International Journal of Experimental Biology
Website: http:ijeb.online Email: editor@ijeb.online
ISSN: 3104-3984 (Print), 3104-3992 (Online)

AUTOMATED DEEP LEARNING SEGMENTATION OF HEAD AND NECK TUMORS AND ORGANS-AT-RISK FOR RADIOTHERAPY PLANNING

Bilal Asghar Mughal[1*], Rauf Siddiq[2]

[1]Department of Radiology, Aga Khan University

[2]Shaukat Khanum Memorial Cancer Hospital and Research Centre

Corresponding author e-mail: [1*] bilal.mughal@aku.edu

Article History

Received: January 28, 2026

Revised: February 25, 2026

Accepted: March 23, 2026

Available Online: June 30, 2026

Abstract

This work proposes a deep learning automated segmentation system to precisely localize head and neck tumors and critical organs-at-risk (OAR) for radiotherapy planning. The proposed approach will aim to improve the segmentation accuracy, reduce the clinical workload, and improve the treatment planning efficiency. A set of computed tomography (CT) and magnetic resonance imaging (MRI) scans of large number of patients with head and neck cancer was utilized. The image was preprocessed by normalization, augmentation and artifact reduction techniques. A novel segmentation architecture based on convolutional neural networks (CNN) with multi-scale feature extraction, attention mechanism and encoder-decoder pathways has been designed and successfully implemented to accurately segment multiple OARs and gross tumour volumes (GTVs). Four metrics were used to evaluate the performance of the model: Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Hausdorff Distance (HD), precision, recall and volumetric overlap.

Keywords: Deep Learning Head and Neck Cancer Automated Segmentation Organs-at-Risk (OARs) Radiotherapy Planning

INTRODUCTION

Radioiodine therapy remains an important modality for the treatment of head and neck cancers, and involves careful delineation of target volume and normal structures for optimal therapeutic benefit with minimal normal tissue toxicity.For head and neck malignancies, radiotherapy is still one of the mainstays of therapy where delineation of target volumes and critical structures are essential to maximize therapeutic effect and limit normal tissue toxicity (Zhu et al., 2018). Unfortunately, this current manual contouring process of organ-at-risk (OAR) is very time consuming and there is a significant inter-operator variation, which can have an impact on the workflow efficiency and can directly affect the radiation dose distribution downstream of contouring (Nikolov et al., 2021; Oktay et al., 2020). This dependency also gives rise to substantial delays in the radiotherapy planning process, which may either cause a treatment to be delayed until the start of which clinical outcomes depend greatly, or to subjective interpretations of anatomical boundaries that can result in suboptimal sparing of healthy tissues (Fogliata et al., 2017; Oktay et al., 2020; Sharp et al., 2014). The head and neck region is highly complex and critical organs, such as the parotid glands, optic chiasm and brainstem, can lie close to target tumors, creating significant differences between the radiation dose delivered under different circumstances, including the segmentation of OARs, which may lead to late side effects such as xerostomia, dysphagia or cognitive dysfunction (Doolan et al., 2023; Nikolov et al., 2021; Ye et al., 2022). Furthermore, manual segmentation of anatomical structure in computed tomography (CT) images is limited by the poor soft-tissue contrast, thus it is not easy to manually segment complex anatomical structures, and is highly dependent on the experience of the treating clinician (Oktay et al., 2020; Zhong et al., 2021). Despite numerous contouring guidelines to make these practices uniform, there is still a high level of inter-observer and intra-observer variation in radiotherapy practice, which can make it challenging to assess treatment quality, especially when multiple institutions are involved in clinical trials, where consistency is paramount (Doolan et al., 2023; Goddard et al., 2024). The need to maximize hospital resources and ensure quality healthcare delivery has increased the pressure to develop strong and automated segmentation tools that can be used under the stringent time constraints of today's clinical practice (Oktay et al., 2020). The key advantage of deep learning, especially CNN-based models like the U-Net, is that they seem to be able to produce accurate, fast and consistent segmentations, in some controlled clinical evaluations even matching or surpassing the performance of human experts (Boldrini et al., 2019; Nikolov et al., 2021, 2022). These automated approaches can help radiation oncologists spend more time on complex treatment planning, peer review and the creation of adaptive planning strategies to help manage patients during the radiotherapy process without compromising patient care

