Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset

Renato Cuocolo, MD, PhD,[1,2] Albert Comelli, PhD,[3] Alessandro Stefano, PhD,[4] Viviana Benfante, MSc,[4] Navdeep Dahiya, PhD,[5] Arnaldo Stanzione, MD,[6*] Anna Castaldo, MD,[6] Davide Raffaele De Lucia, MD,[6] Anthony Yezzi, PhD,[5] and Massimo Imbriaco, MD[6]

Background: Prostate volume, as determined by magnetic resonance imaging (MRI), is a useful biomarker both for distinguishing between benign and malignant pathology and can be used either alone or combined with other parameters such as prostate-specific antigen.
Purpose: This study compared different deep learning methods for whole-gland and zonal prostate segmentation.
Study Type: Retrospective.
Population: A total of 204 patients (train/test = 99/105) from the PROSTATEx public dataset.
Field strength/Sequence: A 3 T, TSE T_2-weighted.
Assessment: Four operators performed manual segmentation of the whole-gland, central zone + anterior stroma + transition zone (TZ), and peripheral zone (PZ). U-net, efficient neural network (ENet), and efficient residual factorized ConvNet (ERFNet) were trained and tuned on the training data through 5-fold cross-validation to segment the whole gland and TZ separately, while PZ automated masks were obtained by the subtraction of the first two.
Statistical Tests: Networks were evaluated on the test set using various accuracy metrics, including the Dice similarity coefficient (DSC). Model DSC was compared in both the training and test sets using the analysis of variance test (ANOVA) and post hoc tests. Parameter number, disk size, training, and inference times determined network computational complexity and were also used to assess the model performance differences. A $P < 0.05$ was selected to indicate the statistical significance.
Results: The best DSC ($P < 0.05$) in the test set was achieved by ENet: 91% ± 4% for the whole gland, 87% ± 5% for the TZ, and 71% ± 8% for the PZ. U-net and ERFNet obtained, respectively, 88% ± 6% and 87% ± 6% for the whole gland, 86% ± 7% and 84% ± 7% for the TZ, and 70% ± 8% and 65 ± 8% for the PZ. Training and inference time were lowest for ENet.
Data Conclusion: Deep learning networks can accurately segment the prostate using T_2-weighted images.
Evidence Level: 4
Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2021.

The assessment of prostate volume (PV) has shown clinical relevance for benign as well as malignant pathologies.[1–3] In particular, PV is employed to calculate biomarkers suggestive of clinically significant prostate cancer (csPCa), usually in combination with serum levels of prostate-specific antigen (PSA).[4,5] Among these, PSA density (PSAd, the ratio between PSA and PV) has been object of numerous investigations, both alone and in combination with MRI, and is included as a

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.27585

Received Dec 16, 2020, Accepted for publication Feb 16, 2021.

*Address reprint requests to: A.S., via Pansini 5, 80131, Naples, Italy. E-mail: arnaldo.stanzione@unina.it

From the [1]Department of Clinical Medicine and Surgery, University of Naples "Federico II", Naples, Italy; [2]Laboratory of Augmented Reality for Health Monitoring (ARHeMLab), Department of Electrical Engineering and Information Technology, University of Naples "Federico II", Naples, Italy; [3]Ri.MED Foundation, Palermo, Italy; [4]Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Cefalù, Italy; [5]Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA; and [6]Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy

Additional supporting information may be found in the online version of this article

valuable biomarker in current prostate cancer guidelines.[4,6–8] Interestingly, studies have proposed transition zone PSA density (TZPSAd) as a more accurate indicator of csPCa, compared to PSAd.[9,10] More recently, the prostate volume index (PVI, the ratio between transition zone and peripheral zone volumes) has been investigated for both inflammatory and neoplastic prostate pathologies, with promising results.[5]

Current Prostate Imaging Reporting and Data System (PI-RADS) guidelines suggest the use of an ellipsoid formula for calculation of PV from MRI exams.[11] While this formula has shown good correlation with PV, obtained from manual segmentation or surgical specimens, it is not totally accurate, especially for grossly hypertrophic glands. For these, other formulas have been proposed, with unsatisfactory results.[12,13] Currently, no recommended methods are available for zonal anatomy segmentation outside of manual, slice-by-slice annotation.

Machine learning (ML), especially deep learning (DL), allows for an automated analysis of medical images. These techniques have been object of great interest in radiology and prostate imaging.[14–16] Among the tasks that can be addressed by ML and DL is that of automated prostate segmentation, as shown in previous works.[17,18] However, DL networks often require high computational power and training time, limiting clinical applicability.[19,20]

The aim of this study was to compare different DL networks in the task of automated whole-gland and zonal anatomy segmentation of prostate MRI exams, taking into account their accuracy, training time, and hardware requirements.

Materials and Methods

Dataset

For the present study, public data from the Radboud University collected in the PROSTATEx dataset and available from "The Cancer Imaging Archive" repository were employed to train and test the DL models.[21,22] Specifically, axial T_2-weighted images were used, with a 0.5×0.5 mm in-plane resolution and 3.6 mm slice thickness. The prostate MRI exams included in this dataset were acquired on 3 T scanners (MAGNETOM Trio and Skyra, Siemens Healthineers, Erlangen, Germany) and are organized in training and test sets. In detail, 99 cases from the PROSTATEx dataset were used as the training set in a 5-fold cross-validation fashion and the remaining 105 patients as the test set. The latter includes patients who have undergone biopsy for suspect of csPCa.

The training dataset was object of a case-by-case quality control paired with manual segmentation of the 1) whole-gland, 2) central zone + anterior stroma + transition zone (TZ), and 3) peripheral zone (PZ). These segmentations were performed by two radiology residents (2 years of experience in genitourinary imaging, A.C. and D.R.D.L.) and two radiologists with over 5 years of experience in genitourinary imaging (R.C. and A.S.), organized in resident-radiologist pairings operating in consensus using a dedicated software (ITKSnap, v3.8).[23] After manual segmentation, the masks underwent an automated cluster removal with the FMRIB Software Library (v5.0).[24] All segmentations are freely available in an online repository (https://rcuocolo.github.io/PROSTATEx_masks/).

Deep Learning

Due to variations in MRI acquisition protocols (i.e. field of view), images had different matrices (ranging from 320×320 to 720×720) while DL networks typically require same-sized inputs. MRI exams were resampled to an isotropic voxel size ($1 \times 1 \times 1$ mm^3) with a matrix resolution of 512×512 using linear interpolation. Manually segmented masks were used as the ground truth (whole-gland, TZ and PZ) were resampled using nearest-neighbor interpolation. Data augmentation was used to reduce overfitting: six transformations were applied by randomly rotating, translating in both x and y directions, and applying shearing, horizontal flip and zooming to the input training slices. Finally, image normalization was performed using training data feature-wise (pixel-wise) mean and standard deviation: image normalized = (image - mean)/standard deviation.

The performances of U-net, where all 3×3 convolutions were replaced by larger 5×5 convolution operators as reported in previous studies, and two DL networks namely efficient neural network (ENet), and efficient residual factorized ConvNet (ERFNet) were compared for the prostate segmentation task.[25–28]

All methods were implemented using the open-source programming language Python (https://www.python.org/) and a high-end HPC system equipped with a GPU (NVIDIA QUADRO P4000 with 8 GB of RAM, 1792 CUDA Cores). A learning rate of 0.0001 for E-Net model and 0.00001 for U-Net models with Adam optimizer was used.[29] A batch size of eight slices for all experiments, and $\alpha = 0.3$ and $\beta = 0.7$ for the loss function were used. During the training process, each network was allowed to train for a maximum of 100 epochs. The training loss was used as a stopping criterion. When the training loss did not decrease for 10 epochs continuously, training was stopped.

A k-fold ($k = 5$), cross-validation was used to train all networks. Within the cross-validation process, slices from the same patient were never included in both training and testing folds. The performance of each DL model was calculated by averaging the results of all validation folds. Finally, the performance metrics for the DL models were calculated on the test set, entirely constituted by previously unseen data. This whole process was performed in parallel for the whole-gland and TZ masks while the PZ masks were obtained by subtraction. Finally, the Tversky loss function was used as loss function to make adjustment of the weights of false positives and false negatives.[30]

Statistical Analysis

Continuous variables are presented as mean ± standard deviation. The mean positive predictive value (PPV), DSC, volume overlap error (VOE), relative volume difference (VD), and average symmetric surface distance (ASSD) were calculated. To test the significance of differences between the three DL algorithms' performance, an analysis of variance (one-way ANOVA) with Tukey HSD, Scheffé, and Bonferroni/Holm post hoc tests were used as multiple comparison correction techniques on the DSC. A P-value ≤ 0.05 was considered as statistically significant, adjusted for multiple comparisons, as appropriate. Finally, four variables (i.e. number of parameters, size on disk, training times, and inference times) commonly used to describe the computational complexity and performance of a DL network were

TABLE 1. Performance Results on the Testing Set Using ENet, UNet, and ERFNet Methods for the Whole Prostate Zone Segmentation

	Sensitivity (%)	PPV (%)	DSC (%)	VOE (%)	VD (%)	ASSD
ENet						
Mean	92.28	89.67	90.63	16.93	3.95	1.21
±std	5.24	7.01	3.90	5.83	14.21	4.87
±CI (95%)	0.46	0.62	0.35	0.52	1.26	0.43
UNet						
Mean	84.49	93.09	88.07	20.83	−8.84	1.06
±std	10.19	4.44	6.19	8.72	13.22	0.54
±CI (95%)	0.90	0.39	0.55	0.77	1.16	0.05
ERFNet						
Mean	92.69	83.82	87.25	22.15	13.17	1.62
±std	8.12	10.05	6.06	8.61	23.07	2.11
±CI (95%)	0.72	0.89	0.54	0.76	2.04	0.19

calculated to evaluate the differences between the three models, using both central processing unit (CPU) and graphics processing unit (GPU) hardware.[25,26]

Results

In total, axial T_2-weighted images of 204 patients were processed obtaining three different zone segmentation masks. Supplementary Tables S1–S3 present all accuracy metrics for each network for whole-gland, TZ, and PZ segmentation in the training set. ENet had the highest DSC compared to both UNet and ERFNet for all segmentations: 90% ± 3% for the whole-gland segmentation, 85% ± 5% for the TZ, and 75 ± 8% for the PZ. Tables 1–3 present the accuracy metrics in the test dataset, with ENet showing the highest DSC. Specifically, average DSC was 91% ± 4% for the whole prostate, 87% ± 5% for the TZ, and 71% ± 8% for the PZ (Figs. 1 and 2).

TABLE 2. Performance Results on the Testing Set Using ENet, UNet, and ERFNet Methods for the Transition Zone Segmentation

	Sensitivity (%)	PPV (%)	DSC (%)	VOE (%)	VD (%)	ASSD
ENet						
Mean	88.57	86.33	86.92	22.77	4.11	1.58
±std	7.43	8.90	5.35	7.69	16.98	5.03
±CI (95%)	0.66	0.79	0.47	0.68	1.51	0.45
UNet						
Mean	81.95	91.36	85.79	24.40	−9.57	1.25
±std	10.28	6.30	6.49	8.65	15.34	0.69
±CI (95%)	1.17	0.72	0.74	0.99	1.75	0.08
ERFNet						
Mean	91.29	79.12	83.89	27.11	19.63	1.83
±std	6.91	12.27	7.44	10.02	29.42	1.59
±CI (95%)	0.69	1.22	0.74	1.00	2.93	0.16

TABLE 3. Performance Results on the Testing Set Using ENet, UNet, and ERFNet Methods for the Peripheral Zone Segmentation

	Sensitivity (%)	PPV (%)	DSC (%)	VOE (%)	VD (%)	ASSD
ENet						
Mean	76.94	67.49	71.42	43.90	16.47	3.19
±std	8.21	10.15	7.75	9.24	21.54	3.20
± CI (95%)	0.76	0.94	0.72	0.85	1.99	0.30
UNet						
Mean	73.81	67.23	69.91	45.63	11.91	2.96
±std	9.00	10.62	8.43	9.80	18.70	2.19
±CI (95%)	1.07	1.26	1.00	1.16	2.21	0.26
ERFNet						
Mean	74.13	58.15	64.55	51.89	30.82	3.72
±std	9.39	9.56	7.57	8.34	26.83	2.74
±CI (95%)	1.01	1.03	0.82	0.90	2.89	0.30

ANOVA showed a significant difference between network DSC ($P < 0.05$). The post hoc analysis showed significant differences in among all DL model pairings in both the training and test sets (Tables 4 and 5; Supplementary Tables S4 and S5).

Supplementary Table S6, through parameters describing the computational complexity and performance of the three models, shows that ENet is faster, more efficient in terms of training time, and more compact than other models. Specifically, E-Net has an order of magnitude lower than the parameters of

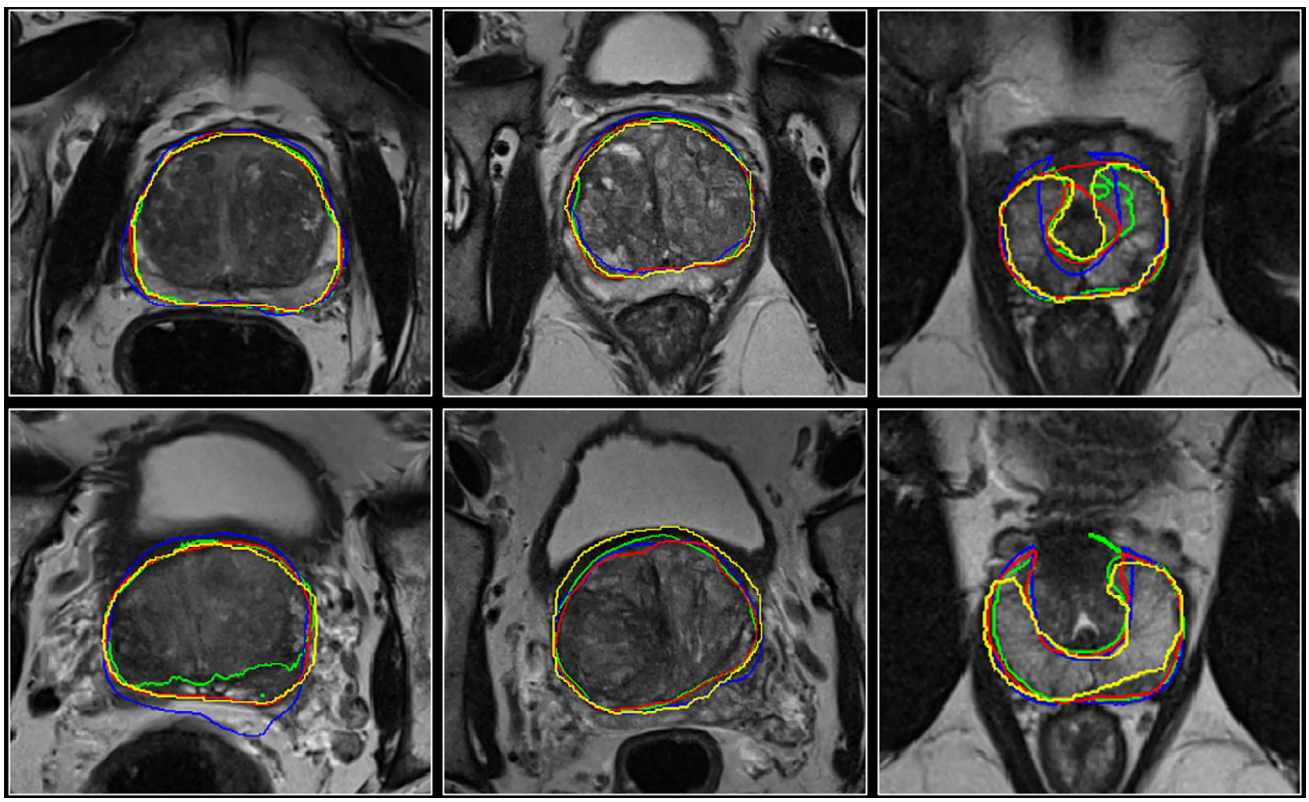

FIGURE 1: Examples of correct segmentation for the deep learning networks for the whole gland (left column), transition (middle column) and peripheral zone (right column). The manual segmentation (yellow), ENet (red), ERFNet (blue), and U-Net (green) are superimposed.

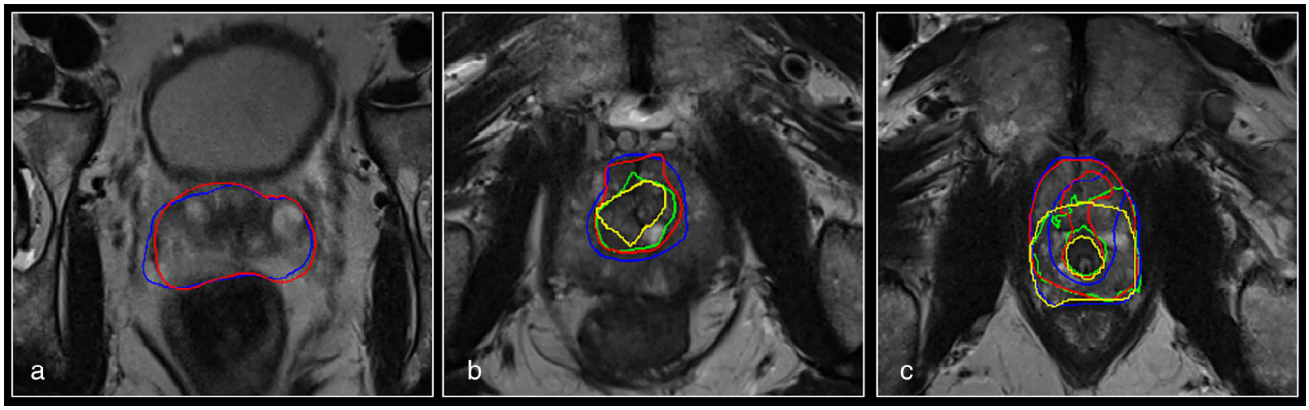

FIGURE 2: Examples of erroneous automated segmentation, possibly due to motion artifacts (a), previous surgery (b), and complexity of the prostate apex region (c). The manual segmentation (yellow), ENet (red), ERFNet (blue), and U-Net (green) are superimposed.

TABLE 4. ANOVA on the DSC Showed Statistical Differences Between Segmentation Methods in the Training Data

ANOVA	F value	F critic value	*P*-value
ENet vs ERFNet vs UNet	25.50	3.028	<0.05
ENet vs ERFNet	53.450	3.891	<0.05
ERFNet vs UNet	13.601	3.891	<0.05
ENet vs UNet	11.089	3.891	<0.05

U-Net. Consequently, the disk size required by the algorithm ranges from about 6 MB (E-Net) to 65 MB (U-Net). The time to train the models using the GPU device (NVIDIA QUADRO P4000, 8 GB VRAM, 1792 CUDA Cores) ranges from 30,780 seconds (E-Net) to 123,420 seconds (U-Net). Using the CPU device (Intel(R) Xeon(R) W-2125 CPU 4.00GHz processor), 308,793 and 5,646,940 seconds for E-Net and U-Net, respectively. Delineation time using the GPU was around a second for all three models. However, because a scientific GPU is not always available in desktop computers due to the high cost, the delineation is usually performed on CPU. In this case, E-Net took about 6 seconds while U-Net was 10 times slower.

In terms of convergence, Figures 3 and 4 show the profiles of DSC and Tversky loss function for the training and test sets, respectively. ENet presents the faster convergence compared to both ERFNet and UNet, reaching a DSC of nearly 0.8 in less than 12 epochs in the training data and 18 epochs in the test one.

Discussion

In this study, the feasibility and efficacy of three DL models for the segmentation of the prostate and its zones were assessed accounting for accuracy of results, speed of training, and hardware requirements. For all networks, the Tversky loss function was used rather than the DSC that is routinely used as a loss function in DL-based medical image segmentation networks.[30] It measures the overlap between predicted and ground truth segmentation masks, calculating the harmonic

TABLE 5. ANOVA on the DSC Showed Statistical Differences Between Segmentation Methods in the Test Set

ANOVA	F value	F critic value	*P*-value
ENet vs ERFNet vs UNet	50.763	3.002	<0.05
ENet vs ERFNet	107.456	3.851	<0.05
ERFNet vs UNet	7.161	3.851	<0.05
ENet vs UNet	53.705	3.851	<0.05

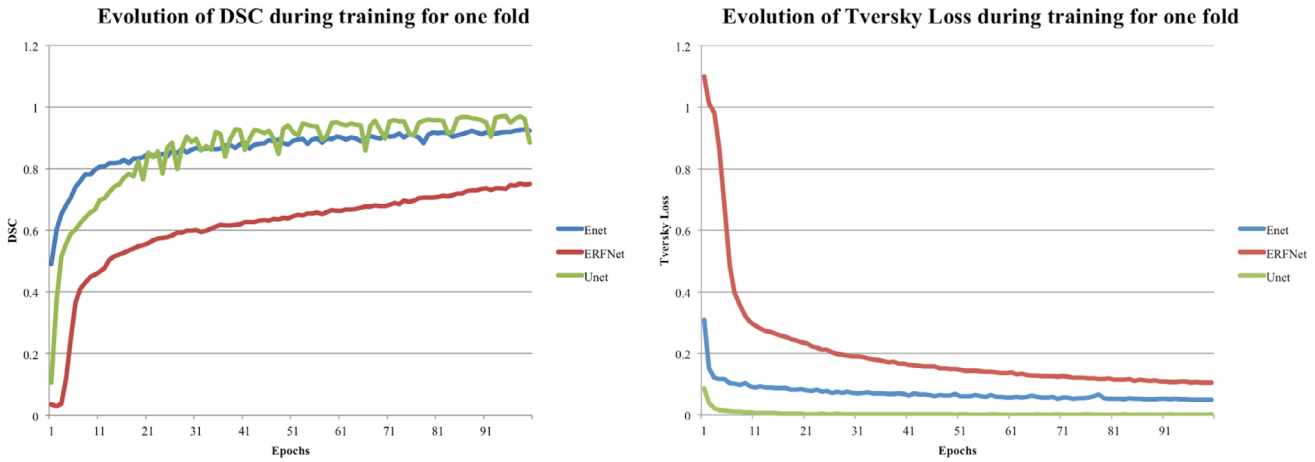

FIGURE 3: Training DSC and Tversky loss plots for the deep learning networks for whole-gland prostate segmentation.

mean of false positives and false negatives and weighting both equally. Nevertheless, DL methods are affected by imbalanced data, i.e. the number of voxels belonging to the target is small compared to the background, as in this study. For this reason, using DSC as loss function, networks may have difficulty in learning a reliable feature representation of the foreground class tending to predict most voxels as belonging to the background class. In this study, to deal with this issue and learn more effectively, the Tversky loss function was used to provide a larger weight to foreground voxels.[30]

All networks were able to accurately segment the whole gland and its zones when compared to manual segmentations. Among tested models, the key differences were 1) the ENet and UNet were more accurate than ERFNet; 2) the ENet converged faster than both the ERFNet and UNet; and 3) the ENet model had an order of magnitude fewer parameters than UNet.

Using CPU hardware, ENet was much faster than UNet. E-Net uses batch normalization layers with not-trainable parameters and their gradients are not backpropagated during training. Consequently, the reduced time and space required by E-Net to calculate delineations makes it more compatible with clinical workflows installed on regular desktop computers or even mobile devices.

In terms of accuracy, results were in line with those found in recent literature. Lee et al reported a DSC of 87% for whole-gland segmentation and 77% for the TZ, while Sanford obtained 93% and 89%, respectively.[17,18] However, these authors employed computationally expensive DL networks trained on cloud-based hardware, also requiring high-speed Internet availability. Findings of the present study show that smaller, easier to train and less computationally expensive networks like ENet can achieve similar results to state-of-the-art solutions. The present findings suggest that more efficient networks should probably receive greater consideration in the future as to lower the barrier of entrance and facilitate their implementation in a wider range of clinical practice settings. Also, the lower training loss of UNet compared to ENet and ERFNet suggests the presence of overfitting, although the number of UNet filters in each layer was reduced.

It is interesting to note that the performance of the models was impaired in those patients previously treated for

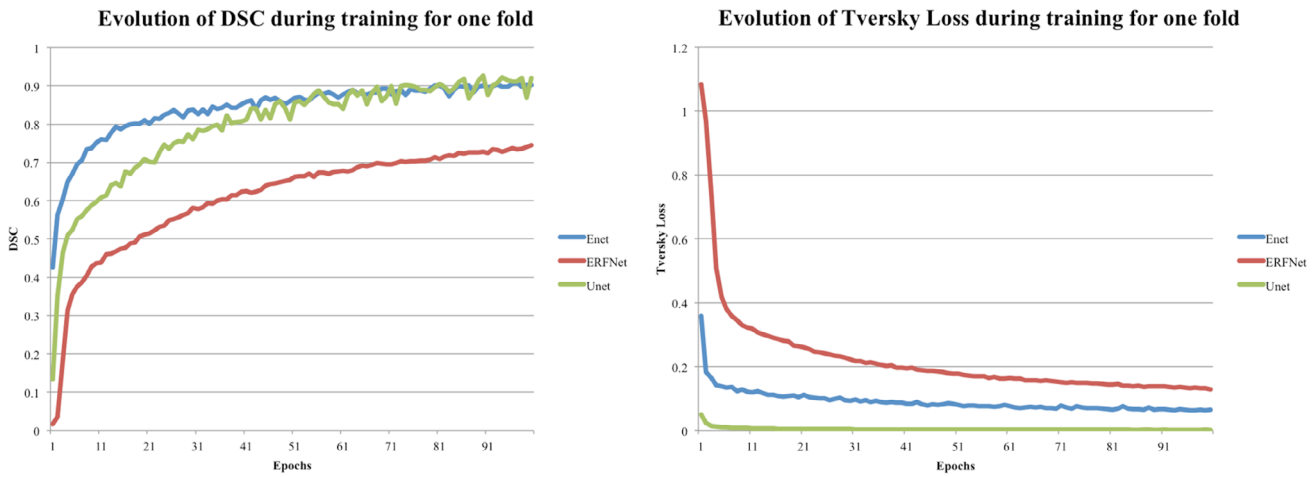

FIGURE 4: Training DSC and Tversky loss plots for the prostate transition zone segmentation.

benign prostatic hyperplasia with transurethral resection of the prostate (TURP). Indeed, TURP-related changes in the transition zone (either partially or completely absent) care observed at MRI.[31] It can be hypothesized that the training data did not contain enough of these cases to allow for correct modeling of this situation. However, since TURP is still considered the surgical cornerstone for benign prostatic hyperplasia treatment, performing mpMRI for csPCa in patients who underwent TURP is not uncommon.[32] Therefore, there are some concerns on the applicability of automated segmentation models to these subjects, which should be considered both in the present work as in future endeavors in this research space. It deserves to be highlighted that TURP-related changes also limit the usefulness of alternative methods to estimate PV, such as the ellipsoid formula. The ideal solution, given sufficient data, could be to develop complementary models for the automated segmentation of patients who have and have not undergone the procedure.

As for the possible clinical implications of this methodology, it can be hypothesized that the fully automated PV calculation could facilitate radiologists in routine daily practice. Indeed, scientific evidence strongly support the adoption of structured reporting for prostate MRI and suggest that the final report should include clinically relevant quantitative data such as PV and PSAd levels.[33,34] Current guidelines recommend estimating PV (and thus PSAd levels) using the maximum diameters measurements and the ellipsoid formula.[11] While this method is fairly accurate in most scenarios and takes radiologists a relatively small time to obtain PV, an automated computation based on whole-gland segmentation would be more efficient and could easily allow to implement structured reporting systems that provide PV and PSAd even before the entire scan is completed since it relies exclusively on axial T_2-weighted images. This appears particularly relevant considering the high volumes of prostate MRI exams currently performed and that are expected to significantly rise in the near future, saving reporting radiologists precious time.[35] Moreover, since a PSAd cut-off for csPCa has been defined, an automated system could also flag suspicious scans to be evaluated with higher priority, further increasing the reporting process efficiency.

The presented models also allow to estimate the TZ and PZ volumes, with noteworthy possible implications for future research endeavors in the field of prostate MRI. In particular, some promising quantitative parameters such as TZPSAd and PVI have been investigated and need to be validated on larger datasets.[5,9,10,36,37] Currently, however, time-efficient methods to accurately estimate TZ volume are lacking, limiting investigations on the topic. By implementing automated tools, it could be possible to rapidly calculate TZPSAd and PVI on multiple, large datasets to verify the actual clinical relevance of such parameters. In this light, this methodology could possibly contribute to further increase the diagnostic value of prostate MRI. Similarly, the Prostate Health Index is another PSA-derived biomarker whose accuracy could be improved by normalization using the PV. Further investigation of this parameter could also be aided by an accurate, automated PV calculation.[38,39]

Limitations

The 3D segmentation was not adopted as this approach requires larger datasets and is more memory demanding. 3D DL networks have extremely prohibitive computational cost. One tool to deal with this is issue is down-sampling. Even with a good scientific GPU, a $128 \times 128 \times 128$ input volume would be the highest resolution feasible in a reasonable amount of time. This lowers the amount of information available in the images and further exacerbates the class imbalance problem, as well as introducing interpolation artifacts. Results also show that a 2D network obtains comparable results to 3D ones. Larger datasets have been used for prostate segmentation in other studies; however, the choice to employ a public dataset allows easier reproducibility of findings as well as a clear benchmark for future studies in this research space.[17,18] In the same vein, all masks are freely available to use by other researchers. To ensure a high quality of the ground truth annotations, a consensus approach by multiple readers was employed. Unfortunately, this did not allow for an assessment of reader variability for the manual segmentation process. Finally, the acquisition parameters were not entirely in line with PI-RADS guidelines. This is a known issue and somewhat limited by the use of 2D images as the in-plane resolution is of diagnostic quality while the main issue was represented by slice thickness (3.6 mm instead of 3 mm).[40] The PROSTATEx dataset was also collected at a high-volume prostate MRI center and is representative of real-world exams.

Conclusion

In conclusion, this study demonstrated that faster and less computationally expensive DL networks can perform accurate prostate segmentation. These tools, if integrated in clinical practice, could facilitate the adoption of novel biomarkers for prostatic oncologic and nononcologic pathology.

References

1. Heidler S, Drerup M, Lusuardi L, et al. The correlation of prostate volume and prostate-specific antigen levels with positive bacterial prostate tissue cultures. Urology 2018;115:151-156.

2. Murphy AB, Nyame YA, Batai K, et al. Does prostate volume correlate with vitamin D deficiency among men undergoing prostate biopsy? Prostate Cancer Prostatic Dis 2017;20:55-60.

3. Kim YM, Park S, Kim J, et al. Role of prostate volume in the early detection of prostate cancer in a cohort with slowly increasing prostate specific antigen. Yonsei Med J 2013;54:1202-1206.

4. Nordström T, Akre O, Aly M, Grönberg H, Eklund M. Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer. Prostate Cancer Prostatic Dis 2018;21:57-63.

5. Porcaro AB, Tafuri A, Sebben M, et al. Prostate volume index is able to differentiate between prostatic chronic inflammation and prostate cancer in patients with Normal digital rectal examination and prostate-specific antigen values <10 ng/mL: Results of 564 biopsy Naïve cases. Urol Int 2019;103:415-422.
6. Fascelli M, Rais-Bahrami S, Sankineni S, et al. Combined Biparametric prostate magnetic resonance imaging and prostate-specific antigen in the detection of prostate cancer: A validation study in a biopsy-naive patient population. Urology 2016;88:125-134.
7. Cuocolo R, Stanzione A, Rusconi G, et al. PSA-density does not improve bi-parametric prostate MR detection of prostate cancer in a biopsy naïve patient population. Eur J Radiol 2018;104:64-70.
8. EAU Guidelines. Edn. presented at the EAU Annual Congress Barcelona 2019 [https://uroweb.org/guideline/prostate-cancer]
9. Schneider AF, Stocker D, Hötker AM, et al. Comparison of PSA-density of the transition zone and whole gland for risk stratification of men with suspected prostate cancer: A retrospective MRI-cohort study. Eur J Radiol 2019;120:108660.
10. Amini E, Pishgar F, Ayati M, et al. Transition zone prostate-specific antigen density could better guide the rebiopsy strategy in men with prostate inflammation at initial biopsy. Urology 2015;86:985-990.
11. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. Eur Urol 2019;76:340-351.
12. Stanzione A, Ponsiglione A, Di Fiore GA, et al. Prostate volume estimation on MRI: Accuracy and effects of ellipsoid and bullet-shaped measurements on PSA density. Acad Radiol 2020.
13. Bezinque A, Moriarity A, Farrell C, Peabody H, Noyes SL, Lane BR. Determination of prostate volume: A comparison of contemporary methods. Acad Radiol 2018;25:1582-1587.
14. Cuocolo R, Cipullo MB, Stanzione A, et al. Machine learning applications in prostate cancer magnetic resonance imaging. Eur Radiol Exp 2019;3:35.
15. Comelli A, Stefano A, Coronnello C, et al. Radiomics: A new biomedical workflow to create a predictive model. In: Papież B, Namburete A, Yaqub M, Noble J, editors. *Med image Underst anal MIUA*, Vol 2020. Cham: Springer; 2020. p 280-293.
16. Giambelluca D, Cannella R, Vernuccio F, et al. PI-RADS 3 lesions: Role of prostate MRI texture analysis in the identification of prostate cancer. Curr Probl Diagn Radiol 2019;50:175-185.
17. Sanford TH, Zhang L, Harmon SA, et al. Data augmentation and transfer learning to improve generalizability of an automated prostate segmentation model. Am J Roentgenol 2020;215:1403-1410.
18. Lee DK, Sung DJ, Kim C-S, et al. Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula. Am J Roentgenol 2020;214:1229-1238.
19. Montagnon E, Cerny M, Cadrin-Chênevert A, et al. Deep learning workflow in radiology: A primer. Insights Imaging 2020;11:22.
20. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018;15:20170387.
21. Armato SG, Huisman H, Drukker K, et al. PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. J Med Imaging 2018;5:1.
22. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. J Digit Imaging 2013;26:1045-1057.
23. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. *2016 38th Annu Int Conf IEEE Eng med biol Soc*: Orlando, FL, USA: IEEE; 2016. p 3342-3345.
24. Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. Neuroimage 2009;45:S173-S186.
25. Comelli A, Coronnello C, Dahiya N, et al. Lung segmentation on high-resolution computerized tomography images using deep learning: A preliminary step for Radiomics studies. J Imaging 2020;6:125.
26. Comelli A, Dahiya N, Stefano A, et al. Deep learning approach for the segmentation of aneurysmal ascending aorta. Biomed Eng Lett 2020.
27. Paszke A, Chaurasia A, Kim S, Culurciello E. ENet: A deep neural network architecture for real-time semantic segmentation. arXiv.org 2016; 1-10.
28. Romera E, Alvarez JM, Bergasa LM, Arroyo R. ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation. IEEE Trans Intell Transp Syst 2018;19:263-272.
29. Kingma DP, Ba J. Adam: A method for stochastic optimization. *3rd International Conference of learn represent ICLR 2015 - Conf track Proc*; San Diego, CA, USA: 2014.
30. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *Lect notes Comput Sci (including Subser Lect notes Artif Intell Lect notes bioinformatics)*; Cham, CH: Springer; 2017. p 379-387.
31. Walker SM, Turkbey B. Role of mpMRI in benign prostatic hyperplasia assessment and treatment. Curr Urol Rep 2020;21:55.
32. Langan RC. Benign prostatic hyperplasia. Prim Care Clin off Pract 2019; 46:223-232.
33. Shaish H. Structured prostate MRI reporting: How and why. Abdom Radiol 2020;45:3969-3973.
34. Stanzione A, Boccadifuoco F, Cuocolo R, et al. State of the art in abdominal MRI structured reporting: A review. Abdom Radiol 2020.
35. Sushentsev N, Caglic I, Sala E, et al. The effect of capped biparametric magnetic resonance imaging slots on weekly prostate cancer imaging workload. Br J Radiol 2020;93:20190929.
36. Porcaro AB, Novella G, Cacciamani G, et al. Prostate volume index associates with a decreased risk of prostate cancer: Results of a large cohort of patients elected to a first biopsy set. Urol Int 2017;98:22-27.
37. Chang T-H, Lin W-R, Tsai W-K, et al. Zonal adjusted PSA density improves prostate cancer detection rates compared with PSA in Taiwanese males with PSA < 20 ng/ml. BMC Urol 2020;20:151.
38. Barisiene M, Bakavicius A, Stanciute D, et al. Prostate health index and prostate health index density as diagnostic tools for improved prostate cancer detection. Biomed Res Int 2020;2020:1-15.
39. Tosoian JJ, Druskin SC, Andreas D, et al. Prostate health index density improves detection of clinically significant prostate cancer. BJU Int 2017;120:793-798.
40. Cuocolo R, Stanzione A, Ponsiglione A, et al. Prostate MRI technical parameters standardization: A systematic review on adherence to PI-RADSv2 acquisition protocol. Eur J Radiol 2019;120:108662.