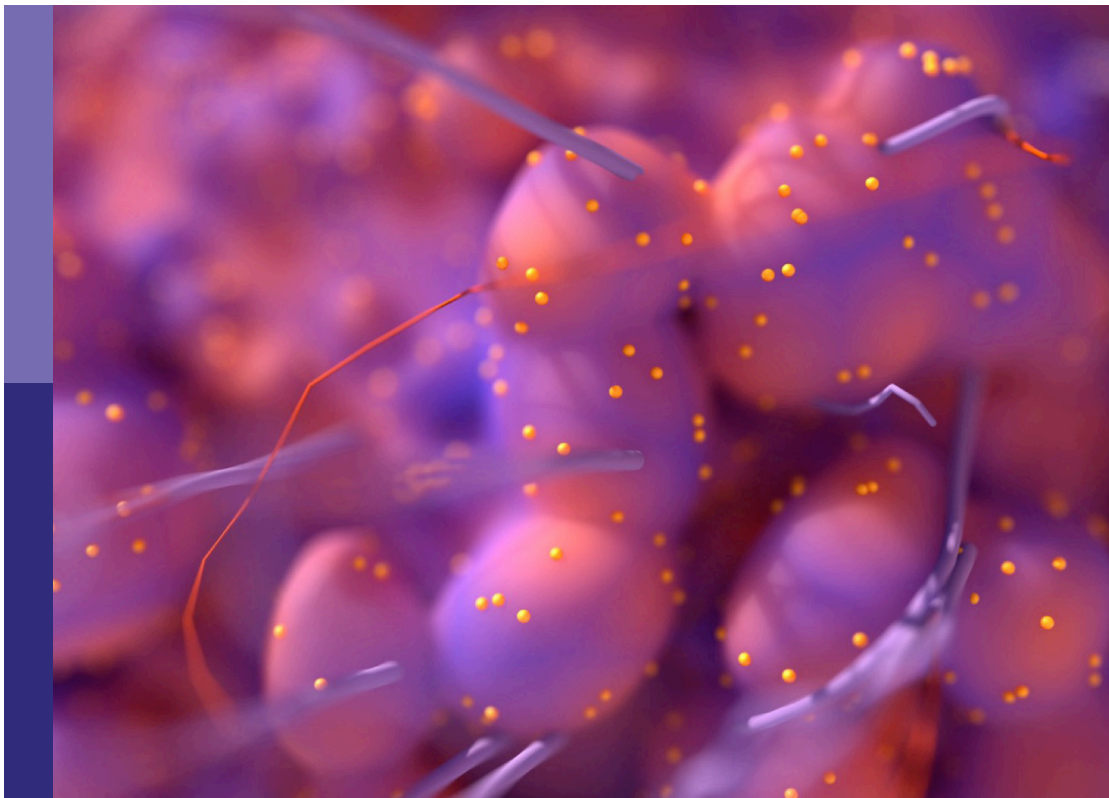
frontiers | Research Topics

# Radiotherapy for head and neck cancers

**Edited by**
Xinyuan Chen, Minglei Kang and Jun-Lin Yi

**Published in**
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-6224-6
DOI 10.3389/978-2-8325-6224-6

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Radiotherapy for head and neck cancers

**Topic editors**

Xinyuan Chen — Center for National Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, China

Minglei Kang — New York Proton Center, United States

Jun-Lin Yi — Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, China

**Citation**

Chen, X., Kang, M., Yi, J.-L., eds. (2025). *Radiotherapy for head and neck cancers*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6224-6

# Table of contents

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 November 2023
DOI 10.3389/fonc.2023.1278180



OPEN ACCESS

EDITED BY
Minglei Kang,
New York Proton Center, United States

REVIEWED BY
Jiajian Shen,
Mayo Clinic Arizona, United States
Matthias Gobbert,
University of Maryland, Baltimore County, United States

*CORRESPONDENCE
Xiaofeng Yang
Xiaofeng.yang@emory.edu

RECEIVED 15 August 2023
ACCEPTED 06 November 2023
PUBLISHED 23 November 2023

CITATION
Gao Y, Chang C-W, Roper J, Axente M, Lei Y, Pan S, Bradley JD, Zhou J, Liu T and Yang X (2023) Single energy CT-based mass density and relative stopping power estimation for proton therapy using deep learning method.
*Front. Oncol.* 13:1278180.
doi: 10.3389/fonc.2023.1278180



# Single energy CT-based mass density and relative stopping power estimation for proton therapy using deep learning method

Yuan Gao[1], Chih-Wei Chang[1], Justin Roper[1], Marian Axente[1], Yang Lei[1], Shaoyan Pan[2], Jeffrey D. Bradley[3], Jun Zhou[1], Tian Liu[1,4] and Xiaofeng Yang[1,2]*

[1]Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA, United States, [2]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States, [3]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [4]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

**Background:** The number of patients undergoing proton therapy has increased in recent years. Current treatment planning systems (TPS) calculate dose maps using three-dimensional (3D) maps of relative stopping power (RSP) and mass density. The patient-specific maps of RSP and mass density were obtained by translating the CT number (HU) acquired using single-energy computed tomography (SECT) with appropriate conversions and coefficients. The proton dose calculation uncertainty of this approach is 2.5%-3.5% plus 1 mm margin. SECT is the major clinical modality for proton therapy treatment planning. It would be intriguing to enhance proton dose calculation accuracy using a deep learning (DL) approach centered on SECT.

**Objectives:** The purpose of this work is to develop a deep learning method to generate mass density and relative stopping power (RSP) maps based on clinical single-energy CT (SECT) data for proton dose calculation in proton therapy treatment.

**Methods:** Artificial neural networks (ANN), fully convolutional neural networks (FCNN), and residual neural networks (ResNet) were used to learn the correlation between voxel-specific mass density, RSP, and SECT CT number (HU). A stoichiometric calibration method based on SECT data and an empirical model based on dual-energy CT (DECT) images were chosen as reference models to evaluate the performance of deep learning neural networks. SECT images of a CIRS 062M electron density phantom were used as the training dataset for deep learning models. CIRS anthropomorphic M701 and M702 phantoms were used to test the performance of deep learning models.

**Results:** For M701, the mean absolute percentage errors (MAPE) of the mass density map by FCNN are 0.39%, 0.92%, 0.68%, 0.92%, and 1.57% on the brain, spinal cord, soft tissue, bone, and lung, respectively, whereas with the SECT stoichiometric method, they are 0.99%, 2.34%, 1.87%, 2.90%, and 12.96%. For RSP

maps, the MAPE of FCNN on M701 are 0.85%, 2.32%, 0.75%, 1.22%, and 1.25%, whereas with the SECT reference model, they are 0.95%, 2.61%, 2.08%, 7.74%, and 8.62%.

**Conclusion:** The results show that deep learning neural networks have the potential to generate accurate voxel-specific material property information, which can be used to improve the accuracy of proton dose calculation.

**Advances in knowledge:** Deep learning-based frameworks are proposed to estimate material mass density and RSP from SECT with improved accuracy compared with conventional methods.

KEYWORDS

deep learning, CT, relative stopping power, mass density, proton therapy

# 1 Introduction

The number of patients receiving proton therapy treatment is rising each year. Current proton treatment planning systems (TPS) incorporate relevant proton energy deposition physics into the process, determining the patient irradiation pattern. All patients treated with radiotherapy (including protons) are simulated using computed tomography (CT) typically acquired using a single energy scanning protocol. This data is utilized for geometrical implementation of the therapy and also to characterize the patient from the point of view of the probability of charged particle interactions. Therefore, proton dose calculation accuracy is dependent on the capability of TPS to characterize patient tissues based on CT imaging (1). This is done by correlating the CT number of tissue substitute phantoms with known material composition with mass density or relative stopping power (RSP) via the stoichiometric calibration method (2). The accuracy of this approach relies on the difference between patient tissue chemical composition and the tissue substitute database used in the calibration (3–5). Since single-energy computed tomography (SECT) can't differentiate changes in CT number as a result of differences in either mass density or material chemical composition (6), the error in RSP calculation can become significant. The accuracy of mass density estimation dominates the uncertainty of RSP (7). Furthermore, tissue heterogeneity, CT image noise, and artifacts can also contribute to the RSP calculation error. The direct consequence of uncertainties associated with material characterization (mass density) from SECT data is a loss of accuracy in the prediction of energy deposition relative to the depth of proton interaction, also called proton range uncertainty. To mitigate range uncertainty, TPS have options to allow for the addition of margins in the proton beam range, the standard being 2.5%-3.5% of the energy-dependent range plus an additional 1mm-1.5mm (8).

To further increase the proton therapy therapeutic ratio advantage, many efforts have been made to decrease the proton range uncertainty. One approach is introducing Monte-Carlo dose calculation algorithms to proton TPS (8–12), which can reduce the margin down to 2.4% plus 1.2 mm (8, 13). Another proposed avenue was to use dual-energy CT (DECT) to build calibration curves between CT number and mass density (14). The methodology allows for the acquisition of CT scans with different X-ray spectra, which in turn can be used to determine relative electron density and mean excitation energy (15). Furthermore, DECT virtual monochromatic image reconstruction techniques can reduce beam hardening artifacts and noise. Numerous algorithms for DECT-based RSP estimation have been developed, and the reported results indicate that the errors in RSP estimates can be reduced to 1% (15–17).

As a robust implementation platform for DECT applications in the area of RSP mapping, machine learning (ML) algorithms have also been applied. Su et al. reported their approach for generating parametric maps using ML, which produced accurate effective atomic numbers, relative electron density, mean excitation energy, and RSP from DECT data (18), and they concluded that artificial neural network (ANN) outperformed other reported ML methods. Building on the potential advantage of increased network depth and compositionality (19, 20), deep learning (DL) is an extension of machine learning, which consists of massive multilayered networks or artificial neurons that can discover useful features in CT images (21). DL methods were successfully applied to improve mass density and RSP mapping from DECT datasets (22).

Despite all DECT-based ML and DL applications for improving proton dose calculation, SECT is still the current standard in clinical CT simulators for proton therapy. Therefore, an accurate and efficient method to reduce range uncertainty based on SECT images would benefit existing proton radiotherapy clinics and workflows. DL networks have high degrees of freedom of modeling, therefore offering the opportunity to improve the accuracy of SECT-based RSP and mass density modeling, as has been shown in DECT-based studies. In this study, we investigate the feasibility of using DL models to correlate SECT-based CT numbers

to voxel-specific mass density and RSP using two types of electron density phantoms.

# 2 Materials and methods

## 2.1 Phantom SECT data sets

A CIRS 062M electron density phantom (Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA) was used to generate the DL training dataset, while Gammex 467-1009 electron density phantom, CIRS ATOM M701 (male) and M702 (female) anthropomorphic phantoms were chosen to generate the DL networks prediction datasets (Figure 1). Table 1 details the mass density and RSP value for the phantoms used. All mass density information was provided by the manufacturer except the bone insert, for which measurements were utilized to produce a reference value (11). All RSP values were calculated using a previously reported method (22). Phantoms were imaged using a Siemens SOMATOM Definition Edge CT scanner and clinical 120 kVp single energy beam acquisition protocols. The electron density phantom was scanned using a standard head-and-neck protocol. A pelvis protocol was used for the Gammex electron density phantom, while the M701 and M702 phantoms were scanned using three different protocols: head-and-neck (HN), thorax, and pelvis protocols. The manufacturer-reported $CTDI_{vol}$ is reported in Table 2, as well as the reconstructed image resolution. All the reconstructed SECT images have a reconstructed field-of-view diameter of 500 mm and a slice thickness of 0.5 mm.

## 2.2 Deep learning models

Three supervised DL models were implemented to demonstrate the capability of artificial intelligence (AI) to improve proton range calculation using SECT (21, 23–25). Figure 2A shows the artificial neural network (ANN) workflow, and Figures 2B, C show the fully convolutional neural network (FCNN) and residual neural network (ResNet). 120 kVp spectra SECT images are used as DL input. The same DL models are utilized to estimate both the mass density and RSP relative to the input SECT CT number values. The DL models were supervised by a loss function, defined as the difference between the true value and predicted value at each voxel. All the DL models were implemented in PyTorch (26). Su et al. reported that ANN with 30 hidden units outperforms traditional ML models in generating quantitative parametric maps based on DECT images (18). Their ANN design was adopted for SECT parametric mapping in this study, with 30 hidden hyperbolic-tangent (tanh) layers and error backpropagation (see Figure 2A). Convolutional neural networks (CNNs) have gained widespread adoption in both regression and classification tasks over the past decade. This popularity is primarily due to their capability to autonomously learn deep, intricate features, a significant advancement over the traditional machine learning models that relied on manually extracted, handcrafted features (27).

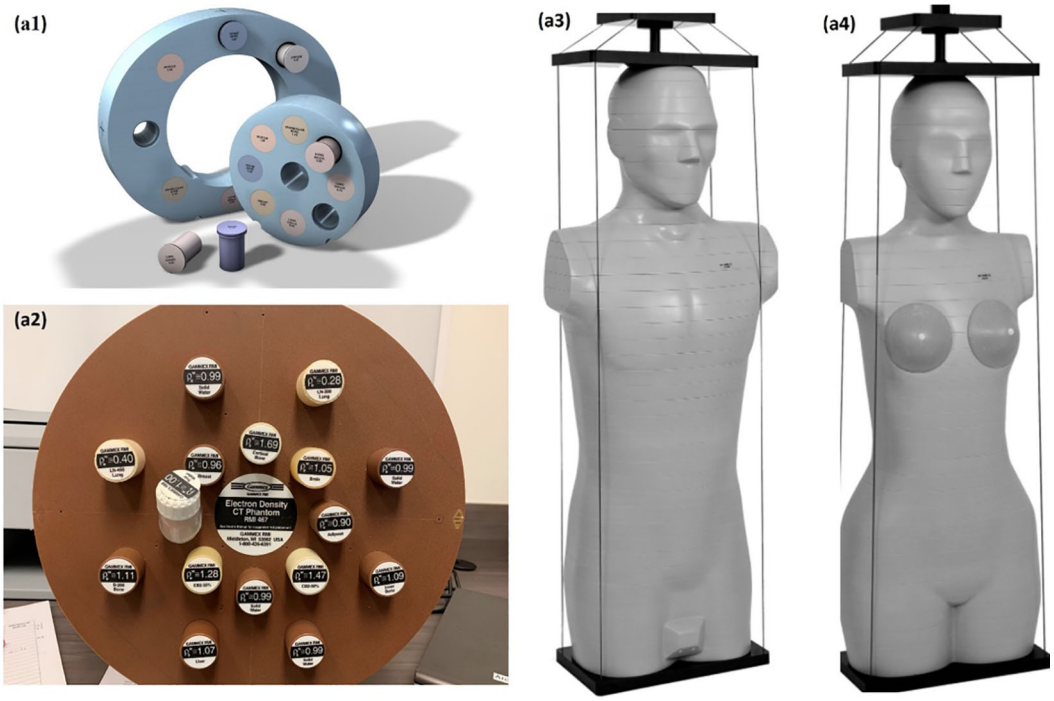



FIGURE 1
Experiment set up, CIRS 062M electron density phantom (Computerized Imaging Reference Systems) **(A1)**, Gammex 468 electron density phantom **(A2)** CIRS M701 **(A3)**, and CIRS M702 (Computerized Imaging Reference Systems) **(A4)** were scanning with Siemens SOMATOM Definition Edge scanner.

TABLE 1 Phantom insert data: mass densities and RSP.

| | Tissue surrogate | $\rho$ (g/cm$^3$) | RSP |
|---|---|---|---|
| CIRS 062M | Lung (Inhale) | 0.203 | 0.202 |
| | Lung (Exhale) | 0.494 | 0.492 |
| | Adipose | 0.965 | 0.977 |
| | Breast Tissue | 0.996 | 1.003 |
| | Muscle | 1.059 | 1.059 |
| | Liver | 1.072 | 1.070 |
| | Bone 200 mg/cc | 1.157 | 1.116 |
| | Bone 800 mg/cc | 1.520 | 1.404 |
| | Bone 1250 mg/cc | 1.830 | 1.647 |
| Gammex 467-1009 | LN450 Lung | 0.450 | 0.448 |
| | Breast | 0.980 | 0.969 |
| | Brain | 1.050 | 1.061 |
| | Liver | 1.090 | 1.090 |
| | B200 Bone Mineral | 1.150 | 1.099 |
| | CB2-50%$CaCO_3$ Bone | 1.560 | 1.422 |
| | SB3 Cortical Bone | 1.820 | 1.614 |
| CIRS M701 & M702 | Lung | 0.202 | 0.201 |
| | Breast | 0.991 | 0.982 |
| | Soft Tissue | 1.055 | 1.041 |
| | Spinal Cord | 1.065 | 1.035 |
| | Brain | 1.069 | 1.049 |
| | Bone | 1.517[1] | 1.410 |

[1]The bone mass density was measured.

A fully connected neural network (FCNN) is a type of CNN that has fully connected hidden layers. A 1D FCNN model was previously implemented to correlate the mass density and RSP map based on DECT parametric maps (22). This 1D FCNN model was adapted to the SECT images dataset input in this study, including seven hidden layers (Figure 2B).

Due to the small size of the training set used in this study, overfitting issues could be associated with trained DL model outputs, which were also considered as vanishing gradient problems for our CNN models. A solution for the latter was proposed: ResNet (25). The ResNet implementation in this study includes a shortcut connection added between input and output after a few weight layers: the residual block. Figure 2C shows the ResNet workflow and Figure 2D shows the detail of the residual block in the ResNet. In Figure 2C, the input is linked to a convolutional layer, which is subsequently connected to four distinct sequences of residual blocks. Each type of residual block varies in terms of channel, kernel, stride, and padding. These details are illustrated in Table 3.

TABLE 2 $CTDI_{vol}$ and voxel grid spacing information at each site with specified standard acquisition protocol.

| | $CTDI_{vol}$ (mGy)/Voxels | | |
|---|---|---|---|
| | HN | Thorax | Pelvis |
| CIRS 062m | 23.3/ 512×512×495 | | |
| Gammex 467 | | | 23.3/512×512×30 |
| CIRS Atom M701 | 23.6/ 512×512×605 | 23.3/ 512×512×609 | 23.3/ 512×512×471 |
| CIRS Atom M702 | 23.6/ 512×512×713 | 23.3/ 512×512×881 | 23.3/ 512×512×809 |

All the networks were trained using an NVIDIA GeForce 3090 GPU with 24 GB of memory, utilizing a batch size of 100. We set the learning rate for the Adam optimizer to 1e-5. During training, each batch optimization consumed 0.5 GB of CPU memory and 2 GB of GPU memory. All networks were implemented in PyTorch 2.1.

## 2.3 SECT stoichiometric method

A conventional physics-based SECT stoichiometric calibration method was utilized (2) to provide a standard set of reference values when comparing the DL models output for mass density and RSP. The X-ray linear attenuation coefficient μ of a material can be calculated using Equation (1):

$$\mu = \rho \frac{N_A}{A}[ZK^{KN}(E) + Z^n K^{SCA}(E) + Z^m K^{PE}(E)] \quad (1)$$

where $\rho$ is mass density, $N_A$ is Avogadro's number, A is the atomic weight, Z is the atomic number, $K^{KN}$, $K^{SCA}$, and $K^{PE}$ are weighting constants for incoherent scattering, coherent scattering, and photoelectric effect, respectively, as a function of the SECT scanner X-ray spectrum (1). m and n are constants and were assigned 4.62 and 2.86 (28) for the energies encountered in kV X-ray imaging and the elements present in human tissue. Taking the ratio of the attenuation coefficient of material of interest to the attenuation coefficient of water, equation (2) can be derived:

$$\frac{HU}{1000} + 1 = \hat{\rho}_{e,w} \frac{[1 + \hat{Z}^{1.86} k_1(E) + \hat{Z}^{3.62} k_2(E)]}{[1 + \hat{Z}_w{}^{1.86} k_1(E) + \hat{Z}_w{}^{3.62} k_2(E)]} \quad (2)$$

where $\hat{\rho}_{e,w}$ is the electron density relative to water, $\hat{Z}_w$ is the effective atomic number of waters, $k_1 = \frac{K^{SCA}}{K^{KN}}$, $k_2 = \frac{K^{PE}}{K^{KN}}$. $k_1$ and $k_2$ were calculated by nonlinear regression, while the CT number values were taken as the mean value of each electron density insert from the SECT scan. Finally, the RSP of each material of interest was calculated using the Bethe equation (29):

$$RSP = \hat{\rho}_{e,w} \ln\left[\frac{2m_e c^2 \beta^2}{I_m(1-\beta^2)} - \beta^2\right] / \ln\left[\frac{2m_e c^2 \beta^2}{I_{water}(1-\beta^2)} - \beta^2\right] \quad (3)$$

where $m_e$ is the electron rest mass, c is the speed of light in vacuum, and β is the proton speed relative to the speed of light. $I_m$ and $I_{water}$ are the mean excitation energies of the material of interest and water. As recommended by ICRU49, $I_{water}$ was set to 75 eV (30). $I_m$ can be calculated from the atomic components using the Bragg additivity rule. Therefore, the stoichiometric calibration method can establish a functional relationship between CT numbers from the SECT datasets and RSP by Equation (2) and (3) (11).

10.3389/fonc.2023.1278180

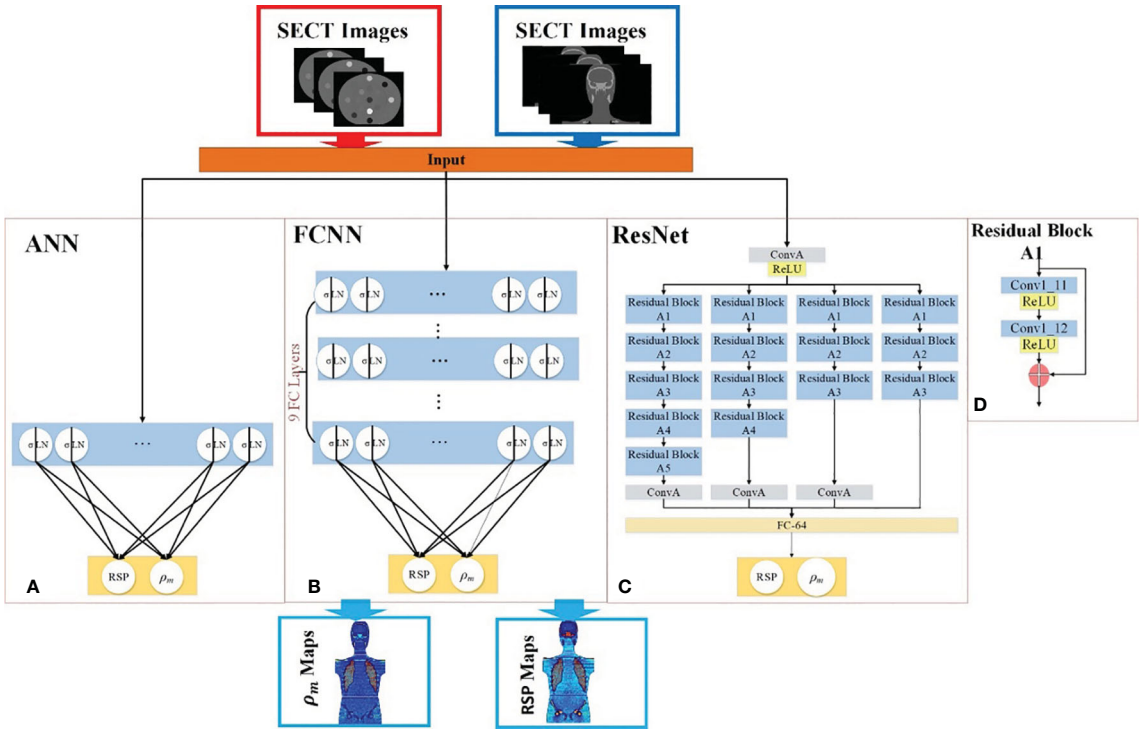



FIGURE 2
Deep learning framework for mass density and RSP map prediction based on SECT images. The red arrows illustrate the training workflows, while the blue arrows depict the prediction workflow. The framework includes **(A)** ANN, **(B)** FCNN, **(C)** ResNet, and **(D)** the residual block of ResNet. During the training phase (indicated by red arrows), the phantom SECT images are input into the three deep learning networks: ANN, FCNN, and ResNet. For the prediction phase, the M701 and M702 phantoms are fed into the trained deep learning networks to predict mass density and RSP.

TABLE 3 Architecture of the 1D convolution components of ResNet (RN).

| Network | Layer | Channel. | Kernel | Stride | Pad |
|---|---|---|---|---|---|
| | ConvA | 64 | 7 | 2 | 3 |
| Residual Block A1 | Conv1_11 | 64 | 3 | 1 | 1 |
| | Conv1_12 | 64 | 3 | 1 | 1 |
| | Residual | 64 | 1 | 2 | 0 |
| Residual Block A2 | Conv1_21 | 128 | 3 | 2 | 1 |
| | Conv1_22 | 128 | 3 | 1 | 1 |
| | Residual | 128 | 1 | 2 | 0 |
| Residual Block A3 | Conv1_31 | 256 | 3 | 2 | 1 |
| | Conv1_32 | 256 | 3 | 1 | 1 |
| | Residual | 256 | 1 | 2 | 0 |
| Residual Block A4 | Conv1_41 | 512 | 2 | 2 | 1 |
| | Conv2_42 | 512 | 2 | 1 | 1 |
| | Residual | 512 | 1 | 2 | 0 |
| Residual Block A5 | Conv1_51 | 1024 | 3 | 2 | 1 |
| | Conv2_52 | 1024 | 3 | 2 | 1 |
| | Residual | 1024 | 1 | 2 | 0 |

## 2.4 Empirical model based on DECT parametric mapping

In addition to the stoichiometric calibration standard, a second set of reference values was obtained to compare the DL models output in terms of mass density and RSP prediction based on SECT images against the manufacturer-defined ground truth. These were based on an empirical model for mass density and RSP based on DECT parametric mapping (22), and corresponding definitions are shown in Equation (4) and (5) (7, 31). Therein, $\rho$ denotes the mass density, and $\rho_e$ and $Z_{eff}$ are relative electron density and effective atomic number and were obtained from the DECT scanner console (Siemens Healthineers, syngo.via, Malvern, PA, USA). As used, the mass density model has a correction for the inflated lung. The RSP model includes corrections for different human tissues classified by various effective atomic numbers. The same phantoms were used in both DECT and SECT scans, and all tissue surrogate-defined contours were kept the same for reference comparison. Detailed information on DECT scans is listed in a previous publication (22).

$$\rho = \begin{cases} -0.1746 + 1.176\rho_e & material \neq inflated\ lung \\ 0.26 & material = inflated\ lung \end{cases} \quad (4)$$

$$RSP = \begin{cases} \rho_{e,} & 0 \le Z_{eff} < 0.5 \\ (1.1114 - 0.0148Z_{eff})\rho_e & 0.5 \le Z_{eff} < 8.5 \\ 0.9905\rho_{e,} & 8.5 \le Z_{eff} < 10 \\ (1.1117 - 0.0116Z_{eff})\rho_{e,} & Z_{eff} \ge 10 \end{cases} \quad (5)$$

## 2.5 Evaluation metrics

The ground truth dataset included the reference mass density and RSP values as shown in Table 1. The CIRS M701 and M702 phantom images were manually contoured in RayStation 9B (RaySearch Laboratories, Stockholm, Sweden) for each of the tissue surrogate inserts, and reference values were assigned to each contour. Absolute percentage error (APE) and mean absolute percentage error (MAPE) were calculated as defined in Equation (6) and (7), where $i$ denotes the $i^{th}$ voxel, x is the mass density or RSP at specific voxel, and N is the total number of voxels. The spatial distribution of error was also generated for the computed APE and displayed in error visualization maps of the whole phantom datasets.

$$APE_i = \left| \frac{x_i - x_{i,REF}}{x_{i,REF}} \right| \times 100\,\% \quad (6)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} APE_i \quad (7)$$

# 3 Results

## 3.1 Gammex electron density phantom analysis using pelvis protocol

Table 4 presents results comparing MAPE of mass densities between DL models and the SECT stoichiometric method over seven Gammex electron density phantom tissue surrogates. As emphasized by the bold values, the DL models perform better on all tissue surrogates. For the lung tissue surrogates (LN300 & LN450), the DL models outperform conventional methods, while for higher density inserts, the FCNN and ResNet perform better than the traditional method in select cases. As seen in Table 5, a similar trend can be observed for the DL RSP predictions.

## 3.2 HN site analysis using HN protocol

The mass densities and RSP predictions from the empirical model based on DECT parametric mapping, the SECT stoichiometric method, and the DL model results analysis are shown in Tables 6, 7. For all four tissues in M701, the DL methods are more accurate than the DECT empirical and SECT stoichiometric methods. The performance of FCNN and ResNet was comparable because of the similarity in the network fundamentals. For M702, the DECT empirical model outperforms the DL models in bone tissue prediction, which proves the advantage of DECT in bone tissue mass density estimation over SECT methods (ResNet has a comparable performance). However, this is not true for M701, the DL models based on SECT can achieve better results with the DECT empirical method in bone and other tissues. Table 7 summarizes the RSP MAPE comparison between the DECT empirical model, the stoichiometric method, and the DL models. For M701, DL methods outperform the traditional

TABLE 4 Mass Density – MAPE [SECT DL vs. SECT stoichiometry].

| Gammex Tissue Surrogates | LN450 Lung | Breast | Brain | Liver | B200 Bone Mineral | CB2-50% $CaCO_3$ Bone | SB3 Cortical Bone |
|---|---|---|---|---|---|---|---|
| SECT Stoichiometric | 12.57 ± 5.14 | 1.63 ± 1.36 | 2.23 ± 1.78 | 1.89 ± 1.03 | 1.18 ± 0.91 | 1.98 ± 1.12 | 2.32 ± 1.09 |
| ANN | 6.92 ± 4.81 | 1.95 ± 1.40 | **1.34 ± 1.04** | **1.18 ± 0.90** | 0.95 ± 0.74 | 6.22 ± 1.26 | 4.85 ± 1.28 |
| FCNN | **6.43 ± 0.16** | 1.18 ± 1.24 | 1.58 ± 1.18 | 1.69 ± 0.65 | **0.45 ± 0.27** | 1.36 ± 0.12 | 0.43 ± 0.30 |
| ResNet | 11.66 ± 0.27 | **1.17 ± 1.37** | 1.72 ± 1.11 | 1.66 ± 0.71 | 0.77 ± 0.23 | **0.85 ± 0.38** | **0.29 ± 0.09** |

Bold values mean best performance.

TABLE 5 RSP – MAPE [SECT DL vs. SECT stoichiometry].

| Gammex Tissue Surrogates | LN450 Lung | Breast | Brain | Liver | B200 Bone Mineral | CB2-50% $CaCO_3$ Bone | SB3 Cortical Bone |
|---|---|---|---|---|---|---|---|
| SECT Stoichiometric | 3.71 ± 2.89 | 3.72 ± 2.09 | 4.50 ± 2.38 | 2.47 ± 1.67 | 1.64 ± 1.61 | 5.33 ± 1.10 | 7.99 ± 1.13 |
| ANN | 4.72 ± 3.54 | 3.33 ± 1.66 | 1.59 ± 1.12 | **1.50 ± 0.92** | 1.87 ± 0.73 | 4.81 ± 1.18 | 2.39 ± 1.10 |
| FCNN | **2.18 ± 0.15** | **2.47 ± 1.47** | **1.16 ± 1.53** | 1.78 ± 0.59 | **1.39 ± 0.16** | **1.08 ± 0.11** | **1.96 ± 0.03** |
| ResNet | 2.26 ± 0.24 | 2.55 ± 1.44 | 1.25 ± 1.46 | 1.73 ± 0.35 | 1.60 ± 0.18 | 1.66 ± 0.71 | 2.69 ± 0.14 |

Bold values mean best performance.

methods, while for M702, the DECT empirical method outperforms the DL methods in the spinal cord.

## 3.3 Chest site analysis using thorax protocol

Table 8 summarizes the MAPE comparison of mass density estimated by empirical and DL models on the chest site using 120 kVp thorax protocol. For the lung site, the DECT empirical model doesn't apply to the normal lung tissue (the empirical model assigns the inflated lung as constant mass density), so the lung tissue mass density prediction is not reported. Table 9 shows the MAPE comparisons of RSP between the reference and DL models. All three DL models outperform the two reference models. The DECT empirical model and the stoichiometric method show better performance than ANN on lung tissue RSP prediction.

## 3.4 Pelvis site analysis using pelvis protocol

Table 10 shows the MAPE values of five models' mass density predictions at the CIRS M701 and M702 phantom pelvis sites. DL methods show better performance in all three tissues and can reduce the error to<1% in the spinal cord and soft tissue. Table 11 shows the MAPE values of three DL models and the two reference models' RSP estimations at the CIRS M701 and M702 phantom pelvis sites. The DL methods have better performance in soft tissue and bone, and all models have comparable results in the spinal cord.

## 3.5 Whole phantom site analysis using HN, thorax, and pelvis protocols

Table 12 shows the MAPE comparisons of mass densities between DL models and the SECT stoichiometric method over the entire phantom site. Table 13 summarizes the MAPE comparisons of RSP between DL models and the SECT stoichiometric method over the entire phantom site.

Figure 3 illustrates the APE maps of the mass density estimation error. Figures 3A17–C1 shows the SECT images of CIRS M701 phantom at three sites, HN, thorax, and pelvis, using 120 kVp corresponding scanning protocols. As presented using the APE color maps, the SECT stoichiometric model results in considerable uncertainty at lung and bone sites. Overall, the analysis for FCNN indicates the lowest error in the mass density estimation compared with other DL models. Figure 4 illustrates the APE maps for RSP error estimation. As with mass density, the SECT stoichiometric method shows considerable error in RSP estimation, especially in bone and lung tissue. ANN improves the estimation accuracy for bone and soft tissue, while FCNN predictions indicate

TABLE 6 Mass Density – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Brain | Spinal Cord | Soft Tissue | Bone |
|---|---|---|---|---|
| DECT empirical model | 1.00 ± 0.80 | 1.40 ± 1.30 | 2.30 ± 2.20 | 2.50 ± 5.20 |
| SECT Stoichiometric method | 0.99 ± 0.66 | 1.53 ± 1.59 | 1.59 ± 1.72 | 4.79 ± 2.20 |
| ANN | 0.64 ± 0.50 | 0.96 ± 1.20 | 1.03 ± 1.08 | 1.95 ± 2.44 |
| FCNN | 0.39 ± 0.31 | **0.59 ± 0.64** | **0.62 ± 0.52** | 1.03 ± 1.30 |
| ResNet | **0.15 ± 0.20** | 0.74 ± 1.20 | 1.46 ± 1.03 | **0.66 ± 0.85** |
| M702 | Brain | Spinal Cord | Soft Tissue | Bone |
| DECT Empirical model | 1.30 ± 0.08 | 1.20 ± 1.00 | 2.20 ± 1.50 | **2.00 ± 1.90** |
| SECT Stoichiometric method | 1.10 ± 0.97 | 1.80 ± 2.01 | 1.64 ± 2.12 | 6.80 ± 4.76 |
| ANN | 0.86 ± 0.81 | 1.28 ± 1.68 | 1.07 ± 1.70 | 4.84 ± 7.00 |
| FCNN | 0.89 ± 0.57 | 0.95 ± 1.20 | **0.57 ± 1.15** | 3.74 ± 6.78 |
| ResNet | **0.19 ± 0.89** | **0.93 ± 2.18** | 1.38 ± 2.44 | 2.16 ± 5.18 |

Bold values mean best performance.

TABLE 7 RSP – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Brain | Spinal Cord | Soft Tissue | Bone |
|---|---|---|---|---|
| DECT Empirical model | 1.20 ± 0.90 | 2.70 ± 1.90 | 2.90 ± 2.60 | 4.20 ± 4.90 |
| SECT Stoichiometric method | 0.95 ± 0.96 | 2.06 ± 1.53 | 2.96 ± 1.64 | 10.22 ± 3.00 |
| ANN | 1.15 ± 0.51 | 2.42 ± 0.85 | 1.11 ± 0.79 | 1.79 ± 1.68 |
| FCNN | **0.85 ± 0.37** | **2.11 ± 0.64** | **1.08 ± 0.51** | 0.67 ± 1.28 |
| ResNet | 1.85 ± 0.20 | 3.14 ± 0.83 | 2.24 ± 1.31 | **0.48 ± 1.18** |
| M702 | Brain | Spinal Cord | Soft Tissue | Bone |
| DECT Empirical model | 0.90 ± 0.80 | **1.60 ± 1.40** | 2.60 ± 2.40 | 3.90 ± 2.20 |
| SECT Stoichiometric method | 1.24 ± 1.32 | 2.19 ± 1.84 | 3.20 ± 2.03 | 10.65 ± 4.47 |
| ANN | 0.99 ± 0.69 | 2.39 ± 1.44 | 1.03 ± 1.49 | 4.35 ± 5.57 |
| FCNN | **0.45 ± 0.50** | 1.80 ± 1.22 | **0.74 ± 1.09** | 3.00 ± 5.67 |
| ResNet | 1.82 ± 0.62 | 3.20 ± 1.56 | 2.18 ± 1.22 | **1.67 ± 4.61** |

Bold values mean best performance.

improvements overall. ResNet shows smaller APE for bone, specifically for the pelvis scan, even when compared with FCNN.

## 4 Discussions

External beam radiotherapy requires an accurate CT characterization of the patient geometry and heterogeneities to deliver accurate therapeutic patient doses. SECT imaging is the current clinical paradigm for generating the necessary information for clinical diagnosis and treatment planning in radiotherapy, including protons. Su et al. demonstrated the capability of machine learning methods to improve the prediction accuracy for mass density and RSP based on DECT imaging (18). In order to approach this methodology while using the more commonly available SECT imaging, we proposed a framework that can accurately predict mass density and RSP parametric maps for proton dose calculation while using SECT imaging as input. All SECT DL methods for mass density and RSP estimation were compared against the ground truth values defined in Table 1 and compared to current standards for mass density and RSP parametrization (DECT and stoichiometry). All parametrization was done on phantom data since the DL network training requires accurate ground truth definition, which excludes patient data. The DECT model MAPE was consistently larger than that of the SECT stoichiometric method for some of the tissues (specifically for the spinal cord). This could be attributed to the fact that the DECT empirical model depends on the relative electron density for mass density prediction and effective atomic number for RSP, and it has not been calibrated for the specific CT scanner used in this work. The SECT DL models demonstrated good or better performance for parametrization based on three different clinical scanning protocols

TABLE 8 Mass density – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Spinal Cord | Soft Tissue | Bone | Lung |
|---|---|---|---|---|
| DECT Empirical model | 3.40 ± 2.60 | 3.10 ± 2.50 | 3.40 ± 3.30 | -[1] |
| SECT Stoichiometric method | 1.53 ± 1.59 | 1.59 ± 1.72 | 3.68 ± 2.17 | 8.29 ± 7.34 |
| ANN | 0.96 ± 1.20 | 1.03 ± 1.08 | 3.64 ± 3.17 | 12.96 ± 7.09 |
| FCNN | 0.59 ± 0.64 | **0.64 ± 1.13** | 1.31 ± 2.36 | **1.57 ± 0.39** |
| ResNet | **0.58 ± 0.61** | 1.46 ± 1.56 | **1.28 ± 1.43** | 2.18 ± 0.73 |
| M702 | Spinal Cord | Soft Tissue | Bone | Lung |
| DECT Empirical model | 2.50 ± 1.90 | 2.70 ± 2.20 | 4.00 ± 6.80 | -[1] |
| SECT Stoichiometric method | 1.72 ± 1.60 | 1.98 ± 4.10 | 5.77 ± 4.80 | 6.91 ± 6.35 |
| ANN | 1.32 ± 3.70 | 1.32 ± 3.70 | 4.55 ± 6.50 | 14.86 ± 7.37 |
| FCNN | 0.89 ± 0.71 | **0.81 ± 3.65** | 3.41 ± 6.54 | **1.52 ± 0.49** |
| ResNet | **0.64 ± 0.83** | 1.53 ± 3.87 | **1.69 ± 4.87** | 2.42 ± 1.97 |

[1]The DECT empirical model doesn't apply to normal lung tissue.
Bold values mean best performance.

TABLE 9 RSP – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Spinal Cord | Soft Tissue | Bone | Lung |
|---|---|---|---|---|
| DECT Empirical model | 2.70 ± 2.60 | 3.20 ± 2.50 | 5.20 ± 3.60 | 7.00 ± 5.80 |
| SECT Stoichiometric method | 2.93 ± 2.35 | 3.06 ± 1.87 | 6.56 ± 2.94 | 8.29 ± 7.34 |
| ANN | 2.86 ± 1.36 | 1.07 ± 1.03 | 3.85 ± 2.88 | 12.36 ± 6.99 |
| FCNN | **2.37 ± 0.95** | **1.04 ± 0.94** | 1.73 ± 2.07 | 1.57 ± 0.43 |
| ResNet | 3.20 ± 0.75 | 2.14 ± 1.22 | **0.72 ± 1.09** | **0.95 ± 0.72** |
| M702 | Spinal Cord | Soft Tissue | Bone | Lung |
| DECT Empirical model | 3.30 ± 3.10 | 2.70 ± 2.20 | 4.70 ± 6.20 | 6.70 ± 5.50 |
| SECT Stoichiometric method | 2.53 ± 1.95 | 3.49 ± 3.86 | 9.23 ± 4.25 | 6.75 ± 6.35 |
| ANN | 2.03 ± 1.56 | 1.32 ± 3.70 | 4.31 ± 5.60 | 14.23 ± 7.37 |
| FCNN | **1.54 ± 1.30** | **0.89 ± 3.21** | 2.94 ± 5.56 | 1.30 ± 0.54 |
| ResNet | 2.90 ± 1.64 | 2.27 ± 3.31 | **1.36 ± 4.28** | **1.10 ± 0.70** |

Bold values mean best performance.

(HN, Chest, Pelvic). Note that DL models were trained with phantom data, while the SECT stoichiometric method and DECT empirical data were optimized for clinical use with both phantom and human tissue, which might lead to worse performance on anthropomorphic phantom mass density and RSP estimation.

Gomà et al. concluded that the Gammex phantom utilized in this study contains tissue substitutes better representative of the human body than the CIRS electron density phantom (17). As shown in Tables 4, 5, we tested our DL models with Gammex 467 electron density phantom tissue surrogates, and the results indicate that the DL models can improve the mass density and RSP estimation relative to the stoichiometric method. Note that we tested our DL models on Gammex 467-1009 electron density phantom to evaluate the performance using different materials; then, we tested our DL models with CIRS M701&M702 phantoms for different materials and patient body sizes.

As shown in Figures 3, 4, the SECT stoichiometric method yields larger error in the lung, skull, pelvic bone, and soft tissue than DL networks. Also, the mass density and RSP maps generated with the SECT stoichiometric method have more noise than those generated by DL models, which indicates that DL models can handle noise and artifact suppression superiorly. It was reported that the uncertainty associated with mass density estimation dominates the proton range calculation uncertainty (7, 15, 32). In this work, DL models were trained with SECT images of electron density phantom, and Table 12 shows that the mass density MAPE can be improved significantly relative to the SECT stoichiometric method. The FCNN and ResNet outperform the ANN and reference models, and their MAPE values are

TABLE 10 Mass density – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Spinal Cord | Soft Tissue | Bone |
|---|---|---|---|
| DECT Empirical model | 3.30 ± 2.70 | 2.50 ± 2.00 | 2.40 ± 1.80 |
| SECT Stoichiometric method | 1.91 ± 1.62 | 2.07 ± 1.74 | 2.18 ± 2.00 |
| ANN | 1.05 ± 0.89 | 1.25 ± 1.03 | 3.15 ± 2.04 |
| FCNN | 0.73 ± 0.61 | **0.75 ± 0.60** | **0.73 ± 1.12** |
| Resnet | **0.51 ± 0.43** | 1.54 ± 1.05 | 1.19 ± 1.11 |
| M702 | Spinal Cord | Soft Tissue | Bone |
| DECT Empirical model | 2.10 ± 1.80 | 2.30 ± 1.90 | 2.20 ± 1.70 |
| SECT Stoichiometric method | 1.82 ± 1.47 | 1.92 ± 1.74 | 3.76 ± 4.60 |
| ANN | 1.03 ± 0.88 | 1.20 ± 1.06 | 4.57 ± 5.09 |
| FCNN | 0.67 ± 0.57 | **0.66 ± 0.54** | 2.61 ± 5.62 |
| Resnet | **0.53 ± 0.63** | 1.41 ± 1.03 | **1.79 ± 4.44** |

Bold values mean best performance.

TABLE 11 RSP – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Spinal Cord | Soft Tissue | Bone |
|---|---|---|---|
| DECT Empirical model | 2.80 ± 2.20 | 2.60 ± 2.10 | 3.50 ± 2.30 |
| SECT Stoichiometric method | **1.93 ± 1.62** | 2.88 ± 1.64 | 6.56 ± 1.71 |
| ANN | 2.40 ± 1.45 | 1.25 ± 0.5 | 3.47 ± 1.91 |
| FCNN | 2.08 ± 1.18 | **1.13 ± 0.64** | 1.33 ± 1.12 |
| Resnet | 2.63 ± 1.31 | 2.22 ± 0.86 | **0.61 ± 0.81** |
| M702 | Spinal Cord | Soft Tissue | Bone |
| DECT Empirical model | 2.50 ± 1.90 | 2.40 ± 1.90 | 3.40 ± 2.10 |
| SECT Stoichiometric method | 1.90 ± 1.23 | 3.06 ± 2.80 | 7.21 ± 3.67 |
| ANN | 2.20 ± 1.27 | 1.05 ± 0.85 | 4.69 ± 4.63 |
| FCNN | **1.62 ± 0.92** | **0.74 ± 0.54** | 2.75 ± 4.75 |
| Resnet | 2.70 ± 1.14 | 2.08 ± 0.86 | **1.27 ± 3.77** |

Bold values mean best performance.

TABLE 12 Mass density – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Brain | Spinal Cord | Soft Tissue | Bone | Lung |
|---|---|---|---|---|---|
| DECT Empirical model | 1.00 ± 0.80 | 2.60 ± 2.40 | 2.70 ± 2.30 | 3.00 ± 3.80 | -[1] |
| SECT Stoichiometric method | 0.99 ± 0.66 | 2.34 ± 2.14 | 1.87 ± 1.86 | 3.17 ± 2.26 | 8.29 ± 7.34 |
| ANN | 0.64 ± 0.52 | 1.46 ± 1.64 | 1.29 ± 1.19 | 2.90 ± 2.36 | 12.96 ± 7.09 |
| FCNN | 0.39 ± 0.31 | **0.92 ± 0.88** | **0.68 ± 0.86** | **0.92 ± 1.65** | **1.57 ± 0.39** |
| ResNet | **0.13 ± 0.20** | 0.97 ± 1.54 | 1.42 ± 1.29 | 1.04 ± 1.45 | 2.18 ± 0.73 |
| M702 | Brain | Spinal Cord | Soft Tissue | Bone | Lung |
| DECT Empirical model | 1.30 ± 0.90 | 1.90 ± 1.70 | 2.40 ± 2.00 | 2.50 ± 3.70 | -[1] |
| SECT Stoichiometric method | 1.10 ± 0.97 | 2.15 ± 1.52 | 1.92 ± 2.79 | 5.29 ± 4.89 | 6.91 ± 6.35 |
| ANN | 0.86 ± 0.81 | 1.53 ± 1.86 | 1.23 ± 2.35 | 4.65 ± 5.73 | 14.86 ± 7.37 |
| FCNN | 0.89 ± 0.57 | 1.09 ± 1.27 | **0.70 ± 2.16** | 3.19 ± 6.24 | **1.52 ± 0.49** |
| ResNet | **0.19 ± 0.89** | **1.08 ± 2.29** | 1.30 ± 2.48 | **1.87 ± 4.82** | 2.42 ± 1.97 |

[1]The DECT empirical model doesn't apply to normal lung tissue.
Bold values mean best performance.

close. A possible reason is that they share the same core network feature, the convolutional layer. Considering the training cost, the FCNN is recommended as the DL model for future dose calculation study. The RSP uncertainty is considerably larger than that of mass density, as demonstrated in Table 13, the DL models still exhibit the potential to enhance the accuracy of RSP estimation. FCNN could reduce the MAPE down to less than 2%, except in the spinal cord of CIRS M701.

Figure 5 illustrates the mass density prediction map generated by FCNN, ResNet, and the SECT stoichiometric method for two patients using HN and pelvis scans. As shown in Figures 5A3, A4, B3, B4, the DL models can reflect the patient's anatomy qualitatively, compared with the SECT images in Figure 5A1. Figures 6A–C shows the comparison of CT number profile and density profile at the position marked in Figures 5A1, B1. Figures 6A–C demonstrates that the mass density profile generated by FCNN and ResNet has a strong agreement with the SECT number profile trend. As shown in Figures 5A2–A4, B2–B4, the mass density maps predicted by DL models have less noise than that predicted by the stoichiometric method, especially in the scan of the soft tissue of the pelvis. This implies that the DL models can suppress CT noises and artifacts, and it is also shown in the mass density profile comparison in Figure 6 that the DL models' mass density lines are smoother than that of the stoichiometric method. The empirical model can provide better contrast information between adipose and bone tissue than the DL models; this might be because the DL models predict a higher density (~0.96 g/cm$^3$) for adipose compared to the stoichiometric method (~0.88 g/cm$^3$). The accurate density of the patient's adipose tissue is not known; DL models estimated it at 8% higher than the stoichiometric method because the training dataset of the DL

TABLE 13 RSP – MAPE [SECT DL vs. SECT stoichiometry vs. DECT model].

| M701 | Brain | Spinal Cord | Soft Tissue | Bone | Lung |
|---|---|---|---|---|---|
| DECT Empirical model | 1.20 ± 0.90 | 4.40 ± 3.50 | 2.09 ± 2.40 | 3.20 ± 3.70 | -[1] |
| SECT Stoichiometric method | 0.95 ± 0.96 | 2.61 ± 2.10 | 2.08 ± 1.90 | 7.74 ± 2.78 | 8.62 ± 7.43 |
| ANN | 1.15 ± 0.51 | 2.74 ± 1.22 | 1.21 ± 1.17 | 3.03 ± 2.07 | 12.52 ± 7.39 |
| FCNN | **0.85 ± 0.37** | **2.32 ± 0.87** | **0.75 ± 0.88** | **1.22 ± 1.34** | 1.25 ± 2.37 |
| ResNet | 1.85 ± 0.20 | 2.73 ± 0.83 | 1.57 ± 1.31 | 1.26 ± 1.02 | **1.01 ± 2.45** |
| M702 | Brain | Spinal Cord | Soft Tissue | Bone | Lung |
| DECT Empirical model | 0.90 ± 0.80 | 3.60 ± 3.00 | 2.50 ± 2.10 | 2.70 ± 3.50 | -[1] |
| SECT Stoichiometric method | 1.24 ± 1.32 | 2.44 ± 2.05 | 2.13 ± 2.81 | 8.87 ± 3.76 | 6.75 ± 1.36 |
| ANN | 0.99 ± 0.69 | 2.28 ± 1.60 | 1.23 ± 2.32 | 4.47 ± 4.65 | 14.23 ± 7.28 |
| FCNN | **0.45 ± 0.50** | **1.71 ± 1.30** | **0.72 ± 2.16** | 2.88 ± 4.70 | 1.30 ± 0.54 |
| ResNet | 1.82 ± 0.62 | 2.07 ± 1.65 | 1.56 ± 2.40 | **1.42 ± 3.73** | **1.10 ± 0.70** |

[1]The DECT empirical model doesn't apply to normal lung tissue.
Bold values mean best performance.

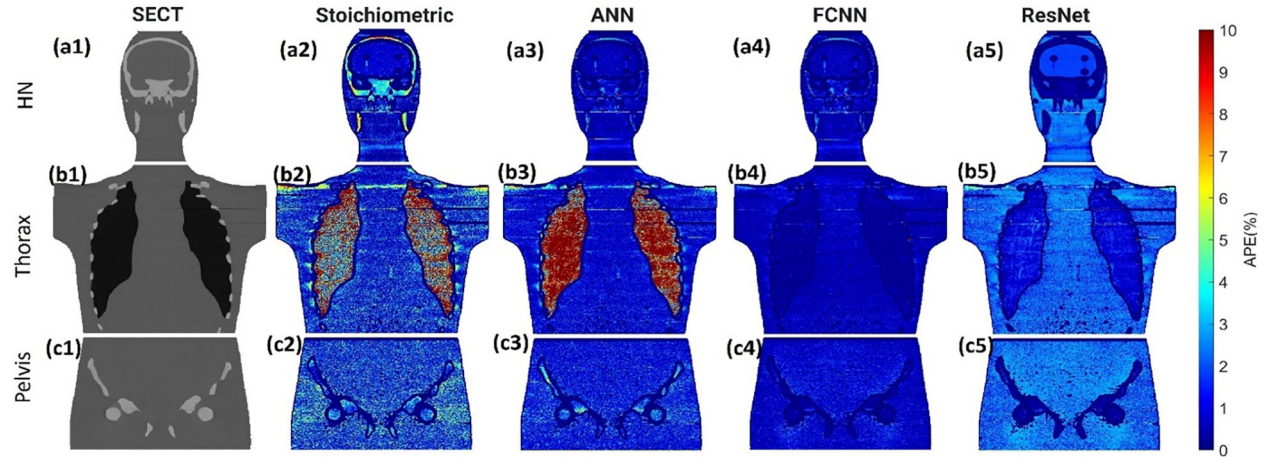



FIGURE 3
**(A1, B1, C1)** SECT images scan using different 120 kVp protocols. APE maps of mass densities between the reference and SECT parametric models at **(A2–A5)** HN, **(B2–B5)** thoracic, and **(C2–C5)** pelvic sites using CIRS M701 phantom.

models doesn't cover the mass density range from 0.5 to 0.95 g/cm$^3$, while the clinically used SECT stoichiometric method CT curve has been calibrated with over 30 materials on various mass densities. If more tissue surrogates can be adapted into the training set of this framework, the robustness of DL models will be significantly improved. As shown in Figures 6A, C, FCNN has better performance in reproducing minor structure information from SECT images, such as the mass density difference in pelvic bone and bone marrow. Because ResNet has a complex model structure, including deep CNN layers, for example, when the dimension of data is insufficient (only SECT image as input), the complex model does not necessarily lead to a robust result (33).

In supervised machine learning, a prevalent challenge is that DL models do not consistently generalize effectively from the training input data to new, unseen data (34). This overfitting phenomenon allows DL models to perform exemplarily in controlled conditions (similar or identical to the training conditions) while performing poorly on the application set. Many reasons can lead to this phenomenon, a lack of training dataset diversity being the predominant one. Only SECT phantom images were adapted into the training set; therefore, the DL networks' infrastructure needed to be designed carefully to avoid overfitting. As shown in Figures 5, 6, the mass density profile estimated by DL models can basically follow the trend of SECT CT number and overlap with the profile of the SECT stoichiometric method. As shown in Table 14, DL models predicted a patient tissue mass density similar to that predicted using traditional methods. Integrating these findings, the design of the DL network successfully addresses and mitigates the overfitting issue.

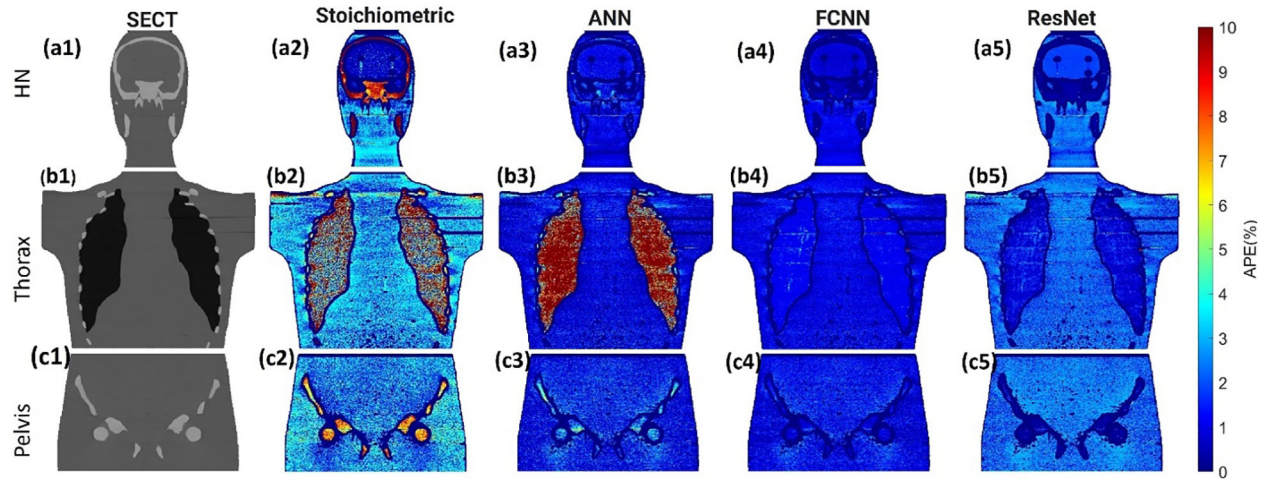



FIGURE 4
**(A1, B1–C1)** SECT images scan using different 120 kVp protocols. APE maps of RSP between the reference and SECT parametric models at **(A2–A5)** HN, **(B2–B5)** thoracic, and **(C2–C5)** pelvic sites using CIRS M701 phantom.

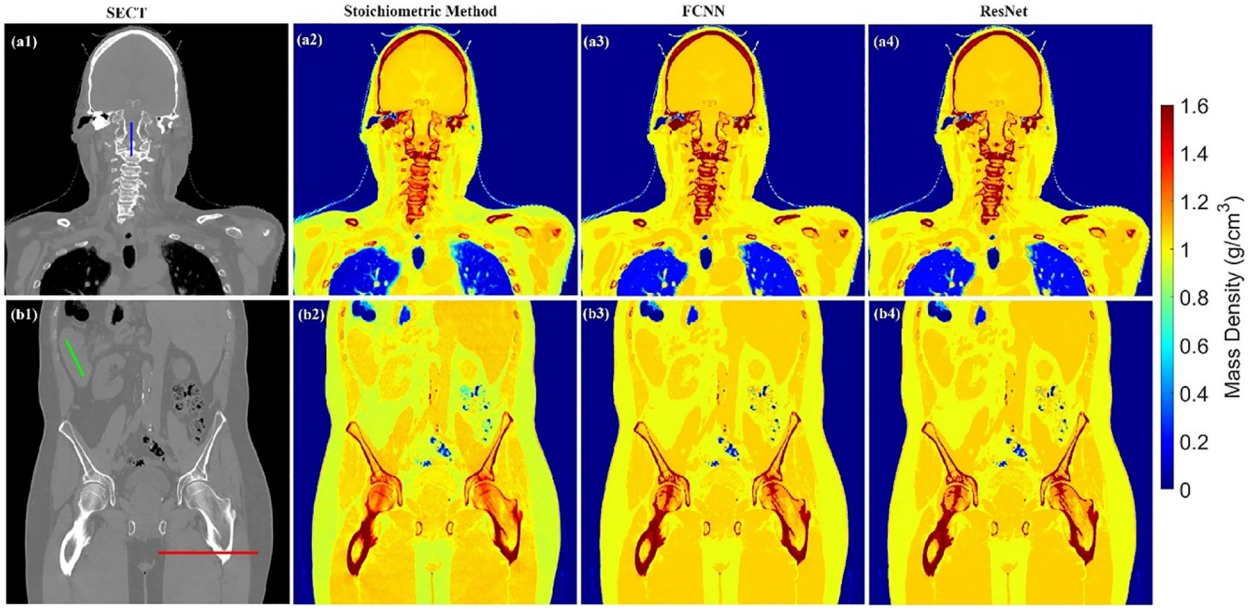



FIGURE 5
Mass density maps of two patients generated by different SECT FCNN and ResNet models at **A1–A4** and **B1–B4** using HN and pelvis scan.

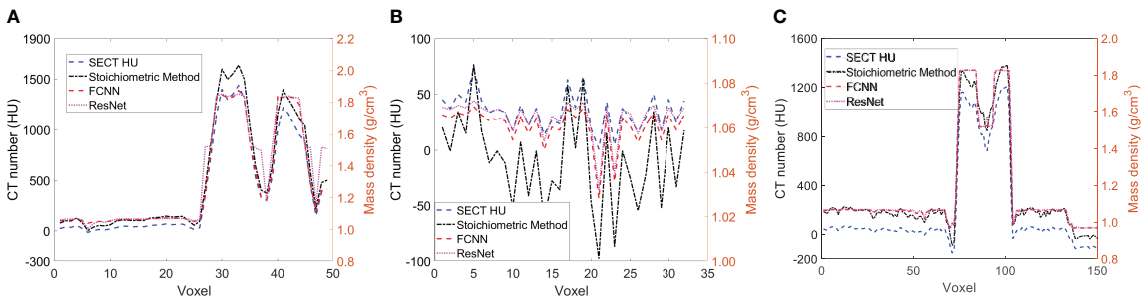



FIGURE 6
**(A)** the line profile from the blue line in Figure 5A1, **(B)** the line profile from green line in Figure 5B1, and **(C)** the line profile from the red line in Figure 5B1.

TABLE 14 Average mass density of three tissues comparison among DL models and the Stoichiometric model based on one patient SECT scan.

| Patient | Brain ($g/cm^3$) | Bone ($g/cm^3$) | Lung ($g/cm^3$) |
|---|---|---|---|
| SECT Stoichiometric method | 1.04 ± 0.04 | 1.69 ± 0.18 | 0.20 ± 0.05 |
| ANN | 1.04 ± 0.02 | 1.62 ± 0.17 | 0.17 ± 0.04 |
| FCNN | 1.05 ± 0.01 | 1.66 ± 0.17 | 0.20 ± 0.01 |
| ResNet | 1.05 ± 0.03 | 1.68 ± 0.15 | 0.21 ± 0.01 |

## 5 Conclusion

A DL framework was proposed to improve the mass density and RSP parametrization for proton dose calculation based on the SECT images. All three DL models outperform the SECT stoichiometric method in tissue substitute except the lung surrogate. FCNN and ResNet improved the mass density and RSP estimation accuracy based on SECT images, outperforming the SECT stoichiometric method over the entire phantom, and outperforming the DECT empirical model. DL models also demonstrate the ability to suppress CT image noise. The proposed DL frameworks have the potential to improve the clinical proton dose calculation based on SECT scans.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The dataset can be available by contact the corresponding author. Requests to access these datasets should be directed to xyang43@emory.edu.

## Ethics statement

The studies involving humans were approved by Emory University review board (IRB#114349). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because this is a retrospective study, and the patient's information was anonymized.

## Author contributions

YG: Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. C-WC: Investigation, Software, Validation, Writing – review & editing. JR: Writing – review & editing. MA: Writing – review & editing. YL: Methodology, Software, Writing – review & editing. SP: Methodology, Writing – review & editing. JB: Supervision, Writing – review & editing. JZ: Investigation, Methodology, Writing – review & editing. TL: Writing – review & editing. XY: Conceptualization, Funding acquisition, Investigation, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported in part by the National Institutes of Health under Award Number R01CA215718, R56EB033332, and R01EB032680.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Zhu J, Penfold SN. Dosimetric comparison of stopping power calibration with dual-energy CT and single-energy CT in proton therapy treatment planning. *Med Phys* (2016) 43:2845–54. doi: 10.1118/1.4948683

2. Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys Med Biol* (1996) 41:111–24. doi: 10.1088/0031-9155/41/1/009

3. Woodard HQ, White DR. The composition of body tissues. *Br J Radiol* (1986) 59:1209–18. doi: 10.1259/0007-1285-59-708-1209

4. International Commission On Radiation, U. & Measurements B. M. D. *Tissue substitutes in radiation dosimetry and measurement*. (Bethesda, MD (United States): International Commission on Radiation Units and Measurements) (1989). Available at: https://www.osti.gov/biblio/10102048.

5. Scott JA. Photon, electron, proton and neutron interaction data for body tissues ICRU report 46. International commission on radiation units and measurements, Bethesda 1992, $40.00. *J Nucl Med* (1993) 34:171–1.

6. McCollough CH, Boedeker K, Cody D, Duan X, Flohr T, Halliburton SS, et al. Principles and applications of multienergy CT: Report of AAPM Task Group 291. *Med Phys* (2020) 47:4. doi: 10.1002/mp.14157

7. Hunemohr N, Krauss B, Tremmel C, Ackermann B, Jakel O, Greilich S. Experimental verification of ion stopping power prediction from dual energy CT data in tissue surrogates. *Phys Med Biol* (2014) 59:83–96. doi: 10.1088/0031-9155/59/1/83

8. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol* (2012) 57:R99–117. doi: 10.1088/0031-9155/57/11/R99

9. Paganetti H, Jiang H, Parodi K, Slopsema R, Engelsman M. Clinical implementation of full Monte Carlo dose calculation in proton beam therapy. *Phys Med Biol* (2008) 53:4825–53. doi: 10.1088/0031-9155/53/17/023

10. Taasti VT, Bäumer C, Dahlgren CV, Deisher AJ, Ellerbrock M, Free J, et al. Inter-centre variability of CT-based stopping-power prediction in particle therapy: Survey-based evaluation. *Phys Imaging Radiat Oncol* (2018) 6:25–30. doi: 10.1016/j.phro.2018.04.006

11. Chang CW, Huang S, Harms J, Zhou J, Zhang R, Dhabaan A, et al. A standardized commissioning framework of Monte Carlo dose calculation algorithms for proton pencil beam scanning treatment planning systems. *Med Phys* (2020) 47:1545–57. doi: 10.1002/mp.14021

12. Peters N, Wohlfahrt P, Dahlgren CV, De Marzi L, Ellerbrock M, Fracchiolla F, et al. Experimental assessment of inter-centre variation in stopping-power and range prediction in particle therapy. *Radiother Oncol* (2021) 163:7–13. doi: 10.1016/j.radonc.2021.07.019

13. Paganetti H. Monte Carlo simulations will change the way we treat patients with proton beams today. *Br J Radiol* (2014) 87:20140293. doi: 10.1259/bjr.20140293

14. Bourque AE, Carrier J-F, Bouchard H. A stoichiometric calibration method for dual energy computed tomography. *Phys Med Biol* (2014) 59:2059–88. doi: 10.1088/0031-9155/59/8/2059

15. Yang M, Virshup G, Clayton J, Zhu XR, Mohan R, Dong L. Theoretical variance analysis of single- and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues. *Phys Med Biol* (2010) 55:1343–62. doi: 10.1088/0031-9155/55/5/006

16. Wohlfahrt P, Mohler C, Hietschold V, Menkel S, Greilich S, Krause M, et al. Clinical implementation of dual-energy CT for proton treatment planning on pseudo-monoenergetic CT scans. *Int J Radiat Oncol Biol Phys* (2017) 97:427–34. doi: 10.1016/j.ijrobp.2016.10.022

17. Gomà C, Almeida IP, Verhaegen F. Revisiting the single-energy CT calibration for proton therapy treatment planning: a critical look at the stoichiometric method. *Phys Med Biol* (2018) 63:235011. doi: 10.1088/1361-6560/aaede5

18. Su KH, Kuo JW, Jordan DW, Van Hedent S, Klahr P, Wei Z, et al. Machine learning-based dual-energy CT parametric mapping. *Phys Med Biol* (2018) 63:125001. doi: 10.1088/1361-6560/aac711

19. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks* (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003

20. 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–e36. doi: 10.1002/mp.13264

21. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539

22. Chang C-W, Gao Y, Wang T, Lei Y, Wang Q, Pan S, et al. Dual-energy CT based mass density and relative stopping power estimation for proton therapy using physics-informed deep learning. *Phys Med Biol* (2022) 67(11):115010. doi: 10.1088/1361-6560/ac6ebc

23. Hassoun MH. *Fundamentals of artificial neural networks*. Cambridge, Massachusetts, United States: MIT press (1995).

24. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* (2012) 25.

25. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. (Springer International Publishing) (2016), 630–45.

26. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* (2019) 32.

27. Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. *Mechanical Syst Signal Process* (2021) 151:107398. doi: 10.1016/j.ymssp.2020.107398

28. Rutherford RA, Pullan BR, Isherwood I. Measurement of effective atomic number and electron density using an EMI scanner. *Neuroradiology* (1976) 11:15–21. doi: 10.1007/BF00327253

29. Bichsel H. *Passage of charged particles through matter*. California Univ., Berkeley (USA), Dept. of Physics: Office of Scientific and Technical Information (OSTI) (1969).

30. Deasy J. ICRU report 49, stopping powers and ranges for protons and alph particles. *Med Phys* (1994) 21:709–10. doi: 10.1118/1.597176

31. Beaulieu L, Tedgren Å.C, Carrier J-F, Davis SD, Mourtada F, Rivard MJ, et al. Report of the Task Group 186 on model-based dose calculation methods in brachytherapy beyond the TG-43 formalism: Current status and recommendations for clinical implementation. *Med Phys* (2012) 39:6208–36. doi: 10.1118/1.4747264

32. Yang M, Zhu XR, Park PC, Titt U, Mohan R, Virshup G, et al. Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. *Phys Med Biol* (2012) 57:4095–115. doi: 10.1088/0031-9155/57/13/4095

33. Chang C-W, Dinh NT. Classification of machine learning frameworks for data-driven thermal fluid models. *Int J Thermal Sci* (2019) 135:559–79. doi: 10.1016/j.ijthermalsci.2018.09.002

34. Ying X. An overview of overfitting and its solutions. *J Physics: Conf Ser* (2019) 1168:022022. doi: 10.1088/1742-6596/1168/2/022022

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 January 2024
DOI 10.3389/fonc.2024.1348291



OPEN ACCESS

EDITED BY
Minglei Kang,
New York Proton Center, United States

REVIEWED BY
Charles Bloch,
University of Washington, United States
Hardev Grewal,
University of Florida, United States

*CORRESPONDENCE
Konstantin Gordon
gordon@mrrc.obninsk.ru

RECEIVED 02 December 2023
ACCEPTED 15 January 2024
PUBLISHED 30 January 2024

CITATION
Gordon K, Gulidov I, Smyk D, Semenov A, Golubev K, Lemaeva A, Koryakin S, Jumaniyazova E, Vishnyakova P, Eremina I, Fatkhudinov T and Kaprin A (2024) Upright proton therapy for esthesioneuroblastoma: a single-institution experience.
*Front. Oncol.* 14:1348291.
doi: 10.3389/fonc.2024.1348291



# Upright proton therapy for esthesioneuroblastoma: a single-institution experience

Konstantin Gordon[1,2]*, Igor Gulidov[1], Daniil Smyk[1,2], Alexey Semenov[1], Kirill Golubev[1], Alyona Lemaeva[1], Sergey Koryakin[1], Enar Jumaniyazova[2], Polina Vishnyakova[2], Irina Eremina[2], Timur Fatkhudinov[2] and Andrey Kaprin[1,2]

[1]Proton Therapy Department, A. Tsyb Medical Radiological Research Center - Branch of the National Medical Radiological Research Center, Obninsk, Russia, [2]Research Institute of Molecular and Cellular Medicine, Medical Institution, P. Lumumba People's Friendship University of Russia, Moscow, Russia

**Aim:** This study presents an analysis (efficacy and toxicity) of outcomes in patients with esthesioneuroblastoma after pencil beam proton therapy with a fixed beamline in the upright position.

**Background:** Esthesioneuroblastoma (ENB) is an extremely rare tumor of sinonasal area located in critical proximity to vital structures. Proton therapy (PT) is often considered the optimal radiation treatment for head-and-neck tumors, although of limited availability. Upright PT delivered using fixed pencil beamline and rotating chair is a fairly promising option.

**Methods:** This is a single-center experience describing the outcomes of PT in 14 patients with ENB treated between January 2016 and October 2022; half of the cases had a history of previous irradiation. The therapy was applied using a fixed pencil beamline with 6D-chair for positioning. The median dose was 63 GyRBE (total range 48–70 GyRBE; based on 1.1 RBE multiplier for protons) with 2.0 GyRBE per fraction. The mean gross tumor volume was 109.5 $cm^3$ (17.1–257.7 $cm^3$). Patient demography, pathology, treatment parameters and toxicity data were analyzed. Radiation-induced reactions were assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) v 4.0.

**Results:** The median follow-up time was 28 months. The 1- and 2-year locoregional control rates constituted 100% and 88.9%, respectively; the median duration of local control was 52 months. The 1- and 2-year progression-free survival (PFS) rates constituted 92.9% and 75.0%, respectively; the median PFS duration was 52 months. The 1- and 2-year overall survival (OS) rates constituted 92.9% and 84.4%, respectively. Two patients died of non-cancer-related causes (coronavirus-induced pneumonia) and 1 patient died of tumor progression. All patients tolerated PT well without any treatment gaps. Serious late toxicity reactions included glaucoma in 1 patient and cataract in 2 patients, in over half a year since irradiation.

10.3389/fonc.2024.1348291

**Conclusion:** PT with upright design of the unit affords promising outcomes in terms of disease control and toxicity rates in ENB, a sinonasal tumor of complicated localization.

KEYWORDS

esthesioneuroblastoma, proton therapy, radiation therapy, reirradiation, brain invasion, upright position

## 1 Introduction

Esthesioneuroblastoma (ENB), also known as olfactory neuroblastoma, is a tumor of neuroectodermal origin that develops from olfactory receptor cells in the nasal cavity (1), wherefrom it tends to invade the adjacent areas of pterygopalatine fossa, skull base and sinuses. ENBs are rare neoplasms accounting for only 3–6% of all tumors of the nasal cavity, which limits the patient recruitment for randomized trials and interferes with the development of a uniform treatment strategy (1, 2). The treatment is further complicated by unspecific symptoms and accordingly a delay in medical attendance and proper diagnosis; hence the prevalence of advanced cases with locoregional and distant metastases (3).

By the spread of the primary disease, ENBs are classified into 4 stages according to the Kadish staging system: A, the tumor is confined to nasal cavity; B, the tumor spreads to paranasal sinuses; C, the tumor invades skull base, pterygopalatine fossa; D, distant metastases are present. Clinical decisions in ENB are typically based on Kadish stage (4), importantly also accounting for the Hyams morphological grading scheme (5). The treatment mainstay for ENB is a multimodal approach using surgery, irradiation and chemotherapy (6).

Radiation therapy (RT) along with surgery is used as curative treatment for ENB in stage A patients and some stage B patients. The 5-year survival rates with this approach reach 29–63% (4, 7, 8). For more common stages C and D, RT is used as neoadjuvant and adjuvant options. Combining RT with chemotherapy potentially can improve survival rates and reduce the probability of relapse (8).

RT for ENB is liable to multiple treatment-related risks; the factors include tumor localization in the nasal cavity, invasion into skull base, the need for higher doses (66–70 Gy) and the proximity to critically vulnerable neural structures (brain stem, optic nerves, chiasm, etc.). Only high-precision RT techniques should be used to alleviate the risks of severe late toxicity (9, 10). In this regard, proton therapy (PT) is particularly promising due to its favorable dose distribution properties that alleviate the radiation burden on risk organs located close to the tumor.

Since 2016 PT has been established as a standard radiation treatment for patients with ENB at Tsyb Medical Radiological Research Center in Obninsk. Here we present experience of ENB treatment with horizontal pencil proton beam in the upright patient position (Prometheus, JSC Protom).

## 2 Materials and methods

The study included 14 patients (pts) with morphologically confirmed diagnosis of ENB receiving a course of PT within the period from January 2016 to October 2022. All patients were approved for inclusion in a retrospective analysis by the Ethical Review Board at the Tsyb Medical Radiological Research Center. Informed consent was waived due to the retrospective nature of the study and anonymous use of the evidence. The patient data are given in Table 1.

PT was performed using a fixed horizontal beam of intensity-modulated protons with the patient sitting in a 6D-movable 360° rotating chair (upright position) (11). A standard immobilization device (thermoplastic mask) was used to fix the patient's position. The patient's positioning was guided each session using built-in cone-beam computed tomography (CB-CT) (12).

The target volumes were delineated as recommended by the international consensus guidelines for head-and-neck tumors (13, 14). For lymph collectors of the neck, the irradiation volumes were also selected in accordance with specific recommendations (15, 16). The contours of target volumes and risk organs were determined based on CT images obtained during simulation, co-registered with magnetic-resonance images (MRI). The target volume included gross tumor volume (GTV) with a margin of 5 to 10 mm (considering anatomical barriers and organs at risk) to obtain the clinical tumor volume (CTV). The margin to the planned tumor volume (PTV) was 3 mm. All patients underwent CT control (typically on $10^{th}$ and $20^{th}$ fractions) in order to identify and correct errors coming from inflammation and tumor response, as charged particles are sensitive to density changes. The treatment data are given in Table 2.

The relative biological effectiveness (RBE) of the protons was accepted as 1.1. Restrictions on risk organs were set considering the RBE according to the QUANTEC group recommendations. The doses were delivered using a conventional regimen, with 2 GyRBE per fraction.

The total doses were selected in accordance with medical histories: 60–70 Gy for primary cases and 48–60 Gy for reirradiation. The PT dose was prescribed to the PTV with the aim of at least 95% coverage, but in the case of meeting OAR limits, dose constraints prioritized PTV coverage. Doses for reRT cases were chosen based on the reserve of the series OARs, and time from

TABLE 1 Patients' and tumors' characteristics.

| Indicator | Number |
|---|---|
| **Median age** | 53 (48- 59) |
| Sex | |
| Male | 4 (28.6 %) |
| Female | 10 (71.4 %) |
| Hyams grade | |
| Grade II | 8 (57.1 %) |
| Grade III | 6 (42.9 %) |
| Kadish stage | |
| A | 1 (7.1 %) |
| B | 5 (35.8 %) |
| C | 8 (57.1 %) |
| Stage (TNM) | |
| I | 1 (7.1 %) |
| II | 1 (7.1 %) |
| III | 2 (14.3 %) |
| IVa | 4 (28.6 %) |
| IVb | 6 (42.9 %) |
| Tumor origin | |
| Nasal cavity | 9 (64.3 %) |
| Paranasal cavity | 2 (14.3 %) |
| Other | 3 (21.4 %) |
| Brain invasion | |
| Yes | 5 (35.7 %) |
| No | 9 (64.3 %) |
| Treatment history | |
| No | 2 (14.3 %) |
| RT | 1 (7.1 %) |
| Surgery | 4 (28.7 %) |
| Surgery + RT | 2 (14.3 %) |
| Surgery + CTX | 1 (7.1 %) |
| Surgery + RT+ CTX | 3 (21.4 %) |
| RT + CTX | 1 (7.1%) |

*CTX, chemotherapy; RT, radiation therapy.

TABLE 2 Proton therapy characteristics.

| Treatment data | Number |
|---|---|
| **Mean total dose (GyRBE)** | 63 (48-70) |
| RT intention | |
| RT only | 9 (64.3 %) |
| Postoperative RT | 5 ( 35.7 %) |
| RT course | |
| Primary RT | 7 (50%) |
| Repeated RT | 7 (50%) |
| **Median time from previous RT (years)** | 7.3 (3-17) |
| **Median previous dose (Gy)** | 62 (50-70) |
| **Median GTV volume ($cm^3$)** | 109.5 (17.1–257.7) |

*GTV, gross tumor volume; RBE, relative biological efficacy; RT, radiation therapy.

prior RT. We used single-field optimized PTV-based plans, usually generated with 4–5 fields.

Representative plan of PT is shown in Figure 1. Treatment planning system ProtomTherapyPlanner ver. 2.14 (JSC Protom) was used.

All patients were examined before the start of the treatment, then every 3 months during the first year after the end of treatment and subsequently at 6 months intervals. The control examination after the course of PT included contrast-enhanced MRI (or CT in cases of MRI contraindication), ultrasound scan of cervical lymph nodes and positron-emission CT with glucose if required. The toxicity was assessed using CTCAE v 4.0 criteria.

The follow-up time was calculated from the end of the treatment to the last clinical assessment. Local control was defined as the absence of tumor growth in the irradiated area. Progression-free survival was defined as the absence of locoregional or distant progression. The overall survival (OS) was calculated from the end of PT to the last visit or date of death. Statistical analysis was carried out using StatTech v3.1.8 (Stattekh LLC). The survival curve was built using the Kaplan-Meyer method. An adjusted $p$-value <0.05 indicated a statistically significant difference; $p$-values ≥0.05, but <0.1 were noted as tendencies. No stratification by disease or treatment parameters was made due to the small size of the cohort.

## 3 Results

The patient and PT data for 14 pts with ENB included in the study are listed in Tables 1, 2. The mean age of participants was 53 years (range 48–59 years). Most of the tumors were located in the nasal cavity (64.3%) and paranasal sinuses (14.3%). Most of the patients had locally advanced disease at the time of the treatment, with intracranial invasion observed in 35.7% of the cases (Figure 2).

Two patients had no previous treatment history, 4 pts had surgery, 1 pt received RT only, 2 pts received surgery and RT, 1 pt received surgery with adjuvant and neoadjuvant chemotherapy, 1 pt received PT followed by chemotherapy and 2 courses of chemoembolization, 3 pts had surgery, and also received radiation and chemotherapy. Thus, 7 out of 14 pts underwent reirradiation.

The median follow-up for the group was 28 months (range 4–64). The 1- and 2-year locoregional control rates were, respectively 100% and 88.9% (95% CI: 43.3–98.4); the median local control duration was 52 months (95% CI: 43.9–60.0). The 1- and 2-year progression-free survival (PFS) constituted, respectively, 92.9%

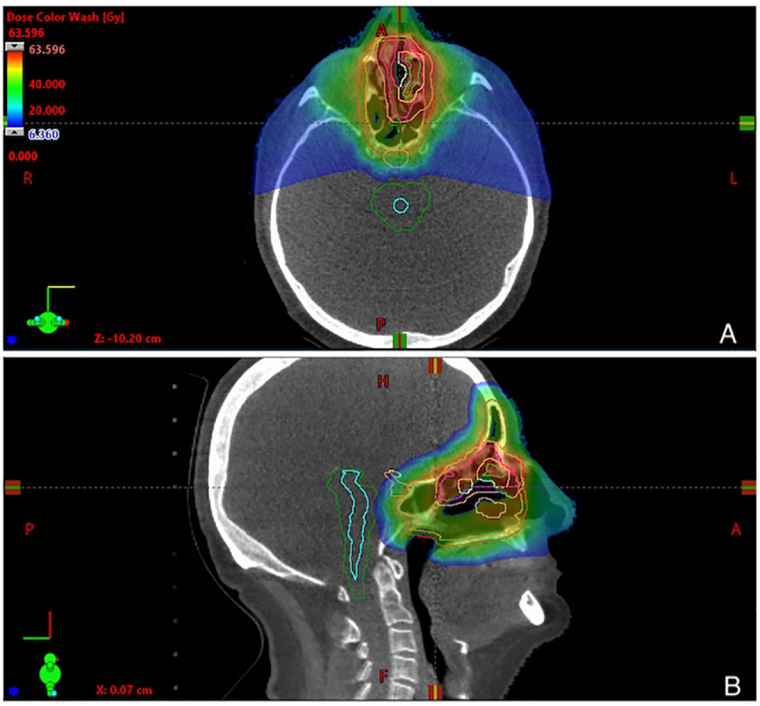



FIGURE 1
Representative 3-fields plan of proton therapy, showing dose distribution in **(A)** axial and **(B)** sagittal view. Physical doses are given in a dose legend **(A)**.

(95% CI: 59.1–99.0) and 75.0% (95% CI: 40.3–91.3); the median PFS was 52 months (95% CI: 18–∞). The 1- and 2-year OS rates constituted, respectively, 92.9% (95% CI: 59.1 – 99.0) and 84.4% (95% CI: 50.4–95.9) (Figures 3, 4).

In the follow-up, disease progression was recorded in 4 pts, including 3 pts with local (in-field) progression and 1 pt with lymph node metastases. Due to inoperability of the local relapses, 2 pts received systemic chemotherapy and 1 pt received chemoembolization, and 1 pt with regional progression underwent extended unilateral cervical lymphadenectomy. It should be noted that in all 4 cases of progression, the patients had stage IVa-b tumor process at the time of PT, and 3/4 pts had intracranial invasion appearing to interfere with the locoregional control after RT (p= 0.09) (Figure 5). No other significant factors influencing the treatment outcome were revealed. Two patients died due to non-cancer reasons (COVID-19 infection), and 1 pt died of disease progression with severe brain edema.

Early toxicity reactions were assessed for the entire cohort (14 pts, 100%). In 6 pts (42.8%) the early toxicity was limited to grade 1, in 7 pts (50.0%) it was grade 2, and 1 pt had acute toxicity of grade 3. The reactions, mostly local, included dermatitis (85.7%), mucositis (57.1%), conjunctivitis (50.0%) and xerostomia (28.5%).

Late toxicity events were recorded in 11 pts (78.5%). Specifically, 9 pts (64.2%) had grade 1-2 reactions including nasal cavity mucosa atrophy (64.2%) and in-field post-radiation fibrosis (35.7%). Also, 1 pt developed glaucoma and 2 pts developed radiation-induced cataract within 6–12 months after the treatment.

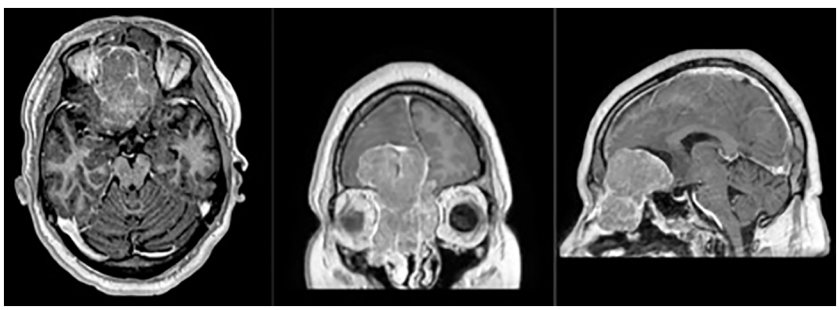


FIGURE 2
Magnetic-resonance (T1 sequence with contrast enhancement, axial, frontal and sagittal views) of the patient with Kadish C tumor stage, with a massive frontal lobes invasion.

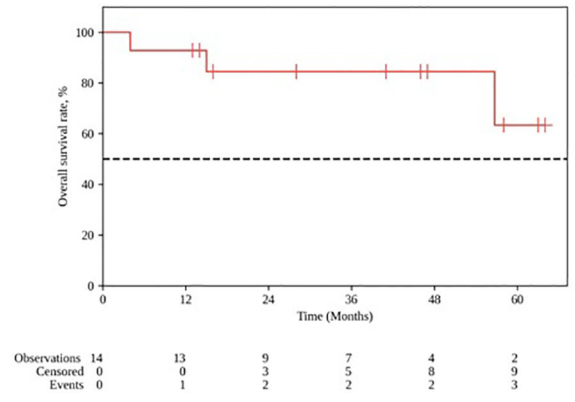



FIGURE 3
Overall survival after PT for ENB. Kaplan-Meyer plot.

## 4 Discussion

Sinonasal esthesioneuroblastomas (ENBs) are rare malignant neoplasms of the head-and-neck region. The lack of uniform clinical approaches for ENBs is due to their rarity. In addition, the principle of clinical decision-making on the basis of tumor morphology plus localization might not be well applicable to ENBs, as these tumors may combine certain properties of neuroendocrine tumors (due to their origin from olfactory neuroepithelium) with those of squamous cell carcinomas. Additional difficulties arise from the established clinical classification for ENBs, which is different from the TNM staging system, and malignancy grade determination for these tumors, which also differs from the typically used grading system.

The default first-line option for ENB is surgery, but, given the tumor localization, curative surgery for ENB is often dismissed as crippling or technically unfeasible. The use of adjuvant modalities, notably a range of chemotherapies from platinum monotherapy to cyclophosphamide, decarbazine, etoposide, vincristine, etc., in various combinations over decades, produced extensive clinical data indicating the doubt of significant benefits of chemotherapy in ENB (8, 17).

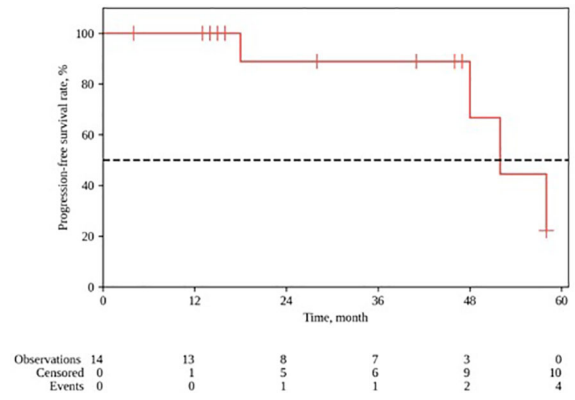



FIGURE 4
Progression-free survival after PT for ENB. Kaplan-Meyer plot.

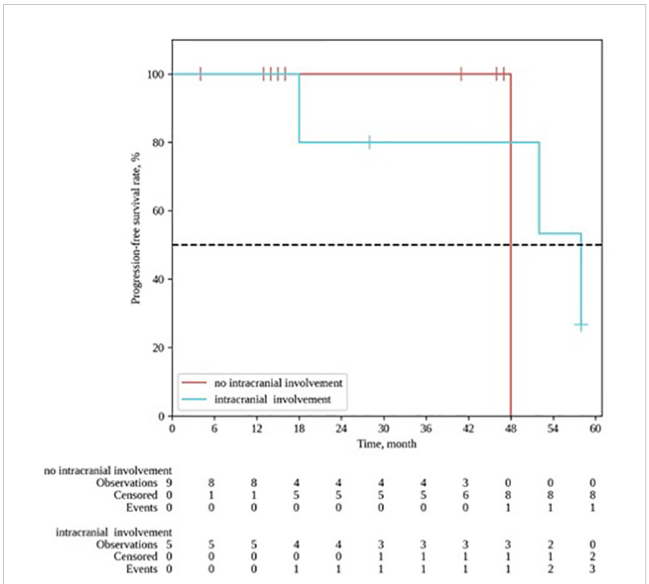



FIGURE 5
Locoregional control in patients with (blue line) / without (red line) frontal lobe invasion. Kaplan-Meyer plot.

The main clinical strategies envisioned for ENB account for its dual nature of neuroendocrine tumor and squamous cell head-and-neck carcinoma by using both the Kaddish staging system and Hyams histological grading. The local Kadish stage A-B and Hyams grade I-II are subject to monotherapy —— either surgery or RT. By contrast, locally advanced tumors of Hyams grade III-IV require a combination of surgery, chemotherapy and RT (optionally a combined chemoradiotherapy) (1). In such cases, RT can be considered as an alternative to surgical treatment (18).

The choice of irradiation method for ENB is extremely important, since the tumor is localized in the facial part of the skull packed with vital structures and also cosmetically important. With clear indications for RT, as ENBs are radiosensitive tumors (19), the risks of radiation complications should be addressed scrupulously.

PT is a generally accepted irradiation method that allows significant reduction of the dose to surrounding tissues, especially relevant for head-and-neck tumors (20). However, its use is limited by complexity, big size and high costs of the equipment. The upright position design of the unit levels some of these disadvantages, making the treatment more accessible for clinical practice while maintaining its quality (21).

Our study enrolled 14 patients who underwent PT for local and locally advanced ENB in upright position with an active scanning beam. The limitations of our finding include small size of the cohort, as well as its considerable heterogeneity. However, most of the published clinical evidence for ENB has similar limitations of small studies. Of note, even the largest analysis of ENB treatment for more than 900 patients provides no detailed data on RT outcomes (17).

The median age of our patients was 53 ( ± 10 years), which is typical for ENB (8). The median total radiation dose was high, amounting to 63 GyRBE. The median survival constituted 52 months. The 1- and 2-year locoregional control rates constituted 100% and 88.9%, respectively. At the time of this analyses, 11/14 pts

remain under observation. Three patients died; in 2 cases the death was related to the consequences of coronavirus infection and 1 pt died from intracranial progression.

As demonstrated by Wang et al., intracranial extension is not an adverse prognostic factor in ENB treatment (22). In our cohort, about one-third of the patients had frontal lobe invasion at the time of the treatment. However, we observed a tendency towards worse treatment outcomes in cases of frontal lobe invasion ($p$ = 0.09); in addition, the local control rates constituted 40.0% vs 88.9% for patients with and without intracranial invasion, respectively.

At the same time, such key parameters of the disease as Hyams malignancy grade, Kadish prevalence stage and the presence of regional metastases showed no significant effect on the local control in ENB (23), although this result may reflect the relatively short follow-ups and the issue should be additionally addressed for longer observation periods.

As demonstrated in some studies, PT not only has dosimetric advantages, but also favorable profiles of treatment-related toxicity (24). In our study, acute radiation toxicity was typically represented by grade 1-2 local mucositis and grade 1-2 dermatitis. The sinus area is one of the most difficult locations to plan and administer irradiation. PT with an active scanning beam is particularly challenging due to multiple density transitions (air-bone-soft tissue) in a relatively small volume, as well as changes in the density and size of the mucous membrane due to extensive inflammatory reaction. We used the strategy of 2 verification cone-beam CT scans after each 20 GyRBE, which allowed us to decrease the potential set-up and density errors and avoid late toxicity, especially from visual structures.

Late complications encountered for the cohort included grade 3 dry mucous membrane in 1 pt, hypoosmia in 2 pts and cataract in 2 pts. Also, 2 pts noted prolonged swelling of the nasal mucosa; 1 pt was diagnosed with ocular melanoma shortly after treatment, which required enucleation.

One of the earliest reports on PT in sinonasal ENB was published by Nishimura et al. in 2007. The authors presented clinical outcomes for 14 patients after PT on a 235 MeV cyclotron with a gantry system, with a dose of 2.5 GyRBE/daily and a total dose of 65 GyRBE. Five-year OS and PFS constituted, respectively, 93% and 84%. The authors encountered acute toxicity grade 1-2 and late toxicity grade 3, and no higher grade radiation toxicity reactions throughout the observation period (25).

Later on, a clinical experience of PT for sinonasal ENBs in a small cohort of 13 patients was reported by researchers at the University of Florida. The treatment was carried out with a passive scattered proton beam, in hyperfractionation mode, 1.2 GyRBE per fraction, which enabled safe delivery of 64.8–74.4 GyRBE doses. In this study, 10 of 13 recipients survived for at least 35 months; the acute toxicity was low and regressed within 4 weeks after the end of irradiation (26).

In 2015, a small experience of PT in 8 patients aged 4 to 21 was presented by Massachusetts General Hospital. Four-year OS in the group was 87.5%. It is important to note that 4 pts (half of the cohort) developed endocrine dysfunctions; other complications included retinopathy in 2 pts and grade III ophthalmopathy in 1 pt (27). The higher tissue radiosensitivity in children compared to adults requires special consideration when planning PT (28).

A cohort of 42 patients described by Nakamura et al. in 2017 can be considered the largest experience of PT for ENB. The authors noted a weak relationship of treatment outcomes to Kadish stage. Thus, 5-year OS rates were 100% and 76% for Kadish stages A and C, respectively, although the difference was statistically non-significant. Treatment failures were primarily due to distant or regional progression (48%) and only 10 pts developed local recurrence. Moreover, age under 50 was a significant favorable factor, even in cases of tumor progression. In contrast to the experience of PT at other centers including ours, the study encountered acute grade 3 reactions including mucositis in 4 pts and dermatitis in 1 pt. Visual complications grade 3-4 emerged in 4 pts; also, 1 pt with Kadish stage C developed grade 4 liquorrhea. Another important note was the lack of effectiveness of adding chemotherapy in Kaddish stage C (10).

In 2018, the University of Heidelberg presented experience with IMRT and carbon-ion therapy (CIT) in a heterogeneous group of 17 patients with primary and recurrent tumors, including 4 cases of re-irradiation, which was identified as a factor for a worse prognosis. In 13 patients without a history of RT, 4-year OS was 100%. The most common radiation toxicity reaction in this cohort was asymptomatic cerebral edema (30%) (29). Similarly with our study, 2 pts died of intracranial tumor growth. A history of RT had no effect on the outcomes probably due to the long interim between the courses (median gap 7.3 years).

Another study of 2018 enrolled 21 patients receiving CIT for $T_4$ ENB in Japan. Three-year OS and LC rates were, respectively, 88.4% and 83%; 3 pts developed severe ophthalmopathy grade 4 (30).

One of the most recent studies on PT in ENB, published by Chang et al. in 2022. assessed post-operative IMPT in 15 patients with ENB; the doses constituted 60–72 GyRBE (median 68 GyRBE). At a median follow-up of 20 months, 2-year OS and relapse-free survival rates were, respectively, 88% and 83%. Severe complications encountered in this study included single cases of brain tissue necrosis and sinus obstruction (31).

In general, due to the rarity of ENBs, no optimal radiation treatment parameters for these tumors have been determined in clinical trials. Importantly, 50% of patients the studied cohort received reirradiation with protons. Thus, PT afforded positive clinical outcomes even for already treated and relapsed cases of ENB. A recent large multi-institutional analysis features PT as an effective and safe option of radiation treatment for tumors of sinonasal localization (including ENB) regardless of tumor morphology (32). Moreover, for sinonasal tumors, repeated irradiation with protons and/or carbon ions allows radiation toxicity risk minimization as compared with other protocols (33).

Upright PT is an old technology almost suspended from clinical use due to the lack of reliable means for precision patient positioning at the beginning of the proton era. With new technological advances in the image guidance, patient positioning, beam delivery, etc., the approach is gaining a renewed interest (21). Our own clinical practice comprises a positive experience of upright PT for similarly complicated tumors including skull base chordomas and chondrosarcomas (34).

Apart from small size and heterogeneous treatment history of the cohort. The second is its retrospective design of the study. All abovementioned studies are heterogeneous in treatment criteria,

outcome assessment. This variability affects the direct comparability of the results. Still, we consider it important to report the outcomes for the state-of-the-art upright PT in ENBs due to their pathogenetic distinctiveness combined to rarity and limited knowledge about optimal management for these tumors.

## 5 Conclusion

The study provides a unique example of upright PT for sinonasal ENB. The outcomes indicate acceptable effectiveness and safety of the treatment independently of irradiation history. Accordingly, the treatment can be considered as a strong alternative to gantry PT in ENB.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

This study was approved by the local ethical committee and the institutional review board of A. Tsyb Medical radiological research center—branch of the National medical research radiological center of the Ministry of Health of Russia, including waver of informed consent due to its retrospective nature. All procedures were performed following the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1964, as revised in 2013.

## Author contributions

KoG: Conceptualization, Validation, Writing – original draft, Writing – review & editing. IG: Conceptualization, Supervision, Writing – review & editing. DS: Data curation, Writing – original draft. AS: Data curation, Formal analysis, Software, Validation, Writing – original draft. KiG: Data curation, Investigation, Methodology, Writing – original draft. AL: Data curation, Formal analysis, Resources, Writing – original draft. SK: Resources, Supervision, Validation, Writing – review & editing. EJ: Data curation, Formal analysis, Validation, Writing – original draft. PV: Formal analysis, Investigation, Supervision, Writing – review & editing. IE: Project administration, Resources, Supervision, Writing – review & editing. TF: Funding acquisition, Project administration, Resources, Writing – review & editing. AK: Funding acquisition, Project administration, Resources, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was carried out with the financial support of the Ministry of Education and Science of Russia; Agreement dated 7 October 2021 No. 075-15-2021-1356 (internal number of the Agreement: 15.SIN.21.0011); (ID: RF 0951.61321X0012).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Fiani B, Quadri SA, Cathel A, Farooqui M, Ramachandran A, Siddiqi I, et al. Esthesioneuroblastoma: A comprehensive review of diagnosis, management, and current treatment options. *World Neurosurg* (2019) 126:194–211. doi: 10.1016/j.wneu.2019.03.014

2. Ghaffar S, Salahuddin I. Olfactory neuroblastoma: A case report and review of the literature. *Ear Nose Throat J* (2005) 84:150–2. doi: 10.1177/014556130508400311

3. Veyrat M, Vérillaud B, Fiaux-Camous D, Froelich S, Bresson D, Nicolai P, et al. Olfactory neuroblastoma. *Adv Otorhinolaryngol* (2020) 84:154–67. doi: 10.1159/000457935

4. Platek ME, Merzianu M, Mashtare TL, Popat SR, Rigual NR, Warren GW, et al. Improved survival following surgery and radiation therapy for olfactory neuroblastoma: analysis of the SEER database. *Radiat Oncol* (2011) 6:41. doi: 10.1186/1748-717X-6-41

5. Saade RE, Hanna EY, Bell D. Prognosis and biology in esthesioneuroblastoma: the emerging role of Hyams grading system. *Curr Oncol Rep* (2015) 17:423. doi: 10.1007/s11912-014-0423-z

6. Sheehan J, Payne R. Esthesioneuroblastomas. In: Winn HR, editor. *Youmans and Winn neurological surgery e-book*. New York: Elsevier Health Sciences (2016). p. 1284–92.

7. Yin Z, Gao L, Luo J, Yi J, Huang X, Qu Y, et al. Long-term outcomes of patients with esthesioneuroblastomas: A cohort from a single institution. *Oral Oncol* (2016) 53:48–53. doi: 10.1016/j.oraloncology.2015.11.021

8. Dulguerov P, Allal AS, Calcaterra TC. Esthesioneuroblastoma: a meta-analysis and review. *Lancet Oncol* (2001) 2:683–90. doi: 10.1016/S1470-2045(01)00558-7

9. Bao C, Hu W, Hu J, Dong Y, Lu JJ, Kong L. Intensity-modulated radiation therapy for esthesioneuroblastoma: 10-year experience of a single institute. *Front Oncol* (2020) 10:1158. doi: 10.3389/fonc.2020.01158

10. Nakamura N, Zenda S, Tahara M, Okano S, Hayashi R, Hojo H, et al. Proton beam therapy for olfactory neuroblastoma. *Radiother Oncol* (2017) 122:368–72. doi: 10.1016/j.radonc.2016.12.020

11. Pryanichnikov AA, Sokunov VV, Shemyakov AE. Some results of the clinical use of the proton therapy complex "Prometheus". *Phys Particles Nuclei Lett* (2018) 15:981–5. doi: 10.1134/S1547477118070592

12. Balakin VE, Belikhin MA, Pryanichnikov AA, Shemyakov AE, Strelnikova NS. Clinical application of new immobilization system in seated position for proton therapy. *KnE Energy* (2018) 3:45. doi: 10.18502/ken.v3i2.1790

13. Li R, Tian S, Zhu Y, Yan L, Zhu W, Quan H, et al. Management of orbital invasion in esthesioneuroblastoma: 14 years' experience. *Radiat Oncol* (2019) 14:107. doi: 10.1186/s13014-019-1313-1

14. Hansen CR, Johansen J, Samsøe E, Andersen E, Petersen JBB, Jensen K, et al. Consequences of introducing geometric GTV to CTV margin expansion in DAHANCA contouring guidelines for head and neck radiotherapy. *Radiother Oncol* (2018) 126:43–7. doi: 10.1016/j.radonc.2017.09.019

15. Vorwerk H, Hess CF. Guidelines for delineation of lymphatic clinical target volumes for high conformal radiotherapy: head and neck region. *Radiat Oncol* (2011) 6:97. doi: 10.1186/1748-717X-6-97

16. Biau J, Lapeyre M, Troussier I, Budach W, Giralt J, Grau C, et al. Selection of lymph node target volumes for definitive head and neck radiation therapy: a 2019 Update. *Radiother Oncol* (2019) 134:1–9. doi: 10.1016/j.radonc.2019.01.018

17. Orton A, Boothe D, Evans D, Lloyd S, Monroe MM, Jensen R, et al. Esthesioneuroblastoma: A patterns-of-care and outcomes analysis of the national cancer database. *Neurosurgery* (2018) 83:940–7. doi: 10.1093/neuros/nyx535

18. Benfari G, Fusconi M, Ciofalo A, Gallo A, Altissimi G, Celani T, et al. Radiotherapy alone for local tumour control in esthesioneuroblastoma. *Acta Otorhinolaryngol Ital* (2008) 28:292–7.

19. Eich HT, Staar S, Micke O, Eich PD, Stützer H, Müller R-P. Radiotherapy of esthesioneuroblastoma. *Int J Radiat Oncol Biol Phys* (2001) 49:155–60. doi: 10.1016/S0360-3016(00)00811-7

20. Gordon KB, Smyk DI, Gulidov IA. Proton therapy in head and neck cancer treatment: state of the problem and development prospects (Review). *Sovremennye tehnologii v Med* (2021) 13:70. doi: 10.17691/stm2021.13.4.08

21. Volz L, Sheng Y, Durante M, Graeff C. Considerations for upright particle therapy patient positioning and associated image guidance. *Front Oncol* (2022) 12:930850. doi: 10.3389/fonc.2022.930850

22. Wang J, Wang L, He H, Li Y, Song X. The treatment outcomes of olfactory neuroblastoma patients with frontal lobe invasion. *Front Oncol* (2021) 11:640892. doi: 10.3389/fonc.2021.640892

23. Konuthula N, Iloreta AM, Miles B, Rhome R, Ozbek U, Genden EM, et al. Prognostic significance of Kadish staging in esthesioneuroblastoma: An analysis of the National Cancer Database. *Head Neck* (2017) 39:1962–8. doi: 10.1002/hed.24770

24. Jumaniyazova E, Smyk D, Vishnyakova P, Fatkhudinov T, Gordon K. Photon- and proton-mediated biological effects: what has been learned? *Life* (2022) 13:30. doi: 10.3390/life13010030

25. Nishimura H, Ogino T, Kawashima M, Nihei K, Arahira S, Onozawa M, et al. Proton-beam therapy for olfactory neuroblastoma. *Int J Radiat Oncol Biol Phys* (2007) 68:758–62. doi: 10.1016/j.ijrobp.2006.12.071

26. Malyapa RS, Mendenhall WM, McKenzie C, Yeung D, Li Z, Mendenhall NP, et al. Proton therapy of esthesioneuroblastoma: the UFPTI experience. *Int J Radiat Oncol Biol Phys* (2011) 81:S524–5. doi: 10.1016/j.ijrobp.2011.06.810

27. Lucas JT, Ladra MM, MacDonald SM, Busse PM, Friedmann AM, Ebb DH, et al. Proton therapy for pediatric and adolescent esthesioneuroblastoma. *Pediatr Blood Cancer* (2015) 62:1523–8. doi: 10.1002/pbc.25494

28. McGovern SL, Okcu MF, Munsell MF, Kumbalasseriyil N, Grosshans DR, McAleer MF, et al. Outcomes and acute toxicities of proton therapy for pediatric atypical teratoid/rhabdoid tumor of the central nervous system. *Int J Radiat Oncol Biol Phys* (2014) 90:1143–52. doi: 10.1016/j.ijrobp.2014.08.354

29. Liermann J, Syed M, Held T, Bernhardt D, Plinkert P, Jungk C, et al. Advanced radiation techniques in the treatment of esthesioneuroblastoma: A 7-year single-institution's clinical experience. *Cancers (Basel)* (2018) 10:457. doi: 10.3390/cancers10110457

30. Hiroaki Suefuji, Koto M, Demizu Y, Saitoh JI, Shioyama Y, Tsuji H, et al. A retrospective multicenter study of carbon ion radiotherapy for locally advanced olfactory neuroblastomas. *Anticancer Res* (2018) 38. doi: 10.21873/anticanres.12399

31. Chang J-H, Wu C-C, Yuan KS-P, Wu ATH, Wu S-Y. Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, and treatment outcomes. *Oncotarget* (2017) 8:55600–12. doi: 10.18632/oncotarget.16340

32. Yu NY, Gamez ME, Hartsell WF, Tsai HK, Laramore GE, Larson GL, et al. A multi-institutional experience of proton beam therapy for sinonasal tumors. *Adv Radiat Oncol* (2019) 4:689–98. doi: 10.1016/j.adro.2019.07.008

33. Vischioni B, Ingargiola R, Bonora M, Ronchi S, Camarda AM, Russo S, et al. Particle reirradiation of Malignant epithelial and neuroectodermal sinonasal tumors: A case series from CNAO. *J Clin Med* (2023) 12:2624. doi: 10.3390/jcm12072624

34. Gordon K, Gulidov I, Koryakin S, Smyk D, Makeenkova T, Gogolin D, et al. Proton therapy with a fixed beamline for skull-base chordomas and chondrosarcomas: outcomes and toxicity. *Radiat Oncol* (2021) 16:238. doi: 10.1186/s13014-021-01961-9

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 February 2024
DOI 10.3389/fonc.2023.1333098



OPEN ACCESS

EDITED BY
Xinyuan Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Alexander Rühle,
University Hospital Leipzig, Germany
Ciprian Camil Mirestean,
University of Medicine and Pharmacy of Craiova, Romania

*CORRESPONDENCE
June Corry
june.corry@genesiscare.com.au

RECEIVED 04 November 2023
ACCEPTED 20 December 2023
PUBLISHED 05 February 2024

CITATION
Corry J, Moore A, Kenny L, Wratten C, Fua T, Lin C, Porceddu S, Liu C, Ruemelin M, Sharkey A, McDowell L, Wilkinson D, Tiong A and Rischin D (2024) Radiotherapy quality assurance in the TROG 12.01 randomised trial and its impact on loco-regional failure. *Front. Oncol.* 13:1333098. doi: 10.3389/fonc.2023.1333098



# Radiotherapy quality assurance in the TROG 12.01 randomised trial and its impact on loco-regional failure

June Corry[1,2,3]*, Alisha Moore[4], Liz Kenny[5,6], Chris Wratten[7], Tsien Fua[3], Charles Lin[5], Sandro Porceddu[8], Chen Liu[3], Michael Ruemelin[9], Amy Sharkey[9], Lachlan McDowell[3], Dean Wilkinson[10], Albert Tiong[3] and Danny Rischin[11,12]

[1]Genesiscare Radiation Oncology Department, St Vincents Hospital, Melbourne, VIC, Australia, [2]Department Medicine, University of Melbourne, Melbourne, VIC, Australia, [3]Department Radiation Oncology, Peter MacCallum Cancer Center, Melbourne, VIC, Australia, [4]Department Radiation Quality Assurance, Trans-Tasman Radiation Oncology Group (TROG), Newcastle, NSW, Australia, [5]Department Radiation Oncology, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia, [6]Faculty Medicine, University of Queensland, Brisbane, QLD, Australia, [7]Department Radiation Oncology, Calvary Mater Hospital and University Newcastle, Newcastle, NSW, Australia, [8]Department Radiation Oncology, Princess Alexander Hospital, Brisbane, QLD, Australia, [9]Department Radiation Therapy, Peter MacCallum Cancer Center, Melbourne, VIC, Australia, [10]Department Radiation Therapy, Illawarra Cancer Care Centre, Wollongong, NSW, Australia, [11]Department Medical Oncology, Peter MacCallum Cancer Center, Melbourne, VIC, Australia, [12]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia

**Introduction:** There is no consensus as to what specifically constitutes head and neck cancer radiotherapy quality assurance (HNC RT QA). The aims of this study are to (1) describe the RT QA processes used in the TROG 12.01 study, (2) review the RT QA processes undertaken for all patients with loco-regional failure (LRF), and (3) provide prospective data to propose a consensus statement regarding the minimal components and optimal timing of HNC RT QA.

**Materials and methods:** All patients undergoing RT QA in the original TROG 12.01 study were included in this substudy. All participating sites completed IMRT credentialling and a clinical benchmark case. Real-time (pre-treatment) RT QA was performed for the first patient of each treating radiation oncologist, and for one in five of subsequent patients. Protocol violations were deemed major if they related to contour and/or dose of gross tumour volume (GTV), high dose planning target volume (PTVhd), or critical organs of risk (spinal cord, mandible, and brachial plexus).

**Results:** Thirty HNROs from 15 institutions accrued 182 patients. There were 28 clinical benchmark cases, 27 pre-treatment RT QA cases, and 38 post-treatment cases. Comprehensive RT QA was performed in 65/182 (36%) treated patients. Major protocol violations were found in 5/28 benchmark cases, 5/27 pre-treatment cases, and 6/38 post-treatment cases. An independent review of all nine LRF cases showed major protocol violations in four of nine cases.

**Conclusion:** Only pre-treatment RT QA can improve patient outcomes. The minimal components of RT QA in HNC are GTVs, PTVhd, and critical organs at risk. What constitutes major dosimetric violations needs to be harmonised.

KEYWORDS

quality assurance, intensity modulated radiotherapy (IMRT), radiotherapy, head and neck (H&N) cancer, human papilloma virus - HPV

## Introduction

Since the publication of the landmark study by Peters et al. (1), head and neck radiation oncologists (HNROs) have been aware of the importance of the quality of radiotherapy in optimising patients' loco-regional control and overall survival. Since that study, some form of radiotherapy quality assurance (RT QA) has been incorporated into the majority of head and neck cancer (HNC) clinical trials.

However, the term "radiotherapy quality assurance" is an umbrella term with no current consensus as to its optimal components, the timing, or what is the optimal percentage of HNC patients who should undergo RT QA.

We have previously retrospectively shown the impact of pre-treatment RT QA of all curative intent cases at a large HNC centre (2). In our study the RT QA consisted of a review of the staging imaging and of the gross tumour volumes (GTVs), planning treatment volumes (PTVs), and critical organs at risk (OARs) by a second RO for *all* patients having curative intent non-surgical treatment. This RT QA occurred prior to dosimetric planning and prior to treatment.

While most HNC published studies include a statement regarding performance of RT QA in their study, the precise details of that RT QA are not always included. There is general agreement in HNC that a major goal of RT QA is to reduce errors that are likely to lead to reduced tumour control probability and/or significant and serious treatment-related toxicity. Questions remain as to what is the optimal RT QA process required to achieve this goal—how do we balance effectiveness, efficiency, and cost within the RT QA process? Clearly, real-time review of every case would be optimal, but this is seen as costly in terms of time and effort and possibly unnecessary.

This prospective study reports fully on the RT QA processes that were used in the randomised trial of weekly cetuximab versus weekly cisplatin and radiation in good prognosis loco-regionally advanced HPV-associated oropharyngeal squamous cell carcinoma —TROG 12.01 (3).

The aims of this study are to (1) give a detailed account of the RT QA processes used in the original study, (2) review the RT QA processes that had been undertaken for all patients who failed loco-regionally, and (3) provide prospective data to propose a consensus statement regarding the minimal components and optimal timing of HNC RT QA.

## Methods

Intensity-modulated radiotherapy (IMRT) was mandatory for participation in the TROG 12.01 study. All participating sites in TROG 12.01 had pre-study credentialing that included completion of a volumetric arc therapy (VMAT)/IMRT facility questionnaire, a Level III dosimetry audit (determining the absorbed dose delivered to selected points within an anthropomorphic phantom; this is an end-to-end audit where the phantom undergoes all steps within the radiotherapy treatment chain), and submission of a library benchmarking case for RT QA review of dosimetry and of all protocol contours by an independent HNRO from a panel of five HNROs. Once the library benchmark case was successfully completed, sites were then eligible to commence patient accrual (Figure 1).

In centres where one or more benchmark case had been submitted by other HNROs from that centre, subsequent benchmarking cases were not required for the remaining HNROs enrolling patients from that centre on the proviso that they had a robust institutional RT QA programme (i.e., defined as a review of all new definitive cases' imaging and contours) and their first case underwent real-time RT QA prior to commencing treatment.

A central real-time pre-treatment RT QA review was performed for the first patient of each radiation oncologist investigator, and post-treatment reviews were performed for one in five patients for each investigator (Figure 2).

Centres sent the diagnostic images and the simulation CT images with all radiotherapy dosimetry and contouring data to the Trans-Tasman Oncology Group (TROG) group. Dosimetry, treatment delivery, and scheduling were reviewed centrally, and the accuracy of all the contouring volumes was reviewed by an independent member of the five-HNRO review panel. Each panel member specialised in HNC and had five or more years of experience.

Each investigator had a copy of the protocol that detailed the RT techniques, including GTV to PTV margins of minimum 10 mm, and the criteria for unilateral RT. In addition, each investigator had the table of items to be reviewed and the acceptable, minor, and major protocol violations for each of these structures (Table 1).

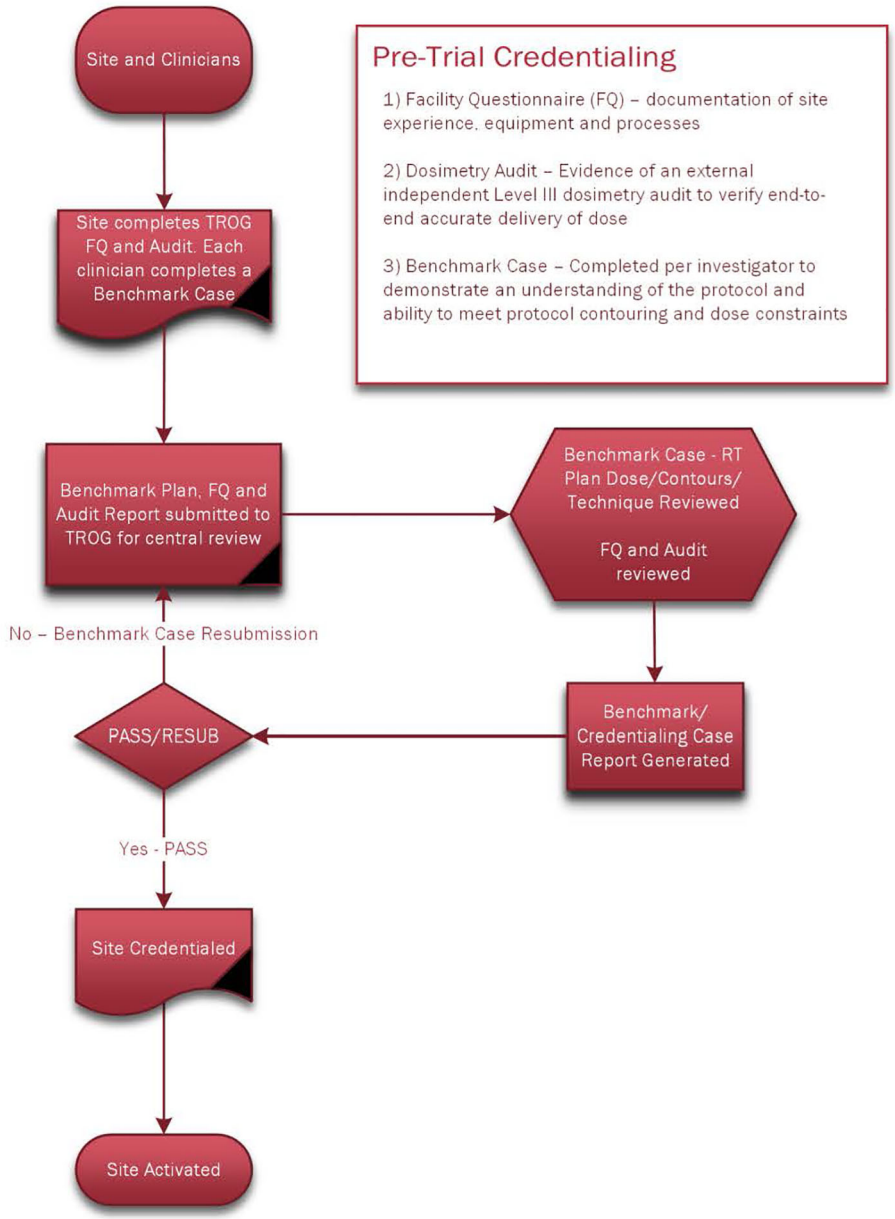



FIGURE 1
Pre-Trial Credentialing.

Major protocol violations were those with major variations in contour and/or dose to the Priority 1 structures (GTVp, GTVn, and PTV70 and critical organs at risk—spinal cord, mandible, and brachial plexus). All other structures were placed in the Priority 2 category (PTV intermediate and low dose, pharyngeal constrictors, larynx, parotids, and oral cavity). The protocol included contouring atlases for brachial plexus and pharyngeal constrictors.

## Statistical analysis

Formal statistical analysis was not required. This report is a presentation of the RT QA results and was limited to descriptive reporting and percentages.

All cases of loco-regional failure (LRF) were reviewed by one of the authors (JC). These cases were blinded for institutional site and responsible radiation oncologist; the actual site of failure (i.e., either local or regional, or both) and the clinical outcomes for each case were also unknown. Once the RT QA review was complete for that list of patients, then the site of failure was made known so that a correlation could then be made between any major protocol violations and the clinical likelihood of that violation (RT dose and/or contour) contributing to the site of failure. For each case, it was also documented what prospective RT QA had been performed in relation to that case, specifically whether a benchmark case had been submitted, and whether pre-treatment or post-treatment RT QA had been performed. We also documented the presence or absence of institutional RT QA for each case.

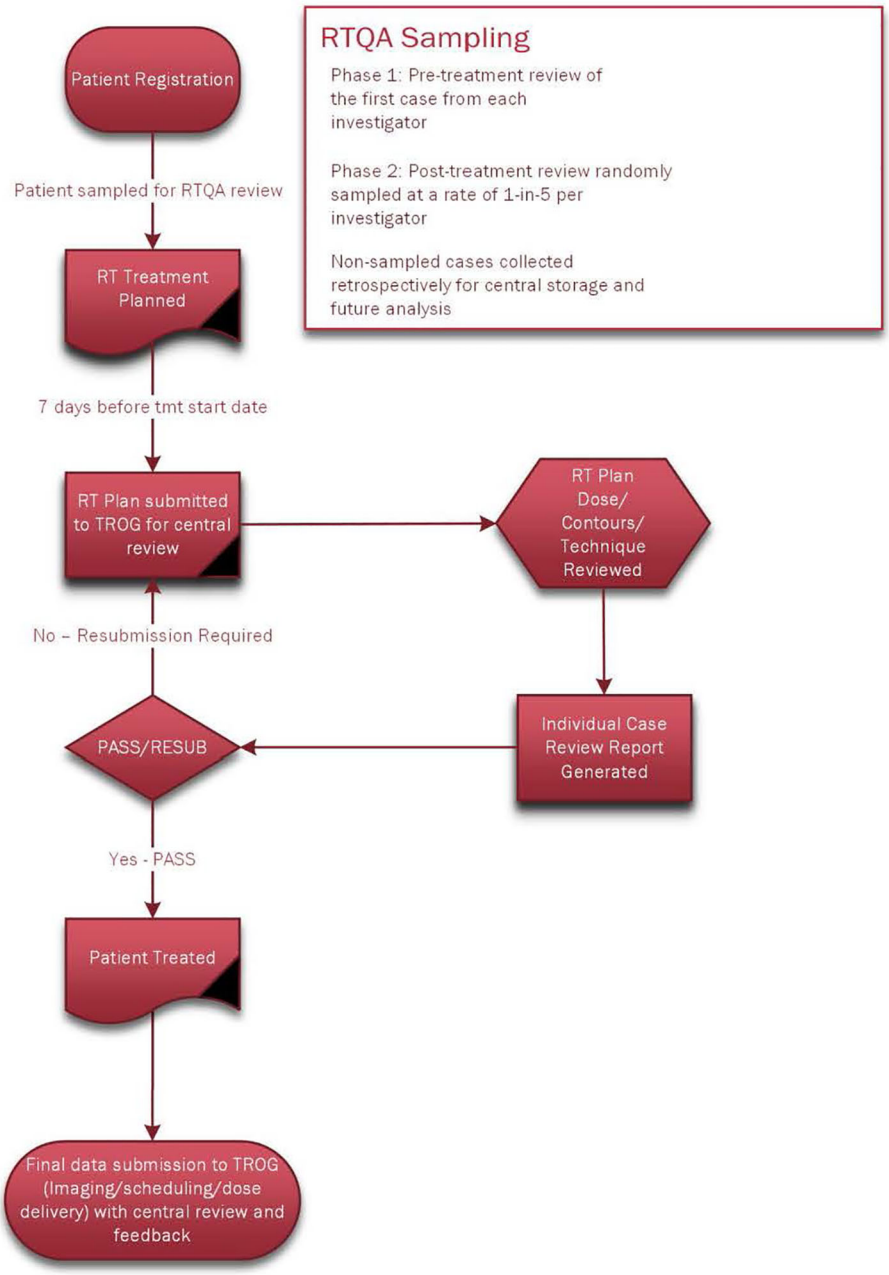



FIGURE 2
On-Trial Radiotherapy Quality Assurance.

## Results

Overall, there were 182 patients available for analysis in the TROG 12.01 study. All except one patient from each arm (180/182, 99%) received the prescribed dose in the prescribed time frame, 70 Gy in 35 fractions over 7 weeks.

Thirty HNROs from 15 institutions accrued patients to TROG 12.01.

There were 28 library benchmark cases, 27 cases underwent real-time RT QA pre-treatment, and 38 cases had their RT QA review performed after completion of their treatment. Thus, in total, there was comprehensive RT QA performed in 93 cases and in 65/182 (36%) treated patients.

### Benchmark cases

There were 28 benchmark cases submitted. Twenty-three cases were protocol compliant, and five cases required resubmission due to major protocol variations in Priority 1 structures (contouring of brachial plexus 2, contour PTV70, and D1% dose to brachial plexus × 4 and spinal cord × 1). Of these, two were corrected by

TABLE 1 Major and minor protocol violations criteria in TROG 12.01.

| Question | Protocol | Baseline | Acceptable | Minor variation | Major variation | Missing/ Invaluable |
|---|---|---|---|---|---|---|
| **Target Volumes—GTV & CTV** | | | | | | |
| GTV-P: Minimum Dose (D100%) | No | >/=66.5 Gy | N/A | N/A | N/A | N/A |
| Percentage GTV-P receiving <95% of the prescribed dose (V95) | Yes | 0 | 0%–2% | 2.01%–7% | >7% | N/A |
| GTV-N: Minimum Dose (D100%) | No | >/=66.5 Gy | N/A | N/A | N/A | N/A |
| Percentage GTV-N receiving <95% of the prescribed dose (V95) | Yes | 0 | 0%–2% | 2.01%–7% | >7% | N/A |
| Is the maximum dose contained within the CTV? | No | Yes | Yes | N/A | N/A | N/A |
| **Target Volumes—PTV** | | | | | | |
| PTV70Gy: D95% | Yes | >/=66.5 Gy | >/=66.5 Gy | 65.1–66.49 Gy | <65.1 Gy | N/A |
| Percentage of PTV70 receiving </=66.5 Gy (V95%) | Yes | 0 | 0%–5% | 5.01%–7% | >7% | N/A |
| PTV70: Near Minimum (D98%) | No | >/=66.5 Gy | N/A | N/A | N/A | N/A |
| PTV70: Median Dose (D50%) | No | 70 Gy | 68.6–71.4 Gy | N/A | N/A | N/A |
| PTV70: Near Maximum (D2%) | Yes | 70 Gy | <110% (76.99 Gy) | 110%–115% (77–80.5 Gy) | >115% (80.51 Gy) | N/A |
| PTV67Gy: D95% | Yes | >/=63.65 Gy | >/=63.65 Gy | 60.3–63.64 Gy | <60.3 Gy | N/A |
| Percentage of PTV67 receiving </=63.65 Gy (V95%) | Yes | 0 | 0-5% | 5.01%–10% | >10% | N/A |
| PTV67: Near Minimum (D98%) | No | >/=63.65 Gy | N/A | N/A | N/A | N/A |
| PTV63Gy: D95% | Yes | >/=59.85 Gy | >/=59.85 Gy | 58.6–59.84 Gy | <58.6 Gy | N/A |
| Percentage of PTV63 receiving </=59.85 Gy (V95%) | Yes | 0 | 0%–5% | 5.01%–10% | >10% | N/A |
| PTV63: Near Minimum (D98%) | No | >/=59.85 Gy | N/A | N/A | N/A | N/A |
| PTV54Gy: D95% | Yes | >/=51.3 Gy | >/=51.3 Gy | 45.9–51.29 Gy | <45.9 Gy | N/A |
| Percentage of PTV54 receiving </=51.3 Gy (V95%) | Yes | 0 | 0%–5% | 5.01%–15% | >15% | N/A |
| PTV54: Near Minimum (D98%) | No | >/=51.3 Gy | N/A | N/A | N/A | N/A |
| **Critical OARs** | | | | | | |
| Spinal Cord: D1% | Yes | </=45 Gy | </=45 Gy | 0%–3% (45.01–46.35 Gy) | >3% (>46.35 Gy) | N/A |
| Spinal Cord: Maximum Point Dose | No | </=45 Gy | N/A | N/A | N/A | N/A |
| Spinal Cord: PRV (Sc + 5 mm): D1% | Yes | </=50 Gy | </=50 Gy | 0%–3% (50.01–51.5 Gy) | 6(>51.5 Gy) | N/A |
| Spinal Cord: PRV (Sc + 5 mm): Maximum Point Dose | No | </=50 Gy | N/A | N/A | N/A | N/A |
| Brachial Plexus Left: D1% | Yes | </=66 Gy | </=66 Gy | 0%–3% (66.01–67.98Gy) | >3% (>67.98Gy) | N/A |
| Brachial Plexus Left: Maximum Point Dose | No | </=66 Gy | N/A | N/A | N/A | N/A |
| Brachial Plexus Right: D1% | | </=66 Gy | </=66 Gy | 0-3% (66.01–67.98Gy) | >3% (>67.98Gy) | N/A |

*(Continued)*

TABLE 1 Continued

| Question | Protocol | Baseline | Acceptable | Minor variation | Major variation | Missing/ Invaluable |
|---|---|---|---|---|---|---|
| Brachial Plexus Right: Maximum Point Dose | No | </=66 Gy | N/A | N/A | N/A | N/A |
| Other OARs | | | | | | |
| Mandible D1% | Yes | </=70 Gy | </=70 Gy | 0-3% (70.01–72.1 Gy) | >3% (>72.1 Gy) | N/A |
| Mandible: Maximum Point Dose | No | </=70 Gy | N/A | N/A | N/A | N/A |
| Parotid Gland (Right): Mean Dose | No | </=26 Gy | </=26 Gy | N/A | N/A | N/A |
| Parotid Gland (Left): Mean Dose | No | </=26 Gy | </=26 Gy | N/A | N/A | N/A |
| Glottic Larynx: Mean Dose | No | </=45 Gy | </=45 Gy | N/A | N/A | N/A |
| Constrictors: Mean Dose | No | </=63 Gy | </=63 Gy | N/A | N/A | N/A |
| Oral Cavity: Mean Dose | No | </=42 Gy | </=42 Gy | N/A | N/A | N/A |

the responsible HNRO and resubmitted and passed RT QA (2/5, 40%), and one was not resubmitted but the first patient from that investigator was reviewed in real time (thus included in the pre-treatment review section). Two were never resubmitted, and these two investigators did not then participate in the study.

## Pre-treatment review

### Priority 1 structures

There were 30 clinicians and, thus, 30 patients were to have real-time pre-treatment review. However, two clinicians did not submit their cases with adequate time for pre-treatment review to be completed (and they were subsequently reviewed post-treatment), and one case was missed from pre-treatment review. Thus, there was a total of 27 real-time pre-treatment RT QA review completed. In these 27 cases, there were six major variations in Priority 1 structures in five patients (four in contouring—GTVp contour × 2, PTV70 contour, and brachial plexus contour, and two in dosimetry to the brachial plexus) (D1% > 68 Gy).

There were 13 minor variations in Priority 1 structures in nine patients (7 in contouring—spinal cord × 2, PTV70 × 3, and brachial plexus × 2, and 6 in dosimetry—mandible D1% × 2, brachial plexus D1% × 1, 95% of PTV70 received less than 65.1 Gy × 2, and 100% GTVn receiving less than 66.5 Gy × 1).

### Priority 2 structures

There were two major variations in two patients in Priority 2 structures (one contour PTV63, one dosimetry 95% of PTV63 received < 58.6 Gy).

There were 20 minor variations in 13 patients in Priority 2 structures (19 in contouring—pharyngeal constrictors × 5, oral cavity × 5, larynx × 2, PTV54 × 3, and PTV63 × 4, and 1 in dosimetry 95% PTV54 receiving less than 51.3 Gy).

Pre-treatment cases with major violations in Priority 1 structures were corrected and resubmitted in three of the five cases. The violations corrected in these resubmitted cases were the GTV contour × 2 and PTV70 contour × 1. The three patients where correction and resubmission were not requested had major violations, namely, contouring of the brachial plexus × 1, and dose to the brachial plexus exceeding 68 Gy × 2.

## Post-treatment review

### Priority 1 structures

In the 38 cases reviewed post-treatment, there were seven major variations in Priority 1 structures in six patients: two in contouring (GTVn and PTV70) and five in dosimetry (95% of PTV70 receiving less than 65.1 Gy × 2, brachial plexus D1% > 68 Gy, mandible D1%, and 100% GTVp receiving < 66.5 Gy).

There were 24 minor violations in Priority 1 structures in 15 patients: 16 in contouring (GTVp × 1, GTVn × 2, brachial plexus × 8, and spinal cord × 5) and 8 in dosimetry (95% of PTV70 received < 66.5 Gy × 4, mandible D1% × 2, and brachial plexus D1% × 2).

### Priority 2 structures

There were five major variations in Priority 2 structures in four patients: four in contouring (parotid × 2, pharyngeal constrictors × 1, and PTV54 × 1) and one in dosimetry, 95% PTV54 received < 45.9 Gy).

There were 39 minor variations in Priority 2 structures in 24 patients: 33 in contouring (PTV54 × 12, PTV 63 × 1, pharyngeal constrictors × 7, oral cavity × 9, and larynx × 4) and 6 in dosimetry (95% of PTV67 received < 60.3 Gy × 2, 95% of PTV63 received < 58.6 Gy × 2, and 95% of PTV54 received < 45.9 Gy × 2).

Overall, there were a total of 10 major violations in either contour or dosimetry in nine patients, 13.8% of the RT QA population. The major variations were two for unacceptable contours (GTVn × 1 and PTV70 × 1) and eight for dosimetry (mandible D1% > 72.1 Gy, Brachial plexus D1% > 68 Gy × 4, 95% PTV70 receiving less than 65.1 Gy × 2, 95% GTVp receiving less than 65.1 Gy).

### RT QA of the patients with loco-regional failure

There were nine LRFs in the 182 treated patients: six regional, two local, and one loco-regional. Of these nine cases, the results of the RT QA associated with each case are tabulated in Table 2.

In these nine cases, four had undergone study RT QA and five had not. In the four cases that had undergone review, three were reviewed post-treatment and one was reviewed pre-treatment. The RT QA reviews in these patients had not shown any major protocol violations, and the second review (JC) was concordant in three cases (75%). The non-concordant case had originally been reviewed post-treatment as having no major protocol violations, but the second review assessment had major protocol violations in the contouring of GTVn and hence PTV70, and this had a significant probability of contributing to the regional failure. This regional recurrence was not resectable and the patent died of disease.

In the five cases not previously reviewed, RT QA revealed major protocol violations in three cases (60%), with a significant probability of contributing to local or regional failure (see Figure 3). In the first case, the GTVn was assessed as under contoured and hence the GTVn-to-PTV70 margin was too small. In addition, the margin from GTVp to PTV70 was 7 mm and the protocol recommended a minimum of 10 mm. This patient died of local and regional failure. In the second case, the margin from GTVn to PTV70 ranged from 4 to 6 mm, and this patient failed in the neck. They had a salvage neck dissection and remain alive with no evidence of disease. In the third case, the GTVp was assessed as under contoured with a subsequent close GTV-to-PTV margin, and this probably contributed to their local failure. They had surgical salvage and remain alive with no evidence of disease. In none of these cases were the less-than-10-mm GTV-to-TV margins related to anatomical boundaries.

## Discussion

The study represents the most detailed report of any RT QA process in an HNC study in the IMRT era.

Over a third of the study patients had comprehensive RT QA. Within that group, 86% had no major protocol violations, and if brachial plexus was omitted as a Priority 1 structure (four cases in three patients), then 90% would have had no major treatment violations.

Thus, the question remains—what items should be included as a minimum for adequate RTQA? As mentioned, the main purpose of HNC RT QA is to optimise cancer control probability and minimise the risk of serious late treatment toxicity. Hence, in terms of items to be assessed, it is not controversial to include GTVp, GTVn, and PTV70 and spinal cord as "Priority 1" structures in oropharyngeal cancer RT QA. However, inclusion of the mandible and brachial plexus, and the maximum doses that constitute a major protocol violation is more controversial. We believed that they were worthy of inclusion because mandibular osteoradionecrosis and brachial plexopathy are serious late treatment toxicities that cause considerable patient morbidity.

TABLE 2 Distribution of the RT QA processes for the loco-regional failure patients ($n$ = 9).

| | Yes | No |
|---|---|---|
| Benchmark case submitted | 8 | 1 |
| Institutional RT QA | 5 | 4 |
| Pre-treatment RT QA review | 1 | 8 |
| Post-treatment RT QA review | 3 | 6 |

There is considerable variation in dosimetry criteria used for RT QA in different protocols. Table 3 compares the RT QA criteria in the protocols of TROG 12.01 and RTOG 1016 (3). It is important to remember that the RTOG 1016 protocol was for 70 Gy in 35 fractions over 6 weeks, with PTV2 56 Gy and PTV3 50 Gy, whereas TROG 12.01 prescribed 70 Gy in 35 fractions over 7 weeks with PTV2 63 Gy and PTV3 54 Gy. However, if we compare "Priority 1 structures", in the RTOG 1016 trial, there was no major protocol violation ascribed to GTV coverage, whereas in the TROG 12.01 trial, the GTV had to receive a minimum of 66.5 Gy. For PTV70, the major deviation was 95% receiving <65.1 Gy in the TROG 12.01, but in RTOG 1016, it was only if <63 Gy. Spinal cord doses were also recorded differently. In TROG 12.01, if the maximum point dose to the spinal cord was >46.4 Gy, it was a major violation, whereas in RTOG 1016, it was >50 Gy, which was considered a major violation.

The brachial plexus D1% maximum dose was a major violation if greater than 68 Gy in TROG 12.01, but it was not mandated in RTOG 1016. The maximum mandible dose in TROG 12.01 was 70 Gy, and D1% maximum dose > 72 Gy was a major violation. In RTOG 1016, it was recommended the maximum dose be less than 66 Gy, but it did not seem to be a major violation. Interestingly the maximum dose allowed within PTV1 in TROG 12.01 was D2% 77 Gy and any higher was a major violation, whereas in RTOG 1016, a PTV1 hot spot was accepted up to 82 Gy. Thus, two experienced HNC trial groups, within the same disease subsite, demonstrate significant differences for structures and dosimetric constraints recorded as major protocol violations. Hence, that needs to be standardised by a consensus statement from major HNC research groups for future studies.

Regarding the number and timing of cases be reviewed, when first planning the RT QA for this study, we had thought completion of the benchmark case may be the most important component of the RT QA process. Successful completion of the benchmark case would be the best way to ensure that all had read and adhered to the study RT protocol prior to entering patients on the study.

However, eight of the nine cases of LRF had had a benchmarking case performed by that clinician; thus, this suggests that it may not be that helpful in reducing specific RT protocol violations.

A significant proportion (3/9, 33%) of the LRF cases had had post-treatment review, but this is of limited value as the patient has completed treatment and any protocol violations cannot be adjusted. It is helpful as a general overview of the quality of the RT delivered, but not for reducing adverse outcomes in any particular patient.

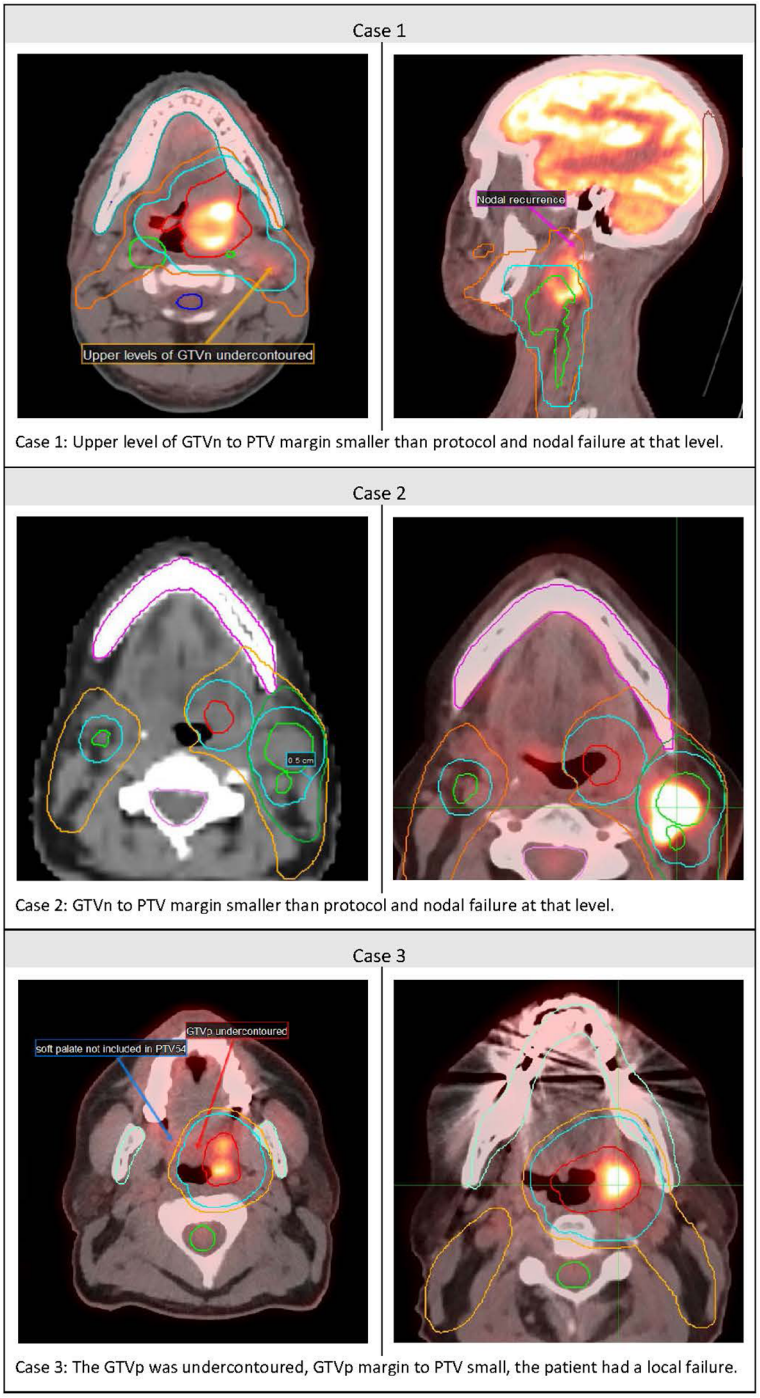



FIGURE 3
Cases of Major Protocol violations and loco-regional failure.

Perhaps the greatest concern was that eight of the nine patients with LRF had not had pre-treatment review. This is clearly the timing that allows corrections before treatment (as per the Peters et al. study) and hence most directly correlates with better oncological outcomes. Thus, pre-RT QA is the area that deserves greater concentration of resources. Pre-treatment review in the TROG 12.01 study resulted in a 40% reduction in Priority 1 major violations, given that three of five cases with major violations in Priority 1 structures were corrected prior to treatment. Theoretically, if the nine patients with LRF had all undergone pre-treatment RT QA review, and the major protocol violations in Priority 1 structures had been corrected, then the LRF rate could *potentially* have been reduced by 4, i.e., from 9 to 5, so a reduction in the LRF rate from 9/182 (4.9%) to 5/182 (2.7%), or a halving of the rate of LRF.

TABLE 3 Comparisons of Items included in HNC RT QA trial protocols.

| Item | TROG 12.01 | RTOG 1016 | Major Deviations TROG 12.01 | Major Deviations RTOG 1016 |
|---|---|---|---|---|
| GTV | Mandatory | Mandatory | Minimum dose > 66.5 Gy | Not stated |
| GTV-PTV margin | Mandatory | Mandatory | 10–15 mm | 10–25 mm |
| PTV70 | Mandatory | Mandatory | 95% PTV70 < 65.1 Gy | 95% PTV70 < 63 Gy |
| PTVid* | Mandatory | Mandatory | 95% PTVid < 58.6 Gy | 95% PTVid < 45 Gy |
| PTVld# | Mandatory | Mandatory | 95% PTVld < 45.9 Gy | 95% PTVld < 40 Gy |
| Spinal cord PRV | Optional | Mandatory | Not stated | >52 Gy |
| Spinal cord | Mandatory | Mandatory | > 46.4 Gy | >50 Gy |
| Brain stem PRV | Optional | Mandatory | Not stated | >52 Gy |
| Mandible | Mandatory | Optional | >72 Gy | >66 Gy |
| Brach Plex | Mandatory | Not required | >68 Gy | Not stated |
| Max dose in PTV1 | | | >77 Gy | >82 Gy |
| | | | Recommended doses | Recommended doses |
| Parotids | Mandatory | Mandatory | Mean < 26 Gy | Mean < 26 Gy |
| Pharynx | Mandatory | Optional | Mean < 63 Gy | Uninvolved mean < 45 Gy |
| Glottis | Mandatory | Optional | Mean < 45 Gy | Mean < 20 Gy |
| SMG | Mandatory | Optional | Mean < 39 Gy | Mean < 39 Gy |
| Oral cavity | Mandatory | Optional | Mean < 42 Gy | Uninvolved mean < 30 Gy |
| Lips | Optional | Optional | Not stated | Mean < 20 Gy |
| Cervical oesoph | Optional | Optional | Not stated | Mean < 30 Gy |

*PTVid, planning target volume intermediate dose; #PTVld,– planning target volume low dose.

The financial cost of RT QA is not insignificant but may vary greatly between different countries. Dosimetry recalculations and dosimetric protocol violations can be recorded automatically via programmes such as CQMS (central quality management system). What cannot currently be automated is the review of GTV and PTV contours. It is possible that in the future, artificial intelligence could be helpful in this area. Currently, the average time taken for RT QA review of imaging and contours by an independent HNRO is approximately 20 min per case (4), or approximately 60 AUD. Quite apart from the emotional cost of salvaging an LRF, the financial cost is high. There are no Australian figures for the cost of managing LRF, but American and European studies suggest that it is in the order of 30,000 AUD (5, 6). In fact, in the TROG 12.01 study, all the HNRO RTQA was performed on an honorary basis. However, if not, the approximate costing of contour reviews for all 182 patients' pre-treatment would have been approximately 11,000 AUD, or roughly a third of managing a single recurrence.

Finally, what proportion of HNC patients need to undergo RTQA for optimal results? There are no data to answer that question. To date, percentages used range around the 10% mark (7), but this is a pragmatic response to available resources rather than a scientific or financial costing of relative benefit.

Ideally, one would review all cases and see if an algorithm could be formulated to determine the optimal percentage of cases needed to undergo pre-treatment RT QA for optimal or most efficient detection of major protocol violations.

## Conclusions

RT QA is important for the optimal management of HNC. This study reinforces the point that pre-treatment peer review with formal RT QA and feedback to the treating HNRO offers the highest likelihood of reducing major protocol violations and improving patient outcomes. There needs to be consensus as to the items to be included in RT QA, but GTV, PTVs, and critical OAR are a good starting point. The percentage of cases that should undergo such review requires further study. Nevertheless, we need to arrest the drift of RT QA being performed post-treatment. Pre-treatment RT QA needs to be a standard procedure during the treatment planning stage and deserves appropriate allocation of resources for the optimal management of HNC patients.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Peter MacCallum Cancer Center, Melbourne, Australia. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

JC: Conceptualization, Writing – original draft, Writing – review & editing. AM: Data curation, Formal Analysis, Writing – review & editing. LK: Writing – review & editing. CW: Writing – review & editing. TF: Writing – review & editing. CLin: Writing – review & editing. SP: Writing – review & editing. CLiu: Writing – review & editing. MR: Writing – review & editing. AS: Writing – review & editing. LM: Writing – review & editing. DW: Writing – review & editing. AT: Writing – review & editing. DR: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Health and Medical Research Council (Project Grant 1047673).

## Acknowledgments

We thank all the patients who participated, all clinicians who contributed to this study, and TROG staff Patrick Wheeler and Olivia Cook. Merck KGaA, Darmstadt, Germany reviewed the manuscript for medical accuracy only before journal submission.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Author Disclaimer

The authors are fully responsible for the content of this manuscript and the views and opinions are described in the publication reflect solely those of the authors.

## References

1. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from the TROG 02.02. *J Clin Oncol* (2010) 28(18):2996–3001. doi: 10.1200/JCO.2009.27.4498

2. Amarasena I, Herschtal A, D'Costa I, Fua T, Tiong A, Geddes V, et al. Outcomes of routine intensity modulated radiation therapy quality assurance in large head and neck cancer center. *Int J Radiat Oncol Biol Phys* (2017) 98(3):541–6. doi: 10.1016/j.ijrobp.2017.02.215

3. Rischin D, King M, Kenny L, Porceddu S, Wratten C, Corry J, et al. Randomised trial of radiotherapy with weekly cisplatin or cetuximab in low risk HPV associated oropharyngeal cancer (TROG 12.01) – a Trans-Tasman Radiation Oncology Group study. *Int J Radiat Oncol Biol Phys* (2021) 111(4):876–86. doi: 10.1016/j.ijrobp.2021.04.015

4. Gillison ML, Trotti AM, Harris J, Eisbruch A, Harari PM, Adelstein DJ, et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *Lancet* (2019) 393:40–50. doi: 10.1016/S0140-6736(18)32779-X

5. Kim LT, Winfree KB, Yang H, Marynchenko M, Yu AP, Frois C, et al. Treatment patterns and economic burden of metastatic and recurrent locally advanced head and neck cancer patients. *J Med Economics* (2012) 15:786–95. doi: 10.3111/13696998.2012.682632

6. Van der Linden N, Buter J, Pescott CP, Lalisang RI, de Boer JP, de Graeff A, et al. Treatments and costs for recurrent and metastatic squamous cell carcinoma of the head and neck in the Netherlands. *Eur Arch Otolaryngol* (2016) 273(2):455–64. doi: 10.1007/s00405-015-3495-y

7. Overgaard J, Mohanti BK, Begum N, Ali R, Agarwal JP, Kuddu M, et al. Five versus six fractions of radiotherapy per week for squamous-cell carcinoma of the head and neck (IAEA-ACC study): a randomised, multicentre trial. *Lancet Oncol* (2010) 11(6):553–60. doi: 10.1016/S1470-2045(10)70072-3

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 28 February 2024
DOI 10.3389/fonc.2024.1346797



OPEN ACCESS

EDITED BY
Xinyuan Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Hong Qi Tan,
National Cancer Centre Singapore, Singapore
Michael Andrew Samuels,
Banner Health, United States

*CORRESPONDENCE
Satomi Shiraishi
shiraishi.satomi@mayo.edu

RECEIVED 29 November 2023
ACCEPTED 14 February 2024
PUBLISHED 28 February 2024

CITATION
Madhavan S, Gamez M, Garces YI, Lester SC, Ma DJ, Mundy DW, Neben Wittich MA, Qian J, Routman DM, Foote RL and Shiraishi S (2024) Impact of radiation dose distribution on nutritional supplementation needs in head and neck cancer radiotherapy: a voxel-based machine learning approach. *Front. Oncol.* 14:1346797. doi: 10.3389/fonc.2024.1346797



# Impact of radiation dose distribution on nutritional supplementation needs in head and neck cancer radiotherapy: a voxel-based machine learning approach

Sudharsan Madhavan, Mauricio Gamez, Yolanda I. Garces, Scott C. Lester, Daniel J. Ma, Daniel W. Mundy, Michelle A. Neben Wittich, Jing Qian, David M. Routman, Robert L. Foote and Satomi Shiraishi*

Department of Radiation Oncology, Mayo Clinic, Rochester, MN, United States

**Objectives:** To investigate the relationship between nutritional supplementation and radiation dose to the pharyngeal constrictor muscles and larynx for head and neck (HN) cancer patients undergoing radiotherapy.

**Methods:** We retrospectively analyzed radiotherapy (RT) dose for 231 HN cancer patients, focusing on the pharyngeal constrictors and larynx. We defined nutritional supplementation as feeding tube utilization or >10% weight loss from baseline within 90 days after radiotherapy completion. Using deformable image registration (DIR), we mapped each patient's anatomical structures to a reference coordinate system, and corresponding deformations were applied to dose matrices. Voxel doses were utilized as features for ridge logistic regression models, optimized through 5-fold cross-validation. Model performance was assessed with area under the curve of a receiver operating curve (AUC) and F1 score. We built and compared models using 1) pharyngeal constrictor voxels, 2) larynx voxels, 3) clinical factors and mean regional dose metrics, and 4) clinical factors and dose-volume histogram metrics. Test set AUCs were compared among the models, and feature importance was evaluated.

**Results:** DIR of the pharyngeal constrictors and larynx yielded mean Dice coefficients of 0.80 and 0.84, respectively. Pharyngeal constrictors voxels and larynx voxel models had AUC of 0.88 and 0.82, respectively. Voxel-based dose modeling identified the superior to middle regions of the pharyngeal constrictors and the superior region of larynx as most predictive of feeding tube use/weight loss. Univariate analysis found treatment setting, treatment laterality, chemotherapy, baseline dysphagia, weight, and socioeconomic status predictive of outcome. An aggregated model using mean doses of pharyngeal constrictors and larynx subregions had an AUC of 0.87 and the model using conventional DVH metrics had an AUC of 0.85 with p-value of 0.04. Feature importance calculations from the regional dose model indicated that mean doses to the superior-middle

pharyngeal constrictor muscles followed by mean dose to the superior larynx were most predictive of nutritional supplementation.

**Conclusions:** Machine learning modeling of voxel-level doses enables identification of subregions within organs that correlate with toxicity. For HN radiotherapy, doses to the superior-middle pharyngeal constrictors are most predictive of feeding tube use/weight loss followed by the doses to superior portion of the larynx.

