BMC Cancer

RESEARCH

Open Access

Deep learning network enhances imaging quality of low-*b*-value diffusion–weighted imaging and improves lesion detection in prostate cancer

Zheng Liu[1,2,3†], Wei-jie Gu[1,2,3†], Fang-ning Wan[1,2,3], Zhang-zhe Chen[2,4], Yun-yi Kong[2,5], Xiao-hang Liu[2,4*], Ding-wei Ye[1,2,3*] and Bo Dai[1,2,3*]

Abstract

Background Diffusion-weighted imaging with higher *b*-value improves detection rate for prostate cancer lesions. However, obtaining high *b*-value DWI requires more advanced hardware and software configuration. Here we use a novel deep learning network, NAFNet, to generate a deep learning reconstructed (DLR_{1500}) images from 800 *b*-value to mimic 1500 *b*-value images, and to evaluate its performance and lesion detection improvements based on whole-slide images (WSI).

Methods We enrolled 303 prostate cancer patients with both 800 and 1500 *b*-values from Fudan University Shanghai Cancer Centre between 2017 and 2020. We assigned these patients to the training and validation set in a 2:1 ratio. The testing set included 36 prostate cancer patients from an independent institute who had only preoperative DWI at 800 *b*-value. Two senior radiology doctors and two junior radiology doctors read and delineated cancer lesions on DLR_{1500}, original 800 and 1500 *b*-values DWI images. WSI were used as the ground truth to assess the lesion detection improvement of DLR_{1500} images in the testing set.

Results After training and generating, within junior radiology doctors, the diagnostic AUC based on DLR_{1500} images is not inferior to that based on 1500 *b*-value images (0.832 (0.788–0.876) vs. 0.821 (0.747–0.899), $P=0.824$). The same phenomenon is also observed in senior radiology doctors. Furthermore, in the testing set, DLR_{1500} images could significantly enhance junior radiology doctors' diagnostic performance than 800 *b*-value images (0.848 (0.758–0.938) vs. 0.752 (0.661–0.843), $P=0.043$).

†Zheng Liu and Wei-jie Gu contributed equally to this work.

*Correspondence:
Xiao-hang Liu
09111230002@fudan.edu.cn
Ding-wei Ye
dwyeli@163.com
Bo Dai
bodai1978@126.com

Full list of author information is available at the end of the article

Conclusions DLR_{1500} DWIs were comparable in quality to original 1500 *b*-value images within both junior and senior radiology doctors. NAFNet based DWI enhancement can significantly improve the image quality of 800 *b*-value DWI, and therefore promote the accuracy of prostate cancer lesion detection for junior radiology doctors.

Keywords Deep learning, Prostate Cancer, Diffusion–weighted imaging, High *b*-Value, Whole-slide imaging.

Background

According to the latest data, the incidence and mortality of prostate cancer are increasing instantly [1]. In the United States, prostate cancer ranks first in the new incidence of men malignant tumors and second in mortality [2]. Multiparametric MRI is widely used in prostate cancer detection and staging [3, 4]. At present, diffusion-weighted imaging (DWI) plays a vital role in prostate cancer diagnosis [5]. Nonetheless, the detection of clinically significant prostate cancer lesions may prove challenging, and high *b*-value DWI imaging can enhance their detection rate [6].

However, not every medical institution has the capability to utilise high *b*-value detection techniques. In clinical practice, obtaining high *b*-value DWI is technically challenging. The obtained DWI often experiences disruptions from low in-plane spatial resolution, distortion, and artifacts. Furthermore, increasing the *b*-value usually means more advanced hardware and software configuration and requires a longer acquisition time [7]. For these reasons, a possible solution is to introduce deep learning reconstruction (DLR) to improve the quality of MRI scans [8–11]. Relevant research has shown the feasibility of deep learning reconstruction in improving image quality and diagnostic performance, including for MR angiography of coronary arteries, spine, brain and other organs [12–14]. Hence, we hypothesize that the utilization of DLR may elevate the quality of prostate DWI scans, thereby improving the diagnostic accuracy in detecting prostate cancer.

At present, deep learning networks have been successfully applied to analyze urological medical images in the fields of cystoscopy, ultrasound, pathology, and radiology [15, 16]. The reconstruction and enhancement of MRI images rely on the in-depth understanding of each pixel within the MRI image. Given that the lesion area usually constitutes a minor portion of the total image, it is imperative that these pixels are thoroughly processed by a well-designed deep learning model to facilitate precision in diagnoses. Here, we introduce NAFNet, a nonlinear activation-free network that simplifies the architecture by replacing complex nonlinear activation functions. Even with this simplification, NAFNet maintains state-of-the-art results while reducing computational costs [17]. Furthermore, NAFNet's capacity for pixel-level feature partitioning enables extraction and identification of subtle differences within MRI imaging. Our previous research has also proven that NAFNet exceeds performance of other prominent networks, such as ResNet, in handling prostate MRI images [18].

Moreover, the Prostate Imaging Reporting and Data System (PI-RADS) version 2.1 recommends the utilization of high *b*-values (greater than 1400) to obtain or compute DWI, and 800–1000 for ADC, enhance the visibility of clinically significant prostate cancers [5]. Therefore, ADC with 800 *b*-value has already met the requirements of PI-RADS v2.1, hence this study is confined to the reconstruction of 1500 *b*-value DWI, while the computation of higher *b*-value ADC is excluded.

Although NAFNet algorithms has already been proposed, this study pioneers the clinical translation of deep learning in prostate MRI through three innovations. Firstly, we establish the implementation of NAFNet architecture for synthesizing high *b*-value (1500 s/mm²) DWI from low *b*-value (800 s/mm²) acquisitions in prostate cancer diagnostics, trying to eliminating hardware dependencies while maintaining diagnostic-grade image accuracy. Second, we utilize prostate cancer lesions identified on WSI as the gold standard to perform lesion annotation on MRI images, establishing our model ground truth. This approach provides superior biological relevance compared to radiologist-defined ground truth commonly used in similar studies. Finally, we conducted validation on a multi-centre external testing set, which proved the robustness and reliability of our algorithm.

Therefore, this study aims to enhance the imaging quality of low *b*-value DWI images using the deep learning network, NAFNet, and simulate them as high *b*-value. In addition, with WSI as the gold standard, we also assess the improvement of the diagnostic ability of prostate cancer lesions due to the deep learning generated image quality.

Methods

Patient cohorts

The enrolled patients met the subsequent criteria: (1) undergone radical prostatectomy (RP) after mpMRI; (2) no prior history of neoadjuvant, adjuvant or other additional therapy; (3) availability of high-quality preoperative prostate MRI images, and (4) identifiable lesions present in the images.

