https://doi.org/10.1186/s13244-023-01439-0

Insights into Imaging
EUROPEAN SOCIETY OF RADIOLOGY

ORIGINAL ARTICLE

Open Access

Anatomically guided self-adapting deep neural network for clinically significant prostate cancer detection on bi-parametric MRI: a multi-center study

Ahmet Karagoz[1,2], Deniz Alis[2,3*], Mustafa Ege Seker[4], Gokberk Zeybel[4], Mert Yergin[2], Ilkay Oksuz[1] and Ercan Karaarslan[3]

Abstract

Objective To evaluate the effectiveness of a self-adapting deep network, trained on large-scale bi-parametric MRI data, in detecting clinically significant prostate cancer (csPCa) in external multi-center data from men of diverse demographics; to investigate the advantages of transfer learning.

Methods We used two samples: (i) Publicly available multi-center and multi-vendor Prostate Imaging: Cancer AI (PI-CAI) training data, consisting of 1500 bi-parametric MRI scans, along with its unseen validation and testing samples; (ii) In-house multi-center testing and transfer learning data, comprising 1036 and 200 bi-parametric MRI scans. We trained a self-adapting 3D nnU-Net model using probabilistic prostate masks on the PI-CAI data and evaluated its performance on the hidden validation and testing samples and the in-house data with and without transfer learning. We used the area under the receiver operating characteristic (AUROC) curve to evaluate patient-level performance in detecting csPCa.

Results The PI-CAI training data had 425 scans with csPCa, while the in-house testing and fine-tuning data had 288 and 50 scans with csPCa, respectively. The nnU-Net model achieved an AUROC of 0.888 and 0.889 on the hidden validation and testing data. The model performed with an AUROC of 0.886 on the in-house testing data, with a slight decrease in performance to 0.870 using transfer learning.

Conclusions The state-of-the-art deep learning method using prostate masks trained on large-scale bi-parametric MRI data provides high performance in detecting csPCa in internal and external testing data with different characteristics, demonstrating the robustness and generalizability of deep learning within and across datasets.

Clinical relevance statement A self-adapting deep network, utilizing prostate masks and trained on large-scale bi-parametric MRI data, is effective in accurately detecting clinically significant prostate cancer across diverse datasets, highlighting the potential of deep learning methods for improving prostate cancer detection in clinical practice.

*Correspondence:
Deniz Alis
deniz.alis@acibadem.edu.tr
Full list of author information is available at the end of the article

Key points

1. We trained a self-adapting deep network on large-scale bi-parametric prostate MRI scans.
2. The model provided a high performance at detecting csPCa on in-distribution tests.
3. The performance did not drop on the external multi-center & multi-vendor data.
4. Transfer learning did not improve the performance in the external test.

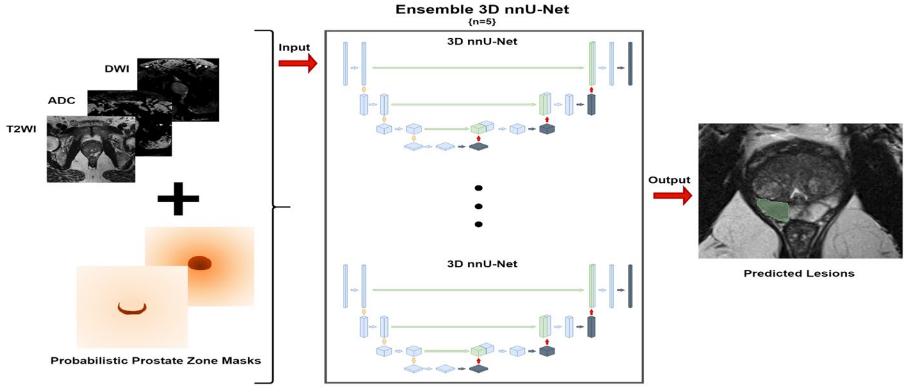
Keywords Deep learning, Magnetic resonance imaging, Prostate cancer

Graphical Abstract

Anatomically guided self-adapting Deep Neural Network for clinically significant prostate cancer detection on bi-parametric MRI: A multi-center study

Insights into Imaging

EUROPEAN SOCIETY OF RADIOLOGY



A self-adapting deep network, utilizing prostate masks and trained on large-scale bi-parametric MRI data, is effective in accurately detecting clinically significant prostate cancer across diverse datasets, highlighting the potential of deep learning methods for improving prostate cancer detection in clinical practice.

Insights Imaging (2023) Karagoz A et al. DOI: 10.1186/s13244-023-01439-0

Introduction

Magnetic resonance imaging (MRI) plays an imperative role in prostate cancer (PCa) diagnostics, and the number of prostate MRI scans is expected to increase significantly as the recent evidence suggests performing pre-biopsy prostate MRI in men with suspicion of PCa [1]. The main objective of prostate MRI is to identify clinically significant PCa (csPCa) (i.e., Gleason Score ≥ 3 + 4) while sparring men with benign lesions or indolent PCa from unnecessary interventions or treatment.

The prostate imaging-reporting and data system (PI-RADS) was introduced in 2012 and most recently updated in 2019 as PI-RADS Version 2.1 to standardize prostate MRI acquisition and interpretation [2]. Though the benefits of the PI-RADS have been well recognized over the years, prostate MRI still suffers from intra-reader and inter-reader differences and non-negligible amounts of false-positive and false-negative results [3–5].

Deep learning (DL) has shown remarkable performance on a broad spectrum of medical imaging tasks in recent years, with prostate cancer diagnostics no exception. However, earlier studies have been hindered by several issues: (i) training, validating, and testing the DL models on the same data obtained [6–16]; (ii) having a small sample size [6–16]; (iii) insufficient details regarding the DL models and/or lack of open-source code sharing [6–12, 15]; (iv) lack of benchmarking DL models on publicly available datasets or challenges [6–13]; using the PI-RADS scores as the reference for performance estimation [16–18].

