2022

Construction of a Repeatable Framework for Prostate Cancer Lesion Binary Semantic Segmentation using Convolutional Neural Networks

Ian Vincent O. Mirasol

Patricia Angela R. Abu

Rosula SJ Reyes

Construction of a Repeatable Framework for Prostate Cancer Lesion Binary Semantic Segmentation using Convolutional Neural Networks

Ian Vincent O. Mirasol, Patricia Angela R. Abu, Rosula S.J. Reyes
Department of Information
Systems and Computer Science
Ateneo de Manila University
Loyola Heights, Philippines

***Abstract*—Prostate cancer is the 3rd most diagnosed cancer overall. Current screening methods such as the prostate-specific antigen test could result in overdiagonosis and overtreatment while other methods such as a transrectal ultrasonography are invasive. Recent medical advancements have allowed the use of multiparametric MRI — a noninvasive and reliable screening process for prostate cancer. However, assessment would still vary from different professionals introducing subjectivity. While convolutional neural network has been used in multiple studies to objectively segment prostate lesions, due to the sensitivity of datasets and varying ground-truth established used in these studies, it is not possible to reproduce and validate the results. In this study, we executed a repeatable framework for segmenting prostate cancer lesions using annotated apparent diffusion coefficient maps from the QIN-PROSTATE-Repeatability dataset — a publicly available dataset that includes multiparametric MRI images of 15 patients that are confirmed or suspected of prostate cancer with two studies each. We used a main architecture of U-Net with batch normalization tested with different encoders, varying data image augmentation combinations, and hyperparameters adopted from various published frameworks to validate which combination of parameters work best for this dataset. The best performing framework was able to achieve a Dice score of 0.47 (0.44-0.49) which is comparable to previously published studies. The results from this study can be objectively compared and improved with further studies whereas this was previously not possible.**

Keywords*—Convolutional neural networks; binary semantic segmentation; prostate cancer; computer vision; deep learning

I. Introduction

Prostate cancer (PCa) is the 3rd most diagnosed and the 8th leading cause of death among all cancers; this constitutes to a 7.3% incidence rate and 3.8% mortality rate from 2020 worldwide cancer cases [1].

While the Prostate-Specific Antigen (PSA) test remains to be an effective test at detecting prostate cancer [2], approximately 40% from this screening method were found to be an occurrence of overdiagnosis [3]. This increases the risk of overtreatment leading to unnecessary procedures which yielded conflicting results in terms of benefit [4].

Recent developments in multiparametric magnetic resonance imaging (mpMRI) have provided a reliable non-invasive screening process for detecting clinically significant cancer with great specificity [5], [6] while being non-inferior to a systematic biopsy [7]. This allows the lesion to be classified in a more specific manner through objective metric scoring measures such as the Prostate Imaging and Reporting Archiving Data System (PI-RADS) [8]. However, the process for scoring requires a qualified professional to observe various images and manually assess the presence and severity of the lesion. It is also worth noting that due to the subjectivity of the assessment, results may vary for different professionals.

Various convolutional neural network (CNN) algorithms have allowed applications in MRI images specific to PCa through prostate organ segmentation and volume estimation, lesion detection, and lesion segmentation [9]. These studies explore the viability of creating an aid for professionals to use as basis and potentially lessen the subjective component of the assessment. But in training and comparing results, the ground-truth is set forth by the professionals' assessment which poses the problem of a standardized ground-truth.

The QIN-PROSTATE-Repeatability (QPR) Dataset [10] is a publicly available annotated dataset. This dataset includes mpMRI images of 15 patients of two studies each with confirmed or suspicion of prostate cancer. These images can be utilized as input and ground-truth.

This study aims to create a pipeline to execute the task of prostate lesion detection and segmentation from an ensemble of frameworks from recent CNN studies to be applied on the QPR dataset for validation.

II. Statement of the Problem

Results from mpMRI images need to be assessed by a qualified professional (e.g. trained radiologist). The procedure remains to be subjective [6] and may vary from person-to-person. The Prostate Imaging and Reporting Archiving Data System (PI-RADS) provides a systemic scoring for image interpretation and lesion detection [8] but does not completely eliminate the subjectivity of human interpretation.

There have been various studies involving the implementation of CNN to prostate mpMRI images [9]. However, most of the datasets used are not publicly available. Moreover, the ground-truth for various studies involving CNN applications for PCa also varies depending on the qualified professional assigned for the assessment. This results in a lack of stan-

dardization. The results from these studies also could not be objectively compared, reproduced, nor improved.

III. Scope and Limitations

The study focused on PCa specifically found in the peripheral zone (PZ) only. About 70% of all PCa are found to be in the PZ having the most amount of prostatic glandular tissue [11]. Patients whose lesions were not found within the PZ were not considered. For sequences and views, only the apparent diffusion coefficient (ADC) map was considered. In PI-RADS scoring for PZ PCa, ADC along with diffusion-weighted imaging (DWI) images are used as references for scoring while dynamic contrast-enhanced (DCE) images are used to distinguish between PI-RADS 3 and 4 having T2-Weighted images used as a supplementary [12]. The DICOM files acquired from the QPR dataset for this study is limited to the ADC map only as the DWI images do not contain any available annotated lesion segmentation. However, these images are closely related wherein the ADC map removes the T2-weighting aspect inherent to the DWI. T2-Weighted images were also not considered as these are used as supplementary in PZ PI-RADS scoring.

IV. Review of Related Literature and Studies

PCa biomedical images are not only scarce and hard to obtain, but they require a lot of resources to manually annotate. While this is the avenue that most studies implement, this introduces variability among datasets. "An overview of publicly available patient-centered prostate cancer datasets" is a paper that summarizes all of publicly available patient-centered PCa datasets with the goal of providing researchers an avenue to select the appropriate dataset needed for their specific field of study [13]. Among the list of datasets within the article, under the imaging category, only the QIN-PROSTATE Repeatability (QPR) dataset had annotated images.

Data image augmentation is a technique used to expand the quantity of images by altering existing images from a dataset. This has been shown to prevent over-fitting and improve the overall accuracy while also being more advantageous to weight decay and dropout which both require fine-tuning of parameters [14]. Isensee et al. achieved the highest score among participants that utilized the U-Net architecture and 2nd place overall for the multi-class brain-tumor segmentation from MRI in the Brain Tumor Segmentation (BraTS) 2018 challenge [15] . This was achieved by using the baseline U-net model while focusing on aggressive data augmentation.

For training modern deep neural networks, it was shown that best results were consistently obtained through smaller batch sizes (2,32) [16]. This was further validated with an experiment on the effects of different batch sizes on a histopathology dataset [17]. Kandel and Castelli ran their model on both Adam and SGD optimizers with a learning rate (LR) set to 1×10^{-4} and 1×10^{-3}. The best AUC was achieved with a batch size of 16, Adam optimizer, and LR set at 1×10^{-4}.

The U-Net architecture showcased in a paper published in 2015 by Ronneberger et al. proposed a model that can be trained with limited images offset by data augmentation [18]. It tackled the main issues presented with working with medical data such as scarcity of an annotated dataset and the importance of proximal location assessment. Advancements in CNN have allowed the option of combining U-Net with a different backbone for the encoder such as ResNet as well as the inclusion of batch normalization.

Multiple studies have applied U-Net on prostate-related tasks. Three U-Net models were applied to prostate MRI in a study by Bardis et al. in segmenting the prostate, PZ, and TZ separately [19]. The ground-truth was set by a radiologist and achieved Dice scores of 0.94, 0.91, and 0.76 for the prostate, TZ, and PZ segmentation respectively. A paper published in 2019 by Yoon et al. [20] used a CNN pipeline for segmenting the prostate organ, lesion detection, and then make a PI-RADS scoring. The study utilized U-Net for prostate organ segmentation then used R-CNN for lesion detection and segmentation. Moreover, a single board-certified radiologist was used as the ground truth for the manual lesion segmentation. The study achieved lesion segmentation with a Dice score of 0.76. The study by Sanford et al. published in 2020 was a development study that used CNN with the goal of predicting PI-RADS classification [21]. The lesions were segmented and bounded by a professional radiologist which was then the input for the CNN model. A study by Youn et al. published in 2021 [22] explored the viability of using deep learning algorithms for PI-RADS scoring. The study reaffirmed that at varying levels of experience of professional radiologists, the assessment would also vary. The study utilized a deep-learning based software called Prostate AI version 1.2.1 which is currently not available for commercial use.

Specific to lesion segmentation, Liu et al. implemented fuzzy Markov random fields and attained a Dice score of 0.62. A pathologist was used as the baseline for ground-truth and the PZ was manually outlined to be used for segmentation [23]. Kohl et al. used U-Net implemented with adversarial network and achieved a Dice score of 0.41. They used a radiologist for this study to establish their ground-truth. [24]. Dai et al. implemented Mask R-CNN and achieved a Dice score of 0.46 while using a clinician to establish their ground-truth [25]. These studies all have varying ways of establishing their ground-truth and the results show that lesion segmentation remains to be an area with promise for improvement.

V. Methodology

A. Image Dataset

The relevant DICOM files were acquired through the QPR dataset [10] which includes segmented mpMRI images of 15 patients with two studies per patient done two months apart. Images extracted that appear to be completely black were removed.

3D Slicer, an open source image computing platform developed for image analysis and visualization [26], was used to view the DICOM files. Within 3D Slicer, the ADC map with their corresponding segmentations were extracted. For each extracted image, the lesion segmentation was set to red within 3D Slicer. This is to maintain a high contrast of the lesion from the original grayscale image to assist in creating the mask.

Even if the patient has a confirmed lesion in the PZ, the lesion can only be seen during specific frames in the ADC sequence. Specifically, only 14.98% of the total images

have the lesion present in them. Only images with the lesion showing was used. The summary of images with lesion and the distribution from the total images can be seen in Table I.

TABLE I. PERCENT DISTRIBUTION OF IMAGES WITH LESION

Total Images	With Lesion	Distribution
434	65	14.98%

The ADC maps are available in grayscale of dimensions 272x672 and was cropped into 256x256. Using OpenCV and HSV threshold, the grayscale images with red lesion segmentations were transformed to images with binary masking having a white mask for the lesion on top of a black background. These masks served as the ground-truth for model training.

The Albumentations library [27] was used to implement the various data augmentations. To assess the effect of varying augmentation techniques for the segmentation model, we trained the model with varying levels of augmentation. The augmentations are categorized as the Isensee et al., and extended realistic augmentations.

The following augmentations were adopted from the Isensee et al., (2018) paper:

1) Horizontal Flip - This flips the image horizontally on the y-axis. This particular augmentation is applicable for the prostate as the original MRI scans are symmetric in nature as opposed to medical images wherein regional location is critical (e.g. images with the heart on the left side of an image).
2) Rotate - This rotates the image randomly in degrees within a lower and upper threshold. The threshold was set at -45 to 45 as to maintain realistic scenarios for mpMRI prostate scans (e.g. It would be unlikely to have a scan that would be 90 degrees).
3) Random Sized Crop - This crops the image and scales it back to its original dimensions. The minimum height was set at 162 and the maximum at 192. This allowed the PZ to be captured regardless of where the cropping would occur. This would then be scaled back to 256x256.
4) Random Gamma - This applies a random gamma pixel-wise adjustment on the image. This augmentation technique is viable for the prostate mpMRI scans caused by varying equipment used by different hospitals.
5) Elastic Transform - This applies a more liberal distortion to the image. This can simulate the varying transformations that naturally occur within the prostate.

Other realistic augmentation techniques were implemented to the dataset. These are augmentations that are realistic scenarios for variations in mpMRI scans.

1) Brightness, (BC) - This alters the image's brightness and contrast. This accounts for the variation in the equipment used.
2) Blur - This blurs the input image using a median filter with a random aperture linear size. This accounts for the variation in quality of scans from different equipment.
3) Grid Distortion (GD) - Random distortions are applied per grid width and height with a maximum magnitude. This accounts for the patient's natural movement during breathing as well as expansion and contraction of body parts such as the rectum.