We retrospectively collected a total of 339 prostate cancer patients from two clinical centres. As shown in Fig. 1A, after screening, 303 patients diagnosed with prostate cancer at Fudan University Shanghai Cancer Centre were involved. These patients had DWI images

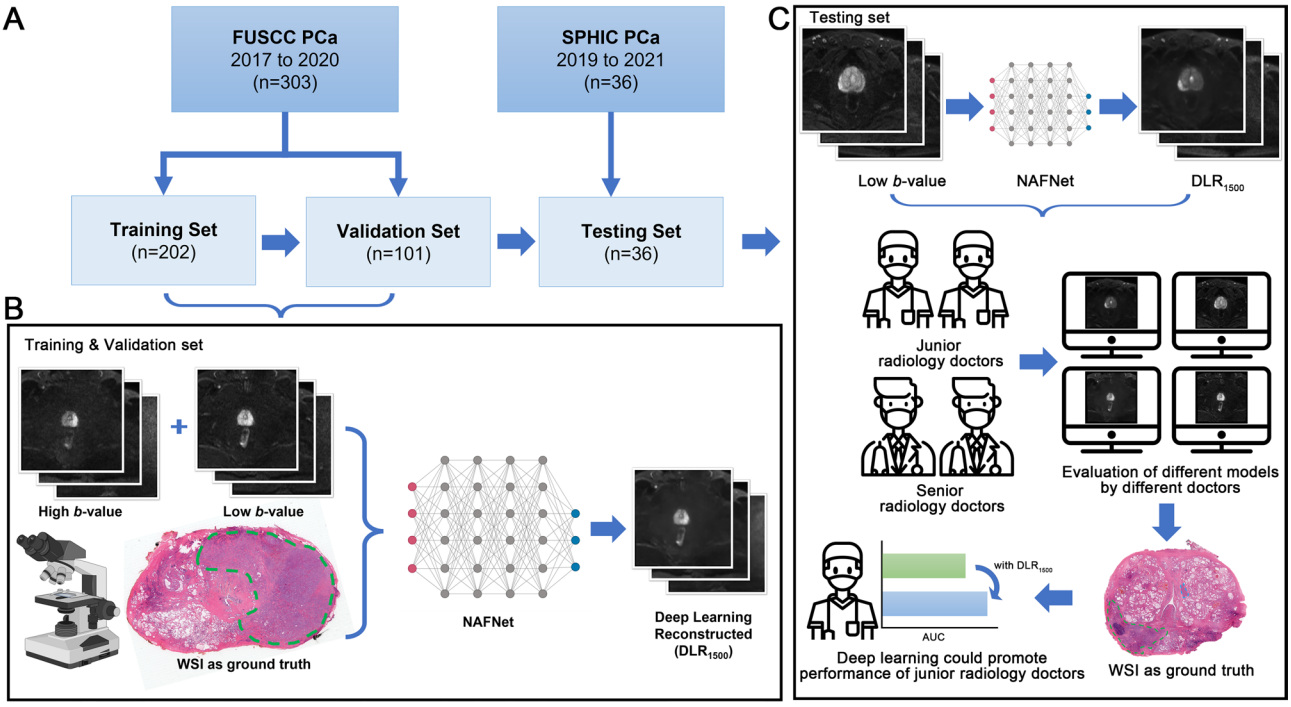

Fig. 1 Overall flowchart and design of this study. (**a**) Flowchart showing the study population of training, validation and testing set in this study. (**b**) Work pipeline of this study illustrating the training and validation of the deep-learning model based on high and low *b*-value DWI images, WSI served as the ground truth. (**c**) After training and validation, we used testing set to generate DLR DWI images based on low *b*-value, and the quality of these images was assessed in comparison with the WSI ground truth by both junior and senior radiology doctors. Abbreviations: DWI, diffusion-weighted imaging; WSI, whole slide images; DLR, deep learning reconstruction

with both 800 *b*-value and 1500 *b*-value, and whole-mount histopathology slide images (WSIs) of RP specimens. The patients were systematically divided into training and validation sets at a ratio of 2:1, thereby balancing different patients' characteristics of the study population. In addition, 36 prostate cancer patients from SPHIC formed an external testing set. These patients did not have high *b*-value DWI images but had preoperative 800 *b*-value images and scanned WSI images.

After training with NAFNet, we incorporated 800 *b*-value DWI images and generated DLR_{1500} DWI images in the validation set, simulating 1500 *b*-value DWI (Fig. 1B). In the testing set, we directly inputted 800 *b*-value DWI images, and produced DLR_{1500} images. Finally, we used the WSI as the gold standard, comparing the diagnostic efficacy of images interpreted by junior and senior radiology doctors. These images included original 800 *b*-value and 1500 *b*-value DWI images, as well as DLR_{1500} images generated by NAFNet. Then, the quality of these three image sets were evaluated (Fig. 1C).

This study protocol adhered to the Declaration of Helsinki and received approval from the Institutional Review Board of Fudan University Shanghai Cancer Centre (Approval No. 050432-4-2108). Each participant provided written consent. The TRIPOD and STARD guidelines were also followed (Supplementary Tables 1 and Supplementary Table 2).

MRI pre-processing and lesion delineation

Prior to the delineation of lesions, patients were subjected to imaging procedures using MRI scanners equipped with phased-array coils. The imaging techniques employed included axial turbo spin-echo T2WI and axial DWI with ADC maps. Detailed parameters for these MRI scanners in this study are listed in Supplementary Table 3.

To ensure the accuracy in building the deep learning model, YY.K., a senior urological pathologist, marked the tumor lesion on WSI images for each patient. Thereafter, these marked lesion on WSI images were used as the ground truth, and XH.L., a senior radiologist, delineated them on the MRI images using ITK-SNAP software for reference in the training set.

At the validation and testing stage, MRI images were subsequently evaluated by two senior radiology doctors (XH.L with 16 years of experience, and ZZ.C with 8 years of experience) and two junior radiology doctors. Using PI-RADS version 2.1 (or version 2.0 in instances where the *b*-value with 800), the images' ROIs were identified on DWI maps. These images expedited the delineation of lesion extent on the MRI images, which were then

used as ROIs. Furthermore, to enable more direct quality assessment of enhanced images, we have introduced the PI-QUAL v2 scoring [19]. This system assists radiologists in rapidly evaluating prostate MRI image quality while allowing separate scoring of DWI sequences. Each doctors assigned PI-QUAL v2 scores to the original high *b*-value DWI images as a baseline reference.

After obtaining the MRI images, we proceeded to normalise them. This process entailed scaling the image values to a range between 0 and 1 by dividing each value by the image's maximum value. Using the 2D images from the DWI map, we then resized them to identical dimensions to facilitate easier model processing. To further diversify our sample set, data augmentation techniques such as flipping and rotating the images were applied. All these steps ensured the improved performance and resilience of our model.

Image enhancing

The NAFNet, a Nonlinear Activation Free Network, is a recently established deep learning network in 2022 [17]. The network has displayed superior results in the image restoration domain, such as denoising, deblurring, and stereo super-resolution. The design of NAFNet involves addressing issues at the pixel level. For instance, NAFNet overcomes pixel accuracy challenges, which typically result from batch normalization, by employing pixel-level layer normalization. This feature enables NAFNet to execute dense predictions or pixelwise predictions, which lead to more precise feature segmentation in the ROI, thereby indicating its potential for MRI enhancement. Our previous study has also demonstrated that NAFNet outperforms existing mainstream algorithms such as ResNet in handling prostate MRI [18]. Therefore, in this study, we continue to use NAFNet for MRI image-enhancement processing.

The adopted image augmentation strategies and optimum configurations are as such: random flipping/rotation were used, and the maps were resized to 384×384 dimensions using bilinear interpolation. Specifically, that data augmentations were dynamically applied during training with horizontal flipping (50% probability), random rotation ($\pm 15°$), and intensity variation ($\pm 10\%$ scaling), across training epochs. The AdamW optimizer was used for network optimization, characterized by a weight decay of 0.05 and a learning rate of 0.00001.

Deep learning experiments were conducted using Python 3.9.5, complemented by PyTorch 1.11.0 and CUDA 11.3. A uniform hardware configuration was maintained for model training, which included a CPU with 16 cores, 64 GB RAM, and a Nvidia A100 GPU possessing 40 GB memory. The hardware resources were provided by the Fudan University Medical Research Data Centre.

Image analysis

For the assessment of deep learning reconstruction (DLR_{1500}) image quality, first, all enhanced images DLR_{1500} image, including those from the validation set and external testing set, were independently reviewed by four radiology doctors (two senior and two junior radiology doctors) who are blinded to whether the images were reconstructed. They then evaluated and scored the DLR_{1500} images using the PI-QUAL v2 scoring system [19], providing a more direct assessment of DLR_{1500} image quality. Next, each radiology doctors performed lesion delineation on the enhanced images using ITK-SNAP software to generate post-enhancement ROIs. Finally, these enhanced image-derived ROIs were algorithmically compared with the ground truth ROIs obtained from high *b*-value DWI to calculate pixel-wise AUC values. This analysis aims to determine whether the enhanced DWI images can improve the diagnostic performance of radiologists with varying levels of experience. Each patient's malignant and benign prostate areas were also determined using marked lesion on WSI as the ground truth.

To evaluate the efficiency of DLR_{1500} images in identifying malignant lesions, we conducted receiver operating characteristic (ROC) analyses on DLR_{1500}, as well as 800 *b*-value and 1500 *b*-value DWI images. Sensitivity, specificity, and accuracy for each set of images were also computed correspondingly.

Additionally, to enable a detailed and intuitive evaluation of the segmentation quality of different images sets on prostate lesions, we also used the Dice coefficient as a key indicator. The Dice coefficient, as a quantifiable standard of measurement, can calculate the percentage of overlap between the image ROI and the WSI ground truth, ranging from 0 to 1, where a Dice coefficient score of 1 represents a perfect and complete overlap between the output and the actual state. The Dice coefficient quantifies the degree of spatial overlap between the predicted segmentation (DLR_{1500}, 800 *b*-value and 1500 *b*-value) and the actual segmentation (WSI), providing a key evaluation indicator for the model's segmentation quality of lesions.

Statistical analysis

Significant differences in areas under the curve (AUCs) between the image sets were tested using DeLong's test [20]. We report the AUCs with 95% CIs and other performance metrics (including accuracy, sensitivity and specificity) in the testing set. The 95% CI was calculated following the documentation of "pROC" package [21]. To achieve the best performance of the binary classification, we obtained the optimal cut-off points by maximizing the accuracy in the ROC curves of each image sets as previously described [22]. All tests were two-sided, and

statistical significance was defined as $P < 0.05$. Statistical analyses were performed using R software version 4.2.1.

Role of funding source

The funders of this study had no role in study design, data collection, data analysis, data interpretation, manuscript preparation and review, or submitting the manuscript for publication.

Results

Patient characteristics

As shown in Table 1, we collected baseline information from patients in three cohorts (Training, Validation, and Testing sets), including clinical features such as age, PSA, PI-RADS score, and surgery type, as well as pathological features like pT stage, pN stage, surgical margin, and Gleason score. The results revealed no significant statistical differences in the distribution of clinicopathological

Table 1 Patient baseline characteristics

	Training Set ($n=202$)		Validation Set ($n=101$)		Testing Set ($n=36$)		*P* value
Factor	**No.**	**%**	**No.**	**%**	**No.**	**%**	
Age at surgery (year)							0.185*
Median (IQR)	69.0 (63.0–74.0)		69.0 (65.0-73.50)		69.50 (67.0–76.0)		
PSA at diagnosis (ng/ml)							0.916*
Median (IQR)	14.00 (8.00-28.59)		13.74 (7.84–32.05)		13.41 (8.87–35.87)		
PSA at diagnosis (ng/ml)							0.974#
≤ 20	133	65.8	66	65.3	23	63.9	
> 20	69	34.2	35	34.7	13	36.1	
PI-RADS v2 score							0.100*
≤3	45	22.3	17	16.8	7	19.4	
4	73	36.1	28	27.7	13	36.1	
5	84	41.6	56	55.4	16	44.4	
Clinical T stage							0.841#
cT2	166	82.2	83	82.2	31	86.1	
≥cT3	36	17.8	18	17.8	5	13.9	
Biopsy Gleason grade group							0.802*
1	48	23.8	17	16.8	9	25.0	
2	31	15.3	18	17.8	7	19.4	
3	34	16.8	20	19.8	7	19.4	
4	51	25.2	30	29.7	4	11.1	
5	38	18.8	16	15.8	9	25.0	
Surgery type							0.143#
Open	146	72.3	70	69.3	23	63.9	
LapRP	22	10.9	9	8.9	1	2.8	
RARP	34	16.8	22	21.8	12	33.3	
Pathological T stage							0.909#
pT2	110	54.5	56	55.4	21	58.3	
≥pT3	92	45.5	45	44.6	15	41.7	
Pathological N stage							0.648#
pN0	194	96.0	99	98.0	35	97.2	
pN1	8	4.0	2	2.0	1	2.8	
Surgical Margin							0.348#
Present	177	87.6	86	85.1	34	94.4	
Absence	25	12.4	15	14.9	2	5.6	
Pathologic ISUP							0.923*
1	20	9.9	7	6.9	3	8.3	
2	53	26.2	26	25.7	11	30.6	
3	45	22.3	37	36.6	8	22.2	
4	36	17.8	10	9.9	4	11.1	
5	48	23.8	21	20.8	10	27.8	

*Kruskal-Wallis test. #Chi-square test.

Abbreviation: IQR: interquartile range; PSA: Prostate specific antigen; PI-RADS: Prostate Imaging–Reporting & Data System; LapRP: laparoscopic radical prostatectomy; RARP: robot-assisted radical prostatectomy; ISUP: International Society of Urological Pathology;

variables among the three patient sets, suggesting similar patient distributions across these groups.

