Received: 5 August 2025 | Revised: 26 October 2025 | Accepted: 31 October 2025

DOI: 10.1002/mp.70160

RESEARCH ARTICLE

MEDICAL PHYSICS

Development of a defacing algorithm to protect the privacy of head and neck cancer patients in publicly-accessible radiotherapy datasets

Kayla O'Sullivan-Steben[1] | **Luc Galarneau**[1,2] | **John Kildea**[1,2,3]

[1]Medical Physics Unit, McGill University, Montreal, Québec, Canada

[2]Research Institute of the McGill University Health Centre, Montreal, Québec, Canada

[3]Gerald Bronfman Department of Oncology, McGill University, Montreal, Québec, Canada

Correspondence
Kayla O'Sullivan-Steben, Medical Physics Unit, McGill University, Montreal, Québec, Canada.
Email: kayla.osullivan-steben@mail.mcgill.ca

Funding information
Natural Sciences and Engineering Research Council of Canada; Ministère de la Santé et des Services sociaux; Fonds de Recherche du Québec - Santé; Rossy Cancer Network

Abstract

Background: The increase in public medical imaging datasets has raised concerns about potential patient reidentification from head CT scans. However, existing defacing algorithms, which help protect patient confidentiality, fail to preserve critical radiotherapy structures, including organs at risk (OARs) and planning target volumes (PTVs) in head and neck cancer (HNC) patients. Furthermore, current algorithms do not address the defacing of DICOM-RT structure set and dose data, which also contain information for facial surface rendering.

Purpose: To develop and validate a novel automated defacing algorithm that preserves OARs and PTVs while removing identifiable features from HNC CTs and DICOM-RT data.

Methods: Eye contours were used as landmarks to automate the removal of CT pixels above the inferior-most slice of the eye and anterior to the midpoint of the eye. Pixels within PTVs were retained if they intersected with the removed region. The body contour and dose map were then reshaped to reflect the defaced image. We validated our approach on 829 HNC CT-simulation scans from 622 patients. To evaluate privacy protection, we applied the FaceNet512 facial recognition algorithm before and after defacing on 3D-rendered CT scan pairs from 70 patients at two time points. To assess research utility, we examined the impact of defacing on auto-contouring performance using LimbusAI and analyzed the locations of PTVs relative to the defaced regions.

Results: Before defacing, the facial recognition algorithm matched 97% of patients' CT scans. After defacing, this rate dropped to just 4%. LimbusAI effectively auto-contoured organs in the defaced CTs, with perfect Dice scores of 1 for OARs below the defaced region, and mean Dice scores exceeding 0.95 for OARs on the same slices as the defaced region. PTV analysis revealed that 86% of PTVs were entirely below the cropped region, 9.1% were on the same slice as the crop without overlap, and only 4.9% extended into the cropped area. All overlapping PTVs were preserved through our algorithm's design.

Conclusions: We developed a novel defacing algorithm that anonymizes HNC CT scans and related DICOM-RT data. Our algorithm balances patient privacy while preserving essential structures for radiotherapy research, facilitating the sharing of HNC imaging datasets for Big Data and AI.

KEYWORDS
CT anonymization, defacing, radiotherapy data-sharing

1 | INTRODUCTION

1.1 | Background

In the present era of Big Data and Artificial Intelligence, there is an increasing demand for publicly accessible imaging datasets for radiation oncology research. When publishing datasets, it is crucial to remove any identifying information to protect patient privacy.[1] However, this task becomes particularly complex in the case of head and neck cancer (HNC) patients, as the body surface renderings of their 3D image scans could potentially be used for facial recognition and re-identification.[1–3] For example, Schwarz and colleagues (2022)[4] found that surface renderings of MRI scans could be matched to a photo of the same patient with 97%–98% accuracy, while CTs were matched at 78% accuracy using an automated facial recognition tool. Therefore, to enable more researchers to contribute without ethical concerns to public datasets for AI and Big Data applications in HNC, it is essential that the community has access to robust de-identification techniques. *Defacing*, in particular, contributes to data anonymization by obscuring identifiable facial features in imaging datasets.

1.2 | Existing defacing algorithms

While defacing is commonly used to address privacy concerns, existing defacing algorithms were developed primarily for neuroimaging research and do not necessarily suit the needs of radiotherapy-related studies.[5] More specifically, these tools typically focus on preserving brain structures but fail to consider other critical structures, such as the many organs at risk (OARs) and planning target volumes (PTVs) in the CT-simulation (CT-sim) and/or cone-beam CT scans of the head and neck region used for radiotherapy delivery. As a result, critical structures can become distorted or removed entirely when conventional defacing techniques are used. This problem was illustrated by Wahid et al. (2022),[6] showing how four state-of-the-art defacing algorithms obscure or remove important HNC OARs like the lymph node levels and salivary glands.

Maintaining the integrity of such structures is crucial for radiotherapy research, which can involve tasks such as auto-segmentation, radiomics, and tracking of tumor volumes and anatomical changes.[7] For instance, Sahlsten et al. (2023)[5] demonstrated how current defacing algorithms impede HNC auto-segmentation research. In their study, they examined eight publicly available defacing algorithms and found that five were incapable of defacing CT images, as they were specifically designed for MRI use. The remaining three tools did deface the images, but caused a significant decline in performance of their auto-segmentation algorithm when it was trained and tested on defaced CTs compared to the original CTs.

As an additional but important consideration in the radiotherapy domain, conventional defacing algorithms do not consider the anonymization of DICOM-RT Structure Set and Dose data that radiotherapy treatment plans are stored in. These data also contain 3D anatomical information that can be used for the surface rendering of a patient's face and so must also be defaced.

1.3 | Our approach

In this work, we aimed to develop an automated defacing algorithm for HNC CT scans that preserves OARs and PTVs while removing identifiable features like the eyes, eyebrows, and forehead. It was also important that our technique extends to defacing DICOM-RT Structure Set and Dose data to ensure an added layer of privacy protection for radiotherapy patients in addition to de-identification. We validated our defacing algorithm by comparing the performance of facial recognition and auto-segmentation algorithms before and after defacing, and by examining the location of HNC tumors relative to the defaced area. We believe that this work can facilitate the safe sharing of HNC imaging datasets by providing a method to anonymize CT images while maintaining their utility for radiotherapy research. To the best of our knowledge, this is the first defacing algorithm designed specifically for HNC radiotherapy data.

