Received: 12 August 2022 | Revised: 17 November 2022 | Accepted: 7 December 2022

DOI: 10.1002/mp.16197

RESEARCH ARTICLE

MEDICAL PHYSICS

HaN-Seg: The head and neck organ-at-risk CT and MR segmentation dataset

Gašper Podobnik[1] | **Primož Strojan**[2] | **Primož Peterlin**[2] | **Bulat Ibragimov**[1,3] | **Tomaž Vrtovec**[1]

[1]Faculty Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

[2]Institute of Oncology Ljubljana, Ljubljana, Slovenia

[3]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Correspondence
Tomaž Vrtovec, Faculty Electrical Engineering, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia.
Email: tomaz.vrtovec@fe.uni-lj.si

Funding information
Javna Agencija za Raziskovalno Dejavnost RS, Grant/Award Numbers: J2-1732, P2-0232, P3-0307; Novo Nordisk Fonden, Grant/Award Number: NFF20OC0062056

Abstract

Purpose: For the cancer in the head and neck (HaN), radiotherapy (RT) represents an important treatment modality. Segmentation of organs-at-risk (OARs) is the starting point of RT planning, however, existing approaches are focused on either computed tomography (CT) or magnetic resonance (MR) images, while multimodal segmentation has not been thoroughly explored yet. We present a dataset of CT and MR images of the same patients with curated reference HaN OAR segmentations for an objective evaluation of segmentation methods.

Acquisition and validation methods: The cohort consists of HaN images of 56 patients that underwent both CT and T1-weighted MR imaging for image-guided RT. For each patient, reference segmentations of up to 30 OARs were obtained by experts performing manual pixel-wise image annotation. By maintaining the distribution of patient age and gender, and annotation type, the patients were randomly split into training Set 1 (42 cases or 75%) and test Set 2 (14 cases or 25%). Baseline auto-segmentation results are also provided by training the publicly available deep nnU-Net architecture on Set 1, and evaluating its performance on Set 2.

Data format and usage notes: The data are publicly available through an open-access repository under the name *HaN-Seg: The Head and Neck Organ-at-Risk CT & MR Segmentation Dataset*. Images and reference segmentations are stored in the NRRD file format, where the OAR filenames correspond to the nomenclature recommended by the American Association of Physicists in Medicine, and OAR and demographics information is stored in separate comma-separated value files.

Potential applications: The *HaN-Seg: The Head and Neck Organ-at-Risk CT & MR Segmentation Challenge* is launched in parallel with the dataset release to promote the development of automated techniques for OAR segmentation in the HaN. Other potential applications include out-of-challenge algorithm development and benchmarking, as well as external validation of the developed algorithms.

KEYWORDS
auto-segmentation, computed tomography, head and neck cancer, image dataset, magnetic resonance, radiation therapy

1 | INTRODUCTION

Cancer and its management are among the key challenges of modern society in the contexts of health and quality of life. Cancer of the head and neck (HaN) region, comprising malignancies of the lips, oral cavity, pharynx, larynx, nasal cavity, paransal sinuses, salivary glands, thyroid, and skin of this region, is potentially associated with swallowing dysfunction, speech and breathing problems, disfigurement, psychosocial distress, and death.[1] With a yearly incidence of above one million cases and prevalence of above four million cases worldwide, it is one of the most frequent cancers, accounting for around 5% of all cancer sites.[2,3]

Radiotherapy (RT) is one of the key treatment modalities for the HaN cancer that aims to deliver a high radiation dose to the targeted cancerous cells while sparing the nearby healthy tissues and organs, that is, organs-at-risk (OARs). In the past years, advances in RT have, in parallel with improvements in less-invasive organ-sparing surgery as well as more intensive multimodal treatments, contributed to the preservation of function and reduced mortality of the HaN cancer.[1] Among these advances, a considerable progress has been made in artificial intelligence (AI) and especially deep learning (DL) as one of its subdomains,[4,5] which has a large spectrum of applications also in radiation oncology. For example, segmentation as the process of partitioning a medical image into multiple anatomical structures is one of the key steps in RT planning,[6] because it provides precise three-dimensional (3D) spatial descriptions of the target volumes as well as OARs that are required for an optimal radiation dose distribution calculation. While manual image segmentation has been recognized as a labor-intensive and time-consuming task that is subjected to intra/interobserver variability,[7] several computerized medical image analysis techniques have been developed for automated segmentation, often referred to as auto-segmentation.[8] In comparison to traditional approaches based on conventional atlases, shape models, and feature classification, the advent of DL has caused a shift towards a superior segmentation performance.[8,9] Although segmentation is primarily performed on computed tomography (CT) images because they contain electron density information used for the calculation of the radiation beam energy absorption, one of the main limitations of CT images is the insufficient image contrast for soft tissues. As a result, the integration of complementary imaging modalities, such as magnetic resonance (MR), is strongly recommended for the HaN region to improve the segmentation of several soft tissue OARs.[10]

Although attempts have been made towards the segmentation of OARs from MR images,[11–15] so far there has been no objective evaluation about the impact the combined analysis of CT and MR images has on the segmentation of OARs in the HaN region. Most of the existing approaches are focused on segmentation from either CT or MR image modality, while multimodal segmentation has not been thoroughly explored yet,[9] probably because CT and MR data of the same patients are not always available or, when they are available, they are not systematically included into the RT planning pipeline. In this paper, we describe a collection of CT and MR images of the same patients with the corresponding curated expert manual segmentations of the complete set of OARs in the HaN region, which aims to promote research in the fields of multimodal and multi-organ auto-segmentation,[16] and MR-guided RT planning.[17,18]

