Classification of Clinical Significance of MRI Prostate Findings Using 3D Convolutional Neural Networks

Alireza Mehrtash[a,b], Alireza Sedghi[c], Mohsen Ghafoorian[a, d], Mehdi Taghipour[a], Clare M. Tempany[a], William M. Wells III[a], Tina Kapur[a], Parvin Mousavi[c], Purang Abolmaesumi[b], and Andriy Fedorov[a]

[a]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[b]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
[c]Medical Informatics Laboratory, Queen's University, Kingston, ON, Canada
[d]Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands

ABSTRACT

Prostate cancer (PCa) remains a leading cause of cancer mortality among American men. Multi-parametric magnetic resonance imaging (mpMRI) is widely used to assist with detection of PCa and characterization of its aggressiveness. Computer-aided diagnosis (CADx) of PCa in MRI can be used as clinical decision support system to aid radiologists in interpretation and reporting of mpMRI. We report on the development of a convolution neural network (CNN) model to support CADx in PCa based on the appearance of prostate tissue in mpMRI, conducted as part of the SPIE-AAPM-NCI PROSTATEx challenge. The performance of different combinations of mpMRI inputs to CNN was assessed and the best result was achieved using DWI and DCE-MRI modalities together with the zonal information of the finding. On the test set, the model achieved an area under the receiver operating characteristic curve of 0.80.

1. INTRODUCTION

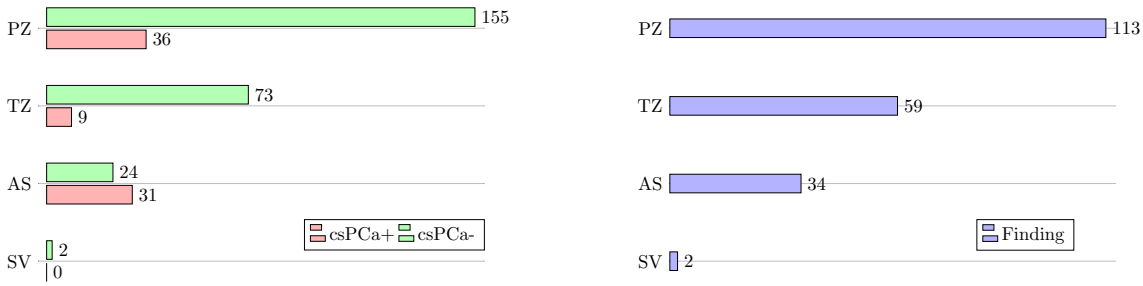
Currently, about one in seven men will be diagnosed with prostate cancer during his lifetime. Estimations show that the number of new cases and deaths from Prostate Cancer (PCa) will be 161,360 and 26,730, respectively, in 2017.[1] Accurate diagnosis and staging of PCa are critical for the selection of the most suitable treatment, and ultimately for reducing PCa morbidity and mortality. Recent advances in prostate multi-parametric magnetic resonance imaging (mpMRI) have improved cancer diagnosis and staging.[2] At present, mpMRI assessment relies on human experts, and requires specialized training and experience. Recently, deep convolutional neural networks (CNN) have been widely used in medical image processing and analysis and have outperformed the conventional computer vision methods in various medical image analysis tasks[3] including detection of microcalcifications in digital breast tomosynthesis,[4] masses in mammography,[5] embolism in CT pulmunary angiography, and lacunes brain MRI.[6] In this paper we present a 3D CNN method developed for the SPIE-AAPM-NCI PROSTATEx challenge, tailored for the task of classification of clinically significant prostate cancer findings in mpMRI.

2. METHODS

2.1 Data

The challenge included the training dataset consisting of 204 patients with 330 suspicious lesion findings, and the test dataset with 140 patients and 208 findings. For each of the findings, assignment of the prostate anatomic region was available. The prostate gland can be sub-divided into four anatomic regions: the peripheral zone (PZ), with 70–80% of the glandular tissue and accounting for about 70% of PCa; the transition zone (TZ), 5%

E-mail: mehrtash@bwh.harvard.edu



(a) Train data.

(b) Test data.

Figure 1: Distribution of training and test datasets of the PROSTATEx challenge. **(a)** Training samples: the distribution of lesion findings shows that the training dataset is not balanced in terms of both zonal distribution and the clinical significance of the finding. **(b)** Test samples are not balanced in terms of zones.

of the glandular tissue, and ≈ 25% of PCa; the central zone, 20% of the glandular tissue and around 5% of PCa; and the non-glandular anterior fibromuscular stroma (AS).[7] The training and test samples in the PROSTATEx challenge were from PZ, TZ, AS and seminal vesicles (SV) as illustrated in Figure 1. After minor data cleaning, we selected 201 subjects with 321 findings for training and validation purposes. In order to augment and balance the training dataset, we used flipping and translation of the original data. As a result of data augmentation, we generated 5-fold cross-validation datasets with 10,000 training and 2,000 validation samples for each fold. The training-validation splitting in each fold was done such that the distribution of the findings across prostate regions was preserved. Image intensities were normalized to be within the range of [0,1]. 3D patches of size $40 \times 40 \times 40$ mm for T2, $32 \times 32 \times 12$ for DWI and DCE-MRI images, centered at finding locations served as inputs to the CNN.

2.2 Network Architecture

Our CNN architecture included three input streams: ADC maps and maximum b-value from DWI, and K^{trans} from DCE-MRI. Similar to the work of Ghafoorian et al.,[8] we added explicit zone information to the first dense layer. The DCE-MRI and DWI streams with input sizes of $(32 \times 32 \times 12)$ had 9 convolutional layers combining of $(3 \times 3 \times 1)$ and $(3 \times 3 \times 3)$ filter sizes. Max-pooling layers of size $2 \times 2 \times 1$ were applied in selected middle layers. At the end of each stream, the output of the last convolutional layer was connected to a dense layer. The neurons of this layer were concatenated with the zonal information of the finding and applied to another set of three fully connected layers. Leaky rectified linear unit[9] function was used as the non-linearity element.

2.3 Training

For training the network, we used the stochastic gradient descent algorithm with the Adam update rule,[10] a mini-batch size of 64, and a binary cross-entropy loss function. We initialized the CNN weights randomly from a Gaussian distribution using the He method.[11] We also batch-normalized[12] the intermediate responses of all layers to accelerate the convergence. To prevent overfitting, in addition to the batch-normalization, we used drop-out with 0.25 probability as well as L_2 regularization with $\lambda_2 = 0.005$ penalty on neuron weights. We used an early stopping policy by monitoring validation performance and picked the best model with the highest accuracy on the validation set. Cross-validation was used to find the best combination of input channels and number of filters for convolutional layers.

3. RESULTS

Our training-validation results indicate that the combination of ADC, maximum B-Value and K^{trans} modalities in combination with zonal information of the lesion leads to the best performance characterized by the area

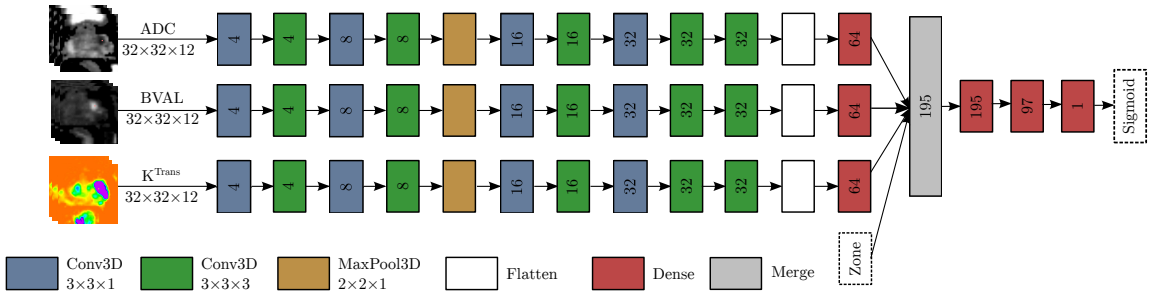


Figure 2: Architecture of the proposed 3D CNN. The network uses combination of ADC map, maximum B-Value (BVAL) from DWI and K^{trans} from DCE-MRI with zone information.

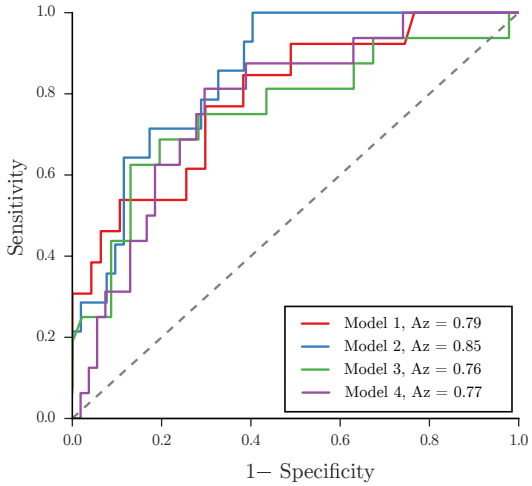


Figure 3: Comparison of classifiers trained with architecture in Figure 2 on different folds of cross-validation.

under curve (A_z) of the receiver operating characteristic (ROC) curve. Figure 3 shows the results of training on different folds of our cross-validation. For test data prediction we combined the prediction of the best 4 out of the 5 models by averaging the outputs of the models. Figure 4 shows an example of a true positive finding in the validation set. The trained model is deployed in the open-source deployment toolkit, the DeepInfer.[13]

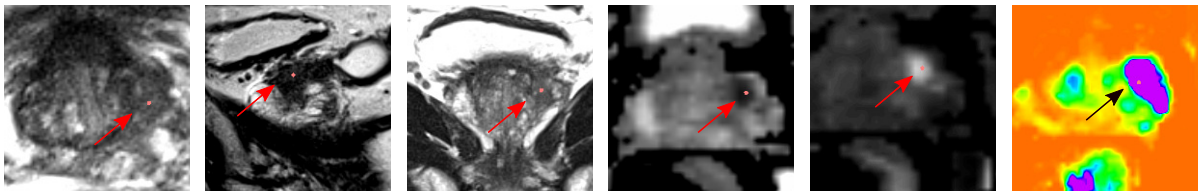
Network was tested using 206 findings from 140 patients. The performance of our model on PROSTATEx challenge data set was reported by organizers as ($A_z = 0.80$). This is in the same range as our validation results, indicating that the proposed model generalizes well on the new data.

4. DISCUSSION AND CONCLUSIONS

In this study we observed that 3D CNNs can be efficiently applied for detecting clinically significant prostate cancer. Our result is comparable with the A_z values achieved by the experienced human reader: 0.79 and 0.83 for PI-RADS v1 and PI-RADS v2, respectively.[14]

5. ACKNOWLEDGEMENTS

Research reported in this publication was supported by NIH Grant No. P41EB015898, Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian Institutes of Health Research (CIHR).



(a) (b) (c) (d) (e) (f)

Figure 4: An example of a PZ true positive in validation set. Only (d-f) modalities with zone information (zone=PZ) were used by the network to predict the clinical significance of the finding.

REFERENCES

[1] Siegel, R. L., Miller, K. D., and Jemal, A., "Cancer statistics," *CA Cancer J Clin* **67**(1), 7–30 (2017).

[2] Hegde, J. V., Mulkern, R. V., Panych, L. P., Fennessy, F. M., Fedorov, A., Maier, S. E., and Tempany, C. M. C., "Multiparametric MRI of prostate cancer: An update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer," *Journal of Magnetic Resonance Imaging* **37**(5), 1035–1054 (2013).

[3] Litjens, G., Kooi, T., Ehteshami Bejnordi, B., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I., "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747* (2017).

[4] Samala, R. K., Chan, H.-P., Hadjiiski, L. M., Cha, K., and Helvie, M. A., "Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis," in [*SPIE Medical Imaging*], 97850Y–97850Y, International Society for Optics and Photonics (2016).

[5] Qiu, Y., Yan, S., Tan, M., Cheng, S., Liu, H., and Zheng, B., "Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study," in [*SPIE Medical Imaging*], 978520–978520, International Society for Optics and Photonics (2016).

[6] Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., Keizer, K., de Leeuw F.E, van Ginneken, B., Marchiori, E., and Platel, B., "Deep multi-scale location-aware 3d convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage: Clinical* (feb 2017).

[7] Liu, L., Tian, Z., Zhang, Z., and Fei, B., "Computer-aided Detection of Prostate Cancer with MRI: Technology and Applications," *Academic Radiology* **23**(8), 1024–1046 (2016).

[8] Ghafoorian, M., Karssemeijer, N., Heskes, T., van Uden, I., Sanchez, C., Litjens, G., de Leeuw, F.-E., van Ginneken, B., Marchiori, E., and Platel, B., "Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities," *arXiv preprint arXiv:1610.04834* (2016).

[9] Maas, A. L., Hannun, A. Y., and Ng, A. Y., "Rectifier nonlinearities improve neural network acoustic models," in [*Proc. ICML*], **30**(1) (2013).

[10] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[11] He, K., Zhang, X., Ren, S., and Sun, J., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in [*Proceedings of the IEEE International Conference on Computer Vision*], 1026–1034 (2015).

[12] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167* (2015).

[13] Mehrtash, A., Pesteie, M., Hetherington, J., Behringer, P. A., Kapur, T., Wells III, W. M., Rohling, R., Fedorov, A., and Abolmaesumi, P., "Deepinfer: Open-source deep learning deployment toolkitfor image-guided therapy," in [*SPIE Medical Imaging*], International Society for Optics and Photonics (2017).

[14] Kasel-Seibert, M., Lehmann, T., Aschenbach, R., Guettler, F. V., Abubrig, M., Grimm, M.-O., Teichgraeber, U., and Franiel, T., "Assessment of pi-rads v2 for the detection of prostate cancer," *European Journal of Radiology* **85**(4), 726–731 (2016).