2 | METHODS

This work was carried out on a retrospective single-center patient dataset of 622 HNC patients who underwent radiotherapy treatment between January 1, 2017 and March 31, 2024. In total, the dataset comprised 829 CT-sim scans along with their associated Structure Sets and Dose maps. The protocol for this retrospective research study was approved by the Research Ethics Board (REB) of the McGill University Health Centre [project number 2025-11285]. All work of the study was conducted in accordance with the Canada Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2). Additionally, all potentially identifying fields of the CT and DICOM-RT data were

TABLE 1 Summary of the needs assessment for our defacing algorithm.

Need	Consideration	Specific Solution
1) Protect patient privacy	Facial recognition algorithms typically rely on facial landmarks, such as the eyes, eyebrows, forehead, nose, and mouth.[8]	Remove facial landmarks used by facial recognition algorithms.
2) Preserve tumor volumes for HNC research	HNC predominantly occurs in the oral cavity, larynx, and pharynx.[9]	Retain the inferior portion of the head starting at least at the oral cavity; preserve all PTV pixels, even if within the proposed cropped region.
3) Preserve OARs for radiotherapy research	The OARs near the surface of the face—which are most susceptible to being removed during defacing—include: eyes, oral cavity, lips, mandible, lymph nodes, submandibular and parotid glands.	Retain the inferior portion of the head, ensuring removal of only those structures that compromise privacy.
4) Ensure utility for neuroimaging studies	The brain structure should remain intact in the image.	Remove only pixels anterior to the centre of the eye.
5) Ensure utility for HNC radiotherapy replanning studies	Replanning often hinges on subtle external changes such as weight loss and local anatomical changes.	Avoid deformation or removal of the skin surface around the tumour, chin, and neck.

Abbreviations: HNC, head and neck cancer; OAR, organ at risk.

anonymized by the Eclipse Treatment Planning System (Varian Medical Systems, Inc. Palo Alto, CA, USA) on export, thus stripping them of any identifiers such as names, dates, and so forth.

2.1 | Selecting the region to deface

In our study, we aimed to strike a balance between ensuring the anonymity of the images and maintaining their utility for radiotherapy research. To achieve this, we conducted a needs assessment, summarized in Table 1, which outlines the key considerations and proposed solutions that guided the algorithm's development. The resulting approach removes the region anterior to the centre of the eye and superior to the bottom of the eyes, while preserving anatomical information related to the PTV, even when located within the defaced area. We chose to remove pixels rather than add tissue to obscure them, as additive approaches would retain some original identifying features that could potentially be recovered.

2.2 | Automated defacing workflow

An overview of our automated defacing workflow is presented in Figure 1. In the first step, the eye contours—as delineated by the dosimetrists/radiation oncologists—are extracted from the DICOM-RT Structure Set data of the CT-sim image. Starting from the inferior-most slice of the eye structure and moving toward the top of the head, the algorithm generates a binary mask (values of 0 and 1) that removes all pixels anterior to the center point of the eye contour from each image slice of the CT-sim. The algorithm then checks for any PTV or brain structures that overlap with the initially cropped area, and, if applicable, modifies the mask to retain the image pixels corresponding to the PTV and brain structures to preserve the target and organ anatomy. The mask is then applied to the image to remove the defaced pixels, and can likewise be applied to any other images (e.g., daily cone-beam CTs) that have been registered to the CT-sim.

In the second step, the array of *x*,*y*,*z*-coordinates of the body contour are updated to reflect the modified cropped images. Additionally, the eye, lens, and cornea contours are removed from the Structure Set. The algorithm then checks whether any other contours (aside from the PTV and brain) protrude into the cropped region and, if so, reshapes them accordingly. The PTV contours are kept in the Structure Set, even when they overlap with the cropped region. Finally, in the third step, the mask is resized and resampled to the size and spacing of the Dose map contained in the RT Dose file. The new mask is then applied to the Dose map to deface it. Once again, in cases where the PTV overlaps with the cropped region, the mask also retains the Dose map voxels corresponding to the overlapping PTV image pixels.

To test our workflow, we ran the automated algorithm over a HNC dataset of 829 unique HNC CT-sim scans corresponding to 622 patients. We visually inspected each defaced CT, Structure Set and Dose map to ensure proper defacing. Additionally, we calculated the average running time required to deface one CT and associated DICOM-RT data to assess computational feasibility. The defacing code was implemented in Python (version 3.8) and is available in the following GitHub repository: https://github.com/kildealab/defaceRT.

2.3 | Validation - patient privacy

To evaluate privacy protection, we conducted facial recognition tests on 2D images (screenshots) of 3D