2 | ACQUISITION AND VALIDATION METHODS

2.1 | Images

Images of 60 patients aged 34–79 years that were appointed for image-guided RT in the HaN region at the Institute of Oncology Ljubljana, Slovenia (Ethics Committee approval no. ERID-EK/139), and underwent both CT and MR image acquisition were retrospectively retrieved from the picture archiving and communication system (PACS). CT images were acquired with either the Philips Brilliance Big Bore (Philips Healthcare, Best, the Netherlands) or Siemens Somatom Definition AS scanner (Siemens Healthineers, Erlangen, Germany) by using standard clinical protocols (x-ray tube voltage: 100–140 kV; field of view [FoV]: $(500 \times 500) - (800 \times 800)$ mm^2). All MR images were acquired with the GE Medical Systems Optima MR450w scanner (GE Healthcare, Chicago, IL, USA) by using either the spin echo (SE; flip angle: 160°, echo time: 8.7 ms, repetition time: 472–743 ms; axial T1-weighted fast SE cross-sections with fat saturation—Ax T1 FSE FS) or gradient echo (GR; flip angle: 12°, echo time: 3.1 ms, repetition time: 6.1–8.3 ms; axial T1-weighted ultra fast spoiled GR cross-sections with water only, fat only, in phase and out of phase echoes and homogeneous fat suppression—Ax T1 WATER LAVA FS) sequence (FoV: $(240 \times 240) - (420 \times 420)$ mm^2). The time intervals between CT and MR image acquisition were relatively short, and amounted on average to 1.8 ± 1.7 days (median: 1 day, range: 0–11 days), with MR images always being acquired after CT images. Both images were acquired in patient standard head-first supine (HFS) position, and by using Posicast® 5-point thermoplastic immobilization masks (CIVCO Radiotherapy, Orange City, IA, USA) that were also used for patient irradiation.

Due to the presence of metal artifacts, several CT images contained regions with Hounsfield units (HU)

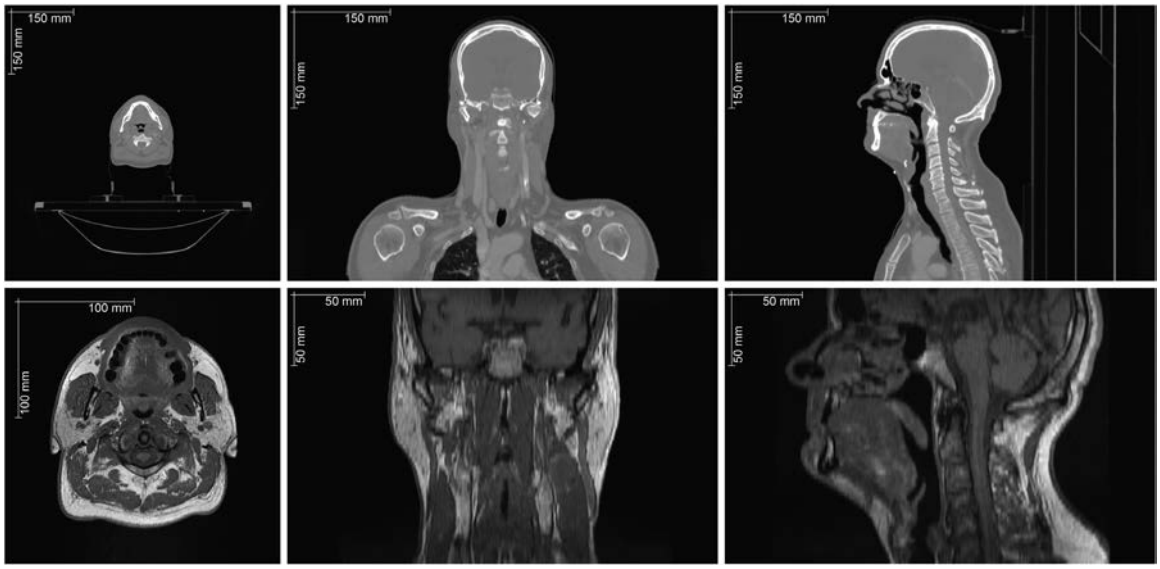
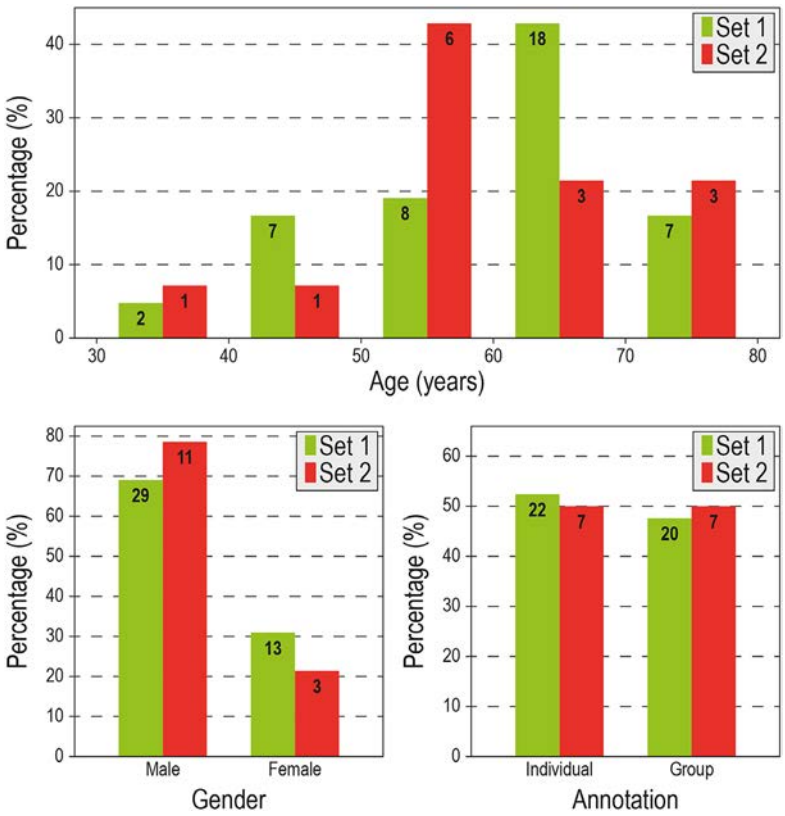


FIGURE 1 Example of a computed tomography (CT, *top*) and T1-weighted magnetic resonance (MR, *bottom*) image of the same patient from the devised dataset (Case 21 from Set 1), displayed in selected axial (*left*), coronal (*middle*), and sagittal (*right*) cross-sections. The distance scale bars are overlaid onto each cross-section.

FIGURE 2 The split of the devised cohort into Sets 1 and 2 was controlled by maintaining the distribution of patient age (*top*), patient gender (*bottom left*), and annotation type (*bottom right*). Bold values denote the number of images assigned to each set.



above 3000, which were consequently set to 3000 to ensure a range of HU that allows easier image visualization. From the initial cohort, one patient was excluded because of poor MR image quality, and three because of the insufficient MR image FoV. The resulting cohort, therefore, consists of 56 cases, where each case is represented by one CT and one T1-weighted MR image of the HaN region of the same patient (Figure 1). Furthermore, the standard arrangement in public competitions of medical image analysis algorithms[19] is to disclose

only part of the data indented for algorithm training, while the nondisclosed data are then used for objective algorithm comparison. By following such an arrangement, the cases were randomly split into two sets by maintaining the distribution of patient age and gender (Figure 2), where 42 cases (75%) from Set 1 are publicly available and 14 cases (25%) from Set 2 are held private. The basic properties of the whole cohort are listed in Table 1.

2.2 | Reference manual segmentation

To enable the development of auto-segmentation methods and evaluation of the obtained segmentation results, we provide reference OAR segmentation masks for CT images in the devised dataset. Initial segmentations were obtained by experts from the Institute of Oncology Ljubljana, Slovenia, who performed manual pixel-wise image annotation of up to 31 OARs (or 23, considering the left and right instances of paired OARs as one instance) in CT images (MR images were used as a support for soft tissues) according to the standard RT planning practice and by following the established guidelines for the delineation of OARs in the HaN region.[10] Approximately half of the cohort (i.e., 29 cases) was annotated by an experienced RT technologist, trained specifically for the delineation of OARs. In this case, each MR image was first registered to the CT image of the same patient, and then OARs were annotated in the reference coordinate system of the CT image. The rest of the cohort (i.e., 27 cases) was annotated by a group of radiation oncologists, experienced in the delineation of OARs. In this case, OARs were annotated in the CT image, however, those OARs that were better visible in the MR modality were annotated in the MR image and then propagated to the reference coordinate system of the CT image (in general, the annotation of the brainstem, lacrimal glands, lips, optic chiasm, optic nerves, parotid glands, pituitary gland, spinal cord, and submandibular glands was supported by MR image information). All annotations as well as image manipulation were performed with the *ARIA Oncology Information System* (OIS) software (Varian Medical Systems, Palo Alto, CA, USA), and were supervised by a senior radiation oncologist. When cases were assigned to Sets 1 and 2, the randomization was, besides patient age and gender, controlled also by maintaining the distribution of individual or group type of annotation in order to reduce the potential bias of different contouring approaches (i.e., RT technologist vs. radiation oncologists). As a result, approximately 50% of cases in Set 1 and exactly 50% of cases in Set 2 originate from each of the two annotation types (Figure 2).

The obtained annotations were then curated by a medical imaging researcher, experienced in annotation of anatomical structures, who manually corrected the

TABLE 1 Properties of patients, and computed tomography (CT) and magnetic resonance (MR) images in the devised HaN-Seg dataset

Sets		Patients		Images			
Set	**No. of cases**	**Gender (M/F)**	**Age (years)**	**Modality**	**Matrix size**	**No. of slices**	**Pixel size (mm × mm)**	**Spacing (mm)**
1 and 2	56 (100%)	40/16 (71%/29%)	59.5 (±11.5)	CT	512 × 512; 1024 × 1024	116–323	0.51 × 0.51–1.56 × 1.56	2.0–3.0
				MR	512 × 512	40–152	0.47 × 0.47–0.82 × 0.82	3.0–5.0
1	42 (75%)	30/12 (71%/29%)	60.0 (±11.4)	CT	512 × 512; 1024 × 1024	116–323	0.52 × 0.52–1.56 × 1.56	2.0–3.0
				MR	512 × 512	40–144	0.47 × 0.47–0.82 × 0.82	3.0–5.0
2	14 (25%)	10/4 (71%/29%)	58.2 (±11.9)	CT	512 × 512; 1024 × 1024	129–217	0.51 × 0.51–1.27 × 1.27	2.0–3.0
				MR	512 × 512	48–152	0.47 × 0.47–0.51 × 0.51	3.0–3.4

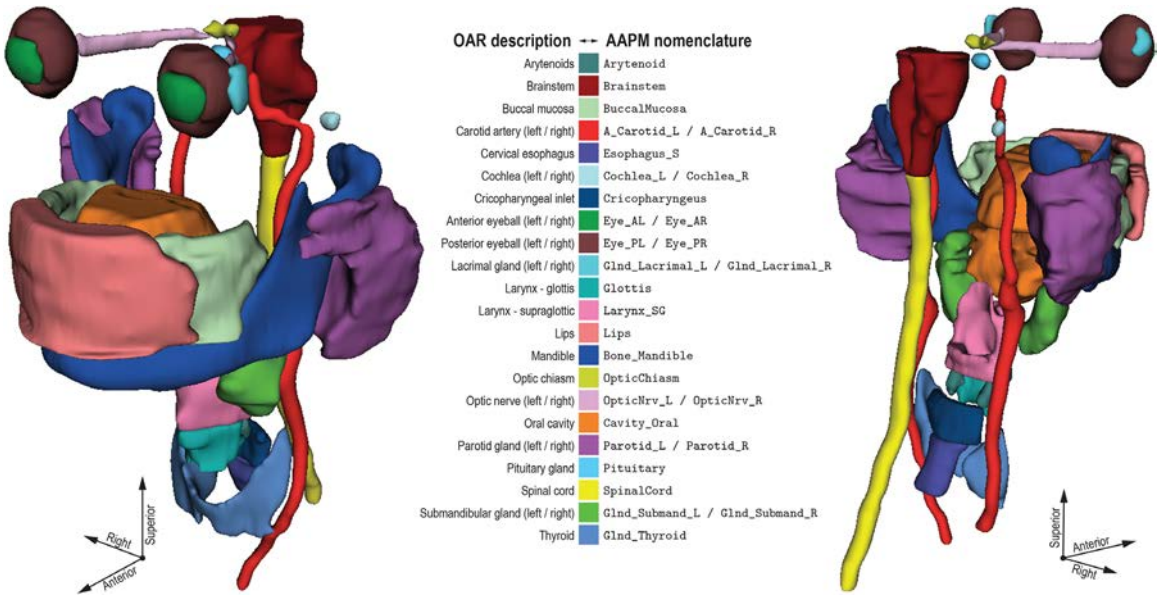


FIGURE 3 Example of reference organ-at-risk (OAR) segmentations for the computed tomography image from Figure 1, displayed as color-coded three-dimensional binary masks.