ProstateX challenge partially addressed the aforementioned problems, yet it did not have the adequate data size to train and test DL models effectively [19]. PI-CAI (Prostate Imaging: Cancer AI) is a new grand challenge encompassing over 10,000 prostate MRI scans [20]. The challenge allows researchers to design, train and test publicly available DL models on large-scale for identifying csPCa on bi-parametric prostate MRI.

We hypothesized that a state-of-the-art self-adapting deep network trained on large-scale bi-parametric MRI data using best practices could provide robust and generalizable performance in detecting csPCa when applied to large-scale multi-center and multi-vendor external datasets. To test our hypothesis, we first trained a self-adapting DL model, nnU-Net, using probabilistic prostate masks on the PI-CAI dataset and tested its performance on the hidden validation and testing set of the challenge. Then, we assessed the performance of the model on a private in-house multi-center dataset comprising men of different demographics. Further, we investigated the benefits of transfer learning on the performance using a small partition of the in-house dataset.

Methods

Study sample

We used two datasets in the present work: Publicly available PI-CAI training data and in-house data. The PI-CAI consists of over 10,000 bi-parametric prostate MRIs, yet only 1500 scans are publicly available. Researchers are also allowed to test their models on the hidden validation set, consisting of 100 scans, and the hidden testing set, consisting of 1000 scans. However, the hidden testing set was only available during a pre-defined period.

Acibadem Mehmet Ali Aydinlar University's review board approved this retrospective study and waived the need for informed consent for the retrospective analysis of medical data. We reviewed consecutive patients who underwent a bi-parametric or multi-parametric prostate MRI scan due to suspicion of PCa (i.e., increased prostate-specific antigen or suspicious digital rectal examination) or active surveillance between January 2015 and December 2021 to create the in-house dataset.

All men in the in-house dataset had undergone whole-mount pathology or biopsy after the MRI scan or were MRI-negative (i.e., PI-RADS score of 1 or 2) with a minimum follow-up of 24 months without any clinical, laboratory, or imaging evidence of PCa. In addition, men who underwent a prostate MRI scan with an endorectal coil were excluded from the study, as were men with a history of any treatment for PCa or prostate operation.

MRI protocols

The scans in the in-house dataset were obtained at nine institutions with 1.5 T (Avanto, Avanto-fit, and Aera, Siemens Healthcare, Erlangen/Germany; Signa HDxt Signa, General Electric Healthcare, Chicago/USA) or 3 T scanners (Prisma, Skyra, and Vida, Siemens Healthcare, Erlangen/Germany; Signa Premier, GE Healthcare, Chicago/USA).

All mpMRI or bi-parametric MRI protocols followed PI-RADS version 2 or 2.1. At a minimum, the bi-parametric prostate MRI protocol encompassed tri-planar T2-weighted and diffusion-weighted imaging. The diffusion-weighted imaging was performed with echo-planar imaging in axial planes with at least three b-values. Some patients had an acquired DWI with a b-value ≥ 1400 s/mm^2, while others had calculated DWI with a b-value of 1400 s/mm^2 following the PI-RADS. The ADC maps were calculated using a linear least-square fitting with all acquired b-value. We did not use dynamic contrast-enhanced images since the challenge organizers did not provide them. Further details of the MRI protocols were omitted for the sake of brevity.

Ground-truth labels

The organizers provided the ground-truth labels for the PI-CAI dataset. For scans harboring csPCa, the organizers provided pixel-level annotations (i.e., lesion masks). For the remaining patients (i.e., those with indolent PCa or benign findings), the organizers only provided scan-level results. All csPCa lesions were annotated by trained investigators under the supervision of three expert radiologists.

The ground-truth labels of the in-house dataset were created following a similar method. First, all available pathology, radiology, and clinical reports of the patients were curated. Then, a radiologist (D.A.), one of the PI-CAI readers with three years of prostate MRI (≤ 150 prostate scans a year) and five years of DL experience, segmented the csPCa on bi-parametric prostate MRI scans. All the segmentations were carried out on a dedicated browser-based platform (https://matrix.md.ai). All segmentations were supervised by a senior radiologist (E.K.) with over 20 years of experience in prostate imaging (≥ 300 scans a year).

In patients with available whole-mount pathology, the radiologist determined the csPCa lesions (i.e., Gleason Score ≥ 3 + 4) using the digitized histopathological images as the reference to ensure a radio-pathological

