in prostate cancer diagnostics: A promising alternative to biopsy? A retrospective study from 104 lesions with histological ground truth. PLoS One 2017;12(10):e0185995.

23. Wang Y, Wang M. Selecting proper combination of mpMRI sequences for prostate cancer classification using multi-input convolutional neuronal network. Phys Med 2020;80:92-100.

24. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. European urology 2016;69(1):16-40.

25. Radtke JP, Schwab C, Wolf MB, et al. Multiparametric Magnetic Resonance Imaging (MRI) and MRI-Transrectal Ultrasound Fusion Biopsy for Index Tumor Detection: Correlation with Radical Prostatectomy Specimen. European urology 2016;70(5):846-853.

26. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. 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. The American journal of surgical pathology 2016;40(2):244-252.

27. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 2006;31(3):1116-1128.

28. Zhu Y, Wei R, Gao G, et al. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. Journal of magnetic resonance imaging : JMRI 2019;49(4):1149-1156.

29. Presti JC. Prostate biopsy: current status and limitations. Reviews in urology 2007;9(3):93-98.

30. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger OJA. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. 2016;abs/1606.06650.

31. Gaur S, Lay N, Harmon SA, et al. Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? a multi-center, multi-reader investigation. Oncotarget 2018;9(73):33804-33817.

32. Giannini V, Mazzetti S, Armando E, et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. European radiology 2017;27(10):4200-4208.

33. 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.

34. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet (London, England) 2014;384(9959):2027-2035.

35. Liu C, Liu SL, Wang ZX, et al. Using the prostate imaging reporting and data system version 2 (PI-RIDS v2) to detect prostate cancer can prevent unnecessary biopsies and invasive treatment. Asian journal of andrology 2018;20(5):459-464.

36. Qi Y, Zhang S, Wei J, et al. Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. Journal of magnetic resonance imaging : JMRI 2020;51(6):1890-1899.

37. Chen Y, Ruan M, Zhou B, et al. Cutoff Values of Prostate Imaging Reporting and Data System Version 2.1 Score in Men With Prostate-specific Antigen Level 4 to 10 ng/mL: Importance of Lesion Location. Clinical genitourinary cancer 2021;19(4):288-295.

38. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin 2019;69(2):127-157.

39. Wake N, Rosenkrantz AB, Huang R, et al. Patient-specific 3D printed and augmented reality kidney and prostate cancer models: impact on patient education. 3D printing in medicine 2019;5(1):4.

40. Yuan Y, Qin W, Buyyounouski M, et al. Prostate cancer classification with multiparametric MRI transfer learning model. Medical physics 2019;46(2):756-765.

41. Wang B, Lei Y, Tian S, et al. Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. Medical physics 2019;46(4):1707-1718.

42. 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.

Table 1**.** Demographic and clinical characteristics of 1628 included men.

|  |  |  |
| --- | --- | --- |
| Variable | Model development dataset (n=1428) | Hold-out test dataset (n=200) |
| Age (y), median (IQR) | 69 (63, 77) | 64.5 (60, 70) |
| Number of cases, n (%) |  |  |
| PCa | 907 (63.5) | 100 (50.0) |
| Non-PCa | 521 (36.5) | 100 (50.0) |
| Total PSA (ng/mL), median (IQR) |  |  |
| PCa | 16.2 (9.6, 33.4) | 7.4 (6.0, 8.5) |
| Non-PCa | 8.4 (6.7, 14.4) | 6.7 (5.5, 7.9) |
| Biopsy distribution per patient |  |  |
| Biopsy-naïve | (72) | (91) |
| Previously biopsied | (18) | (9) |
| Per-patient PI-RADS category in patients with PCa, n (%) |  |  |
| PI-RADS 2 | 26 (2.9) | 6 (6.0) |
| PI-RADS 3 | 50 (5.5) | 5 (5.0) |
| PI-RADS 4 | 272 (30.0) | 66 (66.0) |
| PI-RADS 5 | 559 (61.6) | 23 (23.0) |
| Per-patient PI-RADS category in patients without PCa, n (%) |  |  |
| PI-RADS 2 | 401 (77.0) | 90 (90.0) |
| PI-RADS 3 | 84 (16.1) | 8 (8.0) |
| PI-RADS 4 | 31 (6.0) | 2 (2.0) |
| PI-RADS 5 | 5 (0.9) | 0 (0.0) |
| Per-patient maximum Gleason Score, n (%) |  |  |
| 3+3 | 74 (8.2) | 0 (0.0) |
| 3+4 | 248 (27.3) | 48 (48.0) |
| 4+3 | 203 (22.4) | 31 (31.0) |
| 3+5, 5+3, 4+4 | 161 (17.8) | 16 (16.0) |
| 4+5, 5+4 | 221 (24.3) | 5 (5.0) |
| Number of lesions in patients with PCa\*, n (%) |  |  |
| 1 lesion | 419 (46.2) | 60 (60.0) |
| 2 lesions | 216 (23.8) | 24 (24.0) |
| 3 lesions | 133 (14.7) | 11 (11.0) |
| 4 lesions | 58 (6.4) | 4 (4.0) |
| ≥5 lesions | 81 (8.9) | 1 (1.0) |
| Zone distribution of lesions in patients with PCa\*, n (%) |  |  |
| PZ | 258 (26.5) | 43 (39.5) |
| TZ | 237 (24.4) | 41 (37.6) |
| AFMS | 2 (0.2) | 0 (0.0) |
| CZ | 4 (0.4) | 0 (0.0) |
| MZ | 472 (48.5) | 25 (22.9) |