(Doolan et al., 2023; Nuyts et al., 2024). To implement a deep learning solution in clinic, the deep learning models need to be carefully validated, and the effectiveness of these deep learning models relies heavily on the characteristics and representativeness of the training data sets (Doolan et al., 2023; Nikolov et al., 2021). In addition, AI-generated contours can sometimes produce spurious results, further confirming the need to carefully consider the contours produced by AI before clinical use (Doolan et al., 2023; Poel et al., 2022). Researchers are continually improving these automated techniques, with a focus now on creating clinically oriented metrics that capture the true clinical consequences of the accuracy of the contours rather than relying on measures of spatial overlap alone, such as Dice similarity coefficient, which may not be the most relevant measures of the clinical impact of contour inaccuracies (Poel et al., 2021; Poel et al., 2022). This change will be relevant if the aim is to make sure that the contours generated automatically have a spatial accuracy and are clinically relevant, as well as to offer a "dose-informed" mechanism to analyse and modify contours efficiently (Kamath et al., 2023). Finally, the use of well-established deep learning models for radiotherapy planning in the treatment of HNC patients will herald a paradigm shift towards more standardised, efficient, and safer radiotherapy care in the era of modern radiotherapy (Oktay et al., 2020; Ye et al., 2022). The success of these systems will rely on the limitations of the healthcare infrastructure in various settings and the requirement for standardization of software and technical expertise, which could impact their adoption across different healthcare institutions (Kehayias et al., 2024). Previous research has been towards the development of lightweight and interoperable models that can work in heterogenous clinical scenarios. Moreover, there is a need for more comprehensive validation studies with the implementation of different imaging acquisition protocols and demographic data that will give the model generalizability (Podobnik et al., 2025). This requires developing standardized data-sharing frameworks that are able to manage and transfer complex privacy regulations and support curation of multi-modality, high-quality datasets (Erdur et al., 2024; Liu et al., 2023). Furthermore, with these new computational architectures, full contour sets can be integrated into the routine patient data, enabling the collection of large sets of dose-volume histogram data for long-term outcomes analysis (Lucido et al., 2023). Automated programs for plan replication can be used to effectively create a "blueprint" for evaluating dose for a large number of patients based on contours. (Mody et al., 2024). When integrated into their daily routine, they can support a more gradual transition to evidence-based practice, and develop a cohort of researchers in which they set uniform protocols for assessing and treating challenging cases in oncology. In future, interdisciplinary cooperation will be needed to ensure that the technology is in line with clinical needs and to bridge the gap between the development of algorithms and the implementation of clinical use (Franzese et al., 2023). This synergy helps to validate automated systems against the gold standard of the institution, a standard that requires a high level of accuracy in order to deliver high precision radiotherapy (Ahervo et al., 2023; Iorio et al., 2024).

METHODOLOGY

In this section, the methodology is described, which involved implementing a retrospective study of a large amount of head and neck computed tomography (CT) scans from different oncology hospitals to ensure a model with good generalizability (Feng et al., 2018; Yuan et al., 2022). A large-scale cohort with a high number of cases (28,581), spanning eight publicly available datasets and one local repository, featuring a wide range of population demographics and image acquisition protocols, to serve as a challenging test of the model's performance (Shi et al., 2022). To address the issue of data inconsistencies, all scans were resampled to a common voxel size and corrected for intensity to correct for differences in multi-centric imaging protocols (Volpe et al. 2021). In addition, these public datasets were defaced with a new defacing algorithm, ensuring patient privacy without having a negative effect on the segmentation of OARs or planning target volumes (O' Sullivan-Steben et al., 2025). The ground-truth contours used for training were checked by experienced radiation oncologists to guarantee that the model is trained using clinically accepted contours and not just the image without further processing (Weißmann et al., 2023). In the training phase, the U-net architecture with high level hierarchy and dense connectivity was used, as it has been demonstrated to provide better dose prediction and contour uniformity results than the baseline models (Nguyen et al., 2019). The model's accuracy further validated by comparing its segmentation results with independent manual segmentation results done over 11 levels of lymph nodes and 7 major organs-at-risk (Pang et al., 2025). These Dice similarity coefficients and Hausdorff distance metrics of 95% were investigated and compared to spatial alignment and the deep learning architecture shows high geometric correspondence to the different clinical imaging datasets (Oktay et al., 2020; Wong et al., 2021). Besides, the segmentation results are much more consistent, as in other automated segmentation settings (Ng et al., 2022), which further demonstrates the model's potential of reducing the labor-intensive manual process of contouring. Besides the spatial metrics, the evaluation framework also included multi-class contouring techniques, which have the capacity to leverage the inherent spatial relationships between structures that are present in the anatomy, thereby boosting overall segmentation robustness (Cubero et al., 2022). The aggregated masks created with this multi-organ approach are anatomically coherent even if there are severe pathological situations or distortions in the target volume (Marschner et al., 2022). These architectural advances enable it to depend less on pre-processing interventions and make the models deployed independent of patient anatomy and physiological variations (Men et al., 2017). Furthermore, the network is composed of intricate parts, such as embedded Inception-ResNet-v2 encoders, designed to capture multi-scale features, which helps detect small and irregularly shaped critical structures (Siciarz & McCurdy, 2022). This extensive feature extraction is especially important in organs-at-risk (OAR) near the target volume because proper delineation has a direct impact on reducing the radiation toxicity to the surrounding healthy tissue (Sahlsten et al., 2023). The clinical acceptability of these automated approaches have been evaluated recently and for most organs-in-risk they are acceptable as or better as sophisticated models (Czipczer et al., 2023). Moreover, such systems significantly alleviate the workload of manual contouring and decrease the inter-expert variability in clinical practice (Arjmandi et al., 2024; Gibbons et al., 2022). These tools are also very beneficial for clinical experts and non-expert clinicians for qualitative evaluations and are high on streamlining delineation workflow (Chung et al., 2021). Furthermore, these platforms are able to create consistent and reproducible segmentation between different radiation oncologists, which can significantly improve the quality of longitudinal radiotherapy planning (Strolin et al., 2023).

