Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Hiremath A, Shiradkar R, Fu P, 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 Health 2021; 3: e445–54.
Appendix

Table of contents

A1: Dataset Description ........................................................................................................2
A2: Data Preprocessing........................................................................................................3
A3: Data augmentation .........................................................................................................4
A4: Defining peri-tumoral regions .......................................................................................4
A5: Deep learning implementation details .........................................................................5
A6: QIN-PROSTATE-Repeatability Test-retest repeatability cohort ...............................5
A7: Biochemical free recurrence cohort, D_{test}..................................................................6
Supplementary Figure 1 ..................................................................................................7
Supplementary Figure 2 ..................................................................................................8
Supplementary Figure 3 ..................................................................................................9
Supplementary Figure 4 ................................................................................................10
Supplementary Figure 5 ................................................................................................11
Supplementary Figure 6 ................................................................................................12
Supplementary Figure 7 ................................................................................................13
Supplementary Table 1 ..................................................................................................14
Supplementary Table 2 ..................................................................................................15
Supplementary Table 3 ..................................................................................................16
References .......................................................................................................................17
**A1: Dataset Description**

Detailed information of the dataset from each of the sites is described below. We used bi-parametric MRI (bpMRI) in this study to train the deep learning (DL) classifier since many of the previous studies (1–3), have illustrated the comparable performance of bpMRI and multiparametric MRI (mpMRI) and suggest that bpMRI can be used as an alternative for mpMRI with contrast-enhanced MR imaging protocols.

**D1:** PROSTATEx Challenge (4–6) data was used as D1. The challenge included 204 patients with prostate cancer (PCa). For each of the lesion findings, prostate anatomic region was available. A radiologist with greater than 15 years of experience from University Hospitals Cleveland Medical Center delineated lesions based on the defined anatomic region of the lesions. The ground-truth GGG for this dataset was based on MR-guided biopsies.

**D2:** Image data set from 126 patients with PCa from University Hospitals Cleveland Medical Center who underwent mpMRI was used as part of D2. The ground-truth GGG for this dataset was obtained from 12-core systematic biopsies. A pathologist specialized in Genitourinary (GU) Pathology with 9 years of experience graded the biopsy samples. A radiologist with greater than 15 years of experience delineated the lesions based on the biopsy reports. Additionally, lesions were scored based on PI-RADSv2 (7).

**D3:** This cohort consisted of 38 patients with PCa from the Icahn School of Medicine at Mount Sinai who underwent mpMRI between 2013 and 2017. The reference ground-truth was obtained based on 12-core systematic biopsies and was graded by a highly experienced GU pathologist with greater than 25 years of experience. A radiologist with greater than 15 years of experience from University Hospitals Cleveland Medical Center delineated the lesions based on the biopsy reports and scored the lesions according to PI-RADSv2 (7). A subset of this cohort (N = 31) was used in a previously published study (8) to predict biochemical recurrence (BCR) in PCa patients.
**D4**: Image data set from 151 patients with PCa from Cleveland Clinic who underwent mpMRI between 2009 and 2017 was used a part of D4. All the patients in D4 underwent RP, and the reference ground-truth was obtained with respect to the RP specimens. 12-core systematic biopsy and MR/Ultrasound fusion biopsy data was also available for a subset of 36 patients. Multiple pathologists specialized in GU graded the specimens and the pathologists had experience in the range of 5-15 years. A radiologist with greater than 10 years of experience from Cleveland Clinic delineated the lesions based on the RP specimens and scored the lesions according to PI-RADSv2 (7). A subset of this cohort (N = 127) was used in a previously published study (8) to predict biochemical recurrence (BCR) in PCa patients.

**D5**: This cohort consisted of 73 patients with PCa from Long Island Jewish Medical Center, Northwell Health who underwent mpMRI between 2012 and 2014. The ground-truth GGG for this dataset was obtained with reference to MR/Ultrasound fusion biopsies. A highly experienced GU pathologist with 19 years of experience graded the biopsy samples. A radiologist with greater than 10 years of experience from Northwell Health delineated the lesions based on the biopsy reports and scored the lesions with respect to PI-RADSv2 (7).

**A2: Data Preprocessing**

All the bpMRI images were first de-identified and transferred to the lead site in Cleveland. Lesions were then identified and PI-RADS score was assigned along with annotation of the lesions drawn referring to biopsy/pathological reports.

MRIs were taken before biopsy, and PSA was recorded at the time of MRI. In order to balance the training and the test set and make sure we validate on cases with ground-truth reference to RP specimens, D1, D2 and D3 was chosen as training cohort, D_{train}, and D4, D5 was considered as a testing cohort, D_{test}
T2W MRI images with endorectal coil often exhibit non-uniform intensities due to magnetic field variations rather than the anatomical changes. To remove these non-uniformities, a bias field correction was first applied to all the T2W MRI images with endorectal coil (N=108). Further, to correct for intensity drift artifacts and normalize the intensity to a particular range, a template T2W MRI image with a dynamic range of 0-800 was chosen and all the T2W MRI scans (N=592) were standardized using a previously developed method (9).

In order to be able to provide both T2W MRI and ADC maps as input to the network simultaneously, ADC were co-registered to their corresponding T2W MRI. A rigid followed by an affine registration with a multiresolution framework of “elastix” (10) was for registration. Control points from the entire field of view was used for rigid registration while those from the prostate capsule alone were used for affine registration. Mattes mutual information was used as the similarity metric with number of histogram bins set to 32. The similarity metric was optimized using gradient descent, a maximum of 500 iterations per resolution. A linear interpolator was used.

A3: Data augmentation

All augmentations on the data were performed on the entire 3D volume before extraction of patches. Rotation (3°, 5°, 8°) along the axial plane and shearing was randomly performed to balance and increase the size of the training dataset. For each of the volumes, about 10 different combinations of randomly chosen augmentations were performed.

A4: Defining peri-tumoral regions

Following the previous studies (11,12), we defined multiple scales of patches by including PT region increasing in steps to 3mm. The patches with respect to the first scale (S0) were extracted by drawing a bounding box around the segmented lesion, and the patches with subsequent scales (S1-S4) were extracted by applying a morphological operation (binary dilation) on the lesion
delineation, increasing the field-of-view from periphery of the delineated lesion up to 3 mm, 6 mm, 9 mm and 12 mm respectively

A5: Deep learning implementation details

Since the AlexNet (13) accepts images with input size of 224 x 224 pixels, all the patches were resized to 224 x 224 pixels. AlexNet model available in pytorch (0.4.1) was used for training. The network weights were randomly initialized with a manual seed (14). Binary entropy loss function was used for training the network, and an early stopping criterion was used to stop the network training with respective to the leave one set out cross validation loss. The network training was optimized using an Adam optimizer with a weight decay of $10^{-5}$ with a learning rate of $10^{-6}$. The models were trained with a hardware configuration; CPU with 16 cores, 2.10 GHz clock speed and 128 GB of RAM; 4 GPUs of Nvidia P100 with 3584 CUDA cores, 10.6 TeraFLOPS and 16 GB of memory.

A6: QIN-PROSTATE-Repeatability Test-retest repeatability cohort

To evaluate the stability of DL predictions, QIN-PROSTATE-Repeatability dataset (15) with mpMRI baseline and repeat prostate MRI exams for 15 subjects with a time interval of two weeks between the scans was used. Among the 15 patients, suspected tumor was identified in 11 patients. One patient was excluded due to poor quality of the ADC maps. Repeatability of DIP was assessed using the remaining 10 patients. All the baseline, repeat scans and manual annotations were co-registered and transformed with respect to the baseline T2W MRI. Patches were extracted from the co-registered scans and provided as input to DIP.
A7: Biochemical free recurrence cohort, $D_{\text{test}}$

Patients who (a) underwent RP with open or robotic surgery, b) had documented BCR, (c) or did not have BCR but followed for $\geq 3$ years were used to analyze BCR free survival. After the treatment, each of these patients underwent periodic follow-up with respect to the established clinical protocol (3–6 months in the first year and 6–12 months the following years), which included serial measurements of PSA levels, recording of recurrence dates for BCR+ cases, and date of last follow-up for BCR– cases. The BCR definition was based on two consecutive readings of PSA >0.2 ng/mL for men who underwent RP and as an increase in PSA >2 ng/mL compared to the initial PSA nadir value for men who underwent radiation therapy (RT) (with or without hormonal therapy). BCR free survival was calculated from the date of treatment to the date of recurrence or death, whichever occurred earlier, and was censored at the last follow-up date for those alive without BCR.
Supplementary Figure 1: Chart review of patients with inclusion and exclusion criteria and allocation of the selected patients in the training and test sets.
Supplementary Figure 2: (a) An integrated deep learning and clinical nomogram (DIN) constructed by integrating output probability score of deep learning based imaging predictor (DIP) and clinical variables (lesion volume, prostate volume and prostate specific antigen (PSA)) to distinguish clinically significant (csPCa) and clinically insignificant PCa (ciPCa). (b) Multivariable logistic regression analysis of DIP scoring and clinical variables (lesion volume, prostate volume and PSA).
Supplementary Figure 3: (a) An integrated PI-RADS and clinical nomogram (PIN) constructed by integrating PI-RADS score and clinical variables (lesion volume, prostate volume and prostate specific antigen (PSA)) to distinguish clinically significant (csPCa) and clinically insignificant PCa (ciPCa). (b) Multivariable logistic regression analysis of PI-RADS based scoring and clinical variables (lesion volume, prostate volume and PSA).
Supplementary Figure 4: Decision curve analysis of models; deep learning based imaging predictor (DIP: 2D CNN), integrated deep learning and clinical nomogram (DIN), integrated PI-RADS and clinical nomogram (PIN); and integrated deep learning, PI-RADS and clinical nomogram (ClaD) showing the net-benefit with respect to each other.
Supplementary Figure 5: Comparison of confusion matrices of (a) 12-core systematic biopsies, (b) MR/Ultrasound fusion biopsies, (c) the integrated deep learning, PI-RADS and clinical nomogram (ClaD) using radical prostatectomy surgical specimens as reference ground-truth.
Supplementary Figure 6: Comparison of receiver operating characteristic curves of (a) deep learning imaging predictor (DIP), (b) integrated deep learning and clinical nomogram (c) integrated PI-RADS and clinical nomogram (PIN), (c) integrated deep learning, PI-RADS and clinical nomogram (ClaD) in distinguishing patients with GGG=1 and GGG=2 lesions (N=158 patients).
Supplementary Figure 7: Comparison of receiver operating characteristic curves of (a) integrated deep learning and clinical nomogram (DIN), (b) integrated PI-RADS and clinical nomogram (PIN), (c) integrated deep learning, PI-RADS and clinical nomogram (ClaD) with Prostate Biopsy Collaborative Group (PBCG) risk calculator.
Supplementary Table. 1: Clinical variables, patient demographic information and distribution of clinically significant (csPCa) and insignificant (ciPCa)/benign lesions of data from five institutes (D1 – D5). All patients underwent 3T multi-parametric MRI and bi-parametric MRI (T2W MRI and apparent diffusion coefficient maps) were used. MRIs were acquired either using endorectal coil or surface coil.

|                          | D1   | D2   | D3   | D4   | D5   |
|--------------------------|------|------|------|------|------|
| N=Patients               | 204  | 126  | 38   | 151  | 73   |
| Age (median (IQR))       | NA   | 65.5 (59.0 – 72.0) | 63.0 (59.0 – 68.0) | 62.0 (56.0 – 66.0) | 65.5 (62.0 – 73.0) |
| PSA (ng/ml) (median (IQR)) | NA   | 6.6 (0.25 – 88.2) | 6.7 (5.0 – 10.0) | 5.7 (4.4 – 9.58) | 7.7 (4.8 – 11.3) |
| Lesion Volume (mm³) (median (IQR)) | NA   | 0.23 (0.14 – 0.44) | 0.35 (0.20 – 0.56) | 0.63 (0.28 – 2.39) | 0.29 (0.18 – 0.28) |
| Prostate Volume (mm³) (median (IQR)) | NA   | 45.7 (33.26 – 59.83) | 35.16 (31.30 – 40.05) | 37.12 (29.15 – 47.53) | 39.22 (29.18 – 49.76) |
| N=Lesions                | 325  | 139  | 41   | 168  | 150  |
| Clinical significance (%)|      |      |      |      |      |
| csPCA                    | 76 (23.4%) | 83 (59.7%) | 40 (97.6%) | 154 (91.6%) | 45 (30%) |
| ciPCA/benign             | 249 (76.6%) | 56 (40.3%) | 1 (2.4%) | 14 (8.4%) | 105 (70%) |
Supplementary Table 2: MRI acquisition parameters. The data used in this study consisted of bi-parametric MRI with T2W MRI and apparent diffusion co-efficient (ADC) maps derived from diffusion weighted imaging (DWI). A total of N=560 patients were used in this study with data acquired from five different institutes (D1-D5).

PPAC: pelvic phased-array coil, ERC: Endorectal coil.

| Cohort          | D1    | D2    | D3    | D4    | D5    |
|-----------------|-------|-------|-------|-------|-------|
| **Number of Subjects** | 204   | 126   | 38    | 151   | 73    |
| **Scanner**     |       |       |       |       |       |
| Manufacturer    | Siemens Skyra/ MAGNETOM Trio | Siemens Verio | Siemens Skyra | Philips Achieva | Siemens Verio |
| **Coil type**   | PPAC  | PPAC  | PPAC  | ERC   | ERC   |
| **T2-weighted MRI** |       |       |       |       |       |
| Field-of-view (mm²) | 192 x 192 | 200 x 200 | 160 x 160 | 260 x 260 | 140 x 140 |
| Matrix size     | 384 x 384 | 320 x 320 | 320 x 320 | 256 x 256 | 384 x 384 |
| **Diffusion-Weighted MRI** |       |       |       |       |       |
| Field-of-view (mm²) | 256 x 168 | 260 x 260 | 250 x 250 | 260 x 260 | 260 x 186 |
| Matrix size     | 128 x 84  | 128 x 128 | 114 x 114 | 128 x 128 | 116 x 81  |
| b-values (s/mm²) | 0, 50, 400, 800 | 0, 50, 600, 1000, 1400 | 0, 400, 900, 1500 | 0, 400, 900, 1500 | 0, 50, 500, 1000, 1500 |
Supplementary Table. 3: Area under the receiver operating characteristic (AUC) of deep learning based imaging predictor (DIP) with two base architectures (AlexNet and DenseNet) in identifying clinically significant prostate cancer lesions using five different input configurations. The patches with respect to the first scale (S0) were extracted by drawing a bounding box around the segmented lesion and the patches with subsequent scales (S1-S4) were extracted by extending the bounding box from periphery of the delineated lesion up to 3 mm, 6 mm, 9 mm and 12 mm respectively.

| Architecture | $S_0$   | $S_1$   | $S_2$   | $S_3$   | $S_4$   |
|--------------|---------|---------|---------|---------|---------|
| AlexNet      | 0.736±0.01 | **0.752±0.01** | 0.737±0.01 | 0.749±0.04 | 0.737±0.03 |
| DenseNet     | 0.741±0.03 | **0.748±0.02** | 0.743±0.02 | 0.741±0.03 | 0.729±0.04 |
References

1. Tamada T, Kido A, Yamamoto A, Takeuchi M, Miyaji Y, Moriya T, et al. Comparison of Biparametric and Multiparametric MRI for Clinically Significant Prostate Cancer Detection With PI-RADS Version 2.1. J Magn Reson Imaging. 2021 Jan;53(1):283–91.

2. Cho J, Ahn H, Hwang SI, Lee HJ, Choe G, Byun S-S, et al. Biparametric versus multiparametric imaging of the prostate: detection of clinically significant cancer in a perfect match group. Prostate International [Internet]. 2020 Feb 26 [cited 2020 Oct 23]; Available from: http://www.sciencedirect.com/science/article/pii/S2287888220300088

3. Han C, Liu S, Qin XB, Ma S, Zhu LN, Wang XY. MRI combined with PSA density in detecting clinically significant prostate cancer in patients with PSA serum levels of 4~10ng/mL: Biparametric versus multiparametric MRI. Diagn Interv Imaging. 2020 Apr;101(4):235–44.

4. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. SPIE-AAPM PROSTATEEx Challenge Data [Internet]. The Cancer Imaging Archive; 2017 [cited 2020 Oct 7]. Available from: https://wiki.cancerimagingarchive.net/x/iIFpAQ

5. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. SPIE-AAPM PROSTATEEx Challenge Data [Internet]. The Cancer Imaging Archive; 2017 [cited 2020 Nov 18]. Available from: https://wiki.cancerimagingarchive.net/x/iIFpAQ

6. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. J Digit Imaging. 2013 Dec;26(6):1045–57.

7. Rosenkrantz AB, Oto A, Turkbey B, Westphalen AC. Prostate Imaging Reporting and Data System (PI-RADS), Version 2: A Critical Look. American Journal of Roentgenology. 2016 Feb 25;206(6):1179–83.

8. Li L, Shiradkar R, Leo P, Algohary A, Fu P, Tirumani SH, et al. A novel imaging based Nomogram for predicting post-surgical biochemical recurrence and adverse pathology of prostate cancer from pre-operative bi-parametric MRI. EBioMedicine. 2021 Jan;63:103163.

9. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. Magn Reson Med. 1999 Dec;42(6):1072–81.

10. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging. 2010 Jan;29(1):196–205.

11. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology. 2019 Mar;290(3):783–92.

12. Braman NM, Etessami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res [Internet]. 2017 [cited 2018 Dec 3];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437672/
13. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017 May 24;60(6):84–90.

14. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. :8.

15. Fedorov A, Schwier M, Clunie D, Herz C, Pieper S, Kikinis R, et al. An annotated test-retest collection of prostate multiparametric MRI. Scientific Data. 2018 Dec 4;5(1):180281.