Integration of Free-text Pathology Reports into Models for Automatic Contouring of Clinical Target Volume in Postoperative Head and Neck Cancer

Mu-Hung Tsai[1,2] ACCORDTSAI@GMAIL.COM
[1] *Department of Radiation Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704302, Taiwan*
[2] *Institute of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701401, Taiwan*

Hsi-Huei Lu[3] CLARE2211987@GMAIL.COM
[3] *Division of Nuclear Medicine, Department of Medical Imaging, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704302, Taiwan*
Kun-Hsing Yu[4,5,6] KUN-HSING_YU@HMS.HARVARD.EDU
[4] *Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA*
[5] *Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, USA*
[6] *Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, USA*

Yuan-Hua Wu[1] WUYH@MAIL.NCKU.EDU.TW
Yung-Jen Cheng[1] WAYNEALEX0000@HOTMAIL.COM
Tzu-Hui Pao[1] PTZUHUI@GMAIL.COM
Wei-Ting Hsueh[1] SEINE@MAIL.NCKU.EDU.TW
Irene Tai-Lin Lee[7] TAI.LIN.LEE@EMORY.EDU
[7] *Emory University School of Medicine, Atlanta, Georgia, USA*
Jung-Hsien Chiang[2] JCHIANG@MAIL.NCKU.EDU.TW

Editors: Under Review for MIDL 2026

Abstract

Background: Deep-learning-based automatic contouring systems for generating clinical target volumes (CTVs) have been proposed for several cancers; however, no studies have reported their implementation in the setting of adjuvant radiotherapy for postoperative head and neck cancer. Additionally, standard image-only deep learning models cannot utilize additional data, such as pathology reports. Therefore, we aimed to investigate model implementation and integration of pathology reports in this setting.

Methods: We extracted the data of 153 patients with head and neck cancer who underwent adjuvant radiotherapy after definitive resection. After evaluating the performance of standard image-only segmentation models, we designed a novel architecture, *TransU-Pathology-Net (TransUPNet)*, to process information from pathology reports alongside image features and evaluated its performance using the Dice similarity coefficient (DSC) and 95th percentile of the Hausdorff distance (HD95). Comparisons were made against the best-performing standard image-only baseline model.

Results: The *TransUNet* model produced the best performance among the standard image-only models and was selected as the baseline model for comparison. Incorporating pathology data via the *TransUPNet* architecture improved model performance (DSC: 0.71±0.02, HD95: 33.6±11.6 mm). The importance of pathology input was demonstrated

experimentally, showing that perturbations in pathology report data (truncation, laterality switch, and mismatch) resulted in decreased performance.

Conclusions: We developed a deep learning-based system for CTV generation for adjuvant radiotherapy of head and neck cancers. It has the unique ability to integrate free-text information from pathology reports, which improves model performance and more closely mirrors the physician workflow of integrating multiple information sources to generate CTVs.

Keywords: Head and neck cancer, radiotherapy, pathology, clinical target volume

1. Introduction

Radiation therapy is an essential component of head and neck squamous cell carcinoma treatment. Adjuvant radiotherapy is frequently required in patients who exhibit high-risk pathological features after curative surgery (Machiels et al., 2020).

Contemporary radiotherapy is based on inverse treatment planning, in which the physician carefully contours the area to be treated, which is referred to as the clinical target volume (CTV). Physicians delineate the CTV based on the risk of recurrence, which is guided by pathology reports on surgical specimens and preoperative images. Accurate contouring is a labor- and knowledge-intensive task that requires 60–150 min for each patient (Eller et al., 2015). Postoperative CTV contouring in head and neck cancer is particularly difficult for several reasons. Head and neck cancer encompasses a diverse group of regions containing multiple primary subsites, including the oral cavity, oropharynx, larynx, hypopharynx, and parotids. Surgery is generally tailored to the location of the primary tumor to ensure curative resection while minimizing morbidity, which results in vastly different anatomical defects. The obliteration of common landmarks limits the practicality of atlas-based guidelines in such scenarios. The postoperative scenario is often further complicated by flap reconstruction and postoperative edema (McCarty et al., 2019), which make delineating contours by humans difficult and nearly impossible by traditional atlas-based automatic segmentation algorithms (Ayyalusamy et al., 2019).

In recent years, remarkable progress has been made in image recognition and contextual segmentation owing to the development of deep neural networks, which have promising applications in medical imaging. Such systems have been used to successfully contour organs at risk (OARs) in the chest (Lustberg et al., 2018) and pelvis (Kazemifar et al., 2018)(Liu et al., 2020)(Men et al., 2017b). The contouring of OARs in the head and neck region has been extensively studied and has generally produced acceptable results (Chan et al., 2019)(Ibragimov and Xing, 2017)(Tong et al., 2018)(Zhu et al., 2019). Automatically generated contours are often not identical to contours delineated by experts; however, they have been demonstrated to reduce interobserver variation, resulting in clinically insignificant dose differences (van der Veen et al., 2019)(van Dijk et al., 2020)(van Rooij et al., 2019).