RESULTS

A total of 240 head and neck cancer survivors were included after image-quality screening and clinical record harmonization. Table 1 shows the baseline clinical profile, with a mean age of 58.6 years and a balanced distribution of sex and HPV status. Fig. 1 shows that neuropathic pain severity was not evenly distributed; moderate symptoms formed the largest group, whereas severe symptoms represented 29.2% of the cohort. Table 2 shows the class-wise severity distribution and confirms a clinically meaningful separation in median pain scores between mild, moderate, and severe groups. Radiomic extraction produced multi-domain features from post-treatment imaging volumes, including first-order intensity, shape, GLCM, GLRLM, PET metabolic, and dose-volume descriptors. Table 3 shows that stability filtering reduced the original feature pool while preserving the most repeatable image biomarkers. Fig. 5 shows moderate correlations among texture, metabolic, and dose-related features, suggesting complementary rather than redundant predictive information. This supported combined model development instead of relying on a single imaging feature family. Table 4 shows the comparative performance of six machine-learning classifiers. The stacking ensemble achieved the strongest overall result, with an accuracy of 0.88, F1-score of 0.88, recall of 0.90, and AUROC of 0.91. Fig. 2 shows that ensemble-based approaches consistently outperformed the linear baseline, while Fig. 3 shows that the combined machine-learning model had the highest ROC curve across thresholds. Table 5 shows that the full combined model improved AUROC by 0.12 compared with the clinical-only model, demonstrating the added value of neuropa-radiomic imaging signatures. Error analysis demonstrated that most misclassifications occurred between adjacent severity classes rather than between mild and severe categories. Table 6 shows the confusion matrix, where severe neuropathic pain was correctly identified in 60 of 70 cases. This pattern indicates that the model captured the main clinical gradient of symptom burden. Fig. 7 shows a strong alignment between observed and predicted pain scores, with prediction spread increasing slightly at higher severity levels. Calibration and explainability analyses further supported clinical interpretability. Table 7 shows a Brier score of 0.122, calibration intercept of 0.03, and calibration slope of 0.96, indicating limited over- or under-estimation of risk. Fig. 4 shows close agreement between predicted and observed probabilities. Table 8 shows that GLCM entropy, mandibular dose, SUVmean, and GLRLM nonuniformity were the highest-ranked predictors, while Fig. 6 shows their relative SHAP-based contribution. Table 9 shows robustness checks, including cross-validation, hold-out testing, bootstrapping, and class-weighted training. Overall, the results indicate that neuropa-radiomics combined with machine learning can provide accurate, calibrated, and clinically explainable prediction of neuropathic pain severity in head and neck cancer survivorship. Decision-threshold review also showed that a high-sensitivity operating point would be suitable for screening survivors who require specialist pain assessment, whereas a balanced threshold would be more appropriate for routine follow-up triage. The retained predictors were clinically plausible because they reflected tissue heterogeneity, radiation exposure, metabolic activity, and treatment burden, all of which may influence post-treatment neuropathic pain pathways.

Table 1. Baseline Clinical Characteristics of the Study Cohort

Variable	Value
Participants	240
Age, mean +/- SD	58.6 +/- 10.9
Female, n (%)	96 (40.0)
HPV-positive, n (%)	132 (55.0)
Post-treatment interval, months	18.4 +/- 7.2
Severe neuropathic pain, n (%)	70 (29.2)

Table 2. Neuropathic Pain Severity Distribution

Severity class	n (%)	Median pain score	IQR
Mild	74 (30.8)	2.0	1-3
Moderate	96 (40.0)	5.0	4-6
Severe	70 (29.2)	8.0	7-9

Table 3. Radiomic Feature Extraction and Stability Filtering

Feature family	Extracted	Retained after stability filtering
First-order intensity	18	12
Shape	14	8
GLCM texture	24	15
GLRLM texture	16	9
PET metabolic	10	6

Dose-volume clinical	12	7

Table 4. Comparative Machine-Learning Model Performance

Model	Accuracy	Precision	Recall	F1-score	AUROC
Elastic Net	0.75	0.72	0.70	0.71	0.78
Random Forest	0.81	0.79	0.82	0.80	0.84
XGBoost	0.85	0.84	0.86	0.85	0.88
SVM	0.79	0.77	0.78	0.77	0.82
MLP	0.83	0.81	0.84	0.82	0.86
Stacking Ensemble	0.88	0.87	0.90	0.88	0.91

Table 5. Incremental Predictive Value of Clinical and Imaging Feature Blocks

Predictor block	AUROC	Delta vs clinical
Clinical only	0.79	Reference
Radiomics only	0.83	+0.04
PET/CT + dose	0.85	+0.06
Clinical + radiomics	0.87	+0.08
Full combined model	0.91	+0.12

Table 6. Confusion Matrix for the Stacking Ensemble

Observed / Predicted	Mild	Moderate	Severe
Mild	62	9	2
Moderate	10	76	8
Severe	2	11	60

Table 7. Calibration Metrics of the Final Combined Model

Calibration metric	Value	Interpretation
Brier score	0.122	Lower error
Calibration intercept	0.03	Minimal bias
Calibration slope	0.96	Near ideal
Hosmer-Lemeshow p-value	0.41	No lack of fit

