

RESEARCH

Open Access



# Comparison of data fusion strategies for automated prostate lesion detection using mpMRI correlated with whole mount histology

Deepa Darshini Gunashekar<sup>1,3\*</sup> , Lars Bielik<sup>1,3</sup> , Benedict Oerther<sup>2</sup> , Matthias Benndorf<sup>2</sup> , Andrea Nedelcu<sup>2</sup> , Samantha Hickey<sup>1</sup> , Constantinos Zamboglou<sup>2,4</sup> , Anca-Ligia Grosu<sup>2,3†</sup>  and Michael Bock<sup>1,3†</sup> 

## Abstract

**Background** In this work, we compare input level, feature level and decision level data fusion techniques for automatic detection of clinically significant prostate lesions (csPCa).

**Methods** Multiple deep learning CNN architectures were developed using the Unet as the baseline. The CNNs use both multiparametric MRI images (T2W, ADC, and High b-value) and quantitative clinical data (prostate specific antigen (PSA), PSA density (PSAD), prostate gland volume & gross tumor volume (GTV)), and only mp-MRI images ( $n = 118$ ), as input. In addition, co-registered ground truth data from whole mount histopathology images ( $n = 22$ ) were used as a test set for evaluation.

**Results** The CNNs achieved for early/intermediate / late level fusion a precision of 0.41/0.51/0.61, recall value of 0.18/0.22/0.25, an average precision of 0.13 / 0.19 / 0.27, and F scores of 0.55/0.67/ 0.76. Dice Sorensen Coefficient (DSC) was used to evaluate the influence of combining mpMRI with parametric clinical data for the detection of csPCa. We compared the DSC between the predictions of CNN's trained with mpMRI and parametric clinical and the CNN's trained with only mpMRI images as input with the ground truth. We obtained a DSC of data 0.30/0.34/0.36 and 0.26/0.33/0.34 respectively. Additionally, we evaluated the influence of each mpMRI input channel for the task of csPCa detection and obtained a DSC of 0.14 / 0.25 / 0.28.

**Conclusion** The results show that the decision level fusion network performs better for the task of prostate lesion detection. Combining mpMRI data with quantitative clinical data does not show significant differences between these networks ( $p = 0.26/0.62/0.85$ ). The results show that CNNs trained with all mpMRI data outperform CNNs with less input channels which is consistent with current clinical protocols where the same input is used for PI-RADS lesion scoring.

<sup>†</sup>Anca-Ligia Grosu and Michael Bock contributed equally to this work.

\*Correspondence:

Deepa Darshini Gunashekar  
deepa.darshini.gunashekar@uniklinik-freiburg.de

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Trial registration** The trial was registered retrospectively at the German Register for Clinical Studies (DRKS) under proposal number Nr. 476/14 & 476/19.

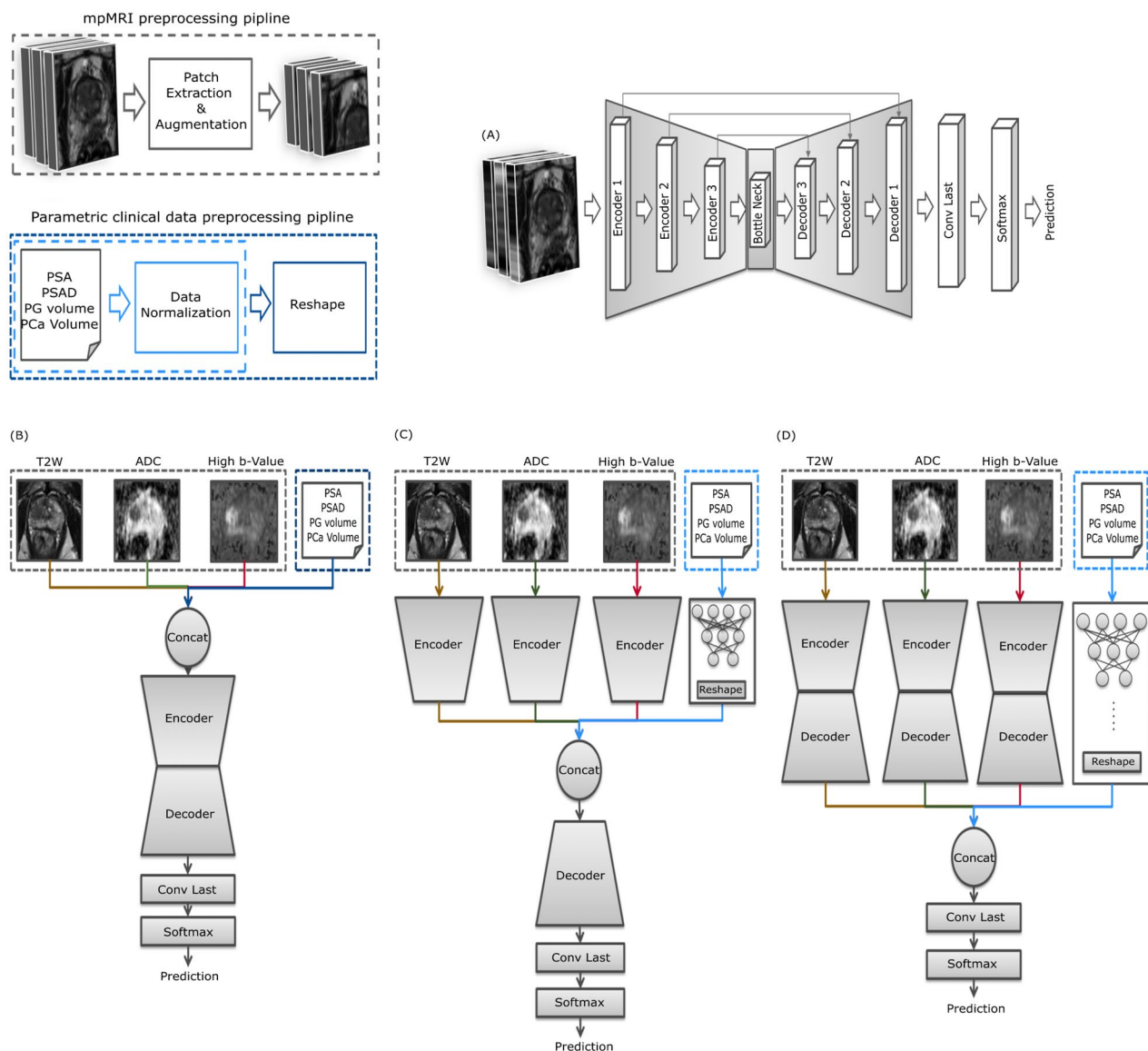
**Keywords** Multiparametric MRI (mpMRI), Prostate specific Antigen (PSA), PSA density (PSAD), Gleason Score, Convolutional neural network, Data fusion, Early Fusion, Intermediate Fusion, Late Fusion, Automated prostate lesion detection, Histological validation