Recently, the application of deep neural networks in CTV contouring has been reported for rectal (Men et al., 2017b)(Larsson et al., 2018), esophageal (Jin et al., 2019), brain (Shusharina et al., 2020), and breast (Chen et al., 2020) cancers. The feasibility of deep-learning-based CTV contouring in the head and neck region has been established for nasopharyngeal (Men et al., 2017a)(He et al., 2022) and oropharyngeal (Cardenas et al., 2018a)(Cardenas et al., 2018b) cancers. However, when the performance of deep neural

networks is compared with physician delineation, their ability to generate CTV contours remains inferior to their relative success in generating OAR contours (Wong et al., 2020).

Postoperative CTV contouring, in contrast, poses a unique challenge, as the visible tumor is removed during surgery. Several researchers have investigated deep-learning-based systems for postoperative CTV generation, particularly in the case of lung, prostate, rectal, and breast cancers (Balagopal et al., 2021)(Bi et al., 2019)(Men et al., 2018)(Song et al., 2020). To the best of our knowledge, no system has been proposed specifically for CTV contouring in postoperative head and neck cancer (Matoska et al., 2024). In this study, we aimed to evaluate the performance of deep-learning segmentation models in generating the CTV for adjuvant radiotherapy in the postoperative head and neck cancer setting and to investigate the extent to which additional information from pathology reports improved model performance.

2. Methods

We retrospectively searched our institutional database for patients with head and neck cancer who underwent definitive resection and adjuvant radiotherapy at our institution. Patients with recurrent disease, those with a prior radiation history, or those receiving active treatment for other cancer sites were excluded. A total of 153 patients with various diagnoses were selected (Table 1). Simulation computed tomography (CT) and physician-delineated CTVs were extracted as ground-truth contours, and patients' surgical pathology reports were also retrieved. This study was approved by the Institutional Review Board of National Cheng Kung University Hospital (B-ER-112-204), and the requirement for informed consent was waived.

Simulation CT was performed with patients in the supine position and thermoplastic mask immobilization. The use of contrast agents was encouraged but not mandated by the institutional guidelines. All CT images were acquired using the helical scan mode on one machine (LightSpeed RT, GE Healthcare), with slice intervals ranging from 2.5–3.75 mm. All images were acquired using a uniform matrix size of 512×512; however, the field of view was varied in each patient to ensure that the entire shoulder region was captured. For this study, we cropped the CT images to the central 224×224 region after verifying that >99% of the contoured pixels were within this region. We elected to use a 2.5D approach, in which each index slice was accompanied by its neighboring cranial and caudal slices to form three channels, which allowed us to use publicly available segmentation models that expected three channels of data as input without further modification.

We first evaluated the performances of several contemporary semantic segmentation models in our task, as such approaches have proven successful in the setting of definitive radiotherapy for head and neck cancers and postoperative radiotherapy for prostate, rectal, and breast cancers. The following model architectures were evaluated: *SegFormer* (Xie et al., 2021), *DPT* (Ranftl et al., 2021), *UPerNet* (Xiao et al., 2018), *BEiT* (Bao et al., 2021), and *TransUNet* (Chen et al., 2021). All models were trained for 100 epochs using an RTX8000 with 48 GB of memory. Image augmentation was performed with random scaling, translation, and rotation.

To integrate information from the pathology reports, we modified *TransUNet* as our segmentation backbone. Briefly, *TransUNet* is an image segmentation model that performs

Table 1: Baseline characteristics of patients included in this study (n = 153)

Characteristics	Number (%)
Sex	
Male	145 (94.8)
Female	8 (5.2)
Diagnosis / Subsite	
Hypopharyngeal cancer	39 (25.5)
Tongue cancer	36 (23.5)
Buccal cancer	31 (20.3)
Gingival cancer	13 (8.5)
Oropharyngeal cancer	11 (7.2)
Laryngeal cancer	9 (5.9)
Mouth floor cancer	6 (3.9)
Oral cavity cancer, not otherwise specified	4 (2.6)
Salivary gland cancer	4 (2.6)
Pathological T classification	
T1	14 (9.2)
T2	44 (28.8)
T3	28 (18.3)
T4	67 (43.8)
Pathological N classification	
N0	43 (28.1)
N1	11 (7.2)
N2	86 (56.2)
N3	13 (8.5)

well with medical images. It is unique because it combines the strengths of transformers and convolutional neural network (CNN) architectures; the transformer components provide global modeling capabilities, whereas CNNs retain strong localization abilities. At the heart of the network, the transformer encoder receives tokenized image patches from the upstream CNN feature map as input and performs global contextual attention, thereby overcoming the shortcomings of the limited receptive fields posed by CNNs.