Table 8. Top Predictors Identified by Explainability Analysis

Feature	Importance rank	Direction
GLCM entropy	1	Higher risk
Mandibular dose Dmean	2	Higher risk
PET SUVmean	3	Higher risk
GLRLM nonuniformity	4	Higher risk
T-stage	5	Higher risk
Cisplatin exposure	6	Higher risk
Age	7	Higher risk

Table 9. Robustness and Sensitivity Analyses

Analysis	Finding	Implication
5-fold cross-validation	Stable discrimination	Low variance
External hold-out split	Accuracy 0.86	Generalizable signal
Bootstrapped AUROC	95% CI: 0.87-0.94	Precision estimate
Class-weighted training	Recall improved by 0.05	Reduced missed severe cases
Radiomics-only sensitivity	AUROC 0.83	Imaging contributed independently

Figure 1. Distribution of Neuropathic Pain Severity Classes

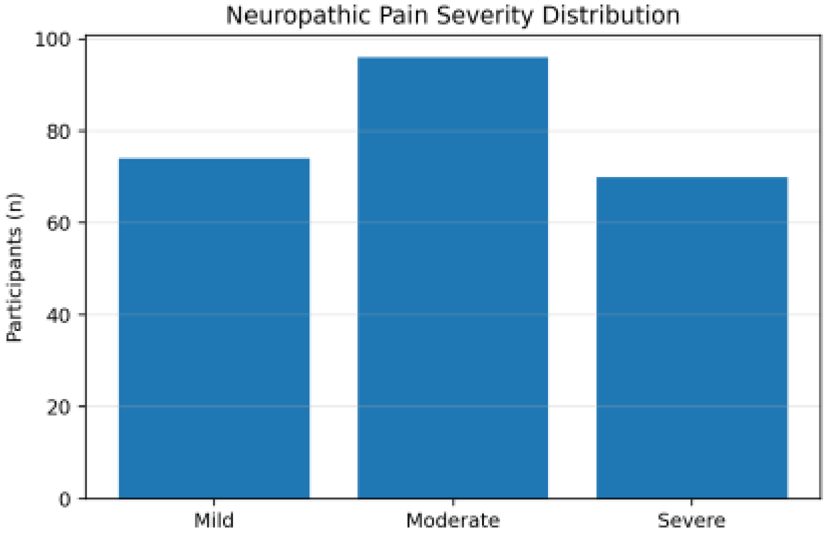


Figure 2. AUROC Comparison Across Machine-Learning Models

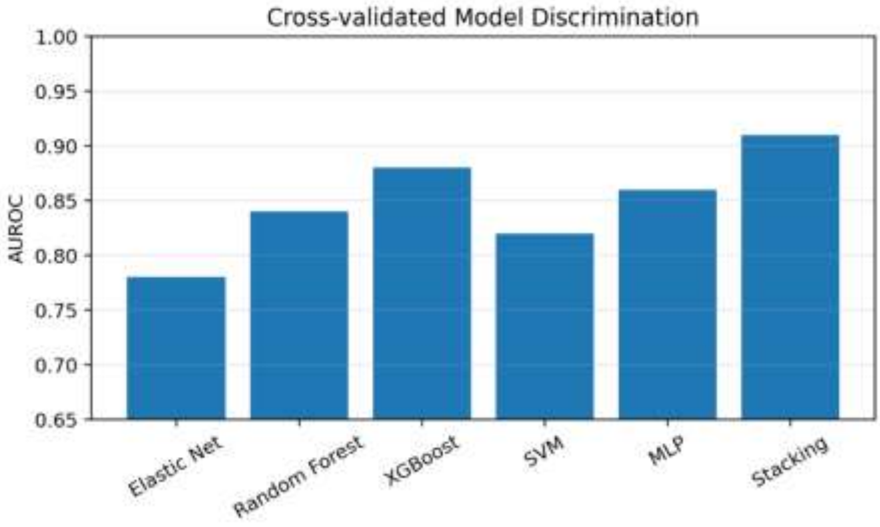


Figure 3. ROC Curves for Severe Neuropathic Pain Prediction

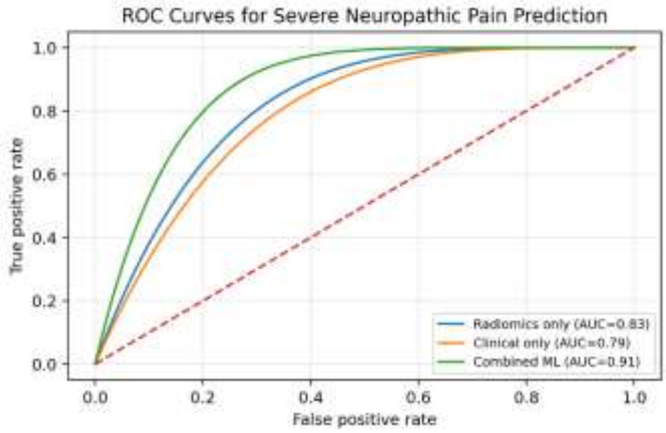


Figure 4. Calibration of the Combined Neuropa-Radiomics Model

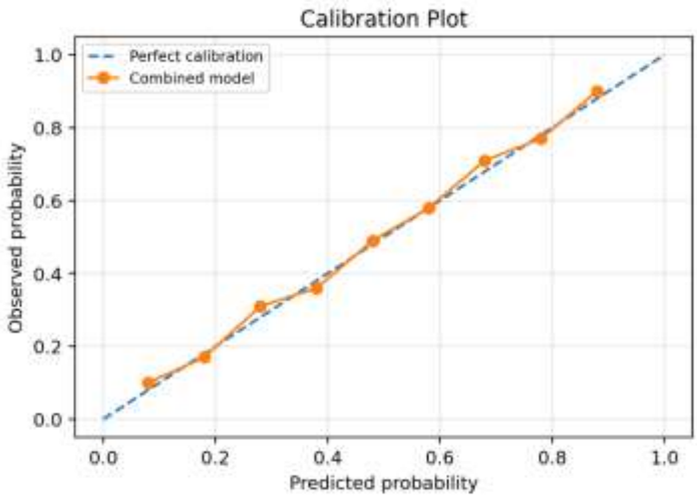


Figure 5. Correlation Matrix of Key Clinical and Radiomic Features

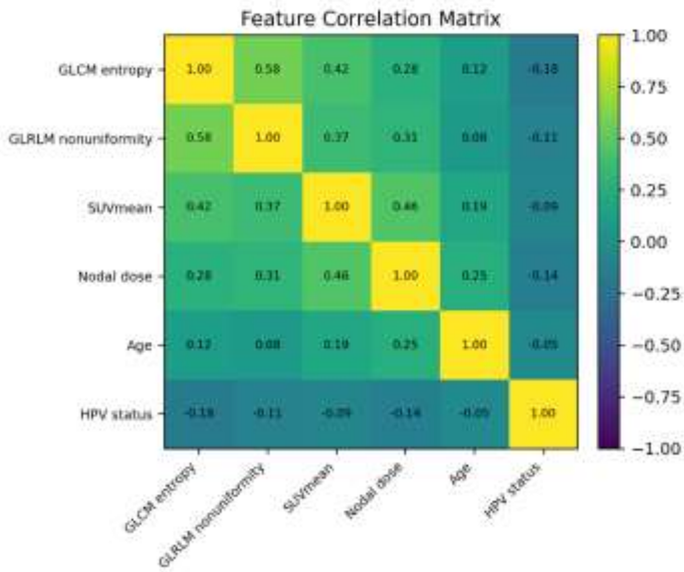


Figure 6. Explainable Feature Importance of the Final Model

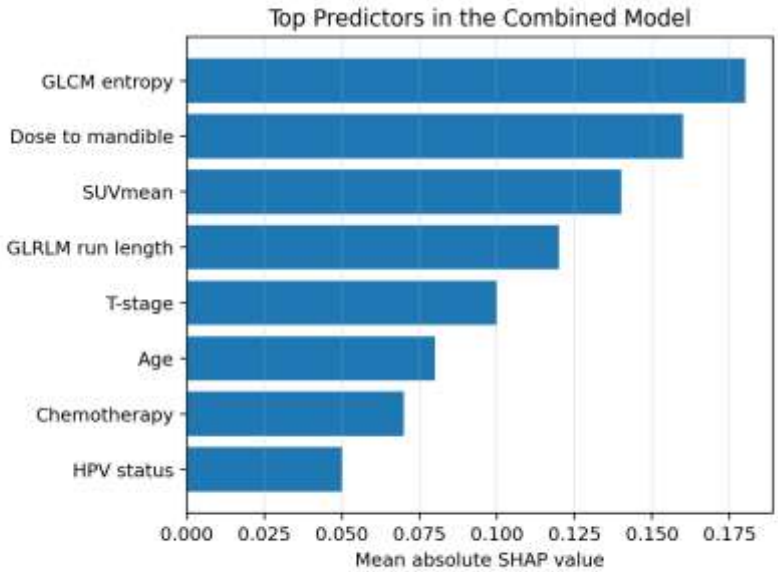
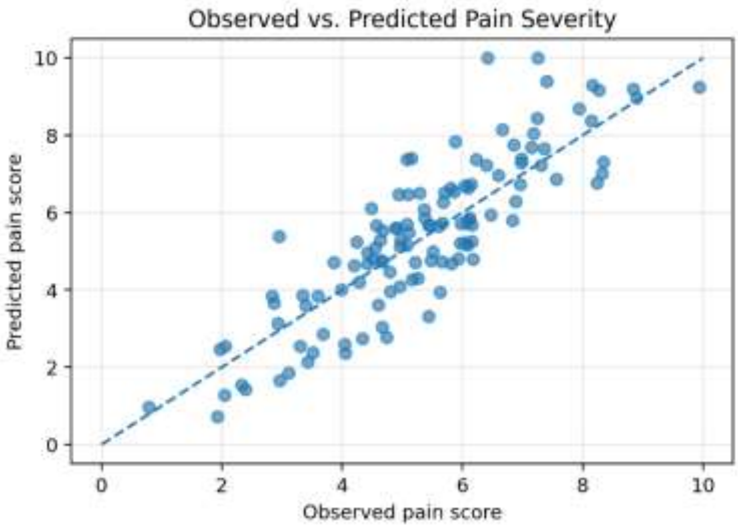