DLR_{1500} images generated by NAFNet are not inferior to 1500 *b*-value

Figure 2 illustrated the representative DLR_{1500} images from NAFNet's output, accompanied by their original 800 *b*-value and 1500 *b*-value images. Red arrows indicate the lesions marked by the radiology doctor. The results demonstrate that in these three patients, the prostate cancer lesions exhibit enhanced visibility and sharper boundary definition on both DLR_{1500} and origin high *b*-value (1500 s/mm^2) DWI images. Furthermore, in Patient C, while only a vaguely lesion was observed on the 800 *b*-value DWI image, three distinct lesions were clearly visualized on both DLR_{1500} and original 1500 *b*-value images, demonstrating that the quality of DLR_{1500} imaging is not inferior to high 1500 *b*-value imaging.

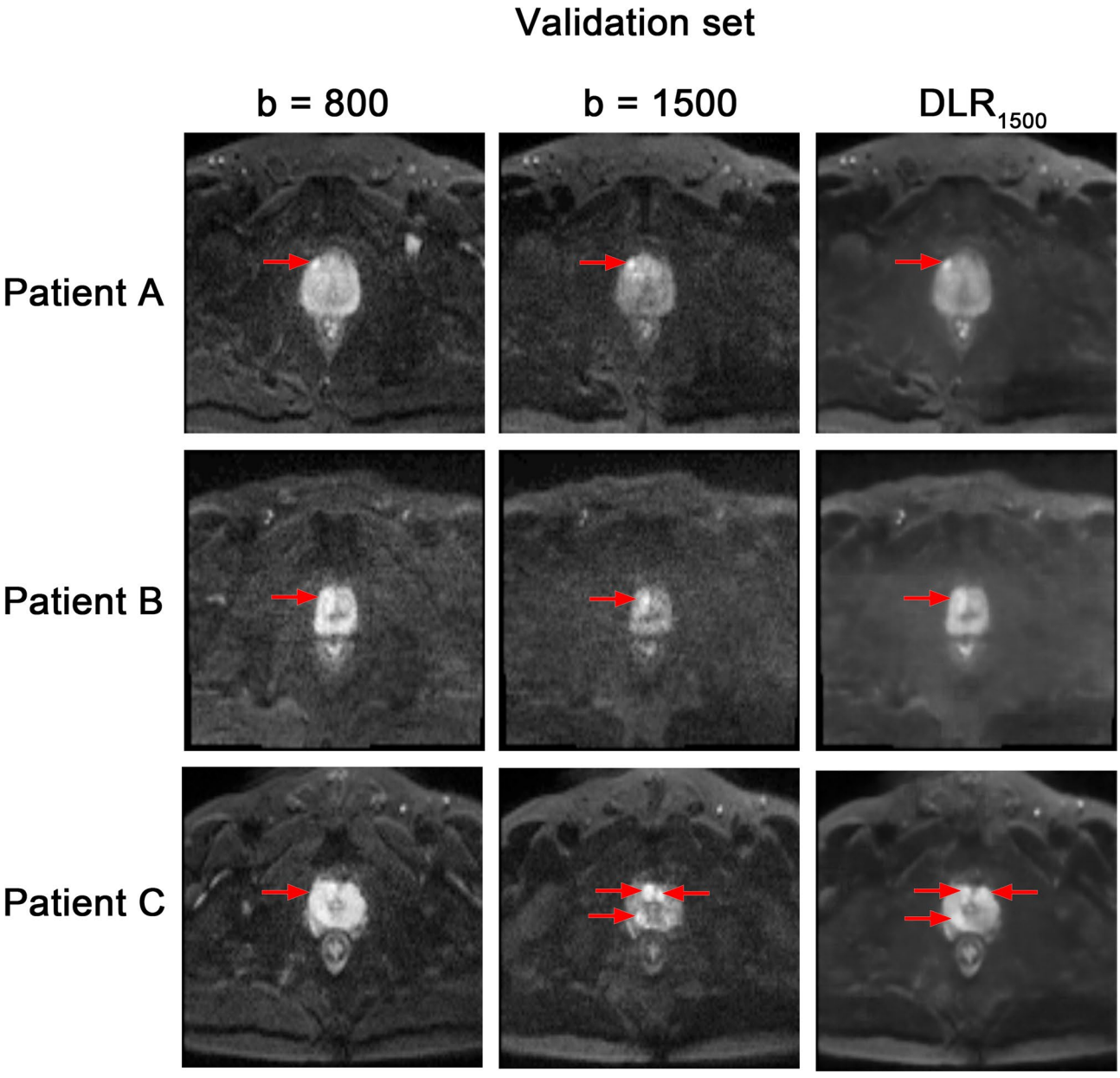

Fig. 2 Representative images of DLR_{1500} images generated by NAFNet within the validation set. Representative images from three patients illustrating the low *b*-value, high *b*-value, and the DLR_{1500} DWI images, with the red arrow indicating the lesions marked by radiology doctors. The prostate cancer lesions are clearer and have more distinct boundaries on both DLR_{1500} and original 1500 *b*-value DWI images. In patient C, although only one lesion can be vaguely observed on the 800 *b*-value DWI image, three lesions can be clearly observed on both the DLR_{1500} and the original 1500 *b*-value images, demonstrating that the performance of our DLR_{1500} images is comparable to that of the original 1500 *b*-value images. Abbreviations: DWI, diffusion-weighted imaging; WSI, whole slide images; DLR, deep learning reconstruction

Table 2 Comparison of various model performances in detecting prostate cancer lesions among Doctors at the same level

Circumstances	AUC on (95% CI) based on WSI	Delong's Test *P*	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Dice coefficient (95% CI)
Validation Set						
Junior radiology doctors						
Original low-*b*	0.753 (0.668–0.838)	**0.043**	0.686 (0.559–0.814)	0.760 (0.642–0.878)	0.723 (0.719–0.727)	0.801 (0.759–0.843)
Original high-*b*	0.821 (0.747–0.899)	0.824	0.882 (0.794–0.971)	0.760 (0.719–0.801)	0.822 (0.819–0.825)	0.885 (0.828–0.942)
DLR_{1500}	0.832 (0.788–0.876)	Reference	0.823 (0.719–0.928)	0.840 (0.738–0.942)	0.832 (0.829–0.834)	0.883 (0.838–0.928)
Senior radiology doctors						
Original low-*b*	0.812 (0.736–0.880)	**0.038**	0.784 (0.671–0.897)	0.820 (0.714–0.926)	0.802 (0.799–0.805)	0.821 (0.770–0.872)
Original high-*b*	0.911 (0.855–0.967)	0.317	0.922 (0.848–0.995)	0.900 (0.817–0.983)	0.911 (0.909–0.912)	0.913 (0.865–0.961)
DLR_{1500}	0.901 (0.842–0.960)	Reference	0.902 (0.820–0.984)	0.900 (0.817–0.983)	0.901 (0.899–0.903)	0.931 (0.910–0.952)
Testing Set						
Junior radiology doctors						
Original low-*b*	0.752 (0.661–0.843)	**0.043**	0.723 (0.564–0.882)	0.712 (0.565–0.859)	0.722 (0.711–0.733)	0.815 (0.759–0.871)
DLR_{1500}	0.848 (0.758–0.938)	Reference	0.770 (0.586–0.954)	0.832 (0.701–0.963)	0.801 (0.797–0.805)	0.835 (0.747–0.888)
Senior radiology doctors						
Original low-*b*	0.823 (0.745–0.901)	0.129	0.872 (0.795–0.949)	0.813 (0.661–0.965)	0.821 (0.810–0.832)	0.865 (0.764–0.966)
DLR_{1500}	0.901 (0.825–0.977)	Reference	0.940 (0.839-1.000)	0.880 (0.744-1.000)	0.917 (0.913–0.921)	0.923 (0.861–0.985)

AUC: area under the receiver operating characteristic curve; CI: confidence interval;

Table 3 Comparison of different Doctors' performances in detecting prostate cancer lesions using the same model

Circumstances	AUC (95% CI) based on WSI	Delong's Test *P*	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Dice coefficient (95% CI)
Validation Set						
Original low-b						
Junior radiology doctors	0.753 (0.668–0.838)	**0.043**	0.686 (0.559–0.814)	0.760 (0.642–0.878)	0.723 (0.719–0.727)	0.801 (0.759–0.843)
Senior radiology doctors	0.812 (0.736–0.880)	Reference	0.784 (0.671–0.897)	0.820 (0.714–0.926)	0.802 (0.799–0.805)	0.821 (0.770–0.872)
Original high-*b*						
Junior radiology doctors	0.821 (0.747–0.899)	**0.045**	0.882 (0.794–0.971)	0.760 (0.719–0.801)	0.822 (0.819–0.825)	0.885 (0.828–0.942)
Senior radiology doctors	0.911 (0.855–0.967)	Reference	0.922 (0.848–0.995)	0.900 (0.817–0.983)	0.911 (0.909–0.912)	0.913 (0.865–0.961)
DLR_{1500}						
Junior radiology doctors	0.832 (0.788–0.876)	**0.011**	0.823 (0.719–0.928)	0.840 (0.738–0.942)	0.832 (0.829–0.834)	0.883 (0.838–0.928)
Senior radiology doctors	0.901 (0.842–0.960)	Reference	0.902 (0.820–0.984)	0.900 (0.817–0.983)	0.901 (0.899–0.903)	0.931 (0.910–0.952)
Testing Set						
Original low-*b*						
Junior radiology doctors	0.752 (0.661–0.843)	**0.040**	0.723 (0.564–0.882)	0.712 (0.565–0.859)	0.722 (0.711–0.733)	0.815 (0.759–0.871)
Senior radiology doctors	0.823 (0.745–0.901)	Reference	0.872 (0.795–0.949)	0.813 (0.661–0.965)	0.821 (0.810–0.832)	0.865 (0.764–0.966)
DLR_{1500}						
Junior radiology doctors	0.848 (0.758–0.938)	**0.042**	0.770 (0.586–0.954)	0.832 (0.701–0.963)	0.801 (0.797–0.805)	0.835 (0.747–0.888)
Senior radiology doctors	0.901 (0.825–0.977)	Reference	0.940 (0.839-1.000)	0.880 (0.744-1.000)	0.917 (0.913–0.921)	0.923 (0.861–0.985)

AUC: area under the receiver operating characteristic curve; CI: confidence interval;

Subsequently, we evaluated the PI-QUAL v2 scores for the original 1500 *b*-value and DLR_{1500} in the validation set. As shown in Supplementary Fig. 1, DLR_{1500} images achieved a 3.757/4.000 PI-QUAL v2 score, significantly higher than origin 1500 *b*-value images (3.708/4.000). This result indicated that visually, radiology doctors considered the quality of DLR_{1500} images to be superior ($P = 0.0002$). Furthermore, we still found that some results of DLR_{1500} were unsatisfactory. As shown in Supplementary Fig. 2, these inferior synthetic images can lead to false positives or false negatives in tumour lesions, and the overall images tend to be blurry.

We then compared the extent of improvement in the diagnostic capabilities of different levels of radiology doctors with DLR_{1500} images. Within the validation set (upper portion of Table 2), among junior radiology doctors, the diagnostic AUC based on DLR_{1500} images is not inferior to the diagnostic AUC based on 1500 *b*-value images (0.832 (0.788–0.876) vs. 0.821 (0.747–0.899), $P = 0.824$). The same phenomenon is also observed in senior radiology doctors (DLR_{1500} vs. 1500 *b*-value: 0.901 (0.842–0.960) vs. 0.911 (0.855–0.967), $P = 0.317$).

As presented in the upper portion of Table 3, independent of the used image sets, the diagnostic performance

metrics (including AUC, sensitivity, specificity, and accuracy) of the senior radiology doctors significantly outperforms that of junior radiology doctors (Original low-*b*: $P = 0.043$, Original high-*b*: $P = 0.045$, DLR_{1500}: $P = 0.011$). Using WSI as the gold standard, the Dice coefficient of the ROI area delineated by senior radiology doctors is also higher than that of the ROI area delineated by junior radiology doctors (Original low-*b*: 0.821 vs. 0.801, Original high-*b*: 0.913 vs. 0.885, DLR_{1500}: 0.931 vs. 0.883). In short, as shown in Table 3, the diagnostic AUC of senior radiology doctors is superior to that of junior radiology doctors, regardless of the image source, which demonstrates the difference at the physician level.

DLR_{1500} images are superior to 800 *b*-value images in detecting lesions

Finally, in the external testing set, we directly input the 800 *b*-value DWI images to obtain the DLR_{1500} images generated by NAFNet. In the testing set (bottom portion of Table 2), we found that, compared to 800 *b*-value images, the DLR_{1500} images made limited improvement to the diagnostic performance of senior radiology doctors (0.901 (0.825–0.977) vs. 0.823 (0.745–0.901), $P = 0.129$). However, for junior radiology doctors, DLR_{1500} images could significantly enhance their diagnostic performance than 800 *b*-value images (0.848 (0.758–0.938) vs. 0.752 (0.661–0.843), $P = 0.043$). Similarly, for both 800 *b*-value images and DLR_{1500} images, the performance metrics (AUC, sensitivity, specificity, accuracy, and Dice coefficient) of senior radiology doctors are superior to those of junior radiology doctors (bottom portion of Table 3).

Furthermore, as shown in Fig. 3, we also found in the testing set that on the DLR_{1500} images generated by NAFNet from 800 *b*-values, prostate lesions appear more clearly. Using WSI as the gold standard (Fig. 3C), the lesions on the DLR_{1500} images (Fig. 3B) match the WSI more closely than the original 800 *b*-value images (Fig. 3A). With the aid of DLR_{1500} images, radiologists were able to identify pathologically confirmed prostate cancer lesions that were undetectable on origin high 1500 *b*-value DWI sequences.

Discussion

In this study, we utilized a deep learning network, NAFNet, to enhance the quality of 800 *b*-value DWI images and subsequently generate DLR_{1500} DWI images that mimic 1500 *b*-value images. Following a comprehensive evaluation by both junior and senior radiologists in the internal validation set, we found that the quality of the generated DLR_{1500} DWI images was not inferior to that of 1500 *b*-value images. Meanwhile, in an independent external testing set, and employing WSI as the gold standard, we observed that the diagnostic performance of the generated DLR_{1500} DWI images significantly superior to that of the 800 *b*-value images, thereby notably improve the diagnostic capabilities of junior radiology doctors.

In the diagnosis of prostate cancer, the importance of DWI is beyond doubt. The Gaussian diffusion model, often employed in DWI, assumes a mono-exponential decay of the diffusion signal. However, this model is limited in its ability to accurately characterize complex tissue structures, particularly in heterogeneous tissues where diffusion is non-Gaussian [23]. This limitation has led to the exploration of non-Gaussian models, which provide a more comprehensive understanding of diffusion processes by accounting for deviations from the mono-exponential decay. For example, in prostate cancer, non-Gaussian diffusion models have demonstrated potential in distinguishing between different grades of cancer, offering a more nuanced assessment compared to traditional Gaussian models [24, 25]. The ability of these non-Gaussian models to capture the complex diffusion behaviour in tissues makes them a powerful tool in enhancing the diagnostic capabilities of prostate MRI DWI.

With the advancements of computer science, artificial intelligence algorithms not only widely used in the segmentation and classification of medical images, but also increasingly applied in the field of image simulation and enhancement [26–29]. These algorithms, including convolutional neural networks and generative adversarial networks, have been proven to accelerate MRI image acquisition, and reduce radiologist reading times [30–33]. In the field of prostate MRI, Ueda et al. [10] used different DWI images of 0, 1000, 3000, 5000 *b*-values from 60 patients, and applied convolutional neural networks for image enhancement. They found that compared to the original images, the images generated by AI have higher signal-to-noise and contrast-to-noise ratios, and the subjective performance index of radiologists is also higher. In another study, Hu et al. [11] used generative adversarial networks to generate images from 800 *b*-value and 1000 *b*-value DWI images to 1500 *b*-value images. The generated images optimized through a series of steps were significantly better than the original 1500 *b*-value images. Further evaluation showed that the optimized generated images can help junior radiology doctors to improve their diagnostic capabilities. Consistent with the findings of these two studies, our study also revealed that the NAFNet-generated DLR_{1500} images significantly enhanced the diagnostic accuracy of junior radiology doctors.

Of note, we have also discovered that images with inferior enhancement often result in diagnostic challenges, such as false positives and false negatives in the identification of tumour lesions. These images are generally characterized by a lack of sharpness and a high level of noise, as shown in Supplementary Fig. 2. While the black-box nature of deep learning models limits our ability to

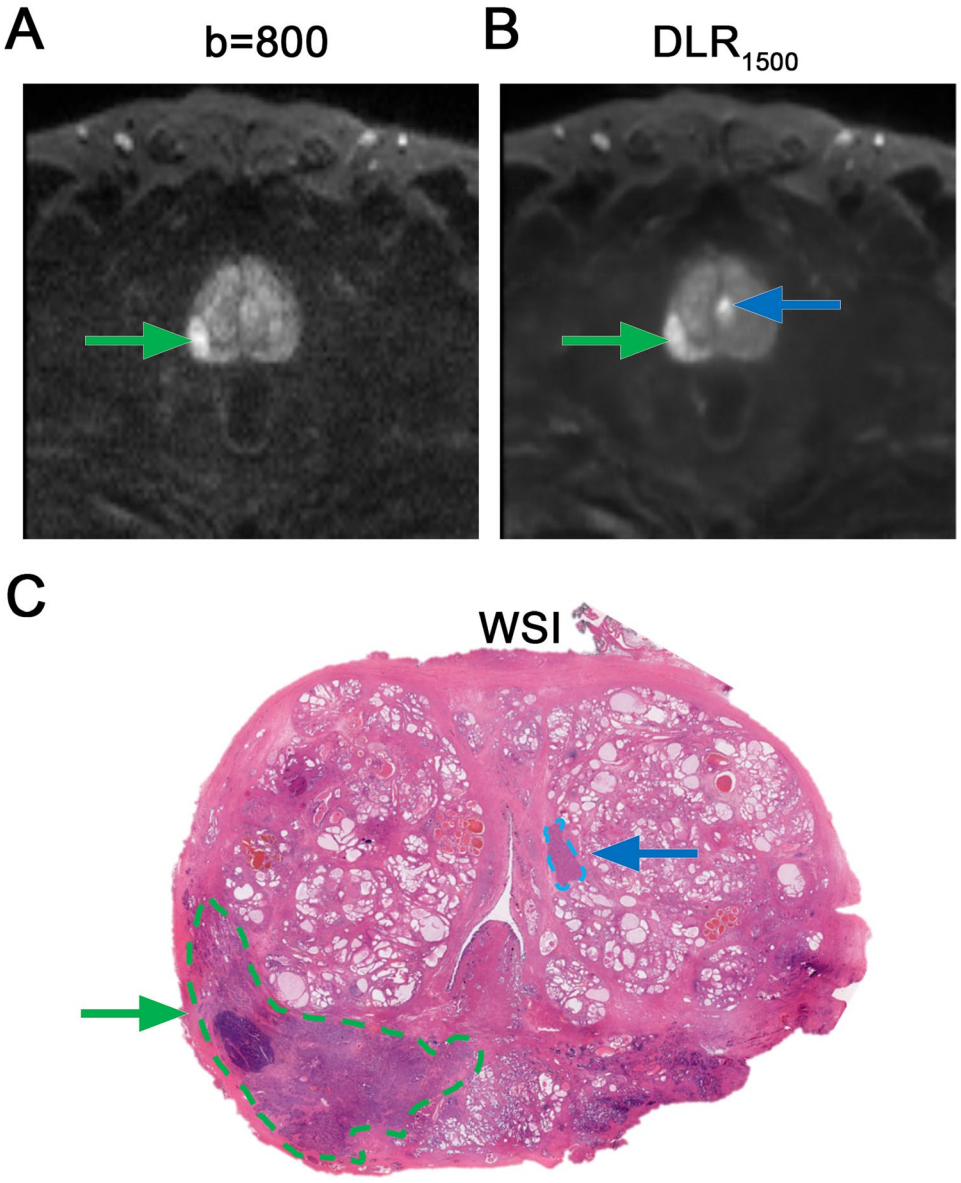

Fig. 3 DLR_{1500} images generated by NAFNet could identify more lesions compared to low *b*-value images within the testing set. (**a**) Original low *b*-value image within the testing set, with a green arrow indicating the lesion marked by radiology doctors. (**b**) DLR_{1500} image within the testing set, with a green arrow and a blue arrow indicating the lesions based on WSI ground truth. (**c**) WSI showing the corresponding lesions in DLR_{1500} images with the green and blue arrow marked by pathologist. With the help of DLR_{1500} images, radiologists can identify prostate cancer lesions that were not detected on the original 1500 *b*-value images but were confirmed pathologically. Abbreviations: WSI, whole slide images; DLR, deep learning reconstruction

fully understand the precise causes of such inferior quality [34], we hypothesize that several factors could be contributing. These include the size of the lesions, which may affect how they are captured and processed; the spatial relationship between lesions and surrounding tissues, which can influence contrast and clarity; and the volume and diversity of training images, which play a crucial role in the model's ability to generalize well across different images. Understanding and addressing these elements could be key to improving image quality and reducing inferior synthetic results.

Compared to previous studies, our research has several advantages. Firstly, our study utilizes a new deep learning framework, NAFNet, as the foundational architecture. NAFNet is a non-linear activation function network that simplifies traditional structures by replacing complicated non-linear activation functions. Despite the simplification, its performance equals or exceeds benchmarks with lower computational costs [17, 18]. Secondly, while most studies used radiologists' subjective assessment as the gold standard for models, we treated pathologic WSI as the ground truth. During the model's training stage,

radiologists and pathologists closely coordinate the ROI from pathological images onto DWI images, ensuring ROI precision in the model. In the validation and testing stage, both junior and senior radiology doctors delineate lesions on DLR_{1500} and original DWI images and compare them with pathological gold standards, verifying more precisely the extent to which the DLR_{1500} images enhance diagnostic capabilities. Finally, we have independently verified our algorithm using data from external institution. The results of external testing set showed that our model has robust resilience, capable of managing different MR scanners.

This study presents several limitations. First, given the nature limitation of retrospective study, our results must be validated in multi-centre prospective cohorts to confirm generalizability. Moreover, the multi-centre nature of our study inevitably introduces inter-scanner heterogeneity; thus, we intend to test these findings in larger, more diverse multi-centre populations. Secondly, due to the inherent challenges in multi-centre data collection, our current testing set comprised only 36 cases. While the results demonstrated robust performance of our algorithm across this multi-centre dataset, validation through larger independent multi-centre cohorts remains significant. In future research, we plan to utilize public databases, such as the TCIA (The Cancer Imaging Archive), to conduct further validation and enhance generalizability. Further, while a considerable number of WSIs were used for registration, the precise alignment of the WSI with MRI could not be entirely assured, which could potentially impact the results. We plan to develop more robust registration technique in future investigations to address this critical technical challenge. Finally, the interpretability of the features extracted by the deep learning network is relatively weak. Compared with traditional machine learning radiomics, NAFNet cannot extract the corresponding interpretable and specific radiographic features for radiologists.

Conclusions

In conclusion, the deep learning algorithm, NAFNet, can utilize low *b*-value DWI images to generate DLR_{1500} images of superior quality than 800 *b*-value, and not inferior to 1500 *b*-value DWI images, leading to the improvement of the diagnostic accuracy of prostate lesions.

Abbreviations

AI	Artificial intelligence
AUC	Area under the curve
DLR	Deep learning reconstruction
DWI	Diffusion–weighted imaging
PI-RADS score	Prostate Imaging Reporting and Data System score
ROI	Region-of-interest
RP	Radical prostatectomy
WSI	Whole-slide images

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14354-y.

Supplementary Material 1

Acknowledgements

We thank the Medical Research Data Center of Fudan University in supporting server and hardware environments.

Author contributions

Z.L and WJ.G for acquisition of data, analysis and interpretation of data, statistical analysis and drafting of the manuscript; FN.W for material support; XH.L and ZZ.C for radiodiagnosis; YY.K for pathological diagnosis; XH.L, DW.Y and B.D for study concept and design, review of the manuscript, obtained funding and study supervision. Z.L and WJ.G contributed equally to this study. All authors read and approved the final manuscript for submission and agreed with the order of presentation of the authors.

Funding

This study was supported financially by the National Nature Science Foundation of China (82373355, 82172703, 82303097), the Shanghai Oriental Talent Program Top Project (BJKJ2024007), the Discipline Leader Project of Shanghai Municipal Health Commission (2022XD013), the AoXiang Project of Shanghai anti-cancer association (SACA-AX202302) and the PARP Inhibitor Cancer Research Fund (Phase 4) of China Anti-Cancer Association (CACA). All these study sponsors have no roles in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data availability

The datasets used and analysed during the current study are available from the corresponding author on reasonable request. Enquiries regarding the datasets should be sent directly to the corresponding author, Prof. M.D. Bo Dai, via email (bodai1978@126.com). A data access request form is required, and adherence to institutional rules and regulations governing data access is mandatory. All code was implemented in Python 3.9 and R 4.3.1. The original code for the NAFNet model can be accessed at https://github.com/megvii-research/NAFNet. The custom Python code specifically relevant to this study is openly accessible at https://github.com/SHMCLz/NAFNet-for-MRI-Enhance.

Declarations

Ethics approval and consent to participate

This study protocol involving human participants was reviewed and approved by the Institutional Review Board of Fudan University Shanghai Cancer Centre (Approval No. 050432-4-2108). Written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Author details

[1]Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[2]Department of Oncology, Shanghai Medical College, Fudan University, 200032 Shanghai, China
[3]Shanghai Genitourinary Cancer Institute, 200032 Shanghai, China
[4]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[5]Department of Pathology, Fudan University Shanghai Cancer Center, 200032 Shanghai, China

Received: 2 December 2024 / Accepted: 19 May 2025

Published online: 27 May 2025

References

1. Dai B, Wang H, Shi B, Xing J, Zhu S, He Z, Zou Q, Wei Q, Bi J, Bian J, et al. CACA guidelines for holistic integrative management of prostate cancer. Holist Integr Oncol. 2024;3(1):47.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.
3. Kayat Bittencourt L, Litjens G, Hulsbergen-van de Kaa CA, Turkbey B, Gasparetto EL, Barentsz JO. Prostate cancer: the European society of urogenital radiology prostate imaging reporting and data system criteria for predicting extraprostatic extension by using 3-T multiparametric MR imaging. Radiology. 2015;276(2):479–89.
4. Padhani AR, Barentsz J, Villeirs G, Rosenkrantz AB, Margolis DJ, Turkbey B, Thoeny HC, Cornud F, Haider MA, Macura KJ, et al. PI-RADS steering committee: the PI-RADS multiparametric MRI and MRI-directed biopsy pathway. Radiology. 2019;292(2):464–74.
5. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, Tempany CM, Choyke PL, Cornud F, Margolis DJ, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. Eur Urol. 2019;76(3):340–51.
6. Metens T, Miranda D, Absil J, Matos C. What is the optimal B value in diffusion-weighted MR imaging to depict prostate cancer at 3T? Eur Radiol. 2012;22(3):703–9.
7. Jendoubi S, Wagner M, Montagne S, Ezziane M, Mespoulet J, Comperat E, Estellat C, Baptiste A, Renard-Penna R. MRI for prostate cancer: can computed high b-value DWI replace native acquisitions? Eur Radiol. 2019;29(10):5197–204.
8. Duan X, Xiong H, Liu R, Duan X, Yu H. Enhanced deep leaning model for detection and grading of lumbar disc herniation from MRI. Med Biol Eng Comput 2024.
9. Bischoff LM, Peeters JM, Weinhold L, Krausewitz P, Ellinger J, Katemann C, Isaak A, Weber OM, Kuetting D, Attenberger U, et al. Deep learning Super-Resolution reconstruction for fast and Motion-Robust T2-weighted prostate MRI. Radiology. 2023;308(3):e230427.
10. Ueda T, Ohno Y, Yamamoto K, Murayama K, Ikedo M, Yui M, Hanamatsu S, Tanaka Y, Obama Y, Ikeda H, et al. Deep learning reconstruction of Diffusion-weighted MRI improves image quality for prostatic imaging. Radiology. 2022;303(2):373–81.
11. Hu L, Zhou DW, Zha YF, Li L, He H, Xu WH, Qian L, Zhang YK, Fu CX, Hu H, et al. Synthesizing High-b-Value Diffusion-weighted imaging of the prostate using generative adversarial networks. Radiol Artif Intell. 2021;3(5):e200237.
12. Yokota Y, Takeda C, Kidoh M, Oda S, Aoki R, Ito K, Morita K, Haraoka K, Yamashita Y, Iizuka H, et al. Effects of deep learning reconstruction technique in High-Resolution Non-contrast magnetic resonance coronary angiography at a 3-Tesla machine. Can Assoc Radiol J. 2021;72(1):120–7.
13. Almansour H, Herrmann J, Gassenmaier S, Afat S, Jacoby J, Koerzdoerfer G, Nickel D, Mostapha M, Nadar M, Othman AE. Deep learning reconstruction for accelerated spine MRI: prospective analysis of interchangeability. Radiology. 2023;306(3):e212922.
14. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal. 2019;58:101552.
15. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nat Rev Urol. 2019;16(7):391–403.
16. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond). 2020;40(4):154–66.
17. Chen L, Chu X, Zhang X, Sun J. Simple Baselines for Image Restoration. *arXiv e-prints* 2022:arXiv:2204.04676.
18. Gu WJ, Liu Z, Yang YJ, Zhang XZ, Chen LY, Wan FN, Liu XH, Chen ZZ, Kong YY, Dai B. A deep learning model, NAFNet, predicts adverse pathology and recurrence in prostate cancer using MRIs. NPJ Precis Oncol. 2023;7(1):134.
19. de Rooij M, Allen C, Twilt JJ, Thijssen LCP, Asbach P, Barrett T, Brembilla G, Emberton M, Gupta RT, Haider MA, et al. PI-QUAL version 2: an update of a standardised scoring system for the assessment of image quality of prostate MRI. Eur Radiol. 2024;34(11):7068–79.
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
22. Hiremath A, Shiradkar R, Fu P, Mahran A, Rastinehad AR, Tewari A, Tirumani SH, Purysko A, Ponsky L, Madabhushi A. 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):e445–54.
23. Zhang G, Wang S, Wen D, Zhang J, Wei X, Ma W, Zhao W, Wang M, Wu G, Zhang J. Comparison of non-Gaussian and Gaussian diffusion models of diffusion weighted imaging of rectal cancer at 3.0 T MRI. Sci Rep. 2016;6:38782.
24. Brancato V, Cavaliere C, Salvatore M, Monti S. Non-Gaussian models of diffusion weighted imaging for detection and characterization of prostate cancer: a systematic review and meta-analysis. Sci Rep. 2019;9(1):16837.
25. Liu G, Lu Y, Dai Y, Xue K, Yi Y, Xu J, Wu D, Wu G. Comparison of mono-exponential, bi-exponential, kurtosis, and fractional-order calculus models of diffusion-weighted imaging in characterizing prostate lesions in transition zone. Abdom Radiol (NY). 2021;46(6):2740–50.
26. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–10.
27. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discov. 2024;14(5):711–26.
28. Bhalla K, Xiao Q, Luna JM, Podany E, Ahmad T, Ademuyiwa FO, Davis A, Bennett DL, Gastounioti A. Radiologic imaging biomarkers in triple-negative breast cancer: a literature review about the role of artificial intelligence and the way forward. *BJR|Artificial Intelligence* 2024, 1(1%@ 2976–8705).
29. Roy S, Bhalla K, Patel R. Mathematical analysis of histogram equalization techniques for medical image enhancement: a tutorial from the perspective of data loss. Multimedia Tools Appl. 2024;83(5):14363–92.
30. Rastogi A, Brugnara G, Foltyn-Dumitru M, Mahmutoglu MA, Preetha CJ, Kobler E, Pfluger I, Schell M, Deike-Hofmann K, Kessler T, et al. Deep-learning-based reconstruction of undersampled MRI to reduce scan times: a multicentre, retrospective, cohort study. Lancet Oncol. 2024;25(3):400–10.
31. Johnson PM, Lin DJ, Zbontar J, Zitnick CL, Sriram A, Muckley M, Babb JS, Kline M, Ciavarra G, Alaia E, et al. Deep learning reconstruction enables prospectively accelerated clinical knee MRI. Radiology. 2023;307(2):e220425.
32. Wilpert C, Neubauer C, Rau A, Schneider H, Benkert T, Weiland E, Strecker R, Reisert M, Benndorf M, Weiss J, et al. Accelerated Diffusion-Weighted imaging in 3 T breast MRI using a deep learning reconstruction algorithm with superresolution processing: A prospective comparative study. Invest Radiol. 2023;58(12):842–52.
33. Huang Y-P, Bhalla K, Chu H-C, Lin Y-C, Kuo H-C, Chu W-J, Lee J-H. Wavelet K-Means clustering and Fuzzy-Based method for segmenting MRI images depicting Parkinson's disease. Int J Fuzzy Syst. 2021;23(6):1600–12.
34. Klauschen F, Dippel J, Keyl P, Jurmeister P, Bockmayr M, Mock A, Buchstab O, Alber M, Ruff L, Montavon G, et al. Toward explainable artificial intelligence for precision pathology. Annu Rev Pathol. 2024;19:541–70.

Publisher's note