We extended the capabilities of *TransUNet* to incorporate information from pathology reports and named it *TransU-Pathology-Net (TransUPNet)*. Because the format of pathology reports varies widely between institutions and is often not structured, we needed a versatile approach to extract information from the reports. We employed text embedding models, which can convert textual data into vector representations called embeddings. These embeddings capture the semantic meaning of text in a high-dimensional vector space. We used pre-trained text-embedding models to convert the free-text pathology reports and then inserted the embeddings alongside the tokenized image patches into the transformer backbone (Figure 1). This allows the attention mechanism of the transformer backbone to consider the features extracted from each image patch, along with the information derived from the pathology report.
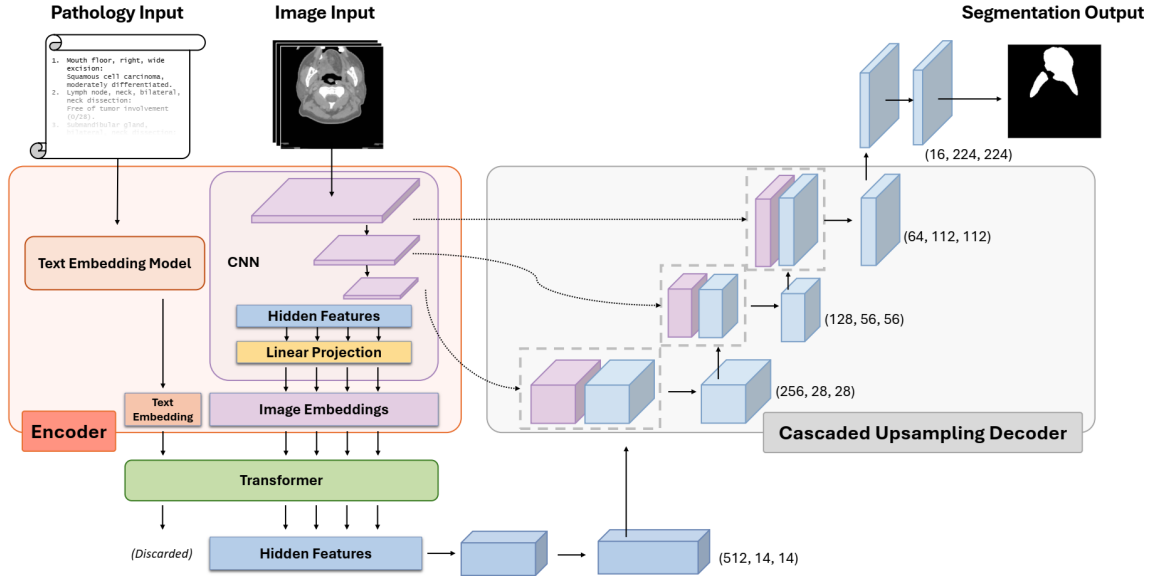

Figure 1: The architecture of our pathology-aware *TransU-Pathology-Net* model. We modified the encoder pathway to incorporate text and image information simultaneously, whereas the decoder pathway remained identical to that of *TransUNet*. Layer dimensions are denoted as *(channel, width, height)*.

With this design, it is evident that the quality of the embeddings (how well they capture the semantic information from the pathology report) is of utmost importance. While maintaining the consistency of the model architecture, we benchmarked the performance of our

model using different text-embedding models. We used pre-trained embedding models to create pathology embeddings without further fine-tuning. Using the *SentenceTransformers* library (Reimers and Gurevych, 2019), we evaluated an open-source general-purpose model (*mpnet*) (HuggingFace, 2024), an open-source model specifically pre-trained on medical domain texts (Deka et al., 2022)(Deka, 2022) (*Bio*), and two closed-source embedding models, text-embedding-3-small (*small*) and text-embedding-3-large (*large*), from OpenAI (OpenAI, 2024). We also designed two pseudo-embeddings that corresponded to the absence of pathology information as the baseline: one with all values equal to zero (*zero*) and another where values were filled in with random numbers (*random*). Identical text-embedding models were used for training and validation in each experiment.

We evaluated the performance of our network using the Dice similarity coefficient (DSC) and the 95th percentile of the Hausdorff distance (HD95). The DSC is a measure that quantifies volumetric overlap, with values ranging from 0 (no overlap) to 1 (complete overlap); HD95 is a robust representation of the overall boundary distance, which is less sensitive to outliers than the maximum Hausdorff distance. To calculate the HD95, we assumed a uniform voxel spacing of 1 mm in the axial plane and 3.75 mm in the craniocaudal direction. Metrics were reported across five-fold cross-validation and were statistically compared with the *TransUNet* baseline values using the Wilcoxon rank-sum test, with an alpha level of 0.05.

3. Results

First, we evaluated the performance of a standard image-only segmentation algorithm (Figure 2). The *TransUNet* model yielded the best performance with a DSC of 0.68 ± 0.02 and HD95 of 42.5 ± 13.8 mm (Table 2). Therefore, we selected this as the baseline model that could be improved upon and used for further statistical comparisons.
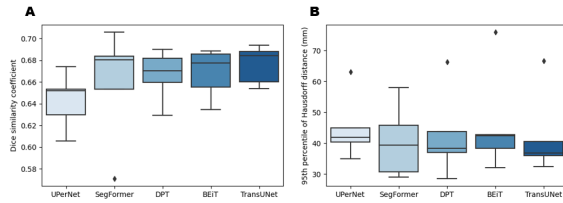

Figure 2: Boxplot using standard image-only segmentation models. Boxplot of (a) dice similarity coefficient (DSC) and (b) 95th percentile Hausdorff distance (HD95) across five-fold validation using standard image-only segmentation models.

Next, we investigated whether the performance was improved by incorporating pathology information and whether the choice of different text-embedding models affected the performance of our system (Figure 3). Compared with baseline *TransUNet*, the incorporation of pathology information resulted in a higher DSC and lower HD95 across all models; however, it was only statistically significant when using text-embedding-3-large embedding, which performed best (DSC = 0.71 ± 0.02 and HD95 = 33.6 ± 11.6 mm). The

Table 2: Model performance using different models in five-fold cross validation.

Model Type	Model	Text Embedding Model	DSC	HD95 (mm)
Standard image-only	UPerNet	-	0.64±0.03	45.0±10.7
	SegFormer	-	0.66±0.05	40.6±11.9
	DPT	-	0.67±0.02	42.8±14.2
	BEiT	-	0.67±0.02	46.3±17.1
	TransUNet	-	0.68±0.02	42.5±13.8
Pathology-aware	TransUPNet	*random*	0.60±0.02	53.7±12.0
		zero	0.68±0.02	42.5±13.1
		mpnet	0.70±0.02	39.4±11.9
		Bio	0.69±0.02	33.8±9.0
		text-embedding-3-small	0.69±0.02	34.9±10.6
		text-embedding-3-large	0.71±0.02*	33.6±11.6

*Difference is statistically significant ($p<0.05$) compared to *TransUNet*.
Abbreviations: DSC, Dice similarity coefficient; HD95, 95th percentile of Hausdorff distance.

open-source models achieved comparable results while avoiding the downside of uploading potentially sensitive information to third parties. The text-embedding model pre-trained on biomedicine-specific data also produced a numerically higher DSC and lower HD95 than the general-purpose model.
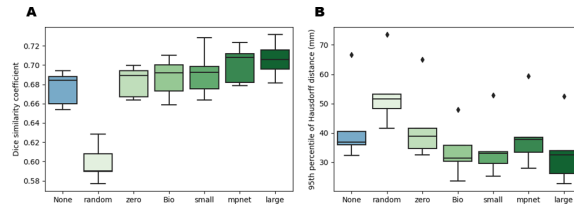

Figure 3: Boxplot using different text embedding models. Boxplot of (a) dice similarity coefficient (DSC) and (b) 95th percentile Hausdorff distance (HD95) across five-fold validation using different text embedding models.

Subsequently, we empirically demonstrated that the increase in model performance was due to the information derived from pathology reports. In this experiment, the models were trained with embeddings from the full report. However, during the inference of the validation set, we intentionally truncated the pathology report to 20%, 40%, 60%, and 80% of its original length. We observed a relative gain in performance as larger portions of the pathology report were made available to the model (Figure 4A). The importance of the information encoded in pathology report embeddings can also be demonstrated by deliberately feeding incorrect or incomplete information into a model. We modified the

pathology report in the following ways: laterality switch, where "right" was changed to "left" and "left" was changed to "right," and mismatch, where an unrelated pathology report from another patient was supplied instead of the original. These perturbations in the data from the pathology report severely disrupted model performance (Figure 4B).
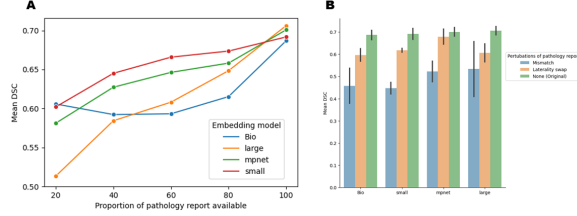

Figure 4: (A) Performance in terms of DSC when only a proportion of pathology reports was available to the model. (B) Performance in terms of DSC when perturbations of the pathology report were performed

We could also directly examine embeddings derived from pathology reports to demonstrate that the embeddings capture vital information in their representation. We employed t-distributed stochastic neighbor embedding in the pathology embeddings to reduce their dimensionality to two for visualization (Figure 5). The embeddings formed distinct clusters between the primary tumor subsites; however, the separation between tumor laterality was less evident. This suggests that the embeddings encoded the primary subsite information from the pathology reports and utilized this information in the downstream task of CTV contouring.
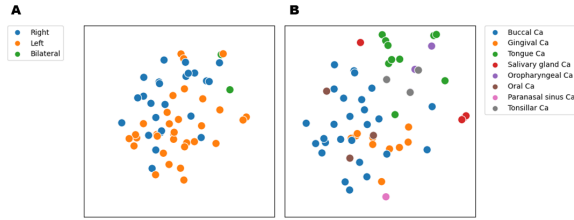

Figure 5: Visualization of pathology report embeddings. Embeddings were reduced to two dimensions via t-distributed stochastic neighbor embedding and labelled in different colors according to (a) primary tumor subsite and (b) laterality.

Finally, we demonstrated the effects of the different pathology embeddings through illustrative examples (Figure 6). The patient in the top row had cancer originating from the right floor of the mouth. After resection and neck dissection, the final diagnosis of a pT3 tumor and one metastatic lymph node with extranodal extension was reached. Bilateral neck drainage from the mouth floor and elevated risk due to the extranodal extension prompted bilateral neck irradiation, which the image-only *TransUNet* failed to delineate.

The patient in he middle row had tongue cancer, and the primary surgical bed was not adequately covered by *TransUNet*. Finally, the patient in the bottom row underwent resection and unilateral neck dissection for left-sided buccal cancer; the contralateral side was spared after confirming negative involvement of the Ia region. The large embedding successfully captured and utilized this information, whereas the *Bio* and *mpnet* embeddings produced similar outputs to those produced by the image-only *TransUNet*.
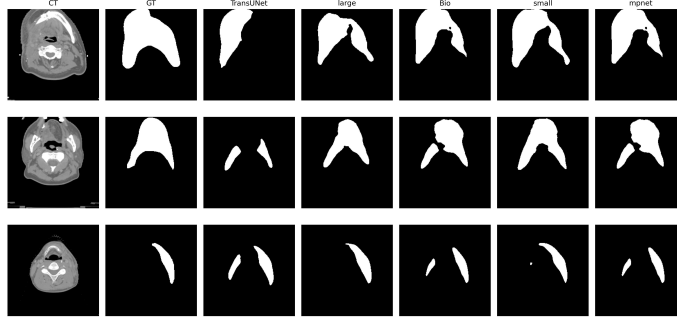

Figure 6: Qualitative examination of the clinical target volume. Qualitative examination of the clinical target volume generated by image-only *TransUNet* and pathology-aware *TransU-Pathology-Net* with different text embedding models.

4. Discussion

In this study, we evaluated the performance of several state-of-the-art segmentation models for generating a CTV for adjuvant radiotherapy in a postoperative head and neck cancer setting. We also developed a pathology-aware deep-learning neural network architecture with the novel ability to incorporate relevant information extracted from free-text pathology reports. Our study showed that *TransUPNet* outperforms standard image-only segmentation algorithms and experimentally demonstrated that it truly incorporates pathological information during CTV generation.

Image-only deep learning algorithms have been successful in CTV generation in definitive radiotherapy for nasopharyngeal (He et al., 2022) and oropharyngeal cancers (Cardenas et al., 2018a); however, we found that the inclusion of pathology information improved the performance in the postoperative setting. The postoperative setting differs from the definitive setting in that pathological lesions are absent on postoperative simulation CT examinations. This is not necessarily a critical limitation, as image-only deep learning algorithms have proven successful in postoperative radiotherapy for lung, prostate, rectal, and breast cancers (Balagopal et al., 2021)(Bi et al., 2019)(Men et al., 2018)(Song et al., 2020). However, it is worth noting that the postoperative CTV contours in these cancers are generally uniform, irrespective of surgical pathological findings. This is exemplified in a rectal cancer study conducted by Song et al. (Song et al., 2020), in which the authors excluded patients with T4b or lymph node metastasis from their sample because these

risk factors would lead to different contours (inclusion of external iliac nodes) according to clinical guidelines. In contrast, the CTVs in postoperative head and neck cancer vary widely depending on the original tumor subsite and laterality, which may explain the performance gain on incorporating pathology reports.

Despite these gains in performance, we recognize that the absolute performance with a DSC of 0.71 falls short of the performances reported in various published papers on automatic CTV segmentation. However, we believe that this number should be considered in the context of postoperative head and neck cancer. Contouring is a particularly challenging task in postoperative head and neck cancer, with wide variation among physicians; however, quantitative studies are scarce in the literature. One study found significant inter-observer variability between radiation oncologists, with a mean DSC of 0.56 and a range of 0.03–0.91; the large range suggests considerable variation in DSC between cases (Ng et al., 2018). Another indirect comparison that could serve as a reference is reported in the study performed by the Danish Head and Neck Cancer group, which investigated interobserver variations in OARs (Nielsen et al., 2024). In their study, when the left buccal mucosa, right buccal mucosa, and oral cavity were contoured as OARs in the definitive setting, expert contour consensus (as measured by mean DSC) was only 0.48, 0.48, and 0.83, respectively. In our study, 29% of the patients had a tumor originating from the buccal or gingival area, while 24% had a tumor originating from the tongue; the postoperative CTV usually encompasses the surgically altered oral cavity area, reconstruction flaps, and involved nodal regions, which further increases the difficulty compared with contouring OARs in a definitive setting.

An intriguing finding of our experiments is that when pathology-aware models are supplied with incorrect pathology information, their performance is often worse than that of standard image-only segmentation models. One explanation for this is that although image-only models can only derive information from images, pathology-aware models have learned to rely on pathology embeddings as an accurate information source. This observation may have important implications for scenarios in which suboptimal text-embedding models are used, thereby providing insufficient information to the model. The limitations of this study include its single-center retrospective design and the use of two-dimensional images as input. All the pathology reports in this study were structured similarly, although we derived the embeddings in a free-text format and did not use structured data elements. We believe that the pathology embedding process should be generalizable to reports from other institutions, even if their report structure or language is dissimilar to ours. Using a full three-dimensional architecture may further improve the performance, with a tradeoff of higher computational costs during training and inference.

5. Conclusions

We developed an automatic contouring system for generating CTV in postoperative head and neck cancers with the unique ability to integrate free-text information from pathology reports. Incorporating pathology information improved the performance of the system, and this approach could be extended to any data in free-text format.

Acknowledgments

References

A. Ayyalusamy, S. Vellaiyan, S. Subramanian, A. Ilamurugu, S. Satpathy, M. Nauman, G. Katta, and A. Madineni. Auto-segmentation of head and neck organs at risk in radiotherapy and its dependence on anatomic similarity. *Radiation Oncology Journal*, 37: 134–142, 2019. doi: 10.3857/roj.2019.00038.

A. Balagopal, D. Nguyen, H. Morgan, Y. Weng, M. Dohopolski, M. H. Lin, A. S. Bark-ousaraie, Y. Gonzalez, A. Garant, N. Desai, R. Hannan, and S. Jiang. A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy. *Medical Image Analysis*, 72: 102101, 2021. doi: 10.1016/j.media.2021.

H. Bao, L. Dong, S. Piao, and F. Wei. Beit: Bert pre-training of image transformers, 2021.

N. Bi, J. Wang, T. Zhang, X. Chen, W. Xia, J. Miao, K. Xu, L. Wu, Q. Fan, L. Wang, Y. Li, Z. Zhou, and J. Dai. Deep learning improved clinical target volume contouring quality and efficiency for postoperative radiation therapy in non-small cell lung cancer. *Frontiers in Oncology*, 9:1192, 2019. doi: 10.3389/fonc.2019.01192.

C. E. Cardenas, B. M. Anderson, M. Aristophanous, J. Yang, D. J. Rhee, R. E. McCarroll, A. S. R. Mohamed, M. Kamal, B. A. Elgohari, H. M. Elhalawani, C. D. Fuller, A. Rao, A. S. Garden, and L. E. Court. Auto-delineation of oropharyngeal clinical target volumes using 3d convolutional neural networks. *Physics in Medicine and Biology*, 63:215026, 2018a. doi: 10.1088/1361-6560/aae8a9.

C. E. Cardenas, R. E. McCarroll, L. E. Court, B. A. Elgohari, H. Elhalawani, C. D. Fuller, M. J. Kamal, M. A. M. Meheissen, A. S. R. Mohamed, A. Rao, B. Williams, A. Wong, J. Yang, and M. Aristophanous. Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in dice similarity coefficient parameter optimization function. *International Journal of Radiation Oncology Biology Physics*, 101: 468–478, 2018b. doi: 10.1016/j.ijrobp.2018.01.114.

J. W. Chan, V. Kearney, S. Haaf, S. Wu, M. Bogdanov, M. Reddick, N. Dixit, A. Sudhyad-hom, J. Chen, S. S. Yom, and T. D. Solberg. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. *Medical Physics*, 46:2204–2213, 2019. doi: 10.1002/mp.13495.

J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation, 2021.

X. Chen, K. Men, B. Chen, Y. Tang, T. Zhang, S. Wang, Y. Li, and J. Dai. Cnn-based quality assurance for automatic segmentation of breast cancer in radiotherapy. *Frontiers in Oncology*, 10:524, 2020. doi: 10.3389/fonc.2020.00524.

P. Deka. Biobert-mnli-snli-scinli-scitail-mednli-stsb. https://huggingface.co/pritamdeka/BioBERT-mnli-snli-scinli-scitail-mednli-stsb, 2022. Accessed: 2024-04-23.

P. Deka, A. Jurek-Loughrey, and P. Deepak. Evidence extraction to validate medical claims in fake news detection. In A. Traina, H. Wang, Y. Zhang, S. Siuly, R. Zhou, and L. Chen, editors, *Health Information Science*, pages 3–15. Springer, 2022. doi: 10.1007/978-3-031-20627-6_1.

Y. Eller, N. Klass, M. Schmuecking, O. Elicin, R. Bigler, J. Tille, S. Fankhauser, N. Mertineit, B. Kleaser, and A. Geretschlaeger. Are contouring time and multimodality imaging prognostic factors for radiation therapy of advanced head and neck cancer? *Cancer Imaging*, 15:P28, 2015. doi: 10.1186/1470-7330-15-S1-P28.

Y. He, S. Zhang, Y. Luo, H. Yu, Y. Fu, Z. Wu, X. Jiang, and P. Li. Quantitative comparisons of deep-learning-based and atlas-based auto-segmentation of the intermediate risk clinical target volume for nasopharyngeal carcinoma. *Current Medical Imaging*, 18:335–345, 2022. doi: 10.2174/1573405617666210827165031.

HuggingFace. all-mpnet-base-v2. https://huggingface.co/sentence-transformers/all-mpnet-base-v2, 2024. Accessed: 2024-04-23.

B. Ibragimov and L. Xing. Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks. *Medical Physics*, 44:547–557, 2017. doi: 10.1002/mp.12045.

D. Jin, D. Guo, T. Y. Ho, A. P. Harrison, J. Xiao, C. Tseng, and L. Lu. Deep esophageal clinical target volume delineation using encoded 3d spatial context of tumors, lymph nodes, and organs at risk. In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, volume 11769, pages 603–612. Springer, 2019. doi: 10.1007/978-3-030-32226-7_67.

S. Kazemifar, A. Balagopal, D. Nguyen, S. McGuire, R. Hannan, S. Jiang, and A. Owrangi. Segmentation of the prostate and organs at risk in male pelvic ct images using deep learning. *Biomedical Physics & Engineering Express*, 4:055003, 2018. doi: 10.1088/2057-1976/aad100.

R. Larsson, J.-F. Xiong, Y. Song, Ling-Fu, Y.-Z. Chen, X. Xiaowei, P. Zhang, and J. Zhao. Automatic delineation of the clinical target volume in rectal cancer for radiation therapy using three-dimensional fully convolutional neural networks. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5898–5901, 2018. doi: 10.1109/EMBC.2018.8513506.

Z. Liu, X. Liu, B. Xiao, S. Wang, Z. Miao, Y. Sun, and F. Zhang. Segmentation of organs-at-risk in cervical cancer ct images with a convolutional neural network. *Physica Medica*, 69:184–191, 2020. doi: 10.1016/j.ejmp.2019.12.008.

T. Lustberg, J. van Soest, M. Gooding, D. Peressutti, P. Aljabar, J. van der Stoep, W. van Elmpt, and A. Dekker. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiotherapy and Oncology*, 126:312–317, 2018. doi: 10.1016/j.radonc.2017.11.012.

J. P. Machiels, C. René Leemans, W. Golusinski, C. Grau, L. Licitra, and V. Gregoire. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: Ehns–esmo–estro clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 31:1462–1475, 2020. doi: 10.1016/j.annonc.2020.07.011.

T. Matoska, M. Patel, H. Liu, and S. Beriwal. Review of deep learning based autosegmentation for clinical target volume: current status and future directions. *Advances in Radiation Oncology*, 9:101470, 2024. doi: 10.1016/j.adro.2024.101470.

J. L. McCarty, A. S. Corey, M. W. El-Deiry, H. M. Baddour, B. M. Cavazuti, and P. A. Hudgins. Imaging of surgical free flaps in head and neck reconstruction. *American Journal of Neuroradiology*, 40:5–13, 2019. doi: 10.3174/ajnr.A5776.

K. Men, X. Chen, Y. Zhang, T. Zhang, J. Dai, J. Yi, and Y. Li. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Frontiers in Oncology*, 7:315, 2017a. doi: 10.3389/fonc.2017.00315.

K. Men, J. Dai, and Y. Li. Automatic segmentation of the clinical target volume and organs at risk in the planning ct for rectal cancer using deep dilated convolutional neural networks. *Medical Physics*, 44:6377–6389, 2017b. doi: 10.1002/mp.12602.

K. Men, T. Zhang, X. Chen, B. Chen, Y. Tang, S. Wang, Y. Li, and J. Dai. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Physica Medica*, 50:13–19, 2018. doi: 10.1016/j.ejmp.2018.05.006.

S. P. Ng, B. A. Dyer, J. Kalpathy-Cramer, A. S. R. Mohamed, M. J. Awan, G. B. Gunn, J. Phan, M. Zafereo, J. M. Debnam, C. M. Lewis, R. R. Colen, M. E. Kupferman, N. Guha-Thakurta, G. Canahuate, G. E. Marai, D. Vock, B. Hamilton, J. Holland, C. E. Cardenas, S. Lai, D. Rosenthal, and C. D. Fuller. A prospective in silico analysis of interdisciplinary and interobserver spatial variability in post-operative target delineation of high-risk oral cavity cancers: does physician specialty matter? *Clinical and Translational Radiation Oncology*, 12:40–46, 2018. doi: 10.1016/j.ctro.2018.07.006.

C. P. Nielsen, E. L. Lorenzen, K. Jensen, J. G. Eriksen, J. Johansen, N. Gyldenkerne, R. Zukauskaite, M. Kjellgren, C. Maare, C. K. Lønkvist, K. Nowicka-Matus, W. M. Szejniuk, M. Farhadi, Z. Ujmajuridze, K. Marienhagen, T. S. Johansen, J. Friborg, J. Overgaard, and C. R. Hansen. Interobserver variation in organs at risk contouring in head and neck cancer according to the dahanca guidelines. *Radiotherapy and Oncology*, 197:110337, 2024. doi: 10.1016/j.radonc.2024.110337.

OpenAI. Embeddings documentation. https://platform.openai.com/docs/guides/embeddings, 2024. Accessed: 2024-04-23.

R. Ranftl, A. Bochkovskiy, and V. Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12159–12168, 2021. doi: 10.1109/ICCV48922.2021.01196.

N. Reimers and I. Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 3982–3992, Hong Kong, China, 2019. Association for Computational Linguistics.

N. Shusharina, J. Söderberg, D. Edmunds, F. Löfman, H. Shih, and T. Bortfeld. Automated delineation of the clinical target volume using anatomically constrained 3d expansion of the gross tumor volume. *Radiotherapy and Oncology*, 146:37–43, 2020. doi: 10.1016/j.radonc.2020.01.028.

Y. Song, J. Hu, Q. Wu, F. Xu, S. Nie, Y. Zhao, S. Bai, and Z. Yi. Automatic delineation of the clinical target volume and organs at risk by deep learning for rectal cancer postoperative radiotherapy. *Radiotherapy and Oncology*, 145:186–192, 2020. doi: 10.1016/j.radonc.2020.01.020.

N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Medical Physics*, 45:4558–4567, 2018. doi: 10.1002/mp.13147.

J. van der Veen, S. Willems, S. Deschuymer, D. Robben, W. Crijns, F. Maes, and S. Nuyts. Benefits of deep learning for delineation of organs at risk in head and neck cancer. *Radiotherapy and Oncology*, 138:68–74, 2019. doi: 10.1016/j.radonc.2019.05.010.

L. V. van Dijk, L. Van den Bosch, P. Aljabar, D. Peressutti, S. Both, R. J. H. M. Steenbakkers, J. A. Langendijk, M. J. Gooding, and C. L. Brouwer. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiotherapy and Oncology*, 142:115–123, 2020. doi: 10.1016/j.radonc.2019.09.022.

W. van Rooij, M. Dahele, H. Ribeiro Brandao, A. R. Delaney, B. J. Slotman, and W. F. Verbakel. Deep learning-based delineation of head and neck organs at risk: geometric and dosimetric evaluation. *International Journal of Radiation Oncology Biology Physics*, 104:677–684, 2019. doi: 10.1016/j.ijrobp.2019.02.040.

J. Wong, A. Fong, N. McVicar, S. Smith, J. Giambattista, D. Wells, C. Kolbeck, J. Giambattista, L. Gondara, and A. Alexander. Comparing deep learning-based autosegmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiotherapy and Oncology*, 144:152–158, 2020. doi: 10.1016/j.radonc.2019.10.019.

T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun. Unified perceptual parsing for scene understanding. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV*, volume 11209, pages 432–448. Springer, 2018. doi: 10.1007/978-3-030-01228-1_26.

E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems*, volume 34, pages 12077–12090, 2021.

W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie. Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical Physics*, 46:576–589, 2019. doi: 10.1002/mp.13300.