

# A Validation Study of the Deep-Learning-Based Prostate Imaging Reporting and Data System Scoring Algorithm

Ryusuke Irie<sup>1,2\*</sup>, Maki Amano<sup>1</sup>, Kazusa Sugeno<sup>1</sup>, Shingo Okada<sup>1</sup>, Ali Kamen<sup>3</sup>, Bin Lou<sup>3</sup>, Heinrich von Busch<sup>4</sup>, Robert Grimm<sup>5</sup>, Dorin Comaniciu<sup>3</sup>, Toshiaki Akashi<sup>1</sup>, Ryohei Kuwatsuru<sup>1</sup>, Shigeo Horie<sup>6</sup>, Kanako K. Kumamaru<sup>1</sup>, Shigeki Aoki<sup>1</sup>

<sup>1</sup>Department of Radiology, Juntendo University School of Medicine, Tokyo, Japan

<sup>2</sup>Department of Radiology, Juntendo University Shizuoka Hospital, Shizuoka, Japan

<sup>3</sup>Digital Technology and Innovation Division, Siemens Healthineers, Princeton, USA

<sup>4</sup>Digital Health, Siemens Healthcare GmbH, Erlangen, Germany

<sup>5</sup>MR Application Predevelopment, Siemens Healthcare GmbH, Erlangen, Germany

<sup>6</sup>Department of Urology, Juntendo University School of Medicine, Tokyo, Japan

Email: \*ririe@juntendo.ac.jp

**How to cite this paper:** Irie, R., Amano, M., Sugeno, K., Okada, S., Kamen, A., Lou, B., von Busch, H., Grimm, R., Comaniciu, D., Akashi, T., Kuwatsuru, R., Horie, S., Kumamaru, K.K. and Aoki, S. (2022) A Validation Study of the Deep-Learning-Based Prostate Imaging Reporting and Data System Scoring Algorithm. *Open Journal of Radiology*, 12, 59-67.

<https://doi.org/10.4236/ojrad.2022.123007>

**Received:** August 6, 2022

**Accepted:** September 6, 2022

**Published:** September 9, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

**Purpose:** The Prostate Imaging Reporting and Data System (PI-RADS) was introduced to standardize prostate cancer diagnosis by MRI. However, the inter-reader agreement by PI-RADS scoring is not always high. The purpose of this study was to validate a deep-learning-based diagnostic algorithm of PI-RADS. **Methods:** We applied a Siemens Healthineers Prostate Artificial Intelligence (AI) prototype (work in progress) for fully automated prostate lesion detection, classification and reporting. More than 2000 bi-parametric MRI studies along with the PI-RADS reports were included as training, validation, and test data. This prospective validation study includes 101 consecutive patients suspected of prostate cancer, and 100 patients were included in the analysis. All subjects underwent a noncontrast-enhanced bi-parametric MRI including T2-weighted and diffusion-weighted imaging. Two board-certified radiologists independently scored the PI-RADS, and if there were disagreements; another radiologist confirmed the diagnosis. We compared the AI results with the interpretation results by the radiologists. **Results:** The sensitivity of our AI model for PI-RADS  $\geq 4$  was 0.76, and the specificity was 0.76. For the cases with PI-RADS  $\geq 3$ , the sensitivity was 0.69, and the specificity was 0.76. In the lesion-based analysis, AI detection rates of PI-RADS 3, 4, 5 lesions in the peripheral zone were 43%, 63%, and 100%, respectively. In the transition zone, AI detection rates of PI-RADS 3, 4, 5 were 30%, 54%, and 100%, respectively. **Conclusion:** Our deep-learning-based algorithm has been

validated and shown to help score PI-RADS.

## Keywords

Deep Learning, Artificial Intelligence, Prostate Cancer, PI-RADS

---

## 1. Introduction

Prostate cancer is the second most frequently diagnosed cancer in males in the world, and it is the most frequently diagnosed cancer among men especially in developed countries [1] [2].