OAR segmentation masks (i.e., jagged boundaries, isolated pixels, mis-annotations, mis-labeling, connected components, etc.). In this process, the following two major inconsistencies were in addition identified. First, the experts often decided not to extend the segmentation masks of the carotid arteries to the optic chiasm, as recommended by the delineation guidelines,[10] but only to the pituitary gland (i.e., approximately 10 mm caudally from the optic chiasm). In order to make the segmentation masks of the carotid arteries consistent among all cases, which is an important aspect when training and evaluating the performance of auto-segmentation methods, they were cut off below the pituitary gland. Second, the segmentation masks of pharyngeal constrictor muscles were often overlapping by a large extent with other OARs (e.g., oral cavity, larynx), moreover, their shape often alternated considerably between adjacent axial slices, therefore not resembling a connected and unified anatomical structure. As a result, the annotations of pharyngeal constrictor muscles were found to be not reliable enough, and were, therefore, not included in the final dataset.

All corrections were performed with the publicly available *Seg3D* software (version 2.5.0, NIH Center for Integrative Biomedical Computing, Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, USA; https://www.sci.utah.edu/cibc-software/seg3d.html). By not including the annotations of pharyngeal constrictor muscles in the final dataset, reference annotations of up to 30 OARs (or 22, considering the left and right instances of paired OARs as one instance) were obtained for each case in the form of 3D binary segmentation masks (Figure 3) and named according to the nomenclature proposed by the Task Group 263 (TG-263) of the American Association of Physicists in Medicine (AAPM),[20] that is, arytenoids (`Arytenoid`), brainstem (`Brainstem`), carotid artery (`A_Carotid`; left/right: `A_Carotid_L`/`A_Carotid_R`), cervical esophagus (`Esophagus_S`), cochlea (`Cochlea`; left/right: `Cochlea_L`/`Cochlea_R`), cricopharyngeal inlet (`Cricopharyngeus`), lacrimal gland (`Glnd_Lacrimal`; left/right: `Glnd_Lacrimal_L`/`Glnd_Lacrimal_R`), larynx—glottis (`Glottis`), larynx—supraglottic (`Larynx_SG`), lips (`Lips`), mandible (`Bone_Mandible`), optic chiasm (`OpticChiasm`), optic nerve (`OpticNrv`; left/right: `OpticNrv_L`/`OpticNrv_R`), oral cavity (`Cavity_Oral`), parotid gland (`Parotids`; left/right: `Parotid_L`/`Parotid_R`), pituitary gland (`Pituitary`), spinal cord (`SpinalCord`), submandibular gland (`Glnd_Submands`; left/right: `Glnd_Submand_L`/`Glnd_Submand_R`), and thyroid (`Glnd_Thyroid`). For OARs that were not found in the AAPM TG-263 nomenclature,[20] that is, buccal mucosa (`BuccalMucosa`), anterior segment of the eyeball (`Eye_A`; left/right: `Eye_AL`/`Eye_AR`), and posterior segment of the eyeball (`Eye_P`; left/right: `Eye_PL`/`Eye_PR`), we followed a similar naming convention that was combined with the nomenclature in the delineation guidelines.[10] An example of the obtained reference segmentations is shown in Figure 4.

2.3 | Baseline auto-segmentation

To provide baseline auto-segmentation results, we applied a publicly available framework for DL-based

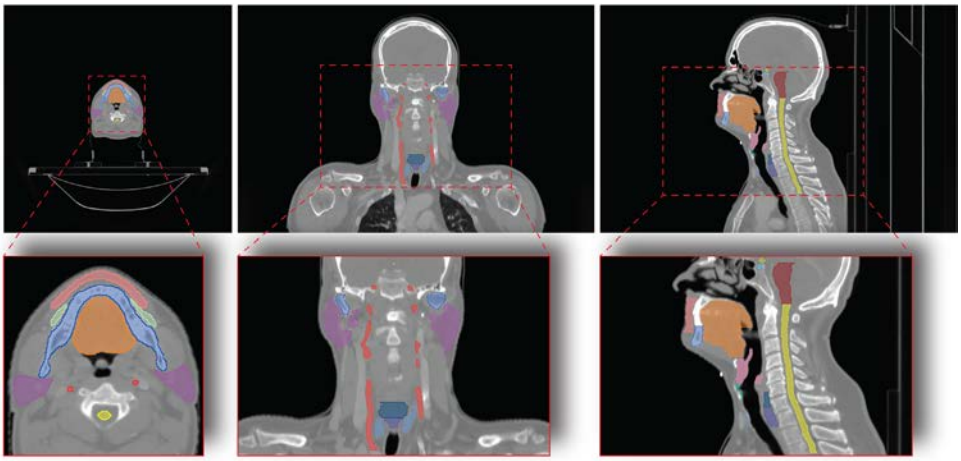


FIGURE 4 Example of reference organ-at-risk (OAR) segmentations for the computed tomography image from Figure 1, displayed in selected axial (*left*), coronal (*middle*), and sagittal (*right*) cross-sections as original images (*top row*), and enlarged regions of interest (*bottom row*). The OAR color-coded overlay corresponds to Figure 3.

segmentation named nnU-Net[21] (https://github.com/MIC-DKFZ/nnUNet). By building on the U-Net architecture,[22] nnU-Net adds self-configurable preprocessing, augmentation, and post-processing techniques, and employs modern strategies for an efficient training, which proved to generate state-of-the-art results on several publicly available datasets used in different biomedical image segmentation challenges.

We trained the nnU-Net framework on images and corresponding reference segmentations from Set 1, for which each pair of CT and MR images was first registered by finding the optimal rigid (i.e., translation and rotation) and nonrigid (i.e., B-splines) geometrical alignment with *SimpleElastix* (version 0.10.0; https://simpleelastix.github.io), an extension of the open-source image registration toolbox *elastix*.[23] The registration quality was evaluated by first placing corresponding landmarks at six anatomical locations that could be reliably identified in both CT and MR images (i.e., at the nasal tip, the posterior edge of the skull at the height of the nasal tip, the posterior edge of the left/right angle of the mandible, and at the approximate junction between the vertebral lamina and left/right transverse process of a C3–C6 vertebral level), and then computing the distance between each pair of corresponding landmarks after registration, that is, the target registration error (TRE). The resulting mean TRE of 1.5 mm (standard deviation: 0.5 mm; maximum: 3.5 mm) is consistent with the AAPM Task Group 132 recommendations,[24] which state that the mean TRE below 2 mm (maximum below 5 mm) is typically desired for the majority of clinical applications, that is, below the tolerance equal to the maximum voxel dimension, which is in our case 2–5 mm.

Before training, all images were also spatially resampled to the same pixel size of $0.5\,\text{mm} \times 0.5\,\text{mm}$ and spacing of 2 mm (Table 1), and HUs were set to the interval between −1000 and 3000 to correct for eventual interpolation errors. A single nnU-Net model was then trained for 1000 epochs on all 42 pairs of CT and MR images (no external images were used), and for all OARs using the `3d_fullres` configuration (the only modification was that the data augmentation for image flipping across the sagittal axis was disabled). The obtained nnU-Net model was then evaluated on images and corresponding reference segmentations from Set 2, and the results are reported in Table 2 in terms of the Dice coefficient (DC), a standard metrics for volumetric mask overlap that measures the harmonic average of the classification precision and recall (i.e., the F_1 score), and Hausdorff distance (HD), a standard metrics for distance measurement that evaluates the mutual proximity of the segmentation mask surfaces, where its 95-percentile version (HD_{95}) is often used to robustly suppress the influence of the outliers.[9] As MR images commonly have, in comparison to CT images, a smaller FoV, CT images were cropped to the FoV of MR images. Consequently, the metrics were computed only for those OARs that retained more than 20% of their original volume after cropping. The whole framework was implemented in the *Python* programming language (version 3.7.10; https://www.python.org).

3 | DATA FORMAT AND USAGE NOTES

The devised Set 1 can be publicly accessed and downloaded through the open-access repository Zenodo (https://zenodo.org) under the collection *HaN-Seg: The Head and Neck Organ-at-Risk CT & MR Segmentation Dataset* (https://doi.org/10.5281/zenodo.7442914), with the condition that any research originating from the

TABLE 2 The baseline auto-segmentation results for the devised dataset, obtained by training nnU-Net[21] on Set 1 and evaluating its performance on organs-at-risk (OARs) in Set 2 (i.e., $N = 28$ for both and 14 for individual instances of left/right OARs, or less when an OAR was not in the field of view). The reported values are mean ± standard deviation of the Dice coefficient (DC) and 95-percentile Hausdorff distance (HD_{95}).

Nomenclature	OAR name	*N*	DC (%)	HD_{95} (mm)
A_Carotid	Carotid arteries (both)	28	84.7 ± 3.4	1.6 ± 1.2
A_Carotid_L	Carotid artery (left)	14	84.4 ± 4.0	1.7 ± 1.2
A_Carotid_R	Carotid artery (right)	14	85.0 ± 2.6	1.5 ± 1.1
Arytenoid	Arytenoids	13	56.3 ± 17.9	3.1 ± 1.5
Bone_Mandible	Mandible	14	94.4 ± 1.6	1.5 ± 0.5
Brainstem	Brainstem	13	86.0 ± 4.2	4.5 ± 1.9
BuccalMucosa	Buccal mucosa	13	69.4 ± 9.7	6.0 ± 2.7
Cavity_Oral	Oral cavity	14	90.2 ± 3.6	4.7 ± 2.1
Cochlea	Cochleae (both)	26	72.8 ± 11.4	1.7 ± 1.0
Cochlea_L	Cochlea (left)	13	74.6 ± 11.1	1.5 ± 0.9
Cochlea_R	Cochlea (right)	13	71.1 ± 11.4	1.8 ± 1.0
Cricopharyngeus	Cricopharyngeal inlet	9	63.2 ± 13.1	5.5 ± 3.0
Esophagus_S	Cervical esophagus	5	64.2 ± 13.0	6.8 ± 2.3
Eye_A	Anterior eyeball segment (both)	15	82.2 ± 4.5	1.8 ± 0.3
Eye_AL	Anterior eyeball segment (left)	7	82.2 ± 3.7	1.7 ± 0.3
Eye_AR	Anterior eyeball segment (right)	8	82.2 ± 5.1	1.9 ± 0.4
Eye_P	Posterior eyeball segment (both)	17	92.6 ± 1.6	1.6 ± 0.5
Eye_PL	Posterior eyeball segment (left)	8	93.0 ± 1.8	1.6 ± 0.5
Eye_PR	Posterior eyeball segment (right)	9	92.3 ± 1.4	1.5 ± 0.4
Glnd_Lacrimal	Lacrimal glands (both)	15	62.6 ± 13.1	3.2 ± 1.3
Glnd_Lacrimal_L	Lacrimal gland (left)	7	63.1 ± 13.0	3.1 ± 1.5
Glnd_Lacrimal_R	Lacrimal gland (right)	8	62.1 ± 13.1	3.3 ± 1.2
Glnd_Submands	Submandibular glands (both)	26	84.4 ± 8.1	3.7 ± 2.6
Glnd_Submand_L	Submandibular gland (left)	13	84.8 ± 9.1	3.2 ± 2.2
Glnd_Submand_R	Submandibular gland (right)	13	84.0 ± 6.9	4.2 ± 2.9
Glnd_Thyroid	Thyroid	6	90.3 ± 2.3	1.5 ± 0.6
Glottis	Larynx—glottis	12	75.2 ± 7.8	2.7 ± 1.3
Larynx_SG	Larynx—supraglottic	13	81.4 ± 4.3	3.5 ± 1.2
Lips	Lips	14	72.2 ± 10.8	6.5 ± 2.8
OpticChiasm	Optic chiasm	9	44.3 ± 8.7	4.0 ± 1.5
OpticNrv	Optic nerves (both)	18	72.2 ± 8.1	2.3 ± 0.7
OpticNrv_L	Optic nerve (left)	9	69.9 ± 8.5	2.6 ± 0.8
OpticNrv_R	Optic nerve (right)	9	74.6 ± 7.0	2.1 ± 0.5
Parotids	Parotid glands (both)	28	86.3 ± 3.6	5.3 ± 3.6
Parotid_L	Parotid gland (left)	14	87.1 ± 2.8	5.9 ± 4.4
Parotid_R	Parotid gland (right)	14	85.6 ± 4.2	4.7 ± 2.5
Pituitary	Pituitary gland	13	70.3 ± 11.9	2.5 ± 1.4
SpinalCord	Spinal cord	13	82.1 ± 4.7	1.7 ± 0.4

usage of the HaN-Seg dataset needs to cite this paper. Images and reference segmentation masks are provided in the compressed nearly raw raster data (NRRD) file format (`.nrrd` extension), a common format for the representation and processing of multidimensional raster data that can be loaded and processed by most medical image analysis software (e.g., ImageJ, Seg3D, ITK, 3D Slicer, etc.) as well as programming languages via dedicated libraries (e.g., Matlab—Image Processing Toolbox, Python—SimpleITK package, etc.). Specifically, for each case (`<nn>` = 01...42), there is one NRRD file containing the 3D CT image (`case_<nn>_IMG_CT.nrrd`)

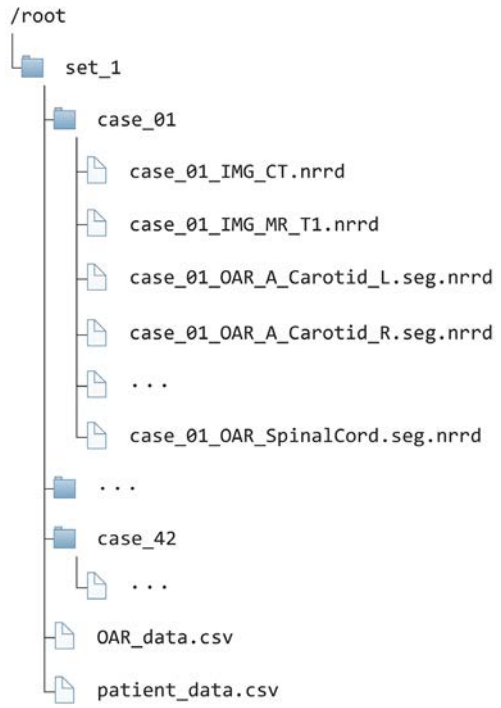


FIGURE 5 The directory structure showing the contents of the publicly available Set 1 of the devised HaN-Seg dataset

and one NRRD file containing the 3D T1-weighted MR image (`case_<nn>_IMG_MR_T1.nrrd`) of the same patient, and up to 30 NRRD files containing reference 3D OAR binary segmentation masks for the CT image (the filenames are composed from OAR labels according to the recommended AAPM nomenclature,[20] i.e., `case_<nn>_OAR_<label>.seg.nrrd`). All information, important for image analysis (e.g., matrix size, pixel size, slice thickness, etc.) can be found in the corresponding NRRD files, while the context demographic information (i.e., gender, age) is provided for all cases in a comma-separated value (CSV; `.csv` extension) file (`patient_data.csv`). An additional CSV file is provided that contains the availability of reference segmentations for each OAR and each case (`OAR_data.csv`), and reports that all segmentations are present (value 1) except for Case 19, where the optic chiasm (`OpticChiasm`) is completely out of FoV (value 0) while the brainstem (`Brainstem`), optic nerves (`OpticNrv_L`, `OpticNrv_R`), and pituitary gland (`Pituitary`) are partially in the FoV (value 0.5). The publicly available data are fully anonymized and free of the protected health information. The directory structure is shown in Figure 5.

4 | POTENTIAL APPLICATIONS

For the devised HaN-Seg dataset, we have identified the following three main potential applications, that is, *A. Computational challenge*, *B. Algorithm development and benchmarking*, and *C. External validation*, although other applications may arise.

4.1 | Computational challenge

In parallel with the release of this dataset, we are also launching the *HaN-Seg: The Head and Neck Organ-at-Risk CT & MR Segmentation Challenge* (https://han-seg2023.grand-challenge.org) to promote the development of new and application of existing state-of-the-art fully automated techniques for OAR segmentation in the HaN region from CT images that exploit the information of multiple imaging modalities, in this case from CT and MR images. The task of the HaN-Seg challenge is to automatically segment up to 30 OARs in the HaN region from CT images in the devised Set 2, consisting of 14 CT and MR images of the same patients, given the availability of Set 1, consisting of 42 CT and MR images of the same patients with reference 3D OAR binary segmentation masks for CT images, that is, the herein described HaN-Seg dataset. From this perspective, Set 2 is held private and therefore not released to the potential participants to prevent algorithm tuning, but instead the algorithms have to be submitted in the form of virtual containers[25] that will be run by organizers on Set 2. The challenge is organized by taking into account the current guidelines for biomedical image analysis competitions, in particular, the recommendations of the Biomedical Image Analysis Challenges (BIAS) initiative for transparent challenge reporting.[19]

4.2 | Algorithm development and benchmarking

Apart from participating in the HaN-Seg challenge described above, researchers can also use the HaN-Seg dataset for algorithm development, and, most importantly, for benchmarking the obtained segmentation results. Specific fields within medical image analysis and RT where CT and MR images of the same patients may be beneficial, are, for example, OAR segmentation in the HaN region,[9,26] multi-organ segmentation,[16,27,28] synthetic MR image generation for segmentation support,[29,30] MR-guided RT planning,[17,18,31] or synthetic CT image generation for MR-only RT.[32–35]

4.3 | External validation

External validation is the process of using independently derived datasets to validate the performance of an AI method, commonly also known as real-world validation.[36] Researchers may use the devised HaN-Seg dataset for external validation[37] as well as for other

topics related to the dataset, such as for evaluating different learning strategies[38] or the impact of the training set size on the performance of an AI method.[39]

5 | DISCUSSION

In radiation oncology, AI models allow automation and optimization of the workflow,[4–6] and DL as a subset of AI has become the dominant methodology for auto-segmentation of medical images.[8] Because it has been recognized as extremely useful for RT planning, OAR auto-segmentation is a very widespread research topic, including approaches focused on the HaN region.[9]

Considering the general incidence and prevalence of the HaN cancer,[2,3] several datasets of HaN images that are used for its diagnosis and treatment have been devised in the past. Some of these datasets are publicly available, for example, via *The Cancer Imaging Archive* (TCIA) that currently hosts 11 collections related to the HaN cancer. However, not all of them are associated with reference OAR segmentations, but were augmented and combined into new publicly available datasets. For example, 48 CT images from one TCIA dataset (*Head-Neck Cetuximab*) were augmented with annotations of nine OARs into the *Public Domain Database for Computational Anatomy (PDDCA)*[40] (http://www.imagenglab.com/newsite/pddca/), 31 CT images from two TCIA datasets (*Head-Neck Cetuximab* and *TCGA-HNSC*) were augmented with annotations of 21 OARs into the *TCIA Test & Validation Radiotherapy CT Planning Scan Dataset*[41] (https://github.com/deepmind/tcia-ct-scan-dataset/), while 140 CT images from two TCIA datasets (*Head-Neck Cetuximab* and *Head-Neck-PET-CT*) and 175 in-house CT images were augmented with annotations of 28 OARs into the *UaNet Dataset*[42] (https://github.com/uci-cbcl/UaNet/). Other HaN datasets that are publicly available but do not originate from TCIA are the *StructSeg Dataset* of 50 CT images with annotations of 22 OARs (https://structseg2019.grand-challenge.org/Dataset/), and the *MRI-RT Dataset*[43] of 15 CT and 15 MR images of the same patients with annotations of 23 OARs (https://figshare.com/s/a5e09113f5c07b3047df).

Similarly as in our case, some of the above listed datasets were devised and released in parallel with computational challenges aiming for a systematic and unbiased auto-segmentation evaluation.[19] For example, the *PDDCA* dataset was used for the *Head and Neck Auto-Segmentation Challenge* during the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) in 2015[40] (http://www.imagenglab.com/wiki/mediawiki/index.php?title=2015_MICCAI_Challenge), while the *StructSeg* dataset was used for the *Automatic Structure Segmentation for Radiotherapy Planning Challenge* during MICCAI 2019 (https://structseg2019.grand-challenge.org). Here, we need to mention that another computational challenge, that is, the *AAPM RT-MAC Challenge*, took place in 2019 during the AAPM Annual Meeting (https://www.aapm.org/GrandChallenge/RT-MAC/), and was based on 55 MR images with annotations of eight OARs.

From the overview of existing datasets and challenges for the evaluation of auto-segmentation in the HaN region, we can observe that most of them are focused on either CT or MR images, with the exception of the *MRI-RT Dataset*[43] that contains CT/MR image pairs of the same patients. However, this dataset is more oriented towards MR-to-CT registration[17] than segmentation, moreover, it is not associated with any challenge that would probably increase its visibility. We can, therefore, conclude that, to the best of our knowledge, to this date no other publicly available dataset exists for the evaluation of auto-segmentation methods given both CT and MR images of the same patients. From the perspective of dataset size, the HaN-Seg dataset consists of 56 CT/MR image pairs, and is, therefore, comparable to other publicly available datasets. On the other hand, from the perspective of OAR annotation, it contains reference segmentations of the most complete set of 30 OARs in the HaN region. Finally, from the perspective of auto-segmentation performance, the baseline results obtained by an off-the-shelf DL solution (i.e., nnU-Net) are inferior to the results reported by existing CT-only DL methods[9] but still comparable to some recent state-of-the-art DL methods.[44] With the release of the HaN-Seg dataset and deployment of the accompanying HaN-Seg challenge, our aim is, therefore, to test the hypothesis that the accuracy and reliability of OAR segmentation can be improved by exploiting the fused information from both CT and MR images, with the objective to design, develop, and evaluate novel auto-segmentation algorithms that rely on robust and accurate registration algorithms and can be benchmarked on a common dataset. Considering the recent growth in the application of AI, especially DL,[45] the devised HaN-Seg dataset has also the potential to contribute to a more objective and trustful reporting of research outcomes. Nevertheless, future versions of the HaN-Seg dataset may be enriched with additional CT/MR image pairs to capture a wider distribution of the anatomical variability among patients[10] or additional reference segmentations to evaluate the interobserver variability of manual annotation,[7] as well as with radiation dose distribution maps as defined by RT planning to evaluate the dosimetric impact of auto-segmentation results.[46]

6 | CONCLUSION

In this paper, we have described the publicly available HaN-Seg dataset consisting of 56 CT and MR images of the same patients with corresponding

reference segmentations of 30 OARs in the HaN region, and the accompanying HaN-Seg challenge, which aim to investigate the potential of multimodal imaging for the development of new and application of existing state-of-the-art OAR auto-segmentation algorithms. The outcomes will indicate whether the utilization of multimodal CT/MR image information is beneficial for obtaining more accurate and reliable OAR segmentation results, and therefore potentially aid clinical practice.

ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency (ARRS) under Grants J2-1732, P2-0232, and P3-0307, and partially by the Novo Nordisk Foundation under Grant NFF20OC0062056. The study was approved by the Ethics Committee of the Institute of Oncology Ljubljana, Slovenia, under No. ERID-EK/139.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The devised Set 1 can be publicly accessed and downloaded through the open-access repository Zenodo (https://zenodo.org) under the collection *HaN-Seg: The Head and Neck Organ-at-Risk CT & MR Segmentation Dataset* (https://doi.org/10.5281/zenodo.7442914).

REFERENCES

1. Chow L. Head and neck cancer. *N Engl J Med*. 2020;382:60-72.
2. Mody M, Rocco J, Yom S, Haddad R, Saba N. Head and neck cancer. *Lancet*. 2021;398:2289-2299.
3. Sung H, Ferlay J, Siegel R, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209-249.
4. Sahiner B, Pezeshk A, Hadjiiski L, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46:e1-e36.
5. Cui S, Tseng H-H, Pakela J, Ten Haken R, El Naqa I. Introduction to machine and deep learning for medical physicists. *Med Phys*. 2020;47:e127-e147.
6. Vandewinckele L, Claessens M, Dinkla A, et al. Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance. *Radiother Oncol*. 2020;153:55-66.
7. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol*. 2020;144:152-158.
8. Seo H, Khuzani M, Vasudevan V, et al. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. *Med Phys*. 2020;47:e148-e167.
9. Vrtovec T, Močnik D, Strojan P, Pernuš F, Ibragimov B. Auto-segmentation of organs at risk for head and neck radiotherapy planning: from atlas-based to deep learning methods. *Med Phys*. 2020;47:e929-e950.
10. Brouwer C, Steenbakkers R, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117:83-90.
11. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. *Med Phys*. 2019;46:2669-2682.
12. Cardenas C, Mohamed A, Yang J, et al. Head and neck cancer patient images for determining auto-segmentation accuracy in T2-weighted magnetic resonance imaging through expert manual segmentations. *Med Phys*. 2020;47:2317-2322.
13. Kieselmann J, Fuller C, Gurney-Champion O, Oelfke U. Cross-modality deep learning: contouring of MRI data from annotated CT data only. *Med Phys*. 2021;48:1673-1684.
14. Korte J, Hardcastle N, Ng S, Clark B, Kron T, Jackson P. Cascaded deep learning-based auto-segmentation for head and neck cancer patients: organs at risk on T2-weighted magnetic resonance imaging. *Med Phys*. 2021;48:7757-7772.
15. Dai X, Lei Y, Wang T, et al. Multi-organ auto-delineation in head-and-neck MRI for radiation therapy using mask scoring regional convolutional neural network. *Phys Med Biol*. 2022;67:025006.
16. Fu Y, Lei Y, Wang T, Curran W, Liu T, Yang X. A review of deep learning based methods for medical image multi-organ segmentation. *Phys Med*. 2021;85:107-122.
17. Boeke S, Mönnich D, van Timmeren J, Balermpas P. MR-guided radiotherapy for head and neck cancer: current developments, perspectives, challenges. *Front Oncol*. 2021;11:616156.
18. Huynh E, Boyle S, Campbell J, et al. Toward implementation of MR-guided radiation therapy for laryngeal cancer with healthy volunteer imaging and a custom MR-CT larynx phantom. *Med Phys*. 2022;49:1814-1821.
19. Maier-Hein L, Reinke A, Kozubek M, et al. BIAS: transparent reporting of biomedical image analysis challenges. *Med Image Anal*. 2020;66:101796.
20. Mayo C, Moran J, Bosch WR, et al. American Association of Physicists in Medicine Task Group 263: standardizing nomenclatures in radiation oncology. *Int J Radiat Oncol Biol Phys*. 2018;100:1057-1066.
21. Isensee F, Jaeger P, Kohl S, Petersen J, Maier-Hein K. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203-211.
22. Ronneberger O, Fischer P, Brox T. U-Net: convolutional neural networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. LNCS*. Vol 9351. Springer; 2015:234-241.
23. Klein S, Staring M, Murphy K, Viergever M, Pluim J. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196-205.
24. Brock K, Mutic S, McNutt T, Li H, Kessler M. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44:e43-e76.
25. Chowdhury F. *The Docker Handbook*. 2021 ed. 2021. https://www.freecodecamp.org/news/the-docker-handbook/
26. Gao Y, Huang R, Yang Y, et al. FocusNetv2: imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images. *Med Image Anal*. 2021;67:101831.
27. Lin H, Li Z, Yang Z, Wang Y. Variance-aware attention U-Net for multi-organ segmentation. *Med Phys*. 2021;48:7864-7876.
28. Jiang J, Elguindi S, Berry S, et al. Nested block self-attention multiple resolution residual network for multiorgan segmentation from CT. *Med Phys*. 2022;49:5244-5257.
29. Liu Y, Lei Y, Fu Y, et al. Head and neck multi-organ auto-segmentation on CT images aided by synthetic MRI. *Med Phys*. 2020;47:4294-4302.
30. Dai X, Lei Y, Wang T, et al. Automated delineation of head and neck organs at risk using synthetic MRI-aided mask scoring regional convolutional neural network. *Med Phys*. 2021;48:5862-5873.

31. Klages P, Benslimane I, Riyahi S, et al. Patch-based generative adversarial neural network models for head and neck MR-only planning. *Med Phys*. 2020;47:626-642.
32. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46:3565-3581.
33. Spadea M, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. *Med Phys*. 2021;48:6537-6566.
34. Qi M, Li Y, Wu A, Lu X, Zhou L, Song T. Multi-sequence MR generated sCT is promising for HNC MR-only RT: a comprehensive evaluation of previously developed sCT generation networks. *Med Phys*. 2022;49:2150-2158.
35. Yuan N, Rao S, Chen Q, Sensoy L, Qi J, Rong Y. Head and neck synthetic CT generated from ultra-low-dose cone-beam CT following Image Gently Protocol using deep neural network. *Med Phys*. 2022;49:3263-3277.
36. Ho S, Phua K, Wong L, Wen W, Goh B. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns*. 2020;1:100129.
37. Brunenberg E, Steinseifer IK, van den Bosch S, et al. External validation of deep learning-based contouring of head and neck organs at risk. *Phys Imaging Radiat Oncol*. 2020;15:8-15.
38. Ho T, Guo D, Jin D, et al. Comprehensive head and neck organs at risk segmentation using stratified learning and neural architecture search. *Int J Radiat Oncol Biol Phys*. 2021;111:e369-e370.
39. Fang Y, Wang J, Ou X, et al. The impact of training sample size on deep learning-based organ auto-segmentation for head-and-neck patients. *Phys Med Biol*. 2021;66:185012.
40. Raudaschl P, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. *Med Phys*. 2017;44:2020-2036.
41. Nikolov S, Blackwell S, Zverovitch A, et al. Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. *J Med Internet Res*. 2021;23:e26151.
42. Tang H, Chen X, Liu Y, et al. Clinically applicable deep learning framework for organs at risk delineation in CT images. *Nat Mach Intell*. 2019;1:480-491.
43. Joint Head and Neck MRI-Radiotherapy Development Cooperative, Kiser K, Meheissen MAM, et al. Prospective quantitative quality assurance and deformation estimation of MRI-CT image registration in simulation of head and neck radiotherapy patients. *Clin Transl Radiat Oncol*. 2019;18:120-127.
44. Udupa J, Liu T, Jin C, et al. Combining natural and artificial intelligence for robust automatic anatomy segmentation: application in neck and thorax auto-contouring. *Med Phys*. 2022;49:7118-7149.
45. El Naqa I, Boone J, Benedict S, et al. AI in medical physics: guidelines for publication. *Med Phys*. 2021;48:4711-4714.
46. Tsuji S, Hwang A, Weinberg V, Yom S, Quivey J, Xia P. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77:707-714.

How to cite this article: Podobnik G, Strojan P, Peterlin P, Ibragimov B, Vrtovec T. HaN-Seg: The head and neck organ-at-risk CT and MR segmentation dataset. *Med Phys*. 2023;1-11. https://doi.org/10.1002/mp.16197