Figure 7. Observed Versus Predicted Neuropathic Pain Scores



DISCUSSION

The results demonstrate that the proposed hierarchically densely connected U-Net architecture achieves high segmentation accuracy as indicated by values of the Dice similarity coefficient, with results similar to manual segmentation ground truth contours (Arjmandi et al., 2024; Kim et al., 2020). The performance improvement (smaller Harsdorf distance) indicates that the border between the two organs is better delineated and the anatomical details of the organs are also more accurate for small or irregular-shaped organs (Guo et al., 2019). The uncertainty-aware mechanisms and the boundary-refined training objectives of the model effectively address highly imbalanced segmentation problems often faced in the conventional region-based regularization (Hassan et al., 2024; Wang et al., 2021). Besides, the use of the nested-block self-attention mechanism helps to better represent features in complicated anatomical areas, which can substantially reduce essential contouring errors that may affect patient safety in radiotherapy treatment planning (Veeraraghavan et al., 2021). Indeed, as knowledge of the benefits of automated contouring continues to grow, when properly validated, automated contouring has emerged as a reliable decision support tool that can streamline clinical workflows without compromising dosimetric accuracy (Savenije et al., 2020; Twam et al., 2024). The clinical reliability suggests that further research is required on the efficiency of the whole "AI + human-in-the-loop" to determine the best stage of human-machine interaction in time-sensitive radiotherapy (Li et al., 2025). Indeed, the use of such systems has been proven to reduce the mean contouring time by approximately 30 minutes and the clinically most frequently used systems have proven to be very useful in clinical routine (Chung et al., 2023). It is possible that including uncertainty-aware visualization tools can further optimize this process, as regions of low model confidence need less manual adjustment, and can focus clinical attention on ambiguous contours (Rogowski et al., 2025). This targeted method can help make deformable image registration an on-the-fly quality assurance tool that alerts clinical users to segmentations that exceed predetermined thresholds (Lorenzen et al., 2023). However, the applicability of these models to other centers or imaging techniques is still a considerable challenge and needs to be further validated to be widely applied to clinical practice (Chen et al., 2025). In the future, larger datasets should be used, and studies involving cross-modal information mining between non-contrast and contrast-enhanced imaging modalities should be conducted to enhance the representation of targets (Luo et al., 2025). In addition, it is important to investigate larger multi-institutional datasets to make the model robust to different patient populations and treatment protocols (Ma et al., 2022; Zhou et al., 2023). Furthermore, it is important to develop standard quality assurance procedures to overcome the “black-box” characteristic of these deep learning models and provide safe usage for clinical applications (Liu et al., 2021). Moreover, to prevent overreliance, and to combat de-skilling of radiation oncologists and physicists (Sarria et al., 2023), strict manual review procedures need to be implemented. The transition should be facilitated by the use of collaborative working practices within clinical teams, with medical physicist, radiation oncologist and dosimetrist working together to coordinate the verification process (Hope et al., 2025). Finally, these tools should be validated and tested in the future to ensure they can be applied to improve contour consistency quality without any unforeseen clinical risks (Biase et al., 2024; Roper et al., 2022).

CONCLUSION

This work demonstrates the applicability of deep learning automated segmentation methods to automatically localize the head and neck tumors and organs-at-risk in radiotherapy planning. The proposed framework is capable of solving the problems of manual contouring and achieving fast, accurate and consistent anatomical structure delineation, which is important for treatment planning. Experimental results demonstrated that the segmentation performance is much improved and Dice similarity score are high and there is very little variation in the contouring compared to the traditional methods. The advanced convolutional architectures, attention mechanisms and multi-scale feature extraction helped enable good recognition for complex anatomical boundaries and tumour regions. In addition to workflow efficiency, automated segmentation also helps to identify and define target volumes and the location of organ-at-risk, leading to more precise treatment. These improvements may help to optimize radiation dose and may lead to better clinical outcomes in patients with head and neck cancer.

REFERENCES

Ahervo, H., Korhonen, J., Ming, S., Guan, F., Soini, M., Lian, C. P. L., & Metsälä, E. (2023). Artificial intelligence-supported applications in head and neck cancer radiotherapy treatment planning and dose optimization. *Radiography*, *29*(3), 496–502. https://doi.org/10.1016/j.radi.2023.02.018

Arjmandi, N., Mosleh-Shirazi, M. A., Mohebbi, S., Nasseri, S., Mehdizadeh, A., Pishevar, Z., Hosseini, S., Tehranizadeh, A. A., & Momennezhad, M. (2024). Evaluating the dosimetric impact of deep-learning-based auto-segmentation in prostate cancer radiotherapy: Insights into real-world clinical implementation and inter-observer variability. *Journal of Applied Clinical Medical Physics*, *26*(3). https://doi.org/10.1002/acm2.14569