The difference in prostate cancer diagnosis rates between regions is largely due to the prevalence of prostate-specific antigen (PSA) testing [3]. PSA testing is widely used in screening for prostate cancer, but there is a certain probability of false positives and false negatives [4]. The definitive diagnosis is a pathological diagnosis by needle biopsy, but is highly invasive [5]. MRI has come to be used as a noninvasive technique supporting the diagnosis and localization of prostate cancer [6]. The Prostate Imaging Reporting and Data System (PI-RADS) was introduced to standardize prostate cancer diagnosis by MRI [7]. However, image interpretation by PI-RADS scoring requires experience, and it has been reported that even if this score system is used, the inter-reader agreement is not always high [8] [9] [10].

In recent years, artificial intelligence (AI) has been actively used in the field of diagnostic imaging [11] [12]. In particular, deep learning can potentially discriminate suspicious and nonsuspicious images with very high accuracy. There are many reports using AI in the field of prostate cancer, such as computer-aided diagnosis of the Gleason score from pathological images [13].

Siemens Healthineers has developed a system that detects and segments prostate lesions and outputs PI-RADS scores using bi-parametric MRI including T2-weighted images (T2WI) and diffusion-weighted images (DWI) as input. Utilization of the AI model is expected to contribute to quick and accurate diagnosis of prostate cancer. In order to operate the developed AI model, it must be validated in an actual clinical setting. The purpose of this study was to validate a deep-learning-based diagnostic algorithm of PI-RADS compared with the interpretation of radiologists.

## 2. Material and Methods

### 2.1. AI Model Development

We applied an AI prototype (Prostate AI Prototype version on December 21, 2019, work in progress, Siemens Healthcare, Erlangen, Germany) for fully automated prostate lesion detection, classification and reporting. The prototype consists of a web-based reading platform for viewing and interpreting the image data and AI-based results, as well as the actual AI preprocessing pipeline and a

component for lesion detection and classification, based on deep learning [14] [15]. The preprocessing stage begins with a fully automated segmentation of the prostate gland and peripheral zone on T2WI using a 3D convolutional neural network (CNN). Then, T2WI and DWI are co-registered, and an apparent diffusion coefficient (ADC) map and calculated DWI at  $b = 2000 \text{ s/mm}^2$  are computed. Using 2D CNNs, Prostate AI automatically detects clinically relevant lesions (PI-RADS 3 or above) within the prostate gland based on the T2WI, ADC and  $b = 2000 \text{ s/mm}^2$  images, followed by a false-positive reduction step using a 2.5D multi-scale neural network. Finally, an independently trained 2.5D convolutional neural network predicts the PI-RADS v2 category of each lesion. 2170 bi-parametric MRI studies from seven different clinical institutions were used during model training, testing and validation.

## 2.2. Sample Selection of the Validation Study

The present prospective analysis was approved by the Institutional Review Board, and 101 consecutive patients suspected of prostate cancer from March to July 2019 were included. The mean age  $\pm$  standard deviation was  $67.0 \pm 10.2$  years. The mean PSA for all patients was  $10.4 \mu\text{g/mL}$  (range 0.018 to  $203 \mu\text{g/mL}$ ).

## 2.3. Validation Study Procedure

All subjects underwent a noncontrast-enhanced bi-parametric MRI including T2WI and DWI ( $b = 0, 1400 \text{ s/mm}^2$ ). MRI scans were conducted using a 3-Tesla clinical scanner (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany). Two board-certified radiologists (R. I. and M. A.) independently scored the PI-RADS score for each case, and if there were disagreements, another radiologist (S. O.) made a final decision and confirmed the diagnosis. When multiple lesions were detected in a single patient, the lesion with the highest category was adopted. We compared the results of the AI model with the interpretation results by the radiologists.

## 3. Results

Of the 101 patients, one was excluded because the misalignment of the T2WI and DWI was so strong due to a gross body movement between the image series that it could not be analyzed by the AI model. In total, 100 patients were included in this study.

As a result of the final diagnosis by the radiologists, the number of cases of PI-RADS v2 category  $\leq 2, 3, 4,$  and  $5$  were 42, 20, 21, and 17, respectively. Inter-reader agreement on the PI-RADS score was substantial (weighted kappa = 0.62). The average reading time per case was 84 seconds and 72 seconds for the two readers. With the AI model, automated prostate segmentation, lesion detection and segmentation as well as PI-RADS classification took about 7 seconds per case.

For the cases with PI-RADS  $\geq 4$ , the AI model correctly identified 29 cases of

those as category  $\geq 4$ . The sensitivity of our AI model for PI-RADS  $\geq 4$  was 0.76, and the specificity was 0.76. For the cases with PI-RADS  $\geq 3$ , the AI model correctly diagnosed 40 cases as category  $\geq 3$ . The sensitivity for PI-RADS  $\geq 3$  was 0.69, and the specificity was 0.76 (Table 1). In the lesion-based analysis, 7 PI-RADS 3, 16 PI-RADS 4, and 10 PI-RADS 5 lesions were identified as PI-RADS  $\geq 3$  in the peripheral zone, with AI detection rates of 43%, 63%, and 100%, respectively. Moreover, 20 PI-RADS 3, 13 PI-RADS 4, and 8 PI-RADS 5 lesions were identified in the transition zone, with AI detection rates of 30%, 54%, and 100%, respectively (Table 2).

#### 4. Discussion

For lesions of category 4 and above, the AI model correctly diagnosed the lesions with an accuracy of 76% and 76% sensitivity/76% specificity.

The AI model correctly diagnosed lesions larger than 15 mm in size (Figure 1), except for one case (Figure 2). Moreover, the lesion was detected even in one miscategorized case. The reason one PI-RADS 5 lesion was diagnosed as category 3 may be that the lesion was too large for the AI model to recognize the boundary of the lesion.

More than half (62%) of the PI-RADS 4 lesions smaller than 15 mm were correctly detected (Figure 3) though 8 of them were classified as PI-RADS 5. Some cases with small lesions could not be detected correctly by the AI model (Figure 4). Small but clinically significant cancers should not be overlooked.

The detection rate of lesions in PI-RADS 3, especially in the transition zone, was low. In our institution, radiologists tended to recognize areas with faint

**Table 1.** Confusion matrix of radiologists and AI diagnosis.

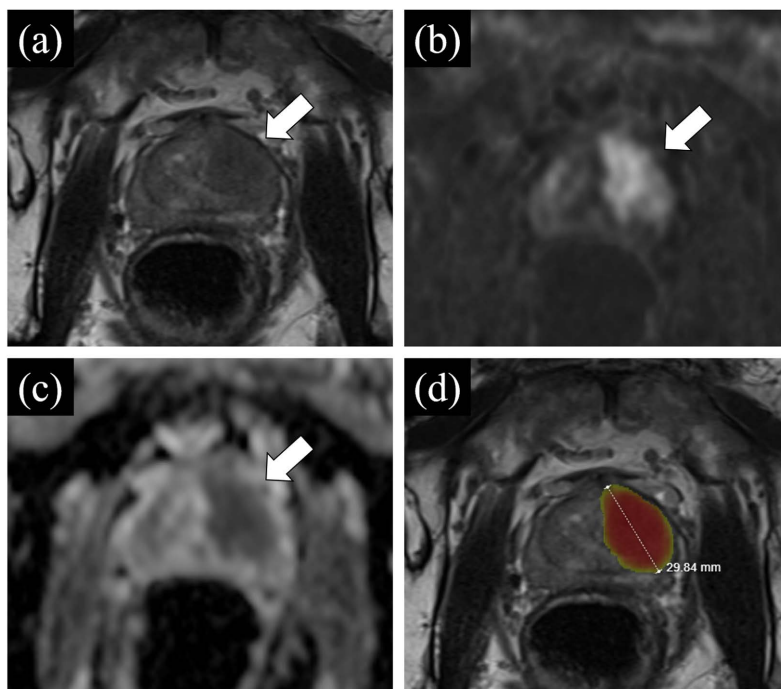
		AI diagnosis			
		PI-RADS $\leq 2$	PI-RADS 3	PI-RADS 4	PI-RADS 5
Radiologists' diagnosis	PI-RADS $\leq 2$	32	3	4	3
	PI-RADS 3	12	0	3	5
	PI-RADS 4	6	2	5	8
	PI-RADS 5	0	1	0	16

AI: artificial intelligence, PI-RADS: Prostate Imaging-Reporting and Data System.

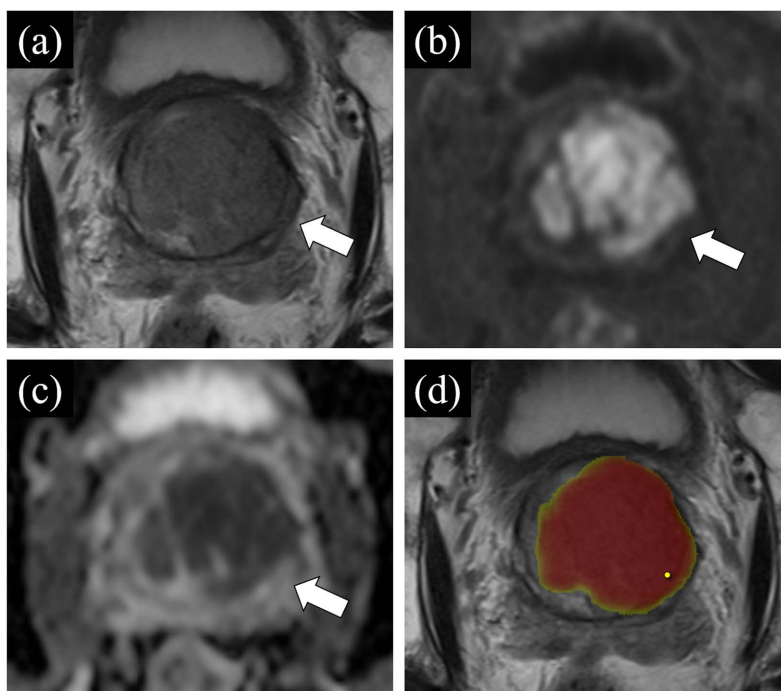
**Table 2.** Detection rate of AI diagnosis for each PI-RADS category.

	Peripheral zone	Transition zone
PI-RADS 3	3/7 (43%)	6/20 (30%)
PI-RADS 4	10/16 (63%)	7/13 (54%)
PI-RADS 5	10/10 (100%)	8/8 (100%)
All	23/33 (70%)	21/41 (51%)

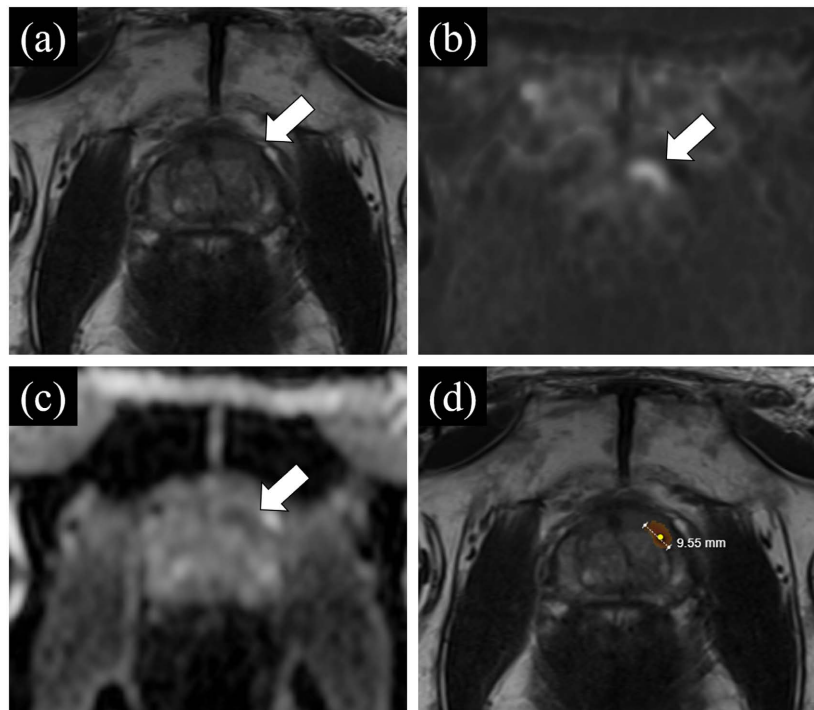
AI: artificial intelligence, PI-RADS: Prostate Imaging-Reporting and Data System.



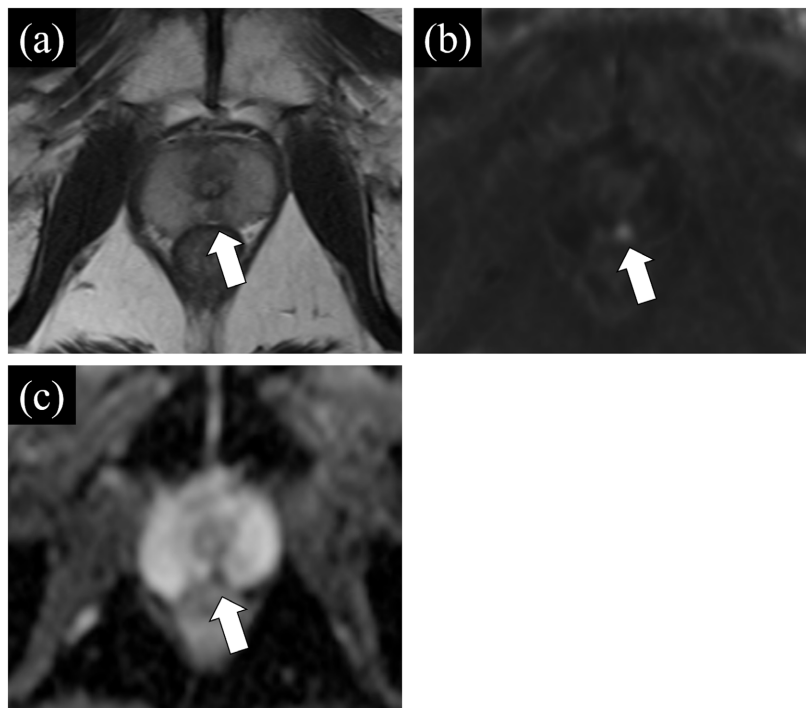
**Figure 1.** A correctly diagnosed lesion in the transition zone. The size of the lesion is 30 mm in a maximum diameter, showing low signal on T2WI (a), high signal on DWI (b), and low ADC value (c). The radiologist's diagnosis was PI-RADS 5, and Prostate AI correctly identified the lesion and diagnosed PI-RADS 5 (d).



**Figure 2.** A correctly identified but miscategorized case. The lesion was mainly in the left transition zone and is widespread. The size of the lesion is 53 mm in a maximum diameter, showing low signal on T2WI (a), high signal on DWI (b), and low ADC value (c). The radiologist's diagnosis was PI-RADS 5. Prostate AI correctly identified the lesion (d) but categorized it as PI-RADS 3.



**Figure 3.** A correctly diagnosed lesion in the peripheral zone. The size of the lesion is 10 mm in a maximum diameter, showing low signal on T2WI (a), high signal on DWI (b), and low ADC value (c). The radiologist’s diagnosis was PI-RADS 4, and Prostate AI correctly identified the lesion and diagnosed PI-RADS 4 (d).



**Figure 4.** A false negative case in the peripheral zone. The size of the lesion is 6 mm in a maximum diameter, showing low signal on T2WI (a), high signal on DWI (b), and low ADC value (c). The radiologist’s diagnosis was PI-RADS 4, but Prostate AI was unable to identify the lesion.

hyperintensity on diffusion-weighted images as lesions, even in the transition zone. Diagnosis of PI-RADS 3 lesions is often controversial among radiologists, so future studies will be needed to assess the actual cancer detection rates based on histopathological samples. Prostate cancer is often not detected pathologically in lesions of PI-RADS 3 [9] [10]. Therefore, it is considered important to correctly diagnose lesions of PI-RADS 4 or higher, and the present result was considered to be acceptable.

False positives in AI diagnosis were caused by BPH nodules, chronic prostatitis, and rectal gas artifact. These conditions cannot be diagnosed by signal intensity alone and require careful consideration of morphology and image properties, which the AI is trained to perform but still does not always get right.

In future clinical practice, the radiologist will make the final diagnosis after AI presents the lesion. If there are many false positives, the confirmation work of radiologists will increase, but if there are many false negatives, there is a possibility that oversights will increase. It is necessary to use AI diagnosis support wisely depending on the situation.

In this validation study, one of the limitations is that no comparison with histopathological diagnosis has been made. This was done by purpose, as we wanted to reflect a clinical, prebiopsy scenario as accurately as possible. The PI-RADS category does not indicate the definite existence of prostate cancer, so the algorithm was trained on detecting radiological lesions and the purpose of the software is to support radiologists during their work. It would also be clinically important to evaluate the pathology-based truth. Second, we used PI-RADS v2 and not v2.1. During the truthing process, v2 was the most recent reference system, and all consequent steps were designed based on this system. Third, the validation study was done on one MR device in one institution. In future studies, further research at more institutions and studies using MRI of different vendors are desired.

## 5. Conclusion

Our deep-learning-based algorithm has been validated and shown to help score PI-RADS.

## Acknowledgements

Data and annotations for the training of the algorithm were provided by Dr. Henkjan Huisman (Radboud University Medical Center, Nijmegen, NL), Dr. David Winkel (Universitätsspital Basel, Basel, Switzerland), Dr. Moon Hyung Choi (Eunpyeong St. Mary's Hospital, Catholic University of Korea, Seoul, Republic of Korea), Prof. Dr. Dieter Szolar (Diagnostikum Graz Süd-West, Graz, Austria), Dr. Evan Johnson and Dr. Andrew Rosenkrantz (New York University, NYC, NY, USA), Dr. Pengyi Xing (Radiology Department, Changhai Hospital of Shanghai, China), Dr. Tobias Penzkofer (Charité, Universitätsmedizin Berlin, Berlin, Germany), Dr. Ivan Shabunin (Patero Clinic, Moscow, Russia), Dr. Fer-

gus Coakley (Diagnostic Radiology, School of Medicine, Oregon Health and Science University, Portland, OR, USA), and Dr. Steven Shea (Department of Radiology, Loyola University Medical Center, Maywood, IL, USA).

This work was supported by AMED under grant number JP19lk1010025h9902.

### Author Information

Ali Kamen is an employee of Siemens Healthineers and has stock ownership and patent royalties/licensing fees (Siemens Healthineers). Bin Lou is an employee of Siemens Healthineers. Heinrich von Busch is an employee of Siemens Healthcare GmbH and has stock ownership (Siemens Healthcare GmbH). Robert Grimm is an employee of Siemens Healthcare GmbH. Dorin Comaniciu is an employee of Siemens Healthineers and has stock ownership (Siemens Healthineers).

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015) Global Cancer Statistics, 2012. *CA: A Cancer Journal for Clinicians*, **65**, 87-108. <https://doi.org/10.3322/caac.21262>
- [2] Bashir, M.N. (2015) Epidemiology of Prostate Cancer. *Asian Pacific Journal of Cancer Prevention*, **16**, 5137-5141. <https://doi.org/10.7314/APJCP.2015.16.13.5137>
- [3] Stamey, T.A., Yang, N., Hay, A.R., McNeal, J.E., Freiha, F.S. and Redwine, E. (1987) Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. *New England Journal of Medicine*, **317**, 909-916. <https://doi.org/10.1056/NEJM198710083171501>
- [4] Barry, M.J. and Simmons, L.H. (2017) Prevention of Prostate Cancer Morbidity and Mortality: Primary Prevention and Early Detection. *Medical Clinics of North America*, **101**, 787-806. <https://doi.org/10.1016/j.mcna.2017.03.009>
- [5] Wade, J., Rosario, D.J., Macefield, R.C., Avery, K.N., Salter, C.E., Goodwin, M.L., *et al.* (2013) Psychological Impact of Prostate Biopsy: Physical Symptoms, Anxiety, and Depression. *Journal of Clinical Oncology*, **31**, 4235-4241. <https://doi.org/10.1200/JCO.2012.45.4801>
- [6] Salami, S.S., Vira, M.A., Turkbey, B., Fakhoury, M., Yaskiv, O., Villani, R., *et al.* (2014) Multiparametric Magnetic Resonance Imaging Outperforms the Prostate Cancer Prevention Trial Risk Calculator in Predicting Clinically Significant Prostate Cancer. *Cancer*, **120**, 2876-2882. <https://doi.org/10.1002/cncr.28790>
- [7] Hamoen, E.H.J., de Rooij, M., Witjes, J.A., Barentsz, J.O. and Rovers, M.M. (2015) Use of the Prostate Imaging Reporting and Data System (PI-RADS) for Prostate Cancer Detection with Multiparametric Magnetic Resonance Imaging: A Diagnostic Meta-Analysis. *European Urology*, **67**, 1112-1121. <https://doi.org/10.1016/j.eururo.2014.10.033>
- [8] Woo, S., Suh, C.H., Kim, S.Y., Cho, J.Y. and Kim, S.H. (2017) Diagnostic Performance of Prostate Imaging Reporting and Data System Version 2 for Detection of



- Prostate Cancer: A Systematic Review and Diagnostic Meta-Analysis. *European Urology*, **72**, 177-188. <https://doi.org/10.1016/j.eururo.2017.01.042>
- [9] Hofbauer, S.L., Maxeiner, A., Kittner, B., Heckmann, R., Reimann, M., Wiemer, L., *et al.* (2018) Validation of Prostate Imaging Reporting and Data System Version 2 for the Detection of Prostate Cancer. *Journal of Urology*, **200**, 767-773. <https://doi.org/10.1016/j.juro.2018.05.003>
- [10] Thai, J.N., Narayanan, H.A., George, A.K., Siddiqui, M.M., Shah, P., Mertan, F.V., *et al.* (2018) Validation of PI-RADS Version 2 in Transition Zone Lesions for the Detection of Prostate Cancer. *Radiology*, **288**, 485-491. <https://doi.org/10.1148/radiol.2018170425>
- [11] Greer, M.D., Lay, N., Shih, J.H., Barrett, T., Bittencourt, L.K., Borofsky, S., *et al.* (2018) Computer-Aided Diagnosis Prior to Conventional Interpretation of Prostate mpMRI: An International Multi-Reader study. *European Radiology*, **28**, 4407-4417. <https://doi.org/10.1007/s00330-018-5374-6>
- [12] Yuan, Y., Qin, W., Buyyounouski, M., Ibragimov, B., Hancock, S., Han, B., *et al.* (2019) Prostate Cancer Classification with Multiparametric MRI Transfer Learning Model. *Medical Physics*, **46**, 756-765. <https://doi.org/10.1002/mp.13367>
- [13] Nagpal, K., Foote, D., Liu, Y., Chen, P.C., Wulczyn, E., Tan, F., *et al.* (2019) Development and Validation of a Deep Learning Algorithm for Improving Gleason Scoring of Prostate Cancer. *NPJ Digital Medicine*, **2**, Article No. 48. <https://doi.org/10.1038/s41746-019-0112-2>
- [14] Yu, X., Lou, B., Shi, B., Winkel, D., Arrahmane, N., Diallo, M., *et al.* (2020) False Positive Reduction Using Multiscale Contextual Features for Prostate Cancer Detection in Multi-Parametric MRI Scans. *Proceedings of 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, 3-7 April 2020, 1355-1359. <https://doi.org/10.1109/ISBI45749.2020.9098338>
- [15] Winkel, D.J., Tong, A., Lou, B., Kamen, A., Comaniciu, D., Disselhorst, J.A., *et al.* (2021) A Novel Deep Learning Based Computer-Aided Diagnosis System Improves the Accuracy and Efficiency of Radiologists in Reading Biparametric Magnetic Resonance Images of the Prostate: Results of a Multireader, Multicase Study. *Investigative Radiology*, **56**, 605-613. <https://doi.org/10.1097/RLI.0000000000000780>