## Introduction

In clinical practice, the detection and management of clinically significant prostate cancer (csPCa) often involves a combination of different diagnostic tests and imaging protocols. For diagnosis, staging, and therapy planning, clinicians typically rely on findings from a digital rectal examination (DRE) [1], a prostate-specific antigen (PSA) test and, additionally, the Prostate Imaging Reporting and Data System (PI-RADS) score derived from multiparametric MRI (mpMRI) [2]. To confirm the diagnosis of csPCa patients are required to undergo targeted and systematic biopsy. During this procedure, tissue samples are harvested and analyzed histologically, and a Gleason score is assigned to each sample, from which the lesions can be classified as clinical significant (Gleason score 7 and above) for further treatment. This parametric clinical data, i.e. PSA [3–6] and Gleason score [7, 8], is combined with image-derived parametric information, such as the prostate volume [8–11], lesion volume, and PSA density [12–16] to retrieve clinically relevant and reliable predictions. However, the multimodal data are not directly correlated among each other [17, 18], due to the different data modalities, and various acquisition techniques. Additionally the variability of the data and the modelling techniques as well as data privacy issues have made it challenging to develop medical multimodal data fusion models [19]. Hence, an analysis of the available image fusion strategies applied to medical imaging and parametric clinical data is urgently needed.

Recently, deep-learning-based multimodal data fusion has gained significant interest in the medical community [20], as the combination of diverse data from various modalities can aid the decision process of a convolutional neural network (CNN) [20–22]. A CNN is a type of AI algorithm used to recognize patterns and features in images, such as shapes, colors, edges and textures to make decisions or predictions. It works by passing the image through multiple layers of filters that help identify important details and combine them to understand the overall image. Detailed information on how the CNN process image data can be found in [23–25]. In deep learning, multimodal data fusion refers to fusing data from various imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), histology images and non-imaging modalities, such as clinical data from electronic health records, to assist in the decision-making

process of CNNs. In the literature, three data fusion techniques have been proposed. (1) In early fusion or input level fusion (EF) (Fig. 1B), the CNN learns a fused feature representation by combining various imaging data and clinical data for decision making. The data from each modality corresponds to one channel in the multi-channel input. (2) In the intermediate fusion or feature level fusion (IF) (Fig. 1C) each channel is passed through multiple CNN layers for feature extraction. These features are then fused and processed further as input to deeper layers of the CNN for decision making. The connection among the early layers enables the CNN to learn the unique feature representation of the corresponding modality, where as the connections from intermediate layers enables the CNN to capture complex relationships between the input modalities, by fully exploiting the feature representation of multimodal images for decision making. (3) In the late fusion method or decision level fusion (LF) (Fig. 1D) the data from each modality is used as input to train independent CNNs. This allows the CNNs to exploit the unique feature representation of the corresponding modality. The outputs of the individual CNNs are then fused and the final prediction is obtained via mean aggregation or by majority voting [26, 27]. Suresh, Harini, et al. [28] used the EF fusion technique to predict onset and weaning of multiple invasive interventions, by integrating data from all available ICU sources (vitals, labs, notes, demographics), Park, Chihyun and colleagues [29] used EF method for prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset. Peng, Chen, et al. [30] used the EF fusion technique to model features based on the capsule network, to identify breast cancer-related genes. Lee, Garam, et al. [31] used the IF approach to predicting Alzheimer's disease progression using multi-modal deep learning approach. Huang, Zhi, et al. [32] for survival prediction for breast cancer. Islam, Md Mohaiminul, et al. [33] used the IF strategy for classification of molecular subtypes of breast cancer. Poirion, Olivier B and coauthors [34] used IF for risk stratification of bladder cancer. Huang and colleagues [35] researched the optimal multimodal fusion strategy for PCa detection using 2D axial T2w imaging and apparent diffusion coefficient (ADC) imaging as the inputs for their multi-modal fusion model to identify a pipeline that works best for automated diagnosis of csPCa. Reda I, et al. [36] presented a noninvasive CAD system using a meta classifier that integrates PSA



**Fig. 1** An overview of the mpMRI & parametric clinical data preprocessing pipeline. **(A)** UNet baseline architecture along with an overview of **(B)** the EF, **(C)** IF and **(D)** LF architectures used in this work

screening results in addition to Diffusion weighted MRI based features for prostate cancer diagnosis using the LF technique. Hiremath A, et al [37] implemented a clinical nomogram combining deep learning-based imaging predictions, PI-RADS score, and the clinical data PSA, prostate volume and lesion volume, using multivariate logistic regression to identify csPCa in bi-parametric MRI. They showed the integrated nomogram could help for risk stratification by identifying patients with very low risk, low risk, and intermediate risk of csPCa for active surveillance and very high-risk patients who might benefit from adjuvant therapy using LF.

In this study, we compare three multimodal fusion strategies for automatic clinically significant prostate

cancer (csPCa) lesion detection and segmentation. For this, we use mpMRI images (T2W, ADC, high b-value) and the corresponding parametric clinical data (PSA, gross tumor volume, gross prostate volume, and PSA density) as input to train EF, IF and LF networks as shown in Fig. 1(B-D), using an architecture that is based on a 3D-Unet [38]- Fig. 1(A). Specifically, we compare the csPCa detection of a CNN trained with only mpMRI data to a CNN trained with both mpMRI and non-imaging data. Additionally, we compare the lesion segmentation of the network to a commercial deep learning algorithm with EF.