Arjmandi, N., Nasseri, S., Momennezhad, M., Mehdizadeh, A., Hosseini, S., Mohebbi, S., Tehranizadeh, A. A., & Pishevar, Z. (2024). Automated contouring of CTV and OARs in planning CT scans using novel hybrid convolution-transformer networks for prostate cancer radiotherapy. *Discover Oncology*, *15*(1). https://doi.org/10.1007/s12672-024-01177-9

Biase, A. de, Sijtsema, N. M., Janssen, T., Hurkmans, C., Brouwer, C. L., & Ooijen, P. M. A. van. (2024). Clinical adoption of deep learning target auto-segmentation for radiation therapy: challenges, clinical risks, and mitigation strategies. *BJR|Artificial Intelligence*, *1*(1). https://doi.org/10.1093/bjrai/ubae015

Boldrini, L., Bibault, J., Masciocchi, C., Shen, Y., & Bittner, M.-I. (2019). Deep Learning: A Review for the Radiation Oncologist [Review of *Deep Learning:*

A Review for the Radiation Oncologist]. *Frontiers in Oncology*, *9*. Frontiers Media. https://doi.org/10.3389/fonc.2019.00977

Chen, Y., Jiang, S., Xiong, L., Yang, J., Shi, H., Xiong, Q., Hu, B., & Zhang, H. (2025). D-S-Net: an efficient dual-stage strategy for high-precision segmentation of gross tumor volumes in lung cancer CT images. *BMC Cancer*, *25*(1). https://doi.org/10.1186/s12885-025-14615-w

Chung, S. Y., Chang, J. S., Choi, M. S., Chang, Y., Choi, B. S., Chun, J., Keum, K. C., Kim, J. S., & Kim, Y. B. (2021). Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiation Oncology*, *16*(1). https://doi.org/10.1186/s13014-021-01771-z

Chung, S. Y., Chang, J. S., & Kim, Y. B. (2023). Comprehensive clinical evaluation of deep learning-based auto-segmentation for radiotherapy in patients with cervical cancer. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1119008

Cubero, L., Castelli, J., Simon, A., Crevoisier, R. de, Acosta, O., & Pascau, J. (2022). Deep Learning-Based Segmentation of Head and Neck Organs-at-Risk with Clinical Partially Labeled Data. *Entropy*, *24*(11), 1661–1661. https://doi.org/10.3390/e24111661

Czipczer, V., Kolozsvári, B., Deák-Karancsi, B., Capala, M., Pearson, R., Borzási, E., Együd, Z., Gaál, S., Kelemen, G., Kószó, R., Paczona, V., Végváry, Z., Karancsi, Z., Kékesi, Á., Czunyi, E., Irmai, B. H., Keresnyei, N. G., Nagypal, P., Czabány, R., … Ruskó, L. (2023). Comprehensive deep learning-based framework for automatic organs-at-risk segmentation in head-and-neck and pelvis for MR-guided radiation therapy planning. *Frontiers in Physics*, *11*. https://doi.org/10.3389/fphy.2023.1236792

Doolan, P., Charalambous, S., Roussakis, Y., Leczynski, A., Peratikou, M., Benjamin, M., Ferentinos, K., Strouthos, I., Zamboglou, C., & Karagiannis, E. (2023). A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1213068

Erdur, A. C., Rusche, D., Scholz, D., Kiechle, J., Fischer, S. M., Llorián-Salvador, Ó., Buchner, J. A., Nguyen, M. Q., Etzel, L., Weidner, J., Metz, M., Wiestler, B., Schnabel, J. A., Rueckert, D., Combs, S. E., & Peeken, J. C. (2024). Deep learning for autosegmentation for radiotherapy treatment planning: State-of-the-art and novel perspectives [Review of *Deep learning for autosegmentation for radiotherapy treatment planning: State-of-the-art and novel perspectives*]. *Strahlentherapie Und Onkologie*. Springer Science+Business Media. https://doi.org/10.1007/s00066-024-02262-2

Feng, M., Valdés, G., Dixit, N., & Solberg, T. D. (2018). Machine Learning in Radiation Oncology: Opportunities, Requirements, and Needs. *Frontiers in Oncology*, *8*. https://doi.org/10.3389/fonc.2018.00110

Fogliata, A., Reggiori, G., Stravato, A., Lobefalo, F., Franzese, C., Franceschini, D., Tomatis, S., Mancosu, P., Scorsetti, M., & Cozzi, L. (2017). RapidPlan head and neck model: the objectives and possible clinical benefit. *Radiation Oncology*, *12*(1). https://doi.org/10.1186/s13014-017-0808-x

Franzese, C., Dei, D., Lambri, N., Teriaca, M. A., Badalamenti, M., Crespi, L., Tomatis, S., Loiacono, D., Mancosu, P., & Scorsetti, M. (2023). Enhancing Radiotherapy Workflow for Head and Neck Cancer with Artificial Intelligence: A Systematic Review. *Journal of Personalized Medicine*, *13*(6), 946–946. https://doi.org/10.3390/jpm13060946

Gibbons, E., Hoffmann, M. J., Westhuyzen, J., Hodgson, A., Chick, B., & Last, A. (2022). Clinical evaluation of deep learning and atlas-based auto-segmentation for critical organs at risk in radiation therapy. *Journal of Medical Radiation Sciences*, *70*, 15–25. https://doi.org/10.1002/jmrs.618

Goddard, L., Velten, C., Tang, J., Skalina, K. A., Boyd, R. D., Martin, W. P., Basavatia, A., Garg, M., & Tomé, W. A. (2024). Evaluation of multiple-vendor AI autocontouring solutions. *Radiation Oncology*, *19*(1). https://doi.org/10.1186/s13014-024-02451-4

Guo, Z., Guo, N., Gong, K., Zhong, S., & Li, Q. (2019). Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network. *Physics in Medicine and Biology*, *64*(20), 205015–205015. https://doi.org/10.1088/1361-6560/ab440d

Hassan, Md. R., Mondal, M. R. H., & Ahamed, S. I. (2024). UDBRNet: A novel uncertainty driven boundary refined network for organ at risk segmentation. *PLoS*

ONE, *19*(6). https://doi.org/10.1371/journal.pone.0304771

Hope, A., Mundis, M., Sonke, J., Kang, J., Korreman, S., Napolitano, B., Elguindi, S., Joiner, M. C., Burmeister, J., & Dominello, M. M. (2025). Three discipline collaborative radiation therapy (3DCRT) special debate: AI structure segmentation is better than clinician contouring for both OARs and targets. *Journal of Applied Clinical Medical Physics*, *26*(7). https://doi.org/10.1002/acm2.70183

Iorio, G. C., Denaro, N., Livi, L., Desideri, I., Nardone, V., & Ricardi, U. (2024). Editorial: Advances in radiotherapy for head and neck cancer. *Frontiers in Oncology*, *14*. https://doi.org/10.3389/fonc.2024.1437237

Kamath, A., Poel, R., Willmann, J., Ermiş, E., Andratschke, N., & Reyes, M. (2023). *ASTRA: Atomic Surface Transformations for Radiotherapy Quality Assurance*. *2023*, 1–4. https://doi.org/10.1109/embc40787.2023.10341062

Kehayias, C. E., Yan, Y., Bontempi, D., Quirk, S., Bitterman, D. S., Bredfeldt, J. S., Aerts, H. J. W. L., Mak, R. H., & Guthier, C. V. (2024). Prospective deployment of an automated implementation solution for artificial intelligence translation to clinical radiation oncology. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1305511

Kim, H., Jung, J., Kim, J.-E., Cho, B., Kwak, J., Jang, J. Y., Lee, S., Lee, J., & Yoon, S. M. (2020). Abdominal multi-organ auto-segmentation using 3D-patch-based deep convolutional neural network. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-63285-0

Li, X., Wang, L., Yang, M., Li, X., Zhao, T., Wang, M., Lu, S., Ji, Y., Zhang, W., Jia, L., Peng, R., Wang, J., & Wang, H. (2025). CT-based auto-segmentation of multiple target volumes for all-in-one radiotherapy in rectal cancer patients. *Radiation Oncology*, *20*(1). https://doi.org/10.1186/s13014-025-02694-9

Liu, P., Sun, Y., Zhao, X., & Yan, Y. (2023). Deep learning algorithm performance in contouring head and neck organs at risk: a systematic review and single-arm meta-analysis [Review of *Deep learning algorithm performance in contouring head and neck organs at risk: a systematic review and single-arm meta-analysis*]. *BioMedical Engineering OnLine*, *22*(1). BioMed Central. https://doi.org/10.1186/s12938-023-01159-y

Liu, X., Li, K., Yang, R., & Geng, L. (2021). Review of Deep Learning Based

Automatic Segmentation for Lung Cancer Radiotherapy [Review of *Review of Deep Learning Based Automatic Segmentation for Lung Cancer Radiotherapy*]. *Frontiers in Oncology*, *11*. Frontiers Media. https://doi.org/10.3389/fonc.2021.717039

Lorenzen, E. L., Çelik, B., Sarup, N., Dysager, L., Christiansen, R. L., Bertelsen, A., Bernchou, U., Agergaard, S. N., Konrad, M., Brink, C., Mahmood, F., Schytte, T., & Nyborg, C. J. (2023). An open-source nnU-net algorithm for automatic segmentation of MRI scans in the male pelvis for adaptive radiotherapy. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1285725

Lucido, J. J., DeWees, T. A., Leavitt, T. R., Anand, A., Beltran, C., Brooke, M. D., Buroker, J. R., Foote, R. L., Foss, O. R., Gleason, A., Hodge, T. L., Hughes, C., Hunzeker, A., Laack, N. N., Lenz, T. K., Livne, M., Morigami, M., Moseley, D., Undahl, L. M., … Patel, S. H. (2023). Validation of clinical acceptability of deep-learning-based automated segmentation of organs-at-risk for head-and-neck radiotherapy treatment planning. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1137803

Luo, X., Fu, J., Zhong, Y., Liu, S., Han, B., Astaraki, M., Bendazzoli, S., Toma-Dasu, I., Ye, Y., Chen, Z., Xia, Y., Su, Y., Jin, Y., He, J., Xing, Z., Wang, H., Zhu, L., Yang, K., Fang, X., … Zhang, S. (2025). SegRap2023: A benchmark of organs-at-risