The Prostate Imaging: Cancer AI (PI-CAI) 2022 Grand Challenge (PIMed Team)

Xinran Li[1,3⋆], Sulaiman Vesal[1,2⋆] (✉), Sara Saunders[1,2⋆], Simon John Christoph Soerensen[1,2], Hassan Jahanandish[1,2], Stefania Moroianu[1], Indrani Bhattacharya[1,2], Richard E. Fan[2], Geoffrey A. Sonn[2⋆⋆], and Mirabela Rusu[1⋆⋆](✉)

[1] Department of Radiology, Stanford University, Stanford CA 94305, USA
[2] Department of Urology, Stanford University, Stanford CA 94305, USA
[3] Institute for Computational and Mathematical Engineering, Stanford CA 94305, USA
svesal@stanford.edu, mirabela.rusu@stanford.edu

Abstract. This paper summarizes our approaches and results for the PI-CAI 2022 Grand Challenge, which focuses on the detection and localization of clinically significant prostate cancer (csPCa) using bi-parametric magnetic resonance imaging (bpMRI). In addition to detecting subtle prostate cancer features on MRI, this particular task presents several other challenges, including a highly imbalanced dataset with only 28% cases having csPCa, images acquired on two different scanners (Siemens and Philips), and cases with imperfect labels. Our proposed model consists of an ensemble of three models including 1) a 2.5D cancer detection model that combines our in-house SPCNet model together with an in-house prostate gland segmentation model to force learning and predictions within the prostate, 2) a novel multi-task model called SPCNet-Decision, that performs voxel-level cancer detection and localization using 2.5D SPCNet, as well as slice-level classification using a decision head classifier, and 3) a 3D UNet model with a decoder incorporating residual connections. We trained all three models using five-fold cross-validation and averaged the predicted probability maps to generate the final ensemble model predictions. Our ensemble model achieved an AUC-ROC of 86.5% and average precision of 68.1% on publicly available validation set which consist of 100 patients. It outperformed the baseline models provided by the challenge organizers by a 3.9% (Our ranking score: 77.3% vs. baseline ranking score: 73.4%).

Keywords: Prostate Cancer Detection · Image Segmentation · Prostate Cancer · Deep learning · bpMRI.

⋆ Equal contribution as first authors
⋆⋆ Equal contribution as senior authors

1 Introduction

Prostate cancer (PCa) is the second leading cause of cancer deaths among men in western countries [1]. Bi-parametric magnetic resonance imaging (bpMRI) is now widely utilized to detect prostate cancer, guide MRI-ultrasound fusion biopsies, and plan treatment. MRI is currently considered as the most sensitive non-invasive imaging approach for the visualization, detection, and localization of prostate cancer. However, MRI interpretation is challenging due to inherent subtle imaging features, and it suffers from wide inter-reader variability. As such, there is a clinical need for standardized MRI interpretation to enable accurate, generalizable and timely clinically significant prostate cancer (csPCa) detection. Many machine learning algorithms have recently been developed to improve computer-aided prostate cancer detection on MRI. However, comparison of these methods requires unbiased assessment on common training and test data. To address these problems, the PI-CAI grand challenge presents a consolidated platform for training and evaluating machine learning algorithms for prostate cancer detection using bpMRI and optional clinical data. The challenge presents 1500 training cases, and standardized evaluation criteria for patient-level diagnosis and lesion-level detection of csPCa (ISUP $\geq$ 2 cancer).

2 Methods

For the PI-CAI challenge, we developed several 3D and 2.5D (including three consecutive MRI slices) deep learning models. This includes the 3D-UNet [2], SPCNet [4] and a novel multi-task learning method with the SPCNet backbone called SPCNet-Decision. In the following, we discuss the methods and data preprocessing steps we used.

2.1 Data Preprocessing

Multiple preprocessing steps were applied to the MRI scans, which were adapted from baseline models provided by challenge organizer [3].

- MRIs were all resampled to the same resolution (0.5mm×0.5mm×3.0mm/voxel) and cropped around the center of the scan to 20 slices of 256×256 each.
- All three MRI sequences (T2, ADC, and DWI) were independently normalized with instance-wise z-score normalization.

2.2 2.5D Models

Our 2.5D models are based on the Stanford Prostate Cancer Network (SPCNet)[4], a convolutional neural network designed to selectively identify aggressive cancer, indolent cancer, and normal tissue on MRI. For the PI-CAI challenge, we modified SPCNet to (a) to detect and localize clinically significant prostate cancer on MRI, and (b) include three input-branches of T2w, ADC

and DWI sequences. We also included an in-house prostate gland segmentation model ProGNet [5] to limit PCa detection to the prostate region. To train our 2.5D prostate gland segmentation model ProGNet, we used ground truth prostate gland segmentations provided by the challenge organizer. The SPCNet takes T2w, ADC, and DWI images as input and has multiple outputs at various image scales, which are then upsampled and fused to form the final output.

The SPCNet predicts more false-positive lesions, so we modified its architecture to include a new classification head to form a multi-task optimization problem. Figure 1 depicts the network architecture. The classification head is designed to determine if a slice contains any clinically significant prostate cancer. The classification head applies a 1x1 convolution layer and fully connected layer to the probability maps generated by SPCNet and determines if cancer is present in that slice. This model, called SPCNet-Decision outperforms the initial SPCNet model significantly at both the patient-level and lesion-level. We trained all variations of the SPCNet model with the Adam optimizer and a learning rate of 0.005 using masked weighted-cross-entropy. The prostate gland segmentation was used to compute the gradient only within the prostate gland boundaries. All models were trained for 15-25 epochs.

2.3 3D Models

We trained several baseline 3D models locally, but the standard 3D UNet with residual connections outperformed other models. This network includes 5 levels and was trained using a multi-class Dice and cross-entropy loss using Adam optimizer with a initial learning rate of 0.001. The models were trained for 200 epochs and the learning rate was adjusted using Cosine Annealing scheduler.

3 Results

For all our experiments, we used five-fold cross-validation on the training data provided by the challenge organizers. We ensembled the predicted probability maps from the U-Net, SPCNet, and SPCNet-decision models into a single model using two different methods: simple averaging and z-score normalization. Simply averaging entailed summing the probability maps of multiple models and dividing by the number of models involved. However, since each model produced probability maps with different distributions, we performed z-score normalization on these results prior to combining them, as follows.

- For each case in the validation set, we calculated the mean and standard deviation of all voxels in the probability map with values greater than 0.01, thresholding to exclude the background.
- We calculated the mean(mean) and mean(std. dev.) of all validation cases to z-score normalize the probability maps output from each model for the test cases:

$$Prob_{norm} = (Prob_{orig} - mean(mean))/mean(std.dev.)$$

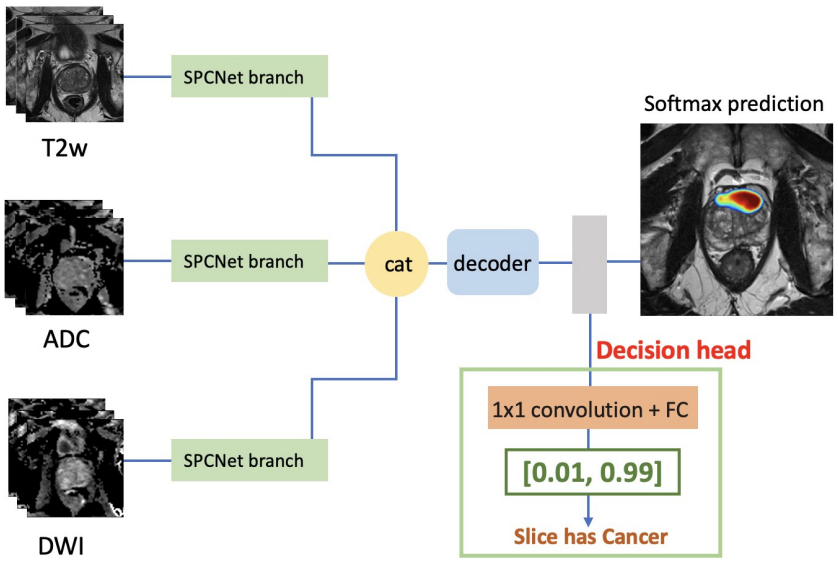

Fig. 1: SPCNet network architecture with decision head.

– Probability maps from all four models were summed together, and this ensembled output re-scaled from 0 to 1, with 0 and 1 corresponding to the cases in which all models predict probability of 0 and 1, respectively.

Thus, the more models that detect a lesion, the higher its probability in an ensembled model. Results are shown in Table 1. Our ensemble model U-Net + SPCNET-dec using averaging ensemble achieved an AUC-ROC of 86.5% and average precision of 68.1% on publicly available validation set which consist of 100 patients. Our ensembled model (U-Net + SPCNET-dec) achieved a ranking score 77.3% vs. baseline ranking score 73.4%. It outperformed the baseline models provided by the challenge organizers by a 3.9% in terms of overall ranking score.

4 Conclusion

In this paper, we presented our work on patient-level diagnosis and lesion-level detection of csPCa in bpMRIs for the PI-CAI grand challenge. We developed a novel multi-task machine learning method, SPCNET-Decision, that incorporates both a classification head and prostate cancer detection head to tackle the challenge of an imbalanced dataset. The classification head determines if cancer is present in a MRI slice. Our final model was an ensemble of the 2.5D SPCNET-Decision with a 3D UNet model, which improved performance and reduced de-

Table 1: The performance results our models for five-fold cross validation and public validation set. Here SPCNET-dec denotes SPCNET-Decision model and avg denotes ensemble averaging. The row with light-gray color is our final model submitted for the hidden test data.

Model	Five-fold Cross Validation			Public Validation Set		
	AUCROC[%]	AP[%]	Rank[%]	AUCROC[%]	AP[%]	Rank[%]
U-Net	80.1	47.0	63.6	81.6	62.8	72.2
SPCNet	80.3	39.1	59.7	80.2	63.0	71.6
SPCNet-dec	82.0	46.0	64.0	86.5	66.4	76.5
U-Net+SPCNet-dec (z-score)	83.2	49.6	66.0	-	-	-
U-Net+SPCNet+SPCNet-dec (z-score)	83.4	50.9	67.2	85.7	66.9	76.3
U-Net + SPCNet-dec (avg)	82.1	48.4	65.3	**86.5**	68.1	**77.3**
U-Net + SPCNet + SPCNet-dec (avg)	**83.4**	**48.9**	**66.1**	86.1	**68.3**	77.2

tection of false positive lesions. The proposed methods achieved promising results when evaluated with AUROC and AP on the validation dataset. In the future, we aim to test our proposed approach on additional multi-institutional data.

5 Acknowledgments

This work was supported by the Department of Radiology and the Department of Urology at Stanford University.

References

1. Bhattacharya, I., Khandwala, Y.S., Vesal, S., Shao, W., Yang, Q., Soerensen, S.J., Fan, R.E., Ghanouni, P., Kunder, C.A., Brooks, J.D., Hu, Y., Rusu, M., Sonn, G.A.: A review of artificial intelligence in prostate cancer detection on imaging. Therapeutic Advances in Urology **14**, 17562872221128791 (2022). https://doi.org/10.1177/17562872221128791
2. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
3. Saha, A., Twilt, J.J., Bosma, J.S., van Ginneken, B., Yakar, D., Elschot, M., Veltman, J., Fütterer, J., de Rooij, M., Huisman, H.: Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol) (May 2022). https://doi.org/10.5281/zenodo.6522364
4. Seetharaman, A., Bhattacharya, I., Chen, L.C., Kunder, C.A., Shao, W., Soerensen, S.J.C., Wang, J.B., Teslovich, N.C., Fan, R.E., Ghanouni, P., Brooks, J.D., Too, K.J., Sonn, G.A., Rusu, M.: Automated detection of aggressive and indolent prostate cancer on magnetic resonance imaging. Medical Physics **n/a**(n/a) (2020). https://doi.org/https://doi.org/10.1002/mp.14855
5. Soerensen, S.J.C., Fan, R.E., Seetharaman, A., Chen, L., Shao, W., Bhattacharya, I., Kim, Y.h., Sood, R., Borre, M., Chung, B.I., Sonn, G.A., Rusu, M.: Deep learning improves speed and accuracy of prostate gland segmentations on mri for targeted biopsy. The Journal of Urology pp. 10–1097 (2021)