KEYWORDS

**voxel-based analysis, head and neck cancer, outcomes modeling, feeding tube, explainable machine learning, larynx, pharyngeal constrictor muscles, weight loss**

## 1 Introduction

Swallowing difficulties are a prevalent side effect of radiotherapy (RT) treatments for head and neck cancers (1–5). RT involves targeting cancers with a three-dimensional (3D) radiation dose. This often leads to the incidental irradiation of nearby organs that play a role in swallowing. In contemporary practices, radiotherapy treatment plans condense the 3D dose distributions inside delineated organs into two-dimensional dose-volume histograms (DVHs). Specific metrics for organs at risk (OARs) and target volumes are scrutinized to reduce the chance of adverse side effects. Physicians qualitatively assess the spatial distribution of radiation doses, focusing on regions that may pose potential toxicities or affect target coverage. This method has proven effective, with a number of toxicities linked to DVH metrics through various normal tissue complication probability (NTCP) modeling, as evidenced in QUANTEC and various clinical trials (1, 5–10). For example, Mavroidis et al. (11), showed that generalized mean dose for superior pharyngeal constrictors to be most predictive of dysphagia at 6 months post-RT. Using Lyman Kutcher Burman (LKB) model which is a popular methodology accounting for seriality of the OAR (12, 13), their study reported $D_{50}$ (dose at which there is 50% chance of complication) of 62.0 Gy, slope parameter m = 0.1, and dose-volume parameter n = 0.49 with AUC of 0.74 for superior pharyngeal constrictor muscles, suggesting moderate sensitivity to subregion damage (14). However, when relying exclusively on DVHs for analysis, delineation of the OARs is necessary, and it is presupposed that every section of an organ has an equal sensitivity to radiation and contributes identically to the overall risk of toxicity. Additionally, Samant et al. reported that machine learning (ML) models can often quantify NTCP better than LKB models, motivating explorations of ML approaches for toxicity analysis (15).

Furthermore, recent studies have shown that the toxicity observed can depend on which subregion of a segmented organ was irradiated (2–4, 16–37). Functionally distinct subregions within a single OAR contour may not be accounted for in current treatment planning. For example, Jiang et al. used ML to identify subregions within the parotid and submandibular glands that correlated with xerostomia (22). Another study by Eisbruch et al. found the pharyngeal constrictor muscles and glottic/supraglottic larynx subsites were most dysphagia-related using videofluoroscopy (38). In clinical practice, it is common for glottic, supraglottic, and subglottic larynx subregions to be grouped under a single larynx segmentation. Likewise, the superior, middle, and inferior pharyngeal constrictor muscles are collectively evaluated as one entity during treatment planning. In addition to the potential presence of distinct subregions, there is a growing number of studies on voxel-based optimization of radiation treatment plans (39–41). These studies have used voxel-based objectives for optimization and opens an opportunity to reflect spatial dose constraints during treatment planning.

Conventionally, investigation of sub-regions of an OAR required radiomics texture extraction based on images and/or time-consuming and resource-intensive manual contouring of each region (42). Manual contouring is also subject to inter-observer variations and inconsistent implementation of contouring guidelines (43–45). In this study, we employed a machine-learning approach to determine the radiation dose's voxel-wise correlation with malnutrition, pinpointing distinctive features in the pharyngeal constrictors and larynx sub-regions without preliminary contouring. We focused on interpretable feature importance to analyze the spatial dose dependency within an organ. This may eventually inform dose sparing of a subregion during treatment planning to reduce malnutrition risk. While methodologies akin to ours have been used in head and neck RT studies focusing on toxicities such as xerostomia and acute dysphagia (28, 29), to the best of our knowledge, there has not been an investigation into the 3D dose distribution's impact on malnutrition, measured by feeding tube (FT) utilization and weight loss. Furthermore, predicting the need for a feeding tube early in radiotherapy can also enhance patient care. Although prophylactic FT placement is occasionally advised (46), its blanket application can be unnecessary or harmful. About half of the preemptively

inserted FTs prove marginally beneficial (47), with complications such as infections being prevalent (48). Moreover, FT placement postpones the transition back to regular diets (49), adversely affecting long-term well-being (18, 50, 51). Therefore, identifying patients in true need of FT can improve quality of care.

# 2 Materials and methods

## 2.1 Cohort selection

After Institutional Review Board approval, we conducted a retrospective analysis of 352 patients treated for head and neck cancer at our institution from January 2016 to November 2020, all of whom had granted consent for their medical records to be used in research. We accessed our department's patient outcomes database (52) and filtered for patients based on International Classification of Diseases (ICD)-9 and 10 codes (53, 54) specific to cancers in the salivary glands, oropharynx, oral cavity, nasopharynx, nasal cavity, sinuses, larynx, and hypopharynx. We excluded patients based on the following criteria: absence of baseline dysphagia assessment before radiotherapy (60 patients), FT insertion before RT (34 patients), radiation doses outside the range of 1.2-2.2 Gy/fraction, discernible disfigurement of the pharyngeal constrictors and larynx due to disease or surgery, and prior RT in an area with potential overlap (27 patients). Ultimately, 231 patients, treated with either photon or proton radiotherapy with prescription dose in the range of 30-81.6 Gy delivered in 15-68 fractions, were deemed suitable for this study.

## 2.2 Data collection

Table 1 summarizes the clinical variables studied in our analysis. We sourced data from our institution's electronic health record reporting database, focusing on parameters such as gender, feeding tube usage, weight, birth date, and the primary address's 9-digit zip code. Feeding tube utilization data was gathered by looking for procedure codes corresponding to the insertion of stomach, gastrostomy, or jejunostomy tubes. Patients who had a feeding tube before starting RT were excluded based on the procedure date, as such utilization is likely attributable to surgery, disease, or both as opposed to RT. Socioeconomic status was inferred using the Area Deprivation Index by Kind and Buckingham (55), derived from the zip code associated with the patient's primary address. This index, ranging from 1 to 100, gauges the socioeconomic disadvantage of a neighborhood, with higher scores denoting greater disadvantage. We also collected baseline dysphagia grades assessed by the care team following the Common Terminology Criteria for Adverse Events (CTCAE v4.03) within ±2 weeks of radiotherapy initiation (56, 57). Additionally, data on smoking habits, concurrent chemotherapy, and treatment context (either primary or post-operative RT) was obtained via chart reviews. Table 1 provides information on clinical variables and on treatment sites of the patients included in the study.

All radiotherapy treatments were planned in the Eclipse treatment planning system (Siemens Healthineers company, Erlangen, Germany). Physical radiation doses in proton plans were scaled by 1.1 (58–60) to account for relative biological effect compared to photon plans. Photon treatments were generally planned to use two to four volumetric modulated arcs. Proton treatments were planned with a pencil beam scanning method utilizing two to five static fields. From our planning system, we exported DICOM files corresponding to radiation doses, CT scans, and structural sets. Of note, the head and neck anatomy of these data sets were retrospectively segmented consistently by a specially trained team of physicians and medical dosimetry assistant as part of a separate project (61). The contouring followed the consensus guidelines in Brouwer et al. (62).

TABLE 1 List of clinical parameters investigated, and treatment site of the patient cohort considered in this study.

| Clinical variables | | Value |
|---|---|---|
| Baseline dysphagia grade | Grades 0-1 | 208 |
| | Grades 2-3 | 23 |
| Age at RT start | Mean | 60.85 years |
| | Range | [23, 89] years |
| ADI | Mean | 45.01 |
| | Range | [1, 98] |
| Treatment setting | Surgery before RT | 146 |
| | No surgery before RT | 85 |
| Smoker status | Smoker | 120 |
| | Never smoker | 111 |
| Gender | Male | 189 |
| | Female | 42 |
| Radiation type | Photon | 131 |
| | Proton | 100 |
| Concurrent chemotherapy | Yes | 156 |
| | No | 75 |
| Treatment site (% of FT/WL) | Salivary glands | 6 (50%) |
| | Oropharynx | 155 (54%) |
| | Oral cavity | 23 (61%) |
| | Nasopharynx | 12 (75%) |
| | Nasal cavity and sinuses | 15 (33%) |
| | Larynx | 17 (35%) |
| | Hypopharynx | 3 (67%) |

## 2.3 Endpoint definition

The endpoint used to characterize malnutrition was feeding tube (FT) utilization or >10% weight loss from baseline within 90 days after radiotherapy completion. Though FT usage typically arises when a patient sheds over 10% of their initial weight, some

clinicians and/or patients opt against it on a case-by-case basis. Our practice is to only recommend a FT if it is clinically indicated (typically >10% weight loss from baseline). We do not use FT prophylactically. Regardless, such patients remain malnourished, indicating a toxicity affecting their quality of life. We chose to merge these observations and study them as one endpoint; for the remainder of the paper, this endpoint will be referred to as FT/WL for simplicity. The baseline weight was defined as the weight recorded closest to the RT commencement, ensuring it was within a ±2-week window of that date.

## 2.4 Analysis

### 2.4.1 Overview

As illustrated in Figure 1, this analysis had two different methodologies after data processing. In the first methodology, 3D voxel-based dose models were trained to identify regions within the pharyngeal constrictors and larynx that better differentiated toxicity endpoints. Because the voxel-based model utilized a large number of input voxels (22,020 for pharyngeal constrictors and 20,814 for larynx), it was prone to overfitting. To validate and confirm the subregion findings, the second methodology investigated models with reduced features where mean doses from segmented sub-regions of the pharyngeal constrictor muscles and larynx along with clinical variables (14 input features). Similarly, another model that combined DVH metrics used in our clinic with the identical clinical parameters (13 input features) was studied as a comparison.

### 2.4.2 Data preprocessing

All patients' pharyngeal constrictors and larynx were deformably registered to a reference patient's corresponding organs to align them in the same coordinate system. The reference patient shown in Figure 2, was chosen based on having a larynx structure size close to the population average, as well as the absence of any disfigurement due to prior surgery or disease involving the pharyngeal constrictors and/or larynx. Deformable registration was performed using the open-source package, Elastix (63–65). The planning images of most patients included in the study had a native resolution of 1.27 x 1.27 $mm^2$ in the axial direction, with slice thicknesses of either 1 or 2 mm. To maintain uniformity, all images were interpolated to a resolution of 1.27 mm x 1.27 mm x 1 mm in the coronal, sagittal, and axial directions, respectively. The DICOM images (66) were then cropped around organ +4 mm using the open-source packages DicomRTTool and ANTs (67, 68). Rigid, affine, and deformable transformations were subsequently applied. The quality of image registrations was evaluated using the Dice coefficient. For the 3D models, we also augmented the training dataset by flipping the OAR contours and dose left-right to create mirrored dose distributions. This is based on the premise that both the pharyngeal constrictors and larynx are midline structures and there are no laterality preferences for one versus the other. The flipped contour and doses were registered to the reference coordinate system in the same manner as the original data. After the deformable image registrations, the same deformation was applied to each patient's dose matrix to obtain dose within the reference coordinate system. Dose to each voxel was extracted, and

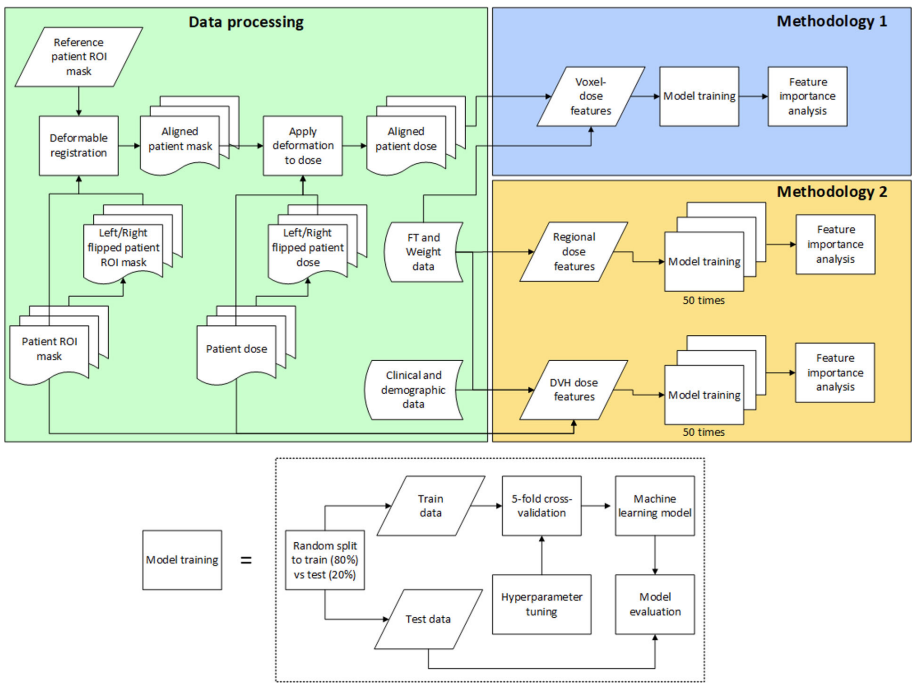



FIGURE 1
Flow diagram of the analysis.

FIGURE 2
Pharyngeal constrictor muscles and larynx of the reference patient.

each numerical input feature was standardized. This standardization process involved adjusting each feature such that it had a mean of zero and a standard deviation of one among the training set. Standardization ensures that our model's performance is not biased by variations in the absolute dose levels, but rather focuses on the relative differences in dose distribution. All categorical features were one-hot encoded and expressed in terms of zeros and ones.

### 2.4.3 Modeling and statistical analysis

We initially evaluated ridge logistic regression, eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) algorithms for modeling. Ridge logistic regression was implemented using the open-source package cuML (69) and scikit-learn (70), while XGBoost (71) and LightGBM (72) models were utilized from their respective repositories. We performed voxel-based analysis using these algorithms with a subset of cohort, and areas under the curves (AUCs) for receiver operating curves (ROCs) were compared. The feature importance maps from XGBoost and lightGBM were sparse and their performance were comparable, so we opted to utilize ridge logistic regression for the remainder of the study.

In Methodology 1, voxel doses were the only input features for the models, with the pharyngeal constrictors and larynx analyzed in separate models. The data set was randomly split into training and test sets in an 80-20% ratio, and the same training patients were used for both pharyngeal constrictors and larynx models. The hyperparameter was tuned using the Optuna package (73) over the 5-fold cross-validation, with ROC AUC as the performance metric. Accuracy and F1 scores evaluated at the threshold of 0.5 were also utilized to compare the general performance of the models. Full training data was used to perform the final fit with the optimized hyperparameter to create the model. For the 3D models, the standardized input features allowed the coefficients from the ridge logistic regression to directly signify feature importance. Specifically, each coefficient from the regression showcases the change in log odds of the outcome for a unit increase in its respective feature. These feature importance coefficients were qualitatively reviewed in a scatter plot for any sub-regions that differentiated FT/WL better than others. In addition to ridge logistic regression, other regularized methods such as LASSO and elastic net were also attempted. However, these L1-based regularization, promoting sparsity combined with highly correlated input parameters yielded spatially fragmented and variable feature importance maps. Therefore, ridge regression was preferred to retain spatial coherence for subregion identification.

In methodology 2, a univariate logistic regression was first performed for each clinical and demographic variable fitting to the binary outcomes of FT/WL, and p-value and odds ratio were calculated. We first developed a model exclusively incorporating the non-dosimetric clinical variables listed in Table 2, which served to establish a baseline for prediction performance. Variables identified as predictive of FT/WL in the univariate analysis were used for modeling with the dosimetric variables. In the final models, there were seven to eight dosimetric variables and six clinical and demographic variables. The clinical and demographic variables were common between the DVH model and the regional dose model: area deprivation index, baseline weight, treatment setting (primary/post-operative), concurrent chemotherapy (yes/no), bilateral treatment (yes/no), and baseline dysphagia grade 0 (yes/no). The dosimetric variables for the DVH metrics model were mean larynx dose, mean pharyngeal constrictor dose, larynx V50Gy and V60Gy, and pharyngeal constrictor V50Gy, V55Gy, and V60Gy, where V$x$Gy represents the percent volume of the organ covered by $x$ Gy or more. These DVH metrics were chosen because they are used to evaluate head and neck treatment plans in our clinic. For the regional dose model, the pharyngeal constrictors were divided into superior, middle and inferior pharyngeal constrictors on the reference patient as well as midline and lateral regions as shown in Figure 3C, grouping the regions based on the feature importance from the voxel models. Following a similar grouping, the larynx was divided into supraglottic larynx and inferior regions as shown in Figure 3F. Mean doses from these subregions were used as the dosimetric variables for the regional dose model. Using these input features, models were trained 50 times with the 5-fold cross-validation process with varying random data splits to confirm that the learned hyperparameters and feature importances were not heavily dependent on a specific random partitioning of the dataset. To assess feature importance, we employed a permutation test (74). We randomized the data of one input variable at a time and evaluated the resulting drop in model performance. The magnitude of performance decline, as measured by the change in ROC AUC from the unaltered data, indicates the importance of that variable to the model's predictive capability. We conducted this permutation process 50 times for every variable, incorporating 5-fold cross-validation on the training dataset in each iteration. The average change in the AUC was used to assess the difference in performance, and the two-tailed Wilcoxon rank sum test was used to evaluate the statistical significance of this change. To compare the general performance

10.3389/fonc.2024.1346797

TABLE 2 P-values and odds ratios from the univariate analysis of non-dosimetric variables correlating with FT/WL.

| Non-dosimetric variables | P-value | Odds Ratios [95% CI] |
|---|---|---|
| **Treatment setting (primary vs post-operative treatment)** | **<0.0001** | **4.62 [2.55 – 8.38]** |
| **Bilateral treatment** | **<0.0001** | **6.72 [3.07 – 14.73]** |
| **Concurrent chemotherapy** | **0.001** | **2.54 [1.44 – 4.49]** |
| **Baseline weight** | **0.01** | **1.014 [1.000-1.028]** |
| **Area deprivation index** | **0.02** | **1.014 [1.002-1.026]** |
| **Baseline dysphagia grade = 0** | **0.03** | **0.51 [0.28 – 0.92]** |
| Gender | 0.16 | 1.63 [0.83-3.20] |
| Treatment modality (photon vs proton) | 0.17 | 0.69 [0.41 – 1.17] |
| Never smoker | 0.34 | 0.78 [0.46 – 1.30] |
| Age | 0.39 | 0.99 [0.97-1.01] |

P-values and odds ratios from the univariate analysis of nondosimetric variables correlating with FT/WL. Variables with p-values below 0.05 (bold) showed a statistically significant correlation with FT/WL placement.

of the regional dose model and the DVH metrics model, test set AUCs from the 50 trials were also compared using the two-tailed Wilcoxon rank sum test.

# 3 Results

Of the 231 patients, 64 patients were found to have utilized FT, and 106 patients lost more than 10% of their baseline weight within 90 days after completing RT. Combining these observations, 122 patients had either FT, WL or both. All registrations in this study yielded a Dice coefficient of greater than 0.7, with a mean and standard deviation of 0.80 ± 0.02 for pharyngeal constrictors and 0.84 ± 0.06 for larynx. Figure 4A, B shows the mean and standard deviation of pharyngeal constrictor dose distributions for the entire cohort. The average mean pharyngeal constrictor dose was 52.3 ± 11.3 Gy and 31.7 ± 15.2 Gy for those who experienced FT/WL and those who did not, respectively. Figure 4C, D show the mean and the standard deviation of larynx dose distributions. Among those who experienced FT/WL, the average mean larynx dose was 41.6 ± 14.4 Gy while that of patients who did not experience FT/WL was 26.5 ± 18.3 Gy.

## 3.1 Voxel-based dose model

Figure 3A shows the ROC curves for the 3D pharyngeal constrictor model. The AUC for the models' training and test data were found to be 0.87 and 0.86, respectively, with cross-validation yielding 0.86 ± 0.02. The accuracy and F1 scores for the test data were 0.83 and 0.85, respectively. Figure 3B shows the feature importance from the model, highlighting the middle to superior pharyngeal constrictors as being more impactful when

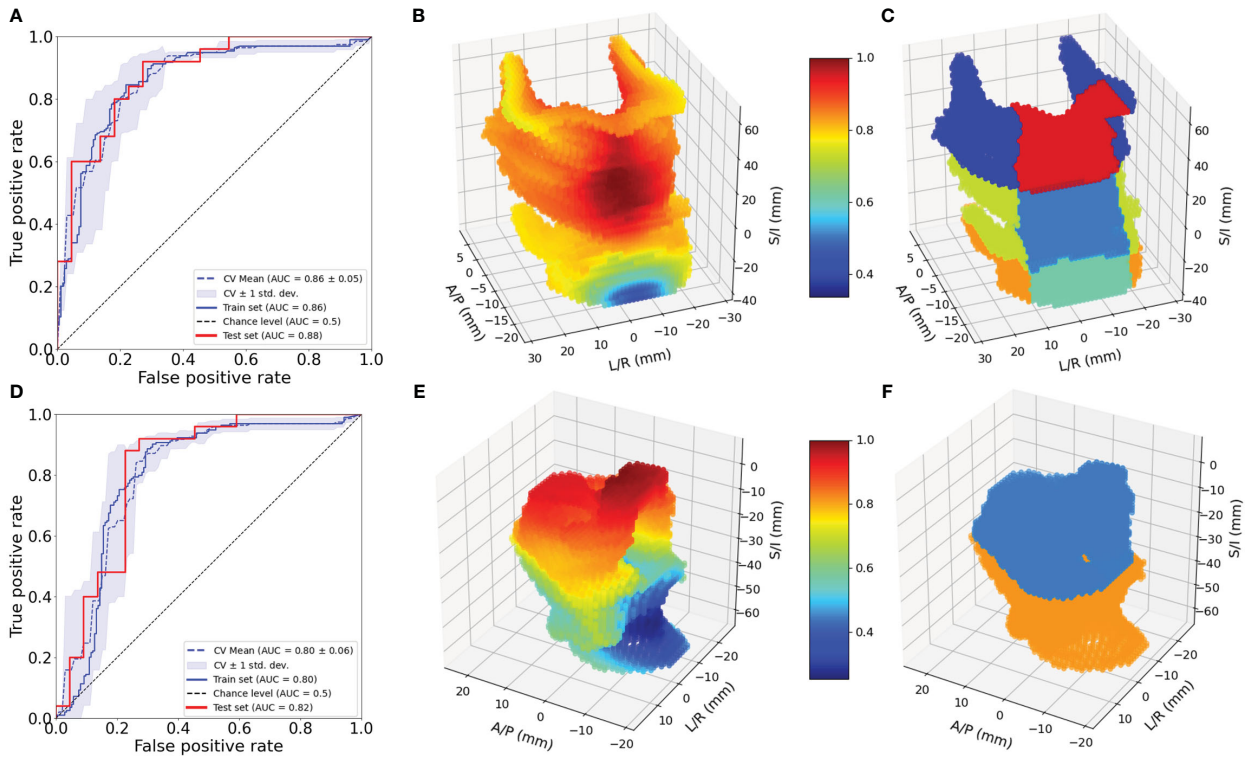



FIGURE 3
Evaluation of 3D models for pharyngeal constrictor and larynx. ROC for pharyngeal constrictor **(A)** and larynx **(D)** 3D models. Feature importance patterns for pharyngeal constrictor **(B)** and larynx **(E)**. Subregions of pharyngeal constrictor **(C)** and larynx **(F)** used for the aggregate models.

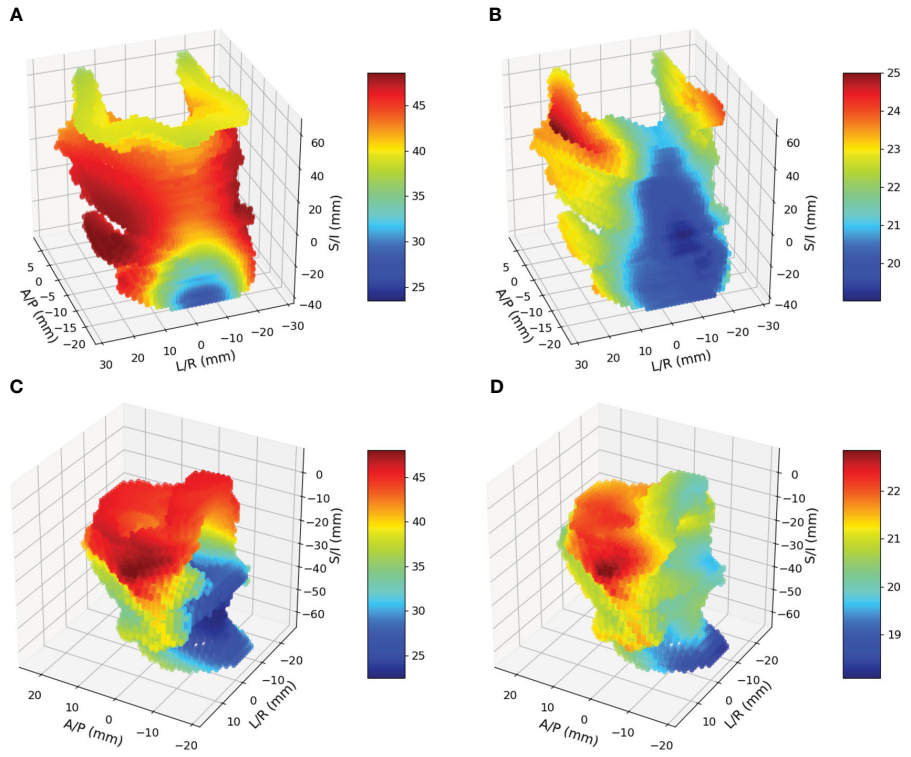



FIGURE 4
**(A)** and **(B)** are mean and standard deviation of pharyngeal constrictor doses. **(C)** and **(D)** are mean and standard deviations of larynx doses.

compared to the inferior region. The midline voxels of the superior and middle pharyngeal constrictors were most important for determining FT/WL for this cohort. Figure 3D, E illustrates ROC and the feature importance pattern from the larynx 3D model. The AUC for the model's training and test data was 0.80 and 0.82, respectively, with cross-validation yielding 0.80 ± 0.06. The accuracy and F1 scores for the test data were both 0.72. The feature importance consistently highlights the superior part of the larynx as a critical factor for predicting FT/WL.

## 3.2 Univariate analysis and aggregated dosimetric model

The results of the univariate fits to non-dosimetric variables are shown in Table 2. There were six variables that showed significant correlation with FT/WL: treatment setting (primary/post-operative), treatment laterality (bilateral/unilateral), concurrent chemotherapy (yes/no), baseline dysphagia grade =0, baseline weight, and ADI. These six non-dosimetric variables were used as part of the aggregated dosimetric models. Gender, treatment modality (photon vs proton), smoking status and age did not show significant correlation with FT/WL. A clinical variables-only model using the non-dosimetric variables in Table 2 achieved an AUC of 0.76 ± 0.01 and 0.74 ± 0.06 for training and test set, respectively. The pharyngeal constrictor and larynx subregions are illustrated in Figure 3C, F, respectively. Test set ROCs from the regional dose and DVH metrics models are shown in Figure 5. The lines indicate the average of the 50 trials, and the shaded area indicates the standard deviations. The regional dose model, with a mean AUC of 0.87 ± 0.05, demonstrated marginally superior performance compared to the DVH metrics model, which had a mean AUC of 0.85 ± 0.05 with a p-value of 0.04. The F1 scores for both models were 0.82 ± 0.04. The incorporation of dosimetric features in the DVH and regional dose models significantly improved performance over the clinical variables-only model ($p < 0.0001$). Table 3 shows the results of a permutation test to evaluate feature importance. Input features that significantly affected the performance of the model when shuffled–as assessed by AUC compared to the actual data–are listed in order of significance. For the regional dose model, the most important features were pharyngeal constrictor doses, particularly the superior to middle region. While dose to the superior part of the larynx significantly contributed to the performance of the regional dose model, none of the larynx DVH metrics showed significant performance gain in the DVH metrics model.

# 4 Discussion

This study delves into the relationship between radiation dose distributions to two organs at risk (OARs) and feeding tube use or weight loss in head and neck cancer patients receiving radiotherapy. Employing deformable image registration and ridge logistic

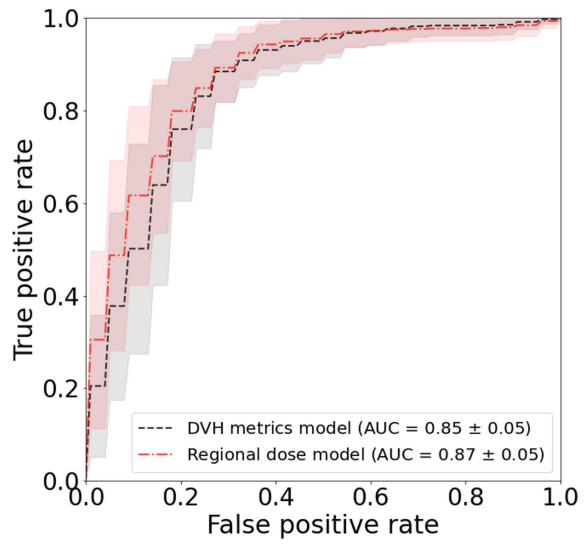



FIGURE 5
Test set ROC for FT/WL prediction using regional dose and dose-volume histogram (DVH) metrics models. Curves show the average area under the ROC curve (AUC) over 50 trials, with shading indicating standard deviation. The regional dose model (mean AUC 0.87 ± 0.05) outperformed the DVH metrics model (0.85 ± 0.05) with p=0.04.

regression, we mapped structures and identified key regions, particularly the superior part of the pharyngeal constrictors and the superior part of the larynx, as crucial determinants for FT/WL. This finding was revealed with 3D voxel-based models and validated using aggregated dose model with reduced input features to minimize risk of overfitting.

Our study highlights that dose to the superior pharyngeal constrictor muscles was most important in predicting FT/WL in our cohort. While we did not find studies investigating the same endpoint, our findings are in general agreement with prior clinical (75) and outcomes studies (43) demonstrating the radiosensitivity of sub-regions in organs and their role in post-radiation dysphagia and aspiration. Feng et al. (76) and Eisbruch et al. (38) also reported that highest correlations of videofluoroscopy based aspiration and dysphagia to the superior pharyngeal constrictor in a prospective study. Petras et al. (77) have evaluated the relationship between dose to larynx subregions and swallowing toxicities assessed by aspiration at one year, and Hedstrom et al. (78) considered dysphagia at 6-months post-treatment. Both studies identified the epiglottis as a critical subregion, in line with our findings.

Overall, the performance of the DVH metrics and the regional dose models were similar. This supports our current standard of care of using DVH for treatment planning. However, we note that feature importance analysis for the DVH metrics model revealed that the model performance did not significantly depend on larynx DVH metrics. On the other hand, the aggregated regional dose model utilized mean doses from superior-middle pharyngeal constrictor and supraglottic larynx. This study motivates further investigation into dose sparing of these subregions. Understanding the dose distribution effects will facilitate voxel-based optimization, evaluation and interpretation of treatment plans that have similar DVH metrics.

The strength of our work includes the number of cohorts, standardized contouring of the pharyngeal constrictors and the larynx, as well as inclusion of patients treated with proton therapy. In contrast to the pharyngeal constrictor and larynx segmentation created during clinical workflow, contours used in this study were retrospectively drawn to achieve high consistency and conformance to the contouring guideline. Consequently, the definition of the organ was consistent across the entire patient cohort. For future studies, automatic segmentation using artificial intelligence-driven algorithms along with quality assurance processes could aid in generating more consistent anatomical segmentations than those available from clinical data. However, automatic segmentation algorithms are typically not trained to segment substructures. As this study demonstrated, a voxel-based approach allows us to eliminate the need for exhaustive contouring of each substructure *a priori* for the entire cohort, thereby streamlining the analytical process.

TABLE 3 Evaluation of feature importance based on a permutation test for the DVH metrics and regional dose models. The table lists input features in descending order of significance, based on their impact on model performance assessed by ROC AUC. (PC: Pharyngeal constrictor muscles).

| | Features from DVH metrics model | Mean AUC loss | Features from regional dose model | Mean AUC loss |
|---|---|---|---|---|
| 1 | PC: Mean dose | -0.021 | PC: superior lateral mean dose | -0.022 |
| 2 | PC: V50Gy (%) | -0.016 | PC: superior midline mean dose | -0.020 |
| 3 | PC: V55Gy (%) | -0.012 | PC: middle midline mean dose | -0.011 |
| 4 | Baseline weight | -0.006 | PC: middle lateral mean dose | -0.009 |
| 5 | PC: V65Gy (%) | -0.006 | Baseline weight | -0.004 |
| 6 | Area Deprivation Index | -0.004 | Larynx: superior mean dose | -0.004 |
| 7 | Concurrent chemotherapy (yes/no) | -0.002 | Area Deprivation Index | -0.003 |
| 8 | Treatment setting | -0.001 | Concurrent chemotherapy (yes/no) | -0.002 |
| 9 | Baseline dysphagia grade = 0 (yes/no) | -0.0006 | Baseline dysphagia grade = 0 (yes/no) | -0.0004 |

Limitations of our work include the nature of single-institution studies and the lack of a treatment planning component to the study, so that it remains to be seen what amount of dose reduction to the mid to superior pharyngeal constrictor and supraglottic larynx could be achieved without compromising treatment quality. An additional limitation is the lack of controlled study on treatment modality (proton/photon). The relative biological effectiveness (RBE) of 1.1 (58–60) for protons used in this study applies to tumor control outcomes, which may not directly translate to functional outcomes in normal tissues. Furthermore, our study included 21 photon patients who received lower prescription doses of 30 Gy. Since our primary objective was to investigate effects of various dose distributions, we opted to include those treatments to increase diversity. However, these choices likely confounded the analysis with respect to treatment modality. While promising, our findings warrant validation in diverse cohorts, treatment modality, treatment planning techniques, and treatment regimens. Nevertheless, our study highlights the potential of interpretable voxel-based modeling to elucidate impact of inhomogeneous dose distributions within an organ.

## 5 Conclusion

In conclusion, the 3D voxel-based analysis and the aggregated regional dose analysis highlighted the superior subregion of the pharyngeal constrictor muscles and the supraglottic larynx as the most important predictor of FT/WL within 90 days of RT.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: protected health information. Requests to access these datasets should be directed to shiraishi.satomi@mayo.edu.

## Ethics statement

The studies involving humans were approved by Mayo Clinic Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

SM: Data curation, Investigation, Methodology, Visualization, Writing – original draft. MG: Writing – review & editing. YG: Writing – review & editing. SL: Writing – review & editing. DJM: Writing – review & editing. DWM: Writing – review & editing. MW: Writing – review & editing. JQ: Methodology, Writing – review & editing. DR: Writing – review & editing. RF: Conceptualization, Data curation, Writing – review & editing. SS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was primarily funded by Varian Medical Systems and partly funded by Mayo Clinic Eagle's Pilot Award.

## Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Caglar HB, Tishler RB, Othus M, Burke E, Li Y, Goguen L, et al. Dose to larynx predicts for swallowing complications after intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. (2008) 72:1110–8. doi: 10.1016/j.ijrobp.2008.02.048

2. Fua TF, Corry J, Milner AD, Cramb J, Walsham SF, Peters LJ. Intensity-modulated radiotherapy for nasopharyngeal carcinoma: clinical correlation of dose to the pharyngo-esophageal axis and dysphagia. *Int J Radiat Oncol Biol Phys*. (2007) 67:976–81. doi: 10.1016/j.ijrobp.2006.10.028

3. Schwartz DL, Hutcheson K, Barringer D, Tucker SL, Kies M, Holsinger FC, et al. Candidate dosimetric predictors of long-term swallowing dysfunction after oropharyngeal intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. (2010) 78:1356–65. doi: 10.1016/j.ijrobp.2009.10.002

4. Humbert-Vidan L, Patel V, Andlauer R, King AP, Guerrero Urbano T. Prediction of mandibular ORN incidence from 3D radiation dose distribution maps using deep learning. *Lecture Notes Comput Sci (including subseries Lecture Notes Artif Intell Lecture Notes Bioinformatics)*. (2022) 13540 LNCS:49–58. doi: 10.1007/978-3-031-17721-7_6/COVER

5. Inada M, Nishimura Y, Ishikura S, Ishikawa K, Murakami N, Kodaira T, et al. Organs-at-risk dose constraints in head and neck intensity-modulated radiation

therapy using a dataset from a multi-institutional clinical trial (JCOG1015A1). *Radiat Oncol*. (2022) 17:133. doi: 10.1186/s13014-022-02105-3

6. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat OncologyBiologyPhys*. (2010) 76:S3–9. doi: 10.1016/j.ijrobp.2009.09.040

7. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys*. (2010) 76:S10–19. doi: 10.1016/j.ijrobp.2009.07.1754

8. Brodin NP, Kabarriti R, Garg MK, Guha C, Tomé WA. Systematic review of normal tissue complication models relevant to standard fractionation radiation therapy of the head and neck region published after the QUANTEC reports. *Int J Radiat OncologyBiologyPhys*. (2018) 100:391–407. doi: 10.1016/j.ijrobp.2017.09.041

9. Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. *Radiother Oncol*. (2012) 105:107–14. doi: 10.1016/j.radonc.2011.08.009

10. Rancati T, Schwarz M, Allen AM, Feng F, Popovtzer A, Mittal B, et al. Radiation dose–volume effects in the larynx and pharynx. *Int J Radiat Oncol Biol Phys*. (2010) 76:S64–9. doi: 10.1016/j.ijrobp.2009.03.079

11. Mavroidis P, Price A, Fried D, Kostich M, Amdur R, Mendenhall W, et al. Dose–volume toxicity modeling for de-intensified chemo-radiation therapy for HPV-positive oropharynx cancer. *Radiother Oncol*. (2017) 124:240–7. doi: 10.1016/j.radonc.2017.06.020

12. Kutcher GJ, Burman C, Brewster L, Goitein M, Mohan R. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. *Int J Radiat Oncol Biol Phys*. (1991) 21:137–46. doi: 10.1016/0360-3016(91)90173-2

13. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl*. (1985) 8:S13–19. doi: 10.2307/3583506

14. Seppenwoolde Y, Lebesque JV, de Jaeger K, Belderbos JSA, Boersma LJ, Schilstra C, et al. Comparing different NTCP models that predict the incidence of radiation pneumonitis. Normal tissue complication probability. *Int J Radiat Oncol Biol Phys*. (2003) 55:724–35. doi: 10.1016/S0360-3016(02)03986-X

15. Samant P, de Ruysscher D, Hoebers F, Canters R, Hall E, Nutting C, et al. Machine learning for normal tissue complication probability prediction: Predictive power with versatility and easy implementation. *Clin Trans Radiat Oncol*. (2023) 39:100595. doi: 10.1016/j.ctro.2023.100595

16. Buettner F, Miah AB, Gulliford SL, Hall E, Harrington KJ, Webb S, et al. Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial. *Radiother Oncol*. (2012) 103:82–7. doi: 10.1016/j.radonc.2012.02.006

17. Cella L, Monti S, Thor M, Rimner A, Deasy JO, Palma G. Radiation-induced dyspnea in lung cancer patients treated with stereotactic body radiation therapy. *Cancers*. (2021) 13:3734. doi: 10.3390/cancers13153734

18. Dohopolski M, Wang K, Morgan H, Sher D, Wang J. Use of deep learning to predict the need for aggressive nutritional supplementation during head and neck radiotherapy. *Radiother Oncol*. (2022) 171:129–38. doi: 10.1016/j.radonc.2022.04.016

19. Ebert MA, Gulliford S, Acosta O, de Crevoisier R, McNutt T, Heemsbergen WD, et al. Spatial descriptions of radiotherapy dose: normal tissue complication models and statistical associations. *Phys Med Biol*. (2021) 66:12TR01. doi: 10.1088/1361-6560/ac0681

20. Guo Y, Jiang W, Lakshminarayanan P, Han P, Cheng Z, Bowers M, et al. Spatial radiation dose influence on xerostomia recovery and its comparison to acute incidence in patients with head and neck cancer. *Adv Radiat Oncol*. (2020) 5:221–30. doi: 10.1016/j.adro.2019.08.009

21. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine learning-based models for prediction of toxicity outcomes in radiotherapy. *Front Oncol*. (2020) 10:790. doi: 10.3389/fonc.2020.00790

22. Jiang W, Lakshminarayanan P, Hui X, Han P, Cheng Z, Bowers M, et al. Machine learning methods uncover radiomorphologic dose patterns in salivary glands that predict xerostomia in patients with head and neck cancer. *Adv Radiat Oncol*. (2019) 4:401–12. doi: 10.1016/j.adro.2018.11.008

23. Kobayashi K, Murakami N, Takahashi K, Inaba K, Igaki H, Hamamoto R, et al. A population-based statistical model for investigating heterogeneous intraprostatic sensitivity to radiation toxicity after 125I Seed Implantation. *In Vivo*. (2019) 33:2103–11. doi: 10.21873/invivo.11710

24. Murr M, Brock KK, Fusella M, Hardcastle N, Hussein M, Jameson MG, et al. Applicability and usage of dose mapping/accumulation in radiotherapy. *Radiother Oncol*. (2023) 182:109527. doi: 10.1016/j.radonc.2023.109527

25. Palma G, Monti S, Conson M, Pacelli R, Cella L. Normal tissue complication probability (NTCP) models for modern radiation therapy. *Semin Oncol*. (2019) 46:210–8. doi: 10.1053/j.seminoncol.2019.07.006

26. Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: A methodological cookbook. *Physica Med*. (2020) 69:192–204. doi: 10.1016/j.ejmp.2019.12.013

27. Voshart DC, Wiedemann J, van Luijk P, Barazzuol L. Regional responses in radiation-induced normal tissue damage. *Cancers*. (2021) 13:1–21. doi: 10.3390/CANCERS13030367

28. Bourbonne V, Lucia F, Jaouen V, Bert J, Rehn M, Pradier O, et al. Development and prospective validation of a spatial dose pattern based model predicting acute pulmonary toxicity in patients treated with volumetric arc-therapy for locally advanced lung cancer. *Radiother Oncol*. (2021) 164:43–9. doi: 10.1016/j.radonc.2021.09.008

29. Bourbonne V, Lucia F, Jaouen V, Pradier O, Visvikis D, Schick U. Combination of radiomics features and functional radiosensitivity enhances prediction of acute pulmonary toxicity in a prospective validation cohort of patients with a locally advanced lung cancer treated with VMAT-radiotherapy. *J Personal Med*. (2022) 12:1926. doi: 10.3390/jpm12111926

30. Cella L, Monti S, Xu T, Liuzzi R, Stanzione A, Durante M, et al. Probing thoracic dose patterns associated to pericardial effusion and mortality in patients treated with photons and protons for locally advanced non-small-cell lung cancer. *Radiother Oncol*. (2021) 160:148–58. doi: 10.1016/j.radonc.2021.04.025

31. Heemsbergen WD, Incrocci L, Pos FJ, Heijmen BJM, Witte MG. Local dose effects for late gastrointestinal toxicity after hypofractionated and conventionally fractionated modern radiotherapy for prostate cancer in the HYPRO trial. *Front Oncol*. (2020) 10:469. doi: 10.3389/fonc.2020.00469

32. Monti S, Xu T, Liao Z, Mohan R, Cella L, Palma G. On the interplay between dosiomics and genomics in radiation-induced lymphopenia of lung cancer patients. *Radiother Oncol*. (2022) 167:219–25. doi: 10.1016/j.radonc.2021.12.038

33. Monti S, Xu T, Mohan R, Liao Z, Palma G, Cella L. Radiation-induced esophagitis in non-small-cell lung cancer patients: voxel-based analysis and NTCP modeling. *Cancers*. (2022) 14:1833. doi: 10.3390/cancers14071833

34. Palma G, Monti S, Pacelli R, Liao Z, Deasy JO, Mohan R, et al. Radiation pneumonitis in thoracic cancer patients: multi-center voxel-based analysis. *Cancers*. (2021) 13:3553. doi: 10.3390/cancers13143553

35. Shelley LEA, Sutcliffe MPF, Thomas SJ, Noble DJ, Romanchikova M, Harrison K, et al. Associations between voxel-level accumulated dose and rectal toxicity in prostate radiotherapy. *Phys Imaging Radiat Oncol*. (2020) 14:87–94. doi: 10.1016/j.phro.2020.05.006

36. Veiga C, Lim P, Anaya VM, Chandy E, Ahmad R, D'Souza D, et al. Atlas construction and spatial normalisation to facilitate radiation-induced late effects research in childhood cancer. *Phys Med Biol*. (2021) 66:105005. doi: 10.1088/1361-6560/abf010

37. Wilson LJ, Bryce-Atkinson A, Green A, Li Y, Merchant TE, van Herk M, et al. Image-based data mining applies to data collected from children. *Physica Med*. (2022) 99:31–43. doi: 10.1016/j.ejmp.2022.05.003

38. Eisbruch A, Kim HM, Feng FY, Lyden TH, Haxer MJ, Feng M, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. *Int J Radiat Oncol Biol Phys*. (2011) 81:e93–99. doi: 10.1016/j.ijrobp.2010.12.067

39. Mai Y, Kong F, Yang Y, Zhou L, Li Y, Song T. Voxel-based automatic multi-criteria optimization for intensity modulated radiation therapy. *Radiat Oncol*. (2018) 13:241. doi: 10.1186/s13014-018-1179-7

40. Cotrutz C, Xing L. Using voxel-dependent importance factors for interactive DVH-based dose optimization. *Phys Med Biol*. (2002) 47:1659–69. doi: 10.1088/0031-9155/47/10/304

41. Eisbruch A, Schwartz M, Rasch C, Vineberg K, Damen E, Van As CJ, et al. Dysphagia and aspiration after chemoradiotherapy for head-and-neck cancer: Which anatomic structures are affected and can they be spared by IMRT? *Int J Radiat OncologyBiologyPhys*. (2004) 60:1425–39. doi: 10.1016/j.ijrobp.2004.05.050

42. Desideri I, Loi M, Francolini G, Becherini C, Livi L, Bonomo P. Application of radiomics for the prediction of radiation-induced toxicity in the IMRT era: current state-of-the-art. *Front Oncol*. (2020) 10:1708. doi: 10.3389/fonc.2020.01708

43. Pepin MD, Anaya S, Foote RL, Garces Y, Hosfield EA, Mundy D, et al. Modeling patient-reported swallowing outcomes with dose to the epiglottis. *Int J Radiat OncologyBiologyPhys*. (2021) 111:e377. doi: 10.1016/j.ijrobp.2021.07.1108

44. Apolle R, Appold S, Bijl HP, Blanchard P, Bussink J, Faivre-Finn C, et al. Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. *Acta Oncol*. (2019) 58:1378–85. doi: 10.1080/0284186X.2019.1629017

45. Petkar I, McQuaid D, Dunlop A, Tyler J, Hall E, Nutting C. Inter-observer variation in delineating the pharyngeal constrictor muscle as organ at risk in radiotherapy for head and neck cancer. *Front Oncol*. (2021) 11:644767. doi: 10.3389/fonc.2021.644767

46. Rutter CE, Yovino S, Taylor R, Wolf J, Cullen KJ, Ord R, et al. Impact of early percutaneous endoscopic gastrostomy tube placement on nutritional status and hospitalization in patients with head and neck cancer receiving definitive chemoradiation therapy. *Head Neck*. (2011) 33:1441–7. doi: 10.1002/hed.21624

47. Madhoun MF, Blankenship MM, Blankenship DM, Krempl GA, Tierney WM. Prophylactic PEG placement in head and neck cancer: How many feeding tubes are unused (and unnecessary)? *World J Gastroenterol*. (2011) 17:1004–8. doi: 10.3748/wjg.v17.i8.1004

48. Grant DG, Bradley PT, Pothier DD, Bailey D, Caldera S, Baldwin DL, et al. Complications following gastrostomy tube insertion in patients with head and neck cancer: a prospective multi-institution study, systematic review and meta-analysis. *Clin Otolaryngol*. (2009) 34:103–12. doi: 10.1111/j.1749-4486.2009.01889.x

49. Langmore S, Krisciunas GP, Miloro KV, Evans SR, Cheng DM. Does PEG use cause dysphagia in head and neck cancer patients? *Dysphagia*. (2012) 27:251–9. doi: 10.1007/s00455-011-9360-2

50. Beaver ME, Matheny KE, Roberts DB, Myers JN. Predictors of weight loss during radiation therapy. *Otolaryngology–Head Neck Surg*. (2001) 125:645–8. doi: 10.1067/mhn.2001.120428

51. Cady J. Nutritional support during radiotherapy for head and neck cancer: the role of prophylactic feeding tube placement. *Clin J Oncol Nurs*. (2007) 11:875–80. doi: 10.1188/07.CJON.875-880

52. Whitaker TJ, Mayo CS, Ma DJ, Haddock MG, Miller RC, Corbin KS, et al. Data collection of patient outcomes: one institution's experience. *J Radiat Res*. (2018) 59:i19–24. doi: 10.1093/jrr/rry013

53. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. *Med Care*. (1997) 35(6):589–602. doi: 10.1097/00005650-199706000-00005

54. Hedegaard H, Johnson RL, Garnett M, Thomas KE. The International classification of diseases, 10th revision, clinical modification (ICD–10–CM): external cause-of-injury framework for categorizing mechanism and intent of injury. (2019) (136):1–22.

55. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible — The neighborhood atlas. *N Engl J Med*. (2018) 378:2456–8. doi: 10.1056/NEJMp1802313

56. Institute NC. Common Terminology Criteria for Adverse Events (CTCAE) (2021). Available online at: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm.

57. Eisbruch A, Lyden T, Bradford CR, Dawson LA, Haxer MJ, Miller AE, et al. Objective assessment of swallowing dysfunction and aspiration after radiation concurrent with chemotherapy for head-and-neck cancer. *Int J Radiat OncologyBiologyPhys*. (2002) 53:23–8. doi: 10.1016/S0360-3016(02)02712-8

58. Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, et al. Relative biological effectiveness (RBE) values for proton beam therapy. *Int J Radiat OncologyBiologyPhys*. (2002) 53:407–21. doi: 10.1016/S0360-3016(02)02754-2

59. Paganetti H. Relating proton treatments to photon treatments via the relative biological effectiveness—Should we revise current clinical practice? *Int J Radiat Oncol Biol Phys*. (2015) 91:892–4. doi: 10.1016/j.ijrobp.2014.11.021

60. Paganetti H, Blakely E, Carabe-Fernandez A, Carlson DJ, Das IJ, Dong L, et al. Report of the AAPM TG-256 on the relative biological effectiveness of proton beams in radiation therapy. *Med Phys*. (2019) 46:e53–78. doi: 10.1002/mp.13390

61. Tryggestad E, Anand A, Beltran C, Brooks J, Cimmiyotti J, Grimaldi N, et al. Scalable radiotherapy data curation infrastructure for deep-learning based autosegmentation of organs-at-risk: A case study in head and neck cancer. *Front Oncol*. (2022) 12:936134. doi: 10.3389/fonc.2022.936134

62. Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, 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. doi: 10.1016/j.radonc.2015.07.041

63. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. (2010) 29:196–205. doi: 10.1109/TMI.2009.2035616

64. Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M, et al. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front neuroinform*. (2014) 7:50. doi: 10.3389/fninf.2013.00050

65. Marstal K, Berendsen F, Staring M, Klein S. (2016). SimpleElastix: A user-friendly, multi-lingual library for medical image registration, *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 574–82.

66. Anderson BM, Wahid KA, Brock KK. Simple python module for conversions between DICOM images and radiation therapy structures, masks, and prediction arrays. *Pract Radiat Oncol*. (2021) 11:226–9. doi: 10.1016/j.prro.2021.02.003

67. Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Sci Rep*. (2021) 11:9068. doi: 10.1038/s41598-021-87564-6

68. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. (2011) 54:2033–44. doi: 10.1016/j.neuroimage.2010.09.025

69. Raschka S, Patterson J, Nolet C. Machine learning in python: main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*. (2020) 11:193. doi: 10.3390/info11040193

70. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in {P}ython. *J Mach Learn Res*. (2011) 12:2825–30.

71. Chen T, Guestrin C. (2016). XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, New York, NY, USA: Association for Computing Machinery. pp. 785–94. doi: 10.1145/2939672.2939785

72. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. (2017) 30:3149–57.

73. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. (2019). Optuna: A next-generation hyperparameter optimization framework, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA. pp. 2623–31. ACM. doi: 10.1145/3292500.3330701

74. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

75. Feng FY, Kim HM, Lyden TH, Haxer MJ, Worden FP, Feng M, et al. Intensity-modulated chemoradiotherapy aiming to reduce dysphagia in patients with oropharyngeal cancer: clinical and functional results. *J Clin Oncol*. (2010) 28:2732–8. doi: 10.1200/JCO.2009.24.6199

76. Feng FY, Kim HM, Lyden TH, Haxer MJ, Feng M, Worden FP, et al. Intensity-modulated radiotherapy of head and neck cancer aiming to reduce dysphagia: early dose-effect relationships for the swallowing structures. *Int J Radiat Oncol Biol Phys*. (2007) 68:1289–98. doi: 10.1016/j.ijrobp.2007.02.049

77. Petras KG, Rademaker AW, Refaat T, Choi M, Thomas TO, Pauloski BR, et al. Dose-volume relationship for laryngeal substructures and aspiration in patients with locally advanced head-and-neck cancer. *Radiat Oncol*. (2019) 14:49. doi: 10.1186/s13014-019-1247-7

78. Hedström J, Tuomi L, Finizia C, Olsson C. Identifying organs at risk for radiation-induced late dysphagia in head and neck cancer patients. *Clin Transl Radiat Oncol*. (2019) 19:87–95. doi: 10.1016/j.ctro.2019.08.005

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 29 February 2024
DOI 10.3389/fonc.2024.1295251



OPEN ACCESS

EDITED BY
Xinyuan Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Ryan T. Hughes,
Wake Forest University, United States
Marina Romanchikova,
National Physical Laboratory, United Kingdom
Yinglin Peng,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Douglas Moseley
Moseley.Douglas@mayo.edu

RECEIVED 15 September 2023
ACCEPTED 05 February 2024
PUBLISHED 29 February 2024

CITATION
Brooks J, Tryggestad E, Anand A, Beltran C, Foote R, Lucido JJ, Laack NN, Routman D, Patel SH, Seetamsetty S and Moseley D (2024) Knowledge-based quality assurance of a comprehensive set of organ at risk contours for head and neck radiotherapy. *Front. Oncol.* 14:1295251. doi: 10.3389/fonc.2024.1295251



# Knowledge-based quality assurance of a comprehensive set of organ at risk contours for head and neck radiotherapy

Jamison Brooks[1], Erik Tryggestad[1], Aman Anand[2], Chris Beltran[3], Robert Foote[1], J. John Lucido[1], Nadia N. Laack[1], David Routman[1], Samir H. Patel[2], Srinivas Seetamsetty[1] and Douglas Moseley[1]*

[1]Department of Radiation Oncology, Mayo Clinic Rochester, Rochester, MN, United States, [2]Department of Radiation Oncology, Mayo Clinic Arizona, Phoenix, AZ, United States, [3]Department of Radiation Oncology, Mayo Clinic Florida, Jacksonville, FL, United States

**Introduction:** Manual review of organ at risk (OAR) contours is crucial for creating safe radiotherapy plans but can be time-consuming and error prone. Statistical and deep learning models show the potential to automatically detect improper contours by identifying outliers using large sets of acceptable data (knowledge-based outlier detection) and may be able to assist human reviewers during review of OAR contours.

**Methods:** This study developed an automated knowledge-based outlier detection method and assessed its ability to detect erroneous contours for all common head and neck (HN) OAR types used clinically at our institution. We utilized 490 accurate CT-based HN structure sets from unique patients, each with forty-two HN OAR contours when anatomically present. The structure sets were distributed as 80% for training, 10% for validation, and 10% for testing. In addition, 190 and 37 simulated contours containing errors were added to the validation and test sets, respectively. Single-contour features, including location, shape, orientation, volume, and CT number, were used to train three single-contour feature models (z-score, Mahalanobis distance [MD], and autoencoder [AE]). Additionally, a novel contour-to-contour relationship (CCR) model was trained using the minimum distance and volumetric overlap between pairs of OAR contours to quantify overlap and separation. Inferences from single-contour feature models were combined with the CCR model inferences and inferences evaluating the number of disconnected parts in a single contour and then compared.

**Results:** In the test dataset, before combination with the CCR model, the area under the curve values were 0.922/0.939/0.939 for the z-score, MD, and AE models respectively for all contours. After combination with CCR model inferences, the z-score, MD, and AE had sensitivities of 0.838/0.892/0.865, specificities of 0.922/0.907/0.887, and balanced accuracies (BA) of 0.880/0.900/0.876 respectively. In the validation dataset, with similar overall performance and no signs of overfitting, model performance for individual

OAR types was assessed. The combined AE model demonstrated minimum, median, and maximum BAs of 0.729, 0.908, and 0.980 across OAR types.

**Discussion:** Our novel knowledge-based method combines models utilizing single-contour and CCR features to effectively detect erroneous OAR contours across a comprehensive set of 42 clinically used OAR types for HN radiotherapy.

KEYWORDS

**contour review, quality assurance, automation, radiotherapy, outlier detection**

# 1 Introduction

Standardized and precise organ at risk (OAR) contours are essential for head and neck (HN) radiation therapy, enabling safe treatments and more consistent dose reporting (1). While manual contouring is time-consuming and prone to user variation, Deep learning (DL) autocontouring methods have demonstrated time savings (2, 3) and reduced variation (4, 5) compared to manual contouring methods. Autocontouring tools generally perform well, however, a variety of clinically relevant failures, ranging from minor to severe, do occur with no warning given from the model-hosting tool (Supplementary Figure S1). Consequently, both contours created manually and with autocontouring tools require thorough quality assurance (QA) review by trained personnel to ensure safe and effective radiotherapy treatments.

The ability of DL autocontouring tools to quickly create many contours enables more contours to be used for a given treatment site and expedites both offline and online adaptive treatment planning. However, it also increases the amount of time spent reviewing contours. Automated approaches to contour review may be able to both decrease review time and improve consistency (6, 7), making them a desirable potential tool for clinical use. Such approaches could be deployed on their own, or in combination with human reviewers to assist them in identifying contours of poor quality.

Several automated algorithmic methods have been proposed for automated OAR contour QA (8–11). One of the most popular approaches utilizes a set of features calculated from high-quality contours to classify contours of unknown quality as similar (acceptable) or different (erroneous). This is referred to as knowledge-based outlier detection using one-class training. Features for this approach include contour volume, shape, orientation, position, and image characteristics. Models for knowledge-based contour classification include statistical approaches looking at several features independent of one another (univariate statistical models) (12, 13), as well as multivariate statistical models, and DL models (14, 15). Most knowledge-based outlier detection methods for OAR contour QA have relied on a few hand-selected features for evaluation which are largely informed by domain experts in radiation oncology. This expertise may allow for comparable performance between simpler statistical models and DL models. Despite several publications, it remains unclear how the performance of univariate models, multivariate statistical models, and DL models compares for knowledge-based OAR contour QA.

In previous studies, knowledge-based contour outlier detection models have used features describing the relationships between different OAR types (henceforth referred to as contour-to-contour relationships or CCRs) to minimize patient-to-patient variation and detect erroneous contours (14–16). Ensuring that contours are appropriately separate, touching, or overlapping is crucial for HN radiation treatment planning due to the precise relationships between many OARs. Neglecting to do so can lead to inaccurately contoured anatomy and unreported dose to OARs because of contour gaps between anatomically touching OARs during IMRT optimization. While CCR relationships are both quantifiable and important, we are not aware of any studies that have directly evaluated the effectiveness of features that quantify contour separation and overlap for the detection of erroneous contours.

To ensure the usefulness of an automated OAR contour quality assurance tool for a specific treatment site, it ideally should have acceptable performance that generalizes to many OAR types (brain, left lung, larynx, etc.) and should encompass various disease types, and patient anatomies. For HN treatment sites, as many as 42 OAR types have been reported to be relevant for HN treatment planning (3). However, existing knowledge-based contour QA studies that have evaluated individual OAR types, assess no more than 17 in any given study (9, 10, 12, 14, 16). This limitation may be attributed to the lack of standardized and curated contours available for model training. Analysis of additional OAR types for HN is needed to demonstrate whether knowledge-based contour outlier detection models can be used for any clinically relevant HN OAR types.

This study investigates the performance and generalizability of knowledge-based, outlier detection methods to identify erroneous contours for 42 HN OAR types used clinically for radiotherapy. This is the largest number of OAR types evaluated for HN in a single study to date. Model training was performed using manually contoured, highly curated, contour sets derived from patients with HN cancer being treated with radiotherapy. Three single-contour feature model types that have not been compared for contour outlier detection in previous work, a univariate statistical model

(z-score) (12), a multivariate statistical Mahalanobis distance (MD) (17, 18) model, and a DL autoencoder (AE) model (19), are compared to identify the model type with the best performance and generalizability to each HN OAR type. As a secondary aim, the study investigates the potential of a novel CCR model, that assesses contour separation and overlap, in combination with the three compared models to enhance performance.

# 2 Materials and methods

## 2.1 Data curation and allocation

The study utilized retrospectively collected data from patients with HN cancers who underwent radiotherapy at Mayo Clinic Rochester and Mayo Clinic Arizona between 2016 and 2020. The dataset encompassed a diverse range of HN disease sites and progressions, including patients with prior resection, representing the current treatment landscape at the institutions. CT images were acquired at simulation before the start of radiotherapy treatment using multiple Somatom Definition AS (Siemens, Munich Germany) CT scanners with voxel dimensions of 1.27 mm x 1.27 mm x 2 mm. The CT images were acquired at 120 kVp and most were reconstructed using iterative metal artifact reconstruction techniques to minimize artifacts caused by dental fillings or other metallic objects commonly present during HN radiotherapy. All CT scanners underwent monthly testing using a CatPhan® phantom *(Phantom Laboratory, Salem New York)* to ensure Hounsfield Unit accuracy (Supplementary Methods and Supplementary Table S1). Head and neck planning CT images and contours used for patient treatment were retrospectively selected and curated to ensure they adhered to institutional guidelines for standardization. This included physician, dosimetrist and physics review and editing during retrospective curation. A thorough description of the dataset and curation efforts has been published (20). The dataset, comprising 490 patient structure sets with corresponding CT images, was considered the gold-standard acceptable patient dataset. These sets were divided into training (80%), validation (10%), and test (10%) subsets. The use of retrospective HN patient data for model training was deemed exempt by our institutional IRB.

Before assessing the performance and generalizability of knowledge-based outlier detection methods to detect erroneous contours, it is essential that the erroneous contours evaluated reflect errors that commonly result from clinical failures. Such errors can occur from both manually created contours, or contours generated using autosegmentation tools. We identified four main categories of such failures that occur: boundary errors, volume errors, non-adjacent slice errors, and positional errors.

Boundary errors encompass instances of accidental border expansion or subtraction, poor delineation of anatomical boundaries, incorrect identification of boundaries based on HU-intensity thresholding, or incorrect propagation of contours from one image set to another due to small deformable or rigid image registration errors.

Volume errors encompass the addition or removal of volumes from an OARs correct volumes. Unlike boundary errors, which pertain to inaccuracies in contour borders, volume errors result from the addition or subtraction of convex shapes from the correct contour. These errors can occur due to incorrect definitions of anatomical boundaries, incomplete contouring, disconnected volumes, or the improper identification of the slices where a contour should start and end.

Non-adjacent slice errors occur due to inadvertent selections of single-slice volumes (i.e. 'misclicks') or inconsistent and 'jagged' delineations of contour boundaries from one slice to another, which may occur during contouring.

Positional errors represent errors resulting in the central location of a contour being substantially misplaced. Such errors arise from mismatched structure labels, errors in manual identification of OARs, or errors made by CT autocontouring tools. Such errors made by autocontouring tools have been observed during the analysis of CT images with abnormal anatomy, positioning, or CT values for several FDA-approved autocontouring tools evaluated by the authors.

## 2.2 Data augmentation

After identifying these common clinical failure modes, manually generated erroneous contours were introduced *ad hoc* by a medical physicist (JB) who edited gold-standard acceptable contours to mimic clinically observed errors encountered during both manual contouring and autocontouring processes. Manually generated erroneous contours were added directly to the validation and test sets after creation. Each erroneous contour error was additionally categorized as moderate or major by the contour editor, providing the ability to assess how the clinical severity of errors influenced the performance of automated outlier detection. Errors categorized as moderate may or may not be clinically relevant depending on clinical context such as the treatment planning approach and the relationship with the target, while errors categorized as major would be relevant in nearly all clinical contexts.

The total number of contours with boundary, volume, non-adjacent slice, and position errors were 74, 99, 14, and 40 respectively. The total number of contours with major and minor errors were 111 and 116. Error types were distributed randomly across OAR types. In the validation set, a minimum of four erroneous contours were created for each OAR type that had left or right counterparts (i.e. left and right lung), and a minimum of five erroneous contours were created for all other OAR types (i.e. brain, larynx, etc.).

## 2.3 Overview of knowledge-based QA framework

In this study, the knowledge-based QA framework was developed through several steps (Figures 1A, B). First, the training dataset, which consisted of acceptable contours only, was

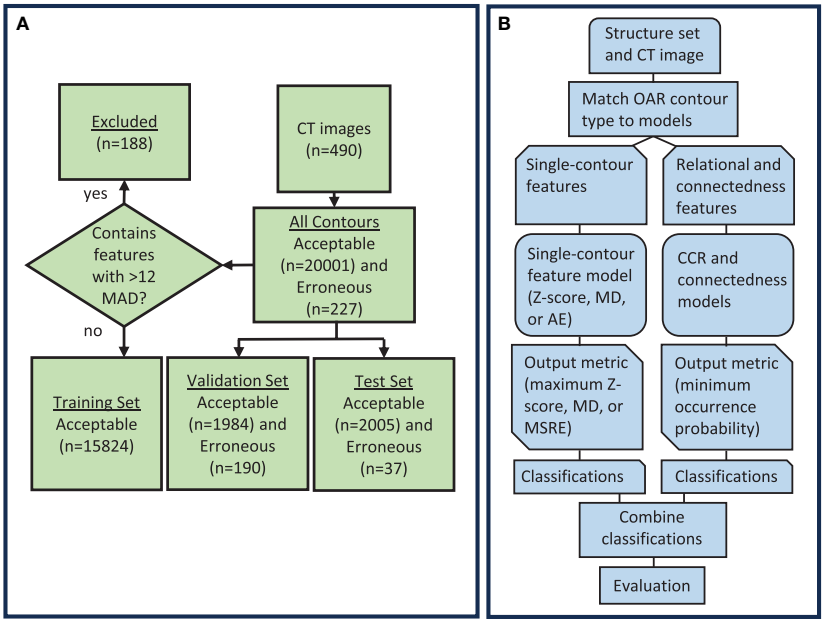



FIGURE 1
**(A)** Diagram showing the distribution of data used to create and evaluate the knowledge-based QA framework. **(B)** Workflow diagram for the knowledge-based QA framework. AE, Autoencoder; MD, Mahalanobis Distance; CCR, contour-to-contour relationship; MSRE, mean squared reconstruction error.

separated by OAR type (i.e. brain, left eye, esophagus). Then, the desired features were calculated for each contour, and model training was performed. Features dependent on only one contour were assigned to each of the three single-contour feature model types (AE, MD, and Z-score), while features involving the relationship between two contours were assigned to the CCR model type. A feature counting the number of disconnected parts in a single contour was included as an additional statistical check outside of the other models as the connectedness model. Separate single-contour feature, CCR, and connectedness models were trained for each OAR type. All models generated a single output metric that indicated the likelihood of a contour being erroneous. Output metrics from models of the same model type were thresholded using a single value to obtain classifications. After model training, the validation set, which included both acceptable and erroneous contours, was used to evaluate the model's performance, select input features, and determine output metric thresholds. Classifications obtained from the single-contour feature models, CCR models, and connectedness models were combined to form the final classification. Lastly, to ensure no overfitting, the models were evaluated on the test set.

## 2.4 Single-contour feature selection

Feature selection for the single-contour feature models was performed initially by choosing features describing contour shape, volume, location, orientation, and CT number. Features were selected to be generalizable to a wide variety of OAR types and were based on common features used in the literature (12–16). A total of 44 features were included and the Pearson correlation coefficient (21) was used to identify and remove features that were strongly correlated, either positively or negatively, across all OARs (Supplementary Figure S2). This was when the correlation was greater than approximately ±0.7. Feature reduction was performed using the validation set to reduce feature correlation while maintaining high classification performance for single-contour feature models. After single-contour feature determination, the same feature set was used for all OAR types and model types (Table 1). This was done to identify a set of features that would generalize well to a wide variety of OARs.

The centroid features in the lateral, vertical, and longitudinal directions (defined as positive x, y, and z respectively) were calculated as the difference between the contour's centroid and the brainstem's centroid. For brainstem contours, it was calculated as the difference in centroid locations between the brainstem contour and the pituitary contour. This accounted for variations in image coordinates between CT images. The brainstem contour was chosen because of its central location and because it is anatomically present in every patient. The extent in x, y, and z was calculated as the difference between the largest and smallest pixel coordinate values for a given contour. Principal component analysis was performed to obtain the eigenvectors and eigenvalues of the principal components (PC) of a contour's pixel coordinates. The x, y, and z components of the first and second PC eigenvectors were used as orientation features, while the ratio of the second and third PC eigenvalues (λ) to the first were used as shape features.

The orientation of a PC eigenvector can be arbitrarily positive or negative (for example $\vec{v} = +0.58\,\hat{x} + 0.58\,\hat{y} + 0.58\,\hat{z}$ or $\vec{v} = -0.58\,\hat{x} - 0.58\,\hat{y} - 0.58\,\hat{z}$). To standardize the orientation of PC1 or PC2 vectors for a given OAR, we identified a representative eigenvector $\vec{r}_{OAR}$ from the training set using Equation 1.

TABLE 1 A list of features used for each model.

| Single-contour feature models (Z-score, MD, AE) | | | | | Connectedness model | CCR model |
|---|---|---|---|---|---|---|
| Location features | Orientation features {0<x<1} | Volume features | Shape features | CT number features | Missing slices or 'ditzel' features | Relational features |
| Centroid X [mm] | $PC1_{\hat{x}}$ | Volume [cc] | $\lambda_{PC2}/\lambda_{PC1}$ *{0<x<1}* | CT minimum | Number of disconnected parts | Minimum distance [mm] *{0<x<∞}* |
| Centroid Y [mm] | $PC1_{\hat{y}}$ | | $\lambda_{PC3}/\lambda_{PC1}$ *{0<x<1}* | CT maximum | | Fractional volume overlap *{0<x<1}* |
| Centroid Z [mm] | $PC1_{\hat{z}}$ | | X extent (mm) | CT mean | | |
| | $PC2_{\hat{x}}$ | | Y extent (mm) | CT std. dev. | | |
| | $PC2_{\hat{y}}$ | | Z extent (mm) | | | |
| | $PC2_{\hat{z}}$ | | Sphericity *{0<x<1}* | | | |

$$\vec{r}_{OAR} = \underset{\vec{v}_i \in V_{OAR}}{\operatorname{argmax}} \left( \sum_{j \neq i}^{n} \left| \vec{v}_i \cdot \vec{v}_j \right| \right) \quad (1)$$

Where $V_{OAR}$ is the set of all PC1 or PC2 eigenvectors for a given OAR type in the training set with number n and $\vec{v}$ is a single PC eigenvector. After identification of $\vec{r}_{OAR}$, the orientation of all eigenvectors in the training, validation, and test set ($V_{OAR}{}^{*}$) were oriented either positive or negative to maximize the dot product between $\vec{r}_{OAR}$ and each eigenvector $\vec{v} \in V_{OAR}{}^{*}$.

Since outlier detection approaches that use one-class training are sensitive to outliers in the training dataset, an outlier removal technique was applied to the training dataset after feature selection. To do this, the training dataset consisting of only acceptable contours was grouped based on its OAR type, and the median absolute deviation (MAD) from the median was calculated for single-contour features. A contour was excluded from the training set if any of its single-contour features deviated from the median by more than twelve MAD. This resulted in the removal of 0% to 3.6% of contours from the training set for each OAR type. The threshold of twelve MAD was determined by evaluating the number of contours removed for each OAR type and the impact of contour removal on model performance for the validation dataset.

## 2.5 Single-contour feature models

After single-contour feature calculation using acceptable contours in the training dataset, an individual model was trained for each OAR type for three single-contour feature model types (z-score, MD, and AE models). The z-score model calculated individual feature z-scores using Equation 2.

$$z = \left| \frac{x - \mu}{\sigma} \right| \quad (2)$$

Where $\mu$ and $\sigma$ are the mean and standard deviation of feature values in the training set. After calculation, the maximum z-score value across all the features is selected as the output metric. The MD model used the Mahalanobis distance of a contour's features with respect to the training dataset features as the output metric (17, 18) and is calculated using Equation 3.

$$D(\vec{x}) = \sqrt{(\vec{x} - \hat{\mu})^T \Sigma^{-1} (\vec{x} - \hat{\mu})} \quad (3)$$

Where $\hat{\mu}$ is a vector containing the mean feature values and $\Sigma^{-1}$ is the inverse of the covariance matrix calculated from the training data set. The output metric for the AE network was the mean squared difference between reconstructed features and input features for a given contour (19). The AE network was trained in MATLAB® using the 'trainAutoencoder' function and consisted of a single hidden layer with 18 neurons and a cost function with a single L2 regularization term. To standardize the feature set, feature z-scores were calculated for input into the AE model. The number of epochs was limited to a maximum of 7000, and the L2 weight regularization coefficient was set to 0.005. The number of hidden layers and L2 regularization coefficient were optimized by evaluating model performance on the validation set across a range of values. The results of each single-contour feature model type were assessed individually and in combination with CCR and connectedness models for the validation and test set.

## 2.6 CCR model

For the CCR model, our objective was to come up with a set of features that could quantify varying degrees of contour-to-contour overlap and separation. To do this, the CCR model utilized the minimum distance between two contours and the fractional volume of overlap of one contour with another as its features. The combination of both features yielded all the information needed to quantify these relationships. A boolean matrix with 42 rows and 43 columns was generated to select the CCRs to include in the CCR model. Rows were associated with the selected contour, while columns were associated with the comparison contour. An additional column was added to allow comparison to the body

contour (Supplementary Figure S3). The selected CCRs primarily focused on OARs that were close to each other. This included OAR types with distinct anatomical boundaries (e.g., cord and brain stem) and cases where one OAR was a subset of another (e.g., brain stem and brain). Well-defined contours in these cases should exhibit consistent anatomical boundaries with each other. In contrast, contours that are not in close proximity to each other may have more uncertainty in their relationship, making them susceptible to false positives.

The minimum distance feature data was fit to a gamma distribution (22) ranging from zero to infinity, while the fractional overlap volume feature data was fit to a beta distribution (23) ranging from zero to one. Distribution types were selected to have the same upper and lower input domains as their representative features and followed the probability distribution of the CCR features. Initial upper and lower outlier cutoffs were determined by taking the upper and lower $99^{th}$ percentile boundaries of the fitted distribution. The percentile boundaries were set manually to minimize the number of false positives detected by the CCR model in the validation set. The determined percentile boundary cutoffs were expanded by 0.02 for fractional volume and 2mm for minimum distance to minimize identification of errors that were present, but small enough to not be clinically relevant.

## 2.7 Connectedness model

For human reviewers, identifying disconnected voxels in a contour can be time-consuming. To improve clarity for potential human reviewers using this QA tool, we separated the connectedness feature from the single-contour feature models and created a separate model including only the number of connected parts in a contour. This enables easy reporting of this feature to reviewers. To establish the maximum number of allowable parts, a statistical threshold of 99.95% was set using a gamma distribution fitted to the training data. The threshold was optimized by evaluating performance on the validation dataset and selected to minimize false positives. A statistical threshold was used instead of setting a predetermined cutoff as some contours were allowed to have multiple parts anatomically (e.g. thyroid) and other contours could have multiple parts due to CT image-related scan truncation (e.g. left and right brachial plexus).

## 2.8 Model combination

To obtain the final combined classifications, if any individual model identified a contour as erroneous, it was classified as such. Thresholds for the connectedness model and CCR model output metrics were set manually and the single-contour features model thresholds were tuned to maximize balanced accuracy for the combined classifications (24). Balanced accuracy is defined as the average of sensitivity and specificity. While accurate detection of erroneous contours is more clinically relevant, the prevalence of erroneous contours will typically be low in the clinical workflow. We estimate a reasonable prevalence of erroneous contours in the clinical workflow to be 10% and the relative severity of incorrectly categorizing erroneous compared to acceptable contours at 9 to 1. In this case, balanced accuracy will be an appropriate optimization metric (25). The values of prevalence and relative severity can easily be adjusted, resulting in different optimal thresholds for future clinical use. The performance of the single-contour feature models without combination with CCR and connectedness models was also evaluated using the same threshold tuning. Thresholds for individual and combined single-contour feature models were not necessarily the same. The test set was assessed using the same thresholds obtained from the validation set.

## 2.9 Statistics

To reduce class imbalance during statistical assessment, we adopted a solution involving random subsampling. Specifically, we selected five acceptable contours at random from the input curated gold-standard contours for each OAR type and merged them with the erroneous validation contours. This approach allowed us to present a single statistical test that was more evenly balanced in terms of its evaluation of performance on both acceptable and erroneous contours. The subsampling included 210 acceptable and 190 erroneous contours. Statistical testing of model performance was performed using the two-sided mid-p value McNemar test with a p-value of less than 0.05 considered to be significantly different (26, 27).

# 3 Results

## 3.1 Single-contour feature method comparison and model combination

Receiver operating curves were used to evaluate the performance of individual z-score, MD, and AE models for all contours. The z-score, MD, and AE models had an area under the curve (AUC) values of 0.922, 0.939, and 0.939 respectively for the test set (Figure 2, Table 2). Combining the single-contour feature models with CCR and connectedness models led to improved performance for all three single-contour feature models. The high specificity of the CCR (0.982) and connectedness (0.990) models made it possible to combine them by identifying a contour as an outlier if any of the models flagged it as one (logical OR) with minimal decrease in combined model specificity. Test set results were similar to the validation set for all models, indicating minimal overfitting due to the single feature selection, model thresholding, and hyperparameter tuning using the validation dataset. In the statistical subset of the validation data, combination of the CCR and connectedness models with the single-contour feature models significantly improved the performance of the z-score (P=0.0007), MD (P=0.0175), and AE (P=0.0201) models (Supplementary Table S2), demonstrating the added benefit of incorporating CCR features for outlier detection.

10.3389/fonc.2024.1295251

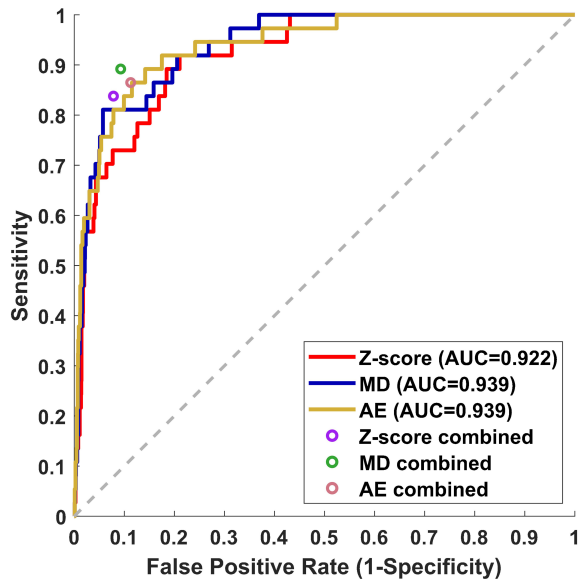



FIGURE 2
Receiver operating curve (ROC) results from the test dataset for three individual single-contour features models, z-score, MD, and AE models. The false positive rate and sensitivity of single-contour feature models combined with CCR and connectedness models are plotted as circles on the ROC plot.

## 3.2 Model performance across OAR types, error types, and error severity

The performance of the knowledge-based QA framework for each OAR type individually was evaluated using the validation dataset without distinguishing between left and right-sidedness for bilateral OARs (Figure 3, Supplementary Figures S4, S5). Analysis of the validation set allowed for an adequate number of contours of each OAR type to be available for classification evaluation. Of the three combined single-contour feature models, the combined AE model had both the highest median and highest minimum BA across all OAR types (Minimum median and maximum BA of 0.729, 0.908, and 0.980 respectively). Combining the single-contour feature models with CCR and connectedness models resulted in an average increase of 0.077 (z-score), 0.055 (MD), and 0.048 (AE) in BA values per OAR type. The improvements in BA were not evenly distributed across all OAR types. The AE model type showed the largest improvements for the spinal cord and oral cavity, while the mandible experienced worse performance (Figure 4), attributable to changes in the optimal single-contour feature thresholds when combined with CCR and connectedness models.

The sensitivity of the combined AE model for boundary, position, non-adjacent slice, and volume error types was 0.867, 0.971, 0.833, and 0.8116 in the validation set. Similar classification accuracies

TABLE 2 Classification results.

| Model | AUC | Balanced Accuracy | Sensitivity | Specificity | True positive | False negative | True negative | False positive |
|---|---|---|---|---|---|---|---|---|
| Validation Set | | | | | | | | |
| Connectedness | – | 0.527 | 0.063 | 0.991 | 12 | 178 | 1966 | 18 |
| CCR | – | 0.730 | 0.474 | 0.987 | 90 | 100 | 1958 | 26 |
| Z-score | 0.852 | 0.794 | 0.684 | 0.904 | 130 | 60 | 1793 | 191 |
| MD | 0.899 | 0.826 | 0.811 | 0.842 | 154 | 36 | 1671 | 313 |
| AE | 0.896 | 0.838 | 0.763 | 0.913 | 145 | 45 | 1812 | 172 |
| Z-score combined | – | 0.866 | 0.816 | 0.916 | 155 | 35 | 1818 | 166 |
| MD combined | – | 0.882 | 0.842 | 0.921 | 160 | 30 | 1828 | 156 |
| AE combined | – | 0.884 | 0.863 | 0.906 | 164 | 26 | 1797 | 187 |
| Test Set | | | | | | | | |
| Connectedness | – | 0.522 | 0.054 | 0.990 | 2 | 35 | 1984 | 21 |
| CCR | – | 0.721 | 0.459 | 0.982 | 17 | 20 | 1969 | 36 |
| Z-score | 0.922 | 0.816 | 0.730 | 0.903 | 27 | 10 | 1811 | 194 |
| MD | 0.939 | 0.851 | 0.865 | 0.837 | 32 | 5 | 1679 | 326 |
| AE | 0.939 | 0.866 | 0.838 | 0.893 | 31 | 6 | 1791 | 214 |
| Z-score combined | – | 0.880 | 0.838 | 0.922 | 31 | 6 | 1848 | 157 |
| MD combined | – | 0.900 | 0.892 | 0.907 | 33 | 4 | 1819 | 186 |
| AE combined | – | 0.876 | 0.865 | 0.887 | 32 | 5 | 1779 | 226 |

10.3389/fonc.2024.1295251

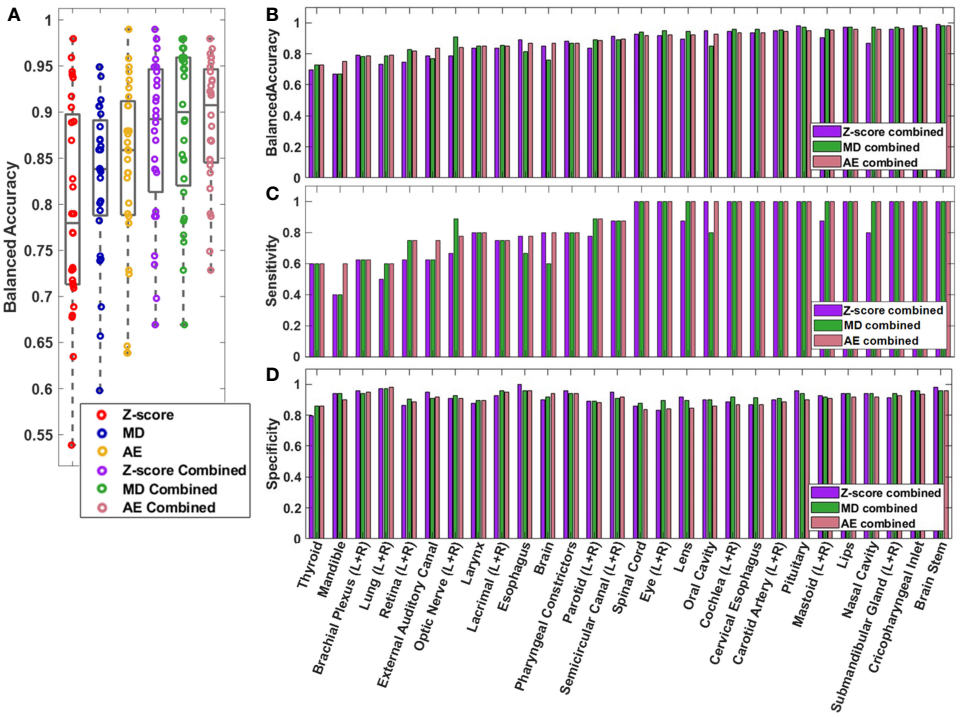



FIGURE 3
**(A)** Box and whisker plot (Box-inner quartile range, whisker-range) of balanced accuracy for each OAR type in the validation dataset. Left and right matching OARs were combined before plotting. **(B)** Balanced accuracy, **(C)** sensitivity, and **(D)** specificity are plotted for each OAR for the combined models.

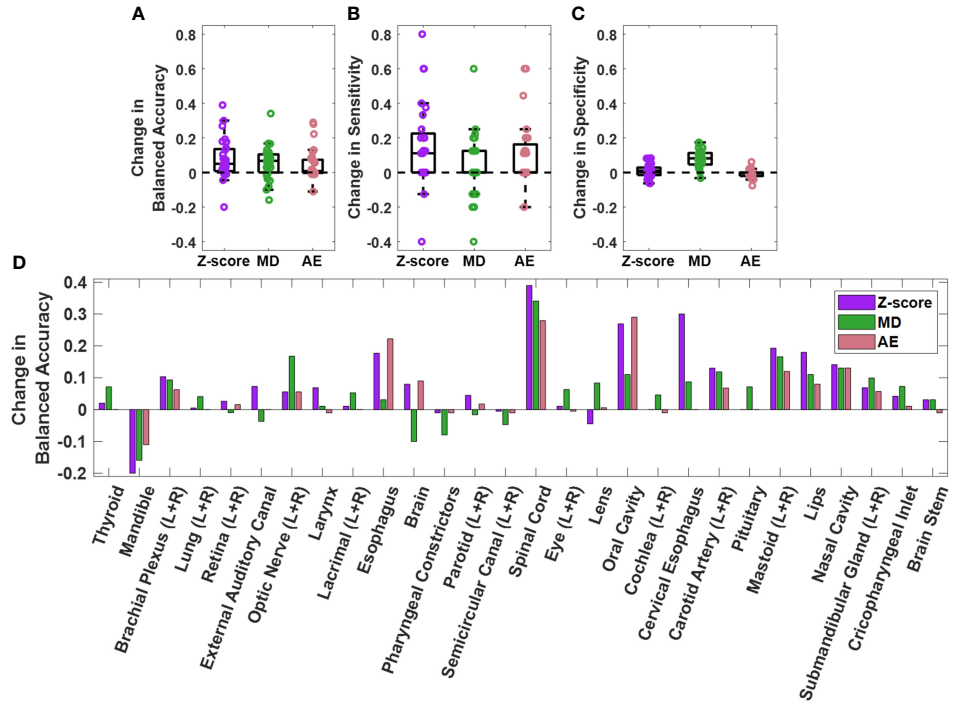



FIGURE 4
Change in **(A)** balanced accuracy, **(B)** sensitivity, and **(C)** specificity for single contour feature models on the validation dataset when single contour feature models are combined with CCR and connectedness models. Each datapoint on the plots represent an OAR type. Change for each OAR type individually is also plotted **(D)**. Positive change indicates improvement in performance after model combination.

across different error types were observed between combined AE, MD, and z-score models in both the test and validation sets (Supplementary Table S3). Higher sensitivity was observed for position errors compared to the other types of errors across all three combined model types, likely due to position errors tending to be more severe than other error types (Figures 5A–D).

The sensitivity of the combined AE model for detecting major and moderate errors was 0.922 and 0.810. Similar differences in sensitivity between major and moderate errors were observed for the combined MD and z-score models (Supplementary Table S4), suggesting that more sever errors are more likely to be detected by the knowledge-based QA framework.

## 3.3 Misclassifications

Some erroneous esophagus, lung, and brachial plexus contours that were incorrectly classified by all models had missing volumes (Figure 5E). These OARs are commonly affected by CT scan truncation which increases variation in volume and shape features. Additional undetected errors included improper boundary delineation (either over-contouring or under-contouring boundary edges) and volume changes that were small relative to the total volume of the contour. Some acceptable contours were identified as erroneous by all combined models. These included contours with clinically insignificant inaccurately delineated boundaries, contours that were anatomically accurate but contoured on patients with abnormal positioning or anatomy, and contours on CT scans with metal artifact-related image quality issues (Supplementary Figure S6). The CCR model was able to identify outliers from improper separation or overlap (Supplementary Figure S7). Out of the 34 erroneous contours from the validation set that were missed by all individual single-contour feature models, the CCR model identified 15.

# 4 Discussion

We have developed a knowledge-based method for detecting clinically relevant erroneous OAR contours in HN radiotherapy.

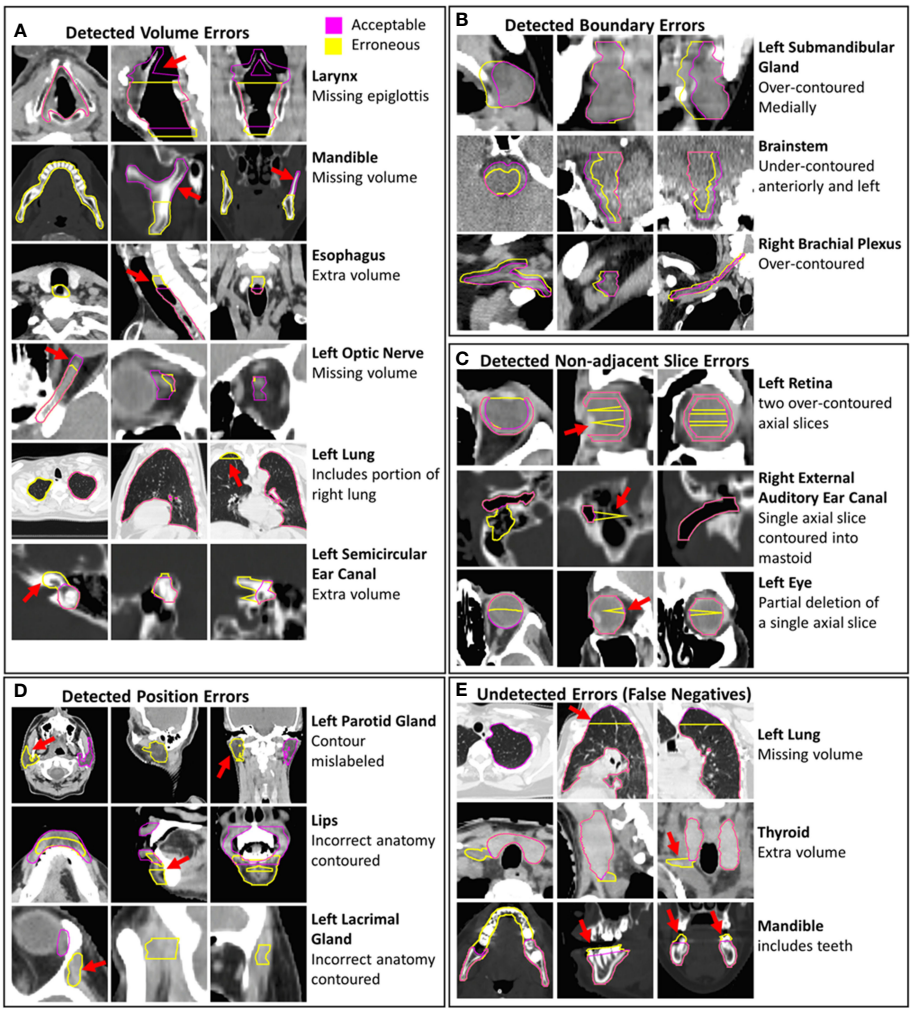



FIGURE 5
Examples of erroneous contours with volume, boundary, non-adjacent slice, and position errors are shown **(A–D)**. Examples of undetected contour errors (false positives are shown **(E)**. The corresponding gold-standard acceptable contours that were edited to create erroneous contours are displayed. Hounsfield unit display ranges were -10 to 70 for images with brainstem contours, -250 to 1500 for images with contours of bone, -1000 to 100 for images with lung contours and -115 to 115 for all other images.

Our method uses models based on single-contour features, as well as CCRs. Combining the single-contour feature model with the CCR and connectedness models significantly improves performance for the z-score, MD, and AE models. The combined AE model achieves a sensitivity of 0.865, specificity of 0.887, and BA of 0.876 for the test set. Similar BA, sensitivity, and specificity were observed for the combined z-score, MD, and AE models for both test and validation datasets, indicating no overfitting in the validation set. Minimum, median, and maximum balanced accuracies across individual OAR types for the AE model were 0.729, 0.908, and 0.980, respectively on the validation set. Our results demonstrate satisfactory model performance for a comprehensive set of OAR types utilized in HN radiotherapy.

Accurately detecting contour errors across a wide range of OAR types is a significant challenge. Many studies examining model performance have been limited to assessing no more than 17 OAR types (9, 10, 12, 14, 16). One study looking at pelvis, abdomen, and thorax regions reported results for 40 OAR types, however, their primary aim was to develop a method for classification of contours to an OAR type or label rather than to detect erroneous contours (15). Furthermore, they did not report the model performance for each OAR type, instead only reporting the overall AUC results. In our study, we use a knowledge-based outlier detection approach with a combined AE model that achieves a minimum sensitivity and specificity of 0.600/0.837 (ignoring left-right distinction) per OAR type for 42 HN OAR types used clinically. The wide variety of contour volumes, and shapes, as well as a large dataset of patients with several different HN disease types and sites, demonstrates that knowledge-based OAR QA for HN radiotherapy is both feasible and generalizable to a wide variety of OARs.

Abnormalities in CT images, caused by factors like CT artifacts, patient positioning, or abnormal anatomy, can contribute to higher false positive rates for knowledge-based outlier detection. However, these images may still result in suboptimal quality for both human-generated and DL-generated contours, highlighting the importance of careful manual review in such situations. Although the knowledge-based quality assurance system may yield false positives when encountering abnormal image data with accurate contours, it provides a rapid and efficient method to aid reviewers in automatically identifying erroneous contours.

The CCR model is a novel tool that can identify incorrect amounts of overlap or separation between two contours. This is crucial in clinical settings for two reasons: first, overlap and separation should be consistent with actual anatomy, and second, gaps between contours that are anatomically touching may result in unreported high doses to the OAR. In this study, we chose CCRs that had consistent anatomical relationships or were close to each other, but the technique can be extended to any CCRs. The high specificity of the CCR model allows for easy deployment as a contour review tool, either on its own or in conjunction with other models.

One limitation of this study is that CCR calculations for both erroneous and acceptable contours were performed only in relation to acceptable contours and never in relation to erroneous contours. This facilitated the identification and quantification of the CCRs model performance. In a real-world application, the CCR model will only detect incorrect CCRs instead of directly identifying incorrect contours. Therefore, in clinical practice, the end user would need to review two contours for each improper CCR to identify a single unacceptable contour.

The exclusion of data from the training set based on the number of MAD from the median provides a way to remove contours of questionable quality in the training dataset. The threshold for data removal can be tuned with a validation dataset. In this work, increases in balanced accuracy for the combined models when implementing outlier removal ranged from 0.00 to 0.06 depending on the model used. For less curated datasets, this approach may have a larger impact on model performance and help improve the generalizability of the QA framework to different datasets.

The knowledge-based QA framework presented in this work has the potential to improve the detection of erroneous contours when used in conjunction with human reviewers. This will require an efficient integration within the clinical contour review workflow, where the QA framework results can be quickly accessed and interpreted by a human reviewer. A script-based approach run directly from the clinical contouring software would be an effective option. This script could allow human reviewers to automatically archive human review labels, model inferences, and contours when run. This data archiving would facilitate model performance tracking, iterative model improvement, and the assessment of the dosimetric impact of erroneous contours.

The use of a large, highly curated HN OAR dataset for model development is a clear foundational strength of our study. However, our modeling also required erroneous H&N contours. This data was not available *a priori*, necessitating fabrication; we recognize that this could be perceived as a weakness in terms of presented model performance evaluation. Our immediate goal is to iteratively develop a clinical solution based on the presented methodology for integration within our contour review workflow. As we detect true erroneous contours during preliminary deployment phases, these erroneous OARs detected "in the wild" can be leveraged for future refinements (iterations in model training/tuning). Thus, we emphasize that the presented model framework, model comparison, and the generalizability of this approach to many OAR types should be recognized as the main focus of this study.

The best-performing combined AE. model can identify erroneous contours but does not identify individual features that are abnormal. To reduce the time spent during human review of contours marked as erroneous by the QA framework, it may be beneficial to identify specific abnormal features along with erroneous contours to guide reviewers more quickly to the errors in the contours. To obtain predictions on abnormal features after identifying erroneous contours using the combined AE model, a separate z-score model could be used *post hoc* to report outlier features. However, this approach may result in both models disagreeing on a contour's classification. Alternatively, more sophisticated model-agnostic tests can be employed to determine the importance of input features in making predictions, which can be useful in identifying features that strongly influence model decisions (28, 29). Additional research is needed to determine whether the identification of erroneous features in this manner would reduce contour review time.

Further research is necessary to evaluate the developed QA framework for other anatomical sites. The QA framework can be extended to other treatment regions with additional sets of curated and outlier data given its adaptability to a variety of OARs for HN. However, it is anticipated that the performance of the CCR model may decrease in the thorax, abdomen, and pelvis due to fewer consistent anatomical relationships between OARs. A better understanding of the amount of curated data needed will become more apparent after the integration of the HN model into our clinical workflow.

In the future, model generalizability to other institutions also needs to be assessed. Several challenges are associated with this, including variation in contour definitions (7), and variation in the determination of clinically relevant contour errors between different institutions. While trained models could be directly deployed in outside institutions, outside institutions could also train their own institution specific model using the same QA framework as illustrated here. This would allow any differences in contour definitions, and contour error definitions to be accounted for. More research is needed to assess the generalizability of this approach to other institutions.

## 5 Conclusion

In this study, we have created a method for knowledge-based QA that utilizes single-contour features and contour-to-contour relationships to identify erroneous contours for forty-two HN OAR types. The effectiveness of multiple models has been evaluated, both in general and for each OAR type. The findings of this study demonstrate the developed framework for knowledge-based QA of HN contours is both feasible and generalizable to a full set of clinical HN OARs.

## Data availability statement

The datasets presented in this article are not readily available because of Mayo Clinic policies on data use and sharing. Requests to access the datasets should be directed to brooks.jamison@mayo.edu.

## Ethics statement

The studies involving humans were approved by Mayo Clinic Rochester Minnesota USA. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

JB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. ET: Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing – review & editing. AA: Data curation, Writing – review & editing. CB: Data curation, Writing – review & editing. RF: Conceptualization, Data curation, Writing – review & editing. JL: Conceptualization, Data curation, Methodology, Software, Writing – review & editing. NL: Data curation, Writing – review & editing. DR: Conceptualization, Writing – review & editing. SP: Data curation, Writing – review & editing. SS: Data curation, Software, Writing – review & editing. DM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

We acknowledge the use of Chat GPT (30) in editing this report. In cases where Chat GPT was utilized, specific portions of the author's written text were modified by Chat GPT to enhance clarity. All modifications made to the text underwent thorough human review before incorporation into the final manuscript. We affirm that we are responsible for all aspects of the code used, data creation, data interpretation, original writing, and citations used in this report.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1295251/full#supplementary-material

# References


1. Brouwer CL, Steenbakkers RJ, Bourhis J, Budach W, Grau C, Grégoire V, 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. doi: 10.1016/j.radonc.2015.07.041

2. Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: A workflow study at two cancer centers. *Radiat Oncol.* (2021) 16:101. doi: 10.1186/s13014-021-01831-4

3. Lucido JJ, DeWees TA, Leavitt TR, Anand A, Beltran CJ, Brooke MD, et al. Validation of clinical acceptability of deep-learning-based automated segmentation of organs-at-risk for head-and-neck radiotherapy treatment planning. *Front Oncol.* (2023) 13. doi: 10.3389/fonc.2023.1137803

4. Byun HK, Chang JS, Choi MS, Chun J, Jung J, Jeong C, et al. Evaluation of deep learning-based autosegmentation in breast cancer radiotherapy. *Radiat Oncol.* (2021) 16:203. doi: 10.1186/s13014-021-01923-1

5. Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, de Souza D, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. *Radiotherapy Oncol.* (2019) 135:130–40. doi: 10.1016/j.radonc.2019.03.004

6. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning. *Radiother Oncol.* (1998) 47:285–92. doi: 10.1016/s0167-8140(98)00021-8

7. van der Veen J, Gulyban A, Willems S, Maes F, Nuyts S. Interobserver variability in organ at risk delineation in head and neck cancer. *Radiat Oncol.* (2021) 16:120. doi: 10.1186/s13014-020-01677-2

8. Rhee DJ, Akinfenwa CPA, Rigaud B, Jhingran A, Cardenas CE, Zhang L, et al. Automatic contouring qa method using a deep learning-based autocontouring system. *J Appl Clin Med Phys.* (2022) 23:e13647. doi: 10.1002/acm2.13647

9. Rhee DJ, Cardenas CE, Elhalawani H, McCarroll R, Zhang L, Yang J, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys.* (2019) 46:5086–97. doi: 10.1002/mp.13814

10. Duan J, Bernard ME, Castle JR, Feng X, Wang C, Kenamond MC, et al. Contouring quality assurance methodology based on multiple geometric features against deep learning auto-segmentation. *Med Phys.* (2023) 10:2715–32. doi: 10.1002/mp.16299

11. Claessens M, Oria CS, Brouwer CL, Ziemer BP, Scholey JE, Lin H, et al. Quality assurance for ai-based applications in radiation therapy. *Semin Radiat Oncol.* (2022) 32:421–31. doi: 10.1016/j.semradonc.2022.06.011

12. Altman MB, Kavanaugh JA, Wooten HO, Green OL, DeWees TA, Gay H, et al. A framework for automated contour quality assurance in radiation therapy including adaptive techniques. *Phys Med Biol.* (2015) 60:5199–209. doi: 10.1088/0031-9155/60/13/5199

13. Hui CB, Nourzadeh H, Watkins WT, Trifiletti DM, Alonso CE, Dutta SW, et al. Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach. *Med Phys.* (2018) 45:2089–96. doi: 10.1002/mp.12835

14. Nourzadeh H, Hui C, Ahmad M, Sadeghzadehyazdi N, Watkins WT, Dutta SW, et al. Knowledge-based quality control of organ delineations in radiation therapy. *Med Phys.* (2022) 49:1368–81. doi: 10.1002/mp.15458

15. McIntosh C, Svistoun I, Purdie TG. Groupwise conditional random forests for automatic shape classification and contour quality assessment in radiotherapy planning. *IEEE Trans Med Imaging.* (2013) 32:1043–57. doi: 10.1109/TMI.2013.2251421

16. Chen HC, Tan J, Dolly S, Kavanaugh J, Anastasio MA, Low DA, et al. Automated contouring error detection based on supervised geometric attribute distribution models for radiation therapy: A general strategy. *Med Phys.* (2015) 42:1048–59. doi: 10.1118/1.4906197

17. Etherington TR. Mahalanobis distances for ecological niche modelling and outlier detection: Implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method. *PeerJ.* (2021) 9:e11436. doi: 10.7717/peerj.11436

18. Mahalanobis PC. On the generalized distance in statistics *Proceedings of the National Institute of Science of India* (1936) 2:49–55. doi: 10.1007/s13171-019-00164-5.

19. Ruff L, Kauffmann JR, Vandermeulen RA, Montavon G, Samek W, Kloft M, et al. A unifying review of deep and shallow anomaly detection. *Proc IEEE.* (2021) 109:756–95. doi: 10.48550/arXiv.2009.11732

20. Tryggestad E, Anand A, Beltran C, Brooks J, Cimmiyotti J, Grimaldi N, et al. Scalable radiotherapy data curation infrastructure for deep-learning based autosegmentation of organs-at-risk: A case study in head and neck cancer. *Front Oncol.* (2022) 12. doi: 10.3389/fonc.2022.936134

21. Cohen I, Huang Y, Chen J, Cohen I. Pearson correlation coefficient. *Noise reduction speech processing.* (2009) 2009:1–4. doi: 10.1007/978-3-642-00296-0_5

22. Hogg RV, Craig AT. Introduction to mathematical statistics.(5"" edition). *Englewood Hills New Jersey.* (1995) 173–86.

23. Johnson NL, Kotz S, Balakrishnan N. *Continuous univariate distributions, volume 2* Vol. 289. New Jersey: John wiley & sons (1995).

24. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. (2010). The balanced accuracy and its posterior distribution, in: *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, (New York, NY, US: Institute for Electrical and Electronics Engineers), 23-26 Aug. 2010.

25. Habibzadeh F. On determining the most appropriate test cut-off value: The case of tests with continuous results. *Biochemia Med.* (2016) 26:297–307. doi: 10.11613/BM.2016.034

26. Fagerland MW, Lydersen S, Laake P. The mcnemar test for binary matched-pairs data: Mid-p and asymptotic are better than exact conditional. *BMC Med Res Method.* (2013) 13:91. doi: 10.1186/1471-2288-13-91

27. Lancaster HO. Significance tests in discrete distributions. *J Am Stat Assoc.* (1961) 56:223–34. doi: 10.1080/01621459.1961.10482105

28. Ribeiro MT, Singh S, Guestrin C. (2016). " Why should i trust you?" Explaining the predictions of any classifier, in: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, (New York, NY, USA: Association for Computing Machinery).

29. Sejr JH, Schneider-Kamp A. Explainable outlier detection: What, for whom and why? *Mach Learn Appl.* (2021) 6:100172. doi: 10.1016/j.mlwa.2021.100172

30. OpenAI. *Chat gpt* [computer program]. OpenAI. (San Francisco, CA, USA) (2023). https://chat.openai.com

frontiers | Frontiers in Oncology

TYPE Brief Research Report
PUBLISHED 14 March 2024
DOI 10.3389/fonc.2024.1358350



OPEN ACCESS

EDITED BY
Minglei Kang,
New York Proton Center, United States

REVIEWED BY
Sunyoung Jang,
The Pennsylvania State University, United States
Yu-Jie Huang,
Kaohsiung Chang Gung Memorial Hospital, Taiwan
Ruijie Yang,
Peking University Third Hospital, China

*CORRESPONDENCE
Vesela Koteva
vesela.koteva@icr.ac.uk

RECEIVED 19 December 2023
ACCEPTED 12 February 2024
PUBLISHED 14 March 2024

CITATION
Koteva V, Eiben B, Dunlop A, Gupta A, Gangil T, Wong KH, Breedveld S, Nill S, Harrington K and Oelfke U (2024) Clinical acceptance and dosimetric impact of automatically delineated elective target and organs at risk for head and neck MR-Linac patients.
*Front. Oncol.* 14:1358350.
doi: 10.3389/fonc.2024.1358350



# Clinical acceptance and dosimetric impact of automatically delineated elective target and organs at risk for head and neck MR-Linac patients

Vesela Koteva[1*], Björn Eiben[1], Alex Dunlop[2], Amit Gupta[3], Tarun Gangil[1], Kee Howe Wong[4], Sebastiaan Breedveld[5], Simeon Nill[1,2], Kevin Harrington[6] and Uwe Oelfke[1,2]

[1]Radiotherapy Physics Modelling, Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [2]The Joint Department of Physics, The Royal Marsden Hospital and The Institute of Cancer Research, London, United Kingdom, [3]Head and Neck Unit, The Royal Marsden National Health Service (NHS) Foundation Trust and The Institute of Cancer Research, London, United Kingdom, [4]Head and Neck Unit, The Royal Marsden National Health Service (NHS) Foundation Trust, London, United Kingdom, [5]Department of Radiotherapy, Erasmus University Medical Center (MC) Rotterdam, Rotterdam, Netherlands, [6]Targeted Radiotherapy, Department of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom

**Background:** MR-Linac allows for daily online treatment adaptation to the observed geometry of tumor targets and organs at risk (OARs). Manual delineation for head and neck cancer (HNC) patients takes 45-75 minutes, making it unsuitable for online adaptive radiotherapy. This study aims to clinically and dosimetrically validate an in-house developed algorithm which automatically delineates the elective target volume and OARs for HNC patients in under a minute.

**Methods:** Auto-contours were generated by an in-house model with 2D U-Net architecture trained and tested on 52 MRI scans via leave-one-out cross-validation. A randomized selection of 684 automated and manual contours (split half-and-half) was presented to an oncologist to perform a blind test and determine the clinical acceptability. The dosimetric impact was investigated for 13 patients evaluating the differences in dosage for all structures.

**Results:** Automated contours were generated in 8 seconds per MRI scan. The blind test concluded that 114 (33%) of auto-contours required adjustments with 85 only minor and 15 (4.4%) of manual contours required adjustments with 12 only minor. Dosimetric analysis showed negligible dosimetric differences between clinically acceptable structures and structures requiring minor changes. The Dice Similarity coefficients for the auto-contours ranged from $0.66 \pm 0.11$ to $0.88 \pm 0.06$ across all structures.

10.3389/fonc.2024.1358350

**Conclusion:** Majority of auto-contours were clinically acceptable and could be used without any adjustments. Majority of structures requiring minor adjustments did not lead to significant dosimetric differences, hence manual adjustments were needed only for structures requiring major changes, which takes no longer than 10 minutes per patient.

KEYWORDS

clinical acceptability, dosimetric impact, MR-Linac, automated delineation, head and neck cancer

# 1 Introduction

Every radiotherapy treatment starts with a generation of a treatment plan specifying a clinically optimized dose distribution and its delivery parameters for each patient. In our head and neck cancer (HNC) radiotherapy protocol, a treatment plan is initially generated based on a CT scan, prescribing 65 Gy to the primary target and 54 Gy to the elective target (the combined volume of the neck lymph nodes excluding the overlap of the nodes and the primary tumor) delivered in 30 fractions. The treatment plan is usually based on a CT scan, acquired one or more weeks before the treatment (1). However, this plan does not consider anatomical changes during treatment, risking compromised clinical goals and increased toxicity (1–7). For instance, the parotid glands may move closer to high-dose regions, posing a risk of overdosing these sensitive organs. Figure 1A shows a patient's scan acquired during their final fraction (fraction 30) with overlayed contours of the parotid glands, spinal cord and brainstem from the patient's initial scan. Hence, if we simply copy the contours from the initial to the final scan the parotid glands would be partially located outside of the patient's external outline and the spinal cord and brainstem would not be at the correct position. Patient anatomy changes, even with radiotherapy masks, necessitate adapting contours to ensure accurate dose delivery.

Adaptive radiotherapy (ART) using an MR-Linac allows for real-time treatment plan adaptation based on daily anatomical changes (8, 9). In order to adapt the treatment plan, the ROIs need to be re-delineated on the daily scan while the patient is on the treatment couch. This requires organ delineation in less than one minute (10). Manual delineation, taking around 45 minutes, is infeasible for HNC patients within acceptable time frames. Current practice involves deformable image registration, but it requires initial manual delineation, lasting 45 to 75 minutes, and is prone to inaccuracies, often requiring additional manual adjustments due to imperfect results (11, 12).

In the past decade numerous groups have investigated automatic delineation using deep convolutional neural networks

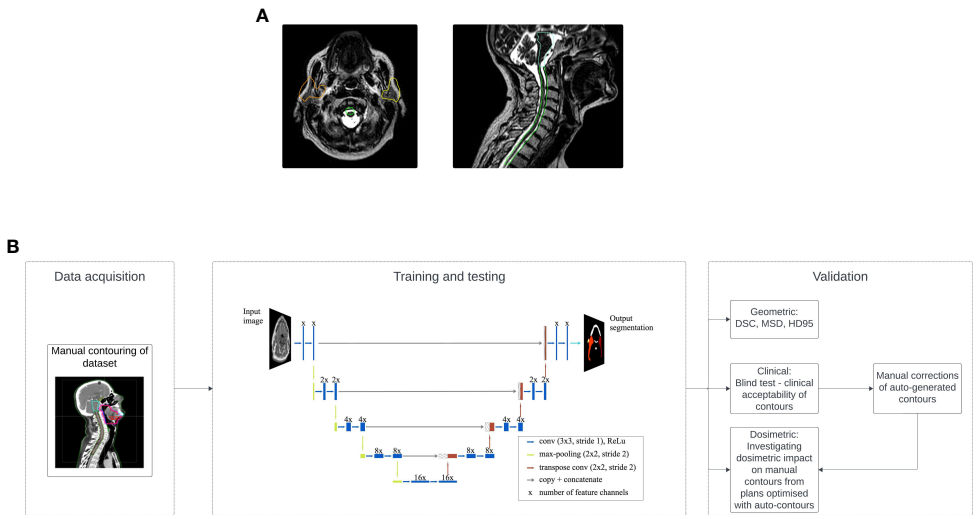



FIGURE 1
**(A)** Example of anatomical changes where contours of the parotid glands (orange: right parotid, yellow: left parotid), spinal cord (green) and brainstem (pale green) obtained from the initial scan are overlayed with the scan of the patient from the last fraction (fraction 30). **(B)** General workflow of the study split into three main groups: data acquisition, where a clinician manually contours all available data, training and testing where the model is trained to learn the manual contours and tested to produce a set of contours on an unseen scan, and last, validation where the model has been validated based on clinical acceptability, dosimetric impact and geometric analysis.

(CNNs) and have shown their great potential (13–16). The assessment of automatically generated contours typically relies on metrics like Dice similarity coefficient (DSC) and geometric measures such as Hausdorff distance. This study seeks to employ an in-house model for automatic delineation of the elective target volume and OARs, reducing delineation time for HNC patients, and assess the clinical acceptability and dosimetric impact of the auto-contours.

## 2 Materials and methods

A graphical representation of the complete workflow is shown in Figure 1B.

### 2.1 In-house model for automated delineation

The dataset employed in this study comprised 52 MR-Linac scans, 14 T1-weighted and 38 T2-weighted, obtained from 52 patients diagnosed with HNC. All scans were acquired using the 7MV (flattening filter free - FFF) Elekta Unity MR-Linac (Elekta AB, Stockholm, Sweden) with magnetic strength of 1.5T. A radiation oncologist thoroughly examined all available scans for each of the 52 patients, selecting a single scan per patient based on optimal imaging quality. Each scan was resampled via SimpleITK (Insight Software Consortium) (17, 18) to cover the HN region with an in-plane resolution of 0.6 x 0.6 $mm^2$, slice thickness of 1.1 mm and dimensions (x, y, z) = (768, 768, 420), where z represents the number of slices, while x and y represent the number of pixels of each slice. From the available scans, 38 originated from the MOMENTUM study (19) and 14 were provided by the Royal Marsden Hospital (RMH), UK. The radiation oncologist delineated the neck nodes (levels 1a-5), parotid glands, spinal cord, brainstem, inferior pharyngeal constrictor muscle (IPCM), superior and middle pharyngeal constrictor muscle (SMPCM) and mandible.

A deep convolutional neural network (CNN) was trained to reproduce the manually-delineated (ground truth) structures. The CNN had a typical 2D U-Net architecture with 58 layers in total including batch normalization and activation layers (Supplementary Material provides more details. The MR scans were fed into the network slice-by-slice (20). This approach makes use of deep learning using Python (version 3.7) and the open-source libraries Tensorflow (21) and Keras both Google, Menlo Park, California, United States (22). The model was trained utilizing the computational power of an NVIDIA Tesla V100 GPU. Although some studies favor 2.5D and 3D U-Nets over 2D U-Nets (23, 24), providing this extra information doesn't consistently enhance accuracy (15). Additionally, 2D CNNs are more computationally efficient than 2.5D or 3D U-Nets, requiring fewer resources for processing. We believe the available MRI scans have sufficient resolution for the task. Training a 3D network would demand decreased resolution and spatial size, risking loss of important features. Furthermore, 2D U-Nets require less data and are less prone to overfitting than 3D U-Nets, potentially leading to better generalization.

Leave-one-out cross-validation was used (25). This technique takes all but one patients as input for training and uses the remaining patient for testing. This is repeated until predictions are made for all patients. All images were downsampled by a factor of 2 before being fed to the network which was trained for 40 epochs with a learning rate of 0.0001. We used the Dice loss, optimizing it with the Adam optimizer (26, 27). Data augmentation was applied through rotation within ±3°, zoom up to ±10% and vertical/horizontal shifts up to 10% of the original image size.

We timed the generation of contours on a 3D scan and evaluated geometrically using DSC, mean surface distance (MSD) and 95$^{th}$ percentile Hausdorff distance (HD95). The DSC shows how good the overlap between the auto-generated and manual contour is (1 for complete, 0 for none). The MSD represents the mean distance between each point of the auto-contour to the closest point from the manual contour. HD95 measures the largest distance among the closest 95% of the points from both contours (28).

### 2.2 Clinical acceptance

A clinical acceptance test by a second oncologist with 13 years of clinical experience assessed 684 contours—half manual, half model-generated. About 57% of model contours had DSC above 0.8, while 9% scored below 0.6. For detailed breakdown of the exact number of structures from the different groups of DSC that were presented to the oncologist we refer the reader to the Supplementary Data. To perform a 'blind test' the oncologist had no prior knowledge which contours were manual and which were auto-generated. The same patient and contours were presented to the oncologist on two separate days without their knowledge. The oncologist stated if the contours are clinically acceptable and if not, they stated the level of adjustments required from 1 to 5 (1 = minimal adjustments, 5 = complete re-contouring), similar to the method presented in (29). Afterwards a detailed breakdown was performed to find how many model-generated contours from each DSC group were classed as clinically acceptable and requiring minor and major adjustments.

### 2.3 Dosimetric impact

The dosimetric impact of the model-generated structures was evaluated. Using our standard clinical template, new treatment plans were optimized using the automatically generated contours and compared to dose distributions derived from the ground truth contours for 13 of the patients. These 13 patients were chosen as contours of the primary target volume were available, whereas contours of the primary target were not provided for the remaining patient population. To create the elective planning target volume, a margin of 3 mm was used around the combined volume of the left and right neck nodes, subtracting the primary planning target volume. The dosimetric impact was evaluated on the neck nodes separately without adding a margin. As quality

indicators we selected the mean dose delivered to the parotid glands, pharyngeal constrictor muscles (PCMs) and mandible, maximum dose delivered to spinal cord and brainstem, and dose delivered to 95% of the volume of the neck nodes for both automated and manual structures. The results are presented as the absolute differences between the two respective doses normalized as a percentage of the prescribed dose of 65 Gy. Our findings are separated into three groups:

1. dosimetric impact on structures classified as clinically acceptable
2. dosimetric impact on structures that required minor changes (levels 1-2)
3. dosimetric impact on structures that required major changes (levels 3+).

Furthermore, we asked the oncologist to perform the required adjustments manually for several patients and recorded the time it would take to amend the contours. We compared the dosimetric results to the average absolute dosimetric difference between dose delivered to manually delineated structures and the corresponding automated and later on manually adjusted contours.

Approximate representation of the dosimetric impact for best, median, and worst algorithm performances was shown, analyzing cases from Figure 2A. The evaluation included median performance for neck nodes and parotid glands, worst and median performance for spinal cord, worst performance for brainstem, and worst, median, and best performance for mandible. Limited target volume information precluded analysis for other examples, so structures with similar DSC values were selected for assessment.

# 3 Results

## 3.1 In-house model for automated delineation

The network took an average of 32 hours to train (range 16-48 hours), while full 3D MRI organ delineation completed within 8 seconds. Figure 2A displays model-generated structures overlaying manually delineated contours, showcasing best, median, and worst performances based on DSC on a contour-by-contour basis. Average DSCs were 0.71 ± 0.17/0.77 ± 0.11, 0.84 ± 0.12/0.85 ± 0.10, 0.75 ± 0.13, 0.88 ± 0.06, 0.66 ± 0.11, 0.63 ± 0.15, and 0.84 ± 0.06 for left/right neck nodes, left/right parotid glands, spinal cord, brainstem, SMPCM, IPCM, and mandible, respectively. Further details on DSC, MSD, and HD95 are in Figure 2B.

## 3.2 Clinical acceptance

The blind test showed that 114 (≈ 33%) of the auto-generated contours required adjustments. The mean score of adjustments was $\overline{M}$ = 1.89 (range 1-4) with a median score of $\tilde{M}$ = 1. From the manually delineated structures 15 (≈ 4.4%) required editing with mean score $\overline{M}$ = 1.60(range 1-3) and median $\tilde{M}$ = 1. Figure 3A shows detailed breakdown of the number of ROIs requiring adjustments.

No clear DSC-adjustment correlation was observed. Generally, DSC>0.8 indicated clinically acceptable contours (except PCMs). DSC between 0.6 and 0.8 showed acceptability or minor changes (3 neck nodes needed major adjustments). DSC < 0.6 usually required major amendments. Due to the small size of the PCMs, most results scored DSC between 0.6 and 0.8 and majority were classed as requiring minor adjustments and when the DSC was below 0.6 majority of contours required major adjustments. Only 2 PCMs had DSC above 0.8 and one of them was clinically acceptable, whereas the other one required minor amendments. Of 196 contours with DSC≥0.8, 180 were acceptable, 16 needed minor adjustments; 39 out of 72 contours (DSC 0.7-0.8) were clinically acceptable, 32 needed minor adjustments, and one (neck node) required major changes. In the next group, 7 out of 43 structures (DSC 0.6-0.7) were acceptable, 31 needed minor adjustments, and 3 required major changes. Lastly, 2 out of 30 structures (DSC<0.6) were acceptable; 5 needed minor adjustments, and 23 required major changes. Detailed breakdown in Supplementary Material. Clinical acceptance test was performed for majority of the best, median and worst performance of the model except for the worst performance for neck nodes and best performance for the parotid glands. Based on the other results, most likely the neck nodes contours for the worst performance would have required major adjustments, while the best performance contours of the parotid glands would have been clinically acceptable. Figure 4 shows detailed outcomes for other cases. All structures (except PCMs) from best and median performance were clinically acceptable; PCMs needed minor adjustments.

No correlation was found between amendment level and manual correction time. Average time for model-generated structure correction is 7.5 minutes per patient: 1min 27s for OAR with minor adjustments (levels 1-2) and 4mins 23s for neck nodes with the same adjustments level. For level 3+ adjustments, it takes 1min 4s for OAR and 4mins 39s for neck nodes. When presenting contours on two days, initially, the oncologist suggested level 1 corrections for SMPCM and right neck nodes, but later deemed all regions clinically acceptable.

## 3.3 Dosimetric impact

The results of the dosimetric impact analysis are shown in Figure 3B. The median absolute difference between dose delivered to auto-generated and manual contours for structures requiring no changes and structures requiring levels 1-2 amendments were very close and under 5% of prescribed dose. The average dosimetric difference between dose delivered to algorithm-generated, manually adjusted contours and manually delineated contours was in most cases higher than the median differences. Dosimetric difference was higher for structures requiring adjustments of levels 3+. Figure 4 shows the dosimetric impact of the structure for which the model had best, median and worst performance or closest to these DSC values if information for the primary target was missing. DSCs of

the structures used for this analysis were in the range of 0.44 - 0.87 for neck nodes, 0.48 - 0.91 for parotid glands, 0.35 - 0.87 for spinal cord, 0.64-0.93 for brainstem, 0.52 - 0.79 for SMPCM, 0.42 - 0.79 for IPCM, and 0.55 - 0.91 for Mandible. For all structures from the best and median performance of the model, the dosimetric difference is under 5% of prescribed dose.

## 4 Discussion

This study investigates the clinical acceptability and dosimetric impact of automatically obtained contours of the elective target volume and OARs required for treatment planning on MR-Linac HNC patients' data. It was found that majority of automated contours (≈ 67%) were clinically acceptable and in general the ones that require minor adjustments do not lead to significant dosimetric differences.

With the increasing interest in deep learning-based strategies for automated segmentation in radiation oncology, numerous groups have developed their own in-house models. Kieselmann et al. (15) have developed a model, similar to the one presented in this study, for delineating the parotid glands on MRI, reporting average DSC of 0.85 ± 0.11, which is comparable to our achieved DSC of 0.84 ± 0.12/0.85 ± 0.10 for left/right parotid glands, respectively. Dai et al. (30) have also developed a very similar model for multi-organ delineation on MRIs, reporting again

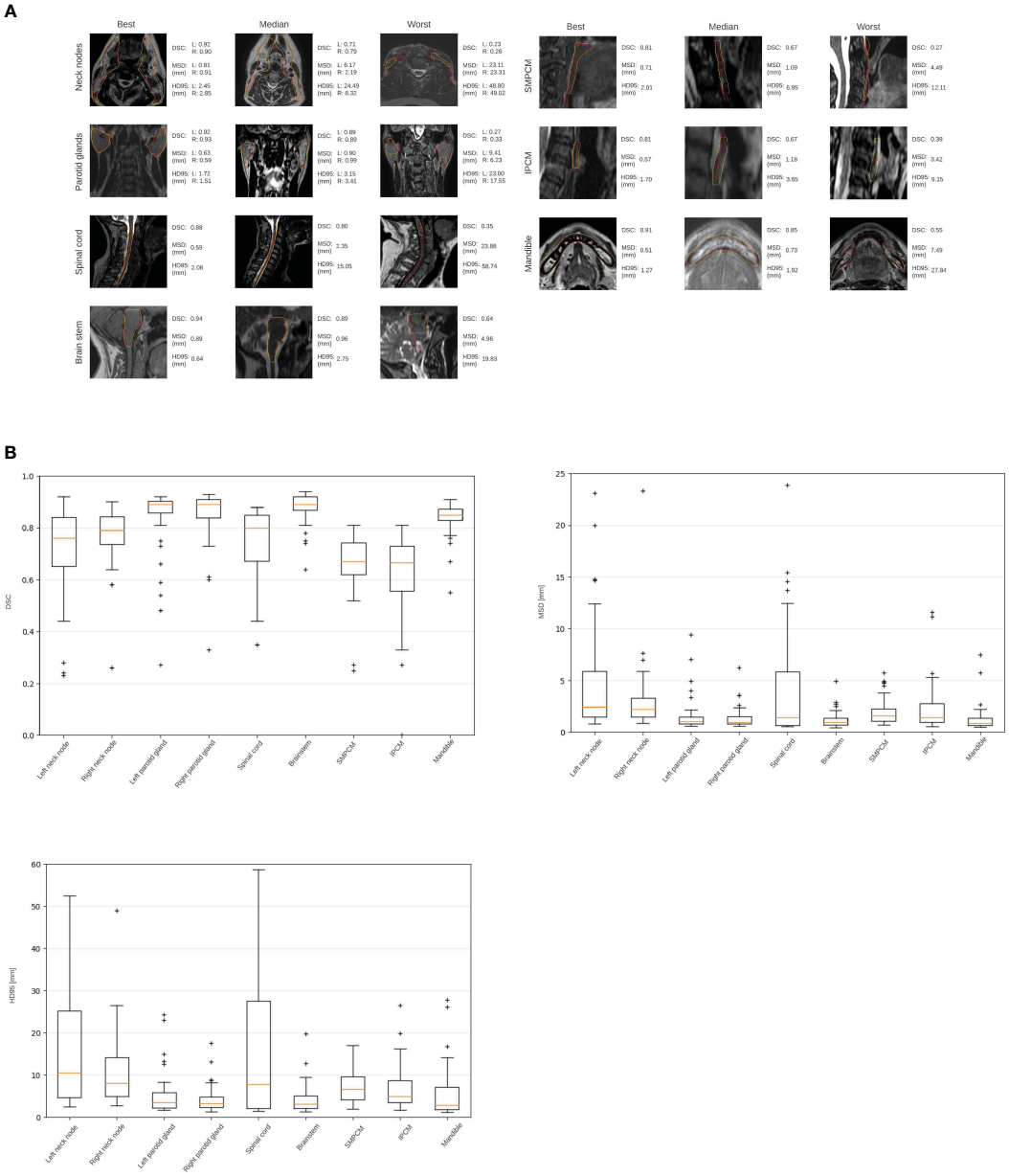



FIGURE 2
**(A)** Examples of the automatically generated contours (yellow) overlaid onto the manually delineated contours (red), representing the model's best, median, and worst performances determined by the DSC scores on a case-by-case basis. **(B)** Box plot showing the obtained range of the DSCs, MSDs, HD95s for all structures.

comparable results with average DSCs of 0.89 ± 0.06, 0.85 ± 0.06/0.86 ± 0.05, 0.77 ± 0.15 and 0.82 ± 0.10 for brainstem, left/right parotid gland, spinal cord and mandible, respectively. Their achieved average DSCs are marginally higher for the brainstem and the spinal cord, however they have not attempted to delineate the elective target volume. Korte et al. (31) and Kawahara et al. (32) have developed models to delineate the parotid gland and elective target volume levels II and III with Korte et al. investigating three different CNN, whereas Kawahara et al. compares CNNs to generative adversarial networks (GANs). Both groups achieve similar results for the parotid glands equivalent to the ones from the other discussed studies. Korte et al. have achieved 0.708 ± 0.053/0.715 ± 0.071 and 0.561 ± 0.100/0.573 ± 0.105 for left/right level II and level III, respectively. Kawahara et al. have shown that GANs have better performance when delineating the elective target volume with DSCs of 0.80/0.81 and 0.77/0.75 for left/right level II and level III, respectively. In comparison, we have achieved 0.71 ± 0.17/0.77 ± 0.11 for left/right neck node, however our contour is the

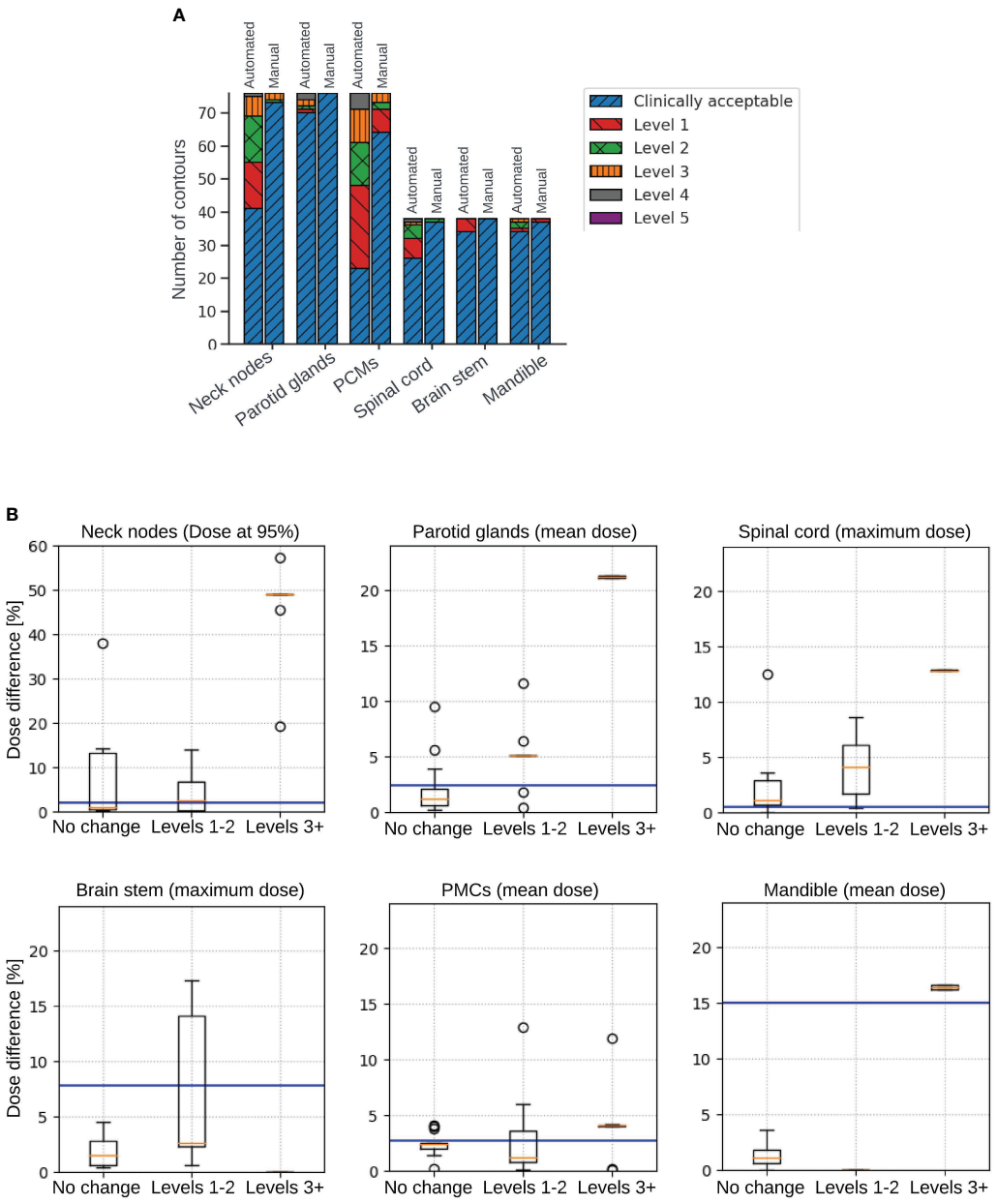



FIGURE 3
**(A)** Detailed breakdown of the number of automated and manual contours that were clinically acceptable, as well as requiring each level of corrections from 1 to 5. **(B)** Box plots illustrating dosimetric impact on structures that were clinically acceptable (not requiring any change), structures requiring minor changes (levels 1-2), and structures requiring changes of levels 3+. Plots show absolute difference in dose delivered to 95% of the volume of the neck nodes (top left), absolute difference in mean dose delivered to the parotid glands (top middle), absolute difference in maximum dose delivered to the spinal cord (top right), absolute difference in maximum dose delivered to the brainstem (bottom left), absolute difference in mean dose delivered to the pharyngeal constrictor muscles (PCMs) (bottom middle), and absolute difference in mean dose delivered to the mandible (bottom right). Blue horizontal lines represent the average dosimetric difference between dose delivered to algorithm-generated manually adjusted contours and manually delineated contours.

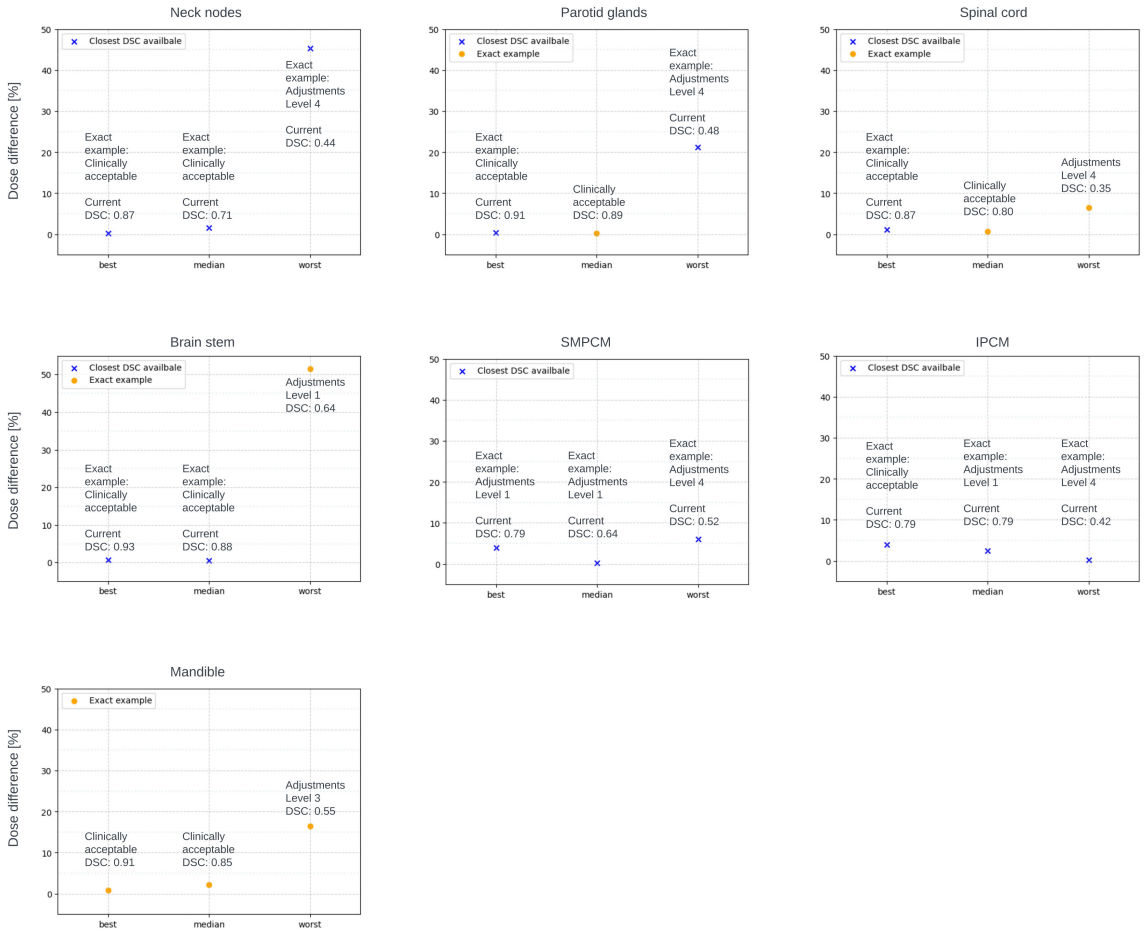



FIGURE 4
Clinical acceptance of exact examples from the best, median and worst performance of the model. Dosimetric difference evaluated for exact examples (orange circle) if information of the primary target was available, or with closest DSC (blue cross) to the exact example if primary target information was not available.

combined volume of all levels 1a-5, therefore we cannot fairly compare our results.

Prior studies have predominantly focused on geometric evaluations, neglecting clinical acceptability. A 2020 review by Vrtovec et al. (33) highlighted this gap, emphasizing the scarcity of studies assessing the clinical viability of automated contours. Recent research post-2020 delves into the impact of auto-contours on HNC patient workflows (29, 34–36). Wong et al. (29) and Zhong et al. (35) mirrored a methodology similar to this study, seeking expert opinions on clinical acceptability or required adjustments for auto-contours. Zhong et al. found a majority of auto-contours clinically acceptable, aligning with this study, while Wong et al. reported that most required minor adjustments. Thor et al. (34) and Radici et al. (36) explored dosimetric impacts, with Thor et al. optimizing treatment plans using auto-generated contours and Radici et al. recalculating doses on auto-contours using original clinical plans. Notably, these studies utilized CT scans. Liu et al. (37) reviewed deep learning-based segmentation in the HN region, finding superior brainstem segmentation accuracy on MR scans (DSC 0.92) than CT (DSC 0.86). Other CT-based studies, He et al. (38) and Zhang et al. (39) reported successful auto-segmentation of HN region organs-at-risk (OARs). Our results showed improved parotid gland segmentation and comparable brainstem performance. Although spinal cord DSC was slightly lower, it remained comparable to inter-observer variability DSC. Strijbis et al. (40) segmented individual levels of the lymph nodes, achieving a combined structure DSC of 0.86, exceeding our model in geometrical evaluation. While the results showcase an impressive performance, it is noteworthy that the sizes of the available datasets for CT scans significantly surpass those for MR images. We expect an enhancement in the performance of our model as the dataset size expands. Moreover, these studies only reported geometrical results without dosimetric or clinical acceptability analyses. This study, to the best of current knowledge, is the first obtaining autocontours for elective target volumes and this set of crucial OARs (per clinical protocol) using MR-Linac HNC patient data. It specifically investigates both clinical acceptability and dosimetric impact, a facet rarely explored in prior research.

The geometric evaluation revealed lower DSC scores for neck nodes and PCMs. Larger HD95 values for the spinal cord and neck nodes suggested misclassified voxels and incomplete delineation. Instances with DSC below 0.6 led to increased Mean Surface

Distance (MSD) and HD95 (see Figure 2). Detailed analysis highlighted misclassifications in various regions across multiple patients (left neck nodes for 10 patients, right neck nodes - 3 patients, left parotid - 4 patients, right parotid - 2 patients, SMPCM - 7 patients, IPCM - 16 patients, and mandible - 1 patient). Larger organs, like the neck nodes and spinal cord, resulted in higher maximum MSD and HD95 values due to their size. In some cases (refer to Figure 2A), correct classification was limited to specific areas, causing substantial differences in MSD and HD95 for lower DSC cases. The dataset's mix of T1- and T2-weighted MRIs, with differing contrasts, most likely impacted the deep learning model's performance. Studies have shown that when multiple oncologists delineate the neck nodes, the DSC ranges between 0.67 - 0.82 (41). Van der Veel et al. (42) have found that the mean DSC of the inter-observer variability is 0.82/0.83, 0.78, 0.88, 0.50/0.53/0.53, and 0.90 for left/right parotid glands, spinalcord, brainstem, superior/middle/inferior pharyngeal constrictor muscles, and mandible, respectively. Expectedly, due to their small size, PCMs had lower DSC values compared to other regions (43). However, the obtained value still closely aligns with the inter-observer variability DSC similar to the rest of the ROIs.

The blind test found that about 67% of model-generated and approximately 95.6% of manual contours were clinically acceptable. Approximately 75% of required adjustments for model-generated contours were only minor (mean level of changes: 1.89), with only around 8% of all aut-contours needing major adjustments (levels 3-4). The oncologist noted that PCMs are generally thin (3 mm), but slight widening was observed on number of presented examples. We explored the relationship between DSC metrics and clinical acceptability criteria, building on Heilemann et al.'s (44) suggestion of a DSC threshold above 0.7 for clinical acceptability. However, due to size-dependent characteristics, smaller ROIs may still be deemed acceptable with DSC below 0.7. Our results indicated that DSC below 0.6 tended to signify major adjustments, and while higher DSC values generally suggested better clinical acceptability, a straightforward correlation between DSC and adjustment levels was not apparent. Corrections didn't consistently correlate with the time required for manual adjustments, averaging 7.5 minutes per patient for non-clinically acceptable contours. Evaluation time averaged about 1.5 minutes per patient for the oncologist. Therefore, the entire process — generating, evaluating, and potentially adjusting some structures - averages under 10 minutes per patient, significantly quicker than manual delineation. The different clinical acceptability outcome for the same patient on different days suggests subjectivity, potentially addressable with deep learning strategies.

The dosimetric impact revealed higher average absolute dosimetric differences for contours needing more adjustments, with outliers stemming from OAR proximity to high-dose regions and steep dose falloff. Achieving the clinical goal for the elective target volume heavily relies on precise contours; expanding the exposed range to cover any shape often meets goals but lacks clinical acceptability. Median dosimetric differences between clinically acceptable contours and those needing minor adjustments (levels 1-2) are quite similar. PCMs and neck nodes requiring minor adjustments show slightly lower median values than clinically acceptable ones. Except for the spinal cord, average dosimetric differences between algorithm-generated manually adjusted contours and manual delineated ones surpass median dosimetric differences between automated and manual contours. In most cases, contours needing minor changes can be used without significant dosimetric impact changes. Higher dosimetric differences are observed for contours needing level 3+ changes, representing only 8% of automated contours. Correcting these takes an average of about 1.5 minutes per organ, varying with organ size. Notably, the dosimetric analysis for best, median, and worst performance echoed general findings, highlighting an intriguing case where brainstem misclassification led to a significant dosimetric difference despite being categorized as needing only level 1 adjustments (see Figure 2A).

After thorough evaluation, we are confident in the algorithm's effectiveness for contouring the parotid glands, brainstem, and mandible. While the outcomes for pharyngeal constrictor muscles were less satisfactory, a detailed dosimetric investigation showed minimal dosimetric differences in most cases. The algorithm shows promise for automating segmentation of the elective target volume and spinal cord, though additional refinements are needed for precision.

A key limitation in this and similar studies conducted on MR scans, is the limited availability of high-quality data. There are vast amounts of delineated CTs available, however consistent planning MRI data collection has only recently started. We utilized the entirety of the accessible data, resulting in a composite dataset with both T1 and T2-weighted MR-Linac scans. This combination may have negatively affected automated segmentation precision. Future research could explore using separate models for T1-weighted and T2-weighted scans, aiming for improved segmentation accuracy through such differentiation. Some of the other limitations of the current study were that primary target information was provided only for 13 of the patients. This allowed us to perform dosimetric analysis only for a small proportion of the patient population and cannot state for certain that the findings will remain the same when tested on larger patient population. Future studies would benefit of primary target information for all patients in order to perform more generalized dosimetric analysis. Furthermore, contouring of the primary target cannot be attempted with the current available data. Delineation of primary target varies among experienced clinicians and requires additional sequences (e.g., T1 post Gd or T2 SPAIR) along with endoscopic findings to aid contouring, accounting for natural anatomical barriers to tumor spread, such as air or bone. Another limitation is our reliance on contours delineated by a single oncologist as the ground truth. The clinical acceptability test showed that not all of these contours would be considered acceptable by another expert, highlighting the influence of inter-observer variability specifically, for smaller ROIs such as the pharyngeal constrictor muscles, the DSC is relatively low, ranging between 0.50 and 0.53. To enhance the model's learning, incorporating contours from multiple experts would be beneficial. Furthermore, evaluating the results by only one oncologist could lead to personal bias. Therefore, incorporating clinical acceptability evaluations by multiple different experts for each task could offer a robust solution to enhance the validity and reliability of our findings. However, our oncologists have been through multiple quality assurance exercises aligned with

established international benchmarks, such as the Gregoire et al. (45) atlas for nodal contouring. This ensures the reliability of the 'clear' pass or fail outcome derived from this assessment.

## 5 Conclusion

Majority (67%) of contours of the elective target volume and organs at risk for HNC patients automatically generated by an in-house developed model were found to be clinically acceptable and could be used for treatment planning without any manual adjustments. Among structures categorized as unfit for clinical use, the majority (≈75%) required only minor adjustments and the dosimetric impact showed that not performing the changes did not lead to significant dosimetric differences in most scenarios. Significant dosimetric differences could be observed for this group only if the ROIs or parts of ROIs were located exactly at the steep dose gradient. The model reliably contoured the parotid glands, brainstem, and mandible. The outcomes for the pharyngeal constrictor muscles were acceptable and the dosimetric impact analysis reveals minimal differences in most cases. While the algorithm shows promise for automating segmentation of the elective target volume and spinal cord, refinements could be performed for acquiring required precision in these areas. The analysis for the structures requiring major adjustments led to the conclusion that the time required for these adjustments to be made is minimal (on average 1min 4s per OAR and 4mins 39s per nCTV). Thus, delineation for HNC patients could be significantly sped up and the presented model could be used for initial delineation and subsequent re-delineation for each treatment fraction.

## Data availability statement

The dataset (MRI scans) used for this article was provided by the MOMENTUM study and the Royal Marsden Hospital for which the authors have directly applied. The dataset itself is not presented in the article. All presented analysis is original and there are no restrictions that apply. Requests to access these datasets should be directed to the Data Management Task Force, MOMENTUM@lygature.org.

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. MOMENTUM (NCT04075305) stands as an observational cohort study, operating as a collaborative effort between various institutions globally and Elekta (Stockholm, Sweden). Approval from the Institutional Review Board (IRB) was obtained at each center involved. Patients were provided with the opportunity to consent to the collection of health-related quality of life (HRQOL) data.

## Author contributions

VK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. BE: Software, Supervision, Writing – review & editing. AD: Conceptualization, Supervision, Validation, Writing – review & editing. AG: Data curation, Writing – review & editing. TG: Writing – review & editing. KW: Validation, Writing – review & editing. SB: Writing – review & editing. SN: Supervision, Writing – review & editing. KH: Supervision, Writing – review & editing. UO: Funding acquisition, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors wish to thank the Oracle Cancer Trust (21-ICR0420) for funding this study. Furthermore, we acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research. The Institute of Cancer Research is supported by Cancer Research UK Programme Grant C33589/A28284.

## Conflict of interest

Authors VK, BE, AD, AG, TG, KHW, SN, KH, and UO declare the following financial interests/personal relationships which may be considered as potential competing interests: The Institute of Cancer Research and The Royal Marsden Hospital are members of the Elekta MR Linac Consortium.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Author disclaimer

The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1358350/full#supplementary-material

# References


1. Sonke JJ, Aznar M, Rasch C. Adaptive radiotherapy for anatomical changes. *Semin radiat. Oncol.* (2019) 29:245–57. doi: 10.1016/j.semradonc.2019.02.007

2. Castadot P, Lee JA, Geets X, Grégoire V. Adaptive radiotherapy of head and neck cancer. *Semin radiat. Oncol.* (2010) 20:84–93. doi: 10.1016/j.semradonc.2009.11.002

3. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol.* (1997) 42:123. doi: 10.1088/0031-9155/42/1/008

4. Hunt A, Hansen V, Oelfke U, Nill S, Hafeez S. Adaptive radiotherapy enabled by MRI guidance. *Clin Oncol.* (2018) 30:711–9. doi: 10.1016/j.clon.2018.08.001

5. Morgan HE, Sher DJ. Adaptive radiotherapy for head and neck cancer. *Cancers Head Neck.* (2020) 5:1–16. doi: 10.1186/s41199-019-0046-z

6. Brunenberg EJ, Steinseifer IK, van den Bosch S, Kaanders JH, Brouwer CL, Gooding MJ, et al. External validation of deep learning-based contouring of head and neck organs at risk. *Phys Imaging radiat. Oncol.* (2020) 15:8–15. doi: 10.1016/j.phro.2020.06.006

7. Trotti A. Toxicity in head and neck cancer: a review of trends and issues. *Int J Radiat. Oncol Biol Phys.* (2000) 47:1–12. doi: 10.1016/S0360-3016(99)00558-1

8. Brady LW, Heilmann H, Molls M. *New technologies in radiation oncology* Vol. 86. Springer Berlin, Heidelberg, Germany:Springer (2006) p. 321–36.

9. Winkel D, Bol GH, Kroon PS, van Asselen B, Hackett SS, Werensteijn-Honingh AM, et al. Adaptive radiotherapy: the Elekta Unity MR-linac concept. *Clin Trans radiat. Oncol.* (2019) 18:54–9. doi: 10.1016/j.ctro.2019.04.001

10. Ahunbay EE, Peng C, Chen GP, Narayanan S, Yu C, Lawton C, et al. An on-line replanning scheme for interfractional variations. *Med Phys.* (2008) 35:3607–15. doi: 10.1118/1.2952443

11. Christiansen RL, Johansen J, Zukauskaite R, Hansen CR, Bertelsen AS, Hansen O, et al. Accuracy of automatic structure propagation for daily magnetic resonance image-guided head and neck radiotherapy. *Acta Oncol.* (2021) 60:589–97. doi: 10.1080/0284186X.2021.1891282

12. Oh S, Kim S. Deformable image registration in radiation therapy. *Radiat Oncol J.* (2017) 35:101. doi: 10.3857/roj.2017.00325

13. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys.* (2017) 44:547–57. doi: 10.1002/mp.12045

14. Fritscher K, Raudaschl P, Zaffino P, Spadea MF, Sharp GC, Schubert R. Deep neural networks for fast segmentation of 3D medical images. In: *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2016: 19th International Conference, Athens, GREECE, October 17-21, 2016, Proceedings, Part II 19*. Springer Cham, Switzerland: Springer (2016). p. 158–65.

15. Kieselmann JP, Fuller CD, Gurney-Champion OJ, Oelfke U. Auto-segmentation of the parotid glands on MR images of head and neck cancer patients with deep learning strategies. *MedRxiv.* (2020) 2020–12. doi: 10.1101/2020.12.19.20248376

16. Yang Q, Zhang S, Sun X, Sun J, Yuan K. (2019). Automatic segmentation of head-neck organs by Multimode CNNs for radiation therapy, in: *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Shenzhen, China. pp. 1–5. IEEE.

17. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of simpleITK. *Front neuroinform.* (2013) 7:45. doi: 10.3389/fninf.2013.00045

18. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R. SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *J Digit. Imaging.* (2018) 31:290–303. doi: 10.1007/s10278-017-0037-8

19. de Mol van Otterloo SR, Christodouleas JP, Blezer EL, Akhiat H, Brown K, Choudhury A, et al. The MOMENTUM study: an international registry for the evidence-based introduction of MR-guided adaptive therapy. *Front Oncol.* (2020) 10:1328. doi: 10.3389/fonc.2020.01328

20. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Int Conf Med image Comput comput-assist. interv.* (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

21. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: large-Scale Machine Learning on Heterogeneous Systems*. arXiv CoRR, Cornwell University: Ithaca, New York (2015). 9351

22. Chollet, Francois. (2015). Keras.

23. Zhang Y, Liao Q, Ding L, Zhang J. Bridging 2D and 3D segmentation networks for computationefficient volumetric medical image segmentation: An empirical study of 2.5 D solutions. *Computerized Med Imaging Graphics.* (2022) 99:102088. doi: 10.1016/j.compmedimag.2022.102088

24. Avesta A, Hossain S, Lin M, Aboian M, Krumholz HM, Aneja S. Comparing 3D, 2.5 D, and 2D approaches to brain image segmentation. *medRxiv.* (2022), 2022–11. doi: 10.1101/2022.11.03.22281923

25. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* (2015) 48:2839–46. doi: 10.1016/j.patcog.2015.03.009

26. Jadon S. (2020). A survey of loss functions for semantic segmentation, in: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Viña del Mar, Chile. pp. 1–7. IEEE.

27. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv prepr. arXiv:1412.6980.* (2014). doi: 10.48550/arXiv.1412.6980

28. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, 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) 144:152–8. doi: 10.1016/j.radonc.2019.10.019

29. Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiat. Oncol.* (2021) 16:101. doi: 10.1186/s13014-021-01831-4

30. Dai X, Lei Y, Wang T, Zhou J, Rudra S, McDonald M, et al. Multi-organ auto-delineation in head-andneck mri for radiation therapy using regional convolutional neural network. *Phys Med Biol.* (2022) 67:025006. doi: 10.1088/1361-6560/ac3b34

31. Korte JC, Hardcastle N, Ng SP, Clark B, Kron T, Jackson P. Cascaded deep learning-based autosegmentation for head and neck cancer patients: organs at risk on T2-weighted magnetic resonance imaging. *Med Phys.* (2021) 48:7757–72. doi: 10.1002/mp.15290

32. Kawahara D, Tsuneda M, Ozawa S, Okamoto H, Nakamura M, Nishio T, et al. Deep learning-based auto segmentation using generative adversarial network on magnetic resonance images obtained for head and neck cancer patients. *J Appl Clin Med Phys.* (2022) 23:e13579. doi: 10.1002/acm2.13579

33. 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:e929–50. doi: 10.1002/mp.14320

34. Thor M, Iyer A, Jiang J, Apte A, Veeraraghavan H, Allgood NB, et al. Deep learning auto-segmentation and automated treatment planning for trismus risk reduction in head and neck cancer radiotherapy. *Phys Imaging Radiat Oncol.* (2021) 19:96–101. doi: 10.1016/j.phro.2021.07.009

35. Zhong Y, Yang Y, Fang Y, Wang J, Hu W. A preliminary experience of implementing deep-learning based auto-segmentation in head and neck cancer: a study on real-world clinical cases. *Front Oncol.* (2021) 11:638197. doi: 10.3389/fonc.2021.638197

36. Radici L, Ferrario S, Borca VC, Cante D, Paolini M, Piva C, et al. Implementation of a commercial deep learning-based auto segmentation software in radiotherapy: evaluation of effectiveness and impact on workflow. *Life.* (2022) 12:2088. doi: 10.3390/life12122088

37. 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) 22:104. doi: 10.1186/s12938-023-01159-y

38. He Y, Song F, Wu W, Tian S, Zhang T, Zhang S, et al. MultiTrans: Multi-scale feature fusion transformer with transfer learning strategy for multiple organs segmentation of head and neck CT images. *Med Novel Technol Devices.* (2023) 18:100235. doi: 10.1016/j.medntd.2023.100235

39. Zhang S, Wang H, Tian S, Zhang X, Li J, Lei R, et al. A slice classification model-facilitated 3D encoder–decoder network for segmenting organs at risk in head and neck cancer. *J Radiat Res.* (2021) 62:94–103. doi: 10.1093/jrr/rraa094

40. Strijbis VI, Dahele M, Gurney-Champion OJ, Blom GJ, Vergeer MR, Slotman BJ, et al. Deep learning for automated elective lymph node level segmentation for head and neck cancer radiotherapy. *Cancers.* (2022) 14:5501. doi: 10.3390/cancers14225501

41. van der Veen J, Gulyban A, Nuyts S. Interobserver variability in delineation of target volumes in head and neck cancer. *Radiother. Oncol.* (2019) 137:9–15. doi: 10.1016/j.radonc.2019.04.006

42. van der Veen J, Gulyban A, Willems S, Maes F, Nuyts S. Interobserver variability in organ at risk delineation in head and neck cancer. *Radiat Oncol.* (2020) 16:1–11. doi: 10.21203/rs.3.rs-39575/v2

43. Turcas A, Leucuta D, Balan C, Clementel E, Gheara C, Kacso A, et al. Deep-learning magnetic resonance imaging-based automatic segmentation for organs-at-risk in the brain: Accuracy and impact on dose distribution. *Phys Imaging Radiat. Oncol.* (2023) 27:100454. doi: 10.1016/j.phro.2023.100454

44. Heilemann G, Buschmann M, Lechner W, Dick V, Eckert F, Heilmann M, et al. Clinical implementation and evaluation of auto-segmentation tools for multi-site contouring in radiotherapy. *Phys Imaging Radiat Oncol.* (2023) 28:100515. doi: 10.1016/j.phro.2023.100515

45. Grégoire V, Ang K, Budach W, Grau C, Hamoir M, Langendijk JA, et al. Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiotherapy Oncol.* (2014) 110:172–81. doi: 10.1016/j.radonc.2013.10.010

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 25 March 2024
DOI 10.3389/fonc.2024.1362025



OPEN ACCESS

EDITED BY
Jun-Lin Yi,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Dominic Moon,
University of Texas Southwestern Medical Center, United States
Carla Pisani,
Azienda Ospedaliero Universitaria Maggiore della Carità, Italy

*CORRESPONDENCE
Elena Riggenbach
elena.riggenbach@insel.ch

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work and share last authorship

RECEIVED 27 December 2023
ACCEPTED 08 March 2024
PUBLISHED 25 March 2024

CITATION
Riggenbach E, Waser M, Mueller SA, Aebersold DM, Giger R and Elicin O (2024) Oncologic outcome with versus without target volume compartmentalization in postoperative radiotherapy for oral cavity squamous cell carcinoma.
*Front. Oncol.* 14:1362025.
doi: 10.3389/fonc.2024.1362025



# Oncologic outcome with versus without target volume compartmentalization in postoperative radiotherapy for oral cavity squamous cell carcinoma

Elena Riggenbach[1*†], Manuel Waser[2†], Simon A. Mueller[2,3], Daniel M. Aebersold[1], Roland Giger[2‡] and Olgun Elicin[1‡]

[1]Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, [2]Department of Otorhinolaryngology, Head and Neck Surgery, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, [3]Department of Otorhinolaryngology Head and Neck Surgery, University Hospital and University of Zurich, Zurich, Switzerland

**Background and purpose:** The volume treated with postoperative radiation therapy (PORT) in patients with oral cavity squamous cell carcinoma (OCSCC) is a mediator of toxicity affecting quality of life. Current guidelines only allow for very limited reduction of PORT volumes. This study investigated the safety and efficacy of de-intensified PORT for patients with OCSCC by refined compartmentalization of the treatment volume.

**Materials and methods:** This retrospective cohort study identified 103 OCSCC patients treated surgically from 2014 to 2019 with a loco-regional risk profile qualifying for PORT according to guidelines. PORT was administered only to the at-risk compartment and according to a refined compartmentalization concept (CC). Oncological outcome of this CC cohort was compared to a historical cohort (HC) of 98 patients treated before the CC was implemented.

**Results:** Median follow-up time was 4.5 and 4.8 years in the CC and HC cohorts, respectively. In the CC cohort, a total of 72 of 103 patients (70%) had a pathological risk profile that allowed for further compartmentalization and, hence, received a reduced treatment volume or omission of PORT altogether. Loco-regional control at 3 and 5 years was 77% and 73% in the CC cohort versus 78% and 73% in the HC ($p$ = 0.93), progression-free survival was 72% and 64% versus75% and 68% ($p$ = 0.58), respectively. Similarly, no statistically significant difference was seen in other outcome measures.

**Conclusions:** De-intensified PORT limiting the treatment volume to the at-risk compartment or avoiding PORT altogether for low-risk patients with OCSCC does not seem to compromise disease control in this retrospective comparison. Based on these hypothesis-generating findings, a prospective study is being planned.

KEYWORDS

**head and neck cancer, oral cavity cancer, head and neck squamous cell carcinoma, postoperative radiotherapy, de-escalation, compartmentalization, radiotherapy, head and neck surgery**

# 1 Introduction

Oral cavity squamous cell carcinoma (OCSCC) represents one of the most frequently diagnosed head and neck malignancy. Despite advances in treatment strategies and technology, OCSCC remains a significant cause of morbidity and mortality (1). Primary surgery with or without postoperative radio(chemo)therapy or primary radio(chemo)therapy are treatment options for patients with OCSCC. While different approaches are effective, they incur long-term morbidity that escalates with treatment intensity (2, 3). Reducing target dose and volume in radiotherapy (RT) or omitting postoperative RT (PORT) altogether are important potential toxicity-mitigation strategies that may improve quality of life (4, 5). To maintain oncological outcomes while reducing the toxicity, an appropriate definition of candidates qualifying for RT volume reduction or even complete omission due to a lower risk of recurrence is essential.

Many factors influence survival and loco-regional tumor control in patients with head and neck cancers. The presence of remaining postoperative microscopic/macroscopic disease at the margins of resection (R1/R2) and/or the presence of extracapsular extension of nodal disease (ECE) in the neck have been clearly defined as poor prognostic features. In patients with these high-risk features, both postoperative radiation therapy (PORT) and additional concurrent chemotherapy (6–9) improve loco-regional control as well as overall survival (OS). The presence of other "minor" adverse risk factors, such as multiple positive lymph nodes (without ECE), perineural invasion (Pn1), vascular invasion (V1), lymphatic invasion (L1), pT3 or pT4 primary, and oral cavity primary cancers with positive lymph nodes in level IV or V, are generally established indications for PORT as well. The direct individual association of each of the minor risk factors on local, regional, or general outcome is, however, not clear.

There is no consensus to whether the primary tumor bed and each hemi-neck of the nodal basin should be considered as separate target compartments when these risk factors arise either only in the primary tumor bed or (hemi-)neck. According to current guidelines (10), the only accepted compartmentalization strategy in PORT for OCSCC is to spare the contralateral neck in case of a lateralized primary with local factor (R+ and/or >1 minor factor) and node negative disease after neck dissection. However, some aspects of PORT target volume definition are based on tradition, rather than evidence. Compared to the recommendations in the current guidelines, our institutional compartmentalization concept (CC) allows further de-intensification of PORT considering the tumor bed and each hemi-neck as three separate compartments for adjuvant RT decisions (Figure 1). Compared to the traditional holistic approach, the intent of the CC is to apply the required dose only to the compartment under risk.

**Abbreviations:** CC, compartmentalization concept; DMFS, distant metastasis-free survival; ECE, extracapsular extension of nodal disease; HC, historical cohort; L1, lymphatic invasion; LC, isolated local control; LRC, loco-regional control; OCSCC, oral cavity squamous cell carcinoma; OS, overall survival; Pn1, perineural invasion; PORT, postoperative radiation therapy; PFS, progression-free survival; RC, isolated regional control; V1, vascular invasion.

The current study investigates further compartmentalization strategies in patients with operated OCSCC mandating PORT and compares their oncological outcome to a historical cohort (HC), where no compartmentalization was implemented.

# 2 Materials and methods

## 2.1 Study design and patient selection

A retrospective cohort study design was pursued following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (11). The study was approved by the regional ethics committee.

Eligible patients were 18 years of age or older with histologically confirmed diagnosis of OCSCC, treated with curatively intended primary surgery from January 2014 to March 2019. The study cohort was limited to patients having one of the following unfavorable loco-regional risk factors: close resection margin of histopathologically less than 5 mm from the tumor, perineural invasion, lympho-vascular space invasion, tumor (T-) stage ≥3, or more than one positive neck lymph node. PORT was administered according to our refined CC detailed below. Oncological outcome of this CC cohort was compared to a HC diagnosed and treated in our institution from January 2007 to December 2013. The period of the HC was started by the standard establishment of concomitant systemic treatment regimens, including cetuximab (12) (as an extrapolation from the definitive RT setting), which was similar to the period of treatment of the CC cohort. All patients in both cohorts were treated with intensity modulated RT techniques.

Patients with a previous head and neck squamous cell carcinoma (HNSCC) or previous RT to the head and neck area before the treatment course under investigation, or an active synchronous cancer at the start of treatment were excluded.

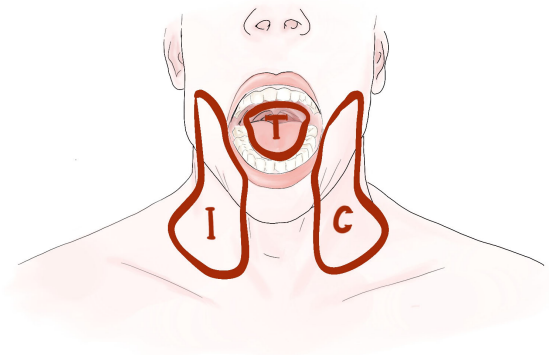



FIGURE 1
Illustrating the three compartments (T = tumor bed, I = ipsilateral neck, and C = contralateral neck) for postoperative radiotherapy in oral cavity squamous cell carcinoma.

## 2.2 Treatment procedures and compartmentalization approach

Resection of the primary tumor was required with additional ipsi- or bilateral neck dissection according to multidisciplinary tumor board decision. All patients in the CC cohort met the criteria for PORT according to international consensus criteria (6, 10, 13). PORT was administered according to a clearly defined risk profile and only to the at-risk compartment (Figure 2). For RT of the primary tumor compartment we seek either an inadequate resection margin (R1/R2 or close) as a major risk factor or at least two minor risk factors (Pn1, L1, V1) with the exception of Pn1 with a nerve diameter >0.1 mm or "named nerve" that is enough as a sole factor warranting irradiation of the primary tumor bed (14, 15). We consider the operated pT3-4 tumor bed as pathological low-risk profile if no other risk factor is present (16, 17). To irradiate each hemi-neck, we seek either the presence of ECE or at least two involved lymph nodes. The theoretical treatment volumes (i.e., the compartments) according to the international consensus were recorded retrospectively for each patient depending on the pathological risk profile.

When treatment to the three compartments (i.e., tumor bed, ipsilateral and contralateral hemi-neck) according to CC was discordant from the guideline-conform PORT, it was classified into five categories: tumor bed spared (CC#1T), one hemi-neck spared (CC #1N), both hemi-necks spared (CC#2NN), tumor bed and one hemi-neck spared (CC#2TN), and all three compartments spared (CC#3TNN).

We detail our CC in Figure 2 for various clinical settings and classify into five different variations of CC, depending on the compartment spared when comparing the treatment to the three compartments according to CC with the treatment corresponding scenarios governed by the pathologic risk profile. Patients in which the application of the refined CC did not result in a deviation from international consensus were classified as "no compartmentalization used".

The concomitant systemic treatment was prescribed according to the international consensus guidelines, namely, based on the results of the EORTC 22931 (9) and RTOG 9501 (7), and Cetuximab (12) was used as substitute for cisplatinum-ineligible patients.

## 2.3 Diagnosis and follow-up

All treatment recommendations were discussed at the head and neck cancer specific multi-disciplinary tumor board after the initial histopathologic confirmation of OCSCC and again postoperatively, concerning the need of an adjuvant treatment. Our standard follow-up protocol is provided in Supplementary Table S1. The sequence and modalities of the diagnostic work-up were similar in the CC cohort and HC. In both cohorts, lymph node levels of the neck dissection were separated and individually marked during surgery before sending off to pathology. The number of positive (with and without ECE) and the total number of harvested lymph nodes were reported separately for each level by the pathologists. Staging for all patients was done according to the 7th edition of UICC (18).

## 2.4 Statistical analysis

The primary endpoint of this retrospective cohort study was to estimate the rate of loco-regional control (LRC), defined as the time from the date of histopathological diagnosis to the first documented local and/or regional recurrence. Median follow-up time was calculated by excluding the deceased patients.

Secondary endpoints included isolated local (LC), isolated regional (RC), LRC, and distant control (DMFS), progression-free survival (PFS) and OS. Kaplan–Meier method was used to depict survival curves for the oncologic endpoints, and the log-rank test for group comparisons. Analyses were performed using JMP® statistical software (Version 16.2.0; SAS Institute Inc., Cary, North Carolina).

# 3 Results

## 3.1 Patient characteristics and treatment variables

The baseline demographic and clinical characteristics were comparable in both groups (Table 1). Of 187 consecutive patients with OCSCC treated with curative intent in our hospital from January 2014 to March 2019, 103 patients had a pathological risk profile based on which, an adjuvant treatment (i.e., PORT with or without concomitant systemic treatment) was indicated according to international guidelines.

The CC cohort comprised these 103 patients. The median age was 62 years (range, 28–95) and 60.2% of patients were male. Median number of total harvested lymph nodes was 48 (range, 17–128). Thirty-six patients (35%) had a node positive disease. Node positive to harvested ratio was median 2% (range, 0%–28%). Treatment strategy was surgery alone (without adjuvant therapy) in 25 patients (24.3%), surgery followed by PORT in 33 patients (32.0%), and PORT with concomitant systemic treatment in 45 patients (43.7%). Low-risk volume in PORT received a median dose of 50 Gy (range, 50–54), intermediate-risk volume 60 Gy (range, 55–60) and high-risk volume 66 Gy (range, 2–68) in 2 Gy daily fractions.

All 98 patients from the HC received standard PORT according to guidelines, and no further compartmentalization strategy was implemented at that time. As with the treatment technique, RT doses in the HC for low-, intermediate-, and high-risk volume did not differ from the CC cohort. Due to the amount of missing data, detailed patient and tumor characteristics such as the anatomical subsites, smoking, and alcohol consumption as well as some pathologic factors (L1, V1, and Pn1) were not extracted.

## 3.2 Compartmentalization concept

Of the CC cohort, 52 patients (50.4%) would have had the theoretical indication for irradiation of the tumor bed and the unilateral neck, while 51 patients (49.5%) would have had the indication to irradiate all three compartments (tumor bed and bilateral neck) according to the standard protocol. Due to the implementation of CC, only 20 patients (19.4%) were effectively

10.3389/fonc.2024.1362025

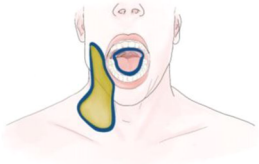
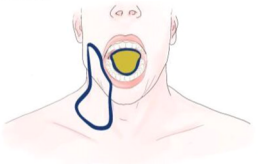
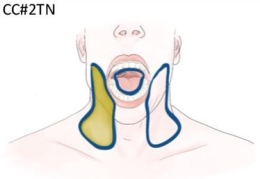
| Risk profile for each of the 3 compartments (tumor bed: RF-P, ipsilateral and contralateral neck: RF-$N_{ipsi}$, RF-$N_{contra}$) | # of CC | Compartmentalization concept (CC) | | RT indication according to guidelines (Evans et al. 2018) |
|---|---|---|---|---|
| | | Irradiated compartment | Spared compartment | |
| **RF-P**: low-risk profile (L1 or V1 allowed) + well lateralized tumor<br>**RF-$N_{ipsi}$**: ECE+ or ≥2+LN<br>**RF-$N_{contra}$**: no RF (pN0-2a w/o ECE) | CC#1T  | $N_{ipsi}$ | $P_{bed}$ | $P_{bed}$ + $N_{ipsi}$ |
| **RF-P**: ≥2 minor RF (Pn1*, L1, V1) or ≥1 major RF (R+ or $R_{close}$) + well lateralized tumor<br>**RF-$N_{ipsi}$**: no RF (pN0-2a w/o ECE)<br>**RF-$N_{contra}$**: no RF (pN0-2a w/o ECE) | CC #1N  | $P_{bed}$ | $N_{ipsi}$ | $P_{bed}$ + $N_{ipsi}$ |
| **RF-P**: ≥2 minor RF (Pn1*, L1, V1) or ≥1 major RF (R+ or $R_{close}$) + not well lateralized and/or midline-crossing tumor<br>**RF-$N_{ipsi}$**: no RF (pN0-2a w/o ECE)<br>**RF-$N_{contra}$**: no RF (pN0-2a w/o ECE) | CC#2NN  | $P_{bed}$ | $N_{ipsi}$ + $N_{contra}$ | $P_{bed}$ + $N_{ipsi}$ + $N_{contra}$ |
| **RF-P**: low-risk profile (L1 or V1 allowed) + not well lateralized or midline tumor<br>**RF-$N_{ipsi}$**: ECE+ OR ≥2+LN<br>**RF-$N_{contra}$**: no RF (pN0-2a w/o ECE) | CC#2TN  | $N_{ipsi}$ | $P_{bed}$ + $N_{contra}$ | $P_{bed}$ + $N_{ipsi}$ + $N_{contra}$ |
| **RF-P**: low-risk profile (L1 or V1 allowed) + not well lateralized or midline tumor<br>**RF-$N_{ipsi}$**: no RF (pN0) w/o ECE-<br>**RF-$N_{contra}$**: no RF (pN0-2a) w/o ECE- | CC#3TNN  | No RT | $P_{bed}$ + $N_{ipsi}$ + $N_{contra}$ | $P_{bed}$ + $N_{ipsi}$ + $N_{contra}$ |

FIGURE 2
Compartmentalization concept (CC): clinical criteria and risk factors in each of the three compartments for one of five strategies of compartmentalization for postoperative radiotherapy in oral cavity squamous cell carcinoma. For staging the 7th edition of the International Union against Cancer (UICC), staging system was used. Each hemi-neck is regarded distinctly for pN-staging. Strategies of CC (#1–5) are shown as images (blue contour = compartment treated according to guidelines, yellow fields = compartments treated according to CC). CC = compartmentalization concept; L1 = lymphatic invasion; $N_{contra}$ = compartment of the contralateral neck; $N_{ipsi}$ = compartment of the ipsilateral neck; $P_{bed}$ = compartment of the primary tumor bed; Pn1 = perineural invasion; R+ = positive resection margin; $R_{close}$ = close resection margin (less than 5 mm from the tumor); risk factors (RF) of the primary tumor bed (RF-P), ipsilateral neck (RF-$N_{ipsi}$) or contralateral neck (RF-$N_{contra}$); V1= vascular invasion. *: exception Pn1 with a nerve diameter >0.1 mm or «named nerve» is enough as a sole factor.

irradiated to all three compartments, while 23 patients (22.3%) received PORT to the tumor bed and unilateral neck, one patient (1.0%) to bilateral neck without the tumor bed, seven patients (6.8%) unilateral neck without the tumor bed, and 27 patients (26.2%) tumor bed only. In 25 patients (24.3%) with indication for PORT according to the standard protocol, PORT was omitted altogether under the CC.

Comparing the treatment and volume applied according to our CC with the guideline-conform indications, a total of 72 patients (69.9%) had a pathological risk profile that allowed further compartmentalization and hence received a reduced treatment volume. The category of compartmentalization of PORT applied in these 72 patients using the CC is shown in Table 2. In the remaining 31 patients (30.1%), no further compartmentalization was deemed safe, and they hence received a treatment volume identical to established guidelines.

## 3.3 Oncologic outcome

Median follow-up time for patients still alive was 4.5 years (range, 0.3–7.4) for the CC cohort and 4.8 years (range, 0.2–8.9) for the HC. None of the oncological outcome measures showed a statistically significant difference when comparing survival curves of the CC cohort with the HC, that is, the null hypothesis of the log-rank test was retained for LC, RC, LRC, PFS, DMFS, and OS

10.3389/fonc.2024.1362025

TABLE 1 Patient characteristics of the study cohorts.

| | Modern cohort (*n* = 103) | Historic cohort (*n* = 98) | *p*-value |
|---|---|---|---|
| | Percent (*n*) or median (range) | Percent (*n*) or median (range) | |
| **Age at diagnosis (y)** | 62 (28–95) | 60 (20–89) | 0.06 |
| **Sex** | | | 0.29 |
| Male | 60.2% (62) | 67.3% (66) | |
| Female | 39.8% (41) | 32.7% (32) | |
| **Primary tumor location within oral cavity** | | | |
| Tongue | 49.5% (51) | NA | |
| Floor of the mouth | 19.4% (20) | NA | |
| Alveolus and gingiva | 12.3% (16) | NA | |
| Buccal mucosa | 4.9% (5) | NA | |
| Hard palate | 1.0% (1) | NA | |
| Unclear/multiple sites infiltrated | 10.0% (10) | NA | |
| **Pathologic AJCC tumor stage (7th ed.)** | | | < 0.01 |
| pT1 | 27.2% (28) | 15.3% (15) | |
| pT2 | 39.8% (41) | 41.8% (41) | |
| pT3 | 7.8% (8) | 8.2% (8) | |
| pT4a | 24.3% (25) | 17.3% (17) | |
| pT4b | 1.0% (1) | 17.3% (17) | |
| **Pathologic AJCC nodal stage (7th ed.)** | | | < 0.01 |
| pN0 | 35.0% (36) | 43.9% (43) | |
| pN1 | 16.5% (17) | 17.3% (17) | |
| pN2a | 1.0% (1) | 3.1% (3) | |
| pN2b | 28.2% (29) | 21.4% (21) | |
| pN2c | 9.7% (10) | 14.3% (14) | |
| pNX | 9.7% (10) | | |
| **Pathologic AJCC stage classification (7th ed.)** | | | < 0.01 |
| Stage I | 15.5% (16) | 10.2% (10) | |
| Stage II | 12.5% (13) | 22.4% (22) | |
| Stage III | 18.4% (19) | 10.2% (10) | |
| Stage IVA | 50.5% (52) | 39.8% (39) | |
| Stage IVB | 2.9% (3) | 17.3% (17) | |
| **Presence of ENE in pN>0** | 46.3% (31) | 45.5% (25) | 0.06 |
| **Median number of LN involved** | 1 (0–16) | 1 (0–9) | 0.12 |
| **Adjuvant treatment** | | | < 0.01 |
| No adjuvant treatment | 24.3% (25) | 0% (0) | |
| Radiotherapy | 32.0% (33) | 46.9% (49) | |
| Radiotherapy with systemic treatment | 43.7% (45) | 53.1% (52) | |
| Cisplatin* 100 mg/m$^2$ three-weekly | 35.9% (37) | 44.9% (44) | |
| Cetuximab* weekly | 7.8% (8) | 8.2% (8) | |

AJCC, American Joint Commission on Cancer; ENE, extranodal extension; LN, Lymph node; RT, radiation therapy; NA, not available.
*Cisplatin concomitant 100 mg/m$^2$ three-weekly, cetuximab 400 mg/m$^2$ loading dose 1 week prior to radiotherapy and concomitant 250 mg/m$^2$ weekly.

(Table 3). Detailed actuarial survival data are presented in Figure 3.

Isolated nodal failure occurred in four out of 67 patients (6%) from the CC cohort, where irradiation to at least one hemi-neck was spared. However, two of those four recurrences occurred in the irradiated volume. Local failures were observed in four out of 33 patients (12%) from the CC cohort, where irradiation of the tumor bed was spared.

## 4 Discussion

In this retrospective study of OCSCC with adverse features, the efficacy and safety of further compartmentalizing PORT was evaluated. Compared to a HC, our approach did not impact loco-regional disease control or survival rates, suggesting our CC to be safe. To our knowledge, this is the first analysis implementing different compartmentalization strategies at once for a relatively large OCSCC cohort.

The indication for PORT for OCSCC is based on the presence of major (ECE, close or positive margins) and minor (Pn1, V1, L1, pT≥3, pN≥2, and lymph node involvement in level IV or V) pathologic risk factors. There is no consensus to whether the primary tumor bed and each hemi-neck of the nodal basin should be considered as separate target compartments when these risk factors arise either only in the (hemi-)neck or in the primary tumor bed. The only accepted compartmentalization strategy in PORT for OCSCC according to current guidelines (10) is, to omit the contralateral neck in case of a lateralized primary with local risk factors (R+ and/or >1 minor factor) and node negative disease after neck dissection.

TABLE 2 Compartmentalization strategy used in the CC (compartmentalization concept) cohort.

| # of CC | Spared compartment* | # of spared compartment(s) | Percent (*n*) |
|---|---|---|---|
| 1 | Tumor bed | 1 | 4.9% (5) |
| 2 | One hemi-neck | 1 | 26.2% (27) |
| 3 | Both hemi-necks | 2 | 11.7% (12) |
| 4 | Tumor bed and one hemi-neck | 2 | 24.3% (25) |
| 5 | All three compartments | 3 | 2.9% (3) |
| 0 | No compartmentalization used | 0 | 30.1% (31) |

*In total, PORT was omitted in a total of 25 (24.3%) patients through the compartmentalization strategy.

A de-escalation of PORT in our cohort of OCSCC was possible in 70% of the patients, either with reduction of the treatment volume by means of sparing the compartment(s) not harboring the corresponding risk factors (45.7%) or omitting PORT altogether (24.3%).

RT is associated with significant acute and long-term toxicities, primarily mediated by treatment volume and prescription dose. Thus, a reduction in radiation volume is expected to have a direct impact on acute and late toxicity as well as patients' quality of life. This, in turn, would potentially allow for selective treatment intensification if deemed necessary. Toxicity mitigation by compartmentalization is to be weighed against the risk of loco-regional recurrence in untreated compartments, but our CC demonstrates this to be feasible without impacting the oncological outcome. Compared to the classical holistic approach, the CC also increases the rate of feasibility in re-irradiation scenarios in terms of the application of the adequate dose in required target volumes.

## 4.1 Omitting the pN0 and/or pN1 neck

The recently demonstrated long-term results of a prospective phase II study supports the safety of omitting the pathologically negative (pN0) neck (19). In this mixed cohort of operated HNSCC ($n$ = 72 patients), including 14 patients with OCSCC, sparing the contralateral pN0 neck (74%) or even the ipsilateral pN0 neck (26%) resulted in an excellent unirradiated neck control of 97%. In our cohort, the omission of PORT to the hemi-neck was additionally allowed in the setting of a single-positive lymph node (pN1) without ECE. An earlier meta-analysis tested this approach but did not allow for general treatment recommendations due to large clinical heterogeneity of included studies (20). A more recent large retrospective study of patients with surgically treated OCSCC or oropharyngeal SCC shows that PORT to the pN1 neck in the absence of other adverse features might be associated with improved survival for pT2 disease or even pT1, especially in those younger than 70 years (21). While another, albeit smaller study only predicted a benefit for pN1 OCSCC if the lymph node yield at levels I–III was less than 20 (22). The hallmarks of a high-quality neck dissection, including at least 18 lymph nodes removed for levels I–III was again stressed in the recent ASCO guideline, making this a prerequisite for considering omission of the pN0 or pN1 neck in the PORT target volume (23). Although omission of RT to the pN1-neck without other adverse features seems to be currently accepted in most clinics, its controversy persists because PORT conferred a survival benefit in a recent large cancer registry-based study independent of adequacy of the neck dissection (24).

The lymph node yield in our CC cohort for patients receiving a neck dissection ensures the required quality. The abovementioned ASCO guideline (23) allows to omit PORT to the pN1 neck unless indications arising in the primary tumor, such as Pn1, L1, V1, or a pT≥3 primary are present. This prompts the question of whether in fact these local factors independently portend a higher risk for nodal recurrence. In a retrospective study on OCSCC, neither L1 nor V1 were independently associated with increased rates of regional or distant recurrence (25). A Japanese study group however revised their strategy of reducing treatment volumes when whole neck-PORT showed to be associated with a better OS, PFS, and LRC compared to limited-field-PORT in a retrospective analysis (26).

Isolated nodal failures in patients, where one or both hemi-necks were spared, was low (6%) in our cohort, suggesting a correct selection of patients with nodal low-risk profile (Figure 2) where irradiation of the nodal compartments might be safely omitted.

TABLE 3 Oncologic outcome with versus without target volume compartmentalization.

| | CC cohort | | Historical cohort | | |
|---|---|---|---|---|---|
| Outcome | 3-year (%) | 5-year (%) | 3-year (%) | 5-year (%) | *p*-value |
| Local control | 88.3 | 83.4 | 82.9 | 77.6 | 0.25 |
| Regional control | 87.6 | 87.6 | 85.8 | 83.8 | 0.64 |
| Locoregional control | 77.3 | 73.0 | 78.5 | 73.3 | 0.93 |
| Progression-free survival | 72.5 | 63.8 | 74.7 | 67.9 | 0.58 |
| Distant metastasis-free survival | 86.6 | 86.6 | 87.6 | 87.6 | 0.72 |
| Overall survival | 76.6 | 63.6 | 79.4 | 70.6 | 0.48 |

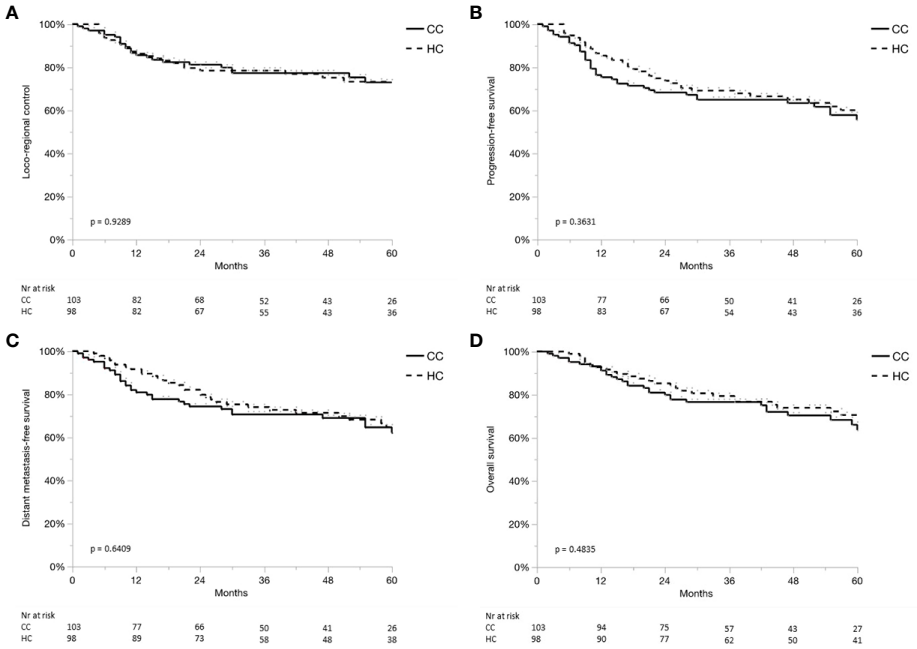



FIGURE 3
**(A–D)** Locoregional control **(A)**, progression-free survival **(B)**, distant metastases-free survival **(C)** and overall survival **(D)** for the compartmentalization cohort (CC, continuous line), and historical control (HC, dashed line).

## 4.2 Omitting the primary tumor bed (CC#1T and CC#2TN)

Another compartmentalization strategy for HNSCC is the omission of the primary tumor bed for patients with a favorable local risk profile that receive PORT to the at-risk areas in the involved neck. Why the postoperative primary tumor bed should be irradiated in the presence of multiple nodal positivity and/or ECE, whereas the same tumor bed would not receive any radiation if the neck is pN0-1 lacks a logical rationale (27). A national patterns of care study revealed no consensus on this issue with 70% of the centers not separating the tumor bed from the dissected nodal levels, and 30% allowing for this type of de-escalation (27).

The recent prospective single-arm phase 2 "AVOID" trial for human papilloma virus-associated oropharyngeal SCC has explored de-intensification for PORT, which resulted in a 2-year local recurrence-free survival of about 98%. The investigators concluded this to be a safe strategy worthy of further study (28). However, one has to be aware that the incidental dose to the primary tumor bed in oropharyngeal SCC is somewhere between 30 Gy and 43 Gy (29) even if only the neck is targeted, which might be high enough to effectively sterilize residual microscopic disease, especially for HPV-associated oropharyngeal SCC. A more significant dose reduction however occurs in the oral cavity due to geometrical relationship and distance between the nodal and primary tumor bed volumes (30) resulting in a more evident compartmentalization effect when sparing the primary tumor bed for OCSCC. In addition to the anatomical, and as a result, dosimetric differences when considering the CC, OCSCC and oropharyngeal SCC (especially HPV+) are known to be distinct diseases from a biological perspective.

According to the recently published AIRO-GORTEC consensus for early stage OCSCC (15), omitting irradiation of the primary tumor bed should be investigated further for OCSCC, given the promising results observed in oropharyngeal SCC. With our current study we provide data that omission of the primary tumor bed may also be safe for selected OCSCC patients.

In terms of limitations, our cohort suffers from the intrinsic problems due to its retrospective nature. Potential unknown confounding bias cannot be eliminated. For the HC, data on pathological risk profile was incomplete and could therefore not be analyzed in the same detail as the CC cohort.

Additionally, owing to lack of randomization there are inherent differences between the two cohorts. Mitigating the impact of this limitation through statistical approaches such as propensity score matching would have been a futile effort with the available sample size and data.

Last, but not least, toxicity and quality of life data is not reported as it was largely missing and not recorded in the same systematic manner in the HC. However, the dose-volume and response (toxicity and quality of life) relationship is well-known, with smaller treatment volumes and lower doses being associated with less treatment-related toxicity in head and neck cancer (31–34). Justifying a de-escalation approach as a potential improvement in quality of life is only warranted as far as the non-inferiority of recurrence is ensured.

Therefore, there is a tremendous international effort for de-escalation of dose and target volumes in HPV-associated oropharyngeal cancer compared to the non-HPV-associated head and neck cancer (35).

Given the extent of missing data in the HC and minor known und potential unknown differences between the two cohorts, our

exploratory comparative result should be considered as supplementary and is not the main emphasis of our paper.

## 5 Conclusions

With implementing a clearly defined strategy of further compartmentalization based on the pathologic risk profile in the respective compartment, a de-escalation of PORT is possible in the majority of OCSCC patients by reducing treatment volume or omitting PORT altogether. No compromise in disease control was seen when compared to a historical control. Based on these hypothesis-generating findings, a prospective trial is being designed.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Regional ethics committee of Bern (2021-00616). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

ER: Conceptualization, Data curation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. MW: Data curation, Writing – original draft, Writing – review & editing. SM: Data curation, Methodology, Writing – original draft, Writing – review & editing. DA: Conceptualization, Methodology, Resources, Writing – review & editing. RG: Conceptualization, Data curation, Investigation, Resources, Supervision, Validation, Writing – review & editing. OE: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1362025/full#supplementary-material

SUPPLEMENTARY TABLE 1
(anonymized and non-anonymized version).

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492

2. Rogers SN, Scott J, Chakrabati A, Lowe D. The patients' account of outcome following primary surgery for oral and oropharyngeal cancer using a "quality of life" questionnaire. *Eur J Cancer Care (Engl)*. (2008) 17:182–8. doi: 10.1111/j.1365-2354.2007.00832.x

3. Trotti A, Pajak TF, Gwede CK, Paulus R, Cooper J, Forastiere A, et al. TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. *Lancet Oncol*. (2007) 8:613–24. doi: 10.1016/S1470-2045(07)70144-4

4. Kelly JR, Husain ZA, Burtness B. Treatment de-intensification strategies for head and neck cancer. *Eur J Cancer*. (2016) 68:125–33. doi: 10.1016/j.ejca.2016.09.006

5. Spencer CR, Gay HA, Haughey BH, Nussenbaum B, Adkins DR, Wildes TM, et al. Eliminating radiotherapy to the contralateral retropharyngeal and high level II lymph nodes in head and neck squamous cell carcinoma is safe and improves quality of life. *Cancer*. (2014) 120:3994–4002. doi: 10.1002/cncr.28938

6. Bernier J, Cooper JS, Pajak TF, Van Glabbeke M, Bourhis J, Forastiere A, et al. Defining risk levels in locally advanced head and neck cancers: A comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (#9501). *Head Neck*. (2005) 27:843–50. doi: 10.1002/hed.20279

7. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. *N Engl J Med*. (2004) 350:1937–44. doi: 10.1056/NEJMoa032646

8. Huang DT, Johnson CR, Schmidt-Ullrich R, Grimes M. Postoperative radiotherapy in head and neck carcinoma with extracapsular lymph node extension and/or positive resection margins: A comparative study. *Int J Radiat Oncol Biol Phys*. (1992) 23:737–42. doi: 10.1016/0360-3016(92)90646-Y

9. Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med*. (2004) 350, 1945–52. doi: 10.1056/NEJMoa032641

10. Evans M, Beasley M. Target delineation for postoperative treatment of head and neck cancer. *Oral Oncol*. (2018) 86:288–95. doi: 10.1016/j.oraloncology.2018.08.011

11. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)

statement: guidelines for reporting observational studies. *J Clin Epidemiol*. (2008) 61:344–9. doi: 10.1016/j.jclinepi.2007.11.008

12. Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med*. (2006) 354:567–78. doi: 10.1056/NEJMoa053422

13. Haddad RI, Hicks WL, Hitchcock YJ, Jimeno A, Leizman D, Pinto HA, et al. NCCN Guidelines Version 1.2023 Head and Neck Cancers Continue NCCN Guidelines Panel Disclosures (2022). Available online at: https://www.nccn.org/home/member- (Accessed April 15, 2023).

14. Bakst RL, Glastonbury CM, Parvathaneni U, Katabi N, Hu KS, Yom SS. Perineural invasion and perineural tumor spread in head and neck cancer. *Int J Radiat Oncol Biol Phys*. (2019) 103:1109–24. doi: 10.1016/j.ijrobp.2018.12.009

15. Merlotti A, Alterio D, Orlandi E, Racadot S, Bonomo P, Franco P, et al. AIRO GORTEC consensus on postoperative radiotherapy (PORT) in low-intermediate risk early stages oral squamous cell cancers (OSCC). *Radiother Oncol*. (2022) 177:95–104. doi: 10.1016/j.radonc.2022.10.035

16. Fleury B, Thariat J, Barnoud R, Buiret G, Lebreton F, Bancel B, et al. Microscopic extensions of head and neck squamous cell carcinomas: Impact for CTV definition. *Cancer/Radiotherapie*. (2014) 18:666–71. doi: 10.1016/j.canrad.2014.04.006

17. Yuen PW, Lam KY, Chan ACL, Wei WI, Lam LK. Clinicopathological analysis of local spread of carcinoma of the tongue. *Am J Surg*. (1998) 175:242–4. doi: 10.1016/S0002-9610(97)00282-1

18. Sobin L, Gospodarowicz M, Wittekind C. *AJCC Cancer Staging Manual*. Oxford: Wiley-Blackwell (2009). doi: 10.1007/978-0-387-88441-7

19. Contreras JA, Spencer C, DeWees T, Haughey B, Henke LE, Chin RI, et al. Eliminating postoperative radiation to the pathologically node-negative neck: Long-term results of a prospective phase II study. *J Clin Oncol*. (2019) 37:2548–55. doi: 10.1200/JCO.19.00186

20. Moergel M, Meurer P, Ingel K, Wendt TG, Al-Nawas B. Effectiveness of postoperative radiotherapy in patients with small oral and oropharyngeal squamous cell carcinoma and concomitant ipsilateral singular cervical lymph node metastasis (pN1): A meta-analysis. *Strahlentherapie und Onkol*. (2011) 187:337–43. doi: 10.1007/s00066-011-2206-x

21. Chen MM, Harris JP, Hara W, Sirjani D, Vasu Divi M. Association of postoperative radiotherapy. *JAMAOtolaryngology–Head&Neck Surg*. (2016) 142:1224–30. doi: 10.1001/jamaoto.2016.3519

22. Feng Z, Xu QS, Qin LZ, Li H, Han Z. Predicting radiotherapy necessity in tongue cancer using lymph node yield. *J Oral Maxillofac Surg*. (2017) 75:1062–70. doi: 10.1016/j.joms.2016.10.005

23. Koyfman SA, Ismaila N, Crook D, D'Cruz A, Rodriguez CP, Sher DJ, et al. Management of the neck in squamous cell carcinoma of the oral cavity and oropharynx: ASCO clinical practice guideline. *J Clin Oncol*. (2019) 37:1753–74. doi: 10.1200/JCO.18.01921

24. Suresh K, Cramer JD. Postoperative radiation therapy vs observation for pN1 oral cavity squamous cell carcinoma. *Head Neck*. (2019) 41:4136–42. doi: 10.1002/hed.25958

25. Adel M, Kao HK, Hsu CL, Huang JJ, Lee LY, Huang Y, et al. Evaluation of lymphatic and vascular invasion in relation to clinicopathological factors and treatment outcome in oral cavity squamous cell carcinoma. *Med (United States)*. (2015) 94:1–7. doi: 10.1097/MD.0000000000001510

26. Makita C, Kodaira T, Daimon T, Tachibana H, Tomita N, Koide Y, et al. Comparisons of the clinical outcomes of different postoperative radiation strategies for treatment of head and neck squamous cell carcinoma. *Jpn J Clin Oncol*. (2017) 47:1141–50. doi: 10.1093/jjco/hyx137

27. Elicin O, Putora PM, Siano M, Broglie MA, Simon C, Zwahlen D, et al. A review of controversial issues in the management of head and neck cancer: A swiss multidisciplinary and multi-institutional patterns of care study—part 2 (radiation oncology). *Front Oncol*. (2019) 9:1126. doi: 10.3389/fonc.2019.01126

28. Swisher-McClure S, Lukens JN, Aggarwal C, Ahn P, Basu D, Bauml JM, et al. A phase 2 trial of alternative volumes of oropharyngeal irradiation for de-intensification (AVOID): omission of the resected primary tumor bed after transoral robotic surgery for human papilloma virus–related squamous cell carcinoma of the oropharynx. *Int J Radiat Oncol Biol Phys*. (2020) 106:725–32. doi: 10.1016/j.ijrobp.2019.11.021

29. Lazarev S, Todorov B, Tam J, Gupta V, Miles BA, Lee N, et al. Adjuvant radiation in the TORS era: Is there a benefit to omitting the tumor bed? *Pract Radiat Oncol*. (2017) 7:93–9. doi: 10.1016/j.prro.2016.08.002

30. Cui T, Ward MC, Kittel JA, Joshi N, Koyfman SA, Xia P. Dosimetric benefits of omitting primary tumor beds in postoperative radiotherapy after transoral robotic surgery using the auto-planning technique. *Cureus*. (2021) 13(9). doi: 10.7759/cureus.18065

31. Murphy BA, Gilbert J, Ridner SH. Systemic and global toxicities of head and neck treatment. *Expert Rev Anticancer Ther*. (2007) 7:1043–53. doi: 10.1586/14737140.7.7.1043

32. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol*. (2008) 26:3770–6. doi: 10.1200/JCO.2007.14.6647

33. Van den Bosch L, van der Schaaf A, van der Laan HP, Hoebers FJP, Wijers OB, van den Hoek JGM, et al. Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. *Radiother Oncol*. (2021) 157:147–54. doi: 10.1016/j.radonc.2021.01.024

34. Sher DJ, Moon DH, Vo D, Wang J, Chen L, Dohopolski M, et al. Efficacy and quality-of-life following involved nodal radiotherapy for head and neck squamous cell carcinoma: the INRT-AIR phase II clinical trial. *Clin Cancer Res*. (2023) 29:3284–91. doi: 10.1158/1078-0432.CCR-23-0334

35. Petrelli F, Luciani A, Ghidini A, Cherri S, Gamba P, Maddalo M, et al. Treatment de-escalation for HPV+ oropharyngeal cancer: A systematic review and meta-analysis. *Head Neck*. (2022) 44:1255–66. doi: 10.1002/hed.27019

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 May 2024
DOI 10.3389/fonc.2024.1394111



OPEN ACCESS

EDITED BY
Jun-Lin Yi,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Ciprian Camil Mirestean,
University of Medicine and Pharmacy of Craiova, Romania
Yu-Jie Huang,
Kaohsiung Chang Gung Memorial Hospital, Taiwan

*CORRESPONDENCE
Chaosu Hu
hucsu62@163.com

†These authors have contributed equally to this work

RECEIVED 01 March 2024
ACCEPTED 10 May 2024
PUBLISHED 30 May 2024

CITATION
Guan X, Peng J, Sun J, Xing X and Hu C (2024) Normal tissue complication probability model of temporal lobe injury following re-irradiation of IMRT for local recurrent nasopharyngeal carcinoma.
*Front. Oncol.* 14:1394111.
doi: 10.3389/fonc.2024.1394111



# Normal tissue complication probability model of temporal lobe injury following re-irradiation of IMRT for local recurrent nasopharyngeal carcinoma

Xiyin Guan[1,2,3†], Jiyou Peng[3,4,5†], Jiayao Sun[1,2,3], Xing Xing[3,4,5] and Chaosu Hu[1,2,3*]

[1]Department of Radiation Oncology, Shanghai Proton and Heavy Ion Center, Fudan University Cancer Hospital, Shanghai, China, [2]Shanghai Key Laboratory of Radiation Oncology, Shanghai, China, [3]Shanghai Engineering Research Center of Proton and Heavy Ion Radiation Therapy, Shanghai, China, [4]Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai, China, [5]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

**Purpose:** We tried to establish the normal tissue complication probability (NTCP) model of temporal lobe injury of recurrent nasopharyngeal carcinoma (NPC) patients after two courses of intensity modulated radiotherapy (IMRT) to provide more reliable dose-volume data reference to set the temporal lobe tolerance dose for recurrent NPC patients in the future.

**Methods and materials:** Recurrent NPC patients were randomly divided into training data set and validation data set in a ratio of 2:1, All the temporal lobes (TLs) were re-contoured as R/L structures and named separately in the MIM system. The dose distribution of the initial IMRT plan was deformed into the second course planning CT via MIM software to get the deformed dose. Equivalent dose of TLs in 2Gy fractions was calculated via linear quadratic model, using an $\alpha/\beta=3$ for temporal lobes. NTCP model that correlated the irradiated volume of the temporal lobe and? the clinical variables were evaluated in a multivariate prediction model using AUC analysis.

**Results:** From Jan. 2010 to Dec. 2020, 78 patients were enrolled into our study. Among which 26 (33.3%) developed TLI. The most important factors affecting TLI was the sum-dose d1.5cc of TL, while the possible clinical factors did not reach statistically significant differences in multivariate analysis. According to NTCP model, the TD5 and TD50 EQD2 dose of sum-dose d1.5cc were 65.26Gy (46.72–80.69Gy) and 125.25Gy (89.51–152.18Gy), respectively. For the accumulated EQD2 dose, the area under ROC shadow was 0.8702 (0.7577–0.9828) in model validation, $p<0.001$.

**Conclusion:** In this study, a NTCP model of temporal lobe injury after a second course of IMRT for recurrent nasopharyngeal carcinoma was established. TD5 and TD50 doses of temporal lobe injury after re-RT were obtained according to the model, and the model was verified by validation set data.

KEYWORDS

recurrent nasopharyngeal carcinoma, intensity modulated radiotherapy, re-irradiation, temporal lobe injury, normal tissue complication probability model

# Introduction

Nasopharyngeal carcinoma (NPC) is prevalent among Asians, particularly in Southern China, and is epidemiologically linked to Epstein-Barr virus infection, where the age-standardized incidence ranges from 15 to 50 cases per 100,000 population (1, 2). Now Intensity modulated radiotherapy (IMRT) is widely used as the primary treatment modality for non-metastasis NPC due to its anatomic location and radio-sensitivity. Local recurrence remains one of the common patterns of treatment failure. Overall, 10% to 20% local failures occur after definitive IMRT. Surgery is a preferred choice for small resectable superficial recurrent lesion (3, 4). Re-irradiation with IMRT remains the mainstay of treatment for advanced stage recurrence. However, due to the considerable critical organs surrounding the tumor, re-irradiation may lead to severe toxic side effects.

Temporal lobe injury (TLI) is a common late complication after re-irradiation for recurrent nasopharyngeal carcinoma, which is often manifested as memory decline, cognitive dysfunction, motor dysfunction, emotional disorders, language disorders, and other related symptoms, leading to a decline in the quality of life. IMRT can effectively limit the high-dose exposure of the temporal lobe. The TLI probability after the first course radiotherapy was about 4.6–16% (5–7). However, in recurrent nasopharyngeal carcinoma, especially for patients with large tumor volume, especially those with skull base invasion or intracranial invasion, a second course of high dose irradiation would be necessary, thus TLI is inevitable (8–10). Currently, there is very limited experience in determining the dose-volume tolerance of the temporal lobe for a second course of radiotherapy. In this study, we retrieved the first and second course of IMRT plan data for recurrent nasopharyngeal carcinoma patients and established the NTCP model of TLI based on clinical and dosimetric parameters. We believe this study would provide a more reliable reference for dose-volume data, and would provide assistance in the decision of temporal lobe dose limitation in the future.

**Abbreviations:** NPC, nasopharyngeal carcinoma; IMRT, Intensity modulated radiotherapy; TLI, Temporal lobe injury; NTCP, Normal tissue complication probability; GTV, Gross tumor volume; CTV, Clinical tumor volume; CT, Computed tomography; PET-CT, Positron emission tomography-computed tomography; MRI, magnetic resonance imaging; KPS, Karnofsky performance scale; DVH, dose volume histogram; EQD, equivalent dose; AUC, Operating characteristic curve; ROC, Operating characteristic curve; TD, Tolerance dose; QUANTEC, Quantitative Analysis of Normal Tissue Effects in the Clinic.

# Method and materials

## Inclusion and exclusion criteria

Inclusion criteria included:(1) Recurrent nasopharyngeal carcinoma confirmed by pathology or at least two imaging methods; (2) Both of the two courses of radiotherapy were using IMRT techniques, and the 2 courses of radiation plans were attainable; (3) Distant metastasis was excluded by chest CT, abdominal ultrasound, emission CT bone scan or whole body positron emission tomography-computed tomography (PET-CT); (4) Karnofsky performance scale (KPS) score ≥70; (5) Patients received complete 2 courses radiotherapy; (6) Patients were regularly followed up in the outpatient clinic with complete magnetic resonance images for at least every 6 months. Exclusion criteria included: (1) TLI occurred before the second course of radiotherapy; (2) The follow up time was less than 6 months; (3) Patient was unable to receive MRI to accurately assess TLI; (4) TLI cannot be differentiated from tumor progression or recurrence; (5) The two courses radiotherapy plan cannot be obtained completely.

## Immobilization and treatment plan

All initial and re-irradiation plans were obtained in Fudan University Shanghai Cancer Center. All patients can proceed to the immobilization, planning, and treatment process only after signing the informed consent for radiotherapy. Patient was immobilized in the supine position with a thermoplastic mask. CT was performed with slice thickness of 5mm after immobilization, ranging from1.5cm above the cranial vertex to at least 2cm below the sternoclavicular joint. The target volumes were delineated on CT images using Pinnacle (Pinnacle 3; Philips Corp, Fitchburg, WI) treatment planning system. Inverse IMRT plans were optimized using Pinnacle. For the initial course IMRT, the

total dose to primary tumor was 66 Gy in 30 fractions for T1 or T2 stage disease, and 70.4 Gy in 32 fractions for T3 or T4. A total dose of 60 Gy and 54 Gy was delivered to the high-risk and low-risk clinical tumor volume (CTV) in 30–32 fractions, respectively. For the re-irradiation course, only recurrent tumor and the positive involved lymph node regions were irradiated. The prescribed doses were 60–70 Gy to the gross tumor volume (GTV)and 50–60 Gy to the CTV, delivered in 25–35 fractions. The normal tissue constraints and plan evaluation were in accordance with the Radiation Therapy Oncology Group 0225 protocol. All the radiation were delivered using a simultaneous integrated boost-IMRT technique using Pinnacle. Patients with advanced T stage disease or positive lymph nodes received cisplatin-based induction or concurrent chemotherapy during IMRT.

## Image assessment and diagnostic criteria for TLIs

All the TLIs were diagnosed based on MRI findings. These abnormalities were verified by two radiologists, and any disagreements were resolved by consensus. Residual or progressive disease was excluded when determining the TLI site. The diagnostic criteria for TLI were as follows: (1) contrast-enhanced lesions, lesions with spotted or patchy enhancement with or without necrosis on post-contrast T1-weighted images; (2) white matter lesions, increased signal intensity on T2-weighted images in white matter; (3) cysts, round or oval lesions of very high signal intensity on T2-weighted images with a thin or imperceptible wall.

## Dose volume histogram data calculation

Both the two courses IMRT plan were imported into the MIM system (MIM software v6.5.9, Cleveland, OH, USA). To ensure precise delineation of the temporal lobe, all temporal lobes were re-contoured by the physician in the re-irradiation plan as R/L structures and named separately using MIM software and cross-checked by another experienced physician. In cases where the tumor infiltrated into the intracranial tissue, this specific region of the temporal lobe was delineated as normal tissue. The dose distribution of the first IMRT plan was deformed into the planning CT of the re-radiotherapy via MIM to get the deformed dose. Since the fractionation of the two IMRT plans were not identical, equivalent dose in 2Gy fractions was calculated via linear quadratic model, using an $\alpha/\beta$=3 for temporal lobes. The equivalent dose ($EQD_3^2$) of the deformed dose and the dose of re-radiotherapy was accumulated to obtain the accumulated dose based on former registration via Python program (v3.9.6). The dose volume histograms of the bilateral temporal lobes and the TLI of the deformed dose, re-irradiated dose, and accumulated dose were exported. Based on the DVH data, the max dose, the dose to 0.5–5cc in 0.5cc. increments were expressed as Dmax, and D0.5-D5cc.

$$EQD2 = Dx\ \frac{\frac{\alpha}{\beta} + d_x}{\frac{\alpha}{\beta} + D_2}$$

## Construction and validation of the NTCP model

Our NTCP model for temporal lobe was constructed based on multivariate logistic regression, formula of which is shown below as equation. x1, x2… xm are different input parameters; β0, β1… βm are the logistic regression coefficients of corresponding input parameters. Both dosimetric parameters and clinical factors were considered as potential input parameters in this model. Dosimetric parameters include D0.5cc-D5cc in 0.5cc increments and Dmax.

$$NTCP = \frac{1}{1 + e^{-(a+\sum_{i=1}^{m} b_i x_i)}}$$

All the patients were randomly divided into training set and validation set at a ratio of 2:1. Training set data were utilized for deriving model parameter, while validation set data were employed for accessing the model. Model construction involved two primary steps. Initially, three modes (linear, quadratic, exponential) were explored to assess the necessity of incorporating the time interval in the combination of doses from two courses. The verification process included the following steps: 1. Defining the range and stride of the parameter based on clinical data and time model formulation; 2. Applying parameter values to the time model to derive the combined dose; 3. Calculating dosimetric indices of the combined dose distribution; 4. Conducting univariate logistic analysis on the dosimetric index and obtaining Nagelkerke's R squared value; 5. Repeating steps 2–4 for different parameter values and various time models.

$$dose\_tol = (1 - a^{*}gap)^{*}dose1 + dose2$$

$$dose\_tol = (1 - a^{*}\ gap^{*}gap)^{*}\ dose1 + dose2$$

$$dose\_tol = \exp(-a^{*}\ gap)^{*}dose1 + dose2$$

Subsequently, multivariate logistic regression was conducted with different sets of factors. Considering diverse clinical scenarios, three protocols were presumed for broader application: 1st, the primary RT plan was unable to obtain, we only consider the 2nd RT plan dosage; 2nd, both the first and second RT plan were available, but we cannot calculate the accumulated dose; 3rd, both the RT plan were available and we use the accumulated EQD2 dose.

To validate the model, the area under the receiver operating characteristic curve (AUC) for the receiver operating characteristic curve (ROC) was calculated. Statistical analyses were performed using SPSS version 26.0 (IBA, SPSS Inc., Chicago, IL).

10.3389/fonc.2024.1394111

## Results

From Jan. 2010 to Dec. 2020, 78 patients were enrolled into our study. These patients were randomly divided into training data set and validation data set in a ratio of 2:1, yielding 52 patients in training set and 26 patients in validation set. There were no significant differences in clinical characteristics between these two sets. Characteristics of patients are listed in Table 1. The median follow-up time was 31 (range, 6–127) months. Among these patients, 26 (33.3%) developed TLI, among which 16 patients experiencing bilateral TLI and 10 patients with unilateral TL. The median latency for development of TLI (from beginning of re-irradiation to first MRI-detected TLI) was 11.5 (range, 3–29) months. The median interval between initial radiotherapy and re-irradiation was 26 (range, 12–108) months.

The 'a' value corresponding to maximum R2 value for various indices in the linear time model was presented in Figure 1. It can be observed that the dose-volume parameter corresponding to the maximum R2 value is D1.5cc, as indicated in Table 2. Results of the ROC curve analysis, using the 'a' value corresponding to the maximum R2 value in the linear time model as the model parameter value, were displayed in Table 3. However, equation incorporating the time factor exhibited minimal deviation from the value obtained by straightforwardly summing temporal lobe doses voxel to voxel, as shown in Figures 1–3. Therefore, the necessity of incorporating interval time into the model was not clearly evident in our data.

Multivariate NTCP model was derived by analyzing dosimetric variables, including relative dose delivered to specific volumes of temporal lobe (in 0.5 cc bins from Dmax to D5cc), and clinical factors, including primary and recurrent tumor stage, RT dose, tumor volume, time interval between two RT courses, chemotherapy, gender, and age. Possible clinical factors did not reach statistically significant differences in multivariate analysis, details were shown in Table 4. According to NTCP model, the TD5 and TD50 EQD2 re-RT dose of d1cc were 13.8Gy (0–20.35Gy), and 62.90Gy (42.49–80.93Gy), respectively. The TD5 and TD50 EQD2 dose of sum-dose d1.5cc were 65.26Gy(46.72–80.69Gy) and 125.25Gy(89.51–152.18Gy), respectively, see in Figure 4.

Model validation: If we consider the 2nd RT plan dose only, the AUC of the verification set was 0.9008, (0.7881–1), p< 0.001; If we consider the first and second RT plan dose individually, without considering their cumulative effect, the AUC was 0.7745(0.6199–0.9292), p=0.0012; For the accumulated EQD2 dose, the AUC was 0.8702 (0.7577–0.9828), p<0.001, as it's shown in Figure 5.

TABLE 1 Clinical characteristics of 78 patients with recurrent nasopharyngeal carcinoma.

| Clinical characteristics | | Training set No. (%) | Validation set No. (%) | P value |
|---|---|---|---|---|
| **Gender** | | | | |
| Male | | 39 (75.0) | 20 (76.9) | 0.924 |
| Female | | 13 (25.0) | 6 (23.1) | |
| Median age (y) (range) | | 45 (29–66) | 49 (30–62) | 0.725 |
| **T stage of primary tumor** | | | | |
| 1–2 | | 29 (55.8) | 12 (46.2) | 0.728 |
| 3 | | 14 (26.9) | 7 (26.9) | |
| 4 | | 9 (17.3) | 7 (26.9) | |
| **N stage of primary tumor** | | | | |
| 0–1 | | 32 (61.5) | 14 (53.8) | 0.261 |
| 2–3 | | 20 (38.5) | 12 (46.1) | |
| **RT dose of primary tumor** | | | | |
| 66 | | 23 (44.2) | 10 (38.5) | 0.493 |
| 70.4 | | 29 (55.8) | 16 (61.5) | |
| Interval between 1st and 2nd RT (range) | | 23.5 (12.1–108.3) | 33.1 (13.8–84.4) | 0.531 |
| **T stage of recurrent tumor** | | | | |
| 0–2 | | 31 (59.6) | 12 (46.1) | 0.742 |
| 3 | | 16 (30.8) | 10 (38.5) | |
| 4 | | 5 (9.6) | 4 (15.4) | |
| **N stage of recurrent tumor** | | | | |
| 0 | | 38 (73.1) | 17 (65.4) | 0.498 |
| 1 | | 13 (25.0) | 9 (34.6) | |
| 2 | | 1 (1.9) | 0 (0) | |
| **RT dose of recurrent tumor** | | | | |
| 60 | | 10 (19.2) | 6 (23.1) | 0.971 |
| 62 | | 3 (5.8) | 2 (7.7) | |
| 64 | | 3 (5.8) | 1 (3.8) | |
| 66 | | 36 (69.2) | 17 (65.4) | |
| Induction chemo | Yes | 25 (48.1) | 12 (46.2) | 0.283 |
| | No | 27 (51.9) | 14 (53.8) | |
| Concurrent chemo | Yes | 13 (25.0) | 5 (19.2) | 0.373 |
| | No | 39 (75.0) | 21 (80.8) | |
| Temporal lobe injury | Yes | 16 (32.7) | 10 (38.5) | 0.774 |
| | No | 36 (67.3) | 16 (61.5) | |

## Discussion

With the widely use of three-dimensional conformal radiotherapy, dose-volume metrics are very important to understand and evaluate the tolerance of normal tissue to dose variation. Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) (11) reported the relationship between the incidence of TLI and the volume dose of radiotherapy. The bioequivalent doses of TD5 and TD50 with TLI at a single dose< 2.5Gy were 120Gy (range: 100–140Gy) and 150Gy (range: 140–170Gy), respectively. However, it was generally believed that the tolerated dose of temporal lobe tissue was higher than the recommended reference dose of QUANTEC. There are very limited number of studies on the DVH probability of TLI after two-course radiotherapy for recurrent nasopharyngeal carcinoma.

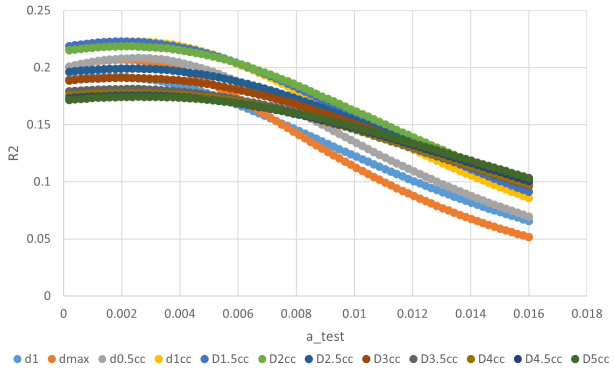



FIGURE 1
The relationship between 'a' value and R square of different indices in linear time model.

Lee et al. (12) analyzed 487 cases of nasopharyngeal carcinoma patients after re-radiotherapy, in which both primary and retreatment radiotherapy were two-dimensional radiotherapy. They found re-radiotherapy significantly increased the incidence of TLI. In the meta-analysis of re-radiotherapy of brain tissue (13), they found the two 2D radiotherapy treatments is safe when the sum of EQD2 is less than 96Gy, and the probability of TLI was 0–3% when the sum of EQD2 is less than 101Gy. Liu et al. (14) conducted dose-volumetric analysis on TLI of 227 patients with recurrent nasopharyngeal carcinoma. In these cases, the first course of radiotherapy was two-dimensional radiotherapy and the second course of radiotherapy was IMRT radiotherapy. However, in this study, the first temporal lobe dose was an estimated dose, and there is a question about the accuracy of directly summing two doses together.

Our study represents a comprehensive evaluation of the NTCP model for TLI in recurrent nasopharyngeal carcinoma. The rigorous approach includes the use of IMRT technology for both primary and recurrent treatment, dual RT plans?, and a minimum 6-month follow-up with available MRI images. Based on the NTCP model, the TD5 and TD50 of D1.5cc for the temporal lobe in our study, derived from the cumulative effect of two radiotherapy plans, was 65.3Gy and 125.3Gy, respectively. Notably, these values were more stringent compared to the EQD2-pMAX dose presented in Liu et al.'s study (14). This discrepancy may be attributed to the overlap of high-dose regions resulting from deformable image registration. TLI is often inevitable when dealing with recurrent tumors with re-irradiation, especially for recurrent tumors invading skull base or cavernous sinus. Consequently, minimizing temporal lobe exposure for the second course treatment planning is important. Consistent with prior research, our study revealed a shorter latency time for TLI post re-radiotherapy, with a median duration of 11.5 months.

The time interval between two radiotherapy treatments influences the incidence of temporal lobe injury, because with the extension of the interval time, the damage of normal tissue will recover gradually. The study by Liu et al. (14) found that the risk of TLI was significantly reduced in patients with nasopharyngeal carcinoma whose interval time was > 26 months. The study by Lee et al. (15) also found that the normal tissue tolerance of patients with an interval of more than two years had a trend of improvement compared with patients with recurrence within two years, but the difference was not statistically significant. Ang et al. (16) carried out

TABLE 2 The 'a' value corresponding to maximum R2 value for different indices in linear time model.

| | d1 | dmax | d0.5cc | d1cc | D1.5cc | D2cc | D2.5cc | D3cc | D3.5cc | D4cc | D4.5cc | D5cc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R2_max | 0.1920 | 0.2069 | 0.2084 | 0.2227 | 0.2228 | 0.2186 | 0.1993 | 0.1911 | 0.1813 | 0.1777 | 0.1755 | 0.1745 |
| max_a | 0.0018 | 0.0020 | 0.0026 | 0.0026 | 0.0022 | 0.0024 | 0.0022 | 0.0020 | 0.0024 | 0.0024 | 0.0022 | 0.0026 |

The volume-dose parameter corresponding to the maximum R2 value is D1.5cc. The value of 0.2228 in the table represent the highest Nagelkerke's R squared values obtained by testing various 'a' values and volume-dose parameters: applying different 'a' values to time models to calculate combined doses, computing dosimetric indices, and performing univariate logistic analysis.

TABLE 3 ROC curve analysis results with the 'a' value corresponding to maximum R2 value in linear time model as model parameter value.

| Variable | Area under ROC curve | β | p | Lower- Upper limit | Cutoff point | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|
| Dmax | 0.8241 | 0.04718 | <0.000 | 0.7317–0.9166 | 125.6 | 74.19 | 92.47 |
| D0.5 | 0.8144 | 0.04600 | <0.0001 | 0.7243–0.9046 | 119.6 | 67.74 | 90.32 |
| D1 | 0.8103 | 0.04559 | <0.0001 | 0.7209–0.8996 | 116.0 | 67.74 | 89.25 |
| D1.5 | 0.8068 | 0.04427 | <0.0001 | 0.7200–0.8936 | 115.3 | 61.29 | 91.4 |
| D2 | 0.8075 | 0.04392 | <0.0001 | 0.7214–0.8936 | 113.1 | 58.06 | 92.47 |
| D2.5 | 0.7971 | 0.04554 | <0.0001 | 0.7078–0.8864 | 109.5 | 58.06 | 87.1 |
| D3 | 0.7867 | 0.04646 | <0.0001 | 0.6956–0.8777 | 94.48 | 80.65 | 63.44 |
| D3.5 | 0.7829 | 0.04657 | <0.0001 | 0.6916–0.8741 | 91.74 | 77.42 | 66.67 |
| D4 | 0.7792 | 0.04674 | <0.0001 | 0.6876–0.8708 | 89.60 | 77.42 | 67.74 |
| D4.5 | 0.7784 | 0.04697 | <0.0001 | 0.6863–0.8704 | 87.27 | 77.42 | 68.82 |
| D5 | 0.7763 | 0.04743 | <0.0001 | 0.6833–0.8692 | 84.06 | 77.42 | 69.89 |

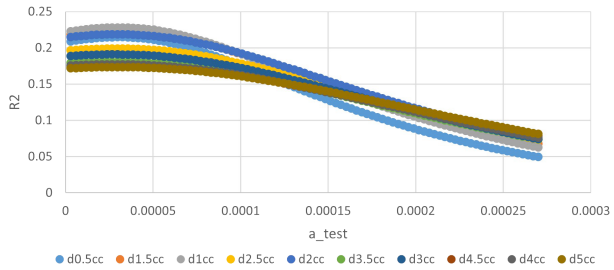



FIGURE 2
The relationship between 'a' value and R square of different indices in quadratic time model.

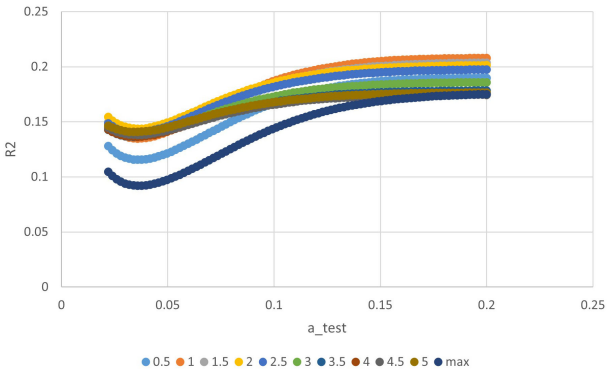



FIGURE 3
The relationship between 'a' value and R square of different indices in exponential time model.

two-course irradiation on the spinal cord of 56 macaque monkeys, and they believed that the spinal cord tolerance recovered significantly within one year after radiotherapy and gradually recovered further in the following years. In this study, we endeavored to formulate two radiotherapy dose superposition models by incorporating interval time variables in various approaches, such as linear, quadratic, and exponential time models. Our findings indicate that the equation incorporating the time factor exhibited minimal deviation from the value obtained by straightforwardly summing the temporal lobe doses voxel to voxel. Furthermore, no statistically significant difference was observed in the impact of interval time within the multifactor regression equation. Therefore, we believe that the two doses can be converted to EQD2 and then added on the corresponding voxels. The possible reason is, the minimum interval between the two radiotherapy treatments was 12 months, and the recovery of temporal lobe tissue was most obvious within one year.

MIM software was applied to map the first dose distribution to the second-course CT by registration of the two CTs. Python was applied to calculate and add EQD2 voxel by voxel. The rigid registration of these two CTs is challenging due to the relatively long interval between CT scans (median 28 months in this study). While there's uncertainty in fusing MRI with CT, the rigid nature of the intracranial temporal lobe minimizes shape changes. The study's method, utilizing MIM software for image fusion, is deemed effective in obtaining relatively

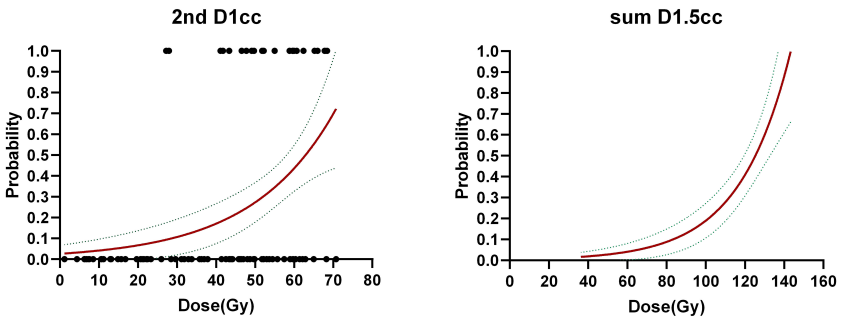



FIGURE 4
Dose-response curves for temporal lobe injury.

TABLE 4 Multivariate logistic regression analysis for temporal lobe injury.

| | P value | Wald | 95% CI | Regression function | Nagelkerke's R squared |
|---|---|---|---|---|---|
| Only second course RT indices | | | | | |
| $2^{nd}$ D1 | <0.001 | 13.841 | 1.036–1.121 | S=0.075×$2^{nd}$D1–4.649 | 0.285 |
| First and second radiation therapy indices, but without dose combination | | | | | |
| $2^{nd}$D1 | <0.001 | 12.190 | 1.032–1.117 | S=0.071×$2^{nd}$D1 + 0.084*$1^{st}$Dmax -10.281 | 0.497 |
| $1^{st}$Dmax | 0.030 | 4.688 | 1.008–1.173 | | |
| Indices of combined dose | | | | | |
| D1.5 | <0.001 | 16.405 | 1.035–1.105 | S=0.067×D1.5–8.216 | 0.330 |

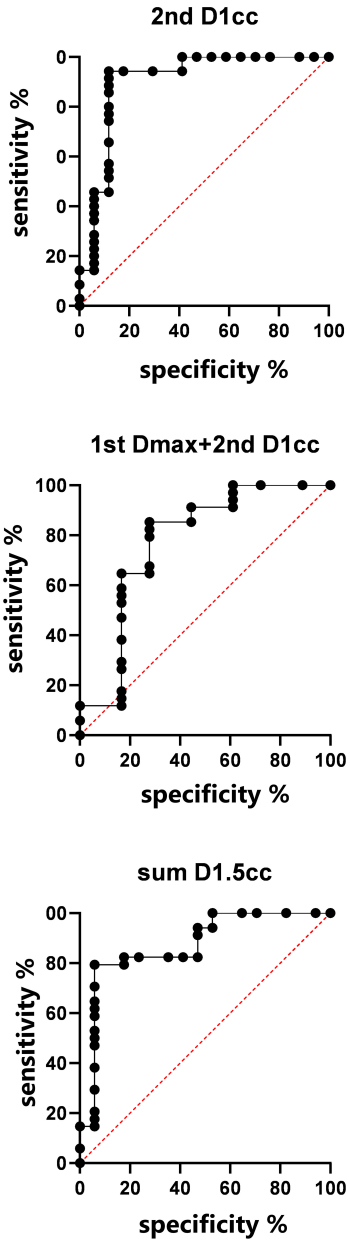



FIGURE 5
ROC curves for temporal lobe injury in validation set.

accurate results under these circumstances (17, 18). To enhance clinical applicability, we addressed challenges such as the absence of the first-course treatment plan or the inability to accumulate initial and recurrent radiotherapy. In such cases, we employed corresponding models for calculations, extending the practicality of our study to accommodate diverse patient scenarios.

Beyond the considerations of volume-dose and interval time's impact on TLI discussed above, correlation analysis was performed to investigate whether TLI was correlated with T stage, tumor volume, KPS score, gender, age, dose prescription, and administration of concurrent chemotherapy. No statistically significant difference was found in the correlation between these clinical factors and the occurrence of TLI. Su et al. (5) reported the incidence of TLI in nasopharyngeal carcinoma patients receiving chemo-radiotherapy is significantly higher than those of patients receiving radiotherapy alone. However, NPC patients received chemotherapy tend to be advanced stage, thus temporal lobe would be exposed to higher dose. In this study, patients with advanced T stage or larger tumor volume tended to have a higher proportion of TLI, but the differences were not statistically significant.

The study acknowledges several limitations, including a small sample size with both primary and secondary plans, inadequacy of follow-up duration (minimum six months in this study), a relatively low two-year overall survival rate (65%) which means that patients may passed away before developing TLI). Furthermore, developing a more comprehensive grading system for TLI grades is crucial, differentiating between mild (grades 1–2) and severe cases (grades 3–4). Achieving precision in these distinctions requires a larger sample size to ensure the accuracy of the results.

## Conclusion

In this study, an NTCP model of temporal lobe injury after re-IMRT radiotherapy for recurrent nasopharyngeal carcinoma was established. The most important factors affecting TLI was the sum-dose d1.5cc of TL. According to NTCP model, the TD5 and TD50 EQD2 dose of sum-dose d1.5cc were 65.26Gy (46.72–80.69Gy) and 125.25Gy (89.51–152.18Gy), respectively. When considering only the re-IMRT dose, the TD5 and TD50 EQD2 re-RT dose of d1cc for TLI were 13.8Gy (0–20.35Gy), and 62.90Gy (42.49–80.93Gy), respectively. Consequently, minimizing temporal lobe exposure during re-RT planning is crucial.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

## Author contributions

XG: Conceptualization, Data curation, Funding acquisition, Investigation, Software, Writing – original draft, Writing – review & editing. JP: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. JS: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. XX: Data curation, Funding acquisition, Writing – review & editing. CH: Conceptualization, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Science and Technology Development Fund of Shanghai Pudong New Area (project no. PKJ2021-Y46), and Shanghai Municipal Health Commission (project no. 20204Y0283).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1394111/full#supplementary-material

## References

1. Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal carcinoma. *Lancet*. (2016) 387:1012–24. doi: 10.1016/S0140–6736(15)00055–0

2. Chen Y-P, Chan ATC, Le Q-T, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet*. (2019) 394:64–80. doi: 10.1016/S0140–6736(19)30956–0

3. Yu KH, Leung SF, Tung SY, Zee B, Chua DTT, Sze WM, et al. Survival outcome of patients with nasopharyngeal carcinoma with first local failure: A study by the Hong Kong Nasopharyngeal Carcinoma Study Group. *Head Neck*. (2005) 27:397–405. doi: 10.1002/hed.20161

4. Lai S-Z, Li W-F, Chen L, Luo W, Chen Y-Y, Liu L-Z, et al. How does intensity-modulated radiotherapy versus conventional two-dimensional radiotherapy influence the treatment results in nasopharyngeal carcinoma patients? *Int J Radiat OncologyBiologyPhysics*. (2011) 80:661–8. doi: 10.1016/j.ijrobp.2010.03.024

5. Su S-F, Huang Y, Xiao W, Huang S-M, Han F, Xie C, et al. Clinical and dosimetric characteristics of temporal lobe injury following intensity modulated radiotherapy of nasopharyngeal carcinoma. *Radiotherapy Oncol*. (2012) 104:312–6. doi: 10.1016/j.radonc.2012.06.012

6. Zhou G-Q, Yu X-L, Chen M, Guo R, Lei Y, Sun Y, et al. Radiation-induced temporal lobe injury for nasopharyngeal carcinoma: A comparison of intensity-modulated radiotherapy and conventional two-dimensional radiotherapy. *PloS One*. (2013) 8:e67488. doi: 10.1371/journal.pone.0067488

7. Yang S-S, OuYang P-Y, Guo J-G, Cai J-J, Zhang J, Peng Q-H, et al. Dosiomics risk model for predicting radiation induced temporal lobe injury and guiding individual intensity-modulated radiation therapy. *Int J Radiat OncologyBiologyPhysics*. (2023) 115:1291–300. doi: 10.1016/j.ijrobp.2022.11.036

8. Qiu S, Lin S, Tham IWK, Pan J, Lu J, Lu JJ. Intensity-modulated radiation therapy in the salvage of locally recurrent nasopharyngeal carcinoma. *Radiat Oncol Biol*. (2012) 83:676–83. doi: 10.1016/j.ijrobp.2011.07.006

9. Perri F, Della Vittoria Scarpati G, Caponigro F, Ionna F, Longo F, Buonopane S, et al. Management of recurrent nasopharyngeal carcinoma: Current perspectives. *Onco Targets Ther*. (2019) 12:1583–91. doi: 10.2147/OTT.S188148

10. Ng WT, Soong YL, Ahn YC, Choi HCW, Corry J, Harrington KJ, et al. International recommendations on re-irradiation by intensity-modulated radiotherapy for locally recurrent nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. (2021) 110(3):682–95. doi: 10.1016/j.ijrobp.2021.01.041

11. Lee T-F, Fang F-M. Quantitative analysis of normal tissue effects in the clinic (QUANTEC) guideline validation using quality of life questionnaire datasets for parotid gland constraints to avoid causing xerostomia during head-and-neck radiotherapy. *Radiotherapy Oncol*. (2013) 106:352–8. doi: 10.1016/j.radonc.2012.11.013

12. Lee AWM, Foo W, Chappell R, Fowler JF, Sze WM, Poon YF, et al. Effect of time, dose, and fractionation on temporal lobe necrosis following radiotherapy for nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. (1998) 40:35–42. doi: 10.1016/S0360–3016(97)00580–4

13. Chen J, Dassarath M, Yin Z, Liu H, Yang K, Wu G. Radiation induced temporal lobe necrosis in patients with nasopharyngeal carcinoma: A review of new avenues in its management. *Radiat Oncol*. (2011) 6:128. doi: 10.1186/1748–717X-6–128

14. Liu S, Lu T, Zhao C, Shen J, Tian Y, Guan Y, et al. Temporal lobe injury after re-irradiation of locally recurrent nasopharyngeal carcinoma using intensity modulated radiotherapy: clinical characteristics and prognostic factors. *J Neurooncol*. (2014) 119:421–8. doi: 10.1007/s11060–014-1513–3

15. Lee AWM, Foo W, Law SCK, Peters LJ, Poon YF, Chappell R, et al. Total biological effect on late reactive tissues following reirradiation for recurrent nasopharyngeal carcinoma. *Int J Radiat OncologyBiologyPhysics*. (2000) 46:865–72. doi: 10.1016/S0360-3016(99)00512-X

16. Ang KK, Jiang GL, Feng Y, Stephens LC, Tucker SL, Price RE. Extent and kinetics of recovery of occult spinal cord injury. *Int J Radiat Oncol Biol Phys*. (2001) 50:1013–20. doi: 10.1016/S0360–3016(01)01599–1

17. Nobnop W, Chitapanarux I, Neamin H, Wanwilairat S, Lorvidhaya V, Sanghangthum T. Evaluation of deformable image registration (DIR) methods for dose accumulation in nasopharyngeal cancer patients during radiotherapy. *Radiol Oncol*. (2017) 51:438–46. doi: 10.1515/raon-2017–0033

18. Yao W, Hu J, Xu P, He M, Fang Y, Liu M, et al. Accumulated dose deviation of rotational and residual setup errors on nasopharyngeal carcinoma using MIM treated by helical tomotherapy. *Technol Cancer Res Treat*. (2023) 22:153303382311696. doi: 10.1177/15330338231169601

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 July 2024
DOI 10.3389/fonc.2024.1375096



OPEN ACCESS

EDITED BY
Xinyuan Chen,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Renjie He,
University of Texas MD Anderson Cancer Center, United States
Matthew Field,
University of New South Wales, Australia

*CORRESPONDENCE
Sheng Huang
RichardHuang2066@outlook.com

RECEIVED 23 January 2024
ACCEPTED 20 June 2024
PUBLISHED 11 July 2024

CITATION
Johnson CL, Press RH, Simone CB 2nd, Shen B, Tsai P, Hu L, Yu F, Apinorasethkul C, Ackerman C, Zhai H, Lin H and Huang S (2024) Clinical validation of commercial deep-learning based auto-segmentation models for organs at risk in the head and neck region: a single institution study. *Front. Oncol.* 14:1375096. doi: 10.3389/fonc.2024.1375096



# Clinical validation of commercial deep-learning based auto-segmentation models for organs at risk in the head and neck region: a single institution study

Casey L. Johnson[1], Robert H. Press[1], Charles B. Simone 2nd[1], Brian Shen[1], Pingfang Tsai[1], Lei Hu[1], Francis Yu[1], Chavanon Apinorasethkul[1], Christopher Ackerman[1], Huifang Zhai[1], Haibo Lin[1] and Sheng Huang[1,2]*

[1]New York Proton Center, New York, NY, United States, [2]National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute & Hospital, Tianjin, China

**Purpose:** To evaluate organ at risk (OAR) auto-segmentation in the head and neck region of computed tomography images using two different commercially available deep-learning-based auto-segmentation (DLAS) tools in a single institutional clinical applications.

**Methods:** Twenty-two OARs were manually contoured by clinicians according to published guidelines on planning computed tomography (pCT) images for 40 clinical head and neck cancer (HNC) cases. Automatic contours were generated for each patient using two deep-learning-based auto-segmentation models—Manteia AccuContour and MIM ProtégéAI. The accuracy and integrity of autocontours (ACs) were then compared to expert contours (ECs) using the Sørensen-Dice similarity coefficient (DSC) and Mean Distance (MD) metrics.

**Results:** ACs were generated for 22 OARs using AccuContour and 17 OARs using ProtégéAI with average contour generation time of 1 min/patient and 5 min/patient respectively. EC and AC agreement was highest for the mandible (DSC 0.90 $\pm$ 0.16) and (DSC 0.91 $\pm$ 0.03), and lowest for the chiasm (DSC 0.28 $\pm$ 0.14) and (DSC 0.30 $\pm$ 0.14) for AccuContour and ProtégéAI respectively. Using AccuContour, the average MD was<1mm for 10 of the 22 OARs contoured, 1-2mm for 6 OARs, and 2-3mm for 6 OARs. For ProtégéAI, the average mean distance was<1mm for 8 out of 17 OARs, 1-2mm for 6 OARs, and 2-3mm for 3 OARs.

**Conclusions:** Both DLAS programs were proven to be valuable tools to significantly reduce the time required to generate large amounts of OAR contours in the head and neck region, even though manual editing of ACs is likely needed prior to implementation into treatment planning. The DSCs and

MDs achieved were similar to those reported in other studies that evaluated various other DLAS solutions. Still, small volume structures with nonideal contrast in CT images, such as nerves, are very challenging and will require additional solutions to achieve sufficient results.

KEYWORDS

deep-learning, autosegmentation, head&neck cancer, OARs, radiotherapy

# 1 Introduction

The evolution of radiation therapy techniques in recent decades has led to major improvements in dose conformality along with precision in dose delivery. Modern methods of dose delivery such as intensity-modulated radiation therapy, volumetric arc therapy, and recently, intensity-modulated proton therapy all have proven to improve local control as well as normal tissue sparing in various tumor types (1–4).

However, to capitalize on the benefits of these treatment modalities, target volumes and surrounding organs-at-risk (OARs) must be carefully delineated on computed tomography (CT) images. This is heavily demonstrated in the case of head and neck cancer (HNC) that often lie in complex anatomical locations surrounded by numerous OARs. The delineation of neighboring structures is a time-consuming manual process, that mandates experienced knowledge of the local anatomy. Furthermore, manual delineation introduces inter-observer variability as evidenced in several recent studies (5–7). OAR delineation guidelines have been published by many authors to combat this but vary widely causing difficulty when comparing dose-volume relationships across studies (8). Consensus guidelines were established in 2015 integrating advice and expertise from radiation oncologists from across the world (9). However, even with established evidence-based guidelines, inter-observer variability still exists as shown in a study conducted by van der Veen et al. (10)in which only around half of the participating radiation oncologists utilized the standardized guidelines.

Methods of utilizing advancements in automatic segmentation techniques have emerged to potentially combat lengthy processing times and wide inter-observer variability. Atlas-based auto-segmentation (ABAS) is one such method in which an 'atlas' of OARs is established by training a software program with a dataset that has OARs already labeled. An explanation of this process has been published by Han et al. (11). This technique has been proven to reduce processing time as well as generating appropriate sedimentation for various OARs (12, 13). Another automated segmentation method currently being investigated is deep-learning-based auto-segmentation (DLAS). DLAS utilizes machine learning to incorporate vast datasets and generate an automated solution. This technique has shown promise in recent studies assessing the efficiency in the head and neck region (14, 15).

There are several commercially available ABAS software programs as well as in-house developed DLAS programs that have been validated on HNC. A comparison study conducted by La Macchia et al. (16) compared three auto-segmentation programs and reported significant reductions in time to generate quality contours when compared to manual processes. However, many studies have shown the contours generated by either ABAS or DLAS methods still require additional manual editing to be clinically acceptable (16–18). In a study conducted by van Dijk et al. (19), time to generate clinically acceptable contours still proved to be significantly less when created using either ABAS or DLAS. Even still, the authors noted that the DLAS method used outperformed ABAS when evaluating on a cohort of HNC patients. However, large cohorts of training data sets are required to train and get an accurate DLAS model. It is not feasible for each center to develop and train their own DLAS model, thus commercially available models or shared DLAS packages would be advantageous. Evidence from clinical validation of DLAS packages will allow for centers to identify solutions that will provide optimal performance for their particular needs. Thus, this study aims to clinically validate the generic models for HNC OAR autocontouring of two commercially available deep-learning-based auto-segmentation software packages, AccuContour (version 3.1, Manteia Medical Technologies, Wisconsin, MI) and ProtègèAI (version 1.0, MIM Software Inc., Cleveland, OH), Both MIM and Manteia's solutions include a generic HNC autocontouring model. Quantitative evaluation on a set of 40 clinical HNC patients will be performed.

# 2 Materials and methods

## 2.1 Expert contour creation

To validate each DLAS model, a cohort of 40 HNC patients were selected who previously received treatment at our institution. All patient data in this retrospective study was approved under an internal review board. All planning CT (pCT) images consisted of 512 pixels × 512 pixels in each slice with voxel size of 0.98mm×0.98mm×1.50mm. All the CT data were acquired on the same version of CT scanner (Somatom Definition AS, Siemens, Forchheim, Germany) without contrast enhancement. In total, 22

OARs were manually contoured to establish the expert contours (ECs) on the pCT images by expert radiation oncologists at our institution according to international consensus guidelines. These OARs were divided into five groups: 1) *Glandular*: submandibular and parotid glands; 2) *Aerodigestive Tract*: oral cavity, larynx, esophagus, constrictor muscle; 3) *Ocular and Aural*: cochlea, lens, eye; 4) *Neural*: brainstem, chiasm, spinal cord, optic nerves, temporal lobes; 5) *Other*: mandible.

## 2.2 DLAS contour creation

In contrast to atlas model-based auto-segmentation which utilizes a trained model of shape and appearance characteristics of anatomy structures and then project onto a new image set through deformable imaging registration, DLAS uses deep neural network architectures with multiple (2 or more) hidden layers to learn features from a dataset by modeling complex nonlinear relationships. These architectures are usually formed by stacking several different-type layers that transform input images to the desired output. The transformation through convolution filters, or kernels, reveals local connectivity between neurons of adjacent layers exploiting spatially local correlation. This permits the networks to learn features both globally and locally allowing the network to detect subtle variations in the input data, which here mean the features of different OARs. The training processes generally utilize supervised learning by back-propagation algorithms, which optimize the node weights to minimize the loss between the predicted and known output through each training iteration to a satisfactory level of accuracy. It remains to be unanswered how many patient scans are optimal to produce clinically acceptable results. One would agree that a robust dataset that includes a large variability of patient anatomies would achieve reasonable and robust model. During past years, more and more commercial DLAS software have emerged and become clinically available. The following two different DLAS packages were implemented in our clinic and assessed in this study:

Manteia AccuContour is a commercial deep-learning-based auto-segmentation software using deep convolutional neural network models based on a U-Net architecture, the design of which follows the work of Ronneberg et al. (20). The training data included in the model consists of 100 HNC image sets acquired from GE, Philips, and Siemens CT scanners. The HNC model was then applied to the same 40 pCTs used in the expert contour creation. As with the expert contour creation, a unique set of autocontours (ACs) was generated for each patient case.

The MIM ProtégéAI generic HNC model is a cloud-based deep learning segmentation model with a similar structure to U-Net. The training data included in the model consists of about 400 HNC images gathered from 31 institutions mainly across the US, but with a few additional institutions located in Europe, Hong Kong, and Australia. MIM's HNC model was also applied to the 40 pCTs used in the expert contour creation. No post-processing was completed after the 3D volume generation for each contour. Again, a unique set of ACs was generated for each patient case. The MIM ProtégéAI autocontour model did not provide contours for temporal lobes, cochlea, or the constrictor muscle. Thus, only 17 OAR contours were generated for each patient.

## 2.3 Evaluation metrics

For performance evaluation, the Sørensen-Dice similarity coefficient (DSC) (21), and the mean distance (MD) between the ECs and ACs for the OARs of each patient were calculated as a comparison metric. The DSC is defined as:

$$DSC(A,B) = \frac{2|A \cap B|}{||A|+|B||} \quad (1)$$

Which describes the overlapping volume between two structures $A$ and $B$. A value of 0 indicates no overlap; a value of 1 indicates complete overlap. The MD is a bi-directional measure of the distance between the surface of two contours and is defined as:

$$MD(A,\ B) = \frac{1}{N(A)} \sum_{i=1}^{N(A)} \min d(a_i, b \in S_B) \quad (2)$$

Where $N(A)$ is the total nodes on the surface of $A$ structure, min $d(a, b \in S_B)$ is the minimum Euclidean distance of node $a_i$ to any point b on surface of $B$. The smaller mean distance indicates the surfaces of $A$ and $B$ are closer to each other. For every clinical case, each OAR delineated on the ECs and ACs was compared, generating a DSC and MD value. Results are presented as averages across the patient cohort for each OAR with ranges describing the performance of grouped OARs and mean ± standard deviation describing the performance of specific OARs.

# 3 Results

## 3.1 Manteia AccuContour evaluation

The time to autocontour all 22 OARs using Manteia AccuContour was 1min/patient. Table 1 lists the average DSC and MD values between ECs and Manteia AccuContour ACs for each OAR.

- *Glandular* OARs: the AccuContour model showed high DSCs and contour agreement (0.73-0.78). The MDs were similarly acceptable for both left and right parotid glands (1.75 ± 1.13, 1.88 ± 1.15mm, respectively). The left submandibular gland MDs were slightly larger than the right submandibular gland (2.05 ± 5.67, 1.72 ± 3.63mm, respectively).
- *Aerodigestive Tract* OARs: the AccuContour model generated the best DSCs for the oral cavity (see Figure 1B) and esophagus (0.80 ± 0.09, 0.75 ± 0.11, respectively), while generating lower DSCs for the constrictor muscle and larynx (0.55 ± 0.08, 0.42 ± 0.10, respectively). The MDs for these OARs were relatively large (1.79-2.65mm).
- *Ocular and Aural* OARs: AccuContour ACs resulted in high DSCs for both eyes (0.83 ± 0.05) (see Figure 1A), and moderate DSCs for the left and right lens and cochlea (0.68-0.70). The MDs for all these OARs were low (0.16-0.70mm).

10.3389/fonc.2024.1375096

TABLE 1 DSCs and MDs for multiple-subject Manteia AccuContour ACs vs. ECs.

| Variable* | OAR | | | | | |
|---|---|---|---|---|---|---|
| | Brainstem | Chiasm | Cochlea L | Cochlea R | Constrictor Muscle | Esophagus |
| **DSC** | 0.84 ± 0.05 | 0.28 ± 0.14 | 0.68 ± 0.13 | 0.68 ± 0.10 | 0.55 ± 0.08 | 0.75 ± 0.11 |
| **MD (mm)** | 0.88 ± 0.32 | 2.20 ± 1.28 | 0.24 ± 0.14 | 0.26 ± 0.12 | 1.98 ± 2.81 | 1.79 ± 6.74 |
| | Eye L | Eye R | Larynx | Lens L | Lens R | Mandible |
| **DSC** | 0.83 ± 0.05 | 0.83 ± 0.05 | 0.42 ± 0.10 | 0.68 ± 0.11 | 0.70 ± 0.10 | 0.90 ± 0.07 |
| **MD (mm)** | 0.70 ± 0.26 | 0.67 ± 0.27 | 2.65 ± 0.79 | 0.19 ± 0.15 | 0.16 ± 0.11 | 0.43 ± 0.15 |
| | Optic Nerve L | Optic Nerve R | Oral Cavity | Parotid L | Parotid R | Spinal Cord |
| **DSC** | 0.64 ± 0.08 | 0.62 ± 0.10 | 0.80 ± 0.09 | 0.78 ± 0.10 | 0.77 ± 0.11 | 0.78 ± 0.06 |
| **MD (mm)** | 0.47 ± 0.30 | 0.96 ± 0.61 | 2.31 ± 0.91 | 1.75 ± 1.13 | 1.88 ± 1.15 | 1.80 ± 5.43 |
| | Submandibular L | Submandibular R | Temporal Lobe L | Temporal Lobe R | | |
| **DSC** | 0.73 ± 0.19 | 0.73 ± 0.17 | 0.78 ± 0.09 | 0.78 ± 0.09 | | |
| **MD (mm)** | 2.05 ± 5.67 | 1.72 ± 3.63 | 2.78 ± 1.95 | 2.80 ± 2.02 | | |

*Mean ± Standard Deviation.

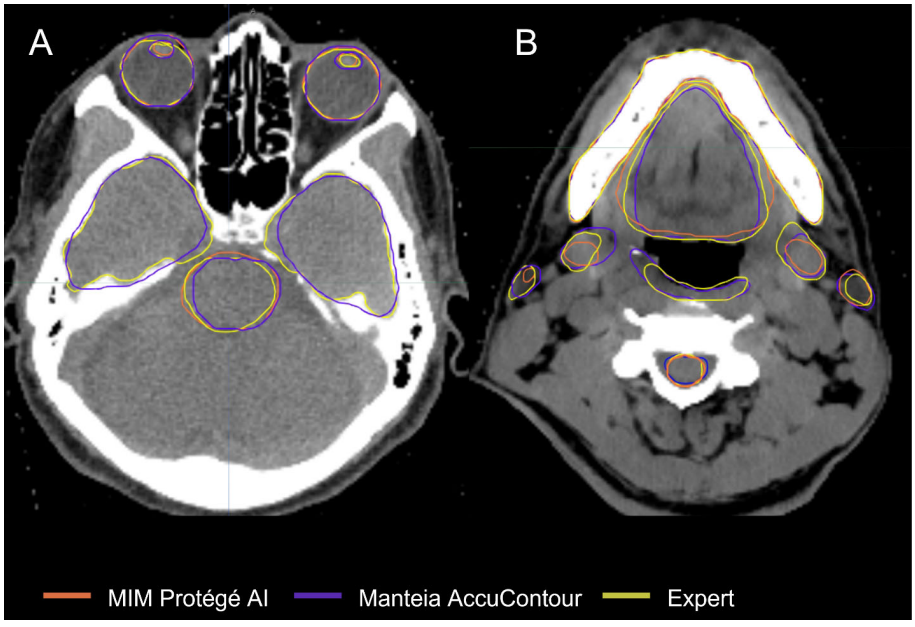



FIGURE 1
Comparison of ACs and ECs for, **(A)** Brainstem, Temporal Lobes, Eyes, Lenses; **(B)** Spinal Cord, Constrictor Muscle, Submandibular Glands, Parotid Glands, Oral Cavity, Mandible.

- *Neural* OARs: the AccuContour model showed high DSCs for the brainstem and spinal cord (0.84 ± 0.05, 0.78 ± 0.06, respectively). Both the right and left optic nerve were lower (0.62 ± 0.10, 0.64 ± 0.08, respectively), and the chiasm performed the lowest (0.28 ± 0.14) (see Figure 1A). The resulting MDs were low for the brainstem and right and left optic nerve (0.47-0.88mm) but were higher for the spinal cord and chiasm (1.80-2.20mm).
- *Mandible* OAR: ACs generated by AccuContour resulted in high DSCs (0.90 ± 0.07) and a similarly acceptable MD (0.43 ± 0.15mm).

## 3.2 MIM ProtègèAI evaluation

The average time to autocontour all 17 OARs using MIM ProtègèAI was 5min/patient. Table 2 lists the average DSC and MD values between ECs and MIM ProtègèAI ACs for each OAR.

- *Glandular* OARs: the ProtégéAI model showed high DSCs for both left and right parotid glands (0.77 ± 0.13, 0.79 ± 0.09, respectively) (Figure 1B). The MDs were similarly

TABLE 2 DSCs and MDs for multiple-subject MIM ProtégéAI ACs vs. ECs.

| Variable* | OAR | | | | | |
|---|---|---|---|---|---|---|
| | Brainstem | Chiasm | Eye L | Eye R | Larynx | Lens L |
| **DSC** | 0.81 ± 0.04 | 0.30 ± 0.14 | 0.89 ± 0.03 | 0.89 ± 0.02 | 0.65 ± 0.09 | 0.55 ± 0.27 |
| **MDs (mm)** | 1.04 ± 0.29 | 1.33 ± 0.95 | 0.47 ± 0.20 | 0.44 ± 0.19 | 2.51 ± 0.97 | 0.38 ± 0.17 |
| | Lens R | Mandible | Optic Nerve L | Optic Nerve R | Oral Cavity | Parotid L |
| **DSC** | 0.54 ± 0.29 | 0.91 ± 0.03 | 0.53 ± 0.09 | 0.63 ± 0.13 | 0.82 ± 0.06 | 0.77 ± 0.13 |
| **MDs (mm)** | 0.43 ± 0.40 | 0.62 ± 0.32 | 0.54 ± 0.34 | 0.40 ± 0.22 | 2.04 ± 0.71 | 1.31 ± 0.74 |
| | Parotid R | Spinal Cord | Submandibular L | Submandibular R | Esophagus | |
| **DSC** | 0.79 ± 0.09 | 0.79 ± 0.05 | 0.68 ± 0.23 | 0.70 ± 0.28 | 0.75 ± 0.11 | |
| **MDs (mm)** | 1.44 ± 1.57 | 1.49 ± 4.55 | 2.41 ± 5.42 | 1.91 ± 3.36 | 0.76 ± 0.50 | |

*Mean ± Standard Deviation.

acceptable for both the left and right parotid glands (1.31 ± 0.74, 1.44 ± 1.57mm, respectively).

- o *Aerodigestive Tract* OARs: the ProtégéAI model produced the best DSCs for the oral cavity (Figure 1B) and esophagus (0.82 ± 0.06, 0.75 ± 0.11, respectively), but resulted in lower DSCs for the larynx (0.65 ± 0.09). The MDs for the esophagus were relatively low (0.76 ± 0.50mm) but were higher for the oral cavity and larynx (2.04 ± 0.71, 2.51 ± 0.97mm respectively).
- o *Ocular and Aural* OARs: ProtégéAI ACs resulted in high DSCs for both right and left eyes (0.89 ± 0.02, 0.89 ± 0.03, respectively) (Figure 1A), but performed less effectively for the left and right lenses (0.55 ± 0.27, 0.54 ± 0.29, respectively). However, the MDs for these OARs were low (0.43-0.62mm).
- o *Neural* OARs: the ProtégéAI model showed high DSCs for the brainstem and spinal cord (0.81 ± 0.04, 0.79 ± 0.05, respectively). Like AccuContour, the right and left optic nerve DSCs were lower (0.63 ± 0.13, 0.53 ± 0.09, respectively), and again, the chiasm performed the lowest (0.30 ± 0.14) (Figure 1A). The resulting MDs were low for the left and right optic nerves (0.54 ± 0.34, 0.40 ± 0.22mm, respectively), but were slightly higher for the brainstem, chiasm, and spinal cord (1.04-1.49mm).
- o *Mandible* OAR: ACs generated by ProtégéAI resulted in high DSCs (0.91 ± 0.03) and similarly acceptable MDs (0.62 ± 0.32mm).

# 4 Discussion

Both ABAS and DLAS methods have shown promise in reducing variability and time required to establish contours (22–25). This study aimed to clinically validate two DLAS commercial software programs by comparing automatically-generated OAR contours with those created manually for a cohort of 40 HNC patients. DLAS contours were evaluated with two gold-standard geometric measures. Reasonable agreement was shown for the glandular OARs, eyes, brainstem, spinal cord, oral cavity, esophagus, and mandible across both autocontour programs. Moderate agreement was shown for the optic nerves and lenses, constrictor muscle, and larynx (ProtégéAI specifically). There was poor agreement for the larynx (AccuContour specifically) and chiasm. The results demonstrate that each DLAS package can adequately contour most HNC OARs efficiently in an independent cohort of patients.

We were able to provide results that closely resemble those reported in other DLAS studies in the head and neck region. There is a consensus on the efficiency of DLAS programs to contour the mandible and brainstem as reported by Brunenberg et al. (26) (DSC 0.95 and 0.87, respectively) validating another commercially-available DLAS program, DLCExpert™. The results for the glandular OARs and optic nerves in this study resemble those reported by Ibragimov and Xing (15) (DSC parotid gland 0.78, submandibular gland 0.73, optic nerve 0.65) using a convolutional neural network (CNN) approach. Willems et al. (27) reported pharyngeal constrictor muscle results (inferior, middle, and superior constrictor muscle average DSC 0.55) that closely resemble the overall constrictor muscle ACs generated in this study. Overall, structures with larger volumes appeared to be easier for the DLAS models to contour as given by the larger DSCs.

Figure 2 shows that all the ACs generated by MIM ProtégéAI and Manteia AccuContour were comparable except for the larynx. The average DSC for the larynx contoured by AccuContour was noticeably lower than that of the larynx contoured using ProtégéAI. When investigated further, it was noted that the larynx AC within AccuContour was consistently omitting the airspace within the larynx when generating a contour (Figure 3). Our institution's standard of practice for contouring the larynx is to include the entire structure as well as the airspace within. This apparent discrepancy led to less agreement between the EC and AC for the larynx using AccuContour, while ProtégéAI was able to contour in a similar fashion to our experts. Fortunately, Mantiea's AccuLearning software allows for the creation of in-house models. Should we want to establish a DLAS model using our institution's method of contouring the larynx, we would be able to do so in the future.

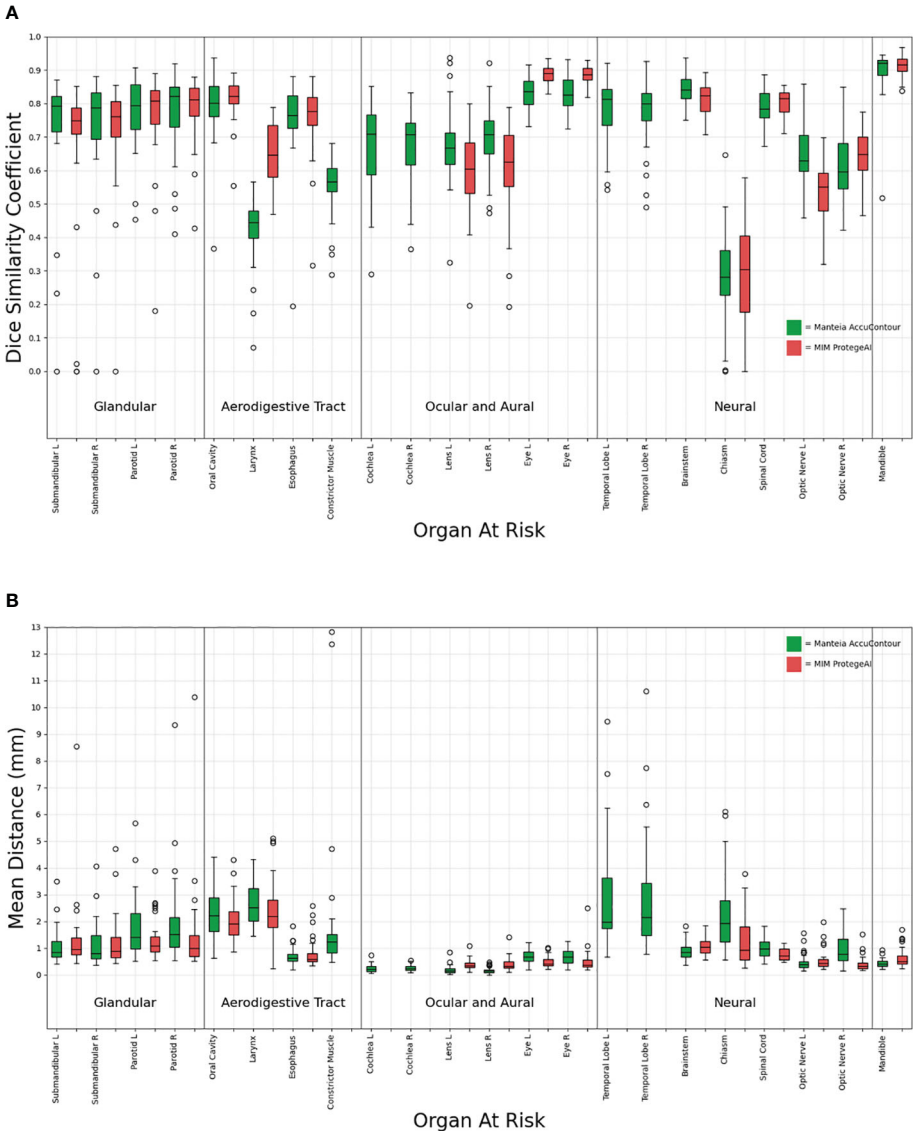



FIGURE 2
Comparison of Manteia AccuContour and MIM ProtègèAI for various OARs Sørensen-Dice similarity coefficients **(A)**, and Mean Distances **(B)**.

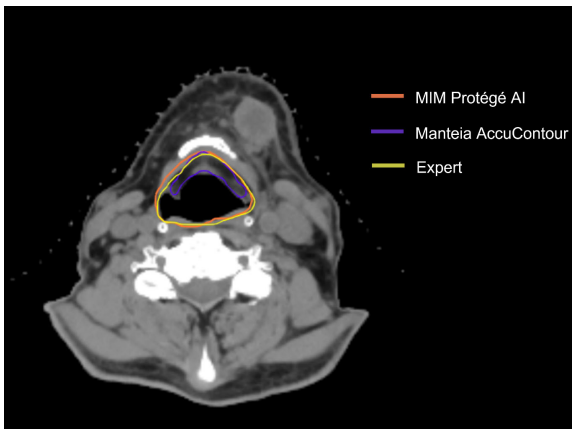



FIGURE 3
Comparison of AC and EC Larynx contour.

The OAR contours from each DLAS model which suffered the most were the chiasm and optic nerves (Figure 4). The low average DSCs of these structures do not appear unique to our study, as several other papers have reported low values for the chiasm and optic nerves (28–30). This can likely be attributed to small volumes of these structures as well as low contrast to the surrounding brain tissue in CT, making it difficult to accurately segment the structures as noted by a study from Ren et al (31). In this study, investigators were able to improve small structure DSCs (including the chiasm and optic nerves) using a specialized 3D CNN approach. In a 2018 study aiming to improve segmentation for small volume structures in the head and neck region, Tong et al. (28) were able to rescue the low DSC of the chiasm and optic nerves to an extent by employing a Shape Representation Model (SRM) to a Fully Convolutional Neural Network (FCNN), improving the average DSC for these small structures. Thus, further effort to adjust autocontouring

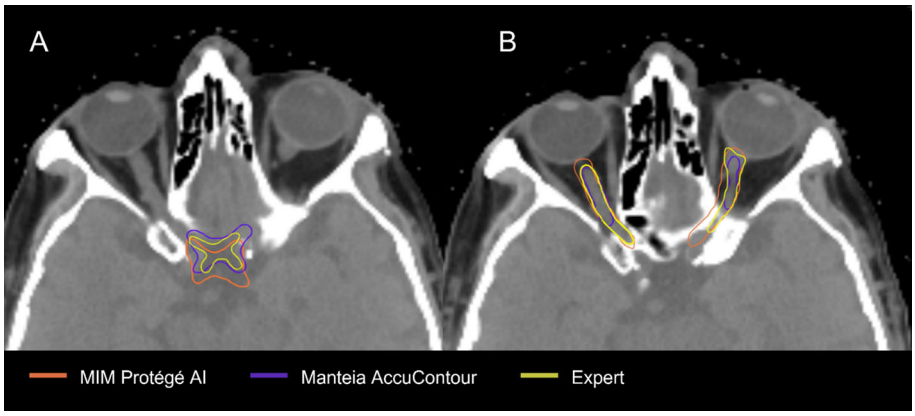



FIGURE 4
Comparison of EC and AC of Chiasm **(A)** and Optic Nerves **(B)** EC.

solutions to accurately contour structures with small volumens appears advantageous.

Our study does have potential limitations. First, the generic HNC model within MIM ProtégéAI does not currently contain the temporal lobes, constrictor muscle, or cochlea. However, MIM's research & development team has confirmed that these structures will be included in future updates and can be clinically validated at that time. We also only validated each model on 40 pCTs from one institution, thus limiting the power and generalizability of our results. Lastly, additional manual editing of ACs after creation was not performed, and only the first iteration of ACs generated was compared.

Thus, one consideration for future use is the practice of assessing ACs with additional manual editing where needed. This practice has proved useful in improving contour acceptability in previous studies. Willems et al. (27) recorded the time required to correct ACs as well as resulting DSCs and found that only an additional 15 minutes total were required to improve the ACs. This was also confirmed in a study by Teguh et al. (22), where the ACs were edited, improving similarity metrics while still requiring less time than manually creating contours from scratch. We did not routinely edit the contours generated by ProtégéAI or AccuContour, but future work could include this practice to assess time sparing and contour improvements.

While the auto-segmentation similarity metrics investigated in this study are commonly used, future research should incorporate advanced comparison metrics such as the surface DSC, as described by Vaassen et al. (32). A comprehensive overview of additional metrics is provided by Kiser et al. (33), who corroborate previous findings that surface DSC may better correlate with clinical applicability. Additionally, this study focused on the performance of commercially available auto-segmentation solutions; however, there is a growing trend of research teams developing open-source solutions, available on platforms like GitHub (34, 35). These open-source solutions often yield competitive results, rivalling commercial software and may be of particularly beneficial for research studies with limited funding or in clinics with constrained resources. Beyond the two commercial artificial intelligence based auto-segmentation software evaluated in this study, other options such Mirada, MVsision, Radformation, Raystation and TheraPanacea also provide contours of comparable quality for the organs at risk (OARs) in the head and neck region (36). The structures covered by each system are adequate for clinical application and can be customized for specific anatomical sites. The reported accuracy metrics of each system should be considered as one of the critical factors in the decision-making process. Additionally, institutions should evaluate the cost, service quality, and integration capability with existing clinical workflows when selecting an auto-segmentation solution.

## 5 Conclusion

Both commercially available DLAS programs were able to significantly reduce the time required to generate OAR contours, even though manual editing of ACs is likely needed prior to implementation into the clinic. The DSCs and MDs achieved were similar to those reported in other studies that evaluated various other DLAS solutions. Still, structures with small volumes are difficult to generate accurate ACs for and will require additional solutions to achieve sufficient contours.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## Author contributions

CJ: Data curation, Formal analysis, Writing – original draft. RP: Data curation, Validation, Visualization, Writing – review & editing. CS: Supervision, Validation, Visualization, Writing – review & editing. BS: Data curation, Formal analysis, Visualization, Writing – review & editing. PT: Formal analysis, Validation, Writing – review & editing. LH: Data curation, Formal analysis, Validation, Writing – review &

editing. FY: Data curation, Formal analysis, Writing – review & editing. CAp: Data curation, Formal analysis, Writing – review & editing. CAc: Data curation, Formal analysis, Writing – review & editing. HZ: Data curation, Formal analysis, Writing – review & editing. HL: Supervision, Writing – original draft, Writing – review & editing. SH: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Tianjin Municipal Education Commission Research Development Program 2022KJ222.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Clark CH, Bidmead Am, Mubata CD, Harrington KJ, Nutting CM. Intensity-modulated radiotherapy improves target coverage, spinal cord sparing and allows dose escalation in patients with locally advanced cancer of the larynx. *Radiotherapy Oncol.* (2004). 0167-8140 (Print). doi: 10.1016/j.radonc.2003.10.012

2. Chao C-J, Kato N, Scott CG, Lopez-Jimenez F, Lin G, Kane GC, et al. Unsupervised machine learning for assessment of left ventricular diastolic function and risk stratification. *J Am Soc Echocardiography*. (2022) 35:1214–25.e8. doi: 10.1016/j.echo.2022.06.013

3. Quan EM, Li X, Li Y, Wang X, Kudchadker RJ, Johnson JL, et al. A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment. *Int J Radiat Oncology Biology Phys.* (2012). 1879-355X (Electronic).

4. Tran A, Zhang J, Woods K, Yu V, Nguyen D, Gustafson G, et al. Treatment planning comparison of IMPT, VMAT and 4π radiotherapy for prostate cases. *Radiat Oncol.* (2017) 12:10–0. doi: 10.1186/s13014-016-0761-0

5. Brouwer CL, Steenbakkers Rj, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol.* (2012). 1748-717X (Electronic). doi: 10.1186/1748-717X-7-32

6. Nelms BE, Tomé Wa, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *Int J Radiat Oncology Biology Phys.* (2012). 1879-355X (Electronic). doi: 10.1016/j.ijrobp.2010.10.019

7. Petkar I, McQuaid D, Dunlop A, Tyler J, Hall E, Nutting C. Inter-observer variation in delineating the pharyngeal constrictor muscle as organ at risk in radiotherapy for head and neck cancer. *Front Oncol.* (2021). 2234-943X (Print). doi: 10.3389/fonc.2021.644767

8. Brouwer CL, Steenbakkers RJ, Gort E, Kamphuis ME, van der Laan HP, Van't Veld AA, et al. Differences in delineation guidelines for head and neck cancer result in inconsistent reported dose and corresponding NTCP. *Radiotherapy Oncol.* (2014). 1879-0887 (Electronic). doi: 10.1016/j.radonc.2014.01.019

9. Brouwer CL, Steenbakkers RJ, Bourhis J, Budach W, Grau C, Grégoire V, 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. *Radiotherapy Oncol.* (2015). 1879-0887 (Electronic). doi: 10.1016/j.radonc.2015.07.041

10. van der Veen JA-O, Gulyban A, Willems S, Maes F, Nuyts S. Interobserver variability in organ at risk delineation in head and neck cancer. *Radiat Oncol.* (2021). 1748-717X (Electronic).

11. Han X, Hoogeman Ms, Levendag PC, Hibbard LS, Teguh DN, Voet P, et al. Atlas-based auto-segmentation of head and neck CT images. *Med Image Computing Computer-Assisted Intervention*. (2008).

12. Kim N, Chang JS, Kim YB, Kim JS. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. *Radiat Oncol.* (2020) 15:106. doi: 10.1186/s13014-020-01562-y

13. Commowick O, Grégoire V, Malandain G. Atlas-based delineation of lymph node levels in head and neck computed tomography images. *Radiotherapy Oncol.* (2008). 0167-8140 (Print). doi: 10.1016/j.radonc.2008.01.018

14. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiotherapy Oncol.* (2020). 1879-0887 (Electronic). doi: 10.1016/j.radonc.2019.10.019

15. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys.* (2017). 2473-4209 (Electronic). doi: 10.1002/mp.12045

16. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol.* (2012). 1748-717X (Electronic). doi: 10.1186/1748-717X-7-160

17. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiotherapy Oncol.* (2018) 126:312–17. doi: 10.1016/j.radonc.2017.11.012

18. Daisne JF, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiat Oncol.* (2013). 1748-717X (Electronic). doi: 10.1186/1748-717X-8-154

19. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, J H M Steenbakkers R, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. *Radiotherapy Oncol.* (2020). 1879-0887 (Electronic). doi: 10.1016/j.radonc.2019.09.022

20. Navab N, Hornegger J, Wells WM, et al eds. *U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* Vol. 2015. Cham: Springer International Publishing (2015).

21. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. (1945) 26:297–302. doi: 10.2307/1932409

22. Teguh DN, Levendag Pc, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncology Biology Phys*. (2011). 1879-355X (Electronic). doi: 10.1016/j.ijrobp.2010.07.009

23. Hwee J, Louie Av, Gaede S, Bauman G, D'Souza D, Sexton T, et al. Technology assessment of automated atlas based segmentation in prostate bed contouring. *Radiat Oncol.* (2011). 1748-717X (Electronic). doi: 10.1186/1748-717X-6-110

24. Nikolov SA-O, Blackwell SA-O, Zverovitch AA-O, Mendes RA-O, Livne MA-O, De Fauw JA-O, et al. Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. *J Med Internet Res.* (2021). 1438-8871 (Electronic). doi: 10.2196/26151

25. Liang J, Wu Q, Yan D. The role of seminal vesicle motion in target margin assessment for online image-guided radiotherapy for prostate cancer. *Int J Radiat Oncology Biology Phys.* (2009) 73:935–43. doi: 10.1016/j.ijrobp.2008.10.019

26. Brunenberg EJL, Steinseifer IK, van den Bosch S, Kaanders J, Brouwer CL, Gooding MJ, et al. External validation of deep learning-based contouring of head and neck organs at risk. *Phys Imaging Radiat Oncol.* (2020). 2405-6316 (Electronic). doi: 10.1016/j.phro.2020.06.006

27. Willems S, Crijns W, La Greca Saint-Esteven A, Van Der Veen J, Robben D, Depuydt T, et al eds. *Clinical Implementation of DeepVoxNet for Auto-Delineation of Organs at Risk in Head and Neck Cancer Patients in Radiotherapy. OR 20 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis* Vol. 2018. Cham: Springer International Publishing (2018).

28. 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). 2473-4209 (Electronic). doi: 10.1002/mp.13147

29. Wang Y, Zhao L, Wang M, Song Z. Organ at risk segmentation in head and neck CT images using a two-stage segmentation framework based on 3D U-net. *IEEE Access*. (2019) 7:144591–602. doi: 10.1109/Access.6287639

30. Tappeiner E, Pröll S, Hönig M, Raudaschl PF, Zaffino P, Spadea MF, et al. Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach. *Int J Comput Assisted Radiol Surg*. (2019). 1861-6429 (Electronic). doi: 10.1007/s11548-019-01922-4

31. Ren X, Xiang L, Nie D, Shao Y, Zhang H, Shen D, et al. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. *Med Phys*. (2018). 2473-4209 (Electronic). doi: 10.1002/mp.12837

32. Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. *Phys Imaging Radiat Oncol*. (2020) 13:1–6. doi: 10.1016/j.phro.2019.12.001

33. Kiser KJ, Barman A, Stieb S, Fuller CD, Giancardo L. Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. *J Digital Imaging*. (2021) 34:541–53. doi: 10.1007/s10278-021-00460-3

34. Weissmann T, Huang Y, Fischer S, Roesch J, Mansoorian S, Ayala Gaona H, et al. Deep learning for automatic head and neck lymph node level delineation provides expert-level accuracy. *Front Oncol*. (2023). 2234-943X (Print). doi: 10.3389/fonc.2023.1115258

35. Asbach JC, Singh AK, Matott LS, Le AH. Deep learning tools for the cancer clinic: an open-source framework with head and neck contour validation. *Radiat Oncol*. (2022) 17:28. doi: 10.1186/s13014-022-01982-y

36. Doolan PJ, Charalambous S, Roussakis Y, Leczynski A, Peratikou M, Benjamin M, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front Oncol*. (2023) 13:1213068. doi: 10.3389/fonc.2023.1213068

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 06 December 2024
DOI 10.3389/fonc.2024.1368187



OPEN ACCESS

EDITED BY
Timothy James Kinsella,
Brown University, United States

REVIEWED BY
Barbara Vischioni,
National Center of Oncological Hadrontherapy, Italy
Xinmao Song,
Fudan University, China

*CORRESPONDENCE
Jianwen Li
Ljwwf112@163.com

RECEIVED 10 January 2024
ACCEPTED 21 October 2024
PUBLISHED 06 December 2024

CITATION
Geng X, Chang X, Wang X, Li S, Han G, Song Z, Hao F and Li J (2024) Consolidation immunotherapy following concurrent chemoradiotherapy in a patient with sinonasal NUT carcinoma: a case report.
*Front. Oncol.* 14:1368187.
doi: 10.3389/fonc.2024.1368187



# Consolidation immunotherapy following concurrent chemoradiotherapy in a patient with sinonasal NUT carcinoma: a case report

Xiaotao Geng[1], Xiaolong Chang[1], Xiaoli Wang[1], Shunjia Li[1], Guiyan Han[2], Zhiyu Song[3], Furong Hao[1] and Jianwen Li[1]*

[1]Department of Radiation Oncology, Weifang People's Hospital, Weifang, China, [2]Department of Pathology, Weifang People's Hospital, Weifang, China, [3]Department of Otolaryngology, Weifang People's Hospital, Weifang, China

**Background:** Nuclear protein in testis (NUT) cancers, also known as midline cancers, tends to occur in organs near the midline, such as the nasal sinuses and mediastinum. NUT carcinoma is very rare and has a poor prognosis.

**Case description:** We report the case of a 44-year-old female patient with sinonasal NUT carcinoma who presented with a soft tissue mass in the left frontal sinus, ethmoid sinus, and left nasal cavity on computed tomography; the tumor was poorly demarcated from the left rectus medialis. After discussion with a multidisciplinary team with expertise on head and neck tumors, the patient was considered inoperable, and definitive concurrent chemoradiotherapy (CCRT) was recommended. The patient underwent CCRT followed by three cycles of consolidation chemotherapy with albumin-bound paclitaxel and nedaplatin. Subsequently, the patient underwent 16 cycles of consolidation therapy with the programmed death–1 (PD-1) inhibitor tislelizumab. The immune-related adverse events included grade 2 hypothyroidism. After CCRT, consolidation chemotherapy, and consolidation immunotherapy, the patient achieved a favorable outcome. The patient survived for 31 months, and there were no signs of recurrence or metastasis during follow-up.

**Conclusion:** At present, there is no clear consensus on the consolidation treatment plan after CCRT for sinonasal NUT cancer. We used consolidation immunotherapy for the first time and achieved good efficacy, providing an innovative and promising treatment plan for refractory sinonasal NUT cancer.

KEYWORDS

**NUT, sinonasal, concurrent chemoradiotherapy, immunotherapy, PD-1 inhibitor**

# 1 Introduction

Nuclear protein in testis (NUT) carcinoma is extremely rare and has mostly been reported in case studies; the main primary tumor sites are the chest and head and neck (1, 2). The main molecular feature is a rearrangement of the testicular nucleoprotein gene (*NUTM1*). Although *NUTM1* can fuse to numerous different partner genes, it most frequently forms a *BRD4-NUTM1* fusion oncogene related to NUT carcinoma (2). The prognosis is very poor, with a median survival of less than 1 year (3). The treatment options for NUT cancer include surgery, radiation therapy, and chemotherapy. Surgery is critical for the treatment of NUT cancer, and surgery combined with postoperative chemoradiotherapy or radiotherapy is associated with improved survival (4). For patients with inoperable tumors or those who refuse surgery, radical concurrent chemoradiotherapy (CCRT) is an alternative. Recent advances in treatment options for head and neck NUT cancer include induction chemotherapy, proton radiotherapy, and immunotherapy. The SINTART 1 study showed that patients with tumor shrinkage greater than or equal to 80% after induction chemotherapy for surgically resectable sinonasal tumors were given the option of radiotherapy and exemption from surgical treatment (5). The SINTART 2 study showed that the addition of induction chemotherapy to treatment regiments for inoperable sinonasal tumors did not significantly improve survival (6). The above two phase II studies of induction chemotherapy did not include NUT cancers, and only one retrospective study of NUT cancers has analyzed the value of induction chemotherapy. Ramesh et al. (7) conducted a retrospective analysis of 12 patients with sinonasal NUT cancer and concluded that induction chemotherapy may be beneficial to patients. Patients with recurrent sinonasal NUT may be considered for proton radiotherapy. Muramatsu et al. (8) reported a case of sinonasal NUT carcinoma with local recurrence followed by reirradiation using proton radiotherapy, which led to complete response. In recent years, the rapid development of immunotherapy has led to the development of new options and useful additions to treatments for NUT cancer, which is often refractory. Currently, treatment with PD-1 and programmed death–ligand 1 (PD-L1) inhibitors has been reported for a small number of patients with lung, thyroid, and parotid NUT cancer (9, 10) but has not yet been reported for sinonasal NUT cancer. Moreover, there is no standard for consolidation regimens after CCRT. This study aimed to explore new treatment options for NUT cancer and strategies for consolidation immunotherapy after CCRT: we report the treatment of one patient with sinonasal NUT cancer with immunotherapy with PD-1 inhibitors after CCRT.

# 2 Case presentation

The patient was a 44-year-old female from Shandong, China. She was first admitted to our hospital on 17 June 2021, with the complaint of left eye pain and headache for 3 months. She had undergone surgery for congenital heart disease 30 years prior. She was admitted to the hospital and underwent relevant examinations. Nasal endoscopy revealed a mass in the left middle nasal meatus adjacent to the left middle nasal turbinate (Figure 1). Biopsy pathology revealed that the tumor cells were blue, rounded, heterogeneous cells, some of which were naked nucleated cells with minimal cytoplasm (Figures 2A, B). Immunohistochemical (IHC) staining revealed tumor cells that were positive for NUT expression (Figure 2C). The Ki-67 mitotic index was nearly 70%. Epstein-Barr virus (EBV)-encoded RNA (EBER) *in situ* hybridization was negative. Based on NUT IHC, a diagnosis of NUT cancer was established. Further testing revealed positive PD-L1 expression in both tumor cells and immune cells (Figure 2D). The percentages of tumor cells and immune cells with PD-L1 positivity were 65% and 1%, respectively. Enhanced computed tomography (CT) of the sinuses revealed that most of the mass was located in the left ethmoid sinus, with the mass invading the frontal sinus upward, invading the medial orbital wall and the rectus medialis to the left, with a discontinuity of bone in the medial orbital wall on the left side of the orbital wall, invading the intracranial area upward, and breaching the wall of the floor of the sieve sinus downward into the middle nasal passages (Figures 3A–C). 18F-fluorodeoxyglucose (18F-FDG) PET/CT demonstrated hypermetabolic activity with increased fluorodeoxyglucose uptake in the ethmoid sinus mass (maximum standardized uptake value, SUVmax 27.7) (Figure 3D). Regional lymph node involvement and metastatic disease were also excluded. Considering the results of nasal endoscopy, paranasal sinus CT, and PET/CT, the final diagnosis was sinonasal NUT carcinoma (cT4bN0M0, stage IVA AJCC-8 version). After discussion with our multidisciplinary team (MDT) of experts on head and neck tumors, including the Department of Radiology, Pathology, Otolaryngology, Radiation Oncology and Medical Oncology, we concluded that the tumor was inoperable and recommended definitive CCRT. The patient agreed with the treatment plan derived from the MDT discussion and underwent definitive CCRT in our department. The gross tumor volume (GTV) contained macroscopic primary tumor detectable on

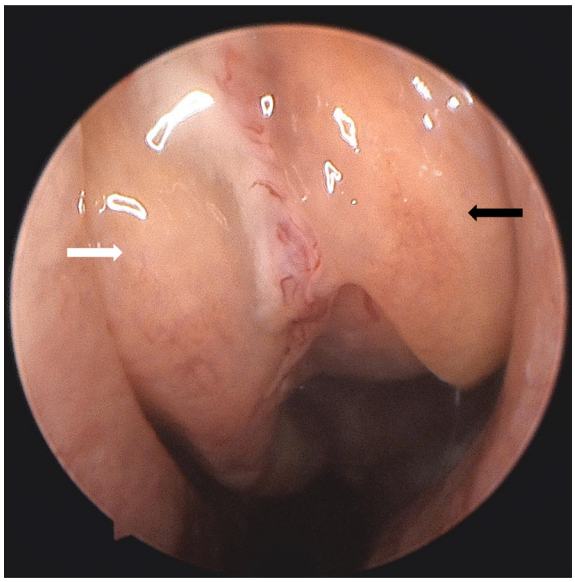


FIGURE 1
Nasal endoscopy revealed a mass (black arrow) in the left middle nasal meatus adjacent to the left middle turbinate (white arrow).

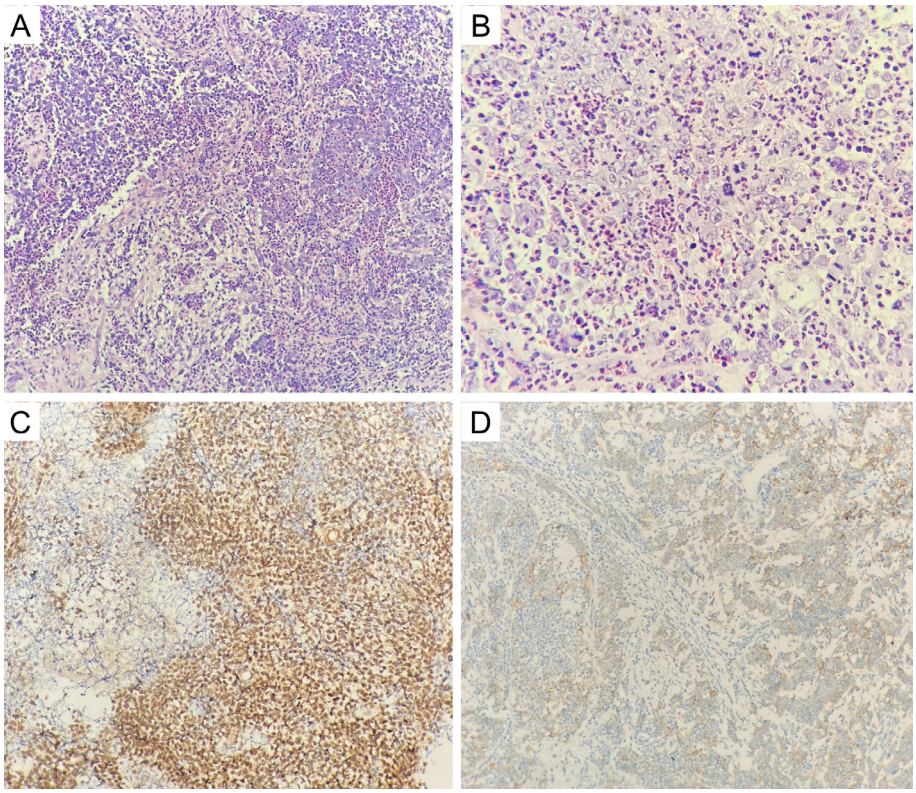



FIGURE 2
**(A, B)** The tumor cells were blue rounded heterogeneous cells, some of which were naked nucleated cells with minimal cytoplasm (hematoxylin and eosin; original magnification, ×200 and ×400); **(C)** positive NUT staining in the nucleus of tumor cells (hematoxylin and eosin; original magnification, ×200); **(D)** positive PD-L1 staining in tumor cells and immune cells (hematoxylin and eosin; original magnification, ×200).

CT imaging. The clinical target volume (CTV) was defined as the GTV plus a 5- to 10-mm margin to encompass the sites of microscopic extension including bilateral parapharyngeal space, bilateral retropharyngeal lymph node drainage area, left II–IV lymph drainage area, right partial II area, left cavernous sinus, bilateral sieve sinus and pterygoid sinus, frontal sinus, left maxillary sinus, nasopharyngeal, oropharyngeal, bilateral intrinsic nasal cavities, cranial base, bilateral pterygoid plate, pterygoid, medial pterygoid muscle, and pterygopalatine fossa. A planning target volume (PTV) was generated by incorporating a three-dimensional margin of 3 mm to the target volume in order to account for the uncertainties associated with treatment setup and internal organ mobility. The prescribed doses were 70 Gy and 60.06 Gy in 33 fractions, for the PTVs derived from GTV and CTV, respectively. Volume-modulated arc therapy technology was used to administer radiotherapy. During radiotherapy, the patient received three cycles of synchronized cisplatin chemotherapy (50 mg/m$^2$ days 1–2 q3w) and sodium glycididazole (1.25 g days 1, 3, and 5 qw). According to the Response Evaluation Criteria in Solid Tumors, the efficacy of chemoradiotherapy (1 month after CCRT) was evaluated as a partial response (PR). Afterward, the patient received three cycles of consolidation chemotherapy with albumin-bound paclitaxel (260 mg/m$^2$ day 1 q3w) and nedaplatin (80 mg/m$^2$ day 1 q3w). The efficacy of three cycles of consolidation chemotherapy was evaluated as PR (Figure 3E). Considering that NUT cancer is a highly malignant tumor with a poor prognosis, subsequent consolidation immunotherapy was agreed upon after thorough communication with the patient. Sixteen cycles of consolidation therapy with intermittent tislelizumab (200 mg day 1) were started on 7 January 2022, and the last immunotherapy treatment was given on 6 December 2023. Details of the timing of the use of tislelizumab are given in the Supplementary File. After 16 rounds of immunotherapy, the ethmoid sinus lesions achieved a state of sustained remission, and the efficacy assessment revealed a PR (Figure 3F). Adverse effects throughout treatment are tolerable. Acute toxicity during radiotherapy is mainly characterized by localized radiodermatitis of the facial skin. According to the Radiation Therapy Oncology Group's acute radiation morbidity scoring criteria for skin, acute radiation dermatitis was grade 1. There was no late toxicity after radiotherapy, and, to date, the patient has not experienced vision loss. The immune-related adverse events included grade 2 hypothyroidism. The recent workup on 6 December 2023 demonstrated no local recurrence or distant metastasis. The entire treatment timeline of the patient is shown in Figure 4.

## 3 Discussion

Sinonasal NUT carcinoma is very rare. There is a lack of large-scale epidemiologic studies and only a few retrospective studies with small sample sizes. A review of the literature by Lee et al. revealed

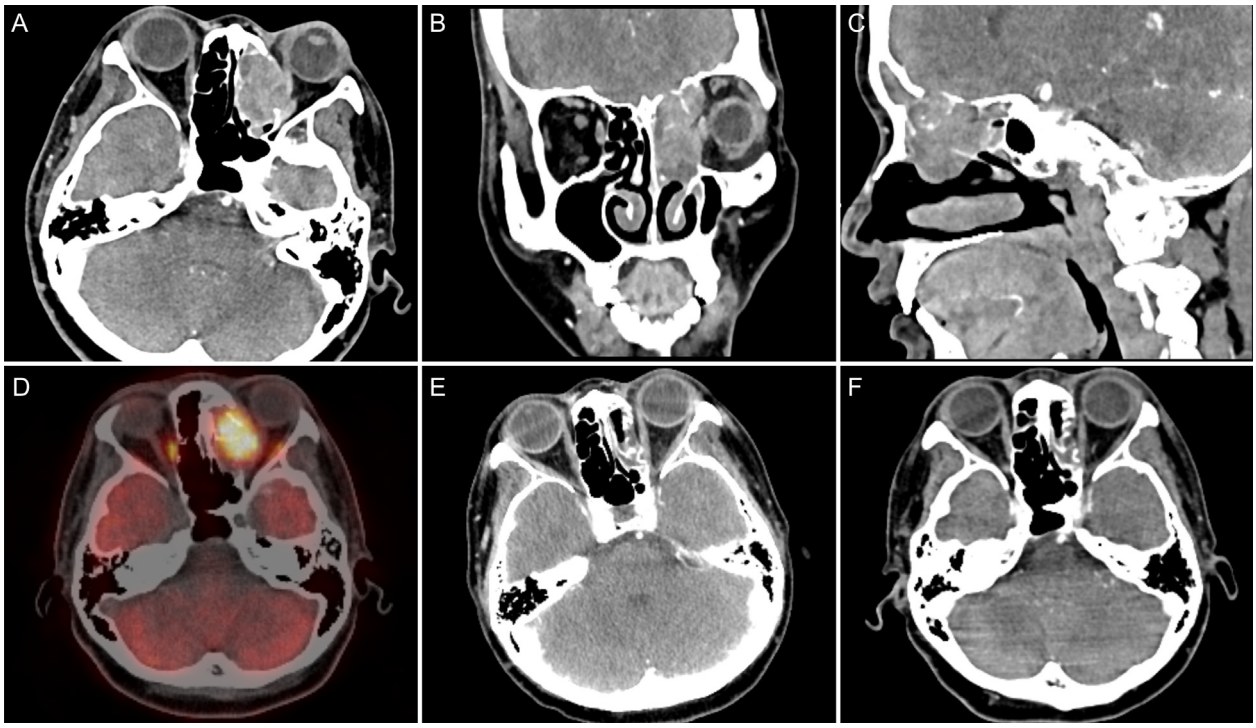



FIGURE 3
**(A)** CT axial view showed a mass in the left ethmoid sinus that was poorly demarcated from the left rectus medialis; **(B)** CT coronal view showed the mass invading the left orbit with intracranial invasion; **(C)** CT sagittal view shows the mass invading the frontal sinus with intracranial invasion; **(D)** PET/CT revealed a hypermetabolic mass with an SUVmax of 27.7 in the left ethmoid sinus; **(E)** CT after three cycles of consolidation chemotherapy revealed that the left ethmoid mass was considerably reduced compared with the previous mass, and the efficacy evaluation was PR; **(F)** the lesions reached a state of sustained remission after 13 cycles of immunotherapy, and the efficacy evaluation remained a PR.

that 4 of the 362 cases of poorly differentiated or undifferentiated carcinomas of the head and neck were sinonasal NUT carcinomas (11). A single-center study from China revealed that 3 of the 145 cases of sinonasal malignancies were NUT carcinomas (12). NUT has a very unfavorable prognosis. The median survival time for patients with primary NUT cancer of the chest is only 4.4 months (13). Compared to that for primary tumors in the lungs, the median survival time for primary NUT cancers of the head and neck is slightly greater, at only 9.7 months (13). In head and neck NUT cancers, survival may also vary depending on the location of the

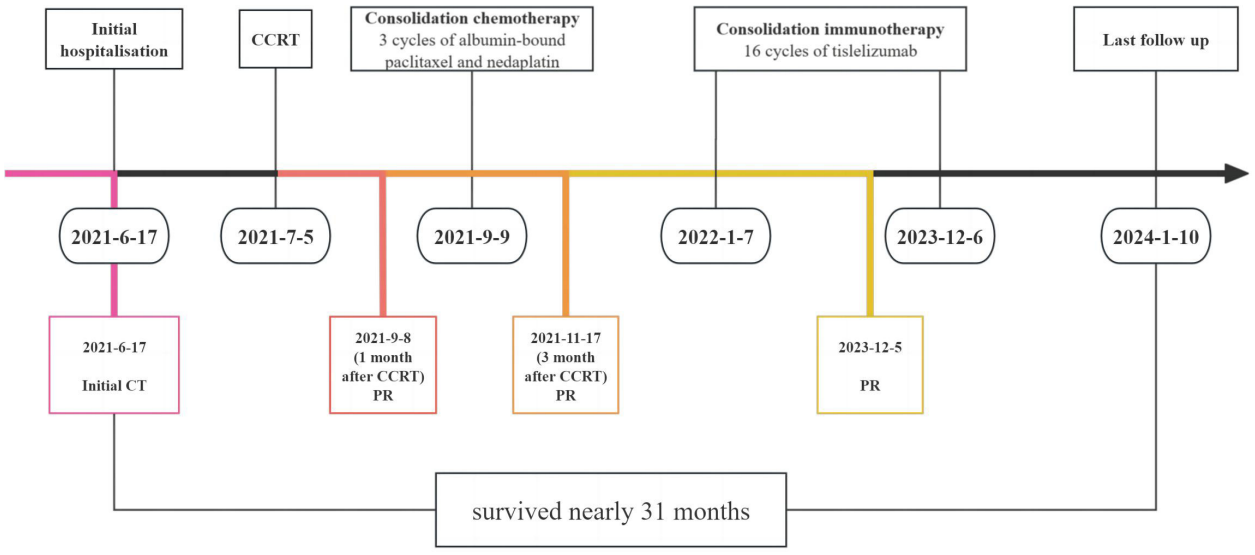



FIGURE 4
Timeline of the treatment. CCRT, concurrent chemoradiotherapy; PR, partial response.

primary focus. A recent single-center study of 12 patients with sinonasal NUT carcinoma revealed median OS and median disease-specific survival times of up to 14.6 months (7).

Pathologic diagnosis of NUT cancer includes IHC and genetic testing, and genetic testing methods include fluorescence *in situ* hybridization (FISH), next-generation sequencing (NGS), and reverse transcription–polymerase chain reaction (RT-PCR) (14). Diffuse (>50%) positive NUT expression on IHC is sufficient for diagnosing NUT cancer (15). The diagnosis of NUTMI molecular rearrangements by FISH and NGS is not necessary, but the above two tests are helpful in determining the prognosis of NUT cancer patients (13). In this case, the patient was diffusely positive for NUT expression according to IHC, and the diagnosis was confirmed on this basis.

We reviewed the literature on sinonasal NUT cancer (7, 8, 12, 16–30) (Table 1). Treatment decisions were made on the basis of the above literature by first assessing the tumor stage. The treatments for locally advanced disease include surgery combined with postoperative adjuvant radiotherapy, definitive chemoradiotherapy, induction chemotherapy combined with surgery, and induction chemotherapy combined with radiotherapy. The treatments for metastatic disease include chemotherapy combined with immunotherapy and debulking surgery and chemotherapy combined with local palliative radiotherapy. Bromodomain and extra-terminal domain inhibitors may be an option after progression on first-line therapy for metastatic NUT cancer. In this case, tumor was considered late stage and considered to be unsuitable for surgical treatment by the MDT; ultimately, CCRT was selected as the treatment strategy. After CCRT, we first performed consolidation chemotherapy include albumin-bound paclitaxel and nedaplatin. There are few previous studies on consolidation chemotherapy after CCRT for sinonasal cancer. The consolidation chemotherapy regimen can refer to the induction chemotherapy

TABLE 1 Summary of previous studies reporting sinonasal NUT cancer.

| Authors | Year | Number of patients | Location | Treatment | Outcome |
|---|---|---|---|---|---|
| Ramesh et al. (7) | 2024 | 12 | Nasal cavity maxillary sinus ethmoid sphenoid | Surgery ± induction chemotherapy; chemoradiotherapy ± induction chemotherapy | 14.6 months (median OS) |
| Caner et al. (16) | 2024 | 1 | Nasal cavity | Induction chemotherapy combined with immunotherapy + surgery + radiotherapy | 5.4 months (OS) |
| Qayum et al. (17) | 2024 | 1 | Maxillary sinus | Endoscopic sinus surgery+ maxillectomy | NM |
| Arai et al. (18) | 2024 | 1 | Maxillary sinus | Induction cheomotherapy + chemoradiotherapy (initial treatment); immunotherapy + targeted therapy and chemotherapy (after progression) | 11 months (OS) |
| Wang et al. (12) | 2023 | 3 | Nasal cavity maxillary sinus ethmoid | Surgery + radiotherapy | One patient 13 months (OS); two survived 12 and 15 months |
| Wartenberg et al. (19) | 2023 | 1 | Maxillary sinus | Surgery + radiotherapy | NM |
| da Costa et al. (20) | 2023 | 1 | Nasal cavity | Chemotherapy | NM |
| Zheng et al. (21) | 2023 | 1 | Maxillary sinus | Chemoradiotherapy + chemotherapy | NM |
| Wei et al. (22) | 2022 | 1 | Nasal sinus; frontal sinus | Surgery + radiotherapy | NM |
| Muramatsu et al. (8) | 2022 | 1 | Ethmoid | Chemoradiotherapy | survived 26 months |
| Patel et al. (23) | 2021 | 1 | Sphenoid | Ehemotherapy + radiotherapy (initial treatment); targeted therapy (after progression) | survived 21 months |
| Vakani et al. (24) | 2020 | 1 | Sphenoid | NM | NM |
| Oliveira et al. (25) | 2019 | 1 | Maxillary sinus | Induction cheomotherapy + chemoradiotherapy; chemotherapy (after progression) | NM |
| Arimizu et al. (26) | 2018 | 1 | Nasal cavity | Chemotherapy + chemoradiotherapy + surgery; chemotherapy + targeted therapy (after progression) | 9 months (OS) |
| Edgar et al. (27) | 2017 | 1 | Nasal cavity | Surgery + radiotherapy and chemotherapy | 3 months (OS) |
| Yang et al. (28) | 2015 | 1 | Nasal cavity | Surgery + radiotherapy and chemotherapy | Survived 10 months |
| Suzuki et al. (29) | 2014 | 1 | Nasal cavity | Chemoradiotherapy | Survived 12 months |
| Hsieh et al. (30) | 2011 | 1 | Nasal cavity | Chemoradiotherapy | NM |

NM, not mentioned.

regimen. The most commonly used induction chemotherapy regimens for sinonasal cancer are docetaxel and platinum (TP) and docetaxel, cisplatin, and fluorouracil (TPF) (31). TP includes paclitaxel and cisplatin, and TPF includes paclitaxel, cisplatin, and fluorouracil. These two consolidation chemotherapy regimens also provide a reference for consolidation chemotherapy for nasal sinus NUT cancer. At the end of CCRT and consolidation chemotherapy, the patient received consolidation immunotherapy based on three considerations. First, recent evidence has confirmed that consolidation therapy with immune checkpoint inhibitors (ICIs) can improve the survival of patients with lung and esophageal cancer after CCRT (32–34). In rare tumors, such as pulmonary sarcomatoid carcinoma (PSC) and pulmonary pleomorphic carcinoma (PPC), there are also case reports of encouraging results with consolidation immunotherapy after CCRT (35, 36). Although there is currently no evidence of survival benefits from ICI consolidation therapy after CCRT in patients with head and neck tumors, these findings in patients with lung and esophageal cancer can guide studies on head and neck tumors. A summary of clinical trials or case reports of consolidation immunotherapy after radiotherapy for malignant tumors is listed in Table 2 (32, 35–42). These clinical trials and case reports provide some reference for the selection of future consolidation ICI regimens after chemoradiotherapy for sinonasal NUT cancer patients. Second, although the patient's treatment efficacy after radiotherapy was evaluated as PR, NUT cancer has a poor prognosis and is prone to recurrence and metastasis, and good recent treatment efficacy may not necessarily indicate a good long-term prognosis. Thus, maintenance therapy may be needed after radical treatment is completed. Maintenance therapy requires the selection of an agent that is both highly effective and less toxic, and PD-1 inhibitor immunotherapy may be an option. Third, the case in this study had high PD-L1 expression in tumor cells. High-dose anti–PD-1/anti–PD-L1 therapy is generally believed to indicate a greater response rate and clinical benefit when PD-L1 is expressed (43). According to a recent report from the NUT symposium, patients with PD-L1 positivity or a high tumor mutation load can receive ICIs in combination with chemotherapy (15). Additionally, there have been recent case reports of the combined use of immunotherapy in head and neck NUT cancers. One patient had thyroid NUT cancer combined with carelizumab immunotherapy in addition to postoperative chemotherapy (9). Another case involved parotid NUT cancer, which was treated with targeted agents combined with sintilimab immunotherapy after multiple postoperative metastases were detected (10).

## 4 Conclusions

NUT carcinoma of the nasal cavity and sinuses is very rare, and only 50 cases have been reported in the literature. Surgery is the first choice for the treatment of NUT carcinoma, and CCRT can be chosen for patients who cannot be treated surgically. Compared with surgery, radiotherapy has the advantage of preserving organ function, thus improving the quality of life of patients. There is no consensus yet on consolidation treatment after CCRT. In this case, we studied a case of ethmoid NUT cancer that was treated with consolidation PD-1 inhibitor therapy after CCRT, which yielded a good therapeutic response. More case studies are needed in the future to validate the efficacy of consolidation immunotherapy after CCRT and to study the underlying mechanisms involved.

TABLE 2 Summary of clinical trials or case reports of consolidation ICI after chemoradiotherapy in patients with different cancer.

| Trial/NCT number/authors | ICI | Target | Dose | Frequency | Duration (time/cycles) | Cancer type |
|---|---|---|---|---|---|---|
| RATIONALE 311 (37) | Tislelizumab | PD-1 | 200 mg | q3w | 24 months | ESCC |
| Geng et al. (our case) | | | | Irregular | 23 months | Sinonasal NUT cancer |
| NCT03671265 (38) | Camrelizumab | PD-1 | 200 mg | q2w | 32 weeks (from the beginning of radiotherapy) | ESCC |
| InTRist (39) | Toripalimab | PD-1 | 240 mg | q3w | 12 months or until progression | NSCLC |
| KEYNOTE-412 (40) | Pembrolizumab | PD-1 | 200 mg | q3w | 14 cycles | HNSCC |
| PACIFIC (32) | Durvalumab | PD-L1 | 10 mg/kg | q2w | 12 months or until progression | NSCLC |
| ADRIATIC (41) | | | 1,500 mg | q4w | 24 months or until PD or intolerable toxicity | SCLC |
| Wang et al. (36) | | | 620 mg | q2w | 12 months | PSC |
| Yorozuya et al. (35) | | | NM | NM | 12 months | PPC |
| JAVELIN Head and Neck 100 (42) | Avelumab | PD-L1 | 10 mg/kg | q2w | 12 months | HNSCC |
| NCT03377400 (34) | Durvalumab | PD-L1 | 1,500 mg | q4w | 24 months | ESCC |
| | Tremelimumab | CTLA4 | 75 mg | | | |

ICI, immune checkpoint inhibitor; NCT, National Clinical Trial; PD-1, programmed cell death–1; PD-L1, programmed cell death–ligand 1; NM, not mentioned; q3w, every 3 weeks; q2w, every 2 weeks; q4w, every 4 weeks; CTLA4, cytotoxic T-lymphocyte antigen–4; ESCC, esophageal squamous cell carcinoma; NSCLC, non–small-cell lung cancer; SCLC, small cell lung cancer; HNSCC, head and neck squamous cell carcinoma; PSC, pulmonary sarcomatoid carcinoma; PPC, pulmonary pleomorphic carcinoma.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by Ethics Committee of Weifang People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

XG: Data curation, Writing – original draft, Writing – review & editing. XC: Data curation, Writing – review & editing. XW: Data curation, Writing – review & editing. SL: Writing – review & editing. GH: Data curation, Writing – review & editing. ZS: Data curation, Writing – review & editing. FH: Supervision, Writing – review & editing. JL: Conceptualization, Data curation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Medical and Health Science and Technology Development Program Project of Shandong Province (Grant No.202109030776).

## Acknowledgments

We thank the patient for her trust in us, her cooperation with the treatment, and her informed consent for the publication of this case report. We would also like to thank all the doctors and nurses who were responsible for this patient's treatment.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1368187/full#supplementary-material

## References

1. Virarkar M, Mallery M, Saleh M, Ramani NS, Morani AC, Bhosale P. Clinical, radiographic, pathologic characterization and survival outcomes of nuclear protein of the testis carcinoma. *J Comput Assist Tomogr*. (2021) 45:431–41. doi: 10.1097/RCT.0000000000001163

2. Chau NG, Ma C, Danga K, Al-Sayegh H, Nardi V, Barrette R, et al. An anatomical site and genetic-based prognostic model for patients with nuclear protein in testis (NUT) midline carcinoma: analysis of 124 patients. *JNCI Cancer Spectr*. (2020) 4: pkz094. doi: 10.1093/jncics/pkz094

3. French CA. The importance of diagnosing NUT midline carcinoma. *Head Neck Pathol*. (2013) 7:11–6. doi: 10.1007/s12105-013-0428-1

4. Chau NG, Hurwitz S, Mitchell CM, Aserlind A, Grunfeld N, Kaplan L, et al. Intensive treatment and survival outcomes in NUT midline carcinoma of the head and neck. *Cancer*. (2016) 122:3632–40. doi: 10.1002/cncr.30242

5. Resteghini C, Castelnuovo P, Nicolai P, Vischioni B, Nicolai P, Castelnuovo P, et al. The SINTART 1 study. A phase II non-randomised controlled trial of induction chemotherapy, surgery, photon-, proton- and carbon ion-based radiotherapy integration in patients with locally advanced resectable sinonasal tumours. *Eur J Cancer*. (2023) 187:185–94. doi: 10.1016/j.ejca.2023.03.033

6. Bossi P, Orlandi E, Resteghini C, Vischioni B, Nicolai P, Castelnuovo P, et al. The SINTART 2 Study. A phase II non-randomised controlled trial of induction chemotherapy, photon-, proton- and carbon-ion-based radiotherapy integration in patients with locally advanced unresectable sinonasal tumours. *Eur J Cancer*. (2023) 187:134–43. doi: 10.1016/j.ejca.2023.03.034

7. Ramesh U, Contrera KJ, Shakibai N, Su SY, Brahimaj B, Roberts D, et al. Sinonasal NUT carcinoma: A consecutive case series and systematic review. *Head Neck*. (2024) 46:29–36. doi: 10.1002/hed.27553

8. Muramatsu J, Takada K, Sugita S, Tsuchiya T, Yamamoto K, Takagi M, et al. Complete response induced by concurrent chemoradiotherapy in a patient with NUT carcinoma. *Intern Med*. (2022) 61:1299–304. doi: 10.2169/internalmedicine.7741-21

9. Zhou J, Duan M, Jiao Q, Chen C, Xing A, Su P, et al. Primary thyroid NUT carcinoma with high PD-L1 expression and novel massive IGKV gene fusions: A case report with treatment implications and literature review. *Front Oncol*. (2021) 11:778296. doi: 10.3389/fonc.2021.778296

10. Fu S, Wang Z, Li C, Li Y, Zhang K, Zhong Z, et al. The whole treatment process and thinking of a patient with NUT carcinoma of the parotid gland: a case report. *Front Oncol*. (2023) 13:1094770. doi: 10.3389/fonc.2023.1094770

11. Lee T, Cho J, Baek CH, Son YI, Jeong HS, Chung MK, et al. Prevalence of NUT carcinoma in head and neck: Analysis of 362 cases with literature review. *Head Neck*. (2020) 42:924–38. doi: 10.1002/hed.26067

12. Wang L, Zhu Z, Wang W, Zha Y, Wang X, Surita A, et al. Sinonasal NUT carcinoma: A retrospective case series from a single institution. *Front Surg*. (2023) 10:1098704. doi: 10.3389/fsurg.2023.1098704

13. French CA, Cheng ML, Hanna GJ, DuBois SG, Chau NG, Hann CL, et al. Report of the first international symposium on NUT carcinoma. *Clin Cancer Res*. (2022) 28:2493–505. doi: 10.1158/1078-0432.CCR-22-0591

14. Zhang Y, Han K, Dong X, Hou Q, Li T, Li L, et al. Case report and literature review: primary pulmonary NUT-midline carcinoma. *Front Oncol*. (2021) 11:700781. doi: 10.3389/fonc.2021.700781

15. Bishop JA. Newly described tumor entities in sinonasal tract pathology. *Head Neck Pathol*. (2016) 10:23–31. doi: 10.1007/s12105-016-0688-7

16. Caner B, Orhan SO, Deligonul A, Evrensel T. Immunotherapy experience in sinonasal NUT midline carcinoma, case report. *J Cancer Res Ther*. (2024) 20:479–81. doi: 10.4103/jcrt.jcrt_1083_22

17. Qayum A, Khan MWZ, Arshad AR, Hasnain S, Tariq MD, Khan S, et al. A rare case of P63-negative sinonasal nut midline carcinoma in the elderly. *Eur J Case Rep Intern Med*. (2024) 11:4265. doi: 10.12890/2024_004265

18. Arai S, Tomioka R, Ueda Y, Shimizu A, Okamoto I, Tsukahara K. Maxillary sinus NUT carcinoma: A case report. *Cancer Diagn Progn*. (2024) 4:370–8. doi: 10.21873/cdp.10334

19. Wartenberg M, Hool S-L, Marrazzini A, Giger R, Rupp NJ. Differentiated papillary NUT carcinoma: an unexpected, deceptively bland presentation of a sinonasal carcinoma. *Head Neck Pathol*. (2023) 17:803–7. doi: 10.1007/s12105-023-01554-w

20. Everton Assunção Ribeiro da Costa R, Luz Santos I, Júlia Andrade Pereira Soares M, Dos Reis de Paula I, Gerônimo da Silva Júnior R, Eduardo Coelho de Sá C. A case of metastatic NUT carcinoma of the nasal cavity. *Oral Oncol*. (2023) 142:106432. doi: 10.1016/j.oraloncology.2023.106432

21. Zheng YY, Cao YY, Li JZ, Chen XM. NUT carcinoma of the maxillary sinus in a child: a case report. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*. (2023) 58:1013–5. doi: 10.3760/cma.j.cn115330-20221202-00728

22. Wei X, Teng X, Zhang Y, Cheng M, Chen G. Case report: NUT carcinoma in an elderly woman with unique morphology and immunophenotype highlights a diagnostic pitfall. *Transl Cancer Res*. (2022) 11:1850–60. doi: 10.21037/tcr-22-364

23. Patel SA, Singer B, Shen C, Zanation AM, Yarbrough WG, Weiss J. A case of metastatic NUT carcinoma with prolonged response on gemcitabine and nab-paclitaxel. *Clin Case Rep*. (2021) 9:e04616. doi: 10.1002/ccr3.4616

24. Vakani PN, Maheshwari J, Maheshwari M, Shah B. Sinonasal NUT midline carcinoma: A new histological entity. *Indian J Pathol Microbiol*. (2020) 63:103–5. doi: 10.4103/ijpm.Ijpm_373_19

25. Oliveira LJC, Gongora ABL, Latancia MT, Barbosa FG, Gregorio JVAM, Testagrossa LA, et al. The first report of molecular characterized BRD4-NUT carcinoma in Brazil: a case report. *J Med Case Rep*. (2019) 13:279. doi: 10.1186/s13256-019-2213-6

26. Arimizu K, Hirano G, Makiyama C, Matsuo M, Sasaguri T, Makiyama A. NUT carcinoma of the nasal cavity that responded to a chemotherapy regimen for Ewing's sarcoma family of tumors: a case report. *BMC Cancer*. (2018) 18:1134. doi: 10.1186/s12885-018-5087-x

27. Edgar M, Caruso AM, Kim E, Foss RD. NUT midline carcinoma of the nasal cavity. *Head Neck Pathol*. (2017) 11:389–92. doi: 10.1007/s12105-016-0763-0

28. Yang L, Yang S. NUT midline carcinoma of sinonasal tract: report of a case. *Zhonghua Bing Li Xue Za Zhi*. (2015) 44:912–3. doi: 10.3760/cma.j.issn.0529-5807.2015.12.018

29. Suzuki S, Kurabe N, Minato H, Ohkubo A, Ohnishi I, Tanioka F, et al. A rare Japanese case with a NUT midline carcinoma in the nasal cavity: a case report with immunohistochemical and genetic analyses. *Pathol Res Pract*. (2014) 210:383–8. doi: 10.1016/j.prp.2014.01.013

30. Hsieh M-S, French CA, Liang C-W, Hsiao C-H. NUT midline carcinoma: case report and review of the literature. *Int J Surg Pathol*. (2011) 19:808–12. doi: 10.1177/1066896909353600

31. Melder KL, Geltzeiler M. Induction chemotherapy for locoregionally advanced sinonasal squamous cell carcinoma and sinonasal undifferentiated carcinoma: A comprehensive review. *Cancers (Basel)*. (2023) 15. doi: 10.3390/cancers15153798

32. Spigel DR, Faivre-Finn C, Gray JE, Vicente D, Planchard D, Paz-Ares L, et al. Five-year survival outcomes from the PACIFIC trial: durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *J Clin Oncol*. (2022) 40:1301–11. doi: 10.1200/JCO.21.01308

33. Zhang Y, Tian Y, Zheng L, Sun X, Zhao Z, Zheng Y, et al. Efficacy and safety of consolidation durvalumab after chemoradiation therapy for stage III non-small-cell lung cancer: a systematic review, meta-analysis, and meta-regression of real-world studies. *Front Pharmacol*. (2023) 14:1103927. doi: 10.3389/fphar.2023.1103927

34. Park S, Oh D, Choi Y-L, Chi SA, Kim K, Ahn MJ, et al. Durvalumab and tremelimumab with definitive chemoradiotherapy for locally advanced esophageal squamous cell carcinoma. *Cancer*. (2022) 128:2148–58. doi: 10.1002/cncr.34176

35. Yorozuya T, Taya T, Yasuda K, Nagano Y, Shioya M, Chiba H, et al. Long-term response with durvalumab after chemoradiotherapy for pulmonary pleomorphic carcinoma: A case report. *Thorac Cancer*. (2020) 11:1090–3. doi: 10.1111/1759-7714.13331

36. Wang Y, Yang L, Wang J, Gui L, Li W, Liu Z, et al. Case report: first case of consolidation immunotherapy after definitive chemoradiotherapy in mediastinal lymph node metastatic sarcomatoid carcinoma. *Front Oncol*. (2021) 11:788856. doi: 10.3389/fonc.2021.788856

37. Yu R, Wang W, Li T, Li J, Zhao K, Wang W, et al. RATIONALE 311: tislelizumab plus concurrent chemoradiotherapy for localized esophageal squamous cell carcinoma. *Future Oncol*. (2021) 17:4081–9. doi: 10.2217/fon-2021-0632

38. Zhang W, Yan C, Zhang T, Chen X, Dong J, Zhao J, et al. Addition of camrelizumab to docetaxel, cisplatin, and radiation therapy in patients with locally advanced esophageal squamous cell carcinoma: a phase 1b study. *Oncoimmunology*. (2021) 10:1971418. doi: 10.1080/2162402x.2021.1971418

39. Wang Y, Deng L, Wang J, Zhang T, Wang W, Wang X, et al. Induction PD-1 inhibitor toripalimab plus chemotherapy followed by concurrent chemoradiotherapy and consolidation toripalimab for bulky locally advanced non-small-cell lung cancer: protocol for a randomized phase II trial (InTRist study). *Front Immunol*. (2023) 14:1341584. doi: 10.3389/fimmu.2023.1341584

40. Machiels J-P, Tao Y, Licitra L, Burtness B, Tahara M, Rischin D, et al. Pembrolizumab plus concurrent chemoradiotherapy versus placebo plus concurrent chemoradiotherapy in patients with locally advanced squamous cell carcinoma of the head and neck (KEYNOTE-412): a randomised, double-blind, phase 3 trial. *Lancet Oncol*. (2024) 25:572–87. doi: 10.1016/S1470-2045(24)00100-1

41. Senan S, Okamoto I, Lee GW, Chen Y, Niho S, Mak G, et al. Design and rationale for a phase III, randomized, placebo-controlled trial of durvalumab with or without tremelimumab after concurrent chemoradiotherapy for patients with limited-stage small-cell lung cancer: the ADRIATIC study. *Clin Lung Cancer Mar*. (2020) 21:e84–8. doi: 10.1016/j.cllc.2019.12.006

42. Lee NY, Ferris RL, Psyrri A, Haddad RI, Tahara M, Bourhis J, et al. Avelumab plus standard-of-care chemoradiotherapy versus chemoradiotherapy alone in patients with locally advanced squamous cell carcinoma of the head and neck: a randomised, double-blind, placebo-controlled, multicentre, phase 3 trial. *Lancet Oncol Apr*. (2021) 22:450–62. doi: 10.1016/s1470-2045(20)30737-3

43. Yi M, Jiao D, Xu H, Liu Q, Zhao W, Han X, et al. Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors. *Mol Cancer*. (2018) 17:129. doi: 10.1186/s12943-018-0864-3

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 24 December 2024
DOI 10.3389/fonc.2024.1397877



OPEN ACCESS

EDITED BY
Timothy James Kinsella,
Brown University, United States

REVIEWED BY
Carla Pisani,
Azienda Ospedaliero Universitaria Maggiore della Carità, Italy
Liangliang Wang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Suning Huang
huangsuning@stu.gxmu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 08 March 2024
ACCEPTED 11 December 2024
PUBLISHED 24 December 2024

CITATION
Lu X, Tan B, Yang L and Huang S (2024) Complete remission after sintilimab combined with chemoradiotherapy in double primary head and neck carcinoma: case report.
*Front. Oncol.* 14:1397877.
doi: 10.3389/fonc.2024.1397877



# Complete remission after sintilimab combined with chemoradiotherapy in double primary head and neck carcinoma: case report

Xiameng Lu[1,2†], Bibo Tan[1†], Liuting Yang[1] and Suning Huang[1*]

[1]Department of Radiation Oncology, Guangxi Medical University Cancer Hospital, Nanning, Guangxi, China, [2]Medical College of Oncology, Guangxi Medical University, Nanning, Guangxi, China

The simultaneous occurrence of head and neck squamous carcinoma in two anatomical sites is rare, posing challenges in treatment selection. This paper presents a clinical case of concurrent hypopharyngeal carcinoma and nasopharyngeal carcinoma, successfully treated with a combination of chemoradiotherapy and an immune checkpoint inhibitor. The patient achieved complete remission and progression-free survival of nearly 3 years, with preserved organ function and minimal toxic side effects, leading to a good quality of life. This case highlights the potential of combined concurrent chemoradiotherapy and immune checkpoint inhibitors in managing double primary HNSCC, offering a promising treatment option for these patients.

KEYWORDS

**head and neck carcinoma, double primary tumor, immune checkpoint inhibitor, chemotherapy, radiotherapy, case report**

## Introduction

Hypopharyngeal cancer, a common malignant tumor with a poor prognosis, poses a significant threat to human health and quality of life. Nasopharyngeal cancer typically presents as locally advanced at diagnosis, with radiotherapy being the primary treatment. Although both hypopharyngeal and nasopharyngeal cancers are classified as head and neck squamous carcinoma (HNSCC), the incidence of double primary cancers arising in these two regions is rare. Consequently, there is a scarcity of treatment protocols that specifically address this uncommon condition. Moreover, the potential of immunotherapy in treating double primary cancers warrants further exploration.

In this report, we present the case of a 70-year-old male patient diagnosed with hypopharyngeal and nasopharyngeal cancers who underwent a treatment regimen consisting of chemoradiotherapy and immunotherapy. The patient achieved complete remission in both tumors and no evidence of recurrence or metastasis during 35 months of

regular follow-up assessments. Furthermore, the patient experienced minimal toxic side effects and maintained good organ function and a high quality of life.

This case highlights the potential efficacy of combining chemoradiotherapy with immune checkpoint inhibitors (ICIs) managing advanced double primary HNSCC, suggesting a promising treatment approach for patients with similar presentations.

## Case presentation

We present a case of a 70-year-old male patient. The patient presented with a sore throat and obstruction sensation. Over the past three months, the patient's throat pain has relapsed repeatedly, with worsening the sensation of swallowing obstruction and difficulty opening the mouth, leading to a consultation at the hospital. A head and neck magnetic resonance imaging (MRI) showed that an abnormal signals in the posterior wall of the hypopharynx, with invasion into the right pyriform sinus, epiglottis, and bilateral aryepiglottic folds. The tumor demonstrated unclear demarcation from the thyroid and compressed the trachea. The nasopharyngeal wall is thickened, with enlarged retropharyngeal lymph nodes (Figure 1). PET-CT scans were performed and showed no signs of metastasis in other organs, except for hypermetabolism in the hypopharyngeal, nasopharynx and metastatic lymph nodes in the neck. Subsequently, surgical procedures performed on the patient included direct laryngoscopy, transoral laryngeal mass excision, rigid esophagoscopy, and tracheostomy. The tracheostomy was performed to establish an open airway due to the trachea compression and to prevent potential suffocation during treatment. During the laryngoscopy, a mass was observed on the posterior wall of the hypopharynx, extending into the right pyriform sinus. A biopsy of the mass was taken for pathological diagnosis. The esophagoscopy was performed, and no abnormalities were noted. Pathology indicated keratinized squamous cell carcinoma of the hypopharynx (Figure 2A). He then underwent a fibro-laryngoscope examination and revealed a nasopharynx mass. A biopsy of the nasopharyngeal confirmed it was non-keratinizing undifferentiated squamous cell carcinoma (Figure 2B), with positive EBERs staining.

The patient was in good physical health prior to treatment, exhibiting no heart, brain, liver, or kidney diseases, and was not on any long-term medication. He worked as a farmer, with no exposure to chemical toxins, industrial dust, or radiation. The patient has a 30-year smoking history, with a smoking index of 450 (30 pack-years). And he has consumed alcohol for 30 years, averaging 50 mg per day. Following his diagnosis in July 2021, he ceased both smoking and alcohol consumption.

In conclusion, the patient presented with a sore throat and odynophagia. Physical examination revealed palpable lymph nodes in the right neck. Imaging examination identified masses in the hypopharynx and the nasopharynx, with local invasion into the thyroid and trachea. Pathological examination indicated different histological type in hypopharynx and nasopharynx. Two anatomical site masses with different histological type raise the certainty of a double primary malignancy. The diagnosis of double primary cancers presents challenges. In terms of hypopharyngeal carcinoma, the tumor was located in the posterior wall of the hypopharynx, invading the pyriform sinus, thyroid and trachea, and was classified as T4a. Bilateral cervical lymph node metastasis leading to a staging of N2c. In terms of NPC staging, the patient exhibited a thickened posterior nasopharyngeal wall without infiltration of the pharyngobasilar fascia, leading to a local classification of T1. Besides, the boundary between the retropharyngeal lymph node and nasopharynx was unclear while the lymph node was anatomically distant from the hypopharyngeal tumor. Therefore, we prefer that nasopharyngeal carcinoma with retropharyngeal lymph node metastasis, leading to a staging of N1. Following the staging criteria for double primary tumors, the diagnosis was as follows: 1. Hypopharyngeal squamous cell carcinoma (T4aN2cM0 IVA, AJCC$^{8th}$); 2. Nasopharyngeal non-keratinizing undifferentiated carcinoma (T1N1M0 II, AJCC$^{8th}$).

Following a multidisciplinary discussion, the treatment approach of "radiotherapy-chemotherapy-immunotherapy" has been established. The patient received three cycles of induction therapy, including albumin-bound paclitaxel combined with cisplatin chemotherapy, along with the ICI sintilimab. The doses were as follows: albumin-bound paclitaxel 260mg/m$^2$ d1, cisplatin 75mg/m$^2$ d1-2, sintilimab 200mg d1. Subsequently, the patient received concurrent chemotherapy and intensified modulated radiotherapy (IMRT). The gross tumor volume (GTV) included tumors in the hypopharynx and nasopharynx, retropharyngeal lymph nodes, and adjacent organs invaded by the tumor. The GTVnd was defined as the metastatic cervical lymph nodes. The GTV was expanded by 1 cm to establish the clinical tumor volume (CTV), which encompassed potential sites of tumor invasion and the drainage area of the bilateral cervical lymph nodes. The prescribed radiation doses were as follows: 6MeV, X-ray, PGTV: 60.2Gy/28f (2.15Gy/f), PGTVnd-L/R: 60.2Gy/28f (2.15Gy/f), PCTV: 50.4Gy/28f (1.8Gy/f), with acceptable dose limits for normal organs (Figure 3). The patient also received the fourth cycle of chemotherapy combined with sintilimab shortly before completing radiotherapy, maintaining the exact dosages as previously mentioned. Throughout the treatment, this patient experienced grade 2 leukopenia, grade 2 acute radiation-induced mucositis, and mild gastrointestinal reactions. A nutrition nursing unit was established in the radiotherapy ward to provide comprehensive nutritional guidance, including intravenous nutrition and oral supplementation. The patient maintained a satisfactory nutritional status, experiencing only a 2 kg weight loss, and therefore did not undergo a gastrostomy.

**Abbreviations:** HNSCC, head and neck squamous cell carcinoma; ICI, immune checkpoint inhibitor; NPC, nasopharyngeal carcinoma; MRI, magnetic resonance imaging; IMRT, intensified modulated radiotherapy; NCCN, National Comprehensive Cancer Network; PFS, progression-free survival; ORR, objective response rate; OS, overall survival; PD-1, programmed cell death protein 1.

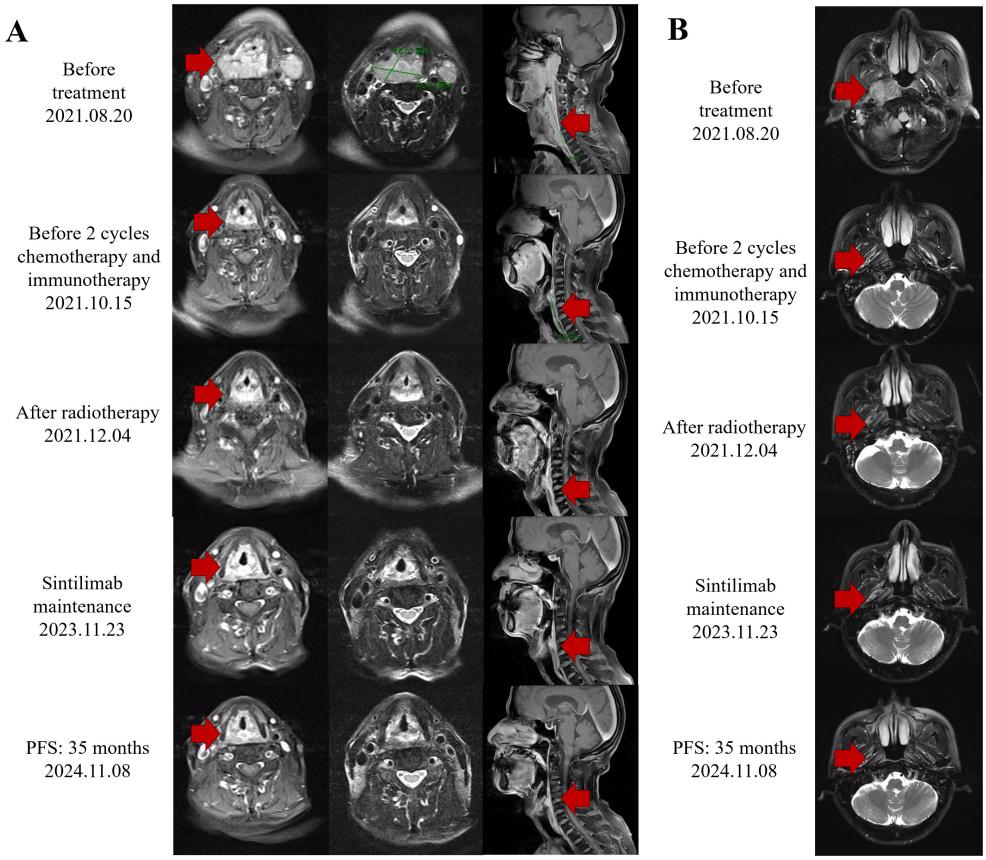



FIGURE 1
Response evaluation during the clinical course. **(A)** Representative images of the MRI scan revealed the decreasing process of both hypopharyngeal primary cancer and metastasis lymph nodes. Red arrows indicate tumor lesions. **(B)** Representative images of the MRI scan revealed the decreasing process of both nasopharyngeal primary cancer. Red arrows indicate tumor lesions.

An MRI re-assessment after comprehensive therapy indicated a complete tumor regression, including hypopharyngeal and nasopharyngeal mass, as well as the adjacent peripheral soft tissue and metastatic lymph nodes (Figure 2). The nasopharyngoscope confirmed that the tumor mass had complete remission. The patient subsequently received regular immunotherapy maintenance for a year. The therapeutic efficacy was assessed to have achieved a complete clinical response (cCR). Over a 35months follow-up period, the patient's self-care abilities and organ functions remained usual, with no adverse events or complications reported. The clinical timeline is shown in Figure 4.

## Discussion

Multiple primary malignant tumors, referring to the patient simultaneously occurring two or more primary malignant tumor, present significant clinical challenges. The pathogenesis of multiple primary tumors has not been clarified but may be related to genetic susceptibility, viral infection, and radiation induction (1). The double primary head and neck cancer in hypopharyngeal and nasopharyngeal is rare and presents a poor prognosis and low quality of life, posing substantial challenges for clinical treatment.

This case involves a patient with double primary HNSCC. Based on the multiple primary tumors treatment guideline in China and the poorer prognosis associated with hypopharyngeal carcinoma, we prioritized the control of hypopharyngeal cancer while incorporating treatment strategies for NPC (2). The treatment approach of "radiotherapy-chemotherapy-immunotherapy" has been established. For one thing, induction chemotherapy followed by concurrent chemoradiotherapy is recommended by National Comprehensive Cancer Network (NCCN) guideline due to the complicated anatomy in nasopharynx and hypopharynx. Specifically, paclitaxel combined with platinum is suitable for both induction and concurrent chemotherapy in hypopharyngeal carcinoma (category 1). Besides, locally advanced HNSCC presents challenges in effectively targeting cells at the center of the tumor, resulting in treatment resistance. Immunotherapy enhances body's anti-tumor immunity, enabling the elimination of these resistance tumor cells.

We finalized a treatment consisting of a induction chemotherapy and partially hypofractionated radiotherapy dose of 60.20 Gy over 28 fractions (2.15 Gy per fraction) with concurrent ICI and chemotherapy. On the one side, in the 2022 NCCN guidelines, the combination of paclitaxel with cisplatin was recommended for induction chemotherapy in HNSCC (category 1). Albumin-bound paclitaxel, approved by the US FDA for the first-line treatment of

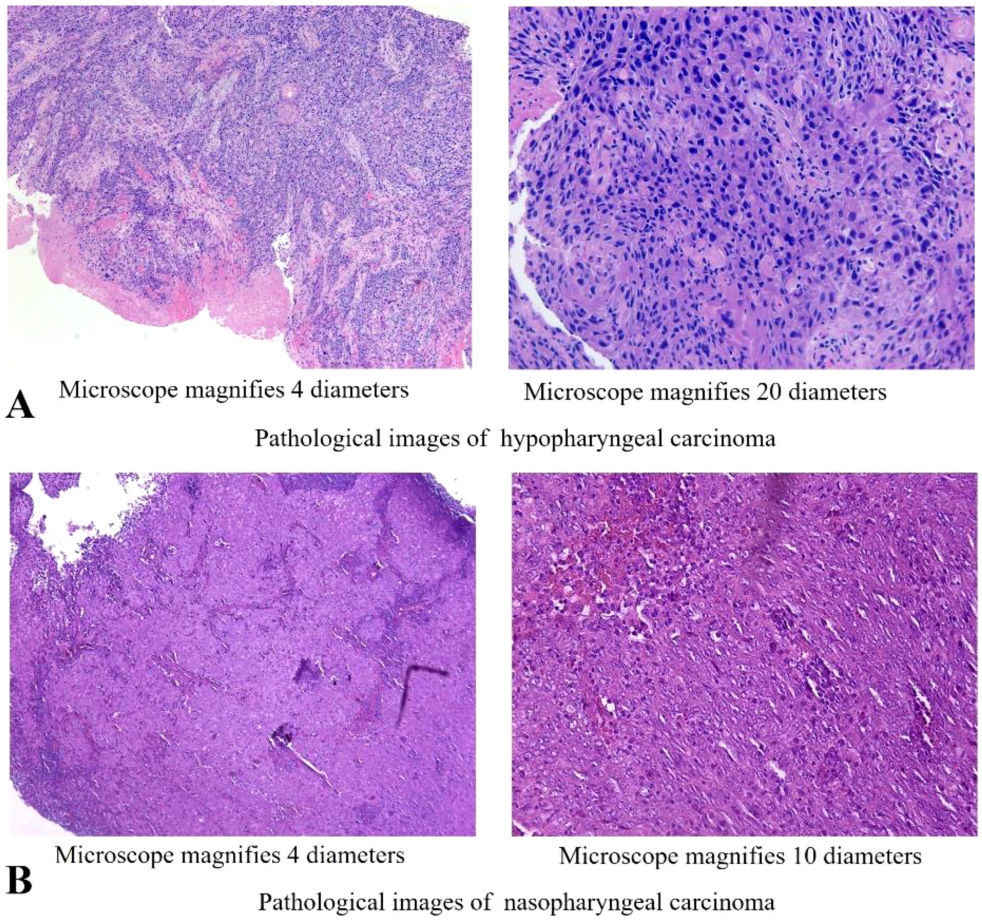



FIGURE 2
Histopathology of the cancer of this patient. **(A)** Microscopic observation (4x,20x) of H&E staining of the hypopharyngeal lesion. **(B)** Microscopic observation (4x, 10x) of H&E staining of the nasopharyngeal lesion.

non-small cell lung cancer (3), breast cancer (4), ovarian cancer (5), and pancreatic cancer (6), has been demonstrated to simplify pre-treatments and reduce allergic events during chemotherapy. The dosage was settled: albumin-bound paclitaxel 260mg/m$^2$ d1, cisplatin 75mg/m$^2$ d1-2. Furthermore, the 2022 NCCN guidelines also recommend paclitaxel and cisplatin as concurrent chemotherapy agents (category 2B) or weekly cisplatin for concurrent chemotherapy following induction chemotherapy (category 2B). Therefore, the patient received three cycles of albumin-bound paclitaxel with cisplatin before radiation and exhibiting a favorable response (Figure 2). Then he received the fourth cycle of albumin-bound paclitaxel with cisplatin based on the tumor response and the guideline.

On the other side, China was in the challenging stage of COVID-19 epidemic prevention and control measures at the time of treatment. Hypofractionated radiotherapy, which shortens the

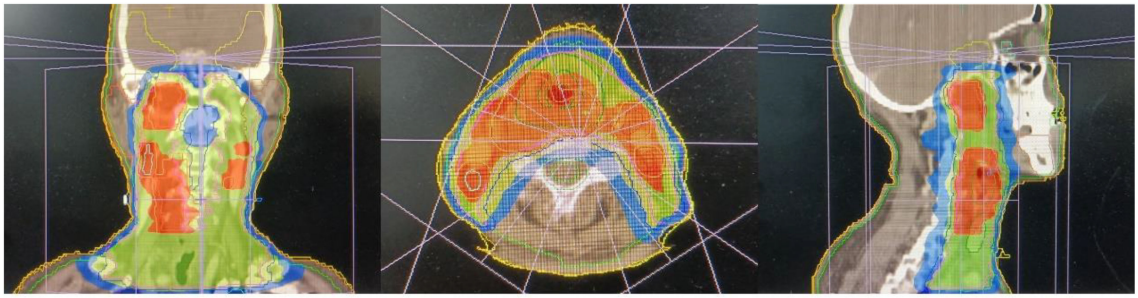


FIGURE 3
Dose profile of the primary cancers and the metastasis lymph nodes.

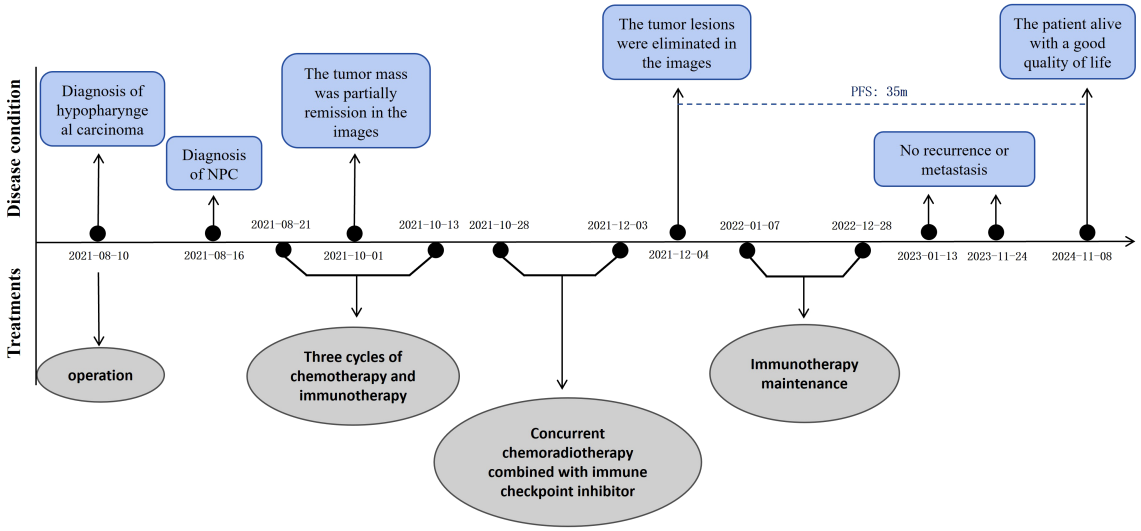



FIGURE 4
The whole clinical timeline of the patient, with major treatment and disease status. PFS, progression-free survival.

overall treatment duration and minimizes patient hospital exposure, was anticipated to lower the risk of COVID-19 infection (7). Evidence indicates that the hypofractionated regimen (2.2-2.75 Gy per fraction, daily for five days a week) demonstrates enhanced efficacy and reduced toxicity in elderly patients with HNSCC (8, 9). The ASTRO and ESTRO guidelines also recommend increasing conventional radiation doses for HNSCC during the pandemic (10). This hypofractionated radiotherapy protocol was determined with the patient's full informed consent.

Despite receiving chemoradiotherapy, the prognostic survival of patients with local advanced HNSCC, especially double primary carcinoma, remains unsatisfactory. Studies have confirmed that the median survival time for double primary cancers is 22.8 months (1). Researches showed that immunotherapy improved patients' long-term survival with locally advanced HNSCC. For immunotherapy, pembrolizumab was recommended as a first-line treatment for unresectable HNSCC. Both sintilimab and pembrolizumab are PD-1 inhibitors that limit tumor evasion and activate the immune response. Sintilimab, an IgG4 monoclonal antibody, mitigates tumor-induced immunosuppression by binding to programmed cell death protein 1 (PD-1), inhibiting its interaction with programmed death ligands[2]. Additionally, sintilimab is economically accessible for patients. Research has consistently shown the antitumor efficacy of sintilimab. In recent years, there has been a deep exploration of programmed cell death receptor 1 (PD-1)/programmed death ligand 1 (PD-L1) immune checkpoint inhibitors. Theoretically, ICIs reduce immune evasion by tumor cells and enhance the function of tumor-specific T-cells, potentially improving overall anti-tumor effectiveness (11). Prospective trials have shown that ICIs significantly improve objective response rate (ORR) and 2-year overall survival (OS) in HNSCC patients. A Phase II clinical trial conducted by Professor Zhou and his colleagues investigated the efficacy and safety of sintilimab in patients with platinum-resistant HNSCC. The trial results demonstrated that sintilimab significantly prolonged progression-free survival (PFS) without increasing the incidence of adverse effects (12). Besides, the KEYNOTE-048 study (NCT02358031) at Yale Cancer Center showed that pembrolizumab combined with chemotherapy had a better OS than the standard treatment in HNSCC patients (13). Checkmate 141 and Keynote-040 studies demonstrated a significant OS benefit with the second line of the immune checkpoint inhibitor PD-1 over standard therapy in recurrent or metastatic HNSCC (14, 15). These research findings indicated that sintilimab can be a promising immune checkpoint inhibitor in locally advanced HNSCC, which offering new treatment options for HNSCC or other carcinoma, thereby expanding the therapeutic possibilities for cancer patients.

Furthermore, the optimal duration of maintenance immunotherapy remains uncertain. Based on previous research findings, a maintenance immunotherapy period of 1 to 2 years is a viable option (16, 17). The patient ultimately received ICI maintenance therapy for 1 year but discontinued it for personal reasons. Clinical trials (NCT04557020, NCT04453826, NCT03700476) are currently enrolling participants, aiming to advance medical applications and guide optimal treatment strategies for HNSCC in the era of immunotherapy.

However, there are limitations to the treatments and dilemmas in strategizing the approach. On the one hand, despite the patient achieving nearly complete tumor regression during induction treatment, as well as the bioequivalent dose of the radiotherapy regimen exceeding 70 Gy, conventional fractionation remains the predominant method for treating HNSCC. Some hypofractionated radiotherapy protocols are still in clinical trials, lacking randomized controlled trials and long-term follow-up data. Nevertheless, upon reviewing the available data, many experts have endorsed partially hypofractionated radiotherapy as the preferred dosage during the pandemic, as it minimizes hospital exposure, particularly for elderly patients, and reduces the risk of COVID-19 infection while maintaining effective tumor control without significantly increasing adverse effects (8, 9). On the other hand, immune checkpoint inhibitors may lead to immune-related adverse events (irAE). Therefore,

monitoring key indicators regarding the patient's heart, liver, kidneys, lungs, thyroid, and other organs is essential throughout the treatment and follow-up. This approach enables prompt identification and management of irAE, facilitating timely interventions or drug discontinuation when necessary.

In all, the treatment strategy takes into account the COVID-19 epidemic and the patient's characteristics while concerning the treatment guidelines. The patient was well-informed about the treatment strategy and demonstrated good compliance, facilitating the treatment process and minimizing side effects. All therapeutic interventions were communicated clearly, and comprehensive informed consent was secured from the patient. After treatment, he regularly received comprehensive examinations, revealing no disfunction in organs and no signs of tumor progression. The patient's quality of life was satisfactory.

## Conclusion

This patient diagnosed with double primary cancer underwent treatment with sintilimab and chemoradiotherapy, resulting in complete regression of the tumor. This case underscores the efficacy and safety of this regimen for hypopharyngeal carcinoma and nasopharyngeal carcinoma. Nevertheless, further meta-analyses and real-world clinical trials are warranted to validate its efficacy in managing multiple primary HNSCC.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

XL: Investigation, Writing – original draft, Writing – review & editing. BT: Investigation, Writing – original draft, Writing – review & editing. LY: Writing – review & editing. SH: Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

We thank the patient who participated in this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Amer M. Multiple neoplasms, single primaries, and patient survival. *Cancer Manag Res.* (2014) 119:119–34. doi: 10.2147/CMAR.S57378

2. Xichun HU, Jianwei LU. China Anti-Cancer Association guideline for diagnosis and treatment of cancer of multiple and unknown primaries (2023 edition). *China Oncol.* (2023) 33:403–22. doi: 10.19401/j.cnki.1007-3639.2023.04.011

3. Socinski MA, Bondarenko I, Karaseva NA, Makhson AM, Vynnychenko I, Okamoto I, et al. Weekly *nab* -paclitaxel in combination with carboplatin versus solvent-based paclitaxel plus carboplatin as first-line therapy in patients with advanced non–small-cell lung cancer: final results of a phase III trial. *J Clin Oncol.* (2012) 30:2055–62. doi: 10.1200/JCO.2011.39.5848

4. Li X, Kwon H. Efficacy and safety of nanoparticle albumin-bound paclitaxel in elderly patients with metastatic breast cancer: A meta-analysis. *J Clin Med.* (2019) 8:1689. doi: 10.3390/jcm8101689

5. Liao JB, Swensen RE, Ovenell KJ, Hitchcock-Bernhardt KM, Reichow JL, Apodaca MC, et al. Phase II trial of albumin-bound paclitaxel and granulocyte macrophage colony-stimulating factor as an immune modulator in recurrent platinum resistant ovarian cancer. *Gynecol Oncol.* (2017) 144:480–5. doi: 10.1016/j.ygyno.2017.01.008

6. Seufferlein T, Uhl W, Kornmann M, Algül H, Friess H, König A, et al. Perioperative or only adjuvant gemcitabine plus nab-paclitaxel for resectable pancreatic cancer (NEONAX)—a randomized phase II trial of the AIO pancreatic cancer group. *Ann Oncol.* (2023) 34:91–100. doi: 10.1016/j.annonc.2022.09.161

7. Nagar H, Formenti SC. Cancer and COVID-19 — potentially deleterious effects of delaying radiotherapy. *Nat Rev Clin Oncol.* (2020) 17:332–4. doi: 10.1038/s41571-020-0375-1

8. Benhmida S, Sun R, Gherga E, Hammoud Y, Rouvier J, Mauvais O, et al. Split-course hypofractionated radiotherapy for aged and frail patients with head and neck cancers. A retrospective study of 75 cases. *Cancer/Radiothèrapie.* (2020) 24:812–9. doi: 10.1016/j.canrad.2020.03.013

9. Bonomo P, Desideri I, Loi M, Lo Russo M, Olmetto E, Maragna V, et al. Elderly patients affected by head and neck squamous cell carcinoma unfit for standard curative treatment: Is de-intensified, hypofractionated radiotherapy a feasible strategy? *Oral Oncol.* (2017) 74:142–7. doi: 10.1016/j.oraloncology.2017.10.004

10. Thomson DJ, Palma D, Guckenberger M, Balermpas P, Beitler JJ, Blanchard P, et al. Practice recommendations for risk-adapted head and neck cancer radiation therapy during the COVID-19 pandemic: an ASTRO-ESTRO consensus statement. *Int J Radiat Oncol Biol Phys*. (2020) 107:618. doi: 10.1016/j.ijrobp.2020.04.016

11. Sun Y, Li WF, Chen NY, Zhang N, Hu GQ, Xie FY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicenter, randomized controlled trial. *Lancet Oncol*. (2016) 17:1509–20. doi: 10.1016/S1470-2045(16)30410-7

12. Li X, Fang Q, Du W, Zhang X, Dai L, Qiao Y. Induction chemotherapy combined with immunotherapy in locally advanced head and neck squamous cell carcinoma. *BMC Cancer*. (2021) 21:622. doi: 10.1186/s12885-021-08373-8

13. Burtness B, Rischin D, Greil R, Soulières D, Tahara M, de Castro G, et al. Pembrolizumab alone or with chemotherapy for recurrent/metastatic head and neck squamous cell carcinoma in KEYNOTE-048: subgroup analysis by programmed death ligand-1 combined positive score. *J Clin Oncol Off J Am Soc Clin Oncol*. (2022) 40:2321–32. doi: 10.1200/JCO.21.02198

14. Cohen EEW, Soulières D, Le Tourneau C, Dinis J, Licitra L, Ahn MJ, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomized, open-label, phase 3 study. *Lancet*. (2019) 393:156–67. doi: 10.1016/S0140-6736(18)33261-6

15. Ferris RL, Blumenschein G, Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med*. (2016) 375:1856–67. doi: 10.1056/NEJMoa1602252

16. Waterhouse DM, Garon EB, Chandler J, McCleod M, Hussein M, Jotte R, et al. Continuous versus 1-year fixed-duration nivolumab in previously treated advanced non–small-cell lung cancer: checkMate 153. *J Clin Oncol*. (2020) 38:3863. doi: 10.1200/JCO.20.00131

17. Sun L, Bleiberg B, Hwang WT, Marmarelis ME, Langer CJ, Singh A, et al. Association between duration of immunotherapy and overall survival in advanced non–small cell lung cancer. *JAMA Oncol*. (2023) 9:1075–82. doi: 10.1001/jamaoncol.2023.1891

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 January 2025
DOI 10.3389/fonc.2025.1382537



OPEN ACCESS

EDITED BY
Timothy James Kinsella,
Brown University, United States

REVIEWED BY
Renjie He,
University of Texas MD Anderson Cancer Center, United States
Prospero Naval,
University of the Philippines, Philippines
Bin Wang,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Gaia Muti
Gaia.Muti@ospedaleniguarda.it

RECEIVED 05 February 2024
ACCEPTED 03 January 2025
PUBLISHED 23 January 2025

CITATION
Muti G, Felisi MMJ, Monti AF, Carsana C, Pellegrini R, Salmeri E, Palazzi M and Colombo PE (2025) Proof of concept of fully automated adaptive workflow for head and neck radiotherapy treatments with a conventional linear accelerator.
*Front. Oncol.* 15:1382537.
doi: 10.3389/fonc.2025.1382537



# Proof of concept of fully automated adaptive workflow for head and neck radiotherapy treatments with a conventional linear accelerator

Gaia Muti[1]*, Marco M. J. Felisi[1], Angelo F. Monti[1], Chiara Carsana[2], Roberto Pellegrini[3], Edoardo Salmeri[3], Mauro Palazzi[2] and Paola E. Colombo[1]

[1]Medical Physics Department, Azienda Socio Sanitaria Territoriale Grande Ospedale Metropolitano (ASST GOM) Niguarda, Milano, Italy, [2]Radioteraphy Department, ASST GOM Niguarda, Milano, Italy, [3]Elekta AB, Medical Affairs, Stockholm, Sweden

**Introduction:** The objective of this study is to evaluate the performance of an automatic workflow for head-and-neck (H&N) radiotherapy using a multi-atlas based auto-contouring software and an a-priori multicriteria plan optimization algorithm and implement an adaptive online approach with CBCT images. Two different modalities are investigated, the fluence-to-position (FTP) and the adapt-to-shape (ATS) approach.

**Materials and methods:** Nine patients are used for the multi-atlas database. The organs at risk (OARs) of the H&N district and five additional structures (air, fat, tissue, bone and patient's exterior) subsequently used for the creation of the synthetic CT are auto-contoured with the Elekta ADMIRE® software. The mCycle algorithm is used for the a-priori multicriteria plan calculation. A total of twenty H&N patients are selected for this step. The automatic plans are compared to manual VMAT plans by assessing differences in planning time, dose delivered to targets and OARs, and calculating the plan quality indexes (PQIs). Two patients are chosen for the retrospective CBCT adaptive online feasibility analysis. To assess the differences for the two adaptive modalities, the clinical goals for targets and OARs and the number of passed constraints are explored. An analysis of the timing for the different steps is carried out to assess its clinical applicability.

**Result:** The dice of the five HU layer structures range between 0.66 and 0.99. The mCycle auto-planning significantly reduces planning time, from 2 hours to 10 minutes. The radiotherapist deems all plans clinically acceptable, and in the majority of cases the automatic plan is the preference choice. The automatic plans enhance OARs sparing and preserve a good target coverage, this is also confirmed by the PQIs result. Comparing FTP and ATS modes in adaptive radiotherapy, ATS exhibits superior outcomes, mostly in the target coverage. In the FTP techniques target coverage is inadequate and statistically different from the accepted values. In the ATS the results align with the initial approved values. Using the ATS mode the planning time takes around 14 minutes and approximately 20 minutes for the entire treatment.

**Conclusion:** This study contributes to the advancement of automatic and adaptive radiotherapy, demonstrating the potential of an automated workflow in H&N treatments.

KEYWORDS

**auto-planning, MCO, online adaptive radiotherapy, offline adaptive radiotherapy, CBCT, adapt to shape, fluence to position**

# 1 Introduction

Head and neck cancer is a challenging site to treat in terms of contour definition, planning technique and anatomical changes between sessions. Anatomical changes may occur from as early as the first irradiation sessions. Inter-fraction changes, that include shrinkage of the tumor and/or normal tissue, result in target movement in different positions relative to other structures (1).

Adaptive radiation therapy (ART) is a process to control anatomical variations over the treatment course. This provides a day-by-day representation of the patient's anatomy to better delineate the target and the OARs volumes (2). ART can be performed online or offline: both accommodate anatomical changes during treatment, but differ in their implementation and timing of treatment plan adaptations. Offline ART relies on periodic, usually CT, imaging sessions, which are separated from the actual treatment. It focuses on adapting the treatment plan periodically for future sessions. Online ART involves the observation of the patient's anatomy using imaging techniques, such as CBCT or MRI, and then it assesses the anatomical or position changes before the treatment. In this way the treatment plan is continually updated on a daily basis to account for the current anatomical configuration. Online ART aims to enhance treatment accuracy by reducing setup uncertainties and improving target localization. In recent years, the rise of MRI-LINACs and the resulting MRI image-guided radiotherapy (MRgRT) (3) has renewed the interest in the field of adaptive online treatment, which also has led to the investigation of ART with CBCT images to promote its applicability on conventional linear accelerators.

The CBCT online ART workflow begins with a CBCT acquisition, creating new reference images. PTVs and OARs can be propagated from the planning CT onto the current CBCT. For the CBCT planning, their inaccuracy in Hounsfield units (HU) and electron densities could induce a non-negligible dose error (4). For this reason the use of a synthetic CT (sCT) in CBCT planning is an essential point. The densities of each volume are calculated from the initial planning CT and subsequently assigned to the contours propagated on the CBCT, resulting in a sCT. The final step is the creation of a new treatment plan that exactly matches the anatomy of the day.

Using CBCT for online ART is a dynamic and iterative treatment process. It requires fast image acquisition, quick outlining of all relevant OARs and targets, and rapid plan creation (5). This can be accomplished by embedding automated methods such as auto-contouring and auto-planning within the workflow.

Atlas-based auto-segmentation and CT-to-CBCT deformable propagation of OAR contours makes the deformable transfer of original contours defined on the initial planning CT to daily CBCT rapid and practical (6). Auto-planning systems such as knowledge-based (KB) (7), protocol-based automatic iterative (8) and multicriteria optimization (MCO) (9) allow the planning process to be optimized while also reducing the timing.

ART is intended as a technological improvement offering potential gains in therapeutic outcomes and reduced adverse effects. Plan adaptation is related to anatomical, physiological, and positioning changes observed during therapy. In head and neck cancer patients, such changes can drastically affect the dose distribution and hence the associated toxicities.

Weight loss during radiotherapy for head and neck cancers leads to changes in body contour, fat distribution, and soft tissue thickness, affecting treatment positioning and accuracy. Minor positional shifts in bony structures can also occur due to changes in soft tissue support, affecting patient alignment. Muscle atrophy or changes in muscle mass around the head and neck area also influence patient stability and positioning. Changes in airway and esophagus positions are noted as surrounding tissues respond to treatment, leading to mucositis and dysphagia. Parotid glands often shrink and deform due to their proximity to the radiation field, altering their position and increasing the risk of xerostomia. Similar changes can occur in the salivary glands. Lymph nodes and tumor shrinkage, a common response to radiation, necessitate adjustments to ensure adequate dosing of the remaining tumor mass. Daily ART accounts for these factors and represents a substantial advance in personalized cancer care. However, the implementation of ART presents challenges. Frequent imaging and plan adjustments require sophisticated technology and organization, increasing the complexity and cost of treatment. Leveraging clinical workflow to incorporate adaptive processes without significant delays or interruptions in patient care is mandatory. Full clinical implementation of ART for head and neck cancer is still limited and requires improvement in both technology and practice guidelines before it becomes a new standard.

In this work, an automated workflow for H&N radiotherapy, using the available resources at our facility is analyzed. An Atlas

Based Auto Segmentation (ABAS) contouring system is employed for automatic contouring, followed by deformable contour propagation to generate sCT images. Planning is carried out using an MCO *a priori* auto-planning system with a wish-list (WL) for the head and neck region. Subsequently, a proof of concept for an automated workflow leveraging the obtained results is tested for the CBCT online adaptation with conventional linear accelerators. A similar strategy conventionally adopted for MRgRT is replicated for two patients, employing the adapt to shape technique (3). We also evaluate a fluence to position approach with the CBCT. For the CBCTgRT, with a C-Arm linac with a 6 degrees of freedom couch, the fluence to position (FTP) is a couch shift and fluence calculation as opposed to the adapt to position (ATP) in the Unity System where the "virtual couch shift" is implemented. To assess the effectiveness of the workflow, the timing of different phases is considered, the robustness of the WL is evaluated, and the acceptability of a treatment plan is examined applying the two adaptive techniques.

## 2 Materials and methods

The workflow starts with the pre-treatment phase, depicted in green in Figure 1. This phase starts with a CT simulation, during which the OARs are defined using automatic tools, as described in Section 2.1. Following this, the reference treatment plan is generated using the automatic planning software outlined in Section 2.2. Both the target/OAR definition and plan creation are performed with the involvement of the physician and medical physicist, ensuring accuracy and quality control during the use of the automatic software. The daily treatment workflow begins with the acquisition of a CBCT, which provides a new set of reference images. These images are then imported into the treatment planning software, where the contours from the planning CT are propagated onto the current CBCT by means of the Deformable Image Registration and contours projection. These propagated contours must be reviewed and, if necessary, manually corrected by the physician. Once the CBCT is loaded into the treatment planning system, the Adapt Setup and Force ED options are used to apply bulk density correction to the CBCT images. After generating the synthetic CT (sCT) from the CBCT, clinicians have two options, depicted in orange and described in Section 2.3, for the next step: they can either use the Fluence-to-Position method to verify the dose delivered by the reference plan in the patient's current position, or they can use the Adapt-to-Shape option to perform an adaptive replan, leveraging the automatic tools described in Section 2.2. Finally, the patient's position is re-checked, and the selected treatment plan, either the reference plan or an adapted one, is delivered.

### 2.1 Auto-contouring

Although Deep Learning or Atlas Based auto-contouring are used to automatically outline the organs at risk, in daily clinical practice a manual inspection and confirmation is always performed by the Radiation Oncologist prior to allow the optimization phase. ADMIRE® software (ADvanced Medical Imaging Registration Engine, research version 3.37, Elekta AB, Sweden) with random forest (RF) algorithm for the multi atlas-based segmentation on CT images is used. The OARs for the head and neck region considered in this study are: cochleae, mandible bone, larynx, oesophagus, oral cavity, brainstem, optic chiasm, optic nerves, lens, eyes, lachrymal glands, pituitary gland, brain, lips, muscles constrictor, parotids, thyroid gland, trachea, brachial plexus, spinal cord and lungs. Additionally, structures essential to manage CBCT images in the planning process are also introduced. These structures represent different HU layers, divided into external body, air, fat, tissue, and bones. To define these structures for the atlas, the semi-automatic "whole-body" tool available in the MIM® software (MIM Software

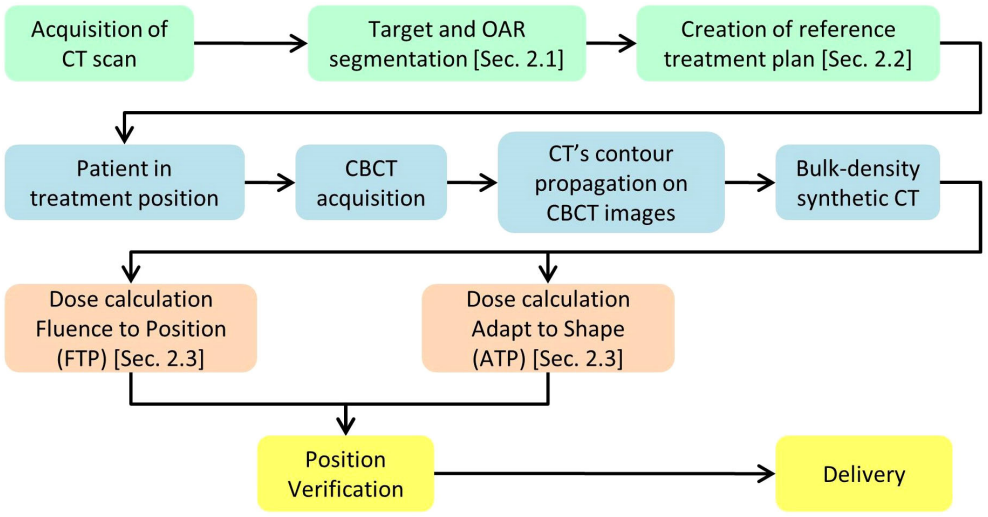



FIGURE 1
Diagram illustrating the key steps of the workflow.

Inc, version 7.2.8, OH) is exploited. The five HU layers are associated with different voxel's HU value. The atlas used for the auto-contouring process contains both the previously listed OAR structures and the HU layers.

To assess the accuracy of the HU layer and the OARS structures generated by the auto-contouring software, the volumetric disparities between semi-automatic contours (ground-truth) and the automatic segmentation of nine patients are measured (Table 1). The analyzed OARs are divided into large-sized structures (volume>15cc) and small-sized structures (volume ≤ 15cc).

For the evaluation, the Dice similarity coefficient (DSC) is considered. To perform the geometrical analyses the Golden Rule software (version 1.2, Canis Lupus LLC, Wisconsin, USA) is employed.

## 2.2 Auto-planning

The mCycle algorithm, recently launched under the name of "ElektaONE AutoPlanning", is used for the a-priori Lexicographic multicriteria plan calculation with a WL for a Simultaneously Integrated Boost (SIB) treatment. mCycle is the Elekta implementation of the iCycle algorithm by Erasmus University and is based on the Lexicographic Approach developed by Sebastian Breedveld (10). The cost functions and the Dose Calculation algorithm are adapted to Monaco. The basis of iCycle is an *a priori* definition of constraints and priority treatment goals. The so-called wish-list can be constructed, goals are optimized sequentially, resulting in a pareto-optimal solution without interactions (11). The overall process is defined as Intelliplan Optimization in the mCycle environment and is summarized in two diagrams included in the Supplementary Material. To exploit the auto-planning process, the WL must be tuned and subsequently validated. For the validation of the WL twenty patients are chosen (Table 1) with a prescribed dose of 70 Gy to the macroscopic tumor and 56 Gy to the nodes in 35 fractions. For each patient a manual and an automatic plan are calculated using the same calculation settings and sequencing parameters. For the validation, the dosimetric quality of the plans and the planning times are analyzed.

Dosimetric quality is assessed both qualitatively and quantitatively. The qualitative assessment is carried out by a physician, based on a blinded dosimetric comparison between plans. Automatic and manual plans are together presented to the physician using identical layouts of plotted dose distributions and dose volume histograms. The physician assigned to each plan a score from 1 to 5 (1-unacceptable; 2-borderline; 3-sufficient; 4-good; 5-excellent) for target coverage, OARs sparing and plan acceptability. The mean value of the physician's score and the percentage of automatic plan choices is then analyzed to determine the effectiveness of automated planning over manual planning. The quantitative assessment of the dosimetric quality is carried out by analyzing the dosimetric objectives for PTV coverage and OARs sparing and introducing a plan quality index (PQI) (12, 13). For the dosimetric scores, the median values distribution (with 1st and 3rd quartile) of each constraint for the two planning modalities is observed.

The PQI defines the overall performance of a plan in an operator independent manner. PQIs for PTVs and OARs are first considered independently and then collectively. The generalized formula for the PQI assessment is shown in Equation 1

$$PQI = \sum w * \frac{D_{x\%}^{goal} - D_{x\%}^{plan}}{D_{x\%}^{plan}} \quad (1)$$

$D_{x\%}$ stands for the dose received by the $x$% of the volume of PTV or OAR, "plan" refers to the dose–volume indexes in the dose plan, "goal" refers to the dose objective and $w$ refers to the weighting factor used as function of clinical relevance of the OAR or PTVs. For the PQI calculation of both PTVs, emphasis is placed on four specific points of the DVH to describe its steepness ($D_{95\%}$, $D_{90\%}$, $D_{50\%}$, $D_{7\%}$) and on the percentage of the volume covered by the 95% of the prescription dose.

The weight $w$ is assessed with the physician on a scale from 1 to 5 to account for the relatively clinical importance assigned by the radiation oncologist team as part of the clinical intent, and it is set equal to 0 for OARs that correspond to the GTVs. Two different templates are defined for the w: one for the NPC cases (wH) and one for the "middle-lower" cases (wL). Table 2 showed the values used for the calculation. The statistical significance ($p < 0.05$) of the dosimetric result is evaluated with a signed-rank Wilcoxon test using Python version 3.10.12.

For the two plan modalities, the time needed for the calculation is measured.

TABLE 1 Patient information for each analysis with cancer diagnosis and staging.

| Auto-contouring (9 patients) | Auto-planning (20 patients) | CBCT Adaptive (2 patients) |
|---|---|---|
| 1 Nasopharynx stage III | 4 Nasopharynx stage III<br>1 Nasopharynx stage IVA | |
| 1 Oropharynx stage I | 2 Oropharynx stage I<br>2 Oropharynx stage IVA | |
| 1 Hypopharynx stage IVB | 1 Hypopharynx stage IVB | 1 Hypopharynx stage IVB |
| 4 Larynx stage III<br>1 Larynx stage IVA | 3 Larynx stage III<br>3 Larynx stage IVA | 1 Larynx stage IVA |
| | 1 Oral cavity stage III | |
| 1 Parotid stage I | 1 Parotid stage IIIB<br>1 Neck stage IVA | |

TABLE 2 OARs weights for the calculation of the PQI.

| OARs | Objective | wH | wL |
|---|---|---|---|
| Bone_Mandible | D0.03cc | 2 | 2 |
| Brachial_Plex | D0.1cc | 2 | 2 |
| Brain | D0.03cc | 3 | 1 |
| | Dmean | 2 | 1 |
| Brainstem | D0.03cc | 5 | 3 |
| Cochlea | D0.03cc | 1 | 1 |
| | Dmean | 4 | 1 |
| Esophagus | D0.03cc | 3 | 3 |
| Eye | D0.03cc | 3 | 1 |
| Glnd_Lacr | D0.03cc | 3 | 1 |
| | Dmean | 2 | 1 |
| Larynx | D0.03cc | 3 | 4 |
| | Dmean | 3 | 4 |
| Lens | D0.03cc | 4 | 1 |
| Lips | D0.03cc | 2 | 3 |
| OpticChiasm | D0.03cc | 5 | 1 |
| OpticNrv | D0.03cc | 5 | 1 |
| Cavity_Oral | V30Gy | 1 | 1 |
| | Dmean | 4 | 4 |
| Parotid | Dmean | 4 | 5 |
| Musc_Constrict | Dmean | 3 | 4 |
| Pituitary | D0.03cc | 4 | 1 |
| SpinalCord | D0.03cc | 5 | 5 |

The use of automated planning with Lexicographic optimization allows a more uniform set of results that are pretty much independent on the planner experience as they rely on the description of the wish-list as a class solution with the personalization both to intra-patient and inter-patient anatomical variations carried on via the Muli-criteria optimization approach. The wish-lists have been created and tweaked to try to achieve the minimum modulation degree able to assure the achievement of the convergence of the optimization (ATS) throughout all the given fractions and the consistency of the Patient Specific Quality Assurance (PSQA) with a clinically acceptable Gamma Index pass rate. All of the above, in order to assure that results can be applied and replicated to the Head&Neck patient class solution.

## 2.3 CBCT adaptive

Once the auto-planning WL is validated, two patients (Table 1) are chosen for a retrospective CBCT adaptive online feasibility analysis: patient A and B. The first seven CBCTs performed on the patients and the first of the following four weeks are selected, for a total of eleven CBCTs. The two modalities of adaptive online investigated are: fluence to position (FTP) and adapt to shape (ATS).

FTP focuses on adjusting the plan isocentre to accommodate daily variations in patient positioning. This mode replicates the current routine clinical practice of treatment delivery where shifts in the x, y, and z coordinates are performed. These shifts are extracted from the R&V (MOSAIQ®, Elekta, Sweden) and applied to the plan isocentres. The dose calculation is performed without further plan optimization, using the same patient-specific template previously saved.

The ATS mode takes into consideration not only changes in patient positioning but also anatomical variations, allowing the adaptation of the treatment plan on the shape of the daily patient's anatomy. For the ATS mode the patient template is calculated and optimized using mCycle without any manual tweaking. The achievements and failures of dose constraints are recorded for the eleven fractions. This assessment involves evaluating the adherence of the adapted plans to the predefined dose constraints established for the OARs and targets. To determine if there are any statistically significant differences between the two adapted plan modalities, the Wilcoxon signed-rank test is performed ($p < 0.05$) using Python version 3.10.12 in a Colab notebook. Any significant differences between the FTP plans and ATS plans are then further investigated. An analysis of the timing for the different steps required to produce an online adaptive plan is also carried out to assess its clinical applicability.

# 3 Result

## 3.1 Auto-contouring

The DSC of the five HU layer structures, large-sized structures and small-sized structures are reported in Table 3. For each layer, the median values and the first (Q1) and third quartile (Q3) are provided.

TABLE 3 Median and interquartile value of the geometric evaluation using DSC index for the HU layer, large-sized and small-sized automatic contours.

| Structures | Air | Bones | Tissue | Fat | External | Large | Small |
|---|---|---|---|---|---|---|---|
| DSC Median | 0.66 | 0.82 | 0.87 | 0.80 | 0.99 | 0.88 | 0.67 |
| DSC [Q1;Q3] | [0.60;0.69] | [0.55;0.89] | [0.85;0.88] | [0.63;0.84] | [0.99;1.00] | [0.83;0.92] | [0.52;0.77] |

TABLE 4 Median physician's scores and percentage of plan preference choices.

| | OARs Sparing | Target Coverage | Plan Acceptability | Plan preference choice |
|---|---|---|---|---|
| Automatic | 5.0 | 4.5 | 5.0 | 65% |
| Manual | 4.5 | 5.0 | 5.0 | 35% |

## 3.2 Auto-planning

To assess the quality of the auto-planning system the physician assigned scores and the percentage blind preference choice are provided in Table 4.

The quantitative dosimetric quality evaluation involves the analysis of the dosimetric score cards and the PQI. Table 5 presents the median values (with Q1 and Q3 in square brackets) of each constraint in both automatic and manual plans. Table 6 shows the PQI values calculated for each plan. The Wilcoxon test p-value results are provided in Tables 5 and 6. Significant p-values are in bold and the best average constraint achieved value is highlighted in gray.

Planning times are drastically reduced through the auto-planning system. Figure 2 shows the time difference between the two planning modalities. Automatic planning reaches an average time of eleven minutes, while manual planning is about two hours.

## 3.3 CBCT adaptive

The number of passed clinical goal for the two modalities are displayed in Table 7, only the constraints that showed a different number between the two modalities are reported. Following the Wilcoxon test to compare the two modes of adaptive planning, the

TABLE 5 Constraint median value results for automatic and manual plans, with Q1 and Q3 in square brackets and Wilcoxon test p-value result.

| Structure | Constraint | Optional | Mandatory | Automatic | Manual | p |
|---|---|---|---|---|---|---|
| $PTV_1$ | $V_{95\%}$ | 98% | 95% | 97.5 [97.0-98.2] | 97.6 [96.9-98.9] | 0.41 |
| $PTV_2$ | $V_{95\%}$ | 98% | 95% | 99.1 [98.8-99.4] | 98.9 [98.7-99.6] | 0.65 |
| $PTV_{2\text{-}1}$ | $V_{95\%}$ | 98% | 95% | 98.7 [98.3-99.1] | 98.6 [98.1-99.4] | 0.65 |
| $PTV_1$ | $D_{50\%}$ | | 70 | 70 [70-70] | 70 [70-70] | 0.41 |
| $PTV_1$ | $D_{7\%}$ | | 73.5 | 72.1 [71.8-72.2] | 71.4 [71.1-71.7] | **<0.01** |
| $PTV_{2\text{-}1}$ | $D_{7\%}$ | 66 | | 66.0 [65.1-66.6] | 66.2 [65.1-66.7] | **0.02** |
| Bone_Mandible | $D_{0.03cc}$ | 70 | 73.5 | 68.6 [58.3-70.2] | 68.5 [55.5-70.5] | 0.33 |
| Brachial_Plex_L | $D_{0.1cc}$ | 60 | 66 | 60.0 [57.8-64.6] | 59.2 [57.4-64.4] | 0.97 |
| Brachial_Plex_R | $D_{0.1cc}$ | 60 | 66 | 58.9 [57.8-61.4] | 57.6 [57.2-60.0] | 0.13 |
| Brain | $D_{0.03cc}$ | 72 | | 34.2 [13.6-50.5] | 30.4 [9.0-50.1] | **0.02** |
| Brain | $D_{mean}$ | 30 | | 1.3 [0.6-4.5] | 1.5 [0.6-4.2] | 0.23 |
| Brainstem | $D_{0.03cc}$ | 54 | 55 | 26.7 [11.1-38.3] | 22.3 [5.9-31.0] | **0.01** |
| Cochlea_L | $D_{0.03cc}$ | 60 | | 1.9 [1.3-12.9] | 2.2 [1.4-16.7] | **0.02** |
| Cochlea_L | $D_{mean}$ | 45 | | 1.7 [1.1-10] | 2.0 [1.3-15.3] | **0.02** |
| Cochlea_R | $D_{0.03cc}$ | 60 | | 2.0 [1.2-23] | 2.5 [1.4-27.3] | **<0.01** |
| Cochlea_R | $D_{mean}$ | 45 | | 1.7 [1.1-18.4] | 2.1 [1.3-24.5] | **<0.01** |
| Esophagus | $D_{0.03cc}$ | 45 | 55 | 52.4 [49.4-53.8] | 53.1 [48.1-56.9] | 0.18 |
| Eye_L | $D_{0.03cc}$ | 40 | 45 | 1.3 [0.9-3] | 1.6 [0.9-4.1] | **<0.01** |
| Eye_R | $D_{0.03cc}$ | 40 | 45 | 1.3 [0.9-3] | 1.7 [1.0-4.4] | **<0.01** |
| Glnd_Lacr_L | $D_{0.03cc}$ | 40 | | 0.7 [0.6-1.9] | 0.8 [0.6-2.1] | **0.01** |
| Glnd_Lacr_L | $D_{mean}$ | 26 | | 0.6 [0.5-1.3] | 0.7 [0.5-1.7] | **<0.01** |
| Glnd_Lacr_R | $D_{0.03cc}$ | 40 | | 0.8 [0.6-1.8] | 0.9 [0.6-1.9] | **<0.01** |

*(Continued)*

TABLE 5 Continued

| Structure | Constraint | Optional | Mandatory | Automatic | Manual | p |
|---|---|---|---|---|---|---|
| Glnd_Lacr_R | $D_{mean}$ | 26 | | 0.7 [0.5-1.5] | 0.8 [0.5-1.6] | **<0.01** |
| Larynx | $D_{0.03cc}$ | | 66 | 68.2 [52.9-70.6] | 64.9 [54.9-72.1] | 0.15 |
| Larynx | $D_{mean}$ | 44 | 50 | 24.2 [20.3-58.7] | 38.8 [37.1-63.8] | **<0.01** |
| Lens_L | $D_{0.03cc}$ | 4 | 10 | 0.8 [0.6-1.4] | 0.9 [0.7-1.9] | **<0.01** |
| Lens_R | $D_{0.03cc}$ | 4 | 10 | 0.8 [0.6-1.7] | 1.0 [0.7-2.2] | **<0.01** |
| Lips | $D_{mean}$ | 30 | 50 | 15.4 [8.3-23.5] | 25.8 [14.5-30.1] | **<0.01** |
| OpticChiasm | $D_{0.03cc}$ | | 55 | 1.2 [0.9-5.2] | 1.3 [0.9-4.8] | 0.94 |
| OpticNrv_L | $D_{0.03cc}$ | | 55 | 1.1 [0.8-4.6] | 1.2 [0.9-4.6] | 0.28 |
| OpticNrv_R | $D_{0.03cc}$ | | 55 | 1.1 [0.8-4.2] | 1.2 [0.9-3.9] | **0.04** |
| Cavity_Oral | $V_{30Gy}$ | 73% | | 45.7 [24.9-73.7] | 73.2 [42.5-99.6] | **<0.01** |
| Cavity_Oral | $D_{mean}$ | 30 | 45 | 30.8 [20.0-43.1] | 41.7 [29.6-50.0] | **<0.01** |
| Parotid_L | $D_{mean}$ | 20 | 25 | 19.8 [17.1-25.2] | 21.7 [15.8-26.6] | 0.65 |
| Parotid_R | $D_{mean}$ | 20 | 25 | 19.4 [16.6-20.9] | 19.3 [14.3-24.0] | 0.97 |
| Musc_Constrict | $D_{mean}$ | 35 | 50 | 51.9 [46.0-55.5] | 56.7 [50.0-58.4] | **<0.01** |
| Pituitary | $D_{0.03cc}$ | | 50 | 1.2 [0.9-7.2] | 1.3 [1.0-6.9] | 0.46 |
| SpinalCord | $D_{0.03cc}$ | 45 | 50 | 29.6 [27.5-30.4] | 25.1 [24.2-31.1] | **<0.01** |

Dose values are reported in Gy. Optional and Mandatory clinical goals are also reported in the table. Significant p-values are written in bold and the better average constraint achieved value is highlighted in gray.

results showed significant differences in the requests for targets and certain OARs. Figure 3 shows the differences in percentage by coverage and hot-spots of the targets in the two modes for the selected patients. Figure 4 shows the dose differences for the mean dose constraints, when the results are significantly different in the two modalities. Figure 5 shows the dose differences for the maximum dose constraints, when results are significantly different in the two modalities. For an evaluation of the clinical feasibility in implementing an adaptive online workflow, the time required for each step of the process is recorded. Table 8 presents the averages times and their standard deviations obtained for the two adaptive modalities. It includes the estimated times for CBCT acquisitions, both for the pre-treatment scan, currently in use with a standard acquisition protocol, and for the "optional" position verification second scan, performed with a fast protocol. The table also presents the total time, offering an esteem of the time required for the two adaptive modalities.

TABLE 6 PQI median values (OARs, PTVs and total) for automatic and manual plans, with Q1 and Q3 in square brackets and Wilcoxon test p-value result.

| | Automatic | Manual | p |
|---|---|---|---|
| PQI_OARs | 0.35 [0.24-0.50] | 0.32 [0.22-0.46] | **0.01** |
| PQI_PTVs | 0.31 [0.29-0.35] | 0.33 [0.31-0.38] | **0.02** |
| PQI_total | 0.68 [0.54-0.83] | 0.67 [0.52-0.79] | 0.32 |

Significant p-value are written in bold and the better average constraint achieved value is highlighted in gray.

# 4 Discussion

In this study, we demonstrated the feasibility of implementing an adaptive online treatment without dedicated systems, but leveraging a combination of available automated software. Our project is the first to leverage this combination of automated software (auto-contouring, auto-planning, sCT creation) on conventional accelerators in an Elekta environment.

## 4.1 Auto-contouring

The software used for the OARs auto-contouring was already validated for clinical application and our results are in agreement with the literature (14, 15). However, no results for the auto-contouring of HU layers are obtained on CT images. Instead, DSC values from MRI images are available (16, 17). Our results for bones, tissue, fat and external reach a DSC median value of over 0.8, while for air the DSC median value is 0.66. The external and bones results are comparable to those reported by Guerreiro (16), and for air, fat, and tissue to those reported by Hsu (17). The auto-contouring process shows promising results and strongly facilitates the creation of sCTs necessary for the following adaptive process.

## 4.2 Auto-planning

The WL definition for the mCycle auto-planning involves a precise iterative tuning process, which may take several days.

10.3389/fonc.2025.1382537

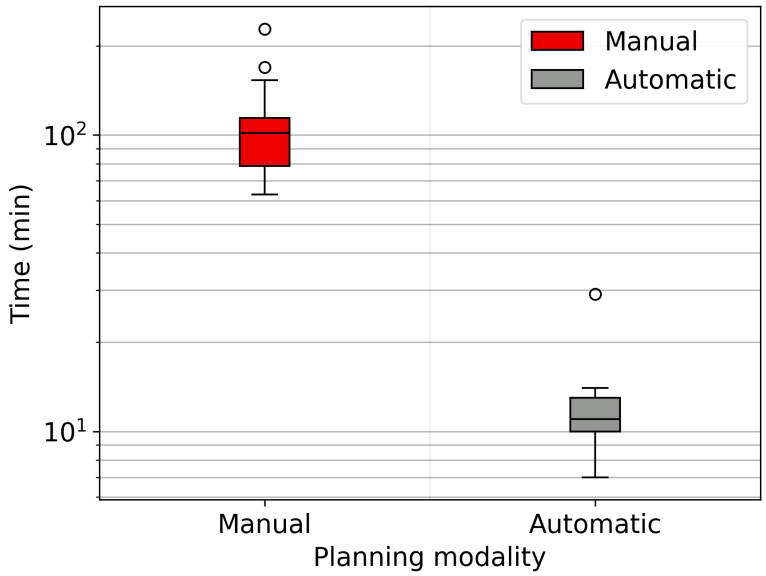



FIGURE 2
Boxplot showing the time spent for the manual and automatic planning, semilog scale graph.

Familiarity with planning in Monaco certainly accelerates the initial steps of this process and the time needed to develop a robust WL is influenced by the level of detail in the defined protocol. When the protocol is more explicit and detailed, mCycle finds easier to get the expected results efficiently. In our case, we took this opportunity to update the clinical constraints commonly used for H&N cases. After the physician's review of all constraints, the first WL is defined and it is then tested and modified until an optimal solution is obtained for all cases. In the qualitative and quantitative dosimetric evaluation of the automatic treatment plans, favorable outcomes have been achieved. The physician deems all plans clinically acceptable, and in the majority of cases (65%), the automatic plan is chosen over the manual one, as shown in Table 4. The automatic plans are preferred by the physician due to their optimal compromise between PTVs coverage and OARs sparing. This finding aligns with the data presented in Table 6. The PQI_OARs value is higher for the automatic plans, indicating its ability to achieve more sparing of the critical organs. This result suggests that the automatic plans offer greater overall sparing of the OARs at the cost of slightly reducing coverage of the PTVs. However Table 5 shows that the requests for PTVs coverage are always well fulfilled in the automatic plans. mCycle auto-planning, enables greater OARs sparing while meeting the target coverage requirements set in the WL. This outcome is consistent with the literature findings on *a priori* MCO auto-planning (9, 10, 13). Looking at the achieved result, we noticed that the constraints of constrictor muscles and larynx are within those modified in the upgrade of our hospital's protocol. Therefore, from the standpoint of the manual planner, there has been an initial response to address these new requirements, and over time better results could be reached. This challenge is not encountered in the investigated auto-planning system, where modifying a clinical protocol involves the adjustment of a parameter in the WL that is automatically optimized during the calculation. The WL approach allows for easy modification as needed, making it highly adaptable to meet specific new requirements or clinical needs. This cannot be asserted for all auto-planning systems, for example, the KB automatic plans rely heavily on the manual plans and the protocol used up until that point (7). The mCycle auto-planning allows real-time adjustments, supporting, rather than replacing, the role of the medical physics expert to achieve optimized results in a short time. The time required for planning is shown in Figure 2, the automatic modality significantly reduces the planning time from about 2 hours in manual mode to about 10 minutes in the automatic mode. The time-saving advantage of auto-planning over manual

TABLE 7 Number of passed clinical objectives for the mandatory and [optional] constraints in the two adaptive modes, for patients A and B out of the 11 CBCT fractions analysed.

| Patient A | | | |
|---|---|---|---|
| Structure | Constraint | FTP pass Mandatory [Optional] | ATS pass Mandatory [Optional] |
| $PTV_1$ | $V_{95\%}$ | 1 [0] | 11 [1] |
| $PTV_2$ | $V_{95\%}$ | 10 [0] | 11 [11] |
| $PTV_{2\text{-}1}$ | $V_{95\%}$ | 7 [0] | 11 [11] |
| PTV1 | $D_{50\%}$ | 10 | 11 |
| Brachial_Plex_L | $D_{0.1cc}$ | 10 [0] | 11 [0] |
| Esophagus | $D_{0.03cc}$ | 0 [0] | 2 [0] |
| Cavity_Oral | $D_{mean}$ | 11 [10] | 11 [11] |
| Parotid_L | $D_{mean}$ | 10 [7] | 11 [7] |
| Parotid_R | $D_{mean}$ | 8 [2] | 11 [11] |

10.3389/fonc.2025.1382537

| Patient B | | | |
|---|---|---|---|
| Structure | Constraint | FTP pass Mandatory [Optional] | ATS pass Mandatory [Optional] |
| $PTV_1$ | $V_{95\%}$ | 3 [0] | 11 [11] |
| $PTV_2$ | $V_{95\%}$ | 2 [0] | 11 [11] |
| $PTV_{2\text{-}1}$ | $V_{95\%}$ | 0 [0] | 11 [11] |
| $PTV_1$ | $D_{50\%}$ | 8 | 11 |
| PTV1 | $D_{7\%}$ | 10 | 11 |
| Bone_Mandible | $D_{0.03cc}$ | 11 [10] | 11 [0] |
| Brachial_Plex_ L | $D_{0.1cc}$ | 11 [8] | 11 [10] |
| BrachialPlex_R | $D_{0.1cc}$ | 2 [0] | 11 [0] |
| Cavity_Oral | $D_{mean}$ | 11 [2] | 11 [4] |
| Parotid_L | $D_{mean}$ | 6 [0] | 11 [7] |
| Parotid_R | $D_{mean}$ | 11 [6] | 11 [11] |

planning is also highlighted in different studies. Focusing on those related to head and neck cases, one study compares the automatic planning times of *a posteriori* MCO, of a protocol-based automatic iterative approach and of a KB methods, resulting in 31 ± 4 minutes, 83 ± 10 minutes, and 27 ± 4 minutes, respectively (18). In other studies focused on KB, planning times of around 30 minutes (19) and 60 minutes (20) are obtained. For the protocol-based automatic iterative approach, planning times exceeding one hour are also reported (20, 21). For *a priori* MCO, one study reports calculation times of around 60 minutes (13).

In our case, the time required for manual planning aligns with the above-cited publications (around 2 hours per patient's plan), while for automatic planning, there is a significant reduction not only compared to all other auto-planning systems, but also when using the same auto-planning system with a different WL. Reducing planning time through auto-planning systems (Figure 2) allows adaptive online sessions to be implemented.

## 4.3 CBCT adaptive

The starting point is the audit of the feasibility of all steps in the workflow, analyzing the difficulties and proposing potential solutions for future clinical applications. The auto-contouring of different layers (anatomical structures and HU layers) facilitates the creation of sCTs, which are made using bulk density override. The quality of CT-adapted contours on daily CBCTs is validated, and reported in literature (6). The time required for the physician to check these contours is not considered in this project. However, the contours propagated are the auto-contoured ones which have already been reviewed and corrected by the physician, thereby errors such as missing or major inaccurate contours that typically require extensive revision, are reduced.

The analysis of the two adaptive modalities "Fluence To Position" (FTP) and "Adapt To Shape" (ATS) reveals a better correspondence of the ATS plans with the reference CT plans compared to the FTP ones. The number of passed clinical goal, reported in Table 7, shows that FTP plans fail to meet the target coverage and also some OARs clinical constraints.

Among the constraints used, the same ones shown in Table 5, there are significant p-values for target coverage, hot-spots and some OARs objectives. The percentage difference for target coverage and hot-spots between the value obtained in each fraction and the reference CT is shown in Figure 3 for both modalities. Moreover, the dose difference with a significant p-value of the mean and max dose objectives for patient A and B

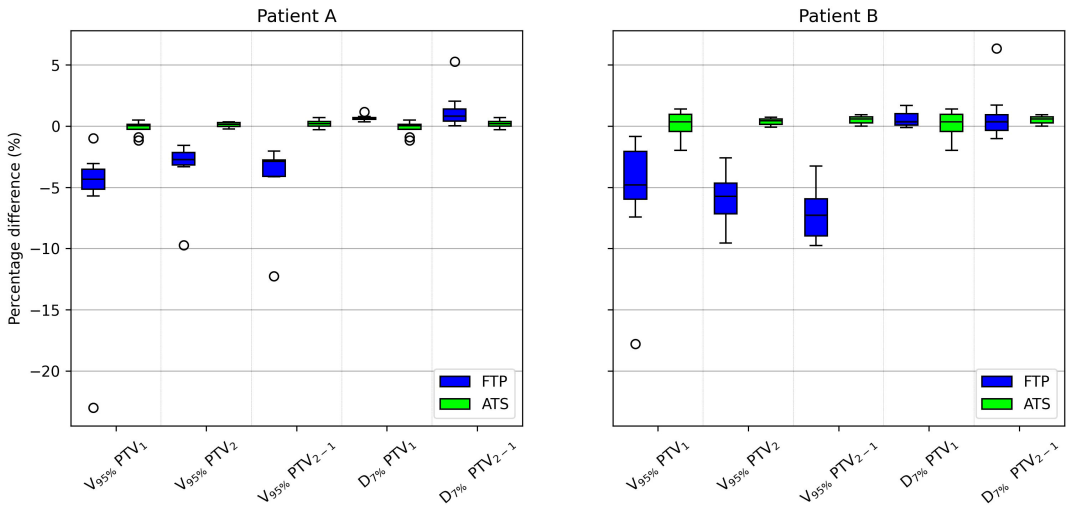



FIGURE 3
Boxplot showing the $PTV_1$, $PTV_2$ and $PTV_{2-1}$ percentage difference coverage and $PTV_1$ and $PTV_{2-1}$ percentage difference hot-spots between the scheduled plan and the FTP plans and between the scheduled plan and the ATS plans, for patient A and B In the boxplot, the inner line denotes the median value, the box the interquartile range and the whiskers the minimum and maximum value excluding the outliers that are presented as single markers.

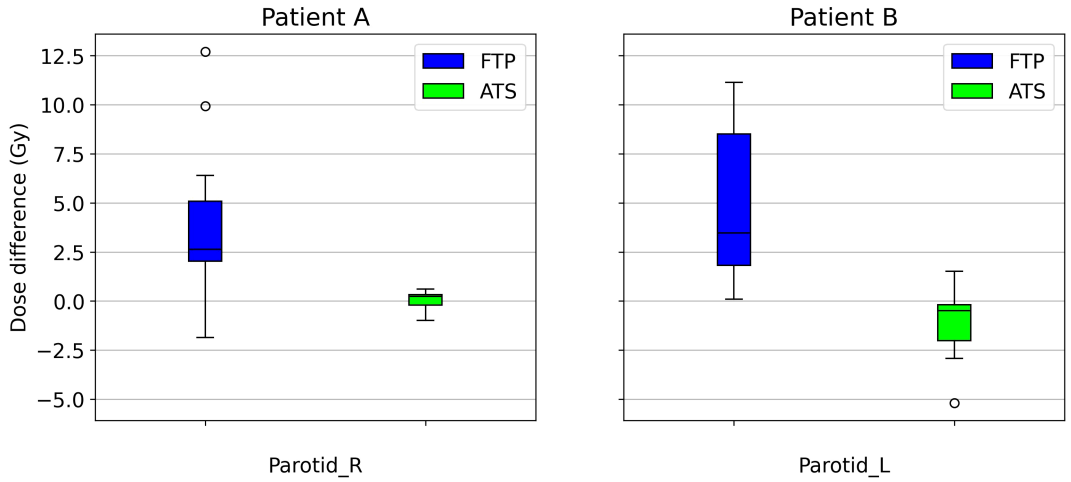



FIGURE 4
Boxplot showing the mean dose difference between the scheduled plan and the FTP plans and between the scheduled plan and the ATS plans, for patient A and B, only for OARs that are significantly different.

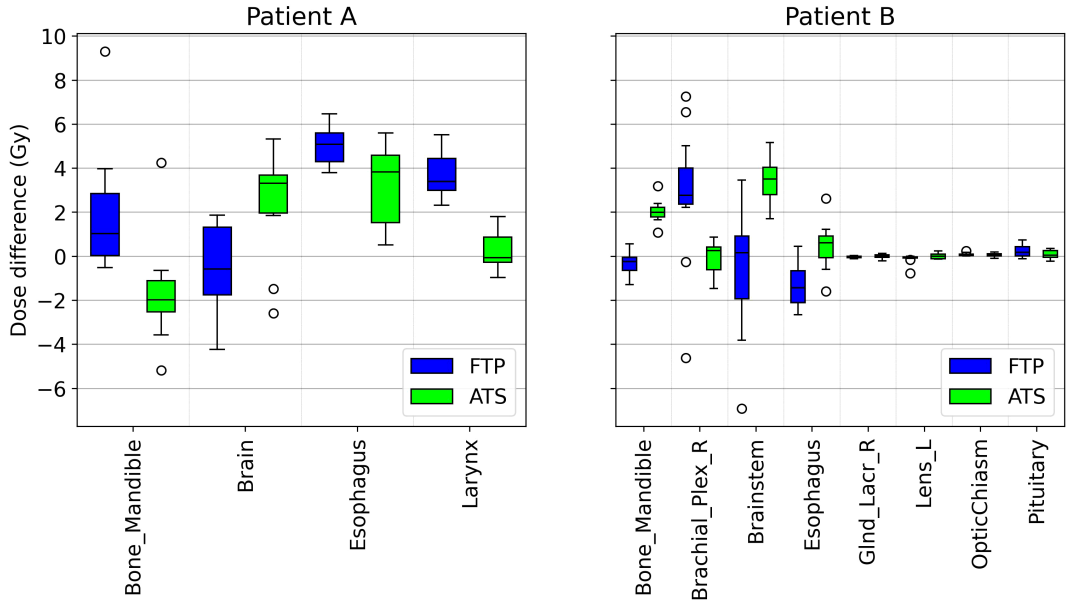



FIGURE 5
Boxplot showing the max dose difference between the scheduled plan and the FTP plans and between the scheduled plan and the ATS plans, for patient A and B, only for OARs that are significantly different.

TABLE 8 Average times ± SD (when available) for each step and total time of the FTP and ATS workflow for both patients.

| | | Pretreatment CBCT (mm:ss) | Adapt anatomy tool-contour (mm:ss) | Replanning time (mm:ss) | Position verification (mm:ss) | Dose delivery (mm:ss) | Total time (min) |
|---|---|---|---|---|---|---|---|
| Patient A | FTP | 01:10 | 01:06 ± 00:04 | 01:32 ± 00:02 | 00:40 | 04:10 | ≈9 |
| | ATS | | | 11:06 ± 00:44 | | | ≈18 |
| Patient B | FTP | 01:10 | 01:05 ± 00:04 | 01:37 ± 00:03 | 00:40 | 04:27 | ≈9 |
| | ATS | | | 13:33 ± 01:05 | | | ≈21 |

are reported respectively in Figures 4 and 5. From the PTV graphs, the inadequate coverage of targets in FTP modality is further highlighted, whereas for ATS, it is optimized in each fraction. Regarding the parotids, which in both patients are close to the PTV, an increase in the mean dose objective is observed in FTP modality, while for ATS, it remains constant. For maximum dose values, both increases and reductions are observed in both modalities. The variation in maximum dose delivered to OARs did not affect the acceptability of the plans. However, the low target coverage achieved in FTP mode negatively impacts their acceptability.

Analysis of the two adaptive modes, FTP and ATS, shows that re-optimizing the plan based on the images acquired at the beginning of the treatment session yields superior results, particularly in enhancing PTV coverage and optimizing sparing of OARs, especially those in close proximity to the target.

Adaptive radiotherapy requires the patient to maintain the treatment position at the linac until the adaptive process is completed. This aspect makes it unsuitable for manual contouring-planning methods. Auto-planning enables the calculation of the treatment plan based on the daily anatomy while maintaining the time limited. For the two analyzed patients, the ATS re-planning time is less than 14 minutes (Table 8), which is also supported by the automatic calculation time results in Figure 2. The estimated workflow total time for online ATS treatment is approximately 20 minutes, which is aligned with data reported in the literature (5, 22). Furthermore, these processing times can be reduced by leveraging more powerful computing systems (e.g. Graphic Processing Units, GPUs). The time required for the clinician to check the contours quality must also be considered in the evaluation of the timing, but in our case this is not assessed. The use of CBCT images for adaptive planning is a topic that has been explored for several years now, exploiting both the use of average structure density override (23, 24), the use of deformed CT images on CBCT (25, 26), and the establishment of Hounsfield numbers versus densities curves (27). The innovation lies in the online implementation of this technique. Currently, this topic is gaining considerable interest, especially due to the introduction of systems that facilitate, automate, and expedite the various steps involved in this process [e.g. ETHOS, Varian Medical Systems (22, 28, 29)].

## 5 Conclusion

This project has undertaken an in-depth analysis of an automated workflow for H&N radiotherapy, using the available resources at our facility.

The auto-contouring approach facilitates the creation of sCTs necessary for the adaptive process. The mCycle auto-planning significantly reduces planning time while maintaining or improving clinical acceptability. In the feasibility study of online adaptive radiotherapy, the ATS mode, which optimizes treatment based on daily anatomy, demonstrates superior outcomes compared to the FTP mode.

Time is a pivotal factor in online adaptive approaches, as all must be managed within the context of a single treatment session. ATS demonstrates efficiency in terms of time, with an estimated total treatment time of about 20 minutes. This outcome marks a preliminary step toward clinical implementation.

The analysis of adaptive methods is focused on feasibility and preliminary evaluation, but the potential of this automated workflow to improve the clinical practice and the patient outcomes remains significant. Further investigations into online adaptive approaches, including retrospective clinical studies with a larger cohort and prospective studies, will contribute to unveiling the complete spectrum of benefits and limitations, particularly regarding the choice between adaptation strategies for each fraction, which remains a critical consideration for future optimization.

This study contributes to the advancement of automatic and adaptive radiotherapy, demonstrating the potential of an automated workflow in challenging cases, such as H&N treatments. The successful validation of auto-contouring and auto-planning software, combined with preliminary findings on online adaptive, underscores the significance of exploiting technology to optimize treatment and improve care for radiotherapy patients.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

## Author contributions

GM: Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing – original draft. MF: Investigation, Writing – review & editing. AM: Validation, Writing – review & editing. CC: Investigation, Writing – review & editing. RP: Conceptualization, Methodology, Writing – review & editing. ES: Data curation, Software, Writing – review & editing. MP: Resources, Writing – review & editing. PC: Project administration, Supervision, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

RP serves as Director of Clinical Science at Elekta, Ing. ES serves as Clinical Science Engineer at Elekta.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2025.1382537/full#supplementary-material

## References


1. Buciuman N, Marcu LG. Adaptive radiotherapy in head and neck cancer using volumetric modulated arc therapy. *J Personalized Med*. (2022) 12:668. doi: 10.3390/jpm12050668

2. Green OL, Henke LE, Hugo GD. Practical clinical workflows for online and offline adaptive radiation therapy. *Semin Radiat Oncol*. (2019) 29:219–27. doi: 10.1016/j.semradonc.2019.02.004

3. Boeke S, Mönnich D, van Timmeren JE, Balermpas P. Mr-guided radiotherapy for head and neck cancer: Current developments, perspectives, and challenges. *Front Oncol*. (2021) 11:616156. doi: 10.3389/fonc.2021.616156

4. Rong Y, Smilowitz J, Tewatia D, Tomé WA, Paliwal B. Dose calculation on kv cone beam ct images: An investigation of the hu-density conversion stability and dose accuracy using the site-specific calibration. *Med Dosimetry*. (2010) 35:195–207. doi: 10.1016/j.meddos.2009.06.001

5. Yoon SW, Lin H, Alonso-Basanta M, Anderson N, Apinorasethkul O, Cooper K, et al. Initial evaluation of a novel cone-beam ct-based semi-automated online adaptive radiotherapy system for head and neck cancer treatment – a timing and automation quality study. *Cureus*. (2020) 12:8. doi: 10.7759/cureus.9660

6. Liu Q, Qin A, Liang J, Yan D. Evaluation of atlas-based auto-segmentation and deformable propagation of organs-at-risk for head-and-neck adaptive radiotherapy. *Recent Patents Topics Imaging*. (2016) 5:79–87. doi: 10.2174/2451827105999160415123925

7. Momin S, Fu Y, Lei Y, Roper J, Bradley JD, Curran WJ, et al. Knowledge-based radiation treatment planning: A data-driven method survey. *J Appl Clin Med Phys*. (2021) 22:16–44. doi: 10.1002/acm2.13337

8. Cilla S, Ianiro A, Romano C, Deodato F, Macchia G, Buwenge M, et al. Template-based automation of treatment planning in advanced radiotherapy: a comprehensive dosimetric and clinical evaluation. *Sci Rep*. (2020) 10:423. doi: 10.1038/s41598-019-56966-y

9. Trivellato S, Caricato P, Pellegrini R, Montanari G, Daniotti MC, Bordigoni B, et al. Comprehensive dosimetric and clinical evaluation of lexicographic optimization-based planning for cervical cancer. *Front Oncol*. (2022) 12:1041839. doi: 10.3389/fonc.2022.1041839

10. Breedveld S. Po-0848: Towards automated treatment planning in radiotherapy. *Radiotherapy Oncol*. (2014) 111:S80. doi: 10.1016/s0167-8140(15)30966-x

11. Breedveld S. Towards automated treatment planning in radiotherapy. *Erasmus MC*. (2013). doi: 10.1016/S0167-8140(15)30966-X

12. Jornet N, Carrasco P, Beltrán M, Calvo JF, Escudé L, Hernández V, et al. Multicentre validation of imrt pre-treatment verification: Comparison of in-house and external audit. *Radiotherapy Oncol*. (2014) 112:381–8. doi: 10.1016/j.radonc.2014.06.016

13. Biston MC, Costea M, Gassa F, Serre AA, Voet P, Larson R, et al. Evaluation of fully automated a *priori* mco treatment planning in vmat for head-and-neck cancer. *Physica Med*. (2021) 87:31–8. doi: 10.1016/j.ejmp.2021.05.037

14. Costea M, Zlate A, Durand M, Baudier T, Grégoire V, Sarrut D, et al. Comparison of atlas-based and deep learning methods for organs at risk delineation on head-and-neck ct images using an automated treatment planning system. *Radiotherapy Oncol*. (2022) 177:61–70. doi: 10.1016/j.radonc.2022

15. Robert C, Munoz A, Moreau D, Mazurier J, Sidorski G, Gasnier A, et al. Clinical implementation of deep-learning based auto-contouring tools–experience of three french radiotherapy centers. *Cancer/Radiothérapie*. (2021) 25:607–16. doi: 10.1016/j.canrad.2021.06.023

16. Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, et al. Evaluation of a multi-atlas ct synthesis approach for mri-only radiotherapy treatment planning. *Physica Med*. (2017) 35:7–17. doi: 10.1016/j.ejmp.2017.02.017

17. Hsu S, DuPre P, Peng Q. Tomé WA. A technique to generate synthetic ct from mri for abdominal radiotherapy. *J Appl Clin Med Phys*. (2020) 21:136–43. doi: 10.1002/acm2.12816

18. Krayenbuehl J, Zamburlini M, Ghandour S, Pachoud M, Tanadini-Lang S, Tol J, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. *Radiat Oncol*. (2018) 13:1–8. doi: 10.1186/s13014-018-1113-z

19. Hu J, Liu B, Xie W, Zhu J, Yu X, Gu H, et al. Quantitative comparison of knowledge-based and manual intensity modulated radiation therapy planning for nasopharyngeal carcinoma. *Front Oncol*. (2021) 10:551763. doi: 10.3389/fonc.2020.551763

20. Chang A, Cheung F, Hung W, Lee M, Ng W. Comparison of planning quality and efficiency between conventional and knowledge-based algorithms in nasopharyngeal cancer patients using intensity modulated radiation therapy. *Int J Radiat OncologyBiologyPhysics*. (2015) 93:E320. doi: 10.1016/j.ijrobp.2015.07.1362

21. Kusters JMAM, Bzdusek K, Kumar P, van Kollenburg PGM, Kunze-Busch MC, Wendling M, et al. Automated imrt planning in pinnacle: A study in head-and-neck cancer. *Strahlentherapie und Onkologie*. (2017) 193:1031–8. doi: 10.1007/s00066-017-1187-9

22. Stanley DN, Harms J, Pogue JA, Belliveau J, Marcrom SR, McDonald AM, et al. A roadmap for implementation of kv-cbct online adaptive radiation therapy and initial first year experiences. *J Appl Clin Med Phys*. (2023) 24:e13961. doi: 10.1002/acm2.13961

23. Fotina I, Hopfgartner J, Stock M, Steininger T, Lütgendorf-Caucig C, Georg D. Feasibility of cbct-based dose calculation: Comparative analysis of hu adjustment techniques. *Radiotherapy Oncol*. (2012) 104:249–56. doi: 10.1016/j.radonc.2012.06.007

24. Wu Q, Chi Y, Chen PY, Krauss DJ, Yan D, Martinez A. Adaptive replanning strategies accounting for shrinkage in head and neck imrt. *Int J Radiat OncologyBiologyPhysics*. (2009) 75:924–32. doi: 10.1016/j.ijrobp.2009.04.047

25. Veiga C, McClelland J, Moinuddin S, Lourenço A, Ricketts K, Annkah J, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using ct-to-cbct deformable registration for "dose of the day" calculations. *Med Phys*. (2014) 41:031703. doi: 10.1118/1.4864240

26. Moteabbed M, Sharp GC, Wang Y, Trofimov A, Efstathiou JA, Lu H. Validation of a deformable image registration technique for cone beam ct-based dose verification. *Med Phys*. (2015) 42:196–205. doi: 10.1118/1.4903292

27. Yang Y, Schreibmann E, Li T, Wang C, Xing L. Evaluation of on-board kv cone beam ct (cbct)-based dose calculation. *Phys Med Biol*. (2007) 52:685–705. doi: 10.1088/0031-9155/52/3/011

28. Yock AD, Ahmed M, Ayala-Peacock D, Chakravarthy AB, Price M. Initial analysis of the dosimetric benefit and clinical resource cost of cbct-based online adaptive radiotherapy for patients with cancers of the cervix or rectum. *J Appl Clin Med Phys*. (2021) 22:210–21. doi: 10.1002/acm2.13425

29. Lemus OMD, Tanny S, Cummings M, Webster M, Wancura J, Jung H, et al. Influence of air mapping errors on the dosimetric accuracy of prostate cbct-guided online adaptive radiation therapy. *J Appl Clin Med Phys*. (2023) 24:e14057. doi: 10.1002/acm2.14057

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 January 2025
DOI 10.3389/fonc.2024.1428329



OPEN ACCESS

EDITED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

REVIEWED BY
Bao-Tian Huang,
Shantou University, China
Li Chen,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Huaqu Zeng
756660752@qq.com

RECEIVED 06 May 2024
ACCEPTED 23 December 2024
PUBLISHED 23 January 2025

CITATION
Zeng H, Zhong M, Chen Z, Tang S and Wen Z (2025) A dosimetric comparison of non-coplanar volumetric modulated arc therapy and non-coplanar fixed field intensity modulated radiation therapy in hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost for brain metastases. *Front. Oncol.* 14:1428329. doi: 10.3389/fonc.2024.1428329



# A dosimetric comparison of non-coplanar volumetric modulated arc therapy and non-coplanar fixed field intensity modulated radiation therapy in hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost for brain metastases

Huaqu Zeng[1]*, MinZhi Zhong[2], Zongyou Chen[1], Shukui Tang[1] and Zunbei Wen[1]

[1]Radiotherapy Center, Gaozhou People's Hospital, Gaozhou, China, [2]Department of Radiology, Guangzhou Red Cross Hospital, Guangzhou, China

**Objective:** The aim of this study was to investigate the dosimetric differences between non-coplanar volumetric modulated arc therapy (VMAT) and non-coplanar fixed-field intensity-modulated radiotherapy (IMRT) in hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost (HA-WBRT+SIB) for brain metastases using the Monaco treatment planning system (TPS).

**Method:** A total of 22 patients with brain metastases were retrospectively enrolled. Two radiotherapy treatment plans were designed for each patient: non-coplanar VMAT and non-coplanar fixed field IMRT. The dose distribution of targets and organs at risk (OAR), the number of monitor units (MUs), and pre-treatment plan verification were compared between the two plans while meeting the prescribed dose requirements of the target volume.

**Results:** There were no significant differences in $V_{50}$, $V_{55}$, $D_{max}$, heterogeneity index (HI) and conformity index (CI) of target PGTV between the two plans ($p>0.05$). For PTV-brain-SIB, there was no significant difference in $D_{98\%}$ between IMRT and VMAT ($p=0.103$). VMAT significantly improved the $V_{30}$ of PTV-brain-SIB ($p<0.001$), decreased HI ($p=0.003$), and increased CI ($p<0.001$). There were no significant differences in the $D_{max}$ to the brain stem, left and right lens, optic chiasm, pituitary gland, and left and right hippocampus between the two plans ($p>0.05$). Compared with IMRT, VMAT significantly reduced the $D_{max}$ to the left and right eyes ($p<0.001$) and significantly increased the $D_{max}$ to the right inner ear ($p=0.010$). There was no significant difference in the $D_{max}$ to the left inner ear between VMAT and IMRT ($p=0.458$). Compared with IMRT, VMAT significantly reduced the $D_{max}$ to the left

optic nerve ($p$=0.006), but significantly increased the $D_{max}$ to the right optic nerve ($p$=0.001). There was no significant difference in the $D_{max}$ to the left and right hippocampus between VMAT and IMRT ($p$>0.05), but VMAT significantly increased the $D_{100\%}$ ($p$<0.05) compared with IMRT. Compared with VMAT, IMRT significantly reduced the MU ($p$<0.001) but VMAT has a higher treatment efficiency than IMRT, with an average reduction of 41 seconds (294.1 ± 16.4 s for VMAT, 335.8 ± 34.9 s for IMRT, $p$<0.001). Under the conditions of 3%/2 mm, and 2%/2 mm, the gamma passing rate of the IMRT QA was improved compared to VMAT, with an average increase of 0.6%, $p$=0.013, and 1.7%, $p$<0.001, respectively.

**Conclusion:** Both non-coplanar VMAT and non-coplanar fixed field IMRT based on the Monaco TPS produce clinically acceptable results for HA-WBRT+SIB in patients with brain metastases. Compared with IMRT, VMAT has better dose distribution in the target volume and treatment efficiency, but IMRT can better protect the hippocampus and reduce the number of MUs.

KEYWORDS

**hippocampus sparing, brain metastases, simultaneous integrated boost, whole brain radiotherapy, volumetric modulated arc therapy, intensity modulated radiotherapy**

## 1 Introduction

Over the past few years, the incidence rate of brain metastases has increased consistently (1). Whole brain radiotherapy (WBRT) for treating brain metastases, prophylactic cranial irradiation for treating small cell lung cancer, and cranial or craniospinal irradiation for treating malignant tumors of the central nervous system in children have all demonstrated clinical efficacy; however, they also increase cognitive neurotoxicity (2–6). Radiation-induced hippocampal damage plays a significant role in the decline of neurocognitive abilities in patients after WBRT (7). The hippocampus is a central element in memory formation, and the degree of atrophy and cognitive deficits are dependent on the delivered dose; thus, maximal protection of the hippocampus is imperative. Furthermore, the risk of brain metastases occurring in the hippocampus is below 5%; this also suggests that hippocampal avoidance during WBRT (HA-WBRT) is safe (8). Therefore, Radiation Therapy Oncology Group (RTOG) report 0933 proposed hippocampal protection during WBRT (9).

The combination of WBRT and a simultaneous integrated boost (SIB) of localized lesions for brain metastases has been shown to have advantages in terms of shortening treatment time, prolonged local control time, and overall survival (10). For patients with non-small cell lung cancer brain metastases, WBRT combined with a stereotactic radiotherapy boost or simultaneous boost may improve their survival rate compared to WBRT alone (11, 12). With the development of radiation therapy techniques, especially the advent of intensity-modulated radiation therapy (IMRT) techniques, it has become possible to protect organs-at-risk (OARs) such as the hippocampus during WBRT for brain metastases (13). IMRT has significant benefits in hippocampal protection for primary brain tumors, preventing neurocognitive decline and reducing the average dose to the hippocampus. Even after 6 months of follow-up post-irradiation, a neurocognitive benefit was seen in most patients (14). Many researchers have studied whether IMRT or volumetric modulated arc therapy (VMAT) alone in HA-WBRT protect the hippocampus during brain metastases and have confirmed that IMRT or VMAT can effectively protect the hippocampus in HA-WBRT (15–18).

Ilinca Popp et al. confirmed that hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost (HA-WBRT+SIB) could be an efficient therapeutic option for patients with multiple brain metastases. It is associated with improved local tumor control of existing metastases, higher intracranial progression-free survival, reduced death rates associated with neurological conditions, and an acceptable risk of radiation necrosis (19). HA-WBRT+SIB is a complex treatment regimen for patients with brain metastases, aimed at reducing adverse neurocognitive effects while increasing tumor control (19).

Developing an effective hippocampal protection plan for HA-WBRT+SIB treatment poses a challenge. Johannes Kraft et al. compared the dose delivered by the Varian Halcyon linear accelerator based on the Eclipse treatment planning system (TPS) and that delivered by the Elekta Synergy linear accelerator based on the Pinnacle TPS for HA-WBRT+SIB using VMAT. The whole-brain prescribed dose was 30 Gy, and the local boost was 51 Gy administered in 12 fractions. In their study, a 7 mm expansion around the hippocampus was implemented to form the hippocampal avoidance region (HAR).The Halcyon and Synergy Agility linear accelerators produced clinically comparable treatment plans for HA-WBRT+SIB in patients with multiple brain metastases (20). R. Vysakh et al. compared the dose distributions

of the coplanar jaw-fixed VMAT (fVMAT) and the conventional coplanar VMAT without the jaw fixed based on the Versa HD linear accelerator and the Monaco TPS for HA-WBRT. With a whole-brain irradiation dose of 30 Gy in 10 fractions, and a 5 mm extracorporeal expansion of the hippocampus forming the HAR, they found that the Elekta Agility™ collimator system and the Monaco TPS can generate superior HA-WBRT plans using the fVMAT technique (21). Xie Xin et al. compared the dosimetric differences between the coplanar dynamic IMRT (dIMRT) and coplanar VMAT plans of the Varian Linear accelerator on the Eclipse TPS in HA-WBRT alone. They found that the hippocampal dose of the dIMRT group was superior to that of the VMAT group, but neither met the standard of RGOT 0933 (22). Fangyu Liu et al. evaluated the potential of the flattening filter-free (FFF) mode of a linear accelerator for patients with HA-WBRT in comparison with the flattened beams (FF) technique in the application of VMAT and IMRT using dosimetric and radiobiological indexes based on the volume of the hippocampus and target. Their study suggests that the FFF mode is feasible and advantageous in HA-WBRT and VMAT-FFF is the optimal solution in terms of dose distribution of the target, sparing OARs, probability of normal tissue complications of the hippocampus,and delivery efficiency compared to the other three techniques. Additionally, the advantages of the FFF technique for VMAT are more prominent in cases with small hippocampal volumes (33).

No comparative study based on the Monaco TPS has been conducted between non-coplanar VMAT and non-coplanar fixed-field IMRT for HA-WBRT+SIB. The purpose of this work is to explore their advantages and provide feasible treatment plans for patient treatment.

# 2 Materials and methods

## 2.1 Patient data

In total, 22 patients with brain metastases who received radiotherapy at our hospital from June 2022 to October 2023 were retrospectively enrolled, including 10 men and 12 women, aged 27 to 83 years. None of the patients had metastases invading the hippocampus. The basic characteristics of the 22 patients are shown in Table 1.

The study was approved by the ethics committee of Gaozhou People's Hospital (GYLLYJ-2022111). Since it is a retrospective study that presents no risk to the participants' health or economic well-being, the ethics committee of Gaozhou People's Hospital granted an exemption from obtaining informed consent.

## 2.2 Computed tomography simulation and target delineation

The patients were immobilized in a supine position using a head and neck thermoplastic mask and immobilization bag. All the patients underwent computed tomography (CT) imaging acquisition using a large-bore CT simulator (Siemens AG, Forchheim, Germany), with a scan slice thickness and slice interval of 1.5 mm. The scanning range extended from the cranial apex to the third cervical vertebra. Additionally, each patient underwent contrast-enhanced T1-weighted magnetic resonance imaging (MRI) (Siemens AG, Forchheim, Germany) with a slice thickness of 1.5 mm. CT and MR images were imported to the Monaco TPS (Elekta, Crawley, England) for fusion. Following the RTOG 0933 delineation guideline, the hippocampus was delineated, expanding the total hippocampus by 5 mm in all directions to create a HAR. Other OARs were also delineated, including whole-brain tissue, brainstem, lens, eyes, optic nerves, optic chiasm, pituitary, and inner ears. An enhanced lesion was defined by the gross tumor volume (GTV), which was expanded by 3 mm in all directions to form the planning gross tumor volume (PGTV). The whole-brain tissue excluding the HAR, and subtracting the PGTV was defined as the whole-brain planning target (PTV-brain-SIB).

## 2.3 Treatment planning

The prescription dose for all patients was 30 Gy to the PTV-brain-SIB in 15 fractions (2 Gy per fraction), and 50 Gy to the PGTV in 15 fractions (3.33 Gy per fraction). Both targets were treated simultaneously. The PGTV requires 100% of the prescription dose to cover at least 95% of the volume ($V_{50}$≥95%).

Two plans were created for each patient, namely, non-coplanar VMAT and IMRT. Both plans were optimized using the Monte Carlo dose calculation algorithm base Monaco 5.4 TPS on an Axesse linear accelerator (Elekta, Crawley, Sweden) with an Agility multileaf collimator, using a 6 MV photon beam. Non-coplanar VMAT included two fields, with the first field being a 360° coplanar rotation arc starting from 180°, with an increment of 20°, the collimator angle was set at 0°, and the treatment couch angle was set at 0°. The second non-coplanar field couch was set at 270°, while the gantry started at 330°, with a rotational span of 210° and an increment of 15°. The collimator angle was set at 0°.

Non-coplanar fixed-field IMRT used nine fields with gantry angles at 5°, 55°, 135°, 165°, 215°, 270°, 315°, 70°, and 30°, where the treatment couch angle for the 70° and 30° fields was 270°, and for the other seven fields couch angle was 0°. The collimator angle for all nine fields was 315°. The choice of field angle and couch angle was based on experience in daily practice.

The planned sequencing parameters, dose deposition calculation properties, and prescription parameters are shown in Figure 1.

The dose limits for the targets and OARs for both plans are shown in Table 2. The same optimization functions and parameters were used for both plans and optimization templates were created to save planning time, as shown in Table A1.

## 2.4 Dosimetric comparison of target and OARS

A comparison was conducted between the VMAT and IMRT plans for the patient cohort. For the PGTV, $V_{50}$ and $V_{55}$, the heterogeneity index (HI), and the conformity index (CI) were compared. For the

TABLE 1 Basic patient characteristics (n= 22).

| Number | Sex | Age | Primary tumor | Number of metastatic lesions | Metastatic lesions total volume($cm^3$) | Hippocampal volume($cm^3$) | Hippocampal avoidance (HA) volume($cm^3$) | PTV-brain-SIB volume($cm^3$) | HA volume as a percentage of total brain volume(%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Female | 38 | Breast | 2 | 50.8 | 8.9 | 38.3 | 1275.2 | 2.81% |
| 2 | Female | 75 | Lung | 5 | 28.8 | 6.7 | 32.0 | 1071.7 | 2.83% |
| 3 | Male | 54 | Lung | 1 | 8.8 | 8.3 | 36.3 | 1443.7 | 2.44% |
| 4 | Female | 54 | Lung | 3 | 28.4 | 7.5 | 34.9 | 1004.6 | 3.27% |
| 5 | Female | 60 | Lung | 1 | 4.2 | 6.3 | 32.0 | 1137.3 | 2.73% |
| 6 | Male | 58 | Lung | 2 | 45.0 | 7.4 | 32.9 | 1398.2 | 2.23% |
| 7 | Male | 72 | Lung | 11 | 40.4 | 8.2 | 37.0 | 1245.3 | 2.80% |
| 8 | Male | 55 | Lung | 6 | 108.1 | 7.5 | 34.6 | 1180.8 | 2.62% |
| 9 | Female | 66 | Sigmoid | 2 | 20.3 | 7.0 | 33.3 | 1234.9 | 2.59% |
| 10 | Female | 63 | Rectum | 1 | 21.6 | 4.4 | 31.0 | 1103.7 | 2.69% |
| 11 | Female | 52 | Lung | 3 | 16.7 | 3.3 | 21.3 | 1134.7 | 1.82% |
| 12 | Male | 40 | Lung | 7 | 60.7 | 6.6 | 32.9 | 1265.8 | 2.42% |
| 13 | Male | 65 | Lung | 4 | 21.5 | 10.8 | 44.2 | 1305.5 | 3.22% |
| 14 | Female | 27 | Lung | 2 | 20.3 | 5.7 | 29.2 | 1335.5 | 2.11% |
| 15 | Female | 83 | Lung | 2 | 88.6 | 5.2 | 27.2 | 1065.8 | 2.31% |
| 16 | Male | 69 | Lung | 1 | 14.8 | 7.8 | 34.8 | 1437.9 | 2.34% |
| 17 | Male | 75 | Lung | 3 | 90.8 | 6.0 | 29.1 | 1348.2 | 1.99% |
| 18 | Female | 55 | Lung | 13 | 34.0 | 6.6 | 34.8 | 1141.8 | 2.88% |
| 19 | Male | 36 | Brain | 1 | 30.7 | 4.1 | 25.9 | 1354.6 | 1.84% |
| 20 | Male | 77 | Ascending colon | 1 | 20.1 | 8.1 | 36.7 | 1312.1 | 2.68% |
| 21 | Female | 46 | Lung | 2 | 39.2 | 9.5 | 38.9 | 1231.2 | 2.97% |
| 22 | Female | 48 | Breast | 1 | 197.1 | 7.3 | 35.0 | 1141.0 | 2.55% |

PTV-brain-SIB, $D_{98\%}$, $V_{30}$, the HI, and the CI were compared. The HI was calculated as $D_{5\%}/D_{95\%}$ (8). A smaller HI indicates a more uniform dose distribution in the target. The CI was calculated as $V^{2RX}/(TV^*V_{RI})$ (23), where $V_{RX}$ is the volume covered by the prescription dose in the target, TV is the volume of the target, and $V_{RI}$ is the volume covered by the prescription dose. The CI ranges from 0 to 1. A higher CI indicates a more conformal dose distribution in the target. For the hippocampus, $D_{100\%}$ and the maximum dose ($D_{max}$) were compared, while for other OARs, the $D_{max}$ was compared (brainstem, lens, eyes, optic nerves, optic chiasm, inner ears, and pituitary).

## 2.5 Deliverability of the two plans

To examine the deliverability, the monitor units (MUs) and beam-on time for the two plans were compared. All plans were delivered in quality assurance (QA) mode, and the beam-on time was recorded using a calibrated stopwatch. The beam-on time only considered the beam irradiation time without considering the gantry rotation time between arcs or fields.

## 2.6 Pre-treatment quality assurance

Pre-treatment plan verification was performed using ArcCheck (SUN NUCLEAR, California, US). We composited all the fields and did not reset the gantry to 0° but reset the couch to 0°. Gamma analysis was conducted on the dose distribution of the planned and measured data, with evaluation criteria of 3%/3 mm and a 10% dose threshold (TH). A pass rate of at least 95% was considered passed. A more stringent gamma analysis was performed using 3%/2 mm or 2%/2 mm to test the two techniques.

## 2.7 Statistical analysis

Statistical analysis was conducted using SPSS 17.0 (IBM, USA). Normality was tested on the data, and for parameters conforming to a normal distribution, data were expressed as mean ± standard deviation ($\bar{x} \pm s$). Those that did not fit the normal distribution were tested using the non-parametric Friedman test for multi-correlated samples, and quantitative data were expressed as medians and 25% and 75%

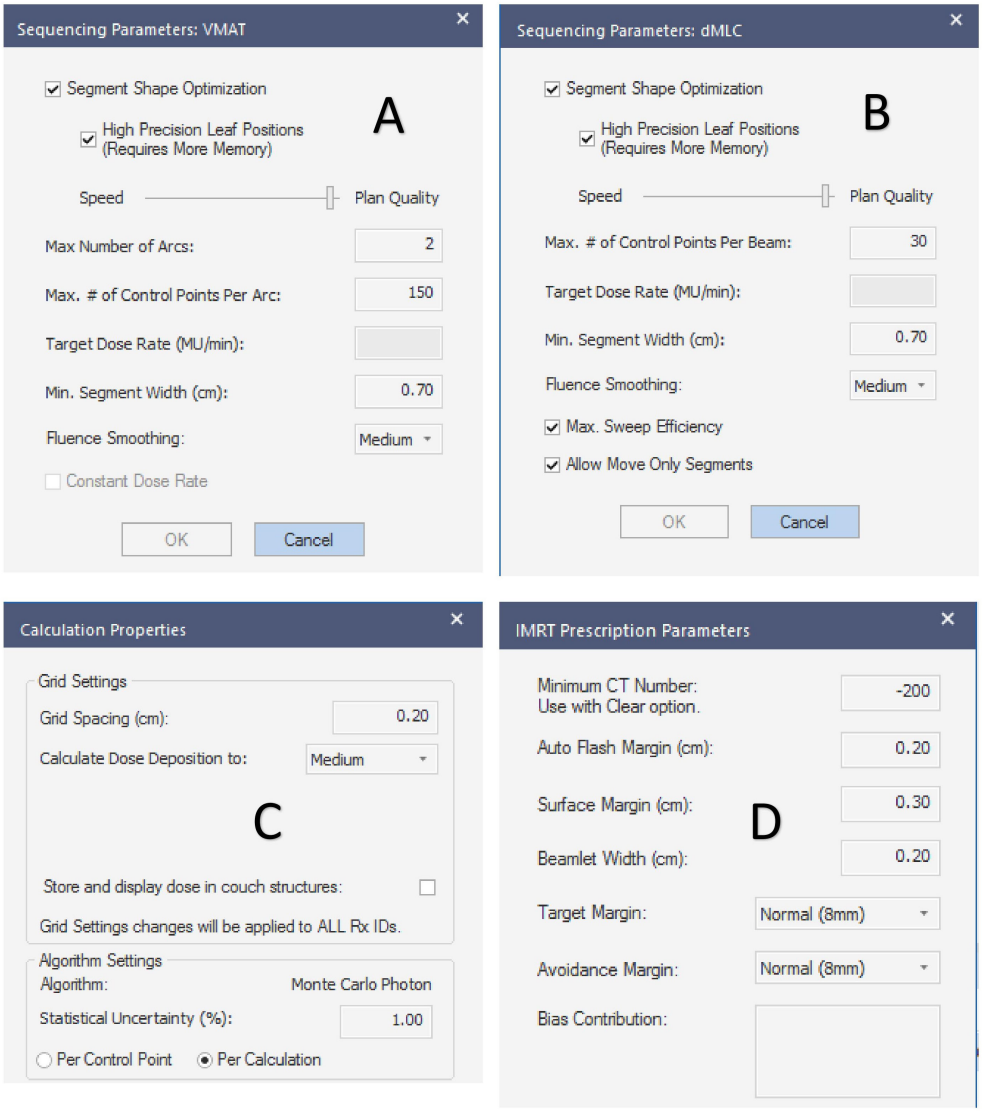



FIGURE 1
Plan setting parameters. **(A–D)** are the sequencing parameters of the VMAT plan, sequencing parameters of the IMRT plan, the calculation properties of the planned dose deposition, and the prescription dose parameters, respectively.

percentiles (P25 and P75). A paired t-test was used for comparisons and a $p$-value<0.05 was considered statistically significant.

# 3 Results

## 3.1 Target dose comparison

The target dose of both plans for the 22 patients met the clinical goals, as shown in Table 2. For the PGTV, there was no significant difference in $V_{50}$, $V_{55}$, $D_{max}$, the HI, and the CI between VMAT and IMRT ($p > 0.05$). For the PTV-brain-SIB, there was no significant difference in $D_{98\%}$ between IMRT and VMAT ($p = 0.103$); VMAT significantly increased the $V_{30}$ for PTV-brain-SIB ($p< 0.001$), reduced the HI ($p = 0.003$), and increased the CI ($p< 0.001$), as shown in Table 3. The dose distribution for a typical patient is illustrated in Figure 2, where the VMAT plan showed a better coverage of the 30 Gy dose, and the IMRT plan had larger cold spots. The dose-volume histogram (DVH) for a representative patient is shown in Figure 3.

## 3.2 Organs-at-risk dose comparison

The dose comparison of OARs for both plans is presented in Table 4. There were no significant differences in the $D_{max}$ for the brainstem, left and right lenses, optic chiasm, pituitary, and left and right hippocampus between the two plans ($p>0.05$). VMAT, relative to IMRT, significantly reduced the $D_{max}$ for the left eye and right eye ($p< 0.001$), significantly increased the $D_{max}$ for the right inner ear ($p = 0.010$), and had no significant difference in the $D_{max}$ for the left inner ear ($p = 0.458$). VMAT significantly reduced the $D_{max}$ for

TABLE 2 Dose criteria for the targets and organs at risk.

| Structure | Dose limits |
|---|---|
| PGTV | $V_{55}$<5%, $V_{50}$≥95% |
| PTV-brain-SIB | $V_{30}$≥90%, $D_{98\%}$≥25 Gy |
| Left and right hippocampus | $D_{100\%}$<9 Gy($D_{100\%}$>10 Gy not acceptable), $D_{max}$<16 Gy($D_{max}$>17 Gy not acceptable) |
| Left and right optic nerves | $D_{max}$<37.5 Gy |
| Optic chiasm | $D_{max}$<37.5 Gy |
| Left and right lens | $D_{max}$<8 Gy |
| Left and right eyes | $D_{max}$<37.5 Gy |
| Brainstem | $D_{max}$<37.5 Gy |
| Left and right inner ears | $D_{max}$<37.5 Gy |
| Pituitary | $D_{max}$<37.5 Gy |

$V_x$ is the volume of the region of interest (ROI) when the dose received is x Gy, $D_{x\%}$ is the dose corresponding to x% of the ROI volume, and $D_{max}$ is the maximum dose.

the left optic nerve ($p$=0.006) but significantly increased the $D_{max}$ for the right optic nerve ($p$ = 0.001) compared to IMRT. There were no significant differences in the $D_{max}$ for the left and right hippocampus between VMAT and IMRT ($p$ > 0.05), but VMAT significantly increased the $D_{100\%}$ for both the left and right hippocampus ($p$< 0.05).

## 3.3 Deliverability of the plans

The average MUs for non-coplanar IMRT and non-coplanar VMAT were 1174 MU (990~1592) and 1326 MU (1112~1660), respectively. IMRT significantly reduced the planned MU compared to VMAT ($p$<0.001). VMAT significantly reduced the beam on time with a mean reduction of 41 seconds (294.1 ± 16.4 seconds vs. 335.8 ± 34.9 seconds, $p$<0.001) compared to IMRT.

## 3.4 Pre-treatment plan quality assurance

The mean gamma passing rates for the VMAT and IMRT plans were 99.4% and 99.5%, respectively, with the 3%/3 mm and 10% threshold criteria, and there was no significant difference ($p$=0.125). However, with the 3%/2 mm, and 2%/2 mm criteria, the passing rate for IMRT was higher than VMAT (99.1 ± 0.56 vs 98.5 ± 0.86, 97.6 ± 1.05 vs 96.0 ± 1.41) with a mean increase of 0.6% ($p$=0.013) and 1.7% ($p$<0.001), respectively.

# 4 Discussion

Several publications have studied HA-WBRT or HA-WBRT +SIB based on different techniques or different linear accelerators (10, 15–17, 21, 24–27). However, there have been no dose comparison studies on HA-WBRT+SIB using non-coplanar VMAT and non-coplanar IMRT based on the Monaco TPS. Different from previous studies (10, 15, 16), the current research protected the brainstem, inner ear, pituitary, and hippocampus while reaching the dose coverage of the target. The $D_{max}$ of the left hippocampus of the current study was 1554.1 ± 249.4 Gy for IMRT and 1515.3 ± 130.1 Gy for VMAT, respectively. The $D_{max}$ of the right hippocampus was 1504.3 ± 129.7 Gy for IMRT and 1496.6 ± 96.7 Gy for VMAT, respectively, which were much lower than that of the study of Johannes Kraft (20) and XIE XIN (22), and the HAR

TABLE 3 PGTV and PTV-brain-SIB dose comparison between the IMRT and VMAT plans for the patient cohort.

| Structure | Parameter | IMRT | VMAT | Difference (%) | $p$-value |
|---|---|---|---|---|---|
| PGTV | $V_{50}$ (%) | 97.26 ± 1.43 | 97.46 ± 1.62 | -0.01 ± 0.01 | 0.350 |
| | $V_{55}$ (%) | 1.35 ± 1.19 | 1.35 ± 0.97 | 0.26 ± 1.33 | 0.983 |
| | $D_{max}$ (cGy) | 5582.8 ± 62.8 | 5578.5 ± 46.3 | 0.08 ± 1.11 | 0.749 |
| | HI | 1.07 ± 0.01 | 1.07 ± 0.01 | 0.13 ± 0.82 | 0.480 |
| | CI | 0.68 ± 0.10 | 0.69 ± 0.09 | -1.37 ± 6.21 | 0.341 |
| PTV-brain-SIB | $D_{98\%}$ (cGy) | 2696.2 ± 82.9 | 2724.5 ± 71.2 | -1.01 ± 2.90 | 0.103 |
| | $V_{30}$ (%) | 92.09 ± 1.25 | 93.32 ± 1.23 | -0.01 ± 0.01 | *<0.001* |
| | HI | 1.46 ± 0.17 | 1.44 ± 0.16 | 1.44 ± 2.03 | *0.003* |
| | CI | 0.77 ± 0.05 | 0.81 ± 0.04 | -4.16 ± 2.44 | *<0.001* |

p value denotes the results of paired t-test between IMRT and VMAT plans. The italicized values indicated p value is less than 0.05.

TABLE 4 Comparison of the OAR doses between the IMRT and VMAT plans.

| Structure | Parameter | IMRT | VMAT | Difference (%) | *p*-value |
|---|---|---|---|---|---|
| Brainstem | $D_{max}$ (cGy) | 3564.7 ± 511.7 | 3551.6 ± 513.2 | 0.40 ± 2.60 | 0.493 |
| Left eye | $D_{max}$ (cGy) | 2625.1 ± 363.1 | 2320.7 ± 271.3 | 13.26 ± 11.23 | *<0.001* |
| Right eye | $D_{max}$ (cGy) | 2672.8 ± 374.0 | 2380.5 ± 442.0 | 13.54 ± 12.52 | *<0.001* |
| Left lens | $D_{max}$ (cGy) | 620.8 ± 32.6 | 630.2 ± 28.5 | -1.39 ± 5.15 | 0.194 |
| Right lens | $D_{max}$ (cGy) | 630.5 ± 52.3 | 646.7 ± 32.8 | -2.34 ± 8.26 | 0.185 |
| Left inner ear | $D_{max}$ (cGy) | 3190.2 ± 105.7 | 3160.8 ± 123.1 | 1.10 ± 5.90 | 0.458 |
| Right inner ear | $D_{max}$ (cGy) | 3073.0 ± 170.2 | 3152.1 ± 169.9 | -2.43 ± 4.08 | *0.010* |
| Left optic nerve | $D_{max}$ (cGy) | 3070.2 ± 125.8 | 2978.3 ± 130.0 | 3.20 ± 4.72 | *0.006* |
| Right optic nerve | $D_{max}$ (cGy) | 2724.5 ± 296.2 | 2943.5 ± 208.4 | -7.30 ± 9.54 | *0.001* |
| Optic chiasm | $D_{max}$ (cGy) | 3246.8 ± 108.8 | 3215.2 ± 110.0 | 1.03 ± 3.30 | 0.183 |
| Pituitary | $D_{max}$ (cGy) | 3034.7 ± 115.4 | 3053.9 ± 67.0 | -6.03 ± 3.74 | 0.445 |
| Left hippocampus | $D_{100\%}$ (cGy) | 757.2 ± 49.6 | 778.8 ± 59.1 | -2.59 ± 5.00 | 0.018 |
| | $D_{max}$ (cGy) | 1554.1 ± 249.4 | 1515.3 ± 130.1 | 2.14 ± 7.00 | 0.186 |
| Right hippocampus | $D_{100\%}$ (cGy) | 742.7 ± 62.6 | 790.0 ± 59.1 | -5.96 ± 4.07 | *<0.001* |
| | $D_{max}$ (cGy) | 1504.3 ± 129.7 | 1496.6 ± 96.7 | 0.49 ± 5.17 | 0.647 |

p value denotes the results of paired t-test between IMRT and VMAT plans. The italicized values inidicate pvalue is less than 0.05.

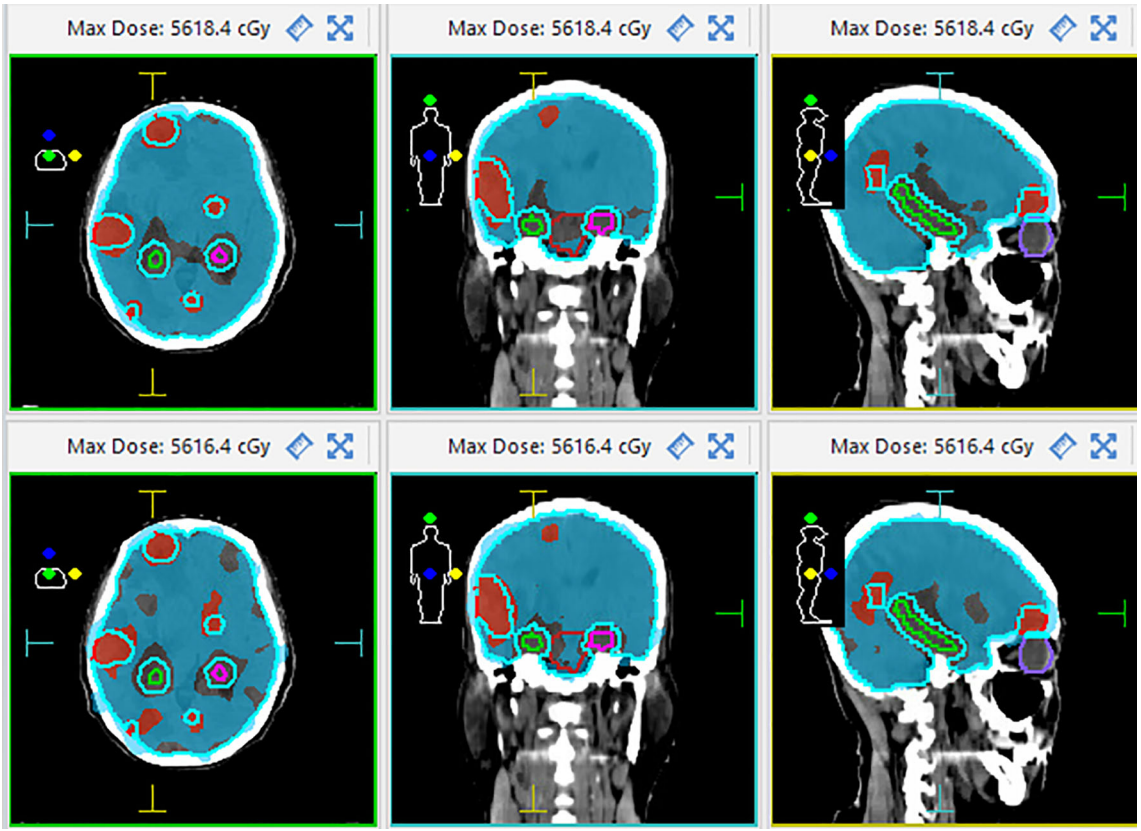



FIGURE 2
The dose distribution of the VMAT (upper) and IMRT (down) plans in cross-section (left), coronal (middle), and sagittal plane (right) for a typical patient. The red area represents the coverage with an isodose of 50 Gy and the blue area represents the coverage with an isodose level at 30 Gy.

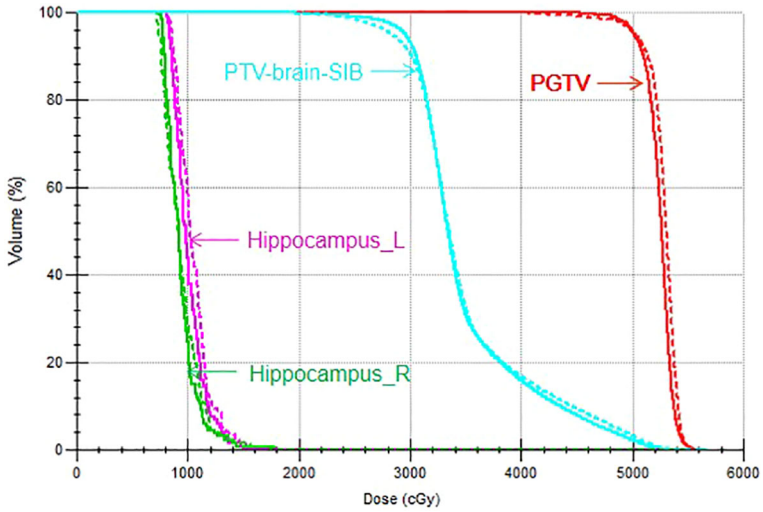



FIGURE 3
DVH of the VMAT and IMRT plans for a representative patient. The solid line indicates the VMAT plan and the dotted line indicates the IMRT plan.

in this study was only a 5 mm outward expansion around the hippocampus. Yu Xiao et al. used the Pinnacle TPS to improve the existing non-coplanar VMAT HA-WBRT plan, and compared the dosimetric differences between the improved non-coplanar VMAT plan and the traditional non-coplanar and coplanar VMAT plans in HA-WBRT (28). For the $D_{max}$ and $D_{100\%}$ in hippocampal tissue, the improved non-coplanar VMAT could be controlled at approximately 14.37 Gy and 8.40 Gy, respectively, which were significantly smaller than the traditional non-coplanar and coplanar plans ($p$<0.05). However, the improved non-coplanar VMAT plan was too complicated, with a total of six arcs and four couch angles (270°, 315°, 45°, and 0°), so the treatment efficiency was low. Shao Wei et al. investigated the dosimetric differences between VMAT with a flattening filter (FF) and flattening filter free (FFF) in HA-WBRT using four half arcs with a prescribed dose of 30 Gy in 10 fractions. The $D_{max}$ of the hippocampus in the FF-VAMT and FFF-VMAT plans was 16.46 ± 0.56 Gy and 15.13 ± 0.38 Gy, respectively, and the $D_{100\%}$ was 7.72 ± 0.28 Gy and 7.12± 0.34 Gy, respectively. The results of the current study are comparable to the two plans, but their study did not have SIB (29).

This study was conducted based on the Monaco TPS and the Elekta Axesse linear accelerator, and a template for field setup and optimization functions was created to save planning time. The most important factor in successful HA-WBRT+SIB planning is the TPS and field setup and the optimization of function criteria. While some plans required fine-tuning for specific cases, such as an SIB region that is too close to the hippocampus (<3 mm) or a large number and volume of brain metastases, the optimization function template provided in this study met the planning requirements for the majority of patients. Among the 22 patients, only one patient had a $D_{max}$ to the left hippocampus that exceeded 17 Gy (26.21 Gy for IMRT and 20.52 Gy for VMAT, respectively). This was due to its proximity to the SIB target, which was only 1 mm from the lesion at 50 Gy. This differs from previous studies that only provided field settings without essential optimization functions (8, 30–32). This study provides detailed planning optimization parameter settings for the clinical implementation of this treatment technique. It protects the pituitary gland, inner ear, and hippocampus. The two plans in this study showed no significant differences in the dose distribution for the target PGTV with SIB. However, VMAT significantly increased the $V_{30}$, conformity, and uniformity for PTV-brain-SIB ($p$<0.05). The plans exhibited no significant differences in the $D_{max}$ to the hippocampus ($p$>0.05), but IMRT significantly reduced the $D_{100\%}$ for the hippocampus, with a lower MU compared to VMAT. Although IMRT reduced the MU, it did not reduce the beam on time, resulting in a time advantage. The reason may be that VMAT can beam on continuously during rotation. When considering the idling time of the gantry, IMRT requires more treatment time than VMAT, so from a cost/benefit perspective, VMAT can treat more patients than IMRT in the same amount of time. IMRT had a significantly higher gamma passing rate than VMAT under the 3%/2 mm and 2%/2 mm criteria, possibly due to the reduced MU that reduced the linear accelerator's head leakage.

Collision avoidance is an important issue in a non-coplanar plan, and the Monaco TPS provides the Room's Eye View function to view the gantry, couch, lighting, decorations, axes, and the active treatment beam in a room to make sure that the patient is in a safe position. Acquiring the patient's CBCT will be a problem when the non-coplanar plan is implemented for treatment; we can only acquire CBCT when the couch angle is at 0°, so it is important to make sure that the center of couch rotation accuracy meets the requirements so that the dose distribution is as expected.

The current study has several limitations. First, it was a retrospective study comparing dosimetry, did not compare biology such as probability of tumor control and probability of normal tissue complications, and was not a prospective study with clinical outcomes. Therefore, the advantages and disadvantages of the two techniques need to be determined by long-term patient follow-up and based on extensive case practice. Second, there were

no comparisons of other IMRT techniques such as helical tomography and proton and carbon ion therapy. Third, the correlation between dose limitation in the hippocampus and symptom reduction in patients has not been adequately demonstrated. Fourth, the TPS and dose calculation algorithms used for planning can also have an impact on outcomes. Thus, different TPSs such as Varian's Eclipse and RaySearch's RayStation and different algorithms such as anisotropy analysis algorithm and pencil beam need to be further investigated. Fifth, the measurement data for all QA measurements in this study were not completed during the same time period, and the deviation of the dose output from the machine each time may have an effect on the QA results. Sixth, the control point limits of the two plans were not uniform, and different control points may have significant effects on dose distribution.

## 5 Conclusion

This study performed a dosimetric comparison of non-coplanar VMAT and non-coplanar IMRT for HA-WBRT+SIB in brain metastasis patients on the Monaco TPS. Both plans demonstrated clinically acceptable results for hippocampal protection. VMAT had advantages in target dose distribution and treatment efficiency, while IMRT protected the hippocampus better and reduced the machine monitor units.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the ethics committee of Gaozhou People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

HZ: Writing – original draft, Writing – review & editing, Conceptualization. MZ: Investigation, Writing – original draft. ZC: Data curation, Writing – review & editing. ST: Software, Writing – original draft. ZW: Resources, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the science and technology plan project of Maoming City, China (Grant number 2022S010).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1428329/full#supplementary-material

## References

1. Singh K, Saxena S, Khosla AA, McDermott MW, Kotecha RR, Ahluwalia MS. Update on the management of brain metastasis. *Neurotherapeutics*. (2022) 19:1772–81. doi: 10.1007/s13311-022-01312-w

2. de Ruiter MB, Groot PFC, Deprez S, Pullens P, Sunaert S, de Ruysscher D, et al. Hippocampal avoidance prophylactic cranial irradiation (HA-PCI) for small cell lung cancer reduces hippocampal atrophy compared to conventional PCI. *Neuro Oncol*. (2022) 25:167–76. doi: 10.1093/neuonc/noac148

3. Hsu F, Carolan H, Nichol A, Cao F, Nuraney N, Lee R, et al. Whole brain radiotherapy with hippocampal avoidance and simultaneous integrated boost for 1-3 brain metastases: a feasibility study using volumetric modulated arc therapy. *Int J Radiat Oncol Biol Phys*. (2010) 76:1480–5. doi: 10.1016/j.ijrobp.2009.03.032

4. Shobana MK, Balasubramanian S. Pediatric craniospinal irradiation – The implementation and use of normal tissue complication probability in comparing photon versus proton planning. *J Med Phys*. (2021) 46:244–52. doi: 10.4103/jmp.jmp_75_21

5. Popp I, Hartong NE, Nieder C, Grosu A-L. PRO: do we still need whole-brain irradiation for brain metastases? *Cancers*. (2023) 15:3193. doi: 10.3390/cancers15123193

6. Ni M, Liu W, Jiang A, Wang Y, Sheng Y, Zeng H, et al. Whole brain radiation therapy plus focal radiation boost may generate better survival benefit for brain metastases from non-small cell lung cancer. *Front Oncol*. (2020) 10:576700. doi: 10.3389/fonc.2020.576700

7. Gondi V, Tomé WA, Mehta MP. Why avoid the hippocampus? A comprehensive review. *Radiotherapy Oncol*. (2010) 97:370–6. doi: 10.1016/j.radonc.2010.09.013

8. Popp I, Grosu AL, Fennell JT, Fischer M, Baltas D, Wiehle R. Optimization of hippocampus sparing during whole brain radiation therapy with simultaneous integrated boost-tutorial and efficacy of complete directional hippocampal blocking. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft* …. (2022) 198:537–46. doi: 10.1007/s00066-022-01916-3

9. Gondi V, Pugh SL, Tome WA, Caine C, Corn B, Kanner A, et al. Preservation of memory with conformal avoidance of the hippocampal neural stem-cell compartment during whole-brain radiotherapy for brain metastases (RTOG 0933): a phase II multi-institutional trial. *J Clin Oncol*. (2014) 32:3810–6. doi: 10.1200/JCO.2014.57.2909

10. Jiang A, Sun W, Zhao F, Wu Z, Shang D, Yu Q, et al. Dosimetric evaluation of four whole brain radiation therapy approaches with hippocampus and inner ear avoidance and simultaneous integrated boost for limited brain metastases. *Radiat Oncol*. (2019) 14(1):46. doi: 10.1186/s13014-019-1255-7

11. Sun H, Xu L, Wang Y, Zhao J, Xu K, Qi J, et al. Additional radiation boost to whole brain radiation therapy may improve the survival of patients with brain metastases in small cell lung cancer. *Radiat Oncol*. (2018) 13:250. doi: 10.1186/s13014-018-1198-4

12. Ni M, Jiang A, Liu W, Sheng Y, Zeng H, Liu N, et al. Whole brain radiation therapy plus focal boost may be a suitable strategy for brain metastases in SCLC patients: a multi-center study. *Radiat Oncol*. (2020) 15:70. doi: 10.1186/s13014-020-01509-3

13. Andreas JJM, Kundapur V. Hippocampus avoidance whole-brain radiation therapy: A practical intensity-modulated radiation therapy planning and delivery approach to RTOG 0933. *J Med Imaging Radiat Sci*. (2015) 46:78–84. doi: 10.1016/j.jmir.2014.09.009

14. Nath DVVD, Shah DS, Vishwanathan B. Hippocampal sparing in radiation therapy to primary brain tumors - and impact on memory function in adults. *Eur J Mol Clin Med*. (2022) 9:70–81.

15. Shen J, Bender E, Yaparpalvi R, Kuo HC, Basavatia A, Hong L, et al. An efficient Volumetric Arc Therapy treatment planning approach for hippocampal-avoidance whole-brain radiation therapy (HA-WBRT). *Med Dosim*. (2015) 40:205–9. doi: 10.1016/j.meddos.2014.11.007

16. Xu Y, Xu Y, Men K, Xiao J, Dai J. Application of piecewise VMAT technique to whole-brain radiotherapy with simultaneous integrated boost for multiple metastases. *Radiat Oncol*. (2022) 17:86. doi: 10.1186/s13014-022-02059-6

17. Zieminski S, Khandekar M, Wang Y. Assessment of multi-criteria optimization (MCO) for volumetric modulated arc therapy (VMAT) in hippocampal avoidance whole brain radiation therapy (HA-WBRT). *J Appl Clin Med Phys*. (2018) 19:184–90. doi: 10.1002/acm2.12277

18. Pokhrel D, Sood S, Lominska C, Kumar P, Badkul R, Jiang H, et al. Potential for reduced radiation-induced toxicity using intensity-modulated arc therapy for whole-brain radiotherapy with hippocampal sparing. *J Appl Clin Med Phys*. (2015) 16:131–41. doi: 10.1120/jacmp.v16i5.5587

19. Popp I, Rau S, Hintz M, Schneider J, Bilger A, Fennell JT, et al. Hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost for multiple brain metastases. *Cancer*. (2020) 126:2694–703. doi: 10.1002/cncr.32787

20. Kraft J, Weick S, Breuer K, Lutyj P, Bratengeier K, Exner F, et al. Treatment plan comparison for irradiation of multiple brain metastases with hippocampal avoidance whole brain radiotherapy and simultaneous integrated boost using the Varian Halcyon and the Elekta Synergy platforms. *Radiat Oncol*. (2022) 17:192. doi: 10.1186/s13014-022-02156-6

21. Vysakh R, Raman RG, Puzhakkal N. Fixed field technique for hippocampal avoidance whole-brain radiotherapy: A feasibility study using Elekta system. *Iranian J Radiat Res*. (2023) 21:105–9. doi: 10.52547/ijrr.21.1.14

22. Xin XIE, Liang LI, Xuemei FAN, Yumei X, Guanchen S, Xiaoxiao L. Dosimetric study of hippocampal protective whole brain radiotherapy based on Varian accelerator. *Chin J Radiol Health*. (2021) 30:326–30. doi: 10.13491/j.issn.1004-714X.2021.03.014

23. Liu F, Peng Y, Li Q, Zhang Q, Shi H, Qie S, et al. Feasibility of flattening filter free beams for hippocampal avoidance whole-brain radiotherapy: a dosimetric and radiobiological analysis. *Front Oncol*. (2023) 13:1290434. doi: 10.3389/fonc.2023.1290434

24. Brown PD, Gondi V, Pugh S, Tome WA, Wefel JS, Armstrong TS, et al. Hippocampal avoidance during whole-brain radiotherapy plus memantine for patients with brain metastases: phase III trial NRG oncology CC001. *J Clin Oncol*. (2020) 38:1019–29. doi: 10.1200/JCO.19

25. Schuermann M, Dzierma Y, Nuesken F, Oertel J, Rube C, Melchior P. Automatic radiotherapy planning for glioblastoma radiotherapy with sparing of the hippocampus and nTMS-defined motor cortex. *Front Neurol*. (2021) 12:787140. doi: 10.3389/fneur.2021.787140

26. Shirata R, Inoue T, Sugimoto S, Saito AI, Omura M, Minagawa Y, et al. Dosimetric investigation of whole-brain radiotherapy with helical intensity modulated radiation therapy and volumetric modulated arc therapy for scalp sparing. *BJR Open*. (2023) 5:20220037. doi: 10.1259/bjro.20220037

27. Gondi V, Tolakanahalli R, Mehta MP, Tewatia D, Rowley H, Kuo JS, et al. Hippocampal-sparing whole-brain radiotherapy: a "how-to" technique using helical tomotherapy and linear accelerator-based intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. (2010) 78:1244–52. doi: 10.1016/j.ijrobp.2010.01.039

28. Chen L-J, Li M-H, Cheng H-W, Kuo C-Y, Sun W-L, Tsai J-T. Hippocampus-sparing whole-brain radiotherapy: dosimetric comparison between non-coplanar and coplanar planning. *Ther Radiol Oncol*. (2021) 5:1–1. doi: 10.21037/tro-20-50

29. Xiao YU, Bing YAN, Lei LIU. and Dosimetric study on improved non-coplanar volumetric modulated arc therapy for hippocampal avoidance whole-brain radiotherapy. *Chin J Med Phys*. (2021) 38:529–33. doi: 10.3969/j.issn.1005-202X.2021.05.001

30. Wei S, Jing L, Yi S, Gang W, Xingzhao G, Duoping S, et al. Dosimetric comparison of hippocampal avoidance wholebrain radiotherapy with flattening filter and flattening filter-free modes. *Chin J Radiol Health*. (2022) 31:740–5. doi: 10.13491/j.issn.1004-714X.2022.06.018

31. Kim KS, Wee CW, Seok JY, Hong JW, Chung JB, Eom KY, et al. Hippocampus-sparing radiotherapy using volumetric modulated arc therapy (VMAT) to the primary brain tumor: the result of dosimetric study and neurocognitive function assessment. *Radiat Oncol*. (2018) 13:29. doi: 10.1186/s13014-018-0975-4

32. Sun W, Chen K, Li Y, Xia W, Dong L, Shi Y, et al. Optimization of collimator angles in dual-arc volumetric modulated arc therapy planning for whole-brain radiotherapy with hippocampus and inner ear sparing. *Sci Rep*. (2021) 11:19035. doi: 10.1038/s41598-021-98530-7

33. Yokoyama K, Kurosaki H, Oyoshi H, Miura K, Utsumi N. Plan quality comparison between hippocampus-sparing whole-brain radiotherapy treated with halcyon and tomotherapy intensity-modulated radiotherapy. *Technol Cancer Res Treat*. (2022) 21:15330338221108529. doi: 10.1177/15330338221108529

# Frontiers in Oncology

**Advances knowledge of carcinogenesis and tumor progression for better treatment and management**

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

---

## Discover the latest Research Topics

See more →


**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact