Journal of Nuclear Medicine, published on November 6, 2020 as doi:10.2967/jnumed.120.254623

Intraprostatic Tumour Segmentation on PSMA-PET Images in Patients with Primary Prostate Cancer with a Convolutional Neural Network

Dejan Kostyszyn [1,2] *, Tobias Fechter [1,3,4] *, Nico Bartl [3,4], Anca L. Grosu [3,4], Christian Gratzke [5], August Sigle [5], Michael Mix [6], Juri Ruf [6], Thomas F. Fassbender [6], Selina Kiefer [7], Alisa S. Bettermann [4], Nils H. Nicolay [3,4], Simon Spohn [3,4], Maria U. Kramer [4], Peter Bronsert [3,7], Hongqian Guo [8], Xuefeng Qiu [8], Feng Wang [9], Christoph Henkenberens [10], Rudolf A. Werner [11], Dimos Baltas [1,3], Philipp T. Meyer [6], Thorsten Derlin [11], Mengxia Chen [8] **, Constantinos Zamboglou [3,4] **

* and ** contributed equally

1. Division of Medical Physics, Department of Radiation Oncology, Medical Center – University of Freiburg, Faculty of Medicine. University of Freiburg, Germany
2. Faculty of Engineering. University of Freiburg, Germany
3. German Cancer Consortium (DKTK). Partner Site Freiburg, Germany
4. Department of Radiation Oncology, Medical Center – University of Freiburg, Faculty of Medicine. University of Freiburg, Germany
5. Department of Urology, Medical Center – University of Freiburg, Faculty of Medicine. University of Freiburg, Germany
6. Department of Nuclear Medicine, Medical Center – University of Freiburg, Faculty of Medicine. University of Freiburg, Germany
7. Institute for Surgical Pathology, Medical Center – University of Freiburg, Faculty of Medicine. University of Freiburg, Germany
8. Department of Urology, Medical School of Nanjing University, Affiliated Drum Tower Hospital, China
9. Department of Nuclear Medicine, Medical School of Nanjing University, Affiliated Drum Tower Hospital, China

10. Department of Radiation Oncology, Hannover Medical School, Hannover, Germany
11. Department of Nuclear Medicine, Hannover Medical School, Hannover, Germany

First author: Dejan Kostyszyn, Division of Medical Physics, Department of Radiation Oncology, Medical Center – University of Freiburg, Robert-Koch Straße 3, 79106 Freiburg, Email: dejan.kostyszyn@uniklinik-freiburg.de, Phone: 0049-761-270-94010

Corresponding author: Dr. Constantinos Zamboglou, Department of Radiation Oncology, Medical Center – University of Freiburg, Robert-Koch Straße 3, 79106 Freiburg, Email: constantinos.zamboglou@uniklinik-freiburg.de, Phone: 0049-761-270-39770

Short running title: Automatic prostate cancer segmentation

Funding: This study was funded from the ERA PerMed call 2018 (BMBF)

Disclosures: All authors confirm that no conflicts of interest exist, that may influence the content of the manuscript.

Word count: 5165

ABSTRACT

Accurate delineation of the intraprostatic gross tumour volume (GTV) is a prerequisite for treatment approaches in patients with primary prostate cancer (PCa). Prostate-specific membrane antigen positron emission tomography (PSMA-PET) may outperform MRI in GTV detection. However, visual GTV delineation underlies interobserver heterogeneity and is time consuming. The aim of this study was to develop a convolutional neural network (CNN) for automated segmentation of intraprostatic tumour (GTV-CNN) in PSMA-PET.

Methods: The CNN (3D U-Net) was trained on 68Ga-PSMA-PET images of 152 patients from two different institutions and the training labels were generated manually using a validated technique. The CNN was tested on two independent internal (cohort 1: 68Ga-PSMA-PET, n=18 and cohort 2: 18F-PSMA-PET, n=19) and one external (cohort 3: 68Ga-PSMA-PET, n=20) test-datasets. Accordance between manual contours and GTV-CNN was assessed with Dice-Sørensen coefficient (DSC). Sensitivity and specificity were calculated for the two internal test-datasets (cohort 1: n=18, cohort 2: n=11) by using whole-mount histology.

Results: Median DSCs for cohorts 1-3 were 0.84 (range: 0.32-0.95), 0.81 (range: 0.28-0.93) and 0.83 (range: 0.32-0.93), respectively. Sensitivities and specificities for GTV-CNN were comparable with manual expert contours: 0.98 and 0.76 (cohort 1) and 1 and 0.57 (cohort 2), respectively. Computation time was around 6 seconds for a standard dataset.

Conclusion: The application of a CNN for automated contouring of intraprostatic GTV in 68Ga-PSMA- and 18F-PSMA-PET images resulted in a high concordance with expert contours and in high sensitivities and specificities in comparison with histology reference. This robust, accurate and fast technique may be implemented for treatment concepts in primary PCa. The trained model and the study's source code are available in an open source repository.

Keywords: PSMA-PET, Convolutional neuronal networks, Segmentation, Prostate cancer, Histopathology

INTRODUCTION

In patients with newly diagnosed PCa, accurate contouring of the intraprostatic gross tumour volume (GTV) is mandatory for successful fusion-biopsy guidance (1). Additionally, focal therapy approaches such as focal dose escalation in radiotherapy (2) rely on an accurate definition of the intraprostatic GTV.

Prostate-specific membrane antigen positron emission tomography (PSMA-PET) has recently been established for initial staging in primary PCa patients (3). It is also increasingly used in order to improve intraprostatic lesion detection (4-6), focal therapy guidance (7) and non-invasive PCa characterization (8). Most of the studies evaluated 68Ga-PSMA-11 as radiopharmaceutical. However, 18F-PSMA-1007 is increasingly used and Kuten et al. reported that 18F-PSMA-1007 may detect additional low-grade lesions (9). In a recent work manual and semi-automatic contouring approaches for 68Ga-PSMA-PET images were validated (10). Although good results (sensitivity and specificity >80%) were obtained for most of the contouring approaches, some methodologies showed a rather poor performance (sensitivity and specificity <70%). This is in line with a dice-index (DSC) varying between 0.56-0.8 for the manual contours, which indicates that PSMA-PET based GTV-definition underlies a substantial interobserver variability. Actually, no validated contouring technique for 18F-PSMA-PET was proposed.

The implementation of an automatic segmentation algorithm may enhance intraprostatic GTV-delineation in PSMA-PET images by extending the two main limits of conventional contouring approaches: interobserver heterogeneity and expenditure of time. Recently, convolutional neural networks (CNNs) based algorithms achieved remarkable results handling this task. In a work by Zhao et al. the pelvic PCa tumour burden in 68Ga-PSMA-PET images was detected by a CNN with 99% precision (11). Although several works already reported the excellent performance of CNNs in prostatic gland delineation on CT images (12) the usage of CNNs for intraprostatic GTV contouring in PSMA-PET was not examined yet. The aim of this

work is to examine the capabilities of CNNs for intraprostatic GTV contouring in 68Ga- and 18F-PSMA-PET.

MATERIALS AND METHODS

Patients

Data from 209 patients with primary PCa from three different centres (Table 1) were included. Inclusion criteria were histologically proven adenocarcinoma of the prostate and no treatment prior to PSMA-PET. The institutional review boards approved this retrospective study and the requirement to obtain written consent was waived.

PET/CT Imaging

A detailed description of the radiolabelling protocol of 68Ga-PSMA-11 and 18F-PSMA-1007 from centre 1-3 can be found in previous studies (6,13-15). One hour (68Ga-PSMA-11) and two hours (18F-PSMA-1007) after intravenous tracer injection, all patients underwent whole body PET scan. In centre 1, protocols were acquired on three cross-calibrated Philips scanners: GEMINI TF TOF64, GEMINI TF16 Big Bore and Vereos. All scanners resulted in a PET image with a voxel size of 2x2x2mm. Centre 2 used an uMI 780 PET/CT scanner (United Imaging Healthcare) with a voxel size of 2.3x2.3x2.7mm. Resampling was performed to obtain a PET image voxel size of 2x2x2mm (tri-linear interpolation in plastimatch v1.8.0) before training of the CNN. Expert contours of intraprostatic GTV and prostate contours were resampled with nearest neighbor interpolation (plastimatch v1.8.0). Centre 3 acquired all studies acquired using a Biograph mCT 128 Flow scanner (Siemens). PET images had a voxel size of 4.1x4.1x5mm. Testing was performed with the original data and with three different resampling methods to obtain a PET image voxel size of 2x2x2mm.

Histopathology and PET/CT Co-registration

For 29 patients from centre 1 (cohort 1: n=18 and cohort 2: n=11) the 3D distribution of the intraprostatic GTV was obtained by histology information from prostatectomy specimen. The resected specimen underwent an ex-vivo CT scan in customized localizer and whole-mount step sections were cut every 4mm using a cutting device. Staining with hematoxylin and eosin was performed and PCa tissue in histology was delineated. Histology slices were registered on ex-vivo CT images and PCa contours were transferred onto the CT images. The contours were interpolated to create a model of the 3D distribution of PCa in histology (GTV-Histo). Ex-vivo CT (including GTV-Histo) was manually registered to in-vivo CT. First, the prostate was delineated in both. Subsequently, ex-vivo CT was oriented in the space of the in-vivo CT and the axes between the apex and the prostatic base in both CTs guided further registration. Rotation was applied for final alignment. The delineations of the prostatic glands in both CTs and intraprostatic markers (e.g. calcifications) served as reference points for anisotropic scaling of the ex-vivo prostate. All co-registration steps were performed by CZ using MITK (Version 2014.10.00).

Contouring of PSMA PET/CT

All GTVs on PET were delineated by two readers (GTV-Exp) from centre 1 in consensus as proposed previously (10): GTVs were delineated manually in every single slice using inverted grey color scale for display, windowed with SUVmin-max: 0-5. In the first step two readers with medium level of experience (ASB or NB, experience approx. 1.5 years) delineated the GTVs under the consideration of the respective PET/CT report. Subsequently, CZ (experience 6 years) reviewed all GTVs independently. In case of discrepancies each individual case was discussed and corrected in order to reach a consensus contour. Additionally, for the patients with histopathology reference in cohorts 1 and 2, threshold-based contouring with 30% of intraprostatic SUVmax was applied (GTV-30%) as proposed previously (16). GTV-30% volumes were created semi-automatically in Eclipse (v15.6). Manual contouring of the prostatic gland on

CT scans was considered as the gold-standard and was done by CZ using the ESTRO-ACROP guidelines (17). All manual delineations were created in 3D Slicer (Version 4.10.0).

Preprocessing

The data (nearly raw raster data format, nrrd) was cropped to a size of 64x64x64 voxels and normalized with $x_i' = \frac{x_i - \overline{x}}{\sigma}$, where x_i is the PET data of patient $i, \overline{x}$ the arithmetic mean and σ the standard deviation within all cropped datasets. The volume of 64x64x64 voxels proved to be large enough to encompass the prostate and its surrounding tissue for all patients and small enough to enable a computation of the whole volume on the GPU.

Due renal excretion it is not always possible to accurately differentiate between prostatic tissue and bladder signal in 68Ga-PSMA-PET. Consequently, only delineations inside the prostatic gland contour were used for computations.

To investigate the impact of a voxel size different from the training voxel size and the usage of different interpolation algorithms we used the PETs from centre 3 in four different ways. First, the original data was fed to our network. In a second setting, the PETs were resampled to a resolution of 2x2x2 mm with three different methods (SimpleITK v1.2.4): B-spline interpolation order 3, tri-linear interpolation and Gaussian interpolation. Prostate contours and ground truth GTVs were resampled with nearest neighbor interpolation.

Convolutional neural network

The current work was based on a 3D variant of the U-Net architecture (18). It consists of 3 down sampling steps with max-pooling, 3 up sampling steps with transposed convolution layers (kernel size:2x2x2, stride:2, padding:1) and skip connections by concatenation. The 18 convolution blocks consist of 3x3x3 convolutions with stride and padding of 1, followed by Batch Normalization and Rectified Linear Unit activation, except for the last convolution where 1x1x1 convolution without padding, Batch Normalization and Sigmoid activation function were used. An

argmax function over the final feature map formed the predicted GTV. The network weights were optimized using adaptive moment estimation (19).

Training. The 152 patients of the training cohort were further split into training (n=142) and evaluation cohort (n=10). The evaluation cohort was used for optimizing the CNN's hyper-parameters during the training process. As input the CNN received a concatenation of the patients' PET and prostate contour. Hyper-parameter optimization was done using a grid search considering: optimizer, learning rate, number of epochs, data augmentation with x-axes flipping and scaling in x-/y-/z-direction. The best performing setting was achieved with adaptive moment estimation $\beta_1 = 0.9 \wedge \beta_2 = 0.999$, a learning rate of 0.0001 and training for 1019 epochs (an epoch means iterating over all training samples once) with a dice loss: $diceloss(X,Y) = 1 - \frac{2\sum_{l=1}^{|L|} w_l \sum_n y_{ln} x_{ln}}{\sum_{l=1}^{|L|} w_l \sum_n y_{ln} + x_{ln}}$ for $|L|$ number of labels, N image elements $x_{0,\dots,N} \in X, y_{0,\dots,N} \in Y$ and without weighting the label classes $w_l = 1$. Grid search was performed without or with data augmentation by (i) flipping the x-axis by 50% chance, (ii) by scaling the data in all directions or (iii) by doing both. For each iteration the original data was pseudo-randomly and independently scaled in x/y/z-direction for ±10 voxels and then cropped as described before. Data augmentation achieved worse or equal results than the settings without augmentation. Consequently no data augmentation was used for further analyses. In Figure 1 visualizations of the training and evaluation curves are presented.

Evaluation. We assessed the agreement between GTV-Exp and GTV-CNN at voxel level using the DSC. Additionally, we considered the Hausdorff distance (HD) and the average symmetric surface distance (ASSD). The sensitivity and specificity for all GTVs based on the histology standard of reference data was calculated as performed previously (20). The prostate in each CT slice (PSMA-PET/CT scans) was divided into four equal segments and the analysis

was performed visually using the GTVs obtained. A median of 52 segments (range: 20-64) were analysed per patient.

Implementation. The network was implemented with pytorch 1.3.1 and torchvision 0.4.2. Gradients for backpropagation were calculated with the pytorch autograd library which keeps track of all operations and builds a computational tree. Please see the provided code: https://gitlab.com/dejankostyszyn/prostate-gtv-segmentation.

Statistical Analysis

The statistical analysis was performed with MedPy's package 'Metric Measures' v0.4.0 and GraphPad Prism v8.1.0 (GraphPad Software). Pairwise comparisons were performed with the Wilcoxon matched-pairs signed rank or Friedman test. Non-pairwise testing was performed with Mann-Whitney test or Chi-square test. The tests were chosen due to non-normal distribution (Shapiro-Wilk test) of the data. Finally, we searched for clinical factors that might impact the CNN performance by influencing the SUV distribution (PSA and Gleason score) or by neighborhood to the bladder (localization): a binary logistic regression analysis was performed to assess the impact of clinical parameter on DSC between GTV-Exp and GTV-CNN. The confidence alpha was set to 5%.

RESULTS

Test Results 68Ga-PSMA-11 PET

On the internal datasets (cohort 1) the network yielded median DSC, HD and ASSD of 0.84 (range: 0.32-0.94), 4mm (range: 1.41-10) and 0.61mm (range: 0.24-1.46), respectively (supplementary Table 1). Considering histology reference (Figure 2) median sensitivity and specificity of 0.98 (range: 0.38-1) and 0.76 (range: 0.13-1) were observed. The achieved

sensitivity and specificity was comparable to GTV-Exp and GTV-30% (Figure 3). The median volumes of the GTVs were: 10.7ml (range: 0.7-101) for GTV-CNN, 11.8ml (range: 0.8-75) for GTV-Exp, 8ml (range: 2.2-41) for GTV-30% and 10.4ml (range: 1.6-103) GTV-histo. No significant differences between absolute volume of GTV-CNN and the three other volumes were observed ($p>0.05$). The GTV-CNN encompassed in median 26.6% of the prostatic gland.

Patients in the external test cohort (cohort 3) had statistically significant differences between Gleason scores but not between PSA values and cT stage (Table 1). Comparison between GTV-CNN and GTV-Exp was performed firstly on non-resampled and secondly on resampled PET images (supplementary Table 1). Friedman test revealed statistically significant ($p<0.01$) differences in DSC, HD and ASSD among the pre-processing procedures and no pre-processing. Post-hoc analyses revealed no statistically significant differences between the three interpolation approaches ($p>0.05$). As datasets with tri-linear interpolation from centre 2 were used in the training cohort, we conducted an additional experiment by training the CNN solely on patients from centre 1 (without interpolation), to exclude a bias. Testing was performed on patients from centre 3 using all three interpolation methods and achieved comparable results to the results shown in supplementary Table 1.

In regression analysis with pooled cohorts 1 and 3 no clinical parameter had an impact on DSC between GTV-Exp and GTV-CNN (supplementary Table 2).

Test Results 18F-PSMA-1007 PET

Median DSC, HD and ASSD for cohort 2 were 0.81 (range: 0.28-0.93), 5mm (range: 1.41-8.49) and 0.51mm (range: 0.26-1.57), respectively (supplementary Table 1). Sensitivity and specificity were 1 (range: 0.86-1) and 0.57 (range: 0.12-1). GTV-CNN had a significant higher sensitivity than GTV-30% ($p=0.01$) but not than GTV-Exp ($p=0.48$). No statistically significant differences in specificity ($p>0.05$) were observed between the three GTVs. Median volume was 3.5ml (range: 0.3-24.4) for GTV-histo, 8.5ml (range: 1.9-38) for GTV-CNN, 3ml (range: 0.6-21.5)

for GTV-30% and 7.2ml (range: 1.2-36) for GTV-Exp. GTV-CNN was statistically significant larger ($p>0.05$) than all other volumes ($p<0.05$) and encompassed in median 32% of the prostate.

Computation Time

For internal test cohorts the segmentation of the GTV of one patient took in median 6 and 6.28 seconds, respectively, including loading and storing the data (supplementary Table 1). This process took 23.3-27.8 seconds for cohort 3. A single forward pass through the CNN took less than a second (approx. 3 µs) for all cohorts.

DISCUSSION

Implementation of automatic GTV-segmentation approaches based on CNN algorithms have already been introduced for several other tumours (21). Although several studies achieved promising results by using CNNs for auto segmentation of the prostatic gland there is limited evidence on the segmentation of the intraprostatic GTV (22). To the best of our knowledge this is the first study analyzing CNNs for intraprostatic GTV delineation based on PET images. We chose PSMA-PET images since several studies reported that PSMA-PET outperforms mpMRI in tumour detection (4-6). Consequently, the use of PSMA-PET for initial staging (3) and intraprostatic GTV detection and contouring (23) has been established and several studies suggested its implementation for treatment individualization in primary PCa (24-27). However, all previous studies used manually or semi-automatically created contours for intraprostatic GTV contouring which may be impeded by low sensitivity/specificity and interobserver heterogeneity (10). Furthermore, manual contouring of intraprostatic GTV is time consuming. Obviously, a fast, robust and accurate workflow for intraprostatic GTV contouring is a prerequisite for a broader deployment of PSMA-PET-based procedures. In this work we proved that CNNs have the ability to delineate the intraprostatic GTV on PSMA-PET with accuracy comparable to human experts

within seconds. Thus, it is likely that PSMA-PET/CT in combination with CNN-based intra- and extraprostatic (11) tumour detection and segmentation may provide a “one-stop shop” tool for tailoring individualized treatment approaches.

The CNN performance for 68Ga-PSMA-PET was tested on two independent datasets and high DSC values (>0.8) between GTV-Exp and GTV-CNN were observed. Bravaccini et al. reported that the PSMA expression correlates with the Gleason score (28). As the two test cohorts had statistically significant differences in Gleason score in biopsy probes our results show that the CNN performance is independent of the Gleason score and suggest that the CNN identified patterns that are independent from absolute accumulation values. Nevertheless, pattern recognition in PSMA-PET images through CNNs may enable non-invasive tumor characterization (e.g. the Gleason score) in the future. In rare cases a high HD was observed despite a high DSC. This was the case when the main parts of CNN and expert GTVs overlapped, but small regions with a high distance to the main tumour, were diagnosed as malignant by the CNN, but not by the human experts. For example, in two patients of cohort 1 the CNN detected small (<5mm in histology) lesions which were missed by GTV-PET. This explains the slightly higher sensitivity of the CNN in cohort 1 although the absolute GTV volumes were comparable. Since HD is sensitive to outliers, we used ASSD as additional metric and achieved comparable results. In comparison with histology reference GTV-CNN achieved high sensitivity and good specificity in 68Ga-PSMA-PET images, which was comparable to manually delineated expert contours and threshold-based contours. Additionally, the absolute volume of GTV-CNN was very similar to the histology reference volume suggesting an adequate coverage of the intraprostatic tumour. Since GTV-CNN encompassed in median 26.6% of the prostatic gland, it is very likely that focal therapy approaches guided by CNN are feasible in most of the patients. 68Ga-PSMA-PET images of the external test cohort were tested with and without previous resampling. Statistically significant differences were observed with better results for the resampled datasets. Hence, when using datasets from different institutions a resampling of the

images to the same voxel size of the training data set should be performed. Although tri-linear interpolation showed a slightly better performance, there was no statistically significant difference between the results of the three methods. Therefore no specific interpolation method can be recommended. Noteworthy, in some patients with 68Ga-PSMA-PET discrepant results between CNN and the other contours were observed. Moreover, PET signal from the adjacent bladder may mislead the CNN in contouring of PCa lesions in the prostatic base. Since no clinical parameters like Gleason Score or tumor localization had an impact on the concordance between GTV-Exp and GTV-CNN, a visual control of the CNN segmentations has to be performed in every patient.

The CNN provided also a high concordance with expert contours (DSC>0.8) in contouring of 18F-PSMA-PET images. Taking into account the differences in physical properties and in bio distribution between both tracers this result is surprising and should be interpreted with caution since no validated approach for contouring was applied. However, considering histology as standard of reference an excellent sensitivity was observed which was comparable to manual contours and better than threshold-based contours. The specificity of GTV-CNN was low which is mainly explained by a significant overestimation of the tumour volume. Thus, the CNN may also be used for GTV contouring in 18F-PSMA-PET images, especially in situations when a complete coverage of the intraprostatic GTV is demanded and a high coverage of non-tumour bearing prostatic tissue is negligible. Surely, further studies implementing 18F-PSMA-PET images and validated expert contours for training and testing are necessary to confirm this observation.

A limitation of our study is the relatively low number of patients used for testing which is explainable by the elaborate co-registration protocol. We assume that the observed results are robust since we used different, independent datasets for evaluation and received comparable results. Another point that supports the robustness is that we did not notice any overfitting in the

training process (Figure 1) which chance was further reduced by hyper parameter optimization in combination with splitting the training data internally. Considering the high value of the two independent datasets used for testing the CNN, no additional approaches for validation were performed (e.g. k-fold cross-validation). Another issue is the uncertainty in correlation of PSMA-PET images and histopathology slices. Thus, it could not be excluded that low coverage of PCa in histology by the PET-derived GTVs is a consequence of mismatch in coregistration or incomplete histopathological coverage. However, as the calculation of sensitivities/specificities was not performed on a voxel-level but on a less stringent slice by slice level, we consider the potential resulting bias negligible. In our study the prostatic gland on CT scans was delineated manually. Following projects should integrate already existing approaches (12) for automatic prostate segmentation with our approach for automatic GTV-delineation, enabling a fully-automated workflow.

CONCLUSION

Our study presents a CNN for automated contouring of intraprostatic GTV in 68Ga- and 18F-PSMA-PET. Likewise, CNN-based GTV delineation is a promising and fast alternative to visual and threshold-based PET image interpretation. The link to the code and trained model of the CNN may be used for focal therapy or targeted-biopsy concepts in primary PCa by providing a GTV-proposal before visual image interpretation. We strongly emphasize that our tool is not clinically validated and not certified, thus a visual control of the CNN contours by experienced experts is obligatory. Furthermore, the CNN may be used as an alternative approach for GTV segmentation in ongoing radiomics and/or deep learning research in the field where certification is not mandatory.

ACKNOWLEDGMENTS

N/A

KEY POINTS

QUESTION: How is the performance of a trained convolutional neuronal network (CNN) for automatic segmentation of intraprostatic tumor volume in PSMA-PET images of primary prostate cancer patients?

PERTINENT FINDINGS: In this multicentre study including 209 patients, the CNN provided comparable results with human expert and threshold-based delineations under consideration of coregistered whole-mount sections as the standard of reference.

IMPLICATIONS FOR PATIENT CARE: The CNN provided a fast and robust auto-segmentation of the intraprostatic tumor and may enhance individualized therapeutic approaches for primary prostate cancer patients like focal therapy or targeted biopsy.

REFERENCES

1. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med.* 2018;378:1767-1777.

2. Perera M, Krishnananthan N, Lindner U, Lawrentschuk N. An update on focal therapy for prostate cancer. *Nature Reviews Urology.* 2016;13:641-653.

3. Hofman MS, Lawrentschuk N, Francis RJ, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *Lancet.* 2020;395:1208-1216.

4. Bettermann AS, Zamboglou C, Kiefer S, et al. [Ga-68-]PSMA-11 PET/CT and multiparametric MRI for gross tumor volume delineation in a slice by slice analysis with whole mount histopathology as a reference standard - Implications for focal radiotherapy planning in primary prostate cancer. *Radiotherapy and Oncology.* 2019;141:214-219.

5. Eiber M, Weirich G, Holzapfel K, et al. Simultaneous (68)Ga-PSMA HBED-CC PET/MRI Improves the Localization of Primary Prostate Cancer. *Eur Urol.* 2016;70:829-836.

6. Chen M, Zhang Q, Zhang C, et al. Combination of (68)Ga-PSMA PET/CT and Multiparametric MRI Improves the Detection of Clinically Significant Prostate Cancer: A Lesion-by-Lesion Analysis. *J Nucl Med.* 2019;60:944-949.

7. Zamboglou C, Sachpazidis I, Koubar K, et al. Evaluation of intensity modulated radiation therapy dose painting for localized prostate cancer using Ga-68-HBED-CC PSMA-PET/CT: A planning study based on histopathology reference. *Radiotherapy and Oncology.* 2017;123:472-477.

8. Zamboglou C, Carles M, Fechter T, et al. Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate-

and high-risk prostate cancer - a comparison study with histology reference. *Theranostics.* 2019;9:2595-2605.

9. Kuten J, Fahoum I, Savin Z, et al. Head-to-Head Comparison of Ga-68-PSMA-11 with F-18-PSMA-1007 PET/CT in Staging Prostate Cancer Using Histopathology and Immunohistochemical Analysis as a Reference Standard. *Journal of Nuclear Medicine.* 2020;61:527-532.

10. Zamboglou C, Fassbender TF, Steffan L, et al. Validation of different PSMA-PET/CT-based contouring techniques for intraprostatic tumor definition using histopathology as standard of reference. *Radiotherapy and Oncology.* 2019;141:208-213.

11. Zhao Y, Gafita A, Vollnberg B, et al. Deep neural network for automatic characterization of lesions on Ga-68-PSMA-11 PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging.* 2020;47:603-613.

12. Liu C, Gardner S, Wen N, Siddiqui F, Movsas B, Chetty I. Automatic Segmentation of the Prostate Gland on Planning CT Images Using Deep Neural Networks (DNN). *Medical Physics.* 2018;45:E464-E464.

13. Zamboglou C, Wieser G, Hennies S, et al. MRI versus Ga-68-PSMA PET/CT for gross tumour volume delineation in radiation treatment planning of primary prostate cancer. *European Journal of Nuclear Medicine and Molecular Imaging.* 2016;43:889-897.

14. Cardinale J, Martin R, Remde Y, et al. Procedures for the GMP-Compliant Production and Quality Control of [F-18]PSMA-1007: A Next Generation Radiofluorinated Tracer for the Detection of Prostate Cancer. *Pharmaceuticals.* 2017;10.

15. Schmuck S, von Klot CA, Henkenberens C, et al. Initial Experience with Volumetric (68)Ga-PSMA I&T PET/CT for Assessment of Whole-Body Tumor Burden as a Quantitative Imaging Biomarker in Patients with Prostate Cancer. *J Nucl Med.* 2017;58:1962-1968.

16. Zamboglou C, Schiller F, Fechter T, et al. (68)Ga-HBED-CC-PSMA PET/CT Versus Histopathology in Primary Localized Prostate Cancer: A Voxel-Wise Comparison. *Theranostics.* 2016;6:1619-1628.

17. Salembier C, Villeirs G, De Bari B, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. *Radiotherapy and Oncology.* 2018;127:49-61.

18. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii.* 2015;9351:234-241.

19. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv.org 2014.

20. Zamboglou C, Drendel V, Jilg CA, et al. Comparison of Ga-68-HBED-CCPSMA-PET/CT and multiparametric MRI for gross tumour volume detection in patients with primary prostate cancer based on slice by slice comparison with histopathology. *Theranostics.* 2017;7:228-237.

21. Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol.* 2019;20:728-740.

22. Cuocolo R, Cipullo MB, Stanzione A, et al. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp.* 2019;3:35.

23. Goodman CD, Fakir H, Pautler S, Chin J, Bauman GS. Dosimetric Evaluation of PSMA PET-Delineated Dominant Intraprostatic Lesion Simultaneous Infield Boosts. *Advances in Radiation Oncology.* 2020;5:212-220.

24. Calais J, Kishan AU, Cao M, et al. Potential Impact of (68)Ga-PSMA-11 PET/CT on the Planning of Definitive Radiation Therapy for Prostate Cancer. *J Nucl Med.* 2018;59:1714-1721.

25. Zamboglou C, Thomann B, Koubar K, et al. Focal dose escalation for prostate cancer using (68)Ga-HBED-CC PSMA PET/CT and MRI: a planning study based on histology reference. *Radiat Oncol.* 2018;13:81.

26. Liu C, Liu T, Zhang Z, et al. (68)Ga-PSMA PET/CT Combined with PET/Ultrasound-Guided Prostate Biopsy Can Diagnose Clinically Significant Prostate Cancer in Men with Previous Negative Biopsy Results. *J Nucl Med.* 2020;61:1314-1319.

27. Zhang LL, Li WC, Xu Z, et al. (68)Ga-PSMA PET/CT targeted biopsy for the diagnosis of clinically significant prostate cancer compared with transrectal ultrasound guided biopsy: a prospective randomized single-centre study. *Eur J Nucl Med Mol Imaging.* 2020.

28. Bravaccini S, Puccetti M, Bocchini M, et al. PSMA expression: a potential ally for the pathologist in prostate cancer diagnosis. *Sci Rep.* 2018;8:4254.

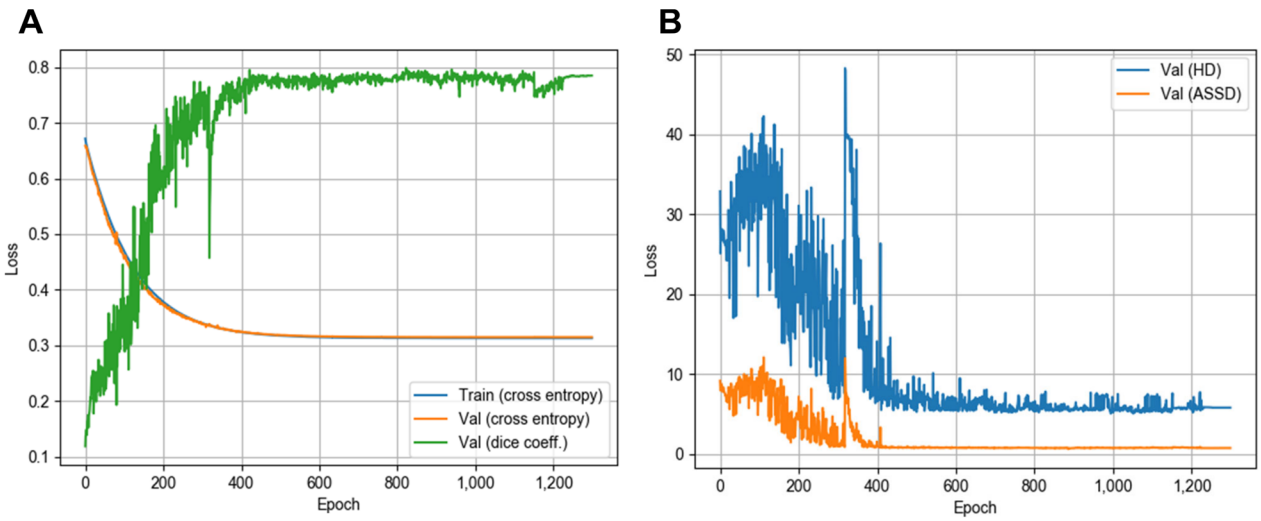


FIGURE 1. Visualization of the training and evaluation curves. Training and evaluation results as dice loss and Dice-Sørensen coefficient (DSC) in A as well as Hausdorff distance (HD) and average symmetric surface distance (ASSD) in B.

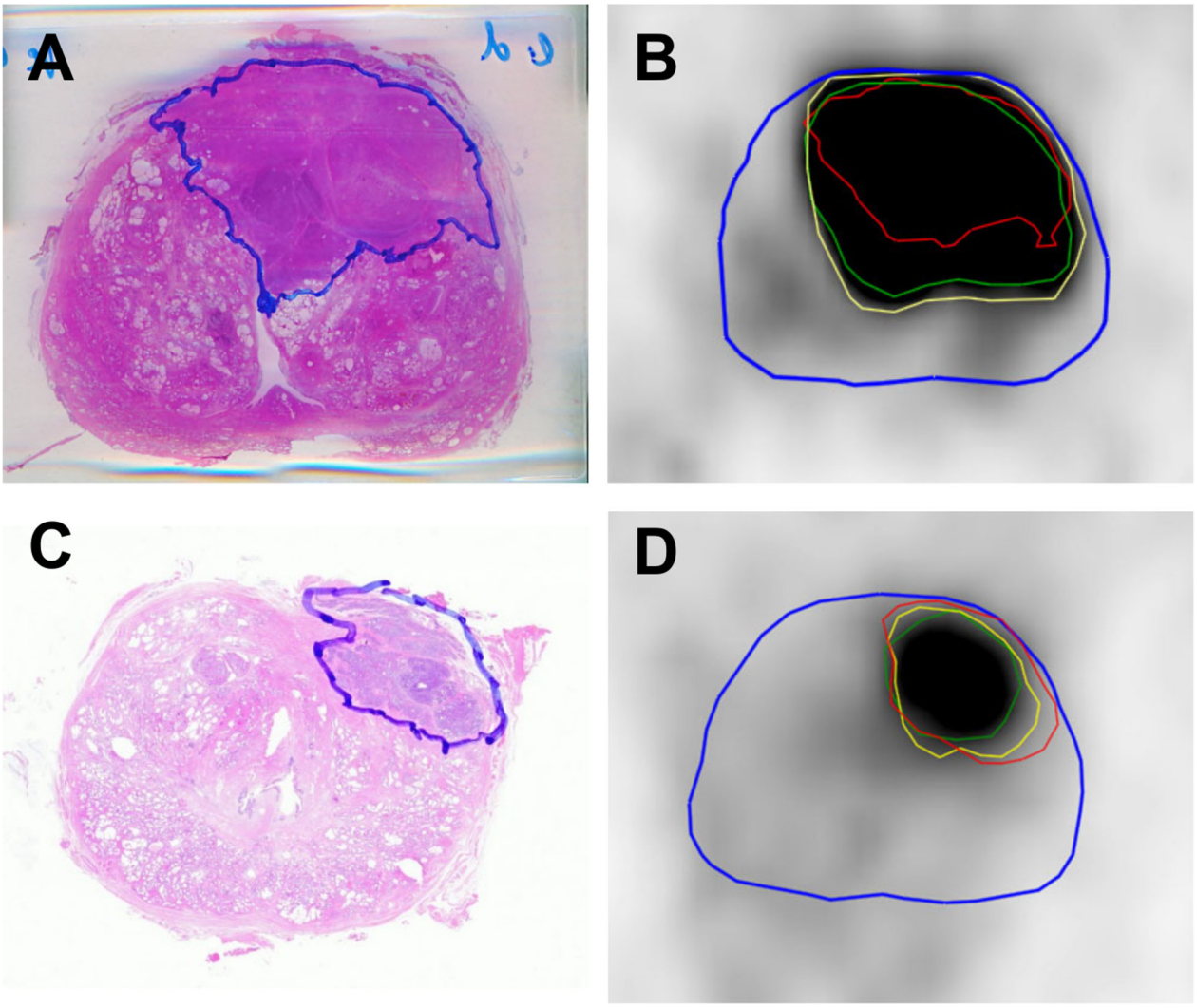


FIGURE 2. Histology reference projected on 68Ga-PSMA-11 (upper row) and 18F-PSMA-1007 (lower row). A and C: Hematoxilin and eosin whole-mount prostate slide with marked PCa lesion. B and D: Axial slide PET-scan (image windowing: SUVmin-max=0-5). Blue contour: prostate. Green contour: histology reference. Green contour: GTV-Exp. Yellow contour: GTV-CNN.

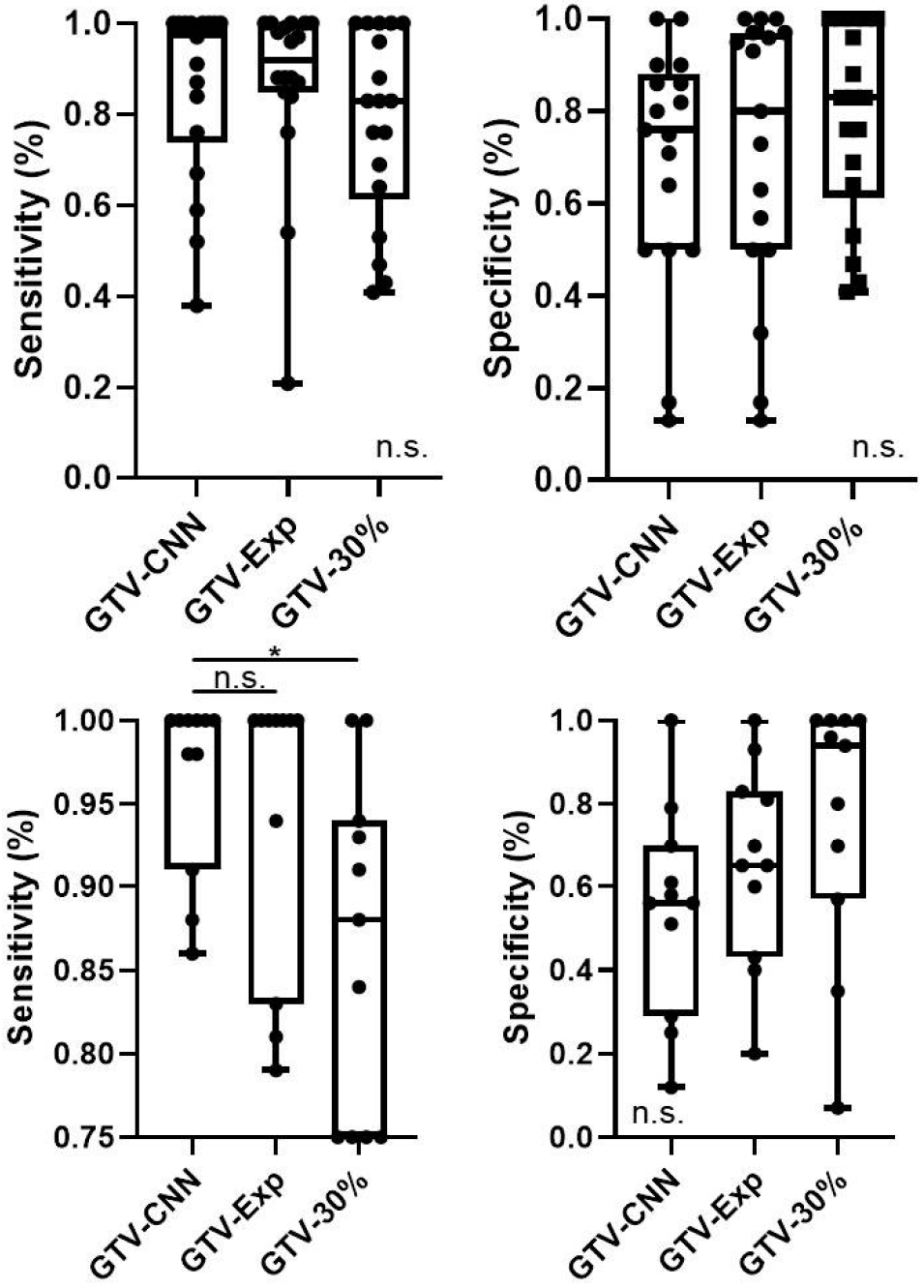


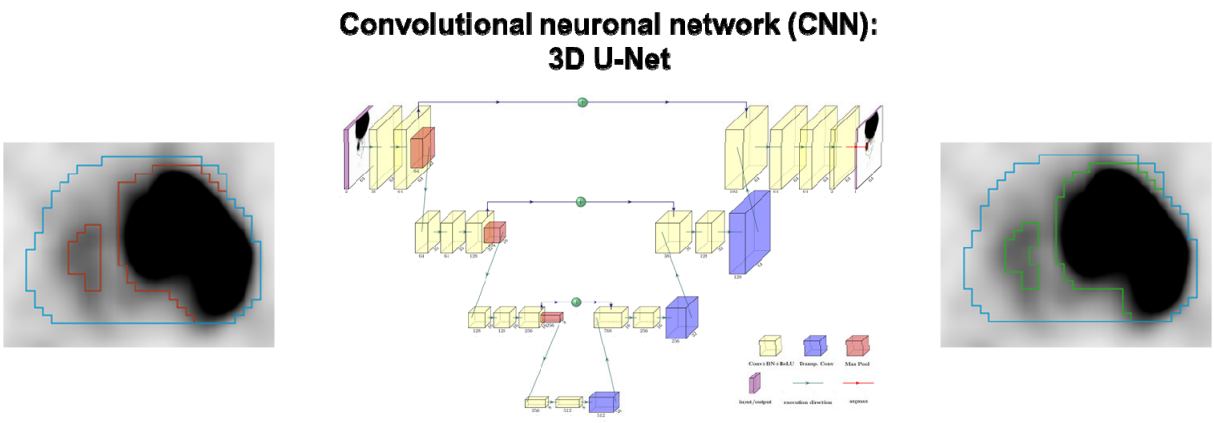
FIGURE 3. Specificity and sensitivity of GTV-CNN, GTV-Exp and GTV-30% based on comparison with histology reference. Upper row: cohort 1 (68Ga-PSMA-11 PET). Lower row: cohort 2 (18F-PSMA-1007 PET). Box plots are presented. Pairwise comparison was performed with Wilcoxon signed-rank test. Abbrevations: n.s.: not significant, *: p=0.05-0.01, CNN: convolutional neural network.

Table 1. Patient characteristics

	Centre 1 Freiburg	Centre 2 Nanjing	Centre 3 Hannover
Median Age in years (range)	70 (48-88)	69 (55-84)	71.5 (59-84)
Median PSA in ng/ml (range)	13.1 (4.4-218)	13.3 (4.04-110)	12.8 (1.91-108.10)
Gleason Score, n and %			
6	5 (3.4)	4 (8.9)	0
7a	44 (30.1)	12 (26.7)	3 (15)
7b	43 (29.9)	12 (26.7)	3 (15)
8	24 (16.7)	10 (22.2)	6 (30)
9	19 (13.2)	7 (15.5)	8 (40)
unknown	9 (6.3)	0	0
cT stage, n and %			
2	89 (61.8)	14 (30.8)	6 (30)
3	55 (38.2)	31 (69.2)	14 (70)
n patients with 68Ga-PSMA-PET/CT, total	125	45	20
n patients training cohort	97	45	
n patients validation cohort	10		
n patients testing cohort	18		20
n patients with histology reference	18		
n patients with 18F-PSMA-PET/CT, total	19		
n patients testing cohort	19		
n patients with histology reference	11		

The CNN was trained on 142 patients with 68Ga-PSMA-11 PET/CT from centres 1 and 2. From centre 1, 10 patients with 68Ga-PSMA-11 PET/CT were pseudo-randomly selected for validation. Finally, internal testing was performed on 18 and 19 patients with 68Ga-PSMA-11 PET/CT (cohort 1) or 18F-PSMA-1007 (cohort 2) PET/CT, respectively. All patients with histological samples were in the testing cohorts. External validation was performed in 20 patients with 68Ga-PSMA-11 PET/CT (cohort 3) from centre 3. Datasets from centre 2 (n=42) were only used for training and from centre 3 (n=20) only for testing. Differences in clinical parameters of the two test cohorts with 68Ga-PSMA-11 PET/CT were analyzed. No differences in PSA and cT stage were observed ($p>0.05$). Patients from centre 3 had statistically significant ($p=0.035$) higher Gleason scores.

GRAPHICAL ABSTRACT



Supplementary Table 1. Testing of the CNN performance

	DSC			HD (mm)			ASSD (mm)			Computation time (sec)		
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
Internal testing 68Ga-PSMA-11	0.84	0.32	0.95	4.03	1.42	10.0	0.61	0.28	1.97	6.28	5.47	7.66
Internal testin 18F-PSMA-1007	0.81	0.28	0.93	5.0	1.41	8.49	0.5	0.26	1.82	6.00	3.53	9.2
External testing 68Ga-PSMA-11												
No resampling	0.78	0.11	0.89	12.57	1.43	32.9	0.62	0.27	4.03	1.93	0.27	2.02
B-spline interpolation	0.82	0.39	0.92	5.83	2.45	22.36	0.55	0.32	2.1	27.79	10.54	30.91
Tri-linear interpolation	0.83	0.32	0.93	4.12	2.01	22.36	0.46	0.28	1.61	23.32	10.55	26.37
Gaussian interpolation	0.81	0.04	0.94	7.35	2.24	20.05	0.55	0.19	3.72	25.13	10.49	28.3

Abbreviations: Dice-Sørensen coefficient (DSC), Hausdorff distance (HD) and average symmetric surface distance (ASSD).

Supplementary table 2. Impact of different clinical parameters on DSC (cut-off: median value of pooled cohorts).

		95% CI		
Parameter	HR	lower	upper	p value
initial PSA (20 ng/ml vs. >20 ng/ml)	1.286	0.343	4.816	0.709
Gleason score (6 + 7a vs. ≥7b)	3.733	0.78	17.88	0.099
cT stage (2 vs. 3)	2.489	0.616	10.056	0.201
Localization (lower half vs. upper half vs. both halves)	1.429	0.622	3.284	0.401

Testing cohorts 1 and 3 were pooled. The best performing resampling step (tri-linear interpolation) in terms of DSC in cohort 3 was used for analyses. The median DSC was used as cut-off point.