**Table 1** Overview of the median along with minimum and maximum age, PSA, PSAD, Prostate gland volume and the prostate lesion volume across the two patient cohorts

Median (Min-Max)	Irradiation & Prostatectomy ( $n_{irr+prost}$ )	Whole mount Histology ( $n_{prost}$ )
Age	74 (56–85)	64 (48–76)
PSA (ng/mL)	8.8 (2.1–66)	17.4 (6.07–218)
PSAD (mL or cc <sup>3</sup> )	0.2 (0.1–2.7)	0.47 (0.15–1.67)
Prostate Volume (mL or cc <sup>3</sup> )	40.8 (10.8–167.7)	42.6 (29.2–130)
PCa Lesion Volume (mL or cc <sup>3</sup> )	1.8 (0.0–38.1)	11.9 (0.5–76)

## Materials & methods

### Clinical data

In this study, mpMRI data along with the corresponding parametric clinical data from primary csPCa patients with histologically confirmed cancer lesions was used. The data consists of two groups, with ( $n_{prost} = 22$ ) and without ( $n_{irr+prost} = 118$ ) whole mount histology data. In the group with whole mount histology data available ( $n_{prost}$ ), patients initially underwent MRI and during the subsequent therapy the prostate gland was surgically removed (prostatectomy). All other patients underwent radiation therapy.

All MRI studies were carried out between 2008 and 2019 on a clinical 1.5T (Avanto, Aera & Symphony, Siemens, Erlangen, Germany) and 3T (Tim TRIO, Siemens, Erlangen, Germany) MRI systems. The MRI protocol consisted of pre-contrast T2-weighted turbo spin echo (TSE) images in transverse, sagittal, and coronal orientations, diffusion-weighted imaging (DWI) with an echo planar imaging sequence in transverse orientation, and dynamic contrast-enhanced (DCE) MRI images. All the images were acquired with surface phased array (body matrix) coils in combination with integrated spine array coils. The DWI data were acquired with b-values of [0, 100, 400, 800] s/mm<sup>2</sup> or [0, 250, 500, 800] s/mm<sup>2</sup> for 1.5T, and [50, 400, 800] s/mm<sup>2</sup> for 3T. From the diffusion-weighted imaging data, a synthetic high b-value image (b=1400 s/mm<sup>2</sup>) was calculated as recommended by the PI-RADS lexicon [2, 39]. To account for the varying diffusion weightings (b-values) across different field strengths, we generated synthetic DWI images with b=1400 s/mm<sup>2</sup>, with original b-values for the 1.5T system being [0, 100, 400, 800] s/mm<sup>2</sup> or [0, 250, 400, 800] s/mm<sup>2</sup>, and for the 3T system, [50, 400, 800] s/mm<sup>2</sup> [40]. While no homogenization method was applied to the T2-weighted images due to field strength-dependent tissue T1 and T2 values, we expect similar contrast in the T2-weighted TSE images from both 1.5T and 3T systems, given the comparable T2-values in a wide range of human tissues and the use of repetition times exceeding 5500 ms to minimize T1 contrast [41]. The parametric clinical data consisted of initial PSA values, PSA density, gross prostate gland

**Table 2** Overview of the number of patients with various Gleason grades from the two patient cohorts

Gleason Grades	6	7a	7b	8	9
Irradiation & Prostatectomy ( $n_{irr+prost}$ )	16	35	38	20	5
Whole mount Histology ( $n_{prost}$ )	2	6	7	4	3

volume, gross lesion volume. Tables 1 and 2 provide an overview of the parametric clinical data. In addition, clinical scores were derived from the biopsy data such as the Gleason grade, the TNM status, and Gleason grade group [42]. The study was approved by the institutional ethics review board (Proposal Nr.476/14 & 476/19) and patients gave written informed consent.

Patient data was separated into a training and a test cohort. The training cohort included a large irradiation and prostatectomy group ( $n_{irr} + n_{prost} = 118$ ), as training the CNN requires a substantial dataset. Due to the limited number of patients, the test cohort consisted solely of the prostatectomy group ( $n_{prost} = 22$ ). For the test cohort, post-operative gleason score and tumor volume information were available, along with the ground truth contours from the whole organ histopathology slices, which were co-registered with pre-operative MRI data. The CNNs were trained on T2-weighted images and apparent diffusion coefficient (ADC) maps together with synthetic high b-value images (b=1400 s/mm<sup>2</sup>). For all  $n_{irr+prost} = 118$  and the  $n_{prost}=22$  in-house mpMRI data sets, the entire gland (PG-Rad), and the tumor within the prostate (PCa-Rad) were contoured by two experienced radiation oncologists during radiation therapy treatment planning. Images of the whole mount histology slices and the ground truth counters from corresponding whole mount histology slices were acquired as described in [40]. The AI-Rad Companion Prostate MR VA20A\_HF02 for biopsy support (Siemens Healthcare AG [43–45]). It performs automated segmentation and automated volume estimation of the prostate and additionally calculates the PSA density, if PSA value is known. The system was used to generate for the csPCa (Rad-AI) contours.

### CNN architecture

As the baseline architecture a patch-based 3D Unet [46] with 3 encoder blocks and 3 decoder blocks (Fig. 1A) was used. Figure 1(B-D) shows the three fusion networks used in this study. For the EF network the input layer consisted of all three mpMRI volumes and parametric clinical data. The parametric clinical data was reshaped to match the dimensions of the image volumes and concatenated with the mpMRI volumes into a single 4D volume. Each channel in the 4D volume corresponds to mpMRI (channel 1 to 3) and parametric clinical data (channel 4 to 7). The IF network consisted of 3 independent encoder heads for each mpMRI volume as input. Each encoder head had 3 encoder blocks, which extracted features from the

corresponding mpMRI volumes independently. A shallow multi-layer perceptron (MLP) with three fully connected layers was used to extract features from parametric clinical data. The features from the 3 encoder heads and the MLP head were concatenated at the bottleneck block of the UNet and processed further as input to the decoder blocks in Fig. 1B. For the decision level fusion methods, 3 UNets were trained separately for each mpMRI volume (see Fig. 1C), and an MLP with parametric clinical data as input was trained individually. The final prediction was obtained by mean aggregation of the predicted individual probability scores of these networks as depicted in Fig. 1D. All CNNs were implemented in MATLAB® (2022b, Math Works, Inc., Natick/MA).

### Data preprocessing

To reduce the computation time, the mpMRI data were cropped to a smaller FOV around the prostate. A chance of 70% for a random 2D-rotation (0-360°) in the axial plane was added for data augmentation. Due to the large sizes of the image volumes which would result in system memory issues, calculations were performed on patches of size 64×64×16 that were randomly chosen with respect to the center location of the original image. Based

on the type of fusion network these data were reshaped to match the image dimensions and concatenated along the 4th dimension.

### Training & testing

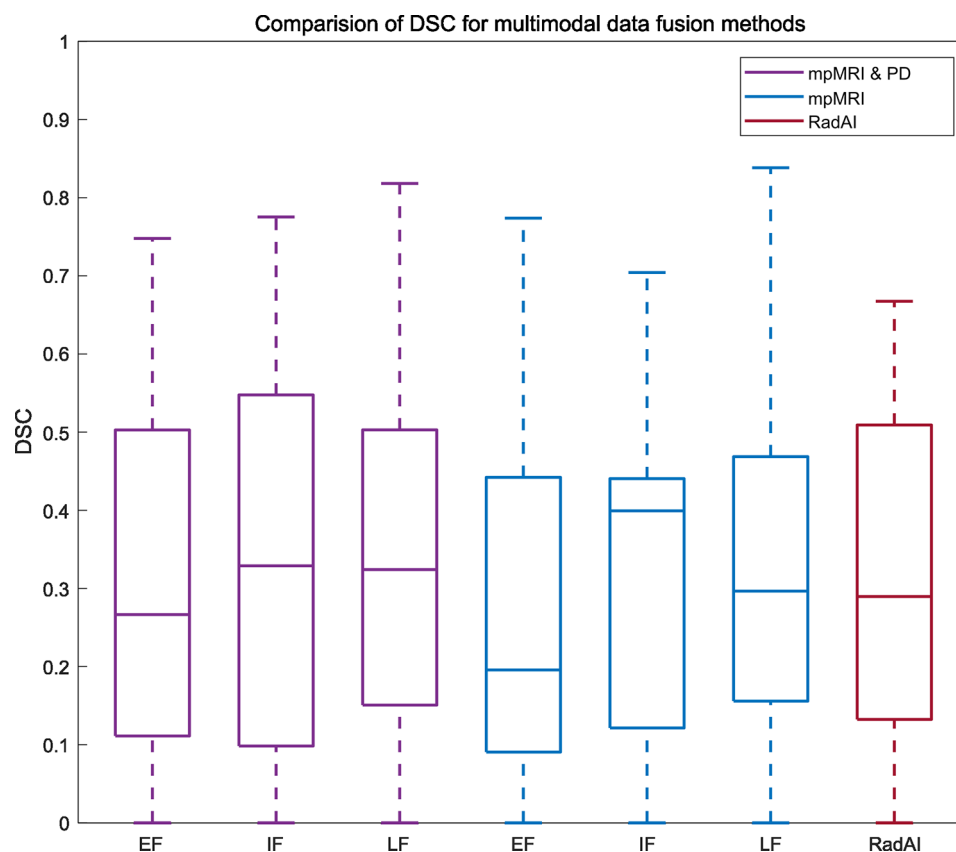
All CNNs were trained for 50 epochs on an NVIDIA RTX2080 GPU with a learning rate of 1e-5, a batch size of 4, patch size of 64×64×16, 50 patches per image, and using the Adam optimizer (Bayesian optimization). The mpMRI data from the prostatectomy group ( $n_{test}=22$ ) were used during the testing phase, as the histologic information could be used as a ground truth.

### Performance evaluation

The performance of each network over different epochs was evaluated using precision recall, average precision (AP), F-Score, and Dice Similarity coefficient (DSC). A t-test was performed to identify differences in the predicted scores for these networks.

### Results

Figure 2 illustrates the overlap between the detected lesions and the ground truth in the test set, using input level, feature level, and decision level fusion strategies.



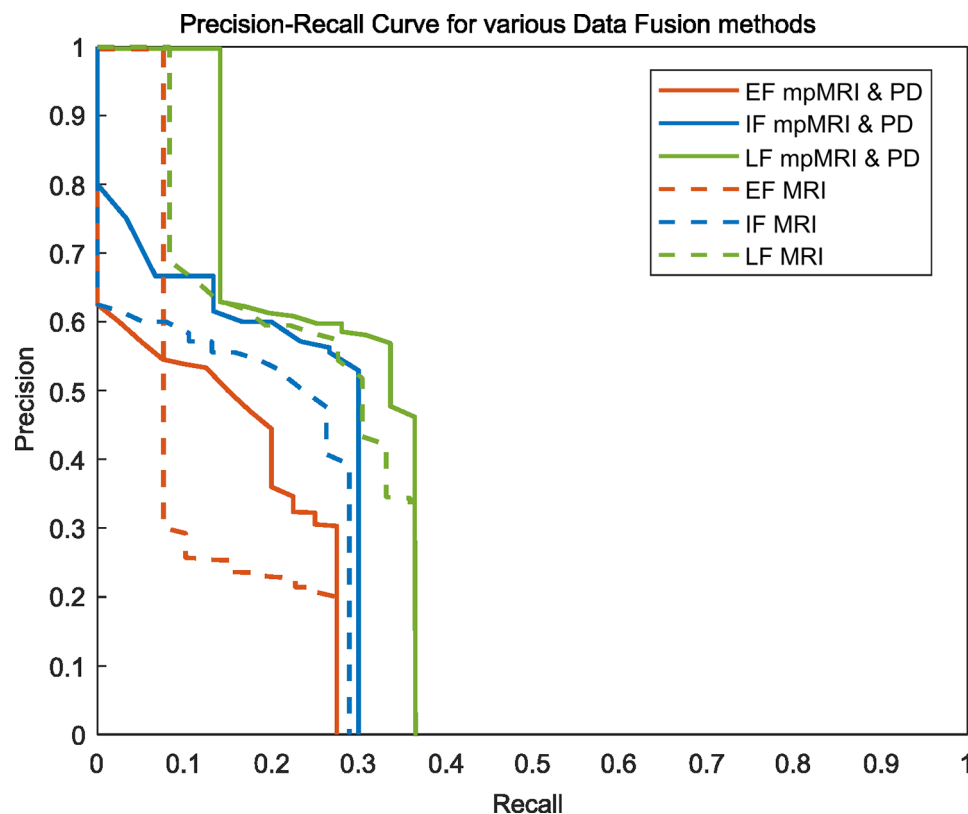
**Fig. 2** Comparison of DSC for the prediction of csPCa lesion on the test cohort ( $n=22$ ) for input level, feature level, and decision level fusion networks using mpMRI & parametric clinical data in purple, mpMRI data only in blue and with RADAI in red with the ground truth

The CNNs achieved a precision of 0.41, 0.51, 0.61 for input level, feature level, and decision level fusion, respectively, while the recall values were 0.18, 0.22, 0.25. Additionally, the AP was 0.13, 0.19, 0.27, and the F scores were 0.55, 0.67, 0.76, respectively for the fusion of image and clinical data. For networks trained only with image data, the CNNs achieved a precision of 0.36, 0.45, 0.56 for input level, feature level, and decision level fusion, respectively, while the recall values were 0.13, 0.21, 0.23. Additionally, the AP was 0.12, 0.16, 0.23, and the F scores were 0.19, 0.26, 0.33, respectively. The network's performance was further summarized using an area under the precision-recall curve (AUC-PR) in Fig. 3. A two-sided student's t-test did not reveal any significant differences between the predicted scores of any two networks ( $p=0.26, 0.62, \text{ and } 0.85$ ). In Fig. 4, the ground truth and the predicted lesion maps are displayed overlaid along with the ground truth on the corresponding input image data for three test patients from the test cohort. We show test cases with multiple lesions as well as single lesions. In patients 1 and 3 all networks detected multiple and single lesions, while in patient 2 the IF network and the RAD-AI detected a suspected lesion additionally in contrast to the ground truth. Table 3 provides an overview of the comparison of mean DSC obtained by comparing the ground truth with predicted segmentation maps

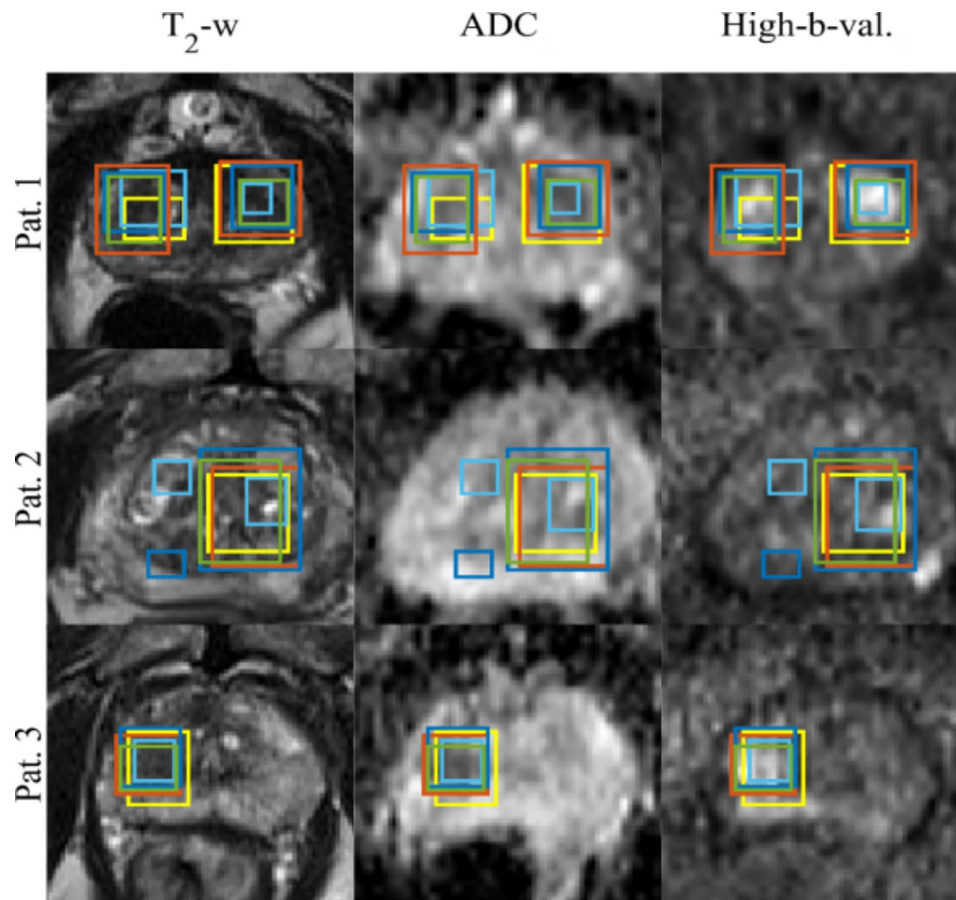
for networks trained with T2w, ADC, and High b-value images only, only with mpMRI data, mpMRI and parametric clinical data and RadAI.

## Discussion

In this study, we investigated multiple deep learning data fusion methods for the automatic detection and segmentation of csPCa lesions using a patch-based 3D UNet. The late fusion network performed better with a mean DSC of  $0.36 \pm 0.24$  when compared to the input and feature level fusion networks  $0.30 \pm 0.23$  &  $0.34 \pm 0.26$  respectively. A statistical t-test showed no significant differences in the prediction ( $p=0.51, 0.10, \text{ and } 0.82$ ). The EF network offers advantages in cost effectiveness and memory efficiency; however, this approach does not exploit the relationships between the different modalities. Rather, as it learns a joint feature representation, it simply fuses the data at the input level. The IF networks instead capture the complex relationships between the input modalities, by learning both individual and joint feature representations. However, both the EF and IF networks are not flexible to train in case of missing data. Moreover, the LF network only learns an independent feature representation, and is flexible to train such network when one or more input data modality is missing, it fails to learn the joint feature representation, since each network is trained



**Fig. 3** Precision-Recall Curve for various data fusion methods EF (orange), IF (blue), and LF (green) for networks trained with mpMRI and parametric clinical data (solid lines) and mpMRI data only (dotted lines)



**Fig. 4** Detected lesions for three patients from the test set by EF fusion network (orange), IF (blue), and LF (green), ground truth (yellow) and RAD-AI (cyan)

**Table 3** Mean DSC between the ground truth (csPCa-Histo) segmentation and the predicted segmentation (csPCa-CNN and csPCa-AI-Rad) for various networks

Network	Input Data	Fusion	Mean DSC
UNet	T2W only		0.14 ± 14
			0.25 ± 20
			0.28 ± 21
	mpMRI & parametric clinical data	EF	0.30 ± 0.23
		IF	0.34 ± 0.26
		LF	0.36 ± 0.24
mpMRI only	EF	0.26 ± 0.22	
	IF	0.33 ± 0.23	
	LF	0.34 ± 0.26	
RADAI	mpMRI only	EF	0.31 ± 0.21

with a particular input data. Though the IF and LF have a better performance; their design leads to an increased number of network parameters, which in turn increase memory consumption and training time, making it computationally expensive to train such networks [17, 47, 48].

In this work we evaluated the influence of combining parametric clinical data with mpMRI images, by comparing the predictions of the networks trained solely on

mpMRI data with networks trained on both mpMRI and parametric clinical. By performing a t-test ( $p$ -value=0.47, 0.88, 0.28), we found no significant difference in the overall prediction for csPCa detection and segmentation. Multimodal imaging alone could be adequate for training CNNs for csPCa detection and segmentation. However, the inclusion of clinical data could be beneficial for csPCa lesion risk classification.

To determine the importance of each mpMRI sequence in the tasks of lesion detection, we relied on predictions from the independently trained CNNs using T2W, ADC and High-b Value images as inputs (Table 3). For the prostate gland detection, the UNet trained with only T2W images performed the best, in comparison to the networks trained using ADC images or High b-Value images only. This is likely a result of the well-defined prostate anatomy in the T2W sequence thus validating the use of T2W images in for prostate gland & prostate zone segmentation [49–52]. For prostate lesion detection and segmentation, the networks trained with only the ADC and High b-Value images showed an 18% improvement in detection csPCa lesions on the test set, in comparison with the UNet trained with only T2W images.

We also showed that the network trained with only T2W images performed better in detecting csPCa lesions with PI-RADS score greater than 4 and 5, a DSC of 0.50, 0.49 & 0.32 and underperformed in detecting csPCa lesions with PI-RADS score 1 to 3, a DSC of 0.17, 0.09, 0.13, 0.0 respectively. We compared the AI-Rad generated segmentation mask with PCa-Histo and PCa-CNN (Early-Fusion) Table 3, indicate that our network performed similarly to RAD-AI [43–45] in segmentation of csPCa lesions.

## Conclusion

In this study, mpMRI and the corresponding clinical data were combined to compare the various data fusion methods for CNN-based csPCa detection, segmentation and risk prediction. We evaluated the significance of including clinical data and found no significant improvement in the predictions of the CNNs for Detection and segmentation. The importance of each mpMRI sequence was analyzed and the results illustrated that all the sequences play a critical role in detection and segmentation of csPCa. Combining parametric clinical data with mpMRI data improves risk prediction. Finally, we compared the performance of our network with RAD-AI [43, 45] and found that our network performs comparably to the DI2IN method [44].

## Abbreviations

csPCa	Clinically significant Prostate carcinoma
DRE	Digital Rectal Examination
PSA	Prostate Specific Antigen
PI-RADS	Prostate Imaging Reporting and Data System
MRI	Magnetic Resonance Imaging
mpMRI	multiparametric MRI
ADC	Apparent Diffusion Coefficient maps
PSAD	PSA density
GTV	Gross Tumor Volume
CT	Computed Tomography
PET	Positron Emission Tomography
CNN	Convolutional Neural Network
MLP	Multi-Layer perceptron
EF	Early Fusion
IF	Intermediate Fusion
LF	Late Fusion
AP	Average Precision
DSC	Dice Similarity Coefficient

## Acknowledgements

Grant support by the Klaus Tschira Stiftung GmbH, Heidelberg, Germany is gratefully acknowledged.

## Author contributions

DDG is the corresponding author and has made substantial contributions in all relevant fields. ALG, CZ, BÖ, MaB and AN have made substantial contributions in the acquisition, analysis and interpretation of the data. They have also made substantial contributions in the conception and design of the patient related part of the work. LB, and SH contributed in major parts to the creation of new software and data processing techniques used in this work. CZ has made substantial contributions in drafting and revising the work. ALG has made substantial contributions in the conception and design of the work, interpretation of the data as well as in drafting and revising the work. MB has made substantial contributions in the conception and design of the

work, interpretation of the data as well as in drafting and revising the work. All authors reviewed the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by a research grant from the Klaus Tschira Stiftung GmbH, 00.014.2019. This work has been supported in parts by the Joint Funding Project "Joint Imaging Platform" of the German Cancer Consortium (DKTK) and the German Science Foundation (DFG) under research grant BO 3025/14 – 1.

## Data availability

No datasets were generated or analysed during the current study.

## Declarations

### Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The trial was registered retrospectively at the German Register for Clinical Studies (DRKS) under proposal number Nr. 476/14 & 476/19. The study was approved by the institutional ethics review board and patients gave written informed consent.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Division of Medical Physics, Department of Diagnostic and Interventional Radiology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

<sup>2</sup>Department of Diagnostic and Interventional Radiology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

<sup>3</sup>German Cancer Consortium (DKTK), Partner Site Freiburg, Freiburg, Germany

<sup>4</sup>German Oncology Center, European University Cyprus, Limassol, Cyprus

Received: 3 January 2024 / Accepted: 14 June 2024

Published online: 29 July 2024

## References

1. Dell'Atti L. The role of the digital rectal examination as diagnostic test for prostate cancer detection in obese patients. *J BUON*. 2015;20(6):1601–5.
2. Hötker A, Donati OF. PI-RADS 2.1 and structured reporting of magnetic resonance imaging of the prostate. Volume 61. *Radiologe*. Springer Medizin; 2021. pp. 802–9.
3. Higashihara E, Nutahara K, Kojima M, Okegawa T, Miura I, Miyata A, et al. Significance of serum free prostate specific antigen in the screening of prostate cancer. *J Urol*. 1996;156(6):1964–68.
4. Dalva I, Akan H, Yildiz O, Telli C, Bingol N. The clinical value of the ratio of free prostate specific antigen to total prostate specific antigen. *Int Urol Nephrol*. 1999;31:675–80.
5. Gurui K, Tewari A, Hemal AK, Wei J, Javidan J, Peabody J, Menon M. The role of prostate specific antigen in screening and management of clinically localized prostate cancer. *Int Urol Nephrol*. 2003;35:107–13.
6. Gupta R, Mahajan M, Sharma P. Correlation between prostate imaging reporting and data system version 2, prostate-specific antigen levels, and local staging in biopsy-proven carcinoma prostate: a retrospective study. *Int J Appl Basic Med Res*. 2021;11(1):32–35.
7. Kim TH, Kim CK, Park BK, Jeon HG, Jeong BC, Seo S, II, et al. Relationship between Gleason score and apparent diffusion coefficients of diffusion-weighted magnetic resonance imaging in prostate cancer patients. *Can Urol Assoc J*. 2016;10(11–12):E377–82.
8. Nepal SP, Nakasato T, Ogawa Y, Naoe M, Shichijo T, Maeda Y et al. Prostate cancer detection rate and Gleason score in relation to prostate volume as



- assessed by magnetic resonance imaging cognitive biopsy and standard biopsy. *Turkish J Urol* [Internet]. 2020 Nov [cited 2023 Jul 6];46(6):449–54. <http://www.ncbi.nlm.nih.gov/pubmed/33052831>
9. Sellers J, Wagstaff R, Helo N, de Riese WTW. Association Between Prostate Size and MRI Determined Quantitative Prostate Zonal Measurements. *Res reports Urol* [Internet]. 2022 [cited 2023 Jul 6];14:265–74. <http://www.ncbi.nlm.nih.gov/pubmed/35795724>
  10. Al-Khalil S, Ibilibor C, Cammack JT, de Riese W. Association of prostate volume with incidence and aggressiveness of prostate cancer. *Res Rep Urol*. 2016;8:201–5.
  11. Knight AS, Sharma P, de Riese WTW. MRI determined prostate volume and the incidence of prostate cancer on MRI-fusion biopsy: a systemic review of reported data for the last 20 years. Vol. 54, *International Urology and Nephrology*. Springer Science and Business Media B.V.; 2022. p. 3047–54.
  12. Bruno SM, Falagarino UG, d'Altilia N, Recchia M, Mancini V, Selvaggio O et al. PSA Density Help to Identify Patients With Elevated PSA Due to Prostate Cancer Rather Than Intraprostatic Inflammation: A Prospective Single Center Study. *Front Oncol* [Internet]. 2021 May 20 [cited 2023 Jul 6];11:693684. <https://www.frontiersin.org/articles/https://doi.org/10.3389/fonc.2021.693684/full>
  13. Iwaki H, Kajita Y, Shimizu Y, Yamauchi T. Predictive value of prostate specific antigen density in the detection of prostate cancer in patients with elevated prostate specific antigen levels and normal digital rectal findings or stage a prostate cancer. *Hinyokika Kyo*. 2001;47(3):169–74.
  14. Morote J, Raventos CX, Lorente JA, Lopez-Pacios MA, Encabo G, De Torres I, et al. Comparison of percent free prostate specific antigen and prostate specific antigen density as methods to enhance prostate specific antigen specificity in early prostate cancer detection in men with normal rectal examination and prostate specific antigen between 4.1 and 10 ng/mL. *J Urol*. 1997;158(2):502–4.
  15. Presti JC, Hovey R, Carroll PR, Shinohara K. Prospective evaluation of prostate specific antigen and prostate specific antigen density in the detection of nonpalpable and stage T1c carcinoma of the prostate. *J Urol*. 1996;156(5):1685–90.
  16. Wang ZB, Wei CG, Zhang YY, Pan P, Dai GC, Tu J et al. The Role of PSA Density among PI-RADS v2.1 Categories to Avoid an Unnecessary Transition Zone Biopsy in Patients with PSA 4–20 ng/mL. *Biomed Res Int*. 2021;2021.
  17. Zhou T, Ruan S, Array SC-. 2019 undefined. A review: Deep learning for medical image segmentation using multi-modality fusion. Elsevier [Internet]. [cited 2023 Dec 19]; <https://www.sciencedirect.com/science/article/pii/S2590005619300049>
  18. Guo Z, Li X, Huang H, ? NG-IT on, 2019 undefined. Deep learning-based image segmentation on multimodal medical imaging. *ieeexplore.ieee.org* Guo, X Li, H Huang, N Guo, Q Li IEEE Trans Radiat Plasma Med Sci 2019-*ieeexplore.ieee.org* [Internet]. [cited 2023 Dec 19]; Available from: <https://ieeexplore.ieee.org/abstract/document/8599078/>
  19. Acosta J, Falcone G, Rajpurkar P, Medicine ET-N. 2022 undefined. Multimodal biomedical AI. *nature.com* JN Acosta, GJ Falcone, P Rajpurkar, EJ Topol *Nature Med* 2022-*nature.com* [Internet]. [cited 2023 Dec 19]; <https://www.nature.com/articles/s41591-022-01981-2>
  20. Azam KSF, Ryabchykov O, Bocklitz T. A review on Data Fusion of Multidimensional Medical and Biomedical Data. *Molecules*. 2022;27(21):7448.
  21. Tiwari P, Viswanath S, Lee G, Madabhushi A. Multi-modal data fusion schemes for integrated classification of imaging and non-imaging biomedical data. In: *Proceedings - International Symposium on Biomedical Imaging*. 2011. pp. 165–8.
  22. Madabhushi A, Doyle S, Lee G, Basavanahally A, Monaco J, Masters S, et al. Integrated diagnostics: a conceptual framework with examples. *Clin Chem Lab Med*. 2010;48:989–98.
  23. Goodfellow I, Bengio Y, Courville A. Deep learning. 775 p.
  24. Ghosh A, Sufian A, Sultana F, Chakrabarti A, De D. Fundamental concepts of convolutional neural network. *Intelligent systems Reference Library*. Springer; 2019. pp. 519–67.
  25. Di W, Bhardwaj A, Wei J. Deep learning essentials: your hands-on guide to the fundamentals of deep learning and neural network modeling. 2018 [cited 2024 Jun 5]. Available from: [https://books.google.co.in/books?hl=en&lr=&id=ISBKDwAAQBA&oi=fnd&pg=PP1&dq=Deep+Learning+essentials:+your+hands-on+guide+to+the+fundamentals+of+deep+learning+and+neural+network+modeling&ots=RaM2L3mk9y&sig=EpO4LNNp6f8BgH87TdXuXW5qF3g&redir\\_esc=y#v=onepage&q=Deep%20learning%20essentials%3A%20your%20hands-on%20guide%20to%20the%20fundamentals%20of%20deep%20learning%20and%20neural%20network%20modeling&f=false](https://books.google.co.in/books?hl=en&lr=&id=ISBKDwAAQBA&oi=fnd&pg=PP1&dq=Deep+Learning+essentials:+your+hands-on+guide+to+the+fundamentals+of+deep+learning+and+neural+network+modeling&ots=RaM2L3mk9y&sig=EpO4LNNp6f8BgH87TdXuXW5qF3g&redir_esc=y#v=onepage&q=Deep%20learning%20essentials%3A%20your%20hands-on%20guide%20to%20the%20fundamentals%20of%20deep%20learning%20and%20neural%20network%20modeling&f=false)
  26. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Sci Rep*. 2022;12(1).
  27. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. Volume 3. *npj Digital Medicine*. Nature Research; 2020. pp. 1–9.
  28. Suresh H, Hunt N, Johnson A, Celi LA, Szolovits P, Ghassemi M. Clinical intervention prediction and understanding using deep networks. 2017. arXiv:1705.08498. Available from: <https://arxiv.org/abs/1705.08498>.
  29. Park C, Ha J, Park S. Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset. *Expert Syst Appl* [Internet]. 2020 [cited 2024 Jun 5];140. <https://www.sciencedirect.com/science/article/pii/S0957417419305834>
  30. Peng C, Zheng Y, Huang DS. Capsule Network based modeling of Multi-omics Data for Discovery of breast Cancer-related genes. *IEEE/ACM Trans Comput Biol Bioinforma*. 2020;17(5):1605–12.
  31. Lee G, Nho K, Kang B, Sohn K, reports DK-S. 2019 undefined. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *nature.com* G Lee, K Nho, B Kang, KA Sohn, D Kim *Scientific reports*, 2019-*nature.com* [Internet]. [cited 2024 Jun 5]; <https://www.nature.com/articles/s41598-018-37769-z>
  32. Huang Z, Zhan X, Xiang S, Johnson TS, Helm B, Yu CY et al. Salmon: Survival analysis learning with multi-omics neural networks on breast cancer. *Front Genet*. 2019;10(MAR).
  33. Islam M, Huang S, Ajwad R, Chi C, Wang Y, Hu P. An integrative deep learning framework for classifying molecular subtypes of breast cancer. *Comput Struct Biotechnol J*. 2020;18:2185–99. Available from: <https://www.sciencedirect.com/science/article/pii/S2001037020303585>
  34. Poirion OB, Chaudhary K, Garmire LX. Deep learning data integration for better risk stratification models of bladder cancer. *AMIA Summits Trans Sci Proc*. 2018;2018:197.
  35. Huang W, Wang X, Huang Y, Lin F, Tang X. Multi-parametric magnetic resonance Imaging Fusion for Automatic classification of prostate Cancer. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2022;2022–July:471–4.
  36. Reda I, Khalil A, Elmogy M, El-Fetouh AA, Shalaby A, El-Ghar MA et al. Deep learning role in early diagnosis of prostate cancer. *Technol Cancer Res Treat*. 2018;17.
  37. Hiremath A, Shiradkar R, Fu P, Mahran A, Rastinehad AR, Tewari A, et al. An integrated nomogram combining deep learning, prostate imaging-reporting and data system (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. *Lancet Digit Heal*. 2021;3(7):e445–54.
  38. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Lecture notes in Computer Science (including subseries lecture notes in Artificial Intelligence and Lecture notes in Bioinformatics)*. Springer; 2015. pp. 234–41.
  39. Agarwal HK, Merten FV, Sankineni S, Bernardo M, Senegas J, Keupp J, et al. Optimal high b-value for diffusion weighted MRI in diagnosing high risk prostate cancers in the peripheral zone. *J Magn Reson Imaging*. 2017;45(1):125–31.
  40. Gunashekar DD, Bielak L, Hägele L, Oerther B, Benndorf M, Grosu AL, et al. Explainable AI for CNN-based prostate tumor segmentation in multiparametric MRI correlated to whole mount histopathology. *Radiat Oncol*. 2022;17(1):65. <https://doi.org/10.1186/s13014-022-02035-0>
  41. Stanisz GJ, Odobina EE, Pun J, Escaravage M, Graham SJ, Bronskill MJ, et al. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magn Reson Med*. 2005;54(3):507–12.
  42. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* [Internet]. 2016 Feb [cited 2024 Jun 5];40(2):244–52. <https://journals.lww.com/0000478-201602000-00010>
  43. Sanford T, Harmon SA, Turkbey EB, Kesani D, Tuncer S, Madariaga M, et al. Deep-learning-based Artificial Intelligence for PI-RADS classification to assist multiparametric prostate MRI interpretation: a Development Study. *J Magn Reson Imaging*. 2020;52(5):1499–507.
  44. Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S et al. Automatic liver segmentation using an adversarial image-to-image network. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* [Internet]. Springer Verlag;