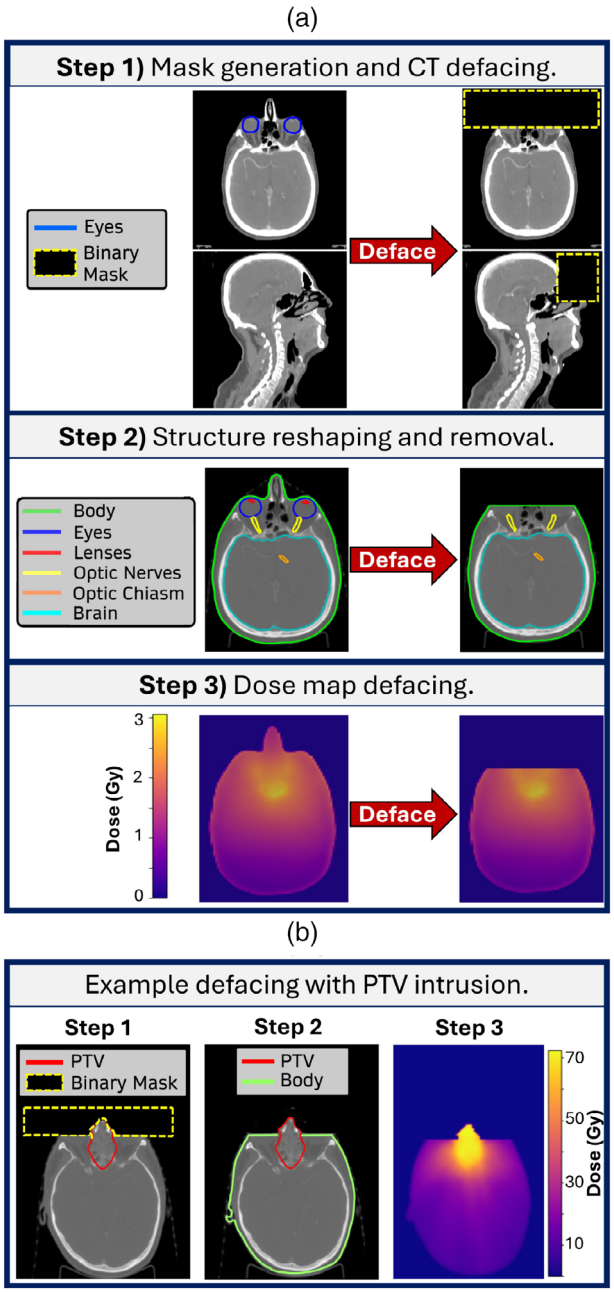

FIGURE 1 (a) Overview of the automated defacing algorithm's workflow. (b) Example of the workflow applied to a patient whose PTV intrudes into the defaced region. PTV, planning target volume.

renders of the CT scans before and after defacing. This methodology is consistent with other studies in the literature investigating facial recognition using CTs.[2,3,10] We performed these tests on a randomly selected subset of 70 patients from our dataset, each of whom had a second independent CT-sim scan available for matching (taken for replanning purposes, typically a week or more after the initial CT). Using two CT scans taken at different time points is essential for replicating a real-life facial recognition scenario. It is akin to comparing two different photos of the same person, whereas comparing two photos or CTs taken in close temporal proximity could yield better (and potentially misleading) matching results.

2.3.1 | Image preparation

To obtain images of the 3D rendered faces for facial recognition, the body contours of the CTs were rendered in 3D in the Eclipse Treatment Planning System (Varian Medical Systems, Inc. Palo Alto, CA, USA). All body contours were set to the same white colour and a 2D screenshot was taken of the render facing forward. This process was repeated to provide three CT scans for each patient: one at the first time point (CT_{t1}), one independent scan at the second time point (CT_{t2}), and one defaced scan corresponding to the first time point (dCT_{t1}).

2.3.2 | Face detection and recognition

Facial recognition algorithms typically comprise three main steps, which we implemented using the open-source DeepFace library[11] (github.com/serengil/deepface). First, a detector model locates the face within an image so that it can be isolated for recognition. In our case, we tested all facial detection methods available in DeepFace and opted to use RetinaFace,[12] as it successfully detected all 140 faces (70 pairs) in our non-defaced CT scans. Next, a recognition model transforms these detected faces into vector embeddings. For this step, we employed the FaceNet512 algorithm,[13] as it has been shown to outperform other existing publicly available models in facial recognition tasks[14,15] and has been used in other CT imaging defacing studies.[16,17] In the final step, a pair of facial embeddings are compared using a distance metric, with closer embeddings (smaller distances) indicating higher similarity and thus a greater likelihood that they represent the same person. For this comparison, we used DeepFace's default cosine distance metric.

We performed facial recognition tests and obtained cosine distances for the following three comparison pairing groups, each of which is visualized in Figure 2:

1. **Same-patient pairing**: comparison between CT scans of the same patient at two different time points ($CT_{t1,patient\ i}$ vs. $CT_{t2,patient\ i}$ for all i).
2. **Different-patient pairing**: comparison between CTs of different patients (($CT_{t1,patient\ i}$ vs. $CT_{t1,patient\ j}$ for all $j = i + 1$ to 70) for all i).
3. **Defaced-same-patient pairing**: comparison between a defaced CT from the first time point and the original CT from the second time point of the same patient ($dCT_{t1,patient\ i}$ vs. $CT_{t2,patient\ i}$ for all i).

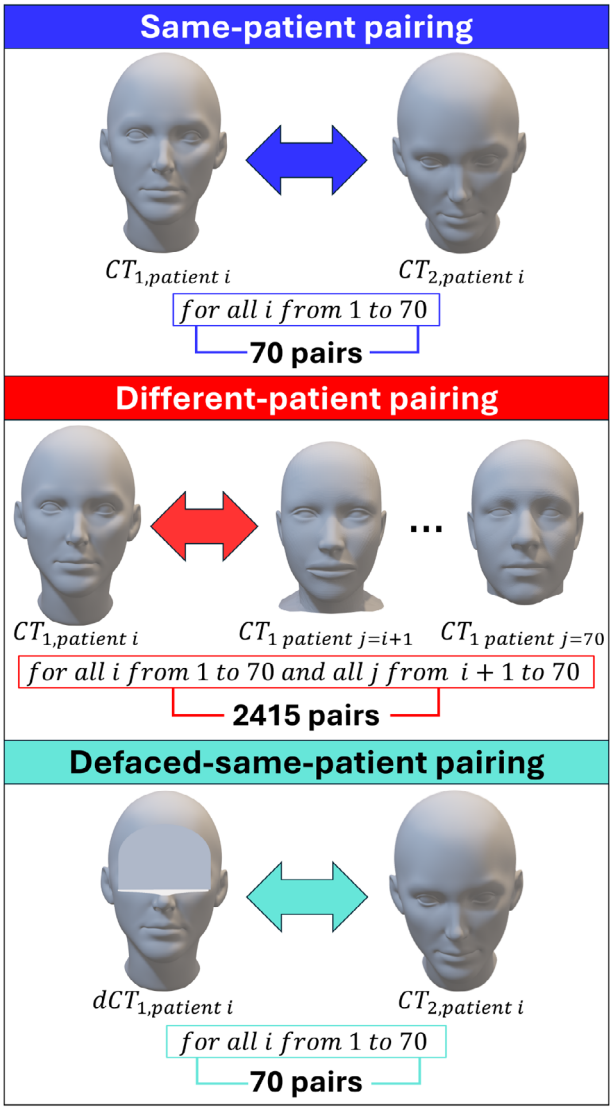

FIGURE 2 Overview of the three groups of facial recognition tests performed. Each pair of image comparisons yields one cosine distance. Note that these images are artist-rendered 3D faces for visualization purposes. They do not represent the CTs of real patients. CT, computed tomography.