PI-RADS = Prostate Imaging Reporting and Data System, PCa = prostate cancer, IQR = interquartile range. PZ = Peripheral zone, TZ = Transition zone, anterior fibromuscular stroma (AFMS), central zone (CZ), MZ = Multi-zonal\* The results were calculated after post-processing.

Table 2. The type and vendor as well as the magnetic resonance imaging protocols.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 3.0T | | | | |  | 1.5T | |
| Scanner | Scanner A | Scanner B | Scanner C | Scanner D | Scanner E | Scanner F | Scanner G |
| Model development dataset, n | 2 | 91 | 983 | 69 | 94 |  | 156 | 33 |
| Hold-out test dataset, n | 0 | 12 | 163 | 2 | 6 | 15 | 2 |
| DWI |  |  |  |  |  |  |  |
| TR (msec), Min-Max | 4300-4300 | 2000-4474 | 2000-3789 | 1800-3000 | 6624-8615 | 4800-7010 | 2462-5314 |
| TE (msec), Min-Max | 58-58 | 64-70 | 55.9-74.2 | 53-67 | 55-56 | 52-63 | 72-83 |
| FOV (cm2), Min-Max | 26×26-26×26 | 18×18-26×26 | 16×16-24×24 | 19×19-22×22 | 22×22-22×22 | 26×26-26×26 | 24×24-28×28 |
| Section thickness (mm), Min-Max | 3.5-3.5 | 4-4 | 3-5.5 | 3-4 | 4-4 | 4-4 | 4-5.5 |
| Slice spacing (mm), Min-Max | 3.85-3.85 | 4-4 | 4-4 | 3.3-4 | 4-5.5 | 4-4 | 4-5.5 |
| Matrix, Min-Max | 110×110-110×110 | 110×99-144×142 | 96×96-256×256 | 96×96-96×96 | 88×88-224×224 | 96×77-96×77 | 102×102-136×136 |
| Flip angle (degree) | 90 | 90 | 90 | 90 | 90 | 180 | 90 |
| b values, (sec/mm2), Min-Max | 1400-1400 | 1200-1400 | 1000-1400 | 1400-2500 | 800-1400 | 800-1400 | 800-1400 |
| T2WI |  |  |  |  |  |  |  |  |
| TR (msec), Min-Max | 2900-2900 | 2000-3176 | 3000-5356 | 3150-4397 | 3000-3000 |  | 2500-3100 | 3790-4908 |
| TE (msec), Min-Max | 90-90 | 90-90 | 85-94 | 145-145 | 115-115 |  | 70-100 | 90-90 |
| FOV (cm2), Min-Max | 26×26-26×26 | 18×18-26×26 | 16×16-24×24 | 19×19-22×22 | 22×22-22×22 |  | 26×26-26×26 | 24×24-28×28 |
| Section thickness (mm), Min-Max | 3.5-3.5 | 4.0-5.0 | 4.0-5.0 | 3.5-4.0 | 4.0-5.0 |  | 4.0-5.0 | 4.0-4.0 |
| Slice spacing (mm), Min-Max | 3.5-3.5 | 4.0-5.0 | 4.0-5.0 | 3.5-4.0 | 4.0-5.0 |  | 4.0-5.0 | 4.0-4.0 |
| Matrix, Min-Max | 320×256-320×256 | 236×226-260×244 | 320×192-320×256 | 320×256-320×256 | 440-340×440-340 |  | 320×240-320×240 | 340×337-360×347 |
| Flip angle (degree) | 90 | 90 | 90 | 90 | 90 |  | 90 | 90 |