B. Segmentation

Due to similarities in the dataset and problem being tackled, we chose U-Net as the main architecture for this study. The flexible architecture also allowed us to modify the encoder with ResNet. The architecture comprises of two major segments: the encoding path which consists of four down-convolutions of 2x2 max pooling and a decoding path which is a set of four 2x2 up-convolutions. Batch normalization was applied before each activation. The γ and β initializers were set to their default values of 1 and 0.

For the model parameters, we used an Adam optimizer with batch size set to 16 and learning rate set at 1×10^{-4}. The initial epoch was set at 300. For the image sampling, we used a train-test-validation split of 70-15-15.

To avoid overfitting the model, EarlyStopping was added to the training on the validation loss metric with a patience of 30 epochs. This meant that if validation loss did not decrease for 30 straight epochs, training will prematurely terminate and not continue to train for the remaining epochs.

A larger learning rate could possibly result in non-convergence of the model. This was mitigated by adding ReduceLROnPlataeu which decreases the learning rate if there is no improvement on the validation loss with a patience of 20 epochs. The maximum decreased learning rate was set at 1×10^{-7} with a reduction factor of 0.1.

Dice score (F1 Score) is a segmentation performance metric that compares the similarity of the predicted mask with the ground-truth mask. This is calculated by having double the intersection of the pixels of the predicted and ground-truth mask divided by the total number of pixels of the predicted mask and the ground-truth mask as shown in (1). The lesion segmentation performance of the model was measured with the average Dice score for the test images.

$$Dice = \frac{2\,|A \cap B|}{|A| + |B|} \tag{1}$$

VI. RESULTS AND DISCUSSION

Different encoders and model variations were applied to the dataset with Isensee augmentations applied. The results can be seen in Table II.

TABLE II. SUMMARY OF SCORES FOR DIFFERENT MODELS

Model + Backbone	Dice
U-Net	0.39
U-Net + ResNet34	0.36
U-Net + SE-ResNet152	0.30
U-Net + SE-ResNet18	0.26
U-Net + ResNet152	0.26
U-Net + Attention	0.06

The baseline U-Net still outperformed the other variations. The addition of Attention was also not beneficial to this dataset. The decrease in scores when the model gets more complicated could be attributed to the low quantity of images in the dataset. This further validated the statement from the Isensee et al. paper that the generic U-Net architecture can be competitive in segmentation given a proper framework [15].

The model was trained on all possible combinations of Isensee (base augmentation), random brightness and contrast (BC), blur, and grid distortion (GD) augmentations to the base U-Net. The effects of adding augmentations to the base U-Net can be seen in Table III.

TABLE III. SUMMARY OF SCORES FOR DIFFERENT AUGMENTATIONS

Augmentations	Dice
Isensee	0.392
Isensee + BC	0.411
Isensee + Blur	0.466
Isensee + GD	0.467
Isensee + BC + Blur	0.416
Isensee + BC + GD	0.411
Isensee + Blur + GD	0.256
Isensee + BC + Blur + GD	0.403

Adding realistic augmentations improved the overall score compared to the baseline. However, the combination of using blur and GD together did not result in an increase in score. We believe that the generated training set from the combination of blur and GD augmentations became too distant from the features that were present within the original images while the BC augmentations were not enough to significantly alter them.

The highest Dice score of 0.467 was achieved with Isensee + GD augmentations followed closely by the Isensee + blur augmentations with a Dice score of 0.466. To verify the performance of the framework with Isensee + GD and Isensee + Blur augmentations, each model was trained thrice. Each cycle had new sets of training augmentations with randomized rotations and distortion values fit within the range and tested on the same test images. Dice scores from the Isensee + GD and Isensee + blur runs can be seen in Table IV and Table V, respectively. The results show that the Isensee + GD runs were more consistent than the Isensee + blur runs and implies that the more aggressive augmentation works better for this particular dataset.

TABLE IV. COMBINATION OF ISENSEE AND GD DICE SCORE

Dice 1	Dice 2	Dice 3	Average
0.47	0.44	0.49	0.47

TABLE V. COMBINATION OF ISENSEE AND BLUR DICE SCORE

Dice 1	Dice 2	Dice 3	Average
0.47	0.37	0.40	0.41

The average Dice score of the Isensee + GD model was then compared to previously done studies. This can be seen in Table VI. For this comparison, only the segmentation results that had a similar approach and methodology were considered. The result from the Liu et al. paper [23] was not considered since the their methodology included a preliminary process of manually outlining the PZ while only considering said pixels for segmentation.

TABLE VI. SUMMARY OF SCORES COMPARED TO LITERATURE

Model	Dice
Mask R-CNN (Yoon et al., 2019)	0.76
U-Net in this study	0.47
Mask R-CNN (Dai et al., 2020)	0.46
U-Net (Kohl et al., 2017)	0.41

While this comparison involved studies that were established using different ground-truths, the aim was to have a more standardized comparison with this framework with potential areas for improvement.

VII. CONCLUSION

In this study, we provided a repeatable framework for prostate lesion segmentation that can be improved and compared with future studies. We used the ADC map and isolated the images that contained lesion as the input for the model trained using various encoders and augmentation combinations. The baseline U-Net with batch normalization trained on images augmented with a combination of Isensee (horizontal flip, random rotation, random sized crop, random gamma, elastic transform) and grid distortion augmentations with batch size of 16 and LR set at 1×10^{-4} using Adam optimizer for Dice loss performed best and achieved an average Dice score of 0.47 (0.44-0.49). Furthermore, the QPR dataset shows promise in being a viable standardized dataset for future testing and benchmarking as shown by the comparison of results to other published studies.

VIII. RECOMMENDATIONS FOR FUTURE WORK

Due to hardware limitations, the study only implemented a hold-out method. However, it is worth considering to implement a k-fold cross-validation technique for performance metrics evaluation and fine-tuning of parameters. Moreover, other models with different encoders could be considered. For the dataset, new augmentation techniques and various combinations may be further explored. Lastly, the use of other mpMRI views aside from the ADC map could be looked into.

REFERENCES

[1] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021.

[2] C Tempany, P Carrol, and M Leapman. The role of magnetic resonance imaging in prostate cancer. *UpToDate. Waltham (MA): UpToDate*, 2018.

[3] Fritz H Schröder, Jonas Hugosson, Monique J Roobol, Teuvo LJ Tammela, Marco Zappa, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Lissa Määttänen, Hans Lilja, et al. The european randomized study of screening for prostate cancer–prostate cancer mortality at 13 years of follow-up. *Lancet*, 384(9959):2027, 2014.

[4] Ian E Haines and George L Gabor Miklos. Prostate-specific antigen screening trials and prostate cancer deaths: the androgen deprivation connection. *Journal of the National Cancer Institute*, 105(20):1534–1539, 2013.

[5] Jared S Winoker, Peter A Pinto, and Ardeshir R Rastinehad. Mri to guide biopsies or avoid biopsies? *Current opinion in urology*, 28(6):522–528, 2018.

[6] Sangeet Ghai and Masoom A Haider. Multiparametric-mri in diagnosis of prostate cancer. *Indian journal of urology: IJU: journal of the Urological Society of India*, 31(3):194, 2015.

[7] Martin Eklund, Fredrik Jäderling, Andrea Discacciati, Martin Bergman, Magnus Annerstedt, Markus Aly, Axel Glaessgen, Stefan Carlsson, Henrik Grönberg, and Tobias Nordström. Mri-targeted or standard biopsy in prostate cancer screening. *New England Journal of Medicine*, 385(10):908–920, 2021.

[8] Andrei S Purysko, Ronaldo H Baroni, Francesco Giganti, Daniel Costa, Raphaël Renard-Penna, Chan Kyo Kim, and Steven S Raman. Pi-rads version 2.1: a critical review, from the ajr special series on radiology reporting and data systems. *American Journal of Roentgenology*, 216(1):20–32, 2021.

[9] Michelle D Bardis, Roozbeh Houshyar, Peter D Chang, Alexander Ushinsky, Justin Glavis-Bloom, Chantal Chahine, Thanh-Lan Bui, Mark Rupasinghe, Christopher G Filippi, and Daniel S Chow. Applications of artificial intelligence to prostate multiparametric mri (mpmri): Current and emerging trends. *Cancers*, 12(5):1204, 2020.

[10] Andriy Fedorov, Michael Schwier, David Clunie, Christian Herz, Steve Pieper, Ron Kikinis, Clare Tempany, and Fiona Fennessy. An annotated test-retest collection of prostate multiparametric mri. *Scientific data*, 5(1):1–13, 2018.

[11] Guan Huang, Gerald Lebovic, and Paraskevi A Vlachou. Diagnostic value of ct in detecting peripheral zone prostate cancer. *American Journal of Roentgenology*, 213(4):831–835, 2019.

[12] Jeffrey C Weinreb, Jelle O Barentsz, Peter L Choyke, Francois Cornud, Masoom A Haider, Katarzyna J Macura, Daniel Margolis, Mitchell D Schnall, Faina Shtern, Clare M Tempany, et al. Pi-rads prostate imaging–reporting and data system: 2015, version 2. *European urology*, 69(1):16–40, 2016.

[13] Tim Hulsen. An overview of publicly available patient-centered prostate cancer datasets. *Translational Andrology and Urology*, 8(Suppl 1):S64, 2019.

[14] Alex Hernández-García and Peter König. Further advantages of data augmentation on convolutional neural networks. In *International Conference on Artificial Neural Networks*, pages 95–103. Springer, 2018.

[15] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No new-net. In *International MICCAI Brainlesion Workshop*, pages 234–244. Springer, 2018.

[16] Dominic Masters and Carlo Luschi. Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*, 2018.

[17] Ibrahem Kandel and Mauro Castelli. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT express*, 6(4):312–315, 2020.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[19] Michelle Bardis, Roozbeh Houshyar, Chanon Chantaduly, Karen Tran-Harding, Alexander Ushinsky, Chantal Chahine, Mark Rupasinghe, Daniel Chow, and Peter Chang. Segmentation of the prostate transition zone and peripheral zone on mr images with deep learning. *Radiology: Imaging Cancer*, 3(3):e200024, 2021.

[20] Choongheon Yoon, Jasper Van, Michelle Bardis, Param Bhatter, Alexander Ushinsky, Justin Glavis-Bloom, Chanon Chantaduly, Daniel S Chow, Roozbeh Houshyar, and Peter Chang. Automated prostate lesion detection and pi-rads assessment with deep learning., 2019.

[21] Thomas Sanford, Stephanie A Harmon, Evrim B Turkbey, Deepak Kesani, Sena Tuncer, Manuel Madariaga, Chris Yang, Jonathan Sackett, Sherif Mehralivand, Pingkun Yan, et al. Deep-learning-based artificial intelligence for pi-rads classification to assist multiparametric prostate mri interpretation: A development study. *Journal of Magnetic Resonance Imaging*, 52(5):1499–1507, 2020.

[22] Seo Yeon Youn, Moon Hyung Choi, Dong Hwan Kim, Young Joon Lee, Henkjan Huisman, Evan Johnson, Tobias Penzkofer, Ivan Shabunin, David Jean Winkel, Pengyi Xing, et al. 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. *European Journal of Radiology*, 142:109894, 2021.

[23] Xin Liu, Deanna L Langer, Masoom A Haider, Yongyi Yang, Miles N Wernick, and Imam Samil Yetik. Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class. *IEEE Transactions on Medical Imaging*, 28(6):906–915, 2009.

[24] Simon Kohl, David Bonekamp, Heinz-Peter Schlemmer, Kaneschka Yaqubi, Markus Hohenfellner, Boris Hadaschik, Jan-Philipp Radtke, and Klaus Maier-Hein. Adversarial networks for the detection of aggressive prostate cancer. *arXiv preprint arXiv:1702.08014*, 2017.

[25] Zhenzhen Dai, Eric Carver, Chang Liu, Joon Lee, Aharon Feldman, Weiwei Zong, Milan Pantelic, Mohamed Elshaikh, and Ning Wen. Segmentation of the prostatic gland and the intraprostatic lesions on multiparametic magnetic resonance imaging using mask region-based convolutional neural networks. *Advances in radiation oncology*, 5(3):473–481, 2020.

[26] Steve Pieper, Michael Halle, and Ron Kikinis. 3d slicer. In *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*, pages 632–635. IEEE, 2004.

[27] Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020.