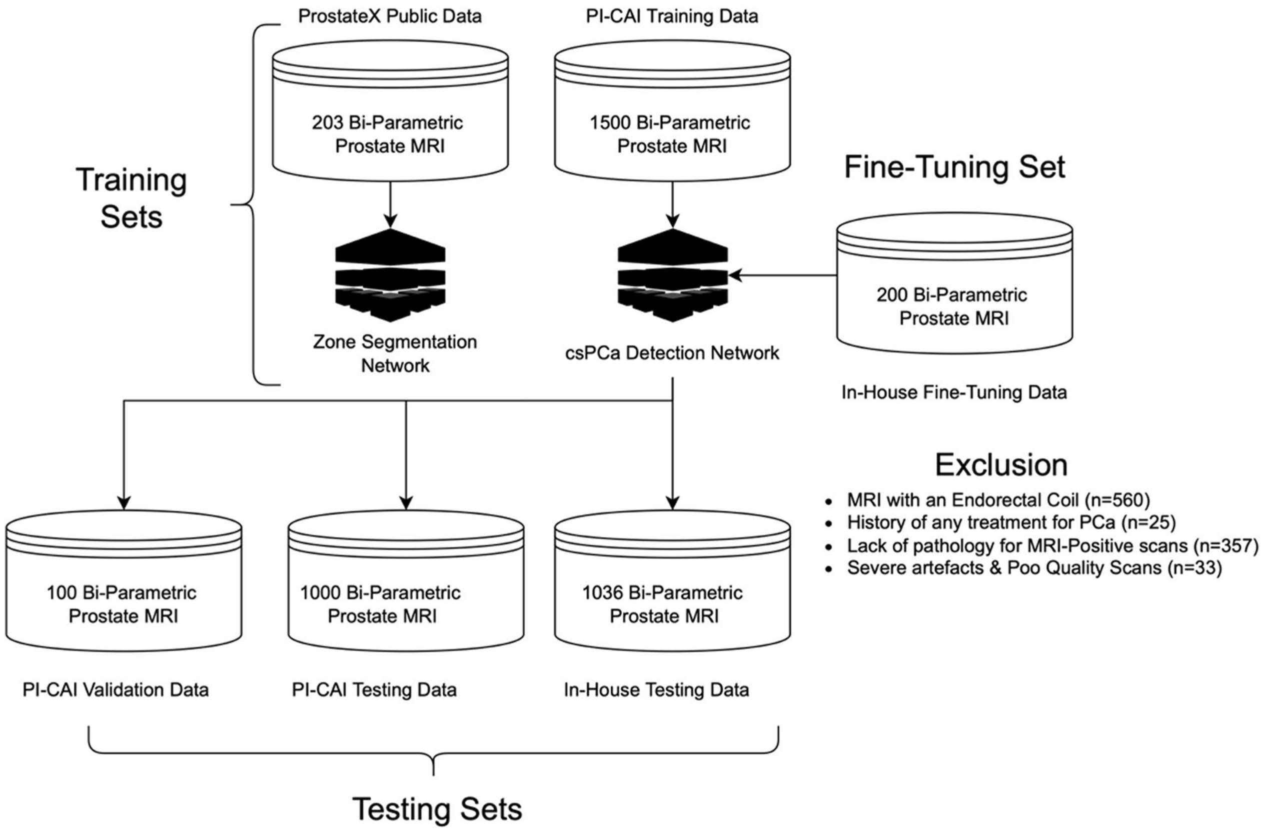

ProstateX Public Data
203 Bi-Parametric Prostate MRI
PI-CAI Training Data
1500 Bi-Parametric Prostate MRI
Training Sets
Fine-Tuning Set
200 Bi-Parametric Prostate MRI
In-House Fine-Tuning Data
Zone Segmentation Network
csPCa Detection Network
Exclusion
MRI with an Endorectal Coil (n=560)
History of any treatment for PCa (n=25)
Lack of pathology for MRI-Positive scans (n=357)
Severe artefacts & Poo Quality Scans (n=33)
100 Bi-Parametric Prostate MRI
PI-CAI Validation Data
1000 Bi-Parametric Prostate MRI
PI-CAI Testing Data
1036 Bi-Parametric Prostate MRI
In-House Testing Data
Testing Sets

Fig. 1 The datasets used in the study. The ProstateX data and Prostate Imaging: Cancer AI [PI-CAI] training data were used for the model training in the study. The PI-CAI data was used to train an ensemble of 3D nnU-Net models to detect clinically significant PCa

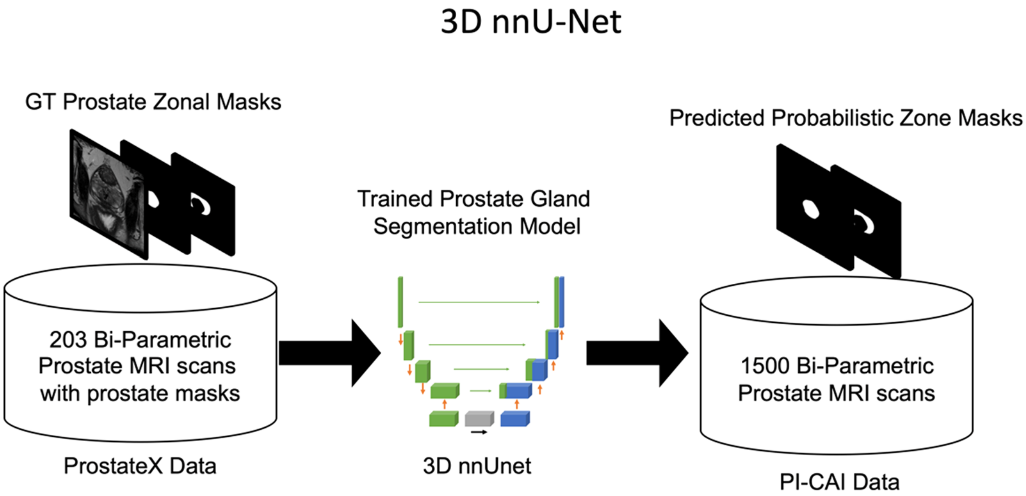

3D nnU-Net
GT Prostate Zonal Masks
203 Bi-Parametric Prostate MRI scans with prostate masks
ProstateX Data
Trained Prostate Gland Segmentation Model
3D nnUnet
Predicted Probabilistic Zone Masks
1500 Bi-Parametric Prostate MRI scans
PI-CAI Data

Fig. 2 Creating the probabilistic zone masks. The ProstateX data was used to train a 3D nnU-Net for creating probabilistic prostate zone masks on the PI-CAI training data. Afterward, the probabilistic masks were used to augment the ensemble model's clinically significant prostate cancer detection performance

match [21]. In patients with biopsy results, the radiologist carefully read the pathology results regarding the location of the lesion. In the in-house sample, all biopsy procedures involved a combination of 3–4-core MRI/ultrasound fusion-guided biopsy followed by an extended transrectal systematic biopsy (Artemis, Eigen) [16]. Men with a benign pathology result or MRI-negative patients (i.e., those with PI-RADS 1 or 2 scans) without any clinical, laboratory, or imaging evidence of PCa were accepted as negative for csPCa. Figure 1 shows the study sample selection.

DL models

First, we trained a prostate zone segmentation model on patients from the publicly available ProstateX dataset [22]. We used a 3D nnU-Net, a self-adapting DL framework, fed the model with T2W images and used the peripheral and central gland masks as the ground truth [23]. Then, we used this network on challenge data to obtain probabilistic central and peripheral gland masks. Figure 2 shows an overview of the gland segmentation network.

We developed 3D nnU-Net for csPCa detection on the challenge data. We fed the model with T2W, ADC, high-b-value DWI, and probabilistic prostate segmentation masks using the csPCa masks as the ground truth.

The nnU-Net model is based on a standard U-Net architecture. This U-Net consists of two sequential encoder-decoder components interconnected via skip connections and a bottleneck layer at the bottom of the model. The encoder layers reduce the spatial resolution of the input and compute representative feature maps for the task at hand, while the decoder increases the spatial resolution, preserving the representative information for precise segmentation. Skip connections between these two layers facilitate information flow and enhance the learning process. During the training process, the U-Net parameters are continuously updated, allowing the model to implicitly learn the essential geometrical and textural features required for successful segmentation of the target masks.

A standardized data preparation and augmentation pipeline of the nnU-Net was used in this work. Additionally, we applied extreme data augmentation and used an ensemble of networks using fivefold cross-validation. The models were trained for a thousand epochs using the loss using a combination of focal and cross-entropy loss. Figure 3 shows the csPCa segmentation models. Further details regarding the model can be found in [24].

We split the in-house data into testing and transfer learning samples by ~ 90%/10%. First, the model trained on the challenge data was used on the in-house testing set without training. Then, we fine-tuned (i.e., transfer learning) the model with a learning rate of 10^{-5} for 100 epochs and tested its performance on the testing set of the in-house data.

Performance evaluation and statistical analyses

The statistical analyses were performed using the SciPy library of the Python programming language. The continuous variables are presented using median and interquartile ranges, and the categorical and ordinal variables are presented with frequencies and percentages. We used the area under the receiver operating characteristic (AUROC) curve to estimate patient-level performance in detecting csPCa.

The lesion-level detection performance was evaluated using the Average precision (AP) metric. True-positive

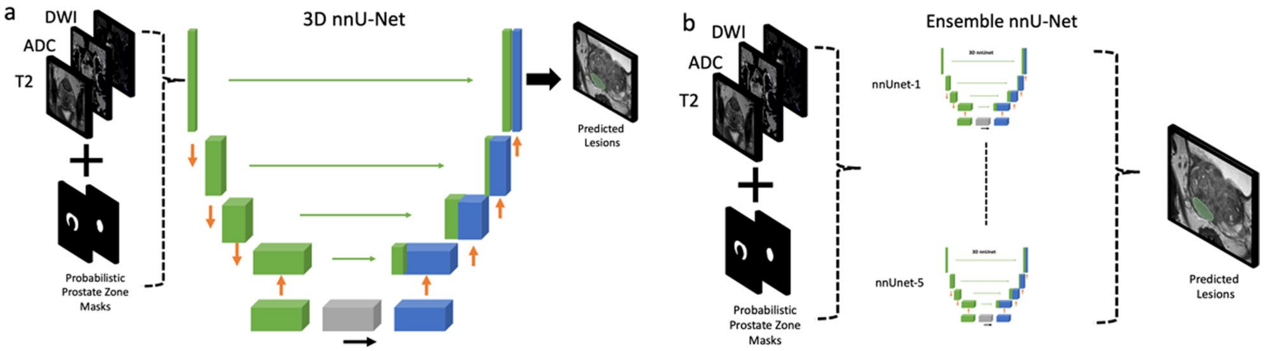


Fig. 3 The 3D nnU-Net model for detecting clinically significant prostate cancer. **a** The 3D nnU-Net was fed with T2W imaging, diffusion-weighted imaging, and apparent diffusion coefficient maps along with probabilistic prostate masks via five different channels. The model was trained on the publicly available Prostate Imaging: Cancer AI training data using the significant cancer masks provided by the organizers as the ground truth. **b** The 3D nnU-Net model was trained using a fivefold cross-validation approach. Then, the ensemble of five nnU-Net models was used to make the final predictions

lesions were the predictions that shared a minimum overlap of 0.10 in 3D space with the ground-truth annotation following the challenge and earlier studies [20, 25]. False-positive lesions were the predictions without a sufficient overlap. We also calculated the free-receiver operating curves (FROC).

We used the permutation test [26] to assess the performance difference of the DL model on the in-house testing data with and without transfer learning. The permutation test briefly shuffles performance metrics across the model with and without fine-tuning and their instances, accounting for potential differences stemming from the training method. A p value less than 0.05 was accepted as showing a significant result.

Results

A total of 1202 men were enrolled in the in-house dataset with a median age of 67 years (IQR, 59–73). The in-house dataset was split into two parts: (i) testing data consisting of 1036 scans; (ii) fine-tuning data consisting of 200 scans. In all, 288 scans in the in-house testing data had csPCa, while the remaining scans had indolent cancer or benign findings. Among 288 scans with csPCa in the in-house testing data, 275 (95.48%) had available whole-mount pathology, while the diagnosis of csPCa was made by a combination of 3–4-core MRI/ultrasound fusion-guided biopsy followed by an extended transrectal systematic biopsy in remaining 13 scans (4.52%). Of the 1,500 mpMRI scans from the PI-CAI public training data, 425 had csPCa. Further details regarding the datasets are shown in Table 1.

The nnU-Net model achieved an AUROC of 0.888 and AP of 0.732 on the hidden validation data of the PI-CAI challenge, being the 1st on the leaderboard at the time of submission. The same model had an AUROC of 0.889 and an AP of 0.614 on the hidden testing data and ranked 3rd on the leaderboard. Since the validation and testing datasets were hidden, we could not draw ROC, FROC, and RP curves for the challenge data.

The same nnU-Net model provided an AUROC of 0.886 and AP of 0.50 on the testing part of the in-house data without fine-tuning. The AUROC of the model was similar to the challenge evaluations, yet AP showed a drop in the in-house data. After fine-tuning the model with transfer learning, the AP slightly increased to 0.539, and AUROC slightly decreased to 0.870, yet the changes were not statistically significant on average ($p = 0.30$). Figure 4 shows the AUROC, FROC, and RP curves of the nnU-Net with and without transfer learning on the in-house testing data. Figure 5 exemplifies the prediction of the model on the in-house testing data.

Discussion

In this study, we trained a state-of-the-art self-adapting nnU-Net model using extensive data augmentations and probabilistic prostate masks on the large-scale PI-CAI

Table 1 The demographics, clinical, and imaging characteristics of the PI-CAI training and in-house data set

Variables	PI-CAI training data	PI-CAI validation data	PI-CAI testing data	In-House testing data	In-house fine-tuning data
# of Patients	1476	100	1000	1002	200
# of Scans	1500	100	1000	1036	200
Age (years)	66 (61–70)	NA	NA	68 (59–73)	67 (58–70)
Prostate Specific Antigen (ng/mL)	8.5 (6–13)	NA	NA	9.2 (5–12)	8.2 (6–10)
# of different MRI scanners	5 Siemens, 2 Philips	6 Siemens, 3 Philips	6 Siemens, 3 Philips	6 Siemens, 2 GE	6 Siemens, 2 GE
# of Centers	3	3	3	9	9
PI-RADS Category of Positive MRI lesions 3 4 5	246 438 403	NA	NA	308 188 240	63 40 46
# of Scans with Benign or Indolent PCa	1075	NA	NA	NA	150
# of Scans csPCa the	425	NA	NA	288	50
# of csPCa Lesions	465	NA	NA	342	52

All continuous variables are presented with median and interquartile ranges

csPCa clinically significant prostate cancer, *PI-CAI* prostate imaging: cancer AI, *PI-RADS* Prostate Imaging-Reporting and Data System

*The PI-CAI training data were obtained from Radboud University Medical Center, Ziekenhuisgroep Twente, University Medical Center Groningen, Norwegian University of Science and Technology. In-house data were obtained from Acibadem Mehmet Ali Aydinlar University's Maslak Hospital, Altunizade Hospital, Atakent Hospital, Adana Hospital, Taksim Hospital, Kozyatagi Hospital, Kocaeli Hospital, Bodrum Hospital, Eskisehir Hospital

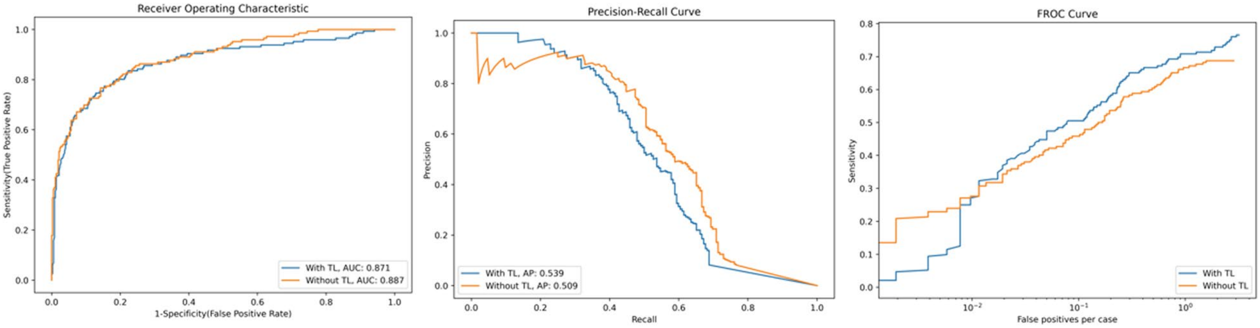


Fig. 4 The AUROC, FROC, and PR curves of the nnU-Net in detecting clinically significant prostate cancer with and without transfer learning. The area under the receiver operating characteristic (AUROC), Free-Response Receiver Operating Characteristic (FROC), and Precision–Recall (PR) curves of the ensemble of five nnU-Net models in detecting clinically significant prostate cancer in the in-house dataset with and without transfer learning. The AUROC and FROC slightly decreased, and average precision slightly increased using transfer learning, not reaching a statistical significance

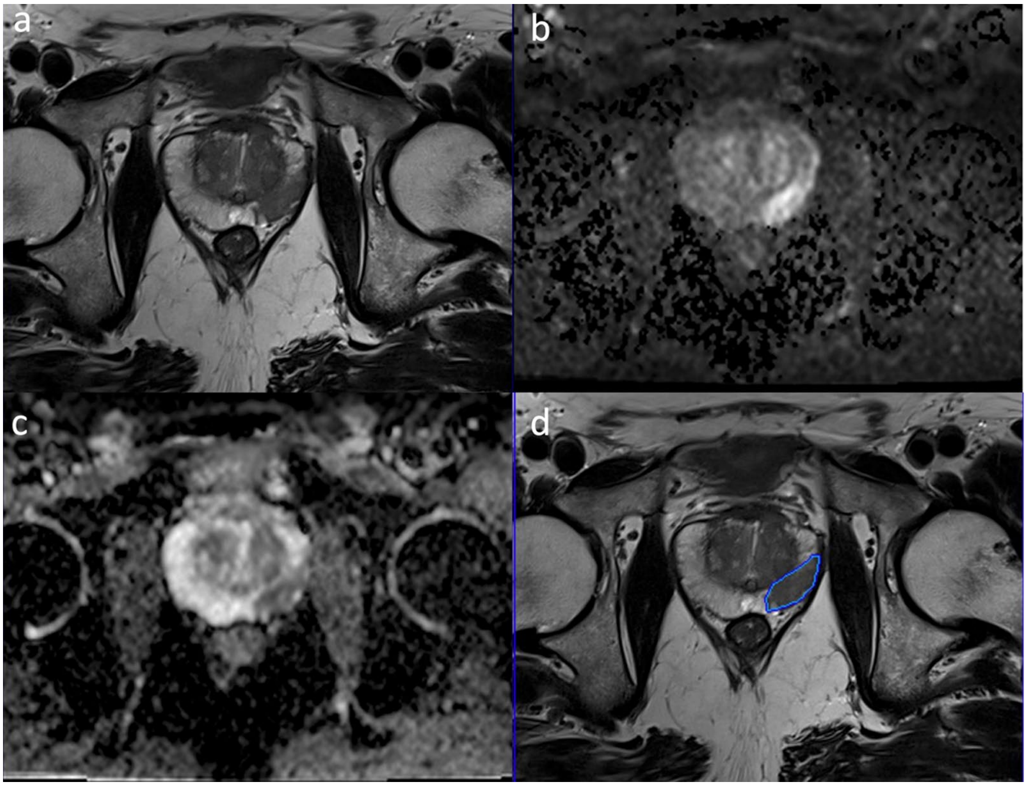


Fig. 5 A patient with clinically significant prostate cancer in the right peripheral zone from the in-house data. The T2W (**a**), diffusion-weighted image with a b-value of 1400 s/mm^2 (**b**), apparent diffusion coefficient map (**c**), and the predictions of the deep learning model overlaid on the T2W image (**d**). The model correctly predicted the lesion and drew its borders

data and reported its performance on the hidden validation and testing sets of the challenge. The model was externally validated on our large-scale multi-center & multi-vendor in-house data, which provided a similar performance in detecting csPCa at the scan level, showing its robustness and generalizability. Notably, transfer learning did not further increase the performance of the model, substantiating its generalizability and robustness against the data shift. Notably, the performance of our model was much higher than the reported median AUC of 0.79 in identifying csPCa in earlier studies [27].

Using testing data from the same data source with the training data, even in the presence of hold-out test sets with temporal split, is a known pitfall in DL applications to medical imaging [28]. Naturally, the performance of DL models evaluated on the same distribution can dramatically degrade when applied to external data due to many factors, including differences in the scanner, acquisition protocols, or patient demographics. Regardless of the cause, appropriate external testing is of utmost importance for performance estimation since clinical translation naturally requires similar diagnostic performance on unseen external data.

Apart from the present work, few other studies have investigated, at least partially, the performance of their DL model on unseen external test data. For example, Castillo et al. [29] trained their in-house model on a single-center data of 271 patients and tested its performance on 371 patients from three external datasets. The authors documented a significant drop in the performance during the external testing of the model. The DL model used in their work was a standard U-net segmentation model, while we implemented a state-of-the-art nnU-Net. Further, their training data was relatively small and derived from a single center. So, during the training, the model might not reach adequate robustness and generalizability. Additionally, in this work, we implemented extensive data augmentations along with probabilistic prostate masks to enhance the generalizability and robustness of the DL models.

Hosseinzadeh et al. [30] designed their DL models on data consisting of 2734 consecutive biopsy-naïve men derived from two centers. Similar to the present work, they used prostate masks to guide the neural nets. They trained their model on the data of the first center and tested its performance on the data of both centers. Hence, their test was not entirely external. Regardless, the DL model achieved an AUC of 0.85 in identifying csPCa in the external test. In contrast to the present work, the authors did not benchmark their models' performance or publish their codes.

Mehta et al. [31] proposed to design a neural network that takes the entire prostate gland with scan-level ground-truth labels. The authors externally tested the performance of their DL model on two different datasets, yielding an AUC of 0.73 and 0.77, which were much lower than the performance of our model. Their follow-up [32] study included lesion-level annotations and zonal prostate masks. Despite lesion-level annotations, their model yielded an AUC of 0.70 on the external testing set, a significant drop from their internal model AUC of 0.85, suggesting the lack of generalizability and robustness across different data distributions. We suggest that a small training sample size, as also suggested by the authors, and the use of standard 2D U-net without extensive data augmentations might lead to low performance in identifying csPCa on the external data.

Netzer et al. [33] trained a nnU-Net on large-scale single-center in-house data. The authors split their in-house data temporally and achieved an AUC of 0.85 on a scan level. The authors also found that the performance of the model decreased from an AUC of 0.85 to 0.81 with the reduced training data size. Further, they benchmarked their models on the ProstateX challenge and achieved an AUC of 0.89. The main drawback of their study was that the study sample and ProstateX were obtained with scanners of the same manufacturer, presumably degrading the generalizability of the results. Indeed, the authors expressed their concerns about the small sample size and the abundance of potentially easier examples in the ProstateX data.

Saha et al. [34] designed an end-to-end csPCa detection network at a large scale. Following best practices, the authors derived training and testing samples from different centers. The authors implemented a 3D U-net leveraging ensemble method, focal loss, and probabilistic prostate zone masks, achieving an AUC of 0.86 on the external independent testing sample. Notably, the authors observed that the use of the ensemble method and probabilistic masks significantly boosted the generalizability of the model. The main drawback of the Saha et al. [34] was that the testing data set was obtained with the same manufacturer's scanners as the training set.

Several limitations to the present work must be acknowledged. First, similar to the PI-CAI challenge, all patients without csPCa did not have a histopathology result in the present work. Likewise, some patients with csPCa had only a biopsy result, which is subject to errors compared with the reference whole-mount pathology. Nevertheless, all biopsies were MRI/ultrasound fusion-guided biopsies followed by an extended transrectal systematic biopsy, reducing the chances of potential sampling errors [16]. Further, we aimed to cover the scans encountered during clinical practice where a DL model needs to interpret many MRI-negative scans, which will not routinely undergo a biopsy.

Second, only MRI-visible csPCa were used for the training and model evaluation. Hence, readers must exercise caution that the DL models used in this work might miss MRI invisible lesions. This limitation might be mitigated by the registration of whole-mount pathology and MRI [35], yet it would significantly reduce the sample size and lead to selection bias by enrolling only patients with whole-mount pathology.

Third, we omitted contrast-enhanced sequences like the PI-CAI challenge. Despite mpMRI being the standard protocol for prostate imaging, evidence has recently emerged showing that the bi-parametric MRI is on par with the multi-parametric one, averting contrast usage and saving time [36, 37]. In a similar vein, we did not include clinical and laboratory findings in our DL models. Thus, creating a DL-based nomogram using prostate MRI along with clinical and laboratory for csPCa might be sought after in the future [38].

Fourth, we did not compare the performance of our DL model with that of radiologists or investigate its benefits to radiologists in reading prostate MRI. Regardless, the PI-CAI challenge organizers plan to compare the performance of the top-ranked models, including the model used in this study, with many radiologists with different experience levels worldwide. Likewise, we also plan to build a browser-based pipeline and invite radiologists with different levels of experience nationwide to read cases of our in-house dataset with and without the DL models. This will allow us to compare our model performance with radiologists in detecting csPCa and show whether it adds value to the readings of radiologists in terms of confidence, accuracy, and effectiveness.

Conclusions

The state-of-the-art DL model trained using extensive data augmentations and probabilistic prostate masks trained on the large-scale PI-CAI data provided high performance in detecting csPCa on the hidden validation and testing sets of the challenge and large-scale multi-center and multi-vendor in-house data consisting of men with different demographics, showing its robustness and generalizability within and across datasets. Notably, implementing transfer learning using a small sample from the in-house data did not further improve the performance, supporting its generalizability and robustness against data shift.

Abbreviations

AP	Average precision
AUROC	The area under the receiver operating curve
csPCa	Clinically significant PCa
DL	Deep learning
FROC	Free-receiver operating curve
MRI	Magnetic resonance imaging
nnU-Net	No-new U-Net
PCa	Prostate cancer
PI-RADS	Prostate Imaging-Reporting and Data System
PR	Precision–recall curve

Author contributions

Each author has made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data. Each author has approved the submitted version (and any substantially modified version that involves the author's contribution to the study). Each author has agreed to be personally accountable for the author's contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. AK: manuscript writing, experiments. DA: manuscript writing, study design, literature Review. MY: experiments. MES: data curation, literature review. GZ: data curation. IO: critical review, manuscript editing. EK: critical review, literature review. All authors read and approved the final manuscript.

Funding

This paper has been produced benefiting from the 1001 Science and Technology Grant Program National Program of TUBITAK (Project No: 122E022). This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C353). However, the entire responsibility of the publication/paper belongs to the authors of the paper. The financial support received from TUBITAK does not mean that the content of the publication is approved in a scientific sense by TUBITAK.

Availability of data and material

The in-house data used in this study is not available. However, the codes used in the present work can be found at https://github.com/ahmetkrgztr/HeviAI_picai. The weights of the models trained on the PI-CAI training data will be shared with the researchers in the upcoming months by the organizers of the challenge.

Declarations

Ethics approval and consent to participate

The local ethics committee approved this retrospective study and waived the need for informed consent for the retrospective evaluation of anonymized medical data (Acıbadem University and Acıbadem Healthcare Institutions Medical Research Ethics Committee, No. 20220312).

Consent for publication

Not applicable.

Competing interests

AK is an employee of Hevi AI Health Tech. DA is the CEO and co-founder of Hevi AI Health Tech. MY is the chief AI scientist and co-founder of Hevi AI Health Tech. IO is on the advisory board of Hevi AI Health Tech. None of Hevi AI's products were used or mentioned in the current work. Furthermore, this paper did not use any commercially available deep-learning software. The remaining authors declare that they have no competing interests.

Author details

[1]Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey. [2]Artificial Intelligence and Information Technologies, Hevi AI Health, Istanbul, Turkey. [3]Department of Radiology, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey. [4]School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey.

Received: 24 January 2023 Accepted: 17 April 2023
Published online: 19 June 2023

References

1. Ahmed HU, El-Shater Bosaily A, Brown LC et al (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 389:815–822. https://doi.org/10.1016/S0140-6736(16)32401-1
2. Turkbey B, Rosenkrantz AB, Haider MA et al (2019) Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. Eur Urol 76:340–351. https://doi.org/10.1016/j.eururo.2019.02.033

3. Sonn GA, Fan RE, Ghanouni P et al (2019) Prostate magnetic resonance imaging interpretation varies substantially across radiologists. Eur Urol Focus 5:592–599. https://doi.org/10.1016/j.euf.2017.11.010
4. Smith CP, Harmon SA, Barrett T et al (2019) Intra-and interreader reproducibility of PI-RADSv2: a multireader study. J Magn Reson Imaging 49:1694–1703. https://doi.org/10.1002/jmri.26555
5. Westphalen AC, McCulloch CE, Anaokar JM et al (2020) Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the society of abdominal radiology prostate cancer disease-focused panel. Radiology 296:76–84. https://doi.org/10.1148/radiol.2020190646
6. Akatsuka J, Yamamoto Y, Sekine T et al (2019) Illuminating clues of cancer buried in prostate MR image: deep learning and expert approaches. Biomolecules 9:673. https://doi.org/10.3390/biom9110673
7. Aldoj N, Lukas S, Dewey M, Penzkofer T (2020) Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. Eur Radiol 30:1243–1253. https://doi.org/10.1007/s00330-019-06417-z
8. Alkadi R, Taher F, El-baz A, Werghi N (2019) A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. J Digit Imaging 32:793–807. https://doi.org/10.1007/s10278-018-0160-1
9. Ayyad SM, Badawy MA, Shehata M et al (2022) A new framework for precise identification of prostatic adenocarcinoma. Sensors 22:1848. https://doi.org/10.3390/s22051848
10. Gunashekar DD, Bielak L, Hägele L et al (2022) Explainable AI for CNN-based prostate tumor segmentation in multi-parametric MRI correlated to whole mount histopathology. Radiat Oncol 17:65. https://doi.org/10.1186/s13014-022-02035-0
11. Liu Y, Zheng H, Liang Z et al (2021) Textured-based deep learning in prostate cancer classification with 3T multiparametric MRI: comparison with PI-RADS-based classification. Diagnostics 11:1785. https://doi.org/10.3390/diagnostics11101785
12. Cao R, Mohammadian Bajgiran A, Afshari Mirak S et al (2019) Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. IEEE Trans Med Imaging 38:2496–2506. https://doi.org/10.1109/TMI.2019.2901928
13. Mehralivand S, Yang D, Harmon SA et al (2022) A cascaded deep learning-based artificial intelligence algorithm for automated lesion detection and classification on biparametric prostate magnetic resonance imaging. Acad Radiol 29:1159–1168. https://doi.org/10.1016/j.acra.2021.08.019
14. de Vente C, Vos P, Hosseinzadeh M et al (2021) Deep learning regression for prostate cancer detection and grading in bi-parametric MRI. IEEE Trans Biomed Eng 68:374–383. https://doi.org/10.1109/TBME.2020.2993528
15. Youn SY, Choi MH, Kim DH et al (2021) Detection and PI-RADS classification of focal lesions in prostate MRI: performance comparison between a deep learning-based algorithm (DLA) and radiologists with various levels of experience. Eur J Radiol 142:109894. https://doi.org/10.1016/j.ejrad.2021.109894
16. Schelb P, Kohl S, Radtke JP et al (2019) Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. Radiology 293:607–617. https://doi.org/10.1148/radiol.2019190938
17. Sanford T, Harmon SA, Turkbey EB et al (2020) Deep-learning-based artificial intelligence for PI-RADS classification to assist multiparametric prostate MRI interpretation: a development study. J Magn Reson Imaging 52:1499–1507. https://doi.org/10.1002/jmri.27204
18. Winkel DJ, Wetterauer C, Matthias MO et al (2020) Autonomous detection and classification of PI-RADS lesions in an MRI screening population incorporating multicenter-labeled deep learning and biparametric imaging: Proof of concept. Diagnostics. https://doi.org/10.3390/diagnostics10110951
19. Litjens G, Debats O, Barentsz J et al (2014) Computer-aided detection of prostate cancer in MRI. IEEE Trans Med Imaging 33:1083–1092. https://doi.org/10.1109/TMI.2014.2303821
20. Saha A, Twilt JJ, Bosma JS et al (2022) Artificial intelligence and radiologists at prostate cancer detection in MRI: the PI-CAI challenge (Study Protocol). https://doi.org/10.5281/ZENODO.6667655
21. Samaratunga H, Montironi R, True L et al (2011) International society of urological pathology (ISUP) consensus conference on handling and staging of radical prostatectomy specimens. Working group 1: specimen handling. Mod Pathol 24:6–15. https://doi.org/10.1038/modpathol.2010.178
22. Cuocolo R, Stanzione A, Castaldo A et al (2021) Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset. Eur J Radiol 138:109647. https://doi.org/10.1016/j.ejrad.2021.109647
23. Isensee F, Jaeger PF, Kohl SAA et al (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
24. Karagoz A, Seker E, Yergin M, Kan AT, Kartal MS, Karaarslan E, Alis D, Oksuz I Prostate Lesion Estimation using Prostate Masks from Biparametric MRI" to be published in Arxiv. https://rumc-gcorg-p-public.s3.amazonaws.com/evaluation-supplementary/642/88bd85e5-ca10-4979-875e-eba7476491d3/Hevi_picai_report.pdf. Accessed 20 Jan 2023
25. Duran A, Dussert G, Rouvière O et al (2022) ProstAttention-Net: a deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. Med Image Anal 77:102347. https://doi.org/10.1016/j.media.2021.102347
26. Bosma J, Saha A, Hosseinzadeh M et al (2022) Annotation-efficient cancer detection with report-guided lesion annotation for deep learning-based prostate cancer detection in bpMRI. arXiv preprint
27. Castillo TJM, Arif M, Niessen WJ et al (2020) Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. Cancers (Basel) 12:1606. https://doi.org/10.3390/cancers12061606
28. Varoquaux G, Cheplygina V (2022) Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med 5:48. https://doi.org/10.1038/s41746-022-00592-y
29. Castillo TJM, Arif M, Starmans MPA, et al. (2021) Classification of clinically significant prostate cancer on multi-parametric MRI: a validation study comparing deep learning and radiomics. Cancers (Basel) 14(1):12. https://doi.org/10.3390/cancers14010012
30. Hosseinzadeh M, Saha A, Brand P et al (2022) Deep learning–assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge. Eur Radiol 32:2224–2234. https://doi.org/10.1007/s00330-021-08320-y
31. Mehta P, Antonelli M, Ahmed HU et al (2021) Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: a patient-level classification framework. Med Image Anal 73:102153. https://doi.org/10.1016/j.media.2021.102153
32. Mehta P, Antonelli M, Singh S et al (2021) AutoProstate: towards automated reporting of prostate MRI for prostate cancer assessment using deep learning. Cancers (Basel) 13:6138. https://doi.org/10.3390/cancers13236138
33. Netzer N, Weißer C, Schelb P et al (2021) Fully automatic deep learning in bi-institutional prostate magnetic resonance imaging: effects of cohort size and heterogeneity. Invest Radiol 56:799–808. https://doi.org/10.1097/RLI.0000000000000791
34. Saha A, Hosseinzadeh M, Huisman H (2021) 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 73:102155. https://doi.org/10.1016/j.media.2021.102155
35. Bhattacharya I, Seetharaman A, Kunder C et al (2022) Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: an MRI-pathology correlation and deep learning framework: CorrSigNIA: an MRI-pathology correlation and deep learning framework. Med Image Anal. https://doi.org/10.1016/j.media.2021.102288
36. Brancato V, di Costanzo G, Basso L et al (2020) Assessment of DCE utility for PCa diagnosis using PI-RADS v.21: effects on diagnostic accuracy and reproducibility. Diagnostics 10:164. https://doi.org/10.3390/diagnostics10030164
37. Bass EJ, Pantovic A, Connor M et al (2021) A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. Prostate Cancer Prostatic Dis 24:596–611. https://doi.org/10.1038/s41391-020-00298-w

38. Hiremath A, Shiradkar R, Fu P et al (2021) 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 3:e445–e454. https://doi.org/10.1016/S2589-7500(21)00082-0

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.