Clinical and Translational Radiation Oncology 47 (2024) 100780

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.sciencedirect.com/journal/clinical-and-translational-radiation-oncology

Design and evaluation of a deep learning-based automatic segmentation of maxillary and mandibular substructures using a 3D U-Net

L. Melerowitz [a], S. Sreenivasa [a], M. Nachbar [a], A. Stsefanenka [a], M. Beck [a], C. Senger [a], N. Predescu [b], S. Ullah Akram [b], V. Budach [a], D. Zips [a], M. Heiland [c], S. Nahles [c], C. Stromberger [a,*]

[a] Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Radiation Oncology, Augustenburger Platz 1, 13353, Berlin, Germany
[b] MVision AI, Paciuksenkatu 29 00270 Helsinki, Finland
[c] Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Oral and Maxillofacial Surgery, Augustenburger Platz 1, 13353, Berlin, Germany

ARTICLE INFO

Keywords:
Jaw
Dental rehabilitation
Osteoradionecrosis
Radiation therapy
Artificial intelligence
Head and neck cancer

ABSTRACT

Background: Current segmentation approaches for radiation treatment planning in head and neck cancer patients (HNCP) typically consider the entire mandible as an organ at risk, whereas segmentation of the maxilla remains uncommon. Accurate risk assessment for osteoradionecrosis (ORN) or implant-based dental rehabilitation after radiation therapy may require a nuanced analysis of dose distribution in specific mandibular and maxillary segments. Manual segmentation is time-consuming and inconsistent, and there is no definition of jaw subsections.

Materials and methods: The mandible and maxilla were divided into 12 substructures. The model was developed from 82 computed tomography (CT) scans of HNCP and adopts an encoder-decoder three-dimensional (3D) U-Net structure. The efficiency and accuracy of the automated method were compared against manual segmentation on an additional set of 20 independent CT scans. The evaluation metrics used were the Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and surface DSC (sDSC).

Results: Automated segmentations were performed in a median of 86 s, compared to manual segmentations, which took a median of 53.5 min. The median DSC per substructure ranged from 0.81 to 0.91, and the median HD95 ranged from 1.61 to 4.22. The number of artifacts did not affect these scores. The maxillary substructures showed lower metrics than the mandibular substructures.

Conclusions: The jaw substructure segmentation demonstrated high accuracy, time efficiency, and promising results in CT scans with and without metal artifacts. This novel model could provide further investigation into dose relationships with ORN or dental implant failure in normal tissue complication prediction models.

1. Introduction

Head and neck cancer (HNC) is the seventh most prevalent malignancy worldwide [1,2]. Radiation therapy (RT) plays a pivotal role in the treatment of HNC, either as monotherapy, concurrent with chemotherapy, or as adjuvant therapy [3]. Survival rates remain poor at approximately 68,5% after five years, depending on cancer stage and diagnosis [4]. However, studies have shown a better chance of long-term survival for human papillomavirus (HPV)-positive cancers at approximately 80 % after five years [5]. Therefore, it is important to address both the short- and long-term side effects of RT. Despite advancements in radiation therapy techniques, the incidence of osteoradionecrosis (ORN) remains high, with reported rates of up to 10 % [6,7].

To prevent odontogenic complications during and after RT, it is necessary to extract loose, periodontally infected, and severely decayed teeth prior to therapy, as extractions are particularly susceptible to the risk of ORN [8–10]. Complications can lead to the discontinuation of RT,

* Corresponding author at: Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Radiation Oncology, Augustenburger Platz 1, 13353, Berlin, Germany.
E-mail address: carmen.stromberger@charite.de (C. Stromberger).

https://doi.org/10.1016/j.ctro.2024.100780
Received 13 November 2023; Received in revised form 9 April 2024; Accepted 17 April 2024
Available online 18 April 2024

which can significantly reduce the probability of cancer cure or contribute to long-term morbidity [11]. Extractions can result in partial or complete edentulism, and health-related quality of life (HRQoL) studies have shown that patients often experience significant difficulties with eating, swallowing, and speaking [2,12,13]. Dental implant (DI)-supported dentures are the preferred solution for restoring oropharyngeal function [14–17]. This approach requires strategically placed dental implants in the mandible and maxilla for secure attachment, potentially necessitating large bone exposures or bone augmentation [18]. These post-radiotherapeutic dentoalveolar procedures also pose a risk for ORN, with the administered radiation dose being a contributing factor that can even affect DI survival [16,19–24].

Historically, the maximum dose for the entire mandible was limited to 70 Gy for a volume of 0.1 cm^3. Recent recommendations suggest using volume constraints. For instance, it is recommended to limit the volume receiving 44 Gy to no more than 42 % or 58 Gy to no more than 25 % [25].

Nevertheless, ORN appears to be more prevalent in the premolar, molar, and retromolar regions than in other regions of the mandible, indicating that the jaws should be considered in a more differentiated approach [6].

Dental implants are considered safe for patients who receive a mean radiation dose of less than 38 Gy. In contrast, a high failure rate has been observed when the radiation dose exceeds 50 Gy, even if the implantation occurs over two years post-RT [16,19–23].

Further research is needed to determine whether a defined dose reduction in specific areas of the jaw will reduce ORN prevalence or increase DI survival. Additionally, a better understanding of jaw tolerances is necessary [25], which might enhance risk management.

Modern RT techniques, such as intensity-modulated RT (IMRT) or volumetric intensity-modulated arc therapy (VMAT), have increasingly replaced three-dimensional (3D) conventional planning, leading to more homogeneous and conformal irradiation and making sparing surrounding risk structures feasible [26,27]. Manual segmentation is necessary to use the maxilla and mandible as organs at risk (OAR) in RT planning. However, this process is resource-intensive, time-consuming, and repetitive. Challenges arise due to high-attenuation materials causing artifacts, variability in the maxillary sinus, patient-specific variations, and the complexity added by the proximity of other bony structures and the overlap of upper and lower teeth in adjacent CT slices [28–30]. Recent technological advancements have introduced automated medical image segmentation to alleviate clinical workload. Several studies have proposed semi-automated or fully automated methods for mandible segmentation in CT or cone beam CT scans [28–35]. These methods serve various purposes, such as OAR delineation, diagnosis, or maxillofacial surgery planning. Among the segmentation techniques, those based on deep learning (DL) architectures, such as U-Net, have shown superior performance in addressing mandible segmentation challenges. Ibragimov and Xing demonstrated the pioneering application of DL techniques using convolutional neural networks (CNN) for segmenting OARs in head and neck CT scans [34]. However, previous research has primarily focused on the mandible, with limited exploration of comprehensive techniques for automatically segmenting the maxilla. Moreover, studies have not yet attempted to segment the jaw into substructures to spare specific areas or illustrate dose distributions. This could further investigate dose relationships with ORN or implant failure in normal tissue complication prediction (NTCP) models [36].

To address this gap, this study presents a new model that automatically segments the maxilla and mandible into twelve anatomically based substructures using a 3D encoder-decoder U-Net [37] architecture.

2. Methods and materials

2.1. Study protocol

Before participating in the study, all patients involved provided informed consent. The study's protocol received approval from the Ethics Committee of Charité-Universitätsmedizin in Berlin, Germany, under the reference number EA2/155/21. The research was conducted following the principles of the Declaration of Helsinki.

2.2. Patient selection

The study used 82 CT scans to train and validate the model. The training set consisted of 64 CT scans, of which 53 were from Charité-Universitätsmedizin Berlin, Germany, and 11 were from MVision AI Oy, Helsinki, Finland, in compliance with the General Data Protection Regulation. The validation set included 18 scans, 15 from Charité and three from MVision AI Oy. After the development phase, the model underwent testing on an additional set of 20 independent CT scans obtained exclusively from Charité. The scans from Charité were from patients diagnosed with head and neck cancer and undergoing (chemo) radiation therapy. The group comprised 72 % males and 28 % females, with an average age of 66.8 years. The scans were collected between January 2013 and October 2020 using Siemens Somatom Sensation Open scanners. The median slice thickness was 2 mm (range 1–3 mm), and metal artifact reduction was not applied. Scans provided by MVision included five males, two females, and four individuals of unspecified gender, with specific age information available for three individuals: 42, 50, and 82 years. Siemens Sensation Open and Siemens Somatom Definition AS scanners were used for imaging, with a slice thickness ranging from 2 to 3 mm. Imaging data was missing for three cases. All scans had a slice size of 512 x 512 pixels, except for one which had a size of 256 x 256 pixels. CT scans with osseous destruction of the jawbone and HNC patients with a tumor site in the oral cavity were excluded.

2.3. Jaw segmentation

Dental reference structures for prosthetic restorations were used to determine segment boundaries in the mandible and maxilla. Therefore, the mandible (lower jaw, LJ) was divided into eight segments and the maxilla (upper jaw, UJ) into four segments. In cases where teeth were present, the dental crown was omitted, and only the bony structure was delineated. An interdisciplinary team of radiation oncologists and dental professionals clearly defined the criteria for delineating each segment. A summary of the segment definition is provided in Table 1. An example of the manual segmentation of the mandible and maxilla is shown graphically in representative CT slices in Fig. 1. Segments were delineated using Eclipse™ version 15.5 (Varian Medical Systems, Inc., Palo Alto, CA, USA) and the MVision AI online annotation platform, version 3.1. Dental status, including metal-related artifacts, was analyzed in each segment.

2.4. Model training and validation

The training dataset underwent extensive data augmentation to enhance model generalization and robustness. Preprocessing included a random selection of three consecutive slices from CT images, followed by cropping and padding to achieve a standardized axial patch size of 512 x 512 pixels. Intensity transformations such as random gamma transform, gamma inversion, additive and multiplicative intensity transform, Gaussian noise, Gaussian blur, and sharpening were applied with specified probabilities. Additionally, spatial transforms were employed, including random deformable transforms and scaling. Finally, intensity values were normalized with specified mean and standard deviation parameters and clamped within a predefined range.

The segmentation model adopts an encoder-decoder U-Net structure,

Table 1
Segments and Boundary Definitions of the 12 segments of the Jaw Model.

	Boundaries and Definition					
Structure	**Cranial**	**Caudal**	**Dorsal**	**Lateral**	**Medial**	**Anterior**
Upper jaw front left and right (UJFL/UJFR)	Base of the anterior nasal spine	Basilar edge of the maxilla	Dorsal edge of the maxilla	Plane tangent to the lateral edge of the nose ("canine line")	Center line through both incisive foramen	Anterior edge of the maxilla
Upper jaw molar left and right (UJML/UJMR)	Base of the anterior nasal spine/base of the maxillary sinus		Connecting line between both greater palatine foramen	Lateral edge of the maxilla	Medial edge of the maxilla	Plane tangent to the lateral edge of the nose ("canine line")
Lower Jaw front left and right (LJFL/LJFR)	Apical edge of the mandible (without teeth)	Basilar edge of the mandible	Plane tangent to the mental foramen	Lateral edge of the mandible	Medial edge of the mandible	Anterior edge of the mandible / mental protuberance
Lower Jaw molar left and right (LJML/LJMR)			Frontal plane: Connecting line between the body of the mandible and the ramus of the mandible			Plane tangent to the mental foramen
Lower Jaw ramus left and right (LJRL/LJRR)	Anterior part: Apical edge ramus of the posterior part of the mandible: Connecting line to LJ condyle left/right		Dorsal edge of the mandible			Frontal plane: Connecting line between the body of the mandible and the ramus of the mandible
Lower Jaw condyle left and right (LJCL/LJCR)	Cranial surface of the condylar head	Connecting line to LJ ramus left/right				Mandibular notch

with the encoder utilizing a Residual Neural Network (ResNet)-type [38] backbone pre-trained on ImageNet [39] to extract hierarchical features from input CT images. These features are symmetrically passed with skip connections to the decoder, comprising multiple Dense Convolutional Network (DenseNet) [40] blocks. Two Upsampling Convolutional (UpConv) [41] layers with 2x upsampling are then employed, followed by a 3D convolutional (Conv 3D) layer to generate final segmentations, with the same spatial size as the original input CT volume. The full architecture is shown in Fig. 2. The DenseNet decoder block is shown in Fig. 3 **(a)**. The output of the previous layer (x_1) is upsampled using bilinear interpolation to match the spatial size of the input provided by the skip connection (x_2). Then, the two tensors are concatenated and passed through a series of Batch Normalization (BN) [42], Rectified Linear Unit (ReLU), and two-dimensional convolutional (Conv 2D) layers. Finally, the output is concatenated to the original input of this sequence of layers. The UpConv 2x layer is shown in Fig. 3 **(b)**. After the input is passed through a BN, a ReLU, and a Conv 2D layer, it is upsampled using bilinear interpolation to have twice the spatial size compared to its original size.

The training procedure employed the sum of Dice loss and weighted cross-entropy loss to optimize the model's performance. The segmentation model utilized the Adam [43] optimizer with a learning rate of 3e-4, momentum of 0.9, and weight decay of 3e-5. The training process spanned 123 epochs, with each epoch comprising 250 iterations. The model was trained with a dual NVIDIA RTX 3090 GPU system (24 GB VRAM each), 64 GB of RAM, and an AMD Ryzen 9 5900X 12-Core CPU. The total training time was approximately 6.5 h.

A ReduceLROnPlateau learning rate scheduler was implemented with a threshold of 1e-2, dynamically adjusting the learning rate during training based on the model's performance plateau.

The cloud-based system uses pseudonymized DICOM data for automated segmentation and is connected to the clinic's RT planning system. A team of six annotators delineated the training and validation datasets, all with backgrounds in medical radiation oncology and dentistry and specific expertise in head and neck radiological anatomy. The bony structures were easily distinguishable from the surrounding tissue, and minor corrections were needed, as judged by two experienced head and neck radiation oncologists.

2.5. Quantitative evaluation

An independent set of 20 CTs was selected from the Charité-Universitätsmedizin Berlin dataset to evaluate the performance of the model (testing). Manual and automated segmentation were performed on each CT. Both segmentations were compared using quantitative scoring with the Dice Similarity Coefficient (DSC), 95 % Hausdorff Distance (HD95), and the surface DSC (sDSC). Manual and automated segmentation times were measured. The test dataset was delineated through manual segmentation by one annotator with a background in dentistry.

2.6. Statistics

The patient cohort was divided into a median split based on the number of segment-related artifacts. Patients with either no metal artifacts or a maximum of one artifact in a segment were compared to patients with more than one artifact in a segment. The number of segments with artifacts was non-parametrically correlated with DSC and HD95 using Spearman's correlation coefficient. In addition, the subgroups created by the median splitting in the patient sample were compared using non-parametric Mann-Whitney U tests. A p value $\leq$ 0.05 was considered statistically significant. MATLAB® R2021a (The MathWorks, Inc., Natick, MA, USA), IBM SPSS Statistics for Windows, version 28.0 (IBM Corp., Armonk, NY, USA), and Microsoft Excel 365 (Microsoft Corporation, Redmond, WA, USA) were used for statistical analysis.

3. Results

3.1. Jaw segmentation

The 3D Jaw Model created after segmentation and color subdivision is shown in Fig. 4.

The dental status of the training and validation dataset is shown in Table 2 and varied between full dentition, partial dentition (one to 15 teeth), and edentulous patients. The metal-related artifacts in the mandible and maxilla were listed.

3.2. Quantitative evaluation

The median number of CT slices in the test dataset of 20 patients was

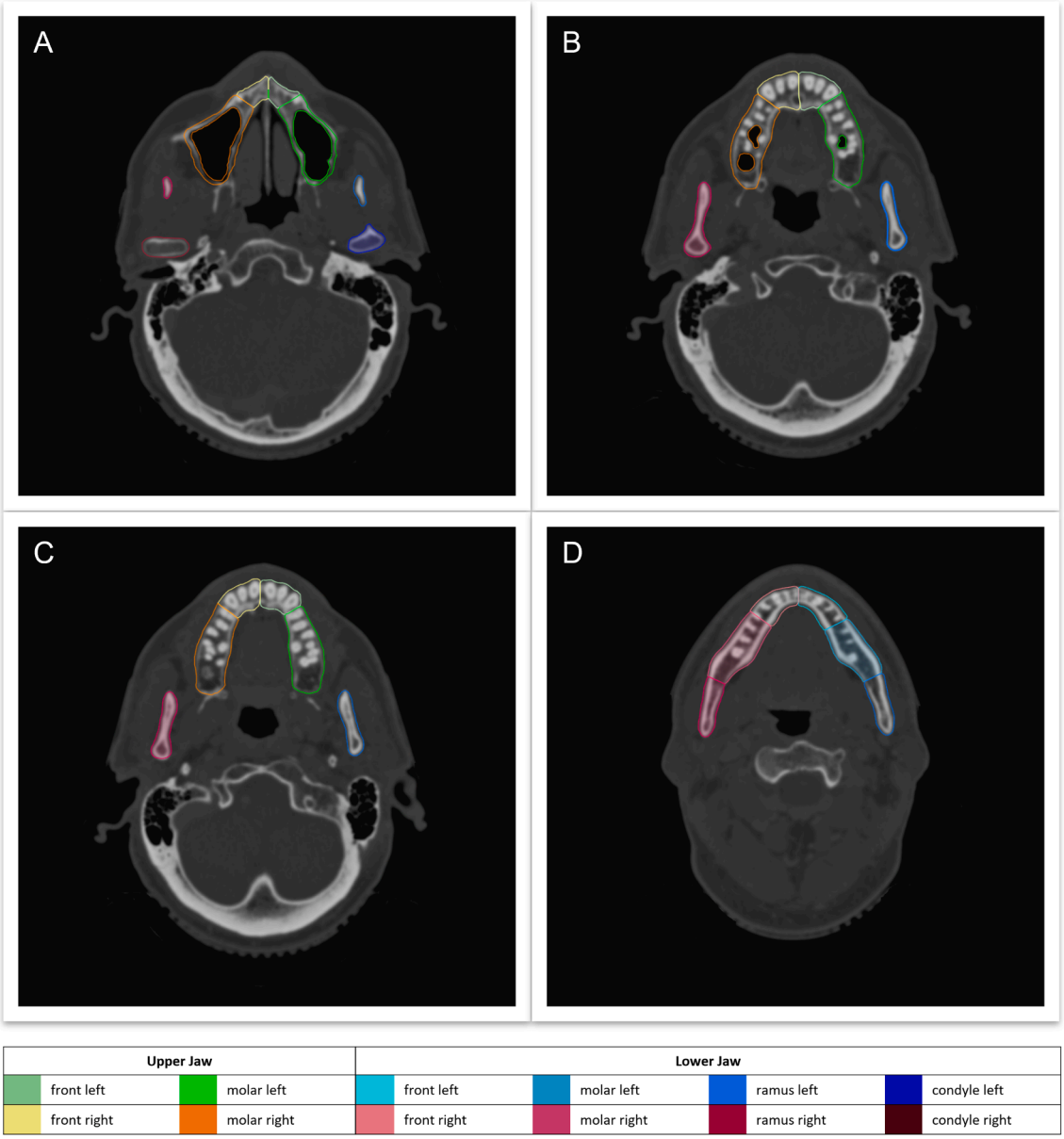

Fig. 1. Axial cross-sectional CT images of the jaw illustrating manual segmentation. (A) shows the maxilla with the maxillary sinus intentionally left unsegmented; (B) and (C) provide additional axial views of the segmented maxilla; (D) shows the axial view of the manually segmented mandible.

191 (range 108–314), and the median slice thickness was 2 mm (range 1–2.0 mm). The automated segmentations were generated in a median of 86 s (range 39–150 s). Conversely, the manual segmentations required a median time of 53.5 min (range 36–75 min). The comparison between manual and automated segmentation showed a maximum median HD95 of 4.22 mm for the LJML segment, while the LJCL segment showed a minimum median HD95 of 1.61 mm. The UJML segment had the widest range for HD95 at 11.01 mm, while the LJFR segment had the narrowest range at 4.57 mm. The maximum and minimum median DSC values were 0.91 for LJRR and 0.81 for UJML. The highest sDSC value (0.98) was observed for LJFR, while the lowest (0.87) was found for UJFR. A comprehensive summary of the results is presented in Table 3, Fig. 5, and Fig. 6.

According to the Mann-Whitney U test, the subgroup analysis using a median split based on the number of artifact-affected segments showed no significant differences between the subgroups in quantitative scores ($p = 0.579$ for DSC and HD95). In addition, non-parametric Spearman correlation analysis revealed no significant relationships between the number of artifact-affected segments and the median DSC and HD95 (DSC: $p = 0.628$; HD95: $p = 0.780$).

4. Discussion

This study presents a novel 3D encoder-decoder U-Net-based model that enables the segmentation of the jaw into 12 substructures.

Automated segmentation has emerged as a pivotal tool in radiation planning, enhancing accuracy, reproducibility, and efficiency, thereby reducing clinical workload and standardizing treatments [28–35,44]. Segmenting the entire mandible, a feature prevalent in many auto-segmentation tools, primarily aims to minimize the risk of ORN. However, the variable tumor locations and the associated diverse dose distributions within the entire mandible expose the limitations of current segmentation in dental rehabilitation post-RT [6]. Furthermore, segmentation of the maxilla is not always performed. In an independent test dataset, the model proved to be effective for segmentation. Manual segmentation of the jaw substructures in our study was time-consuming,

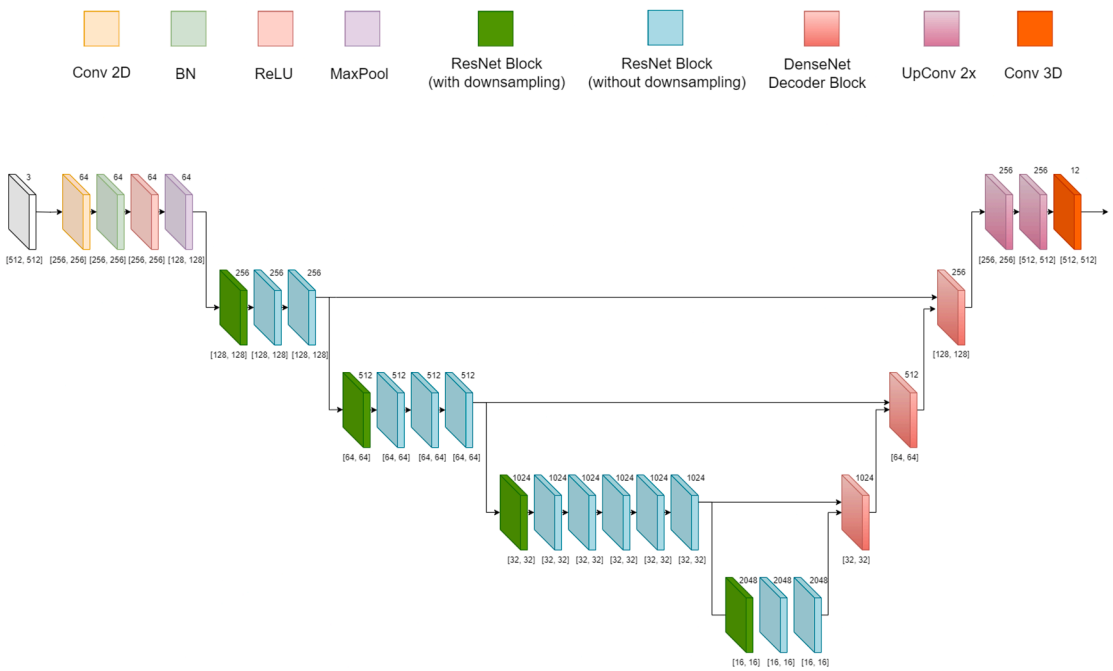

Fig. 2. Schematic representation of the U-Net architecture composed of a ResNet-based encoder and DenseNet decoder for feature extraction and image segmentation. The numbers above each layer denote the output feature size of the respective layer, while the numbers below the layers denote the spatial size of their output. Abbreviations: BN = Batch Normalization, ReLU = Rectified Linear Unit, Conv 2D = Two-Dimensional Convolutional Layer, Conv 3D = Three-Dimensional Convolutional Layer, UpConv = Upsampling Convolutional Layer, MaxPool = Maximum Pooling, ResNet = Residual Neural Network, DensNet = Dense Convolutional Network.

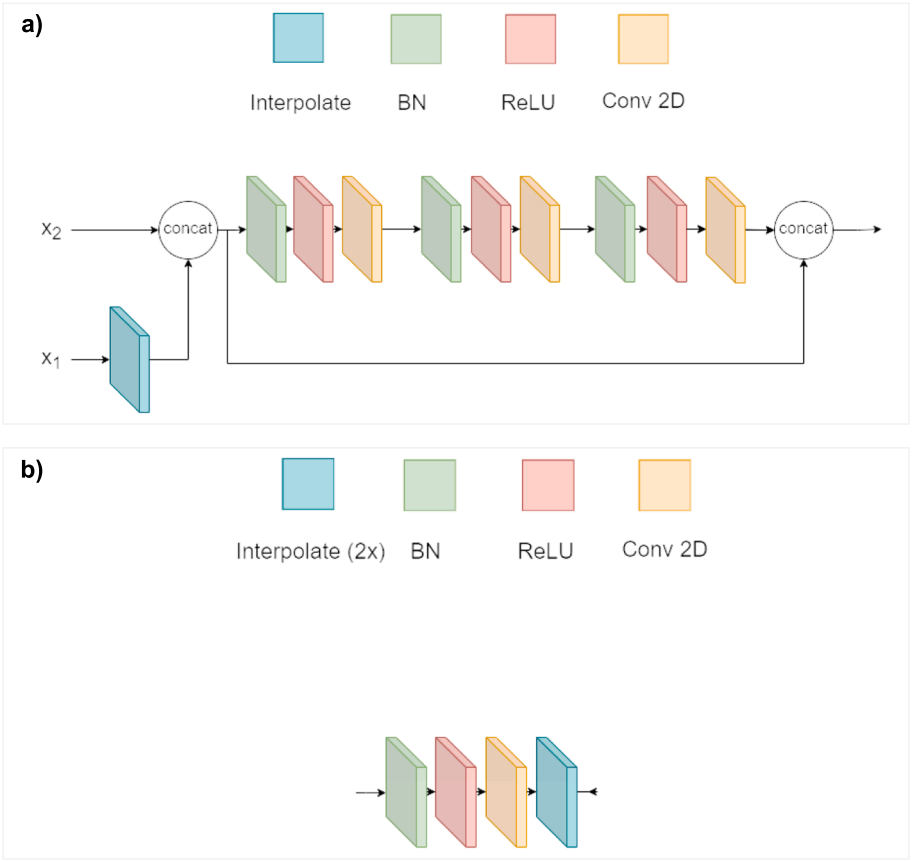

Fig. 3. Detailed illustrations of the decoder components within the U-Net architecture, featuring a) the DenseNet block and b) the UpConv 2x layer. Abbreviations: BN = Batch Normalization, ReLU = Rectified Linear Unit, Conv 2D = Two-Dimensional Convolutional Layer, Conv 3D = Three-Dimensional Convolutional Layer.

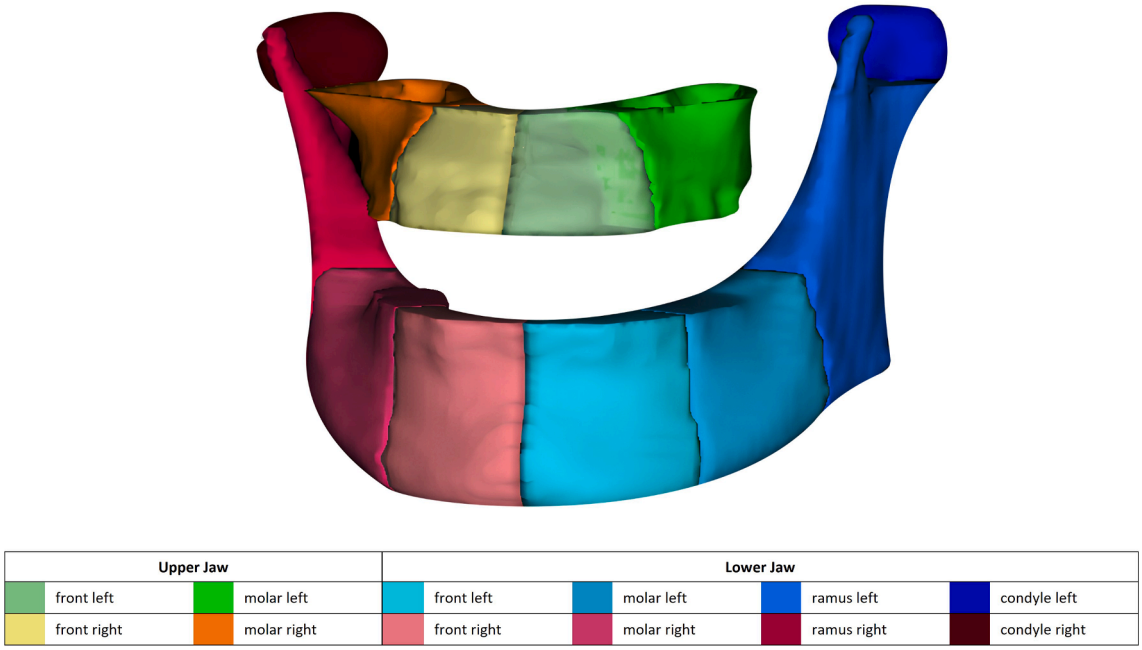

Fig. 4. Jaw Model – Segmentation of the Mandible and Maxilla into 12 substructures. The color-coded legend beneath the model delineates each segment. Upper Jaw (UJ) from left to right: UJ molar right, UJ front right, UJ front left, UJ molar left. Lower Jaw (LJ) from left to right: LJ condyle right, LJ ramus right, LJ molar right, LJ front right, LJ front left, LJ molar left, LJ ramus left, LJ condyle left.

Table 2
Dental Status of the Training, Validation and Test Dataset.

	Upper Jaw	Lower Jaw
Training dataset (n = 64)		
Full dentition	14	13
Partially edentulous	18	22
Edentulous	32	29
Metal-related artifacts	20	25
Validation dataset (n = 18)		
Full dentition	1	1
Partially edentulous	5	9
Edentulous	12	8
Metal-related artifacts	4	7
Test dataset (n = 20)		
Full dentition	12	13
Partially edentulous	7	6
Edentulous	1	1
Metal-related artifacts	13	10

Table 3
Quantitative Evaluation comparing the manual and automated segmentation of the Jaw Model in the test dataset per substructure.

	Performance metric		
	DSC (median, range)	**HD95 [mm] (median, range)**	**sDSC [%] (median, range)**
Whole UJ	0.82	3.31	0.89
Whole LJ	0.89	2.80	0.96
Substructures			
UJFL	0.82 (0.67–0.89)	3.61 (2.29–7.20)	0.87 (0.70–0.98)
UJFR	0.81 (0.66–0.88)	3.16 (2.05–7.24)	0.88 (0.70–0.99)
UJML	0.81 (0.58–0.90)	3.46 (1.43–12.43)	0.90 (0.62–0.99)
UJMR	0.82 (0.52–0.89)	2.87 (2.09–12.18)	0.91 (0.62–0.99)
LJFL	0.88 (0.76–0.94)	2.99 (1.56–6.86)	0.97 (0.75–1.00)
LJFR	0.88 (0.81–0.92)	2.61 (1.93–6.50)	0.98 (0.80–1.00)
LJML	0.88 (0.80–0.98)	4.22 (1.12–8.15)	0.89 (0.81–1.00)
LJMR	0.90 (0.81–0.94)	3.22 (1.76–9.20)	0.93 (0.81–0.99)
LJRL	0.88 (0.79–0.98)	3.02 (0.40–6.40)	0.95 (0.87–1.00)
LJRR	0.91 (0.79–0.95)	2.41 (1.08–10.01)	0.97 (0.85–1.00)
LJCL	0.90 (0.70–0.95)	1.61 (1.01–6.64)	0.96 (0.75–1.00)
LJCR	0.89 (0.63–0.94)	1.62 (1.05–7.46)	0.96 (0.73–1.00)

Legend: DSC = Dice surface coefficient, HD95 = 95 % Hausdorff Distance, sDSC = surface DSC. UJFL = upper jaw front left, UJFR = upper jaw front right, UJML = upper jaw molar left, UJMR = upper jaw molar right, LJFL = lower jaw front left, LJFR = lower jaw front right, LJML = lower jaw molar left, LJMR = lower jaw molar right.

taking approximately 53 min. Artifacts and complex structures such as the maxillary sinus and precise adherence to substructure boundaries contributed to more complex and time-consuming delineation. In contrast, the automated segmentation process in our study took an average of 86 s, surpassing Ibragimov's initial work on OAR segmentation in the head and neck region using CNN networks, which took approximately 4 min [34]. Morita et al., who developed a two-dimensional U-net architecture for facial bone segmentation, were able to reduce the required time to an average of 10 s per region (e.g., maxilla and mandible) [35]. Our longer output time could be explained by the fact that our model is additionally divided into subsections.

The evaluation of the automated segmentation in this study involved multiple metrics, using manual segmentation as the ground truth. An independent test dataset analysis showed a median DSC of 0.88, a median HD95 of 3.00 mm, and a median sDSC of 0.94 % per patient across all substructures. These results demonstrate robust performance despite variations in patient characteristics such as artifact count, dental status, and bone volume. However, this study represents the first automated segmentation of maxillary and mandibular substructures, making direct comparisons with previous studies difficult.

Furthermore, there is no uniform definition of which metric levels are sufficient for clinical use. The multi-atlas segmentation strategy, a traditional method, achieved a DSC of 0.86 for the entire mandible [33]. However, these conventional methods have been surpassed by CNNs due to their inability to handle high noise or metal artifacts and the need for expert analysis, showing poor individualization for mandibular segmentation cases [45].

Ibragimov and Xing observed a median DSC of 0.89 using CNN [34]. The DSC could be improved by combining methods, as shown by Tong et al. [46]. They combined a fully convolutional neural network (FCNN) with a shape representation model (SRM), achieving a DSC of 0.92 for

	Substructure											
Patient	UJFL	UJFR	UJML	UJMR	LJFL	LJFR	LJML	LJMR	LJRL	LJRR	LJCL	LJCR
1	0,79	0,84	0,81	0,81	0,89	0,88	0,91	0,85	0,89	0,80	0,88	0,87
2	0,89	0,82	0,58	0,52	0,81	0,83	0,81	0,89	0,84	0,86	0,82	0,88
3	0,82	0,86	0,78	0,75	0,90	0,91	0,87	0,90	0,84	0,89	0,79	0,84
4	0,83	0,80	0,59	0,73	0,82	0,81	0,84	0,91	0,85	0,93	0,86	0,93
5	0,80	0,80	0,85	0,80	0,89	0,91	0,88	0,94	0,88	0,94	0,88	0,90
6	0,80	0,84	0,87	0,89	0,89	0,88	0,89	0,92	0,88	0,92	0,94	0,89
7	0,82	0,76	0,80	0,86	0,91	0,88	0,90	0,86	0,89	0,88	0,85	0,78
8	0,82	0,82	0,90	0,89	0,89	0,87	0,91	0,90	0,93	0,91	0,93	0,90
9	0,87	0,85	0,87	0,88	0,87	0,87	0,92	0,91	0,90	0,91	0,92	0,89
10	0,67	0,68	0,80	0,86	0,83	0,87	0,85	0,89	0,86	0,91	0,89	0,94
11	0,68	0,80	0,70	0,76	0,94	0,92	0,89	0,84	0,89	0,85	0,85	0,75
12	0,84	0,88	0,81	0,82	0,86	0,87	0,86	0,90	0,87	0,91	0,94	0,94
13	0,81	0,81	0,86	0,86	0,88	0,87	0,90	0,90	0,90	0,93	0,90	0,88
14	0,82	0,86	0,88	0,85	0,90	0,92	0,88	0,94	0,85	0,92	0,89	0,94
15	0,83	0,78	0,80	0,83	0,87	0,89	0,80	0,81	0,79	0,79	0,70	0,63
16	0,75	0,81	0,83	0,82	0,93	0,91	0,91	0,89	0,88	0,85	0,91	0,87
17	0,71	0,66	0,80	0,82	0,87	0,86	0,98	0,93	0,98	0,95	0,92	0,91
18	0,86	0,81	0,87	0,82	0,76	0,82	0,83	0,87	0,89	0,92	0,95	0,94
19	0,84	0,88	0,80	0,88	0,89	0,91	0,87	0,94	0,84	0,88	0,92	0,90
20	0,73	0,70	0,64	0,68	0,88	0,89	0,88	0,90	0,88	0,93	0,93	0,89
Median	**0,82**	**0,81**	**0,81**	**0,82**	**0,88**	**0,88**	**0,88**	**0,90**	**0,88**	**0,91**	**0,90**	**0,89**

Fig. 5. Heatmap Visualization of DSC across all substructures and cases.

the FCNN alone and 0.94 with the SRM-constrained FCNN. Liu et al. evaluated 19 models from 18 studies for automated segmentation of the entire mandible and found a pooled DSC of 0.92 [47]. Previous studies have reported HD95 values ranging from 1.98 to 2.83 [48,49]. The study found that there were performance differences between the substructures, with the maxillary substructures generally having lower metrics than the mandible, as shown in Table 3, Fig. 5 and Fig. 6.

In this study, the median DSC for the entire maxilla was 0.82, compared to 0.89 for the mandible when the substructures were combined. Furthermore, the maxilla's HD95 values were worse than those for the mandible (3.31 vs. 2.8). The maxillary molar substructures showed higher standard deviations in performance metrics, indicating variability in segmentation accuracy due to the maxilla's complex anatomy and a higher incidence of artifacts (65 %). However, artifacts did not significantly affect the overall performance metrics, which is consistent with the results presented by Ilesan et al. [50]. They reported a similar DSC for automated mandibular bone segmentation with and without artifacts.

Morita et al. also had a worse DSC for segmentation of the entire maxilla compared to the entire mandible (0.91 vs. 0.98) [35]. Training and validating the model with additional cases could increase the consistency of maxillary segmentation performance to improve its robustness.

Overall, the metrics' performance is inferior to that of the compared

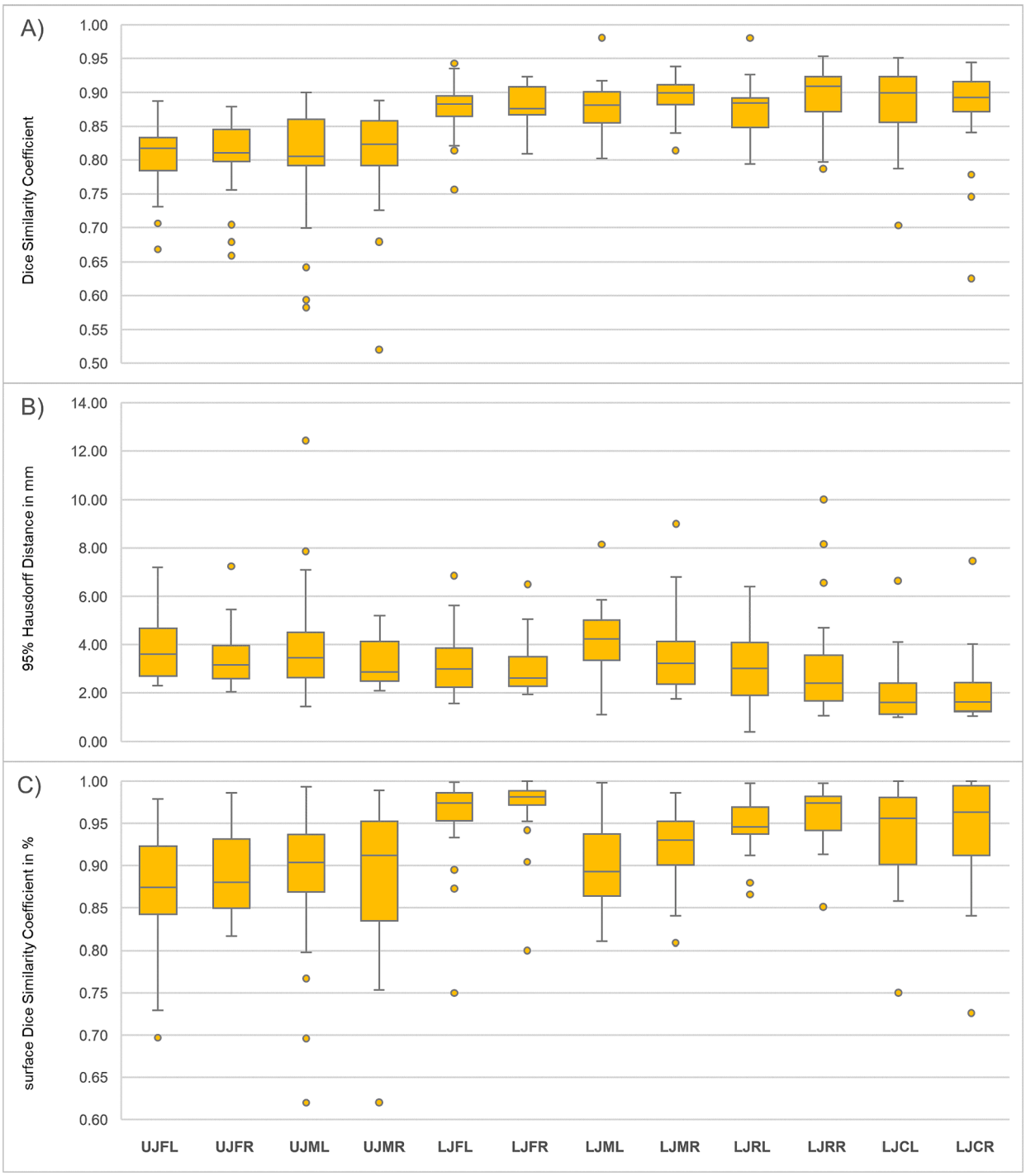

Fig. 6. Box and whisker plots comparing the manual and automated segmentation of the Jaw Model in the test dataset. A) Dice Similarity Coefficient, B) 95 % Hausdorff Distance, C) surface Dice Similarity Coefficient. The median is shown as a horizontal line within each box, while yellow dots represent outliers (exceeding 1.5 times the interquartile range of the upper and lower quartiles). Abbreviations: UJFL = upper jaw front left, UJFR = upper jaw front right, UJML = upper jaw molar left, UJMR = upper jaw molar right, LJFL = lower jaw front left, LJFR = lower jaw front right, LJML = lower jaw molar left, LJMR = lower jaw molar right, LJRL = lower jaw ramus left, LJRR = lower jaw ramus right, LJCL = lower jaw condyle left, LJCR = lower jaw condyle right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

studies. It should be noted that the division into substructures within the same density of tissue is subject to increased variability since the delineation can be more difficult.

In this study, the boundaries are always bone-to-bone structures with equal Hounsfield units compared to bone-to-tissue boundaries. Further investigation is required to determine whether the present results would be better if the maxilla and mandible were considered a single segment.

One limitation of this study is that the Jaw Model mainly uses scans from a single institution. This may limit its ability to adapt to CT scans with different imaging characteristics, potentially affecting the evaluation metrics. Additionally, the method was only applied to intact jawbones with standard dental artifacts, excluding those with complex artifacts such as reconstruction plates that could compromise segmentation integrity. It is important to note that the absence of analysis of

interobserver and intraobserver variability is due to the precise segmentation boundaries requiring minimal adjustments. The evaluation of the model on only 20 CT scans highlights the need for further improvement and broader validation to enhance its robustness and clarify its limitations.

5. Conclusions

The results of this study demonstrate the feasibility of dividing the maxilla and mandible into subsections using the 3D U-Net architecture. The model's ability to generate a larger, more coherent data set promises significant contributions to the research field. This will allow further investigations into ORN development, more detailed dose distributions within the jaws, and, therefore, the feasibility of sparing segments or even improving dental rehabilitation.

CRediT authorship contribution statement

L. Melerowitz: Conceptualization, Writing – original draft, Investigation, Methodology, Validation, Visualization, Formal analysis, Data curation. **S. Sreenivasa:** Conceptualization, Methodology, Writing – review & editing, Formal analysis. **M. Nachbar:** Writing – review & editing, Methodology, Formal analysis. **A. Stsefanenka:** Writing – review & editing, Data curation. **M. Beck:** Writing – review & editing. **C. Senger:** Writing – review & editing. **N. Predescu:** Methodology, Software, Writing – review & editing. **S. Ullah Akram:** Methodology, Software. **V. Budach:** Writing – review & editing. **D. Zips:** Writing – review & editing. **M. Heiland:** Writing – review & editing. **S. Nahles:** Writing – review & editing, Supervision, Conceptualization, Methodology, Project administration. **C. Stromberger:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors disclose institutional collaboration with MVision AI, which was responsible for the technical development of the presented model. MVision AI did not participate in the analysis or evaluation of the test dataset.

Acknowledgement

We would like to express our sincere gratitude to Jarkko Niemelä, Jani Pehkonen, and Szabolcs-Botond Lőrincz-Molnár for their invaluable support and expertise in advancing our research.

References

[1] Gormley M, Creaney G, Schache A, Ingarfield K, Conway DI. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. Br Dent J 2022;233(9):780–6.
[2] Taylor KJ, Amdal CD, Bjordal K, Astrup GL, Herlofson BB, Duprez F, et al. Long-term health-related quality of life in head and neck cancer survivors: a large multinational study. Int J Cancer 2024;154(10):1772–85.
[3] Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Primers 2020;6(1):92.
[4] Barsouk A, Aluru JS, Rawla P, Saginala K, Epidemiology BA. Risk factors, and prevention of head and neck squamous cell carcinoma. Med Sci (Basel) 2023;11(2).
[5] Du E, Mazul AL, Farquhar D, Brennan P, Anantharaman D, Abedi-Ardekani B, et al. Long-term survival in head and neck cancer: impact of site, stage, smoking, and human papillomavirus status. Laryngoscope 2019;129(11):2506–13.
[6] Sijtsema ND, Verduijn GM, Nasserinejad K, van Norden Y, Mast H, van der Lugt A, et al. Development of a local dose-response relationship for osteoradionecrosis within the mandible. Radiother Oncol 2023;186:109736.
[7] Kubota H, Miyawaki D, Mukumoto N, Ishihara T, Matsumura M, Hasegawa T, et al. Risk factors for osteoradionecrosis of the jaw in patients with head and neck squamous cell carcinoma. Radiat Oncol 2021;16(1):1.
[8] Beaumont S, Bhatia N, McDowell L, Fua T, McCullough M, Celentano A, et al. Timing of dental extractions in patients undergoing radiotherapy and the incidence of osteoradionecrosis: a systematic review and meta-analysis. Br J Oral Maxillofac Surg 2021;59(5):511–23.
[9] Jawad H, Hodson NA, Nixon PJ. A review of dental treatment of head and neck cancer patients, before, during and after radiotherapy: part 1. Br Dent J 2015;218(2):65–8.
[10] Jawad H, Hodson NA, Nixon PJ. A review of dental treatment of head and neck cancer patients, before, during and after radiotherapy: part 2. Br Dent J 2015;218(2):69–74.
[11] Hamidi M, Moody JS, Kozak KR. Refusal of radiation therapy and its associated impact on survival. Am J Clin Oncol 2010;33(6):629–32.
[12] Petrovic I, Rosen EB, Matros E, Huryn JM, Shah JP. Oral rehabilitation of the cancer patient: a formidable challenge. J Surg Oncol 2018;117(8):1729–35.
[13] Pieralli S, Spies BC, Schweppe F, Preissner S, Nelson K, Heiland M, et al. Retrospective long-term clinical evaluation of implant-prosthetic rehabilitations after head and neck cancer therapy. Clin Oral Implants Res 2021;32(4):470–86.
[14] Barrowman RA, Wilson PR, Wiesenfeld D. Oral rehabilitation with dental implants after cancer treatment. Aust Dent J 2011;56(2):160–5.
[15] Beech N, Robinson S, Porceddu S, Batstone M. Dental management of patients irradiated for head and neck cancer. Aust Dent J 2014;59(1):20–8.
[16] Raguse JD, Hossamo J, Tinhofer I, Hoffmeister B, Budach V, Jamil B, et al. Patient and treatment-related risk factors for osteoradionecrosis of the jaw in patients with head and neck cancer. Oral Surg Oral Med Oral Pathol Oral Radiol. 2016;121(3):215-21.e1.
[17] Rahman RA, Ngeow WC, Chai WL, Ramli R. Managing complications of radiation therapy in head and neck cancer patients: part III. Provision of Dentures Singapore Dent J 2006;28(1):7–10.
[18] Rouers M, Antoni D, Thompson A, Truntzer P, Haoming QC, Bourrier C, et al. Maxillary and mandible contouring in patients with a head and neck area irradiation. Pract Radiat Oncol 2016;6(3):e61–72.
[19] Ettl T, Junold N, Zeman F, Hautmann M, Hahnel S, Kolbeck C, et al. Implant survival or implant success? Evaluation of implant-based prosthetic rehabilitation in head and neck cancer patients-a prospective observational study. Clin Oral Investig 2020;24(9):3039–47.
[20] Sammartino G, Marenzi G, Cioffi I, Teté S, Mortellaro C. Implant therapy in irradiated patients. J Craniofac Surg 2011;22(2):443–5.
[21] Neckel N, Wagendorf P, Sachse C, Stromberger C, Vach K, Heiland M, et al. Influence of implant-specific radiation doses on peri-implant hard and soft tissue: an observational pilot study. Clin Oral Implants Res 2021;32(2):249–61.
[22] Lee J, Lee JJB, Cha IH, Park KR, Lee CG. Risk factor analysis of dental implants in patients with irradiated head and neck cancer. Head Neck 2022;44(8):1816–24.
[23] Li J, Feng K, Ye L, Liu Y, Sun Y, Wu Y. Influence of radiotherapy on dental implants placed in individuals before diagnosed with head and neck cancer: focus on implant-bed-specific radiation dosage. Clin Oral Investig 2022;26(9):5915–22.
[24] Khoo SC, Nabil S, Fauzi AA, Yunus SSM, Ngeow WC, Ramli R. Predictors of osteoradionecrosis following irradiated tooth extraction. Radiat Oncol 2021;16(1):130.
[25] Lee CT, Litwin S, Yao C, Liu JC, Ridge JA, Galloway TJ. Osteoradionecrosis rate in oropharynx cancer treated with dose volume histogram based constraints. Radiother Oncol 2022;176:215–21.
[26] Das Majumdar SK, Amritt A, Dhar SS, Barik S, Beura SS, Mishra T, et al. A dosimetric study comparing 3D-CRT vs. IMRT vs. VMAT in left-sided breast cancer patients after mastectomy at a tertiary care centre in Eastern India. Cureus. 2022;14(3):e23568.
[27] Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. Br J Radiol 2011;84(1007):967–96.
[28] Qiu B, Guo J, Kraeima J, Glas HH, Borra RJH, Witjes MJH, et al. Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network. Phys Med Biol 2019;64(17):175020.
[29] Qiu B, van der Wel H, Kraeima J, Glas HH, Guo J, Borra RJH, et al. Automatic segmentation of mandible from conventional methods to deep learning-a review. J Pers Med 2021;11(7).
[30] Verhelst PJ, Smolders A, Beznik T, Meewis J, Vandemeulebroucke A, Shaheen E, et al. Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography. J Dent 2021;114:103786.
[31] Xu J, Liu J, Zhang D, Zhou Z, Jiang X, Zhang C, et al. Automatic mandible segmentation from CT image using 3D fully convolutional neural network based on DenseASPP and attention gates. Int J Comput Assist Radiol Surg 2021;16(10):1785–94.
[32] Trimpl MJ, Primakov S, Lambin P, Stride EPJ, Vallis KA, Gooding MJ. Beyond automatic medical image segmentation-the spectrum between fully manual and fully automatic delineation. Phys Med Biol 2022;67(12).
[33] Peroni M, Sharp GC, Golland P, Baroni G. WE-E-213CD-02: Gaussian Weighted Multi-Atlas Based Segmentation for Head and Neck Radiotherapy Planning. Med Phys. 2012;39(6Part27):3959.
[34] Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys 2017;44(2):547–57.
[35] Morita D, Mazen S, Tsujiko S, Otake Y, Sato Y, Numajiri T. Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net. Int J Oral Maxillofac Surg 2023;52(7):787–92.
[36] Humbert-Vidan L, Patel V, Andlauer R, King AP, Guerrero Urbano T, editors. Prediction of mandibular ORN incidence from 3D radiation dose distribution maps using deep learning. International Workshop on Applications of Medical AI; 2022: Springer.

[37] Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18; 2015: Springer.
[38] He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
[39] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L, editors. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; 2009: Ieee.
[40] Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv: 14041869. 2014.
[41] Odena A, Dumoulin V, Olah C. Deconvolution and Checkerboard Artifacts Distill 2016;1(10):e3.
[42] Ioffe S, Szegedy C, editors. Batch normalization: accelerating deep network training by reducing internal covariate shift. International conference on machine learning; 2015: pmlr.
[43] Kingma DP, Ba J. Adam: A method for stochastic optimization. CoRR. 2014;abs/1412.6980.
[44] Kiljunen T, Akram S, Niemelä J, Löyttyniemi E, Seppälä J, Heikkilä J, et al. A deep learning-based automated CT segmentation of prostate cancer anatomy for radiation therapy planning-a retrospective multicenter study. Diagnostics (Basel) 2020;10(11).
[45] Qiu B, Guo J, Kraeima J, Glas HH, Zhang W, Borra RJH, et al. Recurrent convolutional neural networks for 3D mandible segmentation in computed tomography. J Pers Med 2021;11(6).
[46] Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys 2018;45(10): 4558–67.
[47] Liu P, Sun Y, Zhao X, Yan Y. Deep learning algorithm performance in contouring head and neck organs at risk: a systematic review and single-arm meta-analysis. Biomed Eng Online 2023;22(1):104.
[48] Albrecht T, Gass T, Langguth C, Lüthi M. Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning. MIDAS J 2015.
[49] Mannion-Haworth R, Bowes M, Ashman A, Guillard G, Brett A, Vincent G. Fully automatic segmentation of head and neck organs using active appearance models. MIDAS J 2015;6:8.
[50] Ileşan RR, Beyer M, Kunz C, Thieringer FM. Comparison of artificial intelligence-based applications for mandible segmentation: from established platforms to in-house-developed software. Bioengineering (basel) 2023;10(5).