Scanner A: Magnetom Prisma, Siemens Healthcare, Erlangen, Germany, Scanner B: Achieva TX, Philips Healthcare, Best, the Netherlands, Scanner C: Discovery HD 750, Ge Healthcare, Milwaukee, WI, USA, Scanner D: uMR790, UIH, Shanghai, China, Scanner E: Ingenia, Philips Healthcare, Best, The Netherlands, Scanner F: Magnetom Aera, Siemens Healthcare, Erlangen, Germany, and Scanner G: Multiva, Philips Healthcare, Best, The Netherlands.

DWI = diffusion weighted imaging, with a diffusion weighted single-shot gradient echo planar imaging sequence. T2WI =T2-weighted imaging, with a T2-weighted fast spin echo sequence

Table 3 Number of lesions before and after post-processing in Hold-out test

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Patients with csPCa | | | | |  | | Patients without csPCa | | | | | | | | |
|  | Before | Post-processing | After | | |  | |  | | Before | | Post-processing | | After | | |
| Number of lesions | | | |  | | | | | | | | | | |
| Reference standard (n) | 164 | -55 | 109 | |  | |  | | 0 | | 0 | | 0 | | |
| Diffusion model predict | | | |  | | | | | | | | | | |
| TP (n)  FP (n) | 134  53 | -37  -51 | 97  2 | |  | | FP (n) | | 184 | | -159 | | 25 | | |
| Biparametric model predict | | | |  | | | | | | | | | | |
| TP (n)  FP (n) | 131  44 | -24  -40 | 93  4 | |  | | FP (n) | | 210 | | -193 | | 17 | | |
| Number of patients | | | |  | | | | | | | | | | |
| Reference standard (n) | 100 | 0 | 100 | |  | |  | | 100 | | 0 | | 100 | | |
| Diffusion model predict | | | |  | | | | | | | | | | |
| TP (n)  FN (n) | 100  0 | -4  0 | 96  0 | |  | | FP (n)  TN (n) | | 79  21 | | -56  +56 | | 23  77 | | |
| Biparametric model predict | | | |  | | | | | | | | | | |
| TP (n)  FN (n) | 100  0 | -10  0 | 90  0 | |  | | FP (n)  TN (n) | | 93  7 | | -78  +78 | | 15  85 | | |

TP = true positive, FN = false negative, FP = false positive, TN = true negative, csPCa = clinically significant prostate cancer

Table 4. Subgroup sensitivity at the lesion level of the models and PI-RADS assessment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Diffusion model | Biparametric model | PI-RADS assessment | P (Diffusion model vs. Biparametric model) | P (Diffusion model vs. PI-RADS) | P (Biparametric model vs. PI-RADS) |
| Total | 90.0 (97/109) [81.7, 93.6] | 85.3 (93/109) [77.5, 90.8] | 90.8 (99/109) [83.9, 94.9] | 0.289 | 0.625 | 0.754 |
| Based on zone |  |  |  |  |  |  |
| PZ (%) | 93.0 (40/43) [81.4, 97.6] | 88.4 (38/43) [75.5, 94.9] | 93.0 (40/43) [81.4, 97.6] | 0.625 | 1.000 | 0.625 |
| TZ (%) | 90.2 (37/41) [77.5, 96.8] | 80.5 (33/41) [66.0, 89.8] | 78.1 (32/41) [63.3, 88.0] | 0.500 | 0.250 | 1.000 |
| MZ (%) | 88.0 (22/25) [70.0, 95.8] | 88.0 (22/25) [70.0, 95.8] | 92.0 (23/25) [75.0, 97.8] | 1.000 | 1.000 | 1.000 |
| P (PZ vs. TZ) | 0.508 | 0.508 | 0.146 | NA | NA | NA |
| P (PZ vs. MZ) | 1.000 | 1.000 | 1.000 | NA | NA | NA |
| P (TZ vs. MZ) | 1.000 | 1.000 | 0.375 | NA | NA | NA |
| Based on scanner |  |  |  |  |  |  |
| Discovery HD 750 (%) | 87.9 (80/91) [79.6, 93.1] | 83.5 (76/91) [74.6, 89.8] | 86.8 (79/91) [78.4, 92.3] | 0.125 | 0.625 | 0.508 |
| Others (%) | 88.9 (16/18) [67.2, 96.9] | 94.4 (17/18) [74.2, 99.0] | 88.9 (16/18) [67.2 ,96.9] | 1.000 | 1.000 | 1.000 |
| P (Discovery HD 750 vs. Others) | 1.000 | 1.000 | 1.000 | NA | NA | NA |
| Based on field strength |  |  |  |  |  |  |
| 1.5 T (%) | 87.5 (7/8) [52.9, 97.8] | 87.5 (7/8) [52.9, 97.8] | 87.5 (7/8) [52.9, 97.8] | 1.000 | 1.000 | 1.000 |
| 3.0 T (%) | 89.1 (90/101) [81.5, 93.8] | 85.1 (86/101) [76.9, 90.8] | 87.1 (88/101) [79.2, 92.3] | 0.398 | 0.662 | 0.682 |
| P (1.5T vs. 3.0 T) | 0.890 | 0.854 | 0.974 | NA | NA | NA |
| Based on pathology source |  |  |  |  |  |  |
| RP | 87.1 (68/78) [78.0, 92.9] | 84.6 (66/78) [75.0, 91.0] | 88.5 (69/78) [79.5, 93.8] | 0.655 | 0.790 | 0.477 |
| Biopsy | 93.6 (29/31) [77.2, 99.9] | 87.1(27/31) [71.2, 94.9] | 96.8 (30/31) [83.8, 99.4] | 0.390 | 0.559 | 0.163 |
| P (RP vs. Biopsy) | 0.331 | 0.741 | 0.177 | NA | NA | NA |

PI-RADS = Prostate Imaging Reporting and Data System, PZ = Peripheral zone, TZ = Transition zone, MZ = Multiple zones, NA = Not applicable, RP = Radical prostatectomy

\* The difference was statistically significant (P<0.05).

The data shown in brackets represent the 95% confidence intervals (CIs).

Table 5. Diagnostic efficacy and comparisons of the models and the PI-RADS assessment based on patients and sextants

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Diffusion model | Biparametric model | PI-RADS assessment | P (Diffusion model vs. Biparametric model) | P (Diffusion model vs. PI-RADS) | P (Biparametric model vs. PI-RADS) |
| Based on patients |  | | | | | |
| AUC | 0.865 [0.762, 0.872] | 0.875 [0.821, 0.917] | 0.920 [0.873, 0.954] | 0.678 | 0.058 | 0.094 |
| Sensitivity (%) | 96.0 (96/100) [90.1, 98.9] | 90.0 (90/100) [82.4, 95.1] | 94.0 (94/100) [87.4, 97.8] | 0.031\* | 0.625 | 0.289 |
| Specificity (%) | 77.0 (77/100) [(67.5, 84.8] | 85.0 (85/100) [76.5, 91.4] | 90.0 (90/100) [82.4, 95.1] | 0.096 | 0.031\* | 0.383 |
| Accuracy (%) | 86.5 (173/200) | 87.5 (175/200) | 92.0 (184/200) | 0.839 | 0.091 | 0.137 |
| Based on sextants |  |  |  |  |  |  |
| AUC | 0.895 [0.876, 0.912] | 0.893 [0.874, 0.910] | 0.734 [0.708, 0.759] | 0.777 | 0.000\* | 0.000\* |
| Sensitivity (%) | 87.0 (262/301) [82.7, 90.6] | 85.4 (257/301) [80.9, 89.2] | 53.5 (152/301) [47.7, 59.2] | 0.332 | 0.000\* | 0.000\* |
| Specificity (%) | 92.0 (827/899) [89.7, 93.5] | 93.1 (838/899) [91.1, 94.7] | 93.3 (839/899) [91.5, 95.0] | 0.118 | 0.327 | 1.000 |
| Accuracy (%) | 90.8 (1089/1200) [90.0, 92.3] | 91.3 (1095/1200) [89.5, 92.7] | 83.3 (991/1200) [80.3, 84.6] | 0.511 | 0.000\* | 0.000\* |

PI-RADS = Prostate Imaging Reporting and Data System, AUC = Area Under the Curve.

\* The difference was statistically significant (P<0.05).

The data shown in brackets represent the 95% confidence intervals (CIs).

Table 6**.** Diagnostic efficacy and comparisons of the models and the PI-RADS assessment based on zones

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Diffusion model | Biparametric model | PI-RADS assessment | P (Diffusion model vs. Biparametric model) | P (Diffusion model vs. PI-RADS) | P (Biparametric model vs. PI-RADS) |
| Total |  | | | | | |
| AUC | 0.931 [0.911, 0.947] | 0.917 [0.896, 0.935] | 0.863 [0.837, 0.886] | 0.282 | 0.001\* | 0.014\* |
| Sensitivity (%) | 90.4 (113/125) [83.4, 94.4] | 86.4 (108/125) [79.3, 91.3] | 79.2 (99/125) [71.3, 85.4] | 0.180 | 0.007\* | 0.137 |
| Specificity (%) | 95.7 (646/675) [93.4, 97.0] | 97.0 (655/675) [95.5, 98.1] | 93.3 (630/675) [91.2, 95.0] | 0.078 | 0.069 | 0.001\* |
| Accuracy (%) | 94.9 (759/800) [93.1, 96.2] | 95.4 (763/800) [93.7, 96.6] | 91.1 (629/800) [90.0, 93.0] | 0.584 | 0.002\* | 0.000\* |
| PZ |  |  |  |  |  |  |
| AUC | 0.926 [88.1, 95.8] | 0.907 [0.858, 0.944] | 0.910 [0.861, 0.945] | 0.282 | 0.535 | 0.936 |
| Sensitivity (%) | 91.7 (55/60) [81.9, 96.4] | 85.0 (51/60) [73.9, 91.9] | 98.3 (59/60) [91.1, 99.7] | 0.125 | 0.219 | 0.021\* |
| Specificity (%) | 93.6 (131/140) [88.2, 96.6] | 96.4 (135/140) [91.9, 98.5] | 83.6 (117/140) [76.6, 88.8] | 0.500 | 0.001\* | 0.000\* |
| Accuracy (%) | 93.0 (186/200) [88.6, 95.8] | 93.0 (186/200) [88.6, 95.8] | 88.0 (176/200) [82.8, 91.8] | 0.500 | 0.388 | 0.791 |
| TZ |  | | | | | |
| AUC | 0.898 [0.848, 0.936] | 0.903 [0.854, 0.940] | 0.810 [0.748, 0.816] | 0.812 | 0.010\* | 0.007\* |
| Sensitivity (%) | 92.9 (52/56) [83.0, 97.2] | 91.2 (51/56) [80.7, 96.1] | 66.1 (37/56) [53.0, 77.1] | 1.000 | 0.000\* | 0.001\* |
| Specificity (%) | 86.8 (125/144) [80.3, 91.4] | 89.6 (129/144) [83.5, 93.6] | 95.8 (138/144) [91.2, 98.1] | 0.454 | 0.015\* | 0.049\* |
| Accuracy (%) | 88.5 (177/200) [83.3, 92.2] | 90.0 (180/200) [85.1, 93.4] | 87.5 (175/200) [82.2, 91.4] | 0.648 | 0.874 | 0.486 |
| CZ |  | | | | | |
| AUC | 0.697 [0.629, 0.760] | 0.800 [0.738, 0.853] | 0.500 [0.429, 0.571] | 0.305 | 0.107 | 0.014\* |
| Sensitivity (%) | 40.0 (2/5) [11.8, 76.9] | 60.0 (3/5) [23.1, 88.2] | 0 (0/5) [0, 43.5] | 1.000 | 0.500 | 0.250 |
| Specificity (%) | 99.5 (194,195) [97.2, 99.9] | 100.0 (195/195) [98.1, 100.0] | 100.0 (195/195) [98.1, 100.0] | 1.000 | 1.000 | 1.000 |
| Accuracy (%) | 98.0 (196/200) [95.0, 99.2] | 99.0 (198/200) [96.4, 99.7] | 97.5 (195/200) [94.3, 98.9] | 0.500 | 1.000 | 0.250 |
| AFMS |  | | | | | |
| AUC | 1.000 [0.982, 1.000] | 0.875 [0.821, 0.917] | 0.834 [0.775, 0.883] | 0.317 | 0.186 | 0.000\* |
| Sensitivity (%) | 100.0 (4/4) [51.0, 100.0] | 75.0 (3/4) [30.1, 95.4] | 75.0 (3/4) [30.1, 95.4] | 1.000 | 1.000 | 1.000 |
| Specificity (%) | 100.0 (196/196) [98.1, 100.0] | 100.0 (196/196) [98.1, 100.0] | 91.8 (180/196) [87.2, 94.9] | 1.000 | 0.000\* | 0.000\* |
| Accuracy (%) | 100.0 (200/200) [98.1, 100.0] | 99.5 (199/200) [97.2, 99.9] | 91.5 (183/200) [86.8, 94.6] | 1.000 | 0.000\* | 0.000\* |

PI-RADS = Prostate Imaging Reporting and Data System, AUC = Area Under the Curve. PZ = Peripheral zone, TZ = Transition zone, CZ = central zone, AFMS = anterior fibromuscular stroma \* The difference was statistically significant (P<0.05). The data shown in brackets represent the 95% confidence intervals (CI).

**Figure captions**

**Figure 1.** The workflow of patient enrollment and distribution. DRE = digital rectal examination, TRUS = transrectal ultrasound, RP = radical prostatectomy, PCa = prostate cancer.

**Figure 2.** Training and testing process of the models.

**Figure 3.** Notched box plots show the Dice similarity coefficient (DSC) of the diffusion model and biparametric model. Lengths of whiskers are limited to a maximum of 1.5 times the interquartile range. Notches indicate a 95% confidence interval around the median.

**Figure 4.** Examples of csPCa segmentation. **a –** **d** Apparent diffusion coefficient (ADC) map (**a**) and T2-weighted imaging (T2WI) (**c**) showed local low intensity (white arrow) with high intensity (white arrow) on the diffusion-weighted images (DWI) (**b**) in the right peripheral zone, and the prediction results of the diffusion model (**d**, green line) and the biparametric model (**d**, blue line) were highly consistent with manual annotation (**d**, red line). **e – h** ADC (**e**) and T2WI (**g**) showed moderate low intensity (yellow arrow) with moderate high intensity (yellow arrow) on DWI (**f**) in the right peripheral zone of the prostate base. The prediction result of the diffusion model (**h**, green line) was consistent with manual annotation (**h**, red line), while the biparametric model missed the lesion. **i – l** ADC (**i**), DWI (**j**), and T2WI (**k**) showed an asymmetrical central zone (orange arrow). The predicted results were false positives (**l**, green line for the diffusion model and blue line for the biparametric model) in the right central zone. **m – p** ADC (**m**), DWI (**n**), and T2WI (**o**) showed a benign hyperplasia node in the transition zone (red arrow). The predicted results were false positives (**p**, green line for the diffusion model and blue line for the biparametric model).

**Figure 5.** Examples of a cancerous patient with small positive lesions. **a - c** Two small cancer lesions (white arrows) with volumes of 0.25 cc and 0.27 cc in the peripheral zone and one larger lesion in the transition zone (yellow arrow). These lesions showed high intensity in diffusion-weighted imaging (**b**) and low intensity in the apparent diffusion coefficient map (**a**) and T2-weighted imaging (**c**). **d** Original prediction results of the models. The two small true cancer lesions (white arrows) and one larger lesion (yellow arrow) were annotated by ground truth (red line) and were detected by the diffusion model (green line) and the parametric model (blue line). The post-processing (**e**) filtered the two small true cancer lesions (white arrows), and the larger lesion was retained (yellow).

**Figure 6.**  Free response receiver operating characteristic curves of the two models. The longitudinal ordinate is the lesion localization fraction (LLF), which represents the number of true positive lesions divided by the total number of lesions. The transverse ordinate is non-lesion fraction (NLF), which represents the number of predicted false-positive lesions divided by the total number of cases. The curves of the two models were very close and partially crossed, suggesting that the detection efficacy of the two models was similar at the lesion level.

**Figure 7.** Receiver operating characteristic curve analysis of the two models and PI-RADS in the hold-out test dataset based on patients (**a**), sextants (**b**), and anatomic zones (**c**). AUC = area under the curve.