- 2017 [cited 2023 Apr 25]. pp. 507–15. [https://link.springer.com/https://doi.org/10.1007/978-3-319-66179-7\\_58](https://link.springer.com/https://doi.org/10.1007/978-3-319-66179-7_58)
45. Thimansson E, Bengtsson J, Baubeta E, Engman J, Flondell-Sité D, Bjartell A, et al. Deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI. *Eur Radiol.* 2022;33(4):2519–28.
  46. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2015. pp. 234–41.
  47. Chen T, Ma X, Liu X, Wang W, Feng R, Chen J et al. Multi-view learning with feature level fusion for cervical dysplasia diagnosis. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Science and Business Media Deutschland GmbH; 2019. pp. 329–38.
  48. Huang W, Wang X, Huang Y, ? FL-2022 44th A, 2022 undefined. Multi-parametric Magnetic Resonance Imaging Fusion for Automatic Classification of Prostate Cancer. *ieeexplore.ieee.org*W Huang, X Wang, Y Huang, F Lin, X Tang2022 44th Annu Int Conf IEEE Eng 2022•ieeexplore.ieee.org [Internet]. [cited 2023 Dec 19]; Available from: <https://ieeexplore.ieee.org/abstract/document/9871334/>
  49. Bardis M, Houshyar R, ? CC-RI, 2021 undefined. Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning. *pubs.rsna.org*M Bardis, R Houshyar, C Chantaduly, K Tran-Harding, A Ushinsky, C ChahineRadiology Imaging Cancer, 2021•pubs.rsna.org [Internet]. 2021 May 1 [cited 2023 Dec 19];3(3). Available from: <https://doi.org/10.1148/rycan.2021200024>
  50. Motamed S, Gujrathi I, Deniffel D, Oentoro A, Haider MA, Khalvati F, TRANSFER LEARNING FOR AUTOMATED, SEGMENTATION OF PROSTATE WHOLE GLAND AND TRANSITION ZONE IN DIFFUSION WEIGHTED MRI A PREPRINT [Internet]. *arxiv.org*. 2020 [cited 2023 Dec 19]. <https://arxiv.org/abs/1909.09541>
  51. Rundo L, Han C, Zhang J, Hataya R, Nagano Y, Militello C, et al. CNN-Based prostate zonal segmentation on T2-Weighted MR images: a Cross-dataset Study. *Smart Innov Syst Technol.* 2020;151:269–80.
  52. Wong T, Schieda N, Sathiadoss P, Haroon M, Abreu-Gomez J, Ukwatta E. Fully automated detection of prostate transition zone tumors on T2-weighted and apparent diffusion coefficient (ADC) map MR images using U-Net ensemble. *Med Phys.* 2021;48(11):6889–900.

### Publisher's Note

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