



Universiteit
Leiden
The Netherlands

Deep learning for segmentation of the cervical cancer gross tumor volume on magnetic resonance imaging for brachytherapy

Outeiral, R.R.; González, P.J.; Schaake, E.E.; Heide, U.A. van der; Simoes, R.

Citation

Outeiral, R. R., González, P. J., Schaake, E. E., Heide, U. A. van der, & Simoes, R. (2023). Deep learning for segmentation of the cervical cancer gross tumor volume on magnetic resonance imaging for brachytherapy. *Radiation Oncology*, 18(1).
doi:10.1186/s13014-023-02283-8

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3663511>

Note: To cite this publication please use the final published version (if applicable).

RESEARCH

Open Access



Deep learning for segmentation of the cervical cancer gross tumor volume on magnetic resonance imaging for brachytherapy

Roque Rodríguez Outeiral¹, Patrick J. González¹, Eva E. Schaake¹, Uulke A. van der Heide¹ and Rita Simões^{1*}

Abstract

Background Segmentation of the Gross Tumor Volume (GTV) is a crucial step in the brachytherapy (BT) treatment planning workflow. Currently, radiation oncologists segment the GTV manually, which is time-consuming. The time pressure is particularly critical for BT because during the segmentation process the patient waits immobilized in bed with the applicator in place. Automatic segmentation algorithms can potentially reduce both the clinical workload and the patient burden. Although deep learning based automatic segmentation algorithms have been extensively developed for organs at risk, automatic segmentation of the targets is less common. The aim of this study was to automatically segment the cervical cancer GTV on BT MRI images using a state-of-the-art automatic segmentation framework and assess its performance.

Methods A cohort of 195 cervical cancer patients treated between August 2012 and December 2021 was retrospectively collected. A total of 524 separate BT fractions were included and the axial T2-weighted (T2w) MRI sequence was used for this project. The 3D nnU-Net was used as the automatic segmentation framework. The automatic segmentations were compared with the manual segmentations used for clinical practice with Sørensen–Dice coefficient (Dice), 95th Hausdorff distance (95th HD) and mean surface distance (MSD). The dosimetric impact was defined as the difference in D98 ($\Delta D98$) and D90 ($\Delta D90$) between the manual segmentations and the automatic segmentations, evaluated using the clinical dose distribution. The performance of the network was also compared separately depending on FIGO stage and on GTV volume.

Results The network achieved a median Dice of 0.73 (interquartile range (IQR) = 0.50–0.80), median 95th HD of 6.8 mm (IQR = 4.2–12.5 mm) and median MSD of 1.4 mm (IQR = 0.90–2.8 mm). The median $\Delta D90$ and $\Delta D98$ were 0.18 Gy (IQR = -1.38–1.19 Gy) and 0.20 Gy (IQR = -1.10–0.95 Gy) respectively. No significant differences in geometric or dosimetric performance were observed between tumors with different FIGO stages, however significantly improved Dice and dosimetric performance was found for larger tumors.

Conclusions The nnU-Net framework achieved state-of-the-art performance in the segmentation of the cervical cancer GTV on BT MRI images. Reasonable median performance was achieved geometrically and dosimetrically but with high variability among patients.

*Correspondence:

Rita Simões
r.simoes@nki.nl

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Automatic segmentation, Cervical cancer, GTV, Nnu-net, Brachytherapy

Background

For locally advanced cervical cancer the standard of care consists of external beam radiotherapy (EBRT), followed by 3 to 4 fractions of brachytherapy (BT) and concomitant chemotherapy [1]. A key step in both EBRT and BT treatment planning is the segmentation of organs at risk and target volumes. This is mostly performed manually, which is time consuming and suffers from the inherent bias of the observer. To circumvent these issues, automatic segmentation is being widely investigated in the field of radiotherapy [2–4]. For the case of BT, the need for automatic segmentation is even more critical due to the time constraints of the workflow. At each fraction of BT treatment, the applicator is inserted surgically in the patient, after which the MRI images are acquired. The patient then needs to wait, immobilized in bed, while the needed structures (namely organs at risk and target volumes) are manually delineated and a treatment plan is made. The Gynecological (GYN) GEC-ESTRO working group defines the target volumes of interest for BT treatment planning for this cervical cancer as the Gross Tumor Volume (GTV), the high risk Clinical Target Volume (HR-CTV) and the intermediate risk Clinical Target Volume (IR-CTV) [5] and they are currently segmented by radiation oncologists. Automatic image segmentation methods are expected to reduce the clinical workload as well as patient burden.

Automatic segmentation of the targets volumes is still uncommon and it is mostly limited to positron emission tomography (PET) and/or computed tomography (CT) and rarely to magnetic resonance imaging (MRI) [9]. For the particular case of cervical cancer on BT images, automatic segmentation of the organs at risk has been widely investigated [6–10] but literature on the automatic segmentation of the targets, and especially the GTV, is more scarce [7, 9, 10]. Zhang et al. [10] and Wong et al. [9] developed automatic segmentation tools that segmented the HR-CTV (on CT and MRI images, respectively) but to the best of our knowledge, only Yoganathan et al. [7] have studied the automatic segmentation of the gross tumor volume (GTV) on BT MRI images. While they demonstrated that automatic segmentation of the GTV is possible in principle, the cohort was rather small with only 39 patients, resulting in a relatively weak performance with Sørensen–Dice coefficients (Dice) between 0.57 and 0.62. Furthermore, the segmentation architectures used in their project were based on the ResNet50 architecture [11], which is no longer considered state of the art.

A current state-of-the-art framework for the automatic segmentation of medical structures is the nnU-Net

(‘no-new-U-Net’) [12]. The nnU-Net is a deep learning-based framework which automatically configures the parameters needed for training. It has been shown to outperform other approaches on 23 public datasets used on segmentation competitions.

The aim of this study was to assess the quality of the automatic segmentations of the cervical cancer GTV on BT MRI images. We used two methods to determine to what extent the automatic segmentations corresponded to the clinical segmentations performed by an expert radiation oncologist. First, the geometrical correspondence of the automatic and expert delineation was determined using Dice Similarity Coefficient (Dice), 95th Hausdorff Distance (95th HD) and mean surface distance (MSD). Then, to find if the observed geometrical differences between the delineations would have dosimetric consequences, we determined dose-volume parameters D90 and D98 for the automatic and expert delineations using the clinical dose distribution.

Methods

Data

A cohort of 195 histologically proven cervical cancer patients treated in our institution between August 2012 and December 2021 was retrospectively collected. The average age was 53 (standard deviation of 15 years) and tumor stage ranged from IB to IV according to the International Federation of Gynecology and Obstetrics (FIGO) staging [13]. The treatment consisted of external beam radiotherapy (156 patients with 23×2 Gy and 39 patients with 25×1.8 Gy) followed by BT (3×7 Gy) and combined with chemotherapy (cisplatin 40 mg/m², weekly). The institutional review board approved the study (IRBd20276). Informed consent was waived considering the retrospective design.

A total of 524 separate BT fractions were included in this work. For each BT fraction, MRI images of the patient with applicator in place were acquired using a 1.5T (104 scans) or 3T (442 scans) Philips Medical Systems MRI scanner. Axial T2-weighted (T2w) turbo spin-echo images were used (TR = [3500–13,300 ms], TE = [100–120 ms]) with a pixel spacing of 0.39 mm x 0.39 mm (442 scans) or 0.63 mm x 0.63 mm (104 scans) and a slice thickness of 3 mm. The GTV, as segmented for treatment planning by a radiation oncologist on each available MRI, was available as ground truth.

The data set was split into three subsets at the patient level: training set (117 patients, 314 images), validation set (39 patients, 104 images) and test set (39 patients, 106 images). The three subsets were stratified according to

FIGO stage [13], because it is a relevant clinical parameter used to describe gynecological tumors.

Network architecture and training procedure

The nnU-Net framework was used in this work. This framework automatically configures the parameters needed for preprocessing, network architecture and training for each specific task. The loss function was a combination of the Dice loss [14] and cross entropy loss. We used the stochastic gradient descent (SGD) optimizer with learning rate scheduler and early stopping based on the validation loss as criterion to choose the best model. Dropout, data augmentation and weight decay were used as regularization techniques. Further details on the training procedure can be found in the Additional file 1.

Experiment overview

Geometric comparison

The automatic segmentations were compared to the manual segmentations of the GTV that were performed by a radiation oncologist for treatment planning for the patients on the separate test set. The automatic segmentations were compared to the manual segmentations using common segmentation metrics: Dice, 95th HD and MSD, which were implemented using the Python package by DeepMind (<https://github.com/deepmind/>

surface-distance). The segmentation results were additionally compared among patients with different FIGO stage and GTV volume. For the volume analysis, the patients of the test set were allocated to four volume ranges containing the same number of images in each bin.

Attention maps were computed for four different examples to highlight which parts of the input image were relevant for the network to decide on a segmentation. The attention maps were then qualitatively compared to the binary segmentations to investigate if the over-/under-segmentations of the network were on specific areas, therefore highlighting anatomically challenging regions. The attention maps were defined as the activations of the last layer of the nnU-net (i.e. before binarizing).

Dosimetric comparison

To assess if the differences between the automatic segmentations and manual segmentations would result in differences in dose-volume parameters, we calculated the D98 and the D90 for both segmentations on the clinical dose distribution used for the treatment. These dose parameters were chosen in accordance with the Embrace II guidelines [1]. The values for the manual segmentations represent the actual treatment parameters for the patients. The dosimetric impact of using automatically segmented structures was defined as the difference between these parameters compared to the clinical values ($\Delta D90$ and $\Delta D98$). The dosimetric impact was also reported as a relative measure by dividing the absolute difference on the dose parameters by the dose parameter on the manual segmentation ($\Delta D90_{rel}$ and $\Delta D98_{rel}$). The dosimetric results were also compared for patients with different FIGO stage and GTV volume.

Table 1 Patients characteristics in the training, validation and test sets

	Training	Validation	Test
Total	117	39	39
Age (years)			
Mean	53	52	56
Standard deviation	14	15	17
FIGO stage			
FIGO I	12 (10.2%)	6 (15.4%)	6 (15.4%)
FIGO II	70 (59.8%)	22 (56.4%)	22 (56.4%)
FIGO III	21 (18.0%)	8 (20.5%)	8 (20.5%)
FIGO IV	9 (7.7%)	3 (7.7%)	3 (7.7%)
Unknown	5 (4.3%)	0	0
Volume at first BT fraction			
Less than 2.8 cc	22 (18.8%)	8 (12.8%)	5 (20.5%)
between 2.8 and 4.3 cc	7 (6.0%)	7 (17.9%)	7 (17.9%)
between 4.3 and 12.1 cc	53 (42.3%)	12 (30.8%)	13 (33.3%)
More than 12.1 cc	35 (29.9%)	12 (30.8%)	14 (35.9%)
Histopathological type			
Squamous cell carcinoma	97 (82.9%)	31 (79.5%)	33 (84.6%)
Adenocarcinoma	16 (13.7%)	6 (15.4%)	4 (10.3%)
Adeno-squamous cell carcinoma	1 (0.%)	1 (2.6%)	1 (2.6%)
Non specified/unknown	3 (2.5%)	1 (2.5%)	1 (2.5%)
External beam radiotherapy scheme			
23 × 2 Gy	95 (81.2%)	29 (74.4%)	32 (82.0%)
25 × 1.8 Gy	22 (18.8%)	10 (25.6%)	7 (18.0%)

Statistics

The chi-square test for independence was used to confirm that the training, validation and test sets were balanced in terms of FIGO stage. The Kruskal-Wallis H test was used to assess differences among patients of different FIGO stage and GTV volume. If significant differences were found, Dunn's test with Bonferroni correction was used for the post-hoc analysis. A p-value of 0.05 was considered statistically significant. The SciPy Python package (version 1.5.4) and Python 3.9 were used for the statistical analysis.

Results

Patients characteristics of our cohort are described in Table 1. No significant differences were found in the distributions of FIGO stage or volume among the training, validation and test sets. The results are shown for 105 out of the 106 cases of the test set. The remaining case corresponded to a patient that had her uterus removed which

resulted in a deviating anatomy unseen by the trained network.

The network achieved a median Dice of 0.73 (interquartile range (IQR)=0.50–0.80), median 95th HD of 6.8 mm (IQR=4.2–12.5 mm) and median MSD of 1.4 mm (IQR=0.9–2.8 mm). When stratifying for FIGO stage (Fig. 1 - top) no significant differences were found among the different subgroups. When comparing for GTV volume (Fig. 1 - bottom) significant differences were found for the case of Dice (p -value<0.001) but not for the distance-based metrics.

Four examples of automatic segmentations are shown in Fig. 2 (a,c,e,g) with the corresponding attention maps (Fig. 2-b,d,f,h). Even though the network under-/over-segmented the GTV for the last three cases (Fig. 2-c,e,g), the error was in the area surrounding the applicator which is an area that is irradiated anyway. In the case 2c, the applicator was segmented by the network but not by the clinician while in the cases 2e and 2f, the clinicians segmented the applicator but the network did not. Furthermore, the attention map highlighted the undersegmented area of the last case (Fig. 2h), meaning that the network looked at that area when deciding the segmentation.

The median D90 and D98 received by the manually segmented GTV were 12.5 Gy (IQR=11.1–15.5 Gy) and 10.6 Gy (IQR=9.4–13.1 Gy), respectively, in line with Embrace guidelines and the GYN GEC-ESTRO

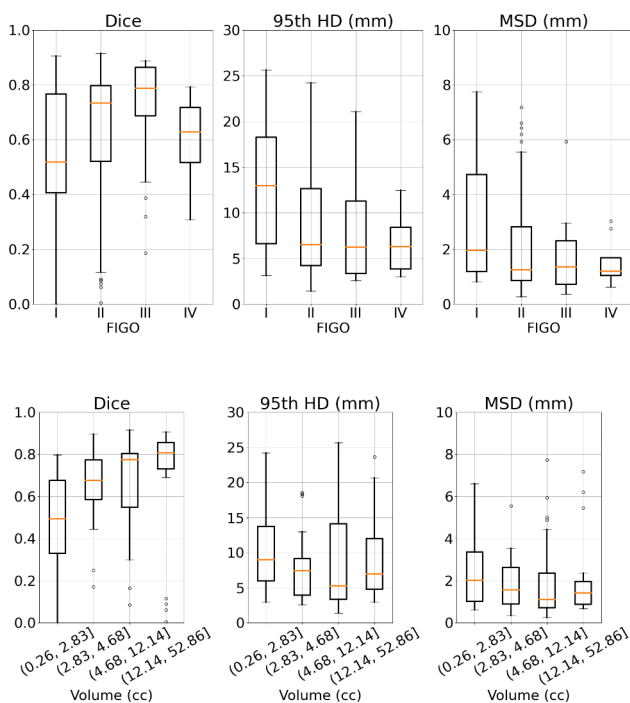


Fig. 1 Geometric comparison by FIGO stage (top) and by volume (bottom). Segmentation performance in terms of Dice, 95th HD and MSD and stratified by FIGO stage (I-IV) and volume

recommendations [15]. The resulting $\Delta D90$ and $\Delta D98$ were 0.18 Gy (IQR = -1.38–1.19 Gy) and 0.20 Gy (IQR = -1.10–0.95 Gy), respectively. The median $\Delta D90_{rel}$ and $\Delta D98_{rel}$ relative differences were 9.6% (QR=4.2–19.28%) and 8.8% (QR=0.15–92.5%), respectively. When stratifying for FIGO stage (Fig. 3 - top), no significant differences were observed among the different subgroups per FIGO stage. When comparing the results for GTV volume (Fig. 3 - bottom), a significantly reduced $\Delta D90$ and $\Delta D98$ (p -value<0.01) was found between the smaller tumors (0.3–2.8 cc) and the largest tumors (12.1–52.9 cc).

Discussion

In this study we investigated the performance of a state-of-the-art automatic framework to segment the cervical cancer GTV on brachytherapy MR images. We used a cohort of patients that for their treatment were segmented manually by a radiation oncologist and compared these manual segmentations to the automatic segmentations. The comparison was performed geometrically and the impact of differences between automatic and manual delineations on dose-volume parameters of the clinical dose distribution was evaluated.

We achieved improved geometric performance when compared to previously published literature and automatic segmentations yielded a $\Delta D90$ and $\Delta D98$ of less than 0.25 Gy. No significant differences in geometric and dosimetric performance were observed when comparing for FIGO stage. When comparing per volume, decreased performance was observed for smaller tumors both for the Dice coefficient and dosimetrically.

To the best of our knowledge, only Yoganathan et al. [7] studied the automatic segmentation of cervical cancer GTV on brachytherapy images. In their work, they implemented and compared four different CNNs for the segmentation of the targets and organs at risk. The geometric performance of our model was considerably higher than what the authors obtained with the models trained with the axial T2w sequence (Dice: 0.56, 95th HD 9.7 mm). Their models were based on the ResNet and Inception architectures, while we use the nnU-Net, currently the state-of-the-art architecture for medical image segmentation. Additionally, we used a larger cohort.

We observed that the median relative $\Delta D90_{rel}$ and $\Delta D98_{rel}$ were lower than 10%. Hellebust et al. [16] showed that the relative $\Delta D90$ between different observers was 9.4% for the GTV, meaning that the average difference dosimetric difference between observers is comparable to using the automatic segmentation tool. However, in some of the cases the dosimetric difference was large. These large differences in dosimetric performance can be partially explained by the marked steepness of the brachytherapy dose distributions, which results in that small

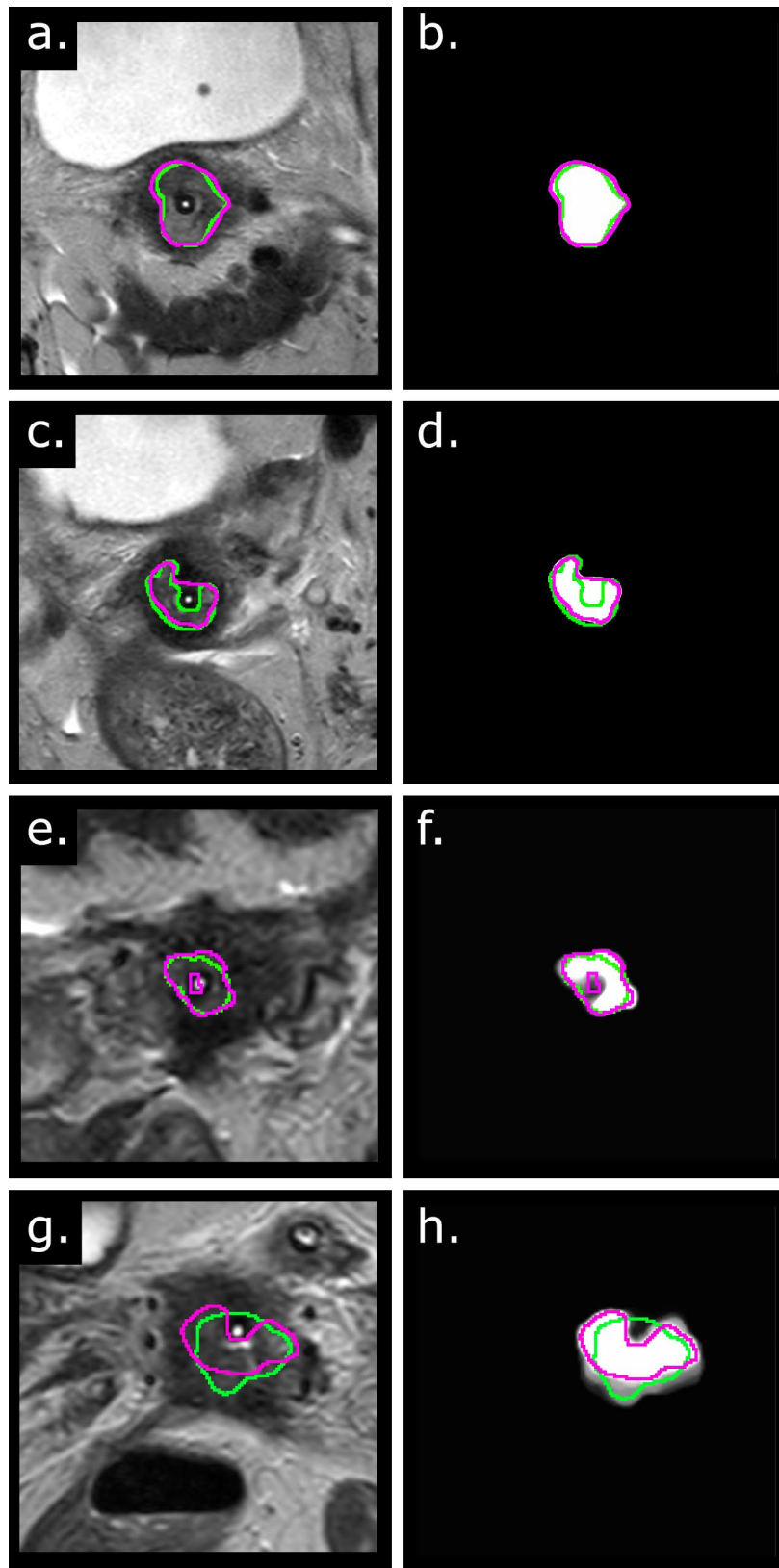


Fig. 2 Qualitative results and attention maps. (Left) Examples of the automatic contours (pink) and the manual clinical contour (green) on four different patients. (Right) The corresponding attention maps for the same patients. The examples are sorted by decreasing Dice

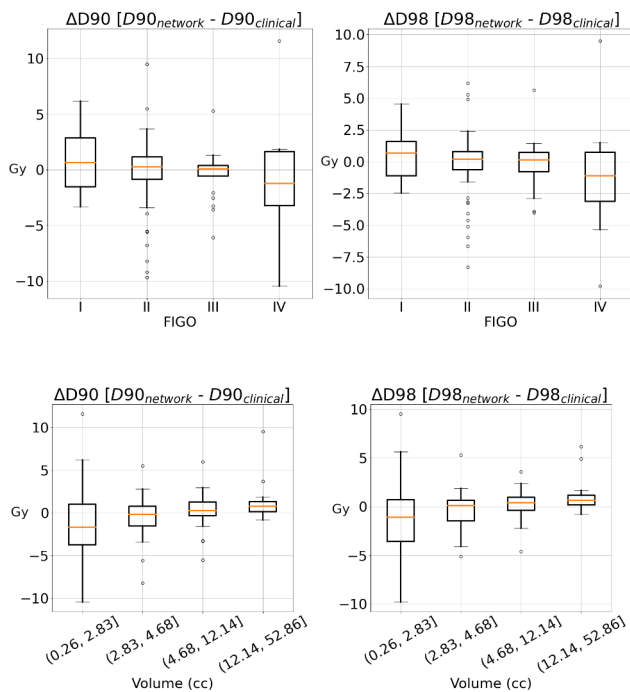


Fig. 3 Dosimetric comparison by FIGO stage (top) and by volume (bottom). Dosimetric impact in terms of $\Delta D90$ and $\Delta D98$ stratified by FIGO stage (I-IV) and volume

geometric errors can lead to large differences in dose parameters.

When comparing the results per FIGO stage, no significant differences were found between the different FIGO stages for neither the geometric nor the dosimetric comparisons. A priori, we would have expected the performance to be different between tumors of different FIGO stages because FIGO stage is an important clinical parameter to describe gynecological tumors. One possible reason is that the FIGO stage is defined at the time of diagnosis and consequently does not take into account the regression of the tumor during the external beam radiotherapy treatment, potentially reducing the differences between FIGO stages. On the other hand, when stratifying per GTV volume, significant differences were found for the Dice and for the dosimetric comparisons. For the Dice, the explanation can be rather trivial, because the Dice is defined as the overlap between the two structures and it therefore favors the bigger structures. However, for the dosimetric impact, larger tumors had lower $\Delta D90$ and $\Delta D98$, and less variability, which suggests that smaller tumors may require more accurate automatic segmentation methods than larger tumors.

This work has the following limitations. Firstly, even though our cohort includes a large amount of patients, patients from only one center were included and fewer patients were included for FIGO I and IV. A multi-center validation study is therefore desirable. Secondly, the GTV segmentations used for training and evaluation

were manually segmented for clinical practice with the treatment plan in mind, meaning that although the segmentations were clinically acceptable, they may contain geometric errors. These geometric errors could potentially lead the network to reproduce these errors and therefore partially bias our geometric analysis. However, we presented the results in terms of dosimetric impact as well and showed that the dosimetric impact of the automatic segmentations is comparable to that derived from the interobserver variability. Finally, the scope of this work was limited to the GTV automatic segmentation while the HR-CTV and the IR-CTV are also needed for treatment planning. The definition of those structures is intrinsically related to the information of the image before external beam radiotherapy (5 weeks before BT) and not only to the information present in the BT image. Therefore in this work we focused solely on the structure that can be found in the BT image.

Conclusions

In this study we evaluated a state-of-the-art framework for the automatic segmentation of the cervical cancer GTV. The quality of the automatic segmentations improved with respect to previously published works. The automatic segmentations yielded similar dose-volume parameters as the manual segmentations used clinically and differences were comparable to the interobserver variability reported in literature.

List of abbreviations

BT	Brachytherapy
CNN	Convolutional Neural Network
CT	Computed Tomography
Dice	Sørensen–Dice coefficient
FIGO	International Federation of Gynecology and Obstetrics
GTV	Gross Tumor Volume
GYN	Gynaecology
95th HD	95th Hausdorff Distance
HR-CTV	High Risk Clinical Target Volume
IR-CTV	Intermediate Risk Clinical Target Volume
MSD	Mean surface distance
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
T2w	T2-weighted

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13014-023-02283-8>.

Supplementary Material 1: Additional File 1. Description: Extra details on the training procedure. Dictionary information as provided by the nnU-Net in the debug.json.

Acknowledgements

Not applicable.

Authors' contributions

RRO collected and processed the data, trained the networks, performed statistical analysis and drafted the manuscript. PJG collected and processed part of the data. EES was part of the clinical team involved in segmentation of

the GTVs and in the treatment of the patients and provided clinically relevant feedback. RS and UAH supervised the study and drafted the manuscript. RRO, EES, UAH and RS contributed significantly to the discussion and interpretation of the results. All co-authors read and revised the manuscript. The final version of the manuscript was approved by all co-authors.

Funding

Not applicable.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to protection of individual patient privacy and the use of an in-house software but are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The institutional review board approved the study (IRBd20276). Informed consent was waived considering the retrospective design.

Consent for publication

Not applicable.

Author details

¹Department of Radiation Oncology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 Amsterdam, CX, The Netherlands

Received: 5 December 2022 / Accepted: 16 May 2023

Published online: 29 May 2023

References

1. Pötter R, Tanderup K, Kirisits C, de Leeuw A, Kirchheiner K, Nout R, et al. The EMBRACE II study: the outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies. *Clin Transl Radiat Oncol*. 2018;9:48–60. <https://doi.org/10.1016/j.ctro.2018.01.001>.
2. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. *Comput Biol Med*. 2018;98:126–46. <https://doi.org/10.1016/j.combiomed.2018.05.018>.
3. Boldrini L, Bibault JE, Masciocchi C, Shen Y, Bittner MI. Deep learning: a review for the Radiation oncologist. *Front Oncol*. 2019;9:977. <https://doi.org/10.3389/fonc.2019.00977>.
4. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46:e1–36. <https://doi.org/10.1002/mp.13264>.
5. Haie-Meder C, Pötter R, van Limbergen E, Briot E, de Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. *Radiother Oncol*. 2005;74:235–45. <https://doi.org/10.1016/j.radonc.2004.12.015>.
6. Lin H, Xiao H, Dong L, Teo KBK, Zou W, Cai J, et al. Deep learning for automatic target volume segmentation in radiation therapy: a review. *Quant Imaging Med Surg*. 2021;11:4847–58. <https://doi.org/10.21037/qims-21-168>.
7. Yoganathan SA, Paul SN, Paloor S, Torfeh T, Chandramouli SH, Hammoud R, et al. Automatic segmentation of magnetic resonance images for high-dose-rate cervical cancer brachytherapy using deep learning. *Med Phys*. 2022;49:1571–84. <https://doi.org/10.1002/mp.15506>.
8. Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I, Zaidi H. Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. *Radiother Oncol*. 2021;159:231–40. <https://doi.org/10.1016/j.radonc.2021.03.030>.
9. Wong J, Kolbeck C, Giambattista J, Giambattista JA, Huang V, Jaswal JK. Deep learning-based Auto-Segmentation for Pelvic Organs at Risk and clinical target volumes in Intracavitary High Dose Rate Brachytherapy. *Int J Radiation Oncology*Biography*Physics*. 2020;108:e284. <https://doi.org/10.1016/j.ijrobp.2020.07.681>.
10. Zhang D, Yang Z, Jiang S, Zhou Z, Meng M, Wang W. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks. *J Appl Clin Med Phys*. 2020;21:158–69. <https://doi.org/10.1002/acm2.13024>.
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Recognition P (CVPR) IEEE. 2016;770–8. <https://doi.org/10.1109/CVPR.2016.90>.
12. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–11. <https://doi.org/10.1038/s41592-020-01008-z>.
13. Tomizawa K, Kaminuma T, Murata K, Noda SE, Irie D, Kumazawa T, et al. FIGO 2018 staging for cervical cancer: influence on stage distribution and outcomes in the 3d-image-guided brachytherapy era. *Cancers (Basel)*. 2020;12:1–10. <https://doi.org/10.3390/cancers12071770>.
14. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV), vol. 6, IEEE; 2016, p. 565–71. <https://doi.org/10.1109/3DV.2016.79>.
15. Pötter R, Haie-Meder C, Van Limbergen E, Barillot I, De Brabandere M, Dimopoulos J, et al. Recommendations from gynaecological (GYN) GEC-ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy – 3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. *Radiother Oncol*. 2006;78:67–77. <https://doi.org/10.1016/j.radonc.2005.11.014>.
16. Hellebust TP, Tanderup K, Lervåg C, Fidarova E, Berger D, Malinen E, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. *Radiother Oncol*. 2013;107:13–9. <https://doi.org/10.1016/j.radonc.2012.12.017>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.