2.3.3 | Baselining exercise

Given that FaceNet512 was originally developed for 2D photographic images and was not explicitly trained on 2D renderings of 3D images, it was important to establish a baseline to ensure that the model can adequately distinguish between two CT scans of the same patient (expected low cosine distances) and two CT scans of different patients (expected high cosine distances). Although the aforementioned studies in the literature[16,17] did not undertake this additional step, we did so to satisfy ourselves that FaceNet512 is an appropriate algorithm to use for our use case involving radiotherapy CT-sim scans.

We performed the Mann–Whitney U test to confirm that the set of cosine distances obtained for the same-patient pairing group and the different-patient pairing group were statistically different. Good differentiation in cosine distances between these two pairing groups would indicate that the FaceNet512 algorithm is capable of distinguishing same patients and different patients.

The optimal cosine distance threshold for separating the two pairing groups was determined by maximizing the Youden Index[18] using scikit-learn's ROC curve implementation,[19] which measures the tradeoff between the true positive rate and false positive rate. Using this threshold, we determined the number of correctly matched patients before defacing, as well as the number of false positive matches when comparing CTs of different patients.

2.3.4 | Defacing evaluation

With the baselining completed, we compared the cosine distances obtained for the defaced-same-patient pairing group with the baseline groups of the same-patient and different-patient pairing groups. We hypothesized that the defaced images would produce cosine distances that are consistent with those of the different-patient pairing group, indicating successful anonymization. Using the previously defined threshold, we determined the number of correctly matched patients after defacing.

We conducted a Wilcoxon signed-rank test to determine if there was a significant difference between the cosine distances for the same patient before and after defacing. Finally, we performed the Mann–Whitney U test to compare the cosine distances of the defaced-same-patient pairing and different-patient pairing groups to determine if these two groups are distinguishable.

A summary of the statistical comparisons and expected outcomes for both the baselining exercise and the defacing evaluation are presented in Table 2. For all statistical tests, a p-value < 0.005 was considered significant.

2.4 | Utility for radiotherapy research

To assess the utility of our defaced images for radiotherapy research, we investigated the effects the defacing had on the two main structure types of interest for radiotherapy research: OARs and PTVs.

2.4.1 | OARs

We used LimbusAI (Limbus AI Inc, Regina, SK, Canada), the auto-segmentation software used in our clinic, to automatically generate contours for head and neck

TABLE 2 Statistical comparison tests and expected outcomes of cosine distances for the different pairing groups.

Test purpose	Cosine distance comparison group	Test Name	Expected Outcome
Baselining	Same-patient pairing (70 pairs) versus Different-patient pairing (2415 pairs)	Mann-Whittney U	Significantly different
Defacing evaluation	Same-patient pairing (70 pairs) versus Defaced-same-patient pairing (70 pairs)	Wilcoxon Signed-Rank	Significantly different
	Different-patient pairing (2415 pairs) versus Defaced-same-patient pairing (70 pairs)	Mann-Whittney U	Indistinguishable

OARs on both the original and defaced CTs of the same subset of 70 patients. Contouring was performed for all 46 OARs defined in our clinic's "Head and Neck" LimbusAI template.

For each patient, we then calculated the Dice coefficient[20] to measure the degree of overlap between the contours on the original and defaced CTs for each OAR contoured. A Dice score of 1 indicates that the two volumes overlap completely, whereas a Dice score of 0 indicates that they do not overlap at all. The intent of calculating these Dice scores was to allow us to examine to what extent each OAR and its surrounding tissues can still be segmented and analysed in radiotherapy research using the defaced CTs.

2.4.2 | PTVs

We analyzed the locations of the PTVs relative to the removed facial regions in our complete dataset of 622 patients. Specifically, we quantified the fractions of PTVs that fell into one of three categories: (1) completely below the cropped region, (2) on the same slices as the cropped region, but not overlapping it, and (3) overlapping with the cropped region. This assessment provided insight into the potential impact of the defacing process on tumor regions, as PTVs located in or on the same slice as the removed region may encounter additional research limitations compared to those completely below it.

3 | RESULTS

3.1 | Real-world defacing

Our algorithm was able to automatically deface 793 (96%) of the 829 CTs and associated Structure Sets and Dose maps in our dataset. The 36 (4%) scans that were not successfully automatically defaced did not have eyes contoured, and thus would require an additional step of either manual or automated contouring before defacing. To best represent real-world conditions, we did not undertake the additional contouring step. All subsequent analyses were thus performed on the 793 automatically defaced CT scans. The code executed with a mean runtime of 13 ± 6 s per CT scan (including defacing of the Structure Set and Dose map) on a machine equipped with a virtual Intel Core Processor (Skylake, IBRS) and 8 GB of RAM. For illustration purposes, Figure 3 shows an example of a public-domain[21] 3D rendered face before and after defacing.

