European Journal of Radiology 175 (2024) 111470

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Using deep learning to optimize the prostate MRI protocol by assessing the diagnostic efficacy of MRI sequences

Stefan J. Fransen [a,*], Christian Roest [a], Quintin Y. Van Lohuizen [a], Joeran S. Bosma [b], Frank F. J. Simonis [c], Thomas C. Kwee [a], Derya Yakar [a], Henkjan Huisman [b]

[a] *University Medical Centre Groningen, Department of Radiology, Hanzeplein 1, 9713 GZ, Groningen, the Netherlands*
[b] *University Medical Centre Nijmegen, DIAG, Geert Grooteplein Zuid 10, 6525 GA, Nijmegen, the Netherlands*
[c] *Technical University Twente, TechMed Centre, Hallenweg 5, 7522 NH, Enschede, the Netherlands*

ARTICLE INFO

Keywords:
Diffusion Magnetic Resonance Imaging
Prostatic Neoplasms
Artificial Intelligence
Acceleration
Diagnostic Imaging

ABSTRACT

Purpose: To explore diagnostic deep learning for optimizing the prostate MRI protocol by assessing the diagnostic efficacy of MRI sequences.
Method: This retrospective study included 840 patients with a biparametric prostate MRI scan. The MRI protocol included a T2-weighted image, three DWI sequences (b50, b400, and b800 s/mm^2), a calculated ADC map, and a calculated b1400 sequence. Two accelerated MRI protocols were simulated, using only two acquired b-values to calculate the ADC and b1400. Deep learning models were trained to detect prostate cancer lesions on accelerated and full protocols. The diagnostic performances of the protocols were compared on the patient-level with the area under the receiver operating characteristic (AUROC), using DeLong's test, and on the lesion-level with the partial area under the free response operating characteristic (pAUFROC), using a permutation test. Validation of the results was performed among expert radiologists.
Results: No significant differences in diagnostic performance were found between the accelerated protocols and the full bpMRI baseline. Omitting b800 reduced 53% DWI scan time, with a performance difference of + 0.01 AUROC ($p = 0.20$) and −0.03 pAUFROC ($p = 0.45$). Omitting b400 reduced 32% DWI scan time, with a performance difference of −0.01 AUROC ($p = 0.65$) and + 0.01 pAUFROC ($p = 0.73$). Multiple expert radiologists underlined the findings.
Conclusions: This study shows that deep learning can assess the diagnostic efficacy of MRI sequences by comparing prostate MRI protocols on diagnostic accuracy. Omitting either the b400 or the b800 DWI sequence can optimize the prostate MRI protocol by reducing scan time without compromising diagnostic quality.

1. Introduction

The increasing diagnostic accuracy of AI allows for studying its usefulness beyond assisting diagnosis [1–3]. With AI able to perform at an expert level, it can also assist in other tasks that the human expert performs, including comparing the efficacy of different imaging protocols. Expert-level AI can quantify and rank different diagnostic imaging protocols by their performance in a clinical task [2]. This underexplored ability could help speed up many diagnostic workflows in radiology by identifying and helping to omit redundant, diagnostically irrelevant MRI sequences.

Prostate MRI protocol efficiency has recently become critical due to changes in the European triage guideline for clinically significant prostate cancer (csPCa), putting prostate MRI scans first in the diagnostic pathway [4,5]. The increasing demand for prostate MRI is part of a bigger trend of a strong increase in radiological examinations. Faster, optimized imaging protocols could help keep up with the vastly increasing demand for MRI and imaging in general. Deep learning is currently studied as an acceleration method using MRI k-space reduction [6–9]. Instead, this paper focuses on automatically identifying

Abbreviations: ADC, apparent diffusion coefficient; AUROC, area under the receiver operating curve; bpMRI, biparametric prostate MRI; DWI, diffusion-weighted imaging; CI, confidence interval; csPCa, clinically significant prostate cancer; pAUFROC, partial area under the free-response operating characteristic curve; PI-RADS, Prostate Imaging Reporting and Data System; PSA, prostate specific antigen; SD, standard deviations.

[*] Corresponding author.
E-mail address: S.j.fransen@umcg.nl (S.J. Fransen).

https://doi.org/10.1016/j.ejrad.2024.111470
Received 5 January 2024; Received in revised form 29 March 2024; Accepted 14 April 2024
Available online 16 April 2024

redundant sequences for clinical tasks using diagnostic deep learning, which could achieve an even higher acceleration than k-space reduction and can also be applied to other imaging modalities.

The current protocol for prostate MRI is based on the Prostate Imaging Reporting and Data System (PI-RADS) [10]. PI-RADS recommends that when using biparametric prostate MRI (bpMRI), a T2-weighted image, a computed high b-value diffusion-weighted imaging (DWI), and a derived apparent diffusion coefficient (ADC) map should be acquired. The T2 image is essential for morphology assessment, including lesion encapsulation [9]. High b-value DWI is recommended to improve the conspicuity and detect subtle, clinically significant cancers [11–13]. ADC map values have been reported to correlate inversely with histologic grades [14]. However, there is disagreement in the literature about how many and which DWI b-values to acquire for calculating high b-value DWI and ADC maps [9,15–19]. Solving this disagreement through reader studies is not feasible because scoring a sufficiently large sample of scans in different protocols is highly time-consuming. This problem is exacerbated by the high reader variability [4,5,20–22]. In contrast, deep learning can easily handle larger study sets and is more reproducible.

This study explores the use of deep learning to optimize the prostate MRI protocol by objectively assessing the diagnostic efficacy of MRI sequences. We developed a novel framework to compare the diagnostic performances of MRI protocols with deep learning. This framework was applied to prostate MRI to investigate an optimized detection protocol for PI-RADS 4 and 5 lesions. The study is a proof of concept for using AI as an independent expert reader in tasks other than assisting in diagnosis.

2. Materials and methods

2.1. Data

Adult patients who underwent prostate MRI scans at a Medical Center in 2016 were eligible for this study. The institutional ethics committee approved the retrospective scientific use of the data and waived the need for informed consent. Included patients had either elevated prostate-specific antigen (PSA) levels or other sources of clinical suspicion of prostate cancer, e.g., family history. Patients were excluded if they had a treatment history, poor scan quality, or incomplete examinations. The study included a cohort of 840 scans of csPCa suspicious patients. All patients underwent bpMRI or multi-parametric MRI on a 3 T MRI system (Skyra or Prisma, Siemens Healthineers, Erlangen, Germany), and imaging protocols followed PI-RADS v2 guidelines [10]. Technical specifications of the T2 weighted sequences were: 2D FSE with spatial saturation, TE 102 (97 – 104) ms, TR 5669 (2020 – 9220) ms, slice thickness 3.6 ± 0.02 mm, in-plane resolution 0.55 ± 0.05 mm, image dimensions 390x390x23, and 2.5 ± 0.9 averages. Technical specifications for the DWI sequence were: 2D EPI with spatial saturation and phase-sensitive fat suppression, TE 64 (46 – 77), TR 3613 (3000 – 6700) ms, slice thickness 3.6 ± 0.02 mm, in-plane resolution 2 mm, image dimensions 128x120x23, and 4.2 ± 3.7, 8.3 ± 4.8, and 12.4 ± 7.2 averages for respectively b50, b400, and b800 images. All cases were read as part of the clinical routine and evaluated by at least one of six experienced radiologists (4–25 years of experience with prostate MRI). To address possible variability in the labels, multidisciplinary team meetings with urologists and technicians were available to aid lesion characterization and risk stratification. Lesions were delineated by trained investigators under the supervision of an experienced radiologist (7 years of experience with prostate MRI). Detected lesions were scored according to PI-RADS guidelines. In line with the negative predictive value of PI-RADS, we considered all lesions with PI-RADS ≥ 4 as csPCa [23]. Patient characteristics for the patient cohort are presented in Table 1.

Table 1
Characteristics of patients and PI-RADS assessment categories. The first category gives the median age in years with an interquartile range (IQR). The second category gives the median PSA levels in μg/l with IQR. The third category gives the number of patients with a certain PI-RADS as the highest lesion (e.g., 77 patients have PI-RADS 3 as the highest lesion). The fourth category gives the number of lesions in the dataset for each PI-RADS score (e.g., within all patients, there are 134 PI-RADS 3 lesions). The last category gives insights into the position of lesions (e.g., 62% of all reported lesions are located in the peripheral zone).

		Dataset N = 840
Age (years)	All patients	66 (IQR 9)
	Non-csPCa patients	65 (IQR 9)
	csPCa patients	67 (IQR 9)
PSA (μg/l)	All patients	8.0 (IQR 6.0)
	Non-csPCa patients	8.0 (IQR 6.0)
	csPCa patients	9.0 (IQR 8.0)
Highest PI-RADS score (patients)	1	118
	2	360
	3	77
	4	143
	5	142
Number of PI-RADS lesions (lesions)	Total	1291
	1	36
	2	676
	3	134
	4	262
	5	183
Lesion position (percentages of total lesions)	Peripheral zone	62
	Transitional zone	26
	Central zone	7
	Anterior fibromuscular stroma	5

2.2. AI system

A deep learning algorithm based on an open-source published AI algorithm was trained to detect and outline PI-RADS 4 and 5 lesions on MRI scans on different inputs [3]. The proven performance in thorough tests and the public availability were the essence of choosing this deep learning algorithm [3]. The algorithm uses a 3D U-Net architecture, channel-wise squeeze-and-excitation modules, residual connections between consecutive convolutional blocks, a binary cross-entropy loss function, and a LeakyReLU activation function [24,25]. Pre-processing involved resampling to uniform voxel spacing of 0.5 x 0.5 mm^2 with 3.6 mm slice thickness using linear interpolation, center cropping all scans to 96 x 96 x 86.4 mm, and z-score-normalization. Rotation (maximum 30-degree angle) and noise (maximum 0.1% multiplication from a uniform distribution) were applied as data augmentations to teach the network the desired invariance and robustness properties. All models finished within 48 h on a 32 GB Tesla V100 GPU Nvidia with an average of 1157 ± 211 epochs. An Adam optimizer with an initial learning rate of 0.0001, a batch size of 12 examinations, and early stopping after 50 epochs were used. Convergence was checked on the validation loss, and the model with the best performance was selected. This algorithm configuration has been extensively tested and demonstrated to outperform similar models and approach expert performance on PI-RADS 4 and 5 lesion detection [3,26,27]. The algorithm was retrained using three inputs: an axial T2-weighted scan, an ADC map, and a b1400 scan. The ADC map and b1400 scan were calculated using different (simulated) b-value DWI protocols for efficacy comparison. Unseen test cases were used to evaluate the performance. In five folds, the data was split into 672 (80%) training cases, 84 (10%) validation cases, and 84 (10%) test cases. A total of 420 individual test cases unseen to AI training were used in the performance assessment. The output of the trained AI model was a heatmap in which each voxel value represents the voxel level likelihood of PI-RADS 4 and 5 presence. Postprocessing and analysis of this heatmap were performed in line with the PI-CAI challenge, which provides an evaluation pipeline for csPCa detection AI [28]. First, the heatmap was condensed to a list of up to five PI-RADS 4 and 5 lesion candidates located automatically by iteratively selecting the connected components containing the highest prediction values [29]. The lesions provide more insight into the AI system decision and enable evaluation at lesion-level performance. Next, a lesion was considered a true positive with a minimum overlap of 10% (Dice score) with the reference segmentation [3]. False positives were defined as predictions with no or insufficient overlap.

2.3. Accelerated protocols

In line with the current discrepancy in the literature about how many and which DWI b-value to acquire for calculating high b-value DWI and ADC maps [9,15–19], we simulated three different DWI protocols to calculate the ADC and b1400 map with a mono-exponential decay model: 1) a baseline protocol including an axial T2-weighted sequence and b-values sequences b50, b400, b800 (6:19 min); 2) an accelerated protocol omitting the DWI b800 sequence (2.57 min); and 3) an accelerated protocol omitting the DWI b400 sequence (4:17 min).

2.4. Comparing protocols

The diagnostic performance of each of the models was assessed on unseen data through 5-fold cross-validation. The diagnostic performances of accelerated bpMRI protocols were compared to the full baseline protocol based on the predicted AI diagnostic performance. The quantitative analysis compared the area under the receiver operating curves (AUROC) and the partial area under the free-response receiver operating characteristic curves (pAUFROC). The AUROC evaluates the diagnostic performance at the patient-level. The statistical comparison between different AUROC was made using DeLong's test [30]. The pAUFROC assesses the diagnostic value of each protocol at the lesion-level. A statistical comparison between different pAUFROC was made using the permutation test between 0 and 1 false positive lesions. In both analyses, the 95% confidence interval (CI) was calculated using the standard deviations (SD) across cross-validation folds. P-values below 0.05 were considered statistically significant. Statistical analysis was performed in R version 4.2.0 using the pROC package [31].

2.5. AI validation

A saliency analysis and radiologist's reader study were performed to validate the findings of our artificial expert reader. The saliency analysis compared the saliency maps of the different MRI protocols. The saliency maps were generated using the method described by Sundararajan et al. [32]. Saliency maps provide an explainable AI strategy that reveals the underlying patterns in the input images that contribute to the final predictions of the AI system [32]. Saliency mapping highlights the regions of the input images that influence the AI system's output the most. A high value in the saliency maps indicates an important region for the final AI model prediction. This information provides valuable insights into the interpretability of the AI system's predictions and can make the predictions more trustworthy (e.g., regions with high values are of more importance for the prediction). Here, the saliency analysis was performed to assess the importance of each input layer for the final prediction. In addition, a radiologist's reader study was performed. Validation of diagnostic accuracy would require a large reader study, which is highly time-consuming. This reader study used 30 randomly selected patients from the included cohort to validate the visual appearance. Four experienced radiologists (6, 10, 13, and 16 years of experience in prostate MRI) were presented with two sets of bpMRI scans, one set calculated with the baseline protocol and one set calculated with the accelerated protocol omitting the DWI b800 sequence. A forced-choice test was used to determine if the radiologists could identify the baseline protocol [33]. In case of low recognition scores below 50%, the radiologists do not outperform random guessing. Additionally, Cohen's Kappa was calculated to determine the agreement among readers' choice of preferred protocol [34].

3. Results

The protocol comparison with AI showed non-significant diagnostic performance differences between the accelerated and the full protocol, as shown in Figures 1 and 2.

At the patient-level, the baseline model with full bpMRI protocol reached an AUROC of 0.77 ± 0.06 (SD). The AUROC analysis showed a non-significant performance difference when omitting the b800 sequence (AUROC 0.78 ± 0.03, p = 0.2) or when omitting the b400 sequence (AUROC 0.76 ± 0.05, p = 0.65).

At the lesion-level, the sensitivity difference compared to the full protocol (pAUFROC = 0.5 ± 0.03) was non-significant when omitting the b800 sequence (pAUFROC = 0.47 ± 0.07, p = 0.45) or omitting the b400 sequence (pAUFROC = 0.51 ± 0.05, p = 0.73).

The saliency maps of the three bpMRI protocols did not reveal any differences. In each protocol, the high b-value image contained the most important information for the prediction. An accelerated DWI protocol does not affect the relative importance of high b-value in the bpMRI protocol for prostate cancer detection. Figures 3 and 4 show representative examples of the saliency maps.

The confirmatory reader experiment results were in line with the findings of the artificial reader. Expert radiologists could not distinguish the baseline from accelerated protocol ADC and High B-value images, underlining the redundance of the DWI b800 sequence. The probability of correct identification was 37%, 53%, 43%, and 67%. The corrected forced-choice recognition scores were below 50%. Additionally, Cohen's Kappa analysis showed strong disagreement (k = 0.09), providing further evidence for protocol similarities.

4. Discussion

This study shows that deep learning can optimize the prostate MRI

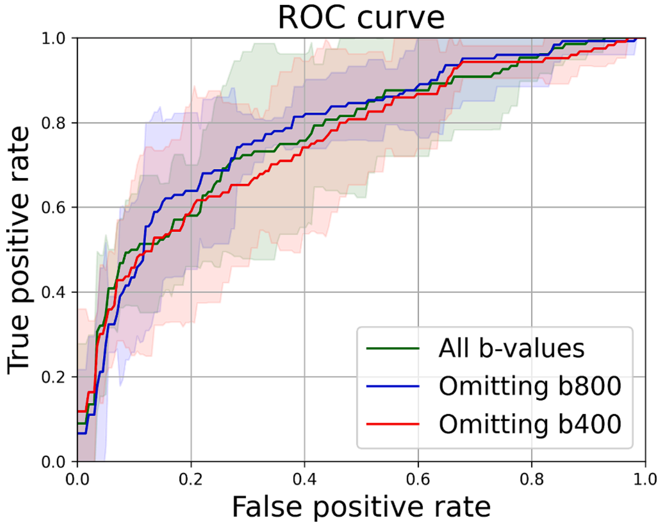

Fig. 1. Diagnostic accuracy for the detection of PI-RADS 4 and 5 lesions at patient-level for the baseline and accelerated protocols. The AUROC plots the false positive rate against the true positive rate. The derived bpMRI is based on DWI b50-b400-b800, b50-b800, and b50-b400 for the green, blue, and red curves, respectively. The shaded areas indicate 95% CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

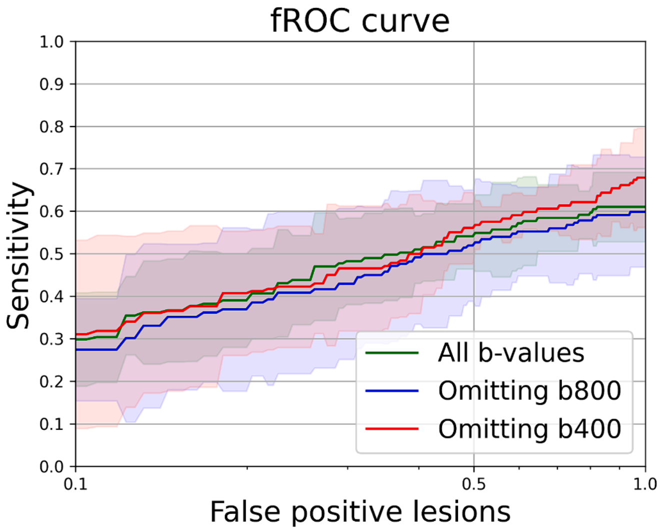

Fig. 2. Diagnostic accuracy for the detection of PI-RADS 4 and 5 lesions at lesion-level for the baseline and accelerated protocols. The pAUFROC plots the false positive lesions per patient on a logarithmic scale against the lesion detection sensitivity. The derived bpMRI is based on DWI b50-b400-b800, b50-b800, and b50-b400 for the green, blue, and red curves, respectively. The shaded areas indicate 95% CIs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

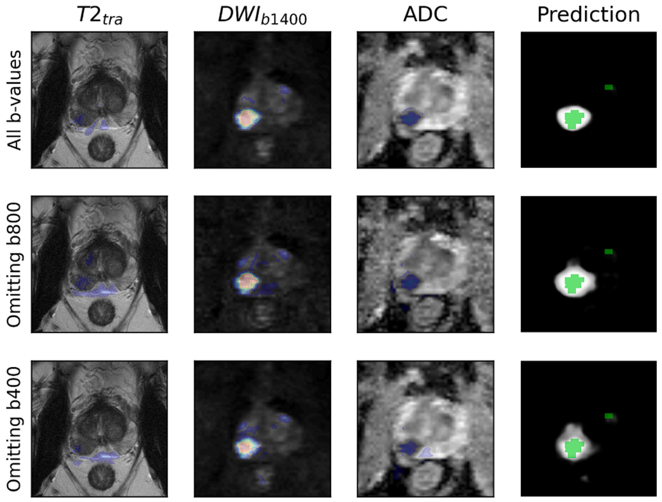

Fig. 3. Very similar AI saliency for the three DWI protocols in a 63-year-old patient with PSA 9 μg/l and a PI-RADS 5 and PI-RADS 4 lesion in the peripheral zone. The DWI and ADC were calculated with three combinations: all b-values (b50, b400, and b800), omitting b800, and omitting b400. The predicted patient-level PI-RADS 4 and 5 lesion likelihood scores (range 0 to 1) were 0.99, 0.98, and 0.97, respectively. The blue and red regions in the input images indicate areas of minor and major importance for the final prediction. The green area in the prediction is the reference segmentation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

protocol by objectively assessing the diagnostic efficacy of MRI sequences on a large data set. The results show that PI-RADS 4 and 5 lesion detection can be performed with fewer b-values without compromising diagnostic quality for the AI reader. The image similarity radiologist reader experiment provided further evidence for diagnostic similarities. Differences in diagnostic performance on patient and lesion levels when omitting the b400 or b800 sequence from the DWI protocol were not statistically significant while providing a decrease in DWI scan time. This time reduction could reduce the patient burden of a long scan time, allow more patients to be scanned, and reduce healthcare costs. The findings indicate that AI employed as an expert reader can be used for more than diagnosis and can objectively assess the diagnostic relevance of MRI sequences.

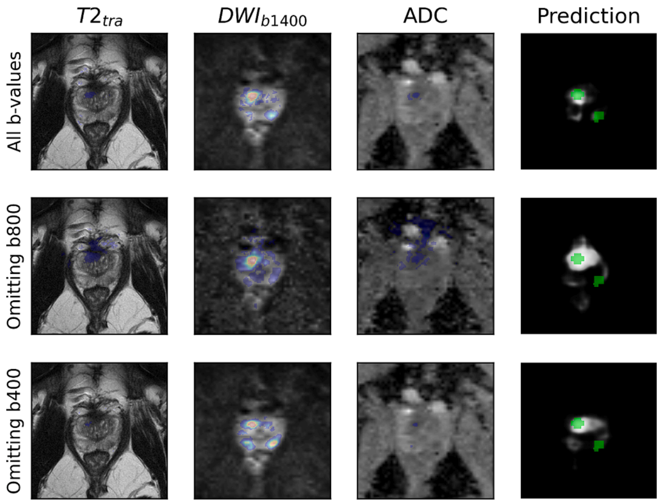

Fig. 4. Slight variations in AI saliency with different DWI protocols in a 69-year-old patient with PSA 5 μg/l and two PI-RADS 4 lesions in the peripheral zone and transition zone. The DWI and ADC were calculated with all b-values (b50, b400, and b800), omitting b800, and omitting b400. The predicted patient-level PI-RADS 4 and 5 lesion likelihood scores (range 0 to 1) were 0.73, 0.97, and 0.85, respectively. The blue and red regions in the input images indicate areas of minor and major importance for the final prediction. The green area in the prediction is the reference segmentation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The findings of this study shed new light on the discussion of optimal b-values by focusing on reducing scan time and using the largest cohort to date for optimal b-value selection. In agreement with the current literature, our method identified two instead of three b-values sufficient for PI-RADS 4 and 5 lesion detection [9,15–19]. In contrast to earlier findings, no evidence was found that agrees with PI-RADS v2 to use b800 instead of b400 [9]. A lower b-value has the benefit of a faster acquisition, a higher signal-to-noise ratio, and fewer artifacts [35]. Adapting the PI-RADS v2 guidelines to recommend b400 instead of b800 would decrease the scan time by 31%.

Healthcare, in general, and MRI, in particular, is under tremendous pressure to improve its workflow without compromising diagnostic accuracy. As shown, deep learning can help to accelerate MRI by identifying redundant MRI sequences. Other technological improvements, such as compressed sensing and AI-based image reconstruction, have also shown the potential to accelerate MRI [36]. However, these methods can introduce misleading artifacts, are often only applicable to T2-weighted images, and the effect on diagnostic quality remains underexplored [35–37]. Our proposed framework applies to all sequences and can accelerate any MRI protocol with a clearly defined diagnostic task without affecting diagnostic quality.

Our study had limitations. The cohort of 840 patients poses a limit to the detectable accuracy differences. A larger dataset could have increased the accuracy of the AI models, possibly finding statistically significant differences in model performance between acquisition protocols. However, we expect the differences will likely be equal or smaller than we observed, which renders them clinically irrelevant. Additionally, our data comes from a single center with a single vendor, which may limit the generalizability of our results. Furthermore, the AI performance depends on the conditions for marking detected lesions as true positive or false negative. Also, the interreader variability in radiologists' assessment might have introduced a label variability and affected the AI performance. A prospective study is needed to study performance across different populations or imaging settings and the effect of performance on PI-RADS 3 diagnosis. Moreover, AI as an artificial reader to identify redundant protocol steps should be tested on protocols of other disciplines to show generalizability. At last, our radiologists' validation did not focus on diagnostic accuracy, which would require a highly time-consuming reader study. Our reader study does show high similarities in visual appearance between the protocols, indicating the presence of redundant MRI sequences and supporting the findings of the artificial reader. It would be interesting to look into the lesion-level segmentation quality of an artificial reader in a future study. We also recommend a multicenter study to validate the robustness of these results for generalizability to multiple centers.

In conclusion, this study shows that deep learning can assess an MRI protocol's efficacy. Omitting the b400 or b800 DWI sequence can reduce DWI scan time without compromising diagnostic quality. Adapting the PI-RADS v2 guidelines to recommend b400 instead of b800 would decrease the scan time by 31%. The outcome of this study shows the utility of deep learning for scan time reduction by assessing the diagnostic efficacy of MRI sequences.

CRediT authorship contribution statement

Stefan J. Fransen: . **Christian Roest:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Quintin Y. Van Lohuizen:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Joeran S. Bosma:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Frank F.J. Simonis:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Thomas C. Kwee:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Derya Yakar:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Henkjan Huisman:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Siemens Healthineers and Health-Holland (grant number: LSHM20103).

References

[1] K.H. Yu, A.L. Beam, I.S. Kohane, Artificial intelligence in healthcare, Nat Biomed Eng 2 (2018) 719–731, https://doi.org/10.1038/s41551-018-0305-z.

[2] R.M. Kronberg, D. Meskelevicius, M. Sabel, M. Kollmann, C. Rubbert, I. Fischer, Optimal acquisition sequence for AI-assisted brain tumor segmentation under the constraint of largest information gain per additional MRI sequence, Neuroscience Informatics 2 (2022) 100053, https://doi.org/10.1016/j.neuri.2022.100053.

[3] A. Saha, M. Hosseinzadeh, H. Huisman, 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 73 (2021), https://doi.org/10.1016/j.media.2021.102155.

[4] H.U. Ahmed, A. El-Shater Bosaily, L.C. Brown, R. Gabe, R. Kaplan, M.K. Parmar, Y. Collaco-Moraes, K. Ward, R.G. Hindley, A. Freeman, A.P. Kirkham, R. Oldroyd, C. Parker, M. Emberton, Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study, The Lancet 389 (2017) 815–822, https://doi.org/10.1016/S0140-6736(16)32401-1.

[5] V. Kasivisvanathan, A.S. Rannikko, M. Borghi, V. Panebianco, L.A. Mynderse, M. H. Vaarala, A. Briganti, L. Budäus, G. Hellawell, R.G. Hindley, M.J. Roobol, S. Eggener, M. Ghei, A. Villers, F. Bladou, G.M. Villeirs, J. Virdi, S. Boxler, G. Robert, P.B. Singh, W. Venderink, B.A. Hadaschik, A. Ruffion, J.C. Hu, D. Margolis, S. Crouzet, L. Klotz, S.S. Taneja, P. Pinto, I. Gill, C. Allen, F. Giganti, A. Freeman, S. Morris, S. Punwani, N.R. Williams, C. Brew-Graves, J. Deeks, Y. Takwoingi, M. Emberton, C.M. Moore, MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis, New England Journal of Medicine (2018), https://doi.org/10.1056/nejmoa1801993.

[6] A. Sriram, J. Zbontar, T. Murrell, C.L. Zitnick, A. Defazio, D.K. Sodickson, GrappaNet: Combining Parallel Imaging with Deep Learning for Multi-Coil MRI Reconstruction, (2019). http://arxiv.org/abs/1910.12325.

[7] M. Uecker, P. Lai, M.J. Murphy, P. Virtue, M. Elad, J.M. Pauly, S.S. Vasanawala, M. Lustig, ESPIRiT - An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA, Magn Reson Med 71 (2014) 990–1001, https://doi.org/10.1002/mrm.24751.

[8] T. Eo, Y. Jun, T. Kim, J. Jang, H.J. Lee, D. Hwang, KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images, Magn Reson Med 80 (2018) 2188–2201, https://doi.org/10.1002/mrm.27201.

[9] N. Mir, S.J. Fransen, J.M. Wolterink, J.J. Fütterer, F.F.J. Simonis, Recent Developments in Speeding up Prostate MRI, Journal of Magnetic Resonance Imaging (2023), https://doi.org/10.1002/jmri.29108.

[10] B. Turkbey, A.B. Rosenkrantz, M.A. Haider, A.R. Padhani, G. Villeirs, K.J. Macura, C.M. Tempany, P.L. Choyke, F. Cornud, D.J. Margolis, H.C. Thoeny, S. Verma, J. Barentsz, J.C. Weinreb, P.I. Reporting, D.S. Version, 2.1,, Update of Prostate Imaging Reporting and Data System Version 2, Eur Urol 76 (2019) (2019) 340–351, https://doi.org/10.1016/j.eururo.2019.02.033.

[11] L.K. Bittencourt, J.O. Barentsz, L.C.D. De Miranda, E.L. Gasparetto, Prostate MRI: Diffusion-weighted imaging at 1.5T correlates better with prostatectomy Gleason grades than TRUS-guided biopsies in peripheral zone tumours, Eur Radiol 22 (2012) 468–475, https://doi.org/10.1007/s00330-011-2269-1.

[12] T. Metens, D. Miranda, J. Absil, C. Matos, What is the optimal b value in diffusion-weighted MR imaging to depict prostate cancer at 3T? Eur Radiol 22 (2012) 703–709, https://doi.org/10.1007/s00330-011-2298-9.

[13] T. Tamada, N. Kanomata, T. Sone, Y. Jo, Y. Miyaji, H. Higashi, A. Yamamoto, K. Ito, High b Value (2,000 s/mm2) Diffusion-WeightedMagnetic Resonance Imaging in Prostate Cancer at 3Tesla: Comparison with 1,000 s/mm2for TumorConspicuity and Discrimination of Aggressiveness, PLoS One 9 (2014), https://doi.org/10.1371/journal.pone.0096619.

[14] X. Wu, P. Reinikainen, A. Vanhanen, M. Kapanen, T. Vierikko, P. Ryymin, S. Hyödynmaa, P.-L. Kollokumpu-Lehtinen, Correlation between apparent diffusion coefficient value on diffusion-weighted MR imaging and Gleason score in prostate cancer, Diagn Interv, Imaging 98 (2017) 261–268, https://doi.org/10.1016/j.diii.2016.08.009.

[15] G. Thörmer, J. Otto, M. Reiss-Zimmermann, M. Seiwerts, M. Moche, N. Garnov, T. Franz, M. Do, J.U. Stolzenburg, L.C. Horn, T. Kahn, H. Busse, Diagnostic value of ADC in patients with prostate cancer: Influence of the choice of b values, Eur Radiol 22 (2012) 1820–1828, https://doi.org/10.1007/s00330-012-2432-3.

[16] T. de Perrot, M. Scheffler, J. Boto, B.M.A. Delattre, C. Combescure, M. Pusztaszeri, J.C. Tille, C. Iselin, J.P. Vallée, Diffusion in prostate cancer detection on a 3T scanner: How many b-values are needed? Journal of Magnetic Resonance Imaging 44 (2016) 601–609, https://doi.org/10.1002/jmri.25206.

[17] N. Adubeiro, M.L. Nogueira, R.G. Nunes, H.A. Ferreira, E. Ribeiro, J.M.F. la Fuente, Apparent diffusion coefficient in the analysis of prostate cancer: determination of optimal b-value pair to differentiate normal from malignant tissue, Clin Imaging 47 (2018) 90–95, https://doi.org/10.1016/j.clinimag.2017.09.004.

[18] S.E. Maier, J. Wallström, F. Langkilde, J. Johansson, S. Kuczera, J. Hugosson, M. Hellström, Prostate Cancer Diffusion-Weighted Magnetic Resonance Imaging: Does the Choice of Diffusion-Weighting Level Matter? Journal of Magnetic Resonance Imaging 55 (2022) 842–853, https://doi.org/10.1002/jmri.27895.

[19] H.J. Rogers, S. Singh, A. Barnes, N.A. Obuchowski, D.J. Margolis, D.I. Malyarenko, T.L. Chenevert, A. Shukla-Dave, M.A. Boss, S. Punwani, Test-retest repeatability of ADC in prostate using the multi b-Value VERDICT acquisition, Eur J Radiol 162 (2023), https://doi.org/10.1016/j.ejrad.2023.110782.

[20] C.P. Smith, S.A. Harmon, T. Barrett, L.K. Bittencourt, Y.M. Law, H. Shebel, J.Y. An, M. Czarniecki, S. Mehralivand, M. Coskun, B.J. Wood, P.A. Pinto, J.H. Shih, P. L. Choyke, B. Turkbey, Intra- and interreader reproducibility of PI-RADSv2: A multireader study, Journal of Magnetic Resonance Imaging 49 (2019) 1694–1703, https://doi.org/10.1002/jmri.26555.

[21] J.N. Thai, H.A. Narayanan, A.K. George, M.M. Siddiqui, P. Shah, F.V. Mertan, M. J. Merino, P.A. Pinto, P.L. Choyke, B.J. Wood, B. Turkbey, Validation of PI-RADS version 2 in transition zone lesions for the detection of prostate cancer, Radiology 288 (2018) 485–491, https://doi.org/10.1148/radiol.2018170425.

[22] A.B. Rosenkrantz, L.A. Ginocchio, D. Cornfeld, A.T. Froemming, R.T. Gupta, B. Turkbey, A.C. Westphalen, J.S. Babb, D.J. Margolis, Interobserver reproducibility of the PI-RADS version 2 lexicon: A multicenter study of six experienced prostate radiologists, Radiology 280 (2016) 793–804, https://doi.org/10.1148/radiol.2016152542.

[23] R. Itatani, T. Namimoto, S. Atsuji, K. Katahira, S. Morishita, K. Kitani, Y. Hamada, M. Kitaoka, T. Nakaura, Y. Yamashita, Negative predictive value of multiparametric MRI for prostate cancer detection: outcome of 5-year follow-up in men with negative findings on initial MRI studies, Eur J Radiol 83 (2014) 1740–1745, https://doi.org/10.1016/j.ejrad.2014.06.026.

[24] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2015: pp. 234–241. Doi: 10.1007/978-3-319-24574-4_28.

[25] J. Hu, L. Shen, G. Sun, Squeeze-and-Excitation Networks, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141, https://doi.org/10.48550/arXiv.1709.01507.

[26] C. Roest, T.C. Kwee, A. Saha, J.J. Fütterer, D. Yakar, H. Huisman, AI-assisted biparametric MRI surveillance of prostate cancer: feasibility study, Eur Radiol (2022) 89–96, https://doi.org/10.1007/s00330-022-09032-7.

[27] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M. S. Nobile, C. Ferretti, D. Besozzi, M.C. Gilardi, S. Vitabile, G. Mauri, H. Nakayama, P. Cazzaniga, USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets, Neurocomputing 365 (2019) 31–43, https://doi.org/10.1016/j.neucom.2019.07.006.

[28] A. Saha, J.S. Bosma, J.J. Twilt, B. Van Ginneken, D. Yakar, D.Y. Nl, M. Elschot, J. Veltman, J. Fütterer, H. Huisman, H.H. Nl, Medical Imaging with Deep Learning-Under Review 2023 Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI-The PI-CAI Challenge Maarten de Rooij, 2023. https://pi-cai.grand-challenge.org/.

[29] J.S. Bosma, A. Saha, M. Hosseinzadeh, I. Slootweg, M. de Rooij, H. Huisman, Semisupervised Learning with Report-guided Pseudo Labels for Deep Learning–based Prostate Cancer Detection Using Biparametric MRI, Radiol Artif Intell 5 (2023), https://doi.org/10.1148/ryai.230031.

[30] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach, Biometrics 44 (1988) 837, https://doi.org/10.2307/2531595.

[31] N. Turck, L. Vutskits, P. Sanchez-Pena, X. Robin, A. Hainard, M. Gex-Fabry, C. Fouda, H. Bassem, M. Mueller, F. Lisacek, L. Puybasset, J.-C. Sanchez, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinformatics 8 (2011) 12–77. http://link.springer.com/10.1007/s00134-009-1641-y.

[32] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, 34th International Conference on Machine Learning, ICML 2017 7 (2017) 5109–5118. Doi: Doi: 10.48550/arXiv.1703.01365.

[33] S.N. Singh, C.A. Cole, Forced-Choice Recognition Tests: A Critical Review, Source, Journal of Advertising 14 (1985) 52–58, https://doi.org/10.1080/00913367.1985.10672958.

[34] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem Med (zagreb) 22 (2012) 276–282, https://doi.org/10.11613/BM.2012.031.

[35] D. le Bihan, C. Poupon, A. Amadon, F. Lethimonnier, Artifacts and pitfalls in diffusion MRI, Journal of Magnetic Resonance Imaging 24 (2006) 478–488, https://doi.org/10.1002/jmri.20683.

[36] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M.J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K.J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C.L. Zitnick, M.P. Recht, D. K. Sodickson, Y.W. Lui, fastMRI: An Open Dataset and Benchmarks for Accelerated MRI, ArXiv Preprint (2018), https://doi.org/10.48550/arXiv.1811.08839.

[37] V. Antun, F. Renna, C. Poon, B. Adcock, A.C. Hansen, On instabilities of deep learning in image reconstruction - Does AI come at a cost? ArXiv Preprint (2019) https://doi.org/10.1073/pnas.1907377117.