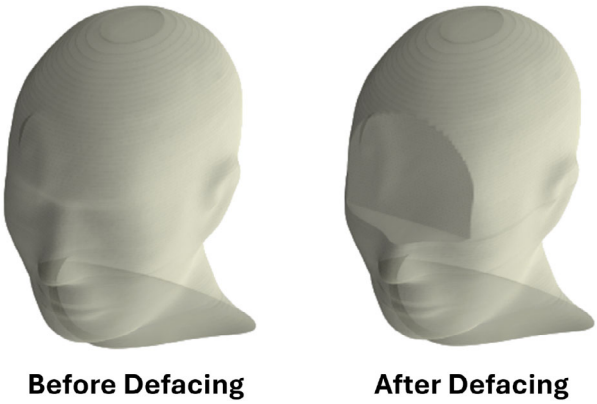

FIGURE 3 3D surface rendering of the reconstructed face of a head phantom before and after defacing. The head phantom data were retrieved from the SlicerRtData GitHub repository.[21]

3.2 | Privacy evaluation

Figure 4a presents histograms of the cosine distances for the three pairing groups tested. The cosine distances for the same-patient pairing group ($CT_{t1,patient\ i}$ vs. $CT_{t2,patient\ i}$) were significantly different from the cosine distances for the different-patient pairing group ($CT_{t1,patient\ i}$ vs. $CT_{t1,patient\ j}$), with a p-value of $p < 0.001$. This significant difference indicates that FaceNet512 can reliably distinguish the renderings of CT scans of the same patient from CT scans of different patients. After defacing, we found that the cosine distances of the defaced-same-patient pairing group ($dCT_{t1,patient\ i}$ vs. $CT_{t2,patient\ i}$) were significantly different from the cosine distances of the same-patient pairing group ($CT_{t1,patient\ i}$ vs. $CT_{t2,patient\ i}$), with a p-value of $p < 0.001$. Furthermore, the cosine distances for the defaced-same-patient pairing group were statistically indistinguishable from the cosine distances of different-patient pairing group (p-value of 0.40).

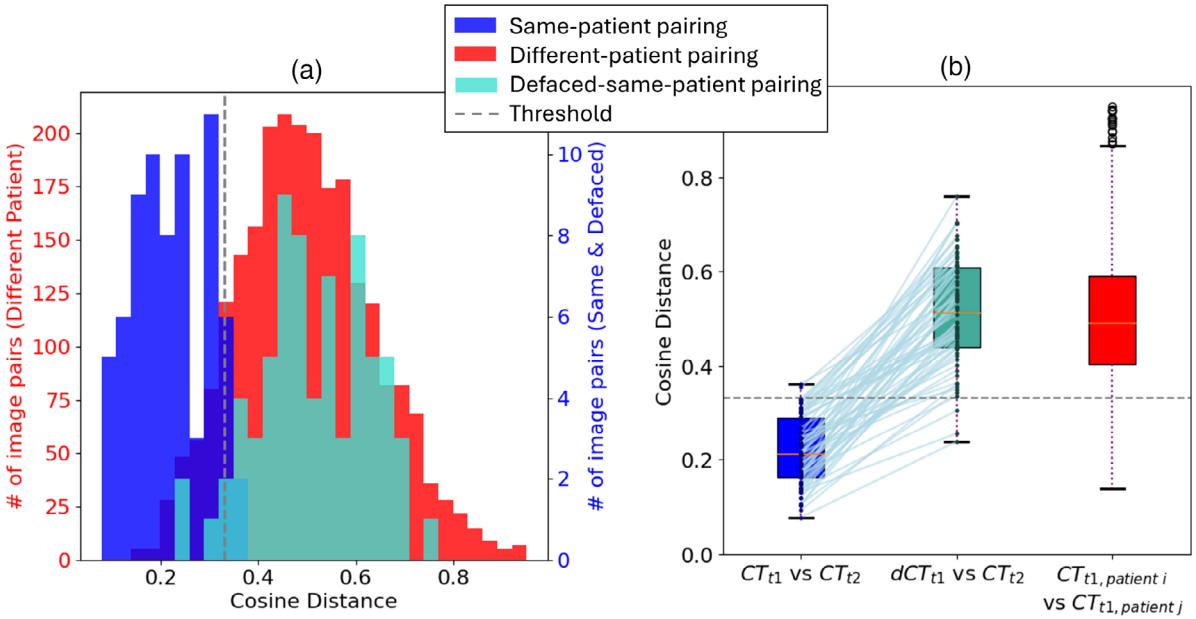

FIGURE 4 Results of the FaceNet512 facial recognition algorithm on 70 patients for our three pairing groups. Lower cosine distances indicate a higher likelihood that two scans are from the same patient. (a) presents histograms of the cosine distances of the three pairing groups tested. (b) shows the same data presented in whisker plots, with blue lines connecting data for the same patient.

Using the maximized Youden Index, the threshold cosine distance was determined to be 0.331. Using this value, the baseline match rate for same-patient pairs (i.e., before defacing) was 97% (68/70) with a false positive rate of 11% (258/2415). After defacing, the match rate decreased to 4% (3/70), which is notably lower than the false positive rate. Figure 4b illustrates the increase in cosine distances (i.e., decrease in match likelihood) between each pair of CTs before and after defacing.

3.3 | Evaluation of utility for radiotherapy research

3.3.1 | OARs

The LimbusAI software was able to generate contours on all of the 70 original and 70 defaced CT scans. A visualization of the contours before and after defacing on a sample patient are provided in Figure 5.

The mean Dice scores comparing each of the 46 auto-contoured OARs on the original and defaced CT scans are presented in Figure 6. As expected, the six anatomical structures that were partially or completely removed exhibited low Dice scores, indicating poor overlap. Specifically, the lenses and corneas had mean Dice scores of 0 ($SD = 0$), while the left and right eyes had mean Dice scores of 0.5 ($SD = 0.1$). These results are consistent with the fact that the corneas and lenses were entirely removed, and about half of the eyes were removed.

For the eight anatomical structures located on the same slice as the cropped region, but not overlapping it, Dice scores were all close to or equal to 1, indicating almost perfect overlap. Specifically, the brain, optic chiasm, left optic nerve, and right optic nerve had mean Dice scores of 0.999 ($SD = 0.002$), 0.95 ($SD = 0.05$), 0.95 ($SD = 0.04$), and 0.96 ($SD = 0.03$), respectively. The left and right hippocampi, brainstem, and pituitary all had perfect Dice scores of 1 ($SD = 0$).

Finally, all of the 32 OARs inferior to the cropped region were auto-contoured identically in the defaced and original CTs, yielding perfect Dice scores of 1 ($SD = 0$). Notably, this includes the 10 lymph node levels, which are routinely used by radiation oncologists to delineate clinical target volumes (CTVs).

3.3.2 | PTVs

We found that for 86.0% (682/793) of the CT scans that were successfully defaced, the PTVs were located entirely inferior to the cropped region, indicating that the beam entry points were unaffected by the defacing process for the majority of patients. 9.1% (72/793) of CTs had a PTV that was partially on the same slice as the defaced region, but not overlapping the cropped pixels. Only 4.9% (39/793) of the CT scans had a PTV that overlapped with the initial cropped region. These cases corresponded to patients with diagnoses in the nasal cavity (22/39), the sinuses (11/39), the palate (3/39), the cheek (2/39), and the nasopharynx (1/39).

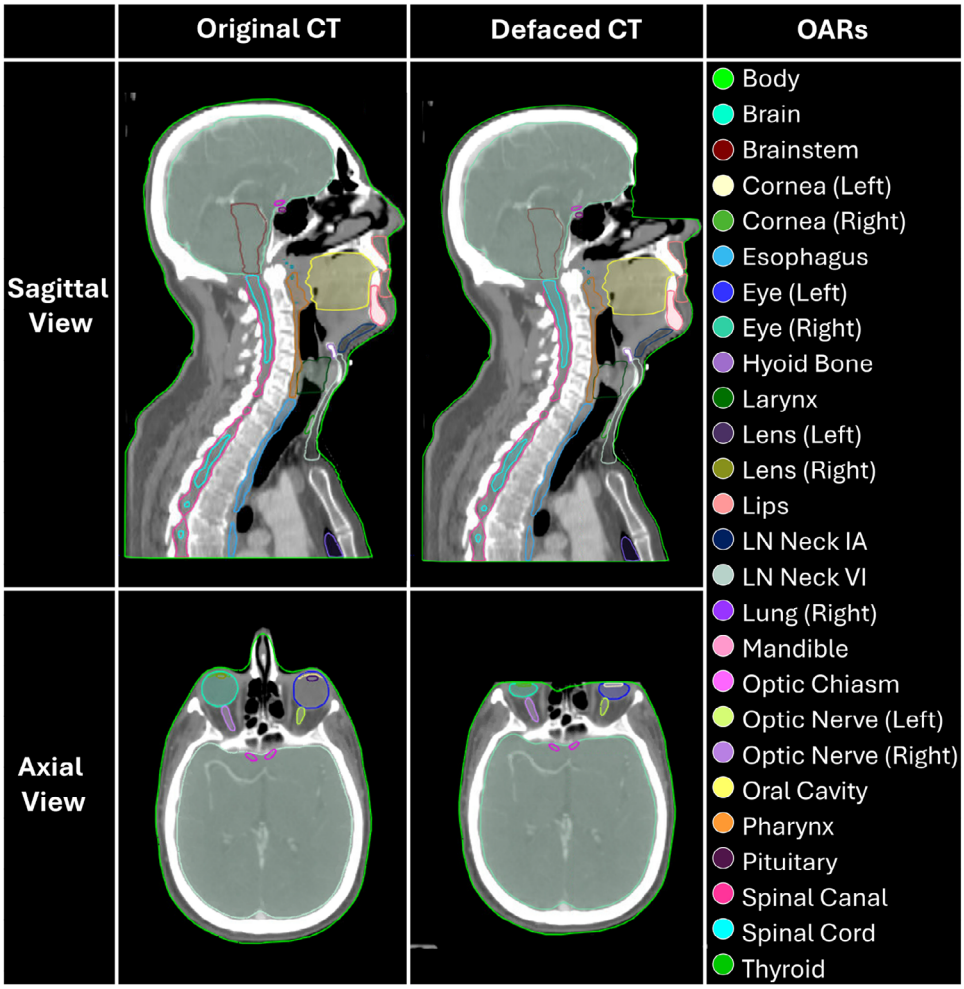

FIGURE 5 Visualization of LimbusAI's auto-contoured OARs before and after defacing on a sample patient. Not pictured are the brachial plexuses, clavicles, cochleas, hippocampi, left lung, submandibular glands, parotid glands, and the following right and left Lymph Node (LN) levels: Neck, Neck 2347AB, Neck IB, and Neck V. OAR, organ at risk.

4 | DISCUSSION

In this study, we developed a novel defacing algorithm for HNC CT scans and associated DICOM-RT data. Our algorithm automatically removes identifiable facial features—namely the eyes, eyebrows, and forehead—while preserving critical anatomical structures needed for radiotherapy research. Additionally, our algorithm is fast and requires minimal computational power. To our knowledge, this is the first implementation of a defacing algorithm specifically designed for HNC CT data that also includes the defacing of Structure Sets and Dose maps, addressing an important privacy vulnerability in radiotherapy data sharing.

Our defacing algorithm successfully addressed all identified privacy concerns, with facial recognition rates (between defaced and non-defaced images) decreasing from 97% to 4% following defacing. These results align with similar studies, such as Schwarz et al. (2022)[4] and Selfridge et al. (2023),[16] who reported decreases from 78% to 5% and 93% to 7% with their respective defacing algorithms on CT scans but without radiotherapy considerations. As such, our approach achieves comparable privacy to existing algorithms, while offering additional advantages for radiotherapy research.

A key advantage of our approach is the preservation of OARs essential for radiotherapy research. Previous studies have demonstrated that conventional defacing algorithms, which were primarily designed to preserve only the brain structure for neuroimaging studies, obscure or remove important HNC structures, including lymph node levels and salivary glands.[5,6] This OAR degradation can have downstream impacts on research that uses these defaced images. For instance, Sahlsten et al. (2023)[5] found that autosegmentation models trained on original CT scans performed poorly on defaced images, and models trained on defaced scans were less effective when tested on the original data.

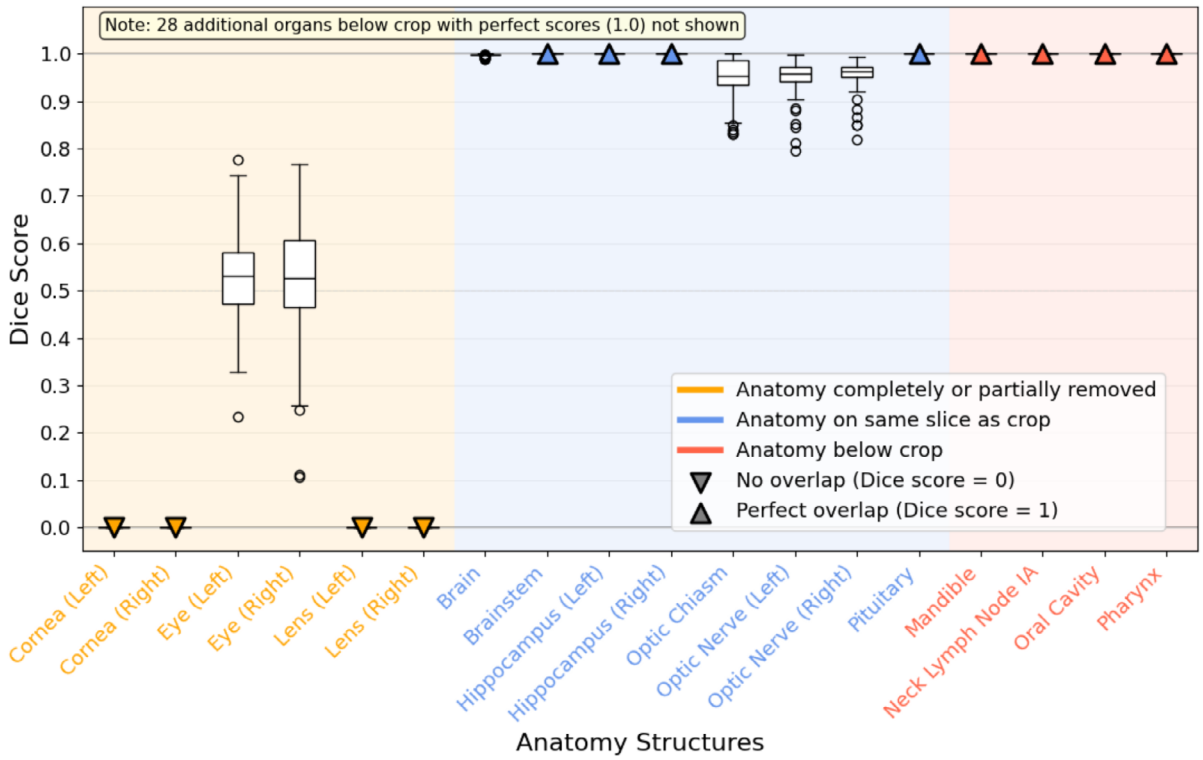

FIGURE 6 Mean Dice scores measuring overlap between auto-contoured OARs on the defaced CTs and the original CTs for 70 patients. A value of 1 indicates that the contours were identical on the original and defaced CTs. Additional OARs below the cropped region are not shown in this graph. They include brachial plexuses, clavicles, cochleas, esophagus, hyoid bone, larynx, lips, lungs, neck lymph node levels, parotid glands, spinal canal, spinal cord, submandibular glands, and thyroid, each of which had perfect Dice scores of 1 ($SD = 0$). CT, computed tomography; OAR, organ at risk.

In contrast, our defacing algorithm not only preserves these critical HNC structures by design, but also maintains sufficient surrounding anatomical context to enable accurate auto-segmentation. Our analysis revealed that auto-segmentation is virtually unaffected by defacing—aside from the intentionally removed structures—with perfect Dice scores of 1.0 for structures inferior to the cropped area and near-perfect mean scores (> 0.95) for structures on the same slice as the crop.

Another advantage of our defacing algorithm is the systematic preservation of PTV structures. Not only are all PTV pixels retained in the image, but our evaluation confirmed that the majority of the PTVs (86.0%) lie entirely below the cropped regions, with only 4.9% extruding into the defaced area. To our knowledge, no other study has investigated the impact of defacing on PTVs, despite their critical value to radiotherapy research applications. However, given that other defacing algorithms remove or deform important OARs, it is reasonable to assume that PTVs would similarly be compromised by them.

While most existing defacing techniques were originally developed for MRI, a few more recent algorithms have been proposed specifically for CT scans. For example, Mahmutoglu et al. (2024)[17] and Lindholz et al. (2025)[22] report deep-learning based defacing algorithms, but both remove a substantial portion of important HNC OARs around the mouth, resulting in the same limitation of many of the aforementioned MRI-focused defacing techniques.

Rather than removing pixels entirely, some researchers have explored deformation and blurring methods that attempt to preserve facial resemblance. Uchida et al. (2023)[10] proposed a deformation-based de-identification method that manipulates head CT Images according to 400 control points set on the surface rendering of the patient's face. This approach maintains facial resemblance, but the process is manual and thus not feasible for large-scale datasets. Selfridge et al. (2023)[16] proposed a blurring method that, although automatic, severely deforms the facial structure. Importantly, both studies only investigated the effect of defacing on the brain structure, while appearing to deform many of the OARs near the anterior of the head. These anatomical distortions can interfere with radiotherapy applications requiring precise anatomical measurements, such as adaptive HNC treatment planning studies that rely on tracking body shrinkage and skin separation around the lower face and neck.

By comparison, our pixel-removal approach maintains clean undeformed anatomical boundaries in the regions below the defaced region.

Overall, in addressing the limitations of current defacing techniques, our algorithm supports a broad spectrum of research applications in radiotherapy that rely on anatomical structural information, such as autosegmentation model development, radiomics analyses, dosimetric studies, anatomical change modeling, tumor volume tracking, and treatment planning.

Limitations

Our study has several limitations. First, the algorithm relies on pre-contoured eye structures for radiotherapy treatment planning, which were absent in about 4% of our dataset. However, given the widespread availability of auto-contouring tools and the ease of contouring the eyes, this limitation can be relatively easily overcome.

Second, for the small subset of patients (4.9% in our dataset) with PTVs extending into the defaced region (primarily those with nasal cavity and sinus tumours), comprehensive dosimetric studies may be limited since nearby OARs like the eyes and lenses are of higher dosimetric importance in these cases. For these cases, further investigations are required to properly quantify the dosimetric impact of defacing.

Third, although we demonstrated that FaceNet512 provided a reasonable baseline for facial recognition using 2D renderings of 3D CT scans, it was not specifically trained to do so. However, to our knowledge, no other algorithms have been trained in such a context. As such, we believe that any publicly available algorithm would suffer from the same limitation. Moreover, while our study focused on facial recognition between two CT renderings, real-world scenarios may involve comparisons between a CT rendering and publicly available photographs, which can be explored in future work.

Lastly, while our results indicate that our defacing algorithm substantially minimizes reidentification risks, there is always an inherent risk in sharing any healthcare data publicly. We also acknowledge the possibility that new facial recognition algorithms may be created in the future that are better at recognizing defaced patients. Furthermore, our defacing algorithm specifically addresses re-identification via facial renderings and does not protect against all theoretically possible re-identification methods. While the risk appears minimal, patients could in principle be re-identified using other 3D anatomical features, such as the shapes and relative locations of contoured structures, or dental patterns. We therefore advise that all publicly shared head CT images—whether defaced or not—be distributed only via secure imaging archives with signed user agreements prohibiting the use of these images for non-research purposes, including 3D rendering for facial recognition.

5 | CONCLUSION

In conclusion, we developed a defacing algorithm specifically for the defacing of HNC CT scans and their related DICOM-RT data. Our algorithm balances the need for patient privacy with the preservation of critical OARs and target structures that are crucial for radiotherapy research. By enabling the secure de-identification of imaging data while maintaining their research utility, this work addresses an important need in the era of Big Data and AI. Overall, this work can facilitate the sharing of HNC imaging datasets, which in turn can enable broader collaboration and accelerate advancements in radiotherapy research.

ACKNOWLEDGMENTS

We gratefully acknowledge Victor Matassa for extracting the HNC patient IDs that met our inclusion criteria from the clinical database. We also thank Odette Rios-Ibacache, who, along with K.O., exported the CT-sim images and DICOM-RT data used in this project. K.O. acknowledges financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Québec Ministère de la Santé et des Services Sociaux (MSSS), and the CREATE Responsible Health and Healthcare Data Science (SDRDS) grant of NSERC. This work was also supported by the Fonds de recherche du Québec–Santé dual-chair in AI and digital health held by J.K. and a research grant from the Rossy Cancer Network.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Parker W, Jaremko JL, Cicero M, et al. Canadian association of radiologists white paper on de-identification of medical imaging: part 2, practical considerations. *Can Assoc Radiol J*. 2021;72:25-34.
2. Parks CL, Monson KL. Automated facial recognition of computed tomography-derived facial images: patient privacy implications. *J Digit Imaging*. 2017;30:204-214.
3. Mazura JC, Juluru K, Chen JJ, Morgan TA, John M, Siegel EL. Facial recognition software success rates for the identification of 3d surface reconstructed facial images: implications for patient privacy and security. *J Digit Imaging*. 2012;25:347-351.
4. Schwarz CG, Kremers WK, Lowe VJ, et al. Face recognition from research brain PET: an unexpected PET problem. *NeuroImage*. 2022;258:119357.
5. Sahlsten J, Wahid KA, Glerean E, et al. Segmentation stability of human head and neck cancer medical images for radiotherapy applications under de-identification conditions: Benchmarking data sharing and artificial intelligence use-cases. *Front Oncol*. 2023;13:1120392.
6. Wahid KA, Glerean E, Sahlsten J, et al. Artificial intelligence for radiation oncology applications using public datasets. *Semin Radiat Oncol*. 2022;32:400-414.

7. Volpe S, Pepa M, Zaffaroni M, et al. Machine learning for head and neck cancer: a safe bet?—a clinically oriented systematic review for the radiation oncologist. *Front Oncol*. 2021;11:772663.
8. Adjabi I, Ouahabi A, Benzaoui A, Taleb-Ahmed A. Past, present, and future of face recognition: a review. *Electronics*. 2020;9:1188.
9. Barsouk A, Aluru JS, Rawla P, Saginala K, Barsouk A. Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Med Sci*. 2023;11:42.
10. Uchida T, Kin T, Saito T, et al. De-Identification technique with facial deformation in head CT Images. *Neuroinformatics*. 2023;21:575-587.
11. Serengil SI, Ozpinar A. LightFace: A Hybrid Deep Face Recognition Framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)* IEEE; 2020:1-5. Accessed November 19, 2025. https://ieeexplore.ieee.org/document/9259802
12. Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: Single-Stage Dense Face Localisation in the Wild. 2019, arXiv:1905.00641 [cs].
13. Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015:815-823, arXiv:1503.03832 [cs].
14. Firmansyah A, Kusumasari TF, Alam EN. Comparison of Face Recognition Accuracy of ArcFace, Facenet and Facenet512 Models on Deepface Framework. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*; 2023:535-539.
15. Serengil S, Özpinar A. A Benchmark of facial recognition pipelines and co-usability performances of modules. *Bilişim Teknolojileri Dergisi*. 2024;17:95-107.
16. Selfridge AR, Spencer BA, Abdelhafez YG, Nakagawa K, Tupin JD, Badawi RD. Facial anonymization and privacy concerns in total-body PET/. *J Nucl Med*. 2023; 64:1304-1309.
17. Mahmutoglu MA, Rastogi A, Schell M, et al. Deep learning-based defacing tool for CT angiography: CTA-DEFACE. *Europe Radiol Exp*. 2024;8:1-7.
18. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-35.
19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
20. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11:178-189.
21. SlicerRtData/eclipse-8.1.20-phantom-ent at master · SlicerRt/SlicerRtData. GitHub. Accessed October 2, 2025. https://github.com/SlicerRt/SlicerRtData/tree/master/eclipse-8.1.20-phantom-ent
22. Lindholz M, Ruppel R, Schulze-Weddige S, et al. Analyzing the TotalSegmentator for facial feature removal in head CT scans. *Radiography*. 2025;31:372-378.

How to cite this article: O'Sullivan-Steben K, Galarneau L, Kildea J. Development of a defacing algorithm to protect the privacy of head and neck cancer patients in publicly-accessible radiotherapy datasets. *Med Phys*. 2025;52:e70160. https://doi.org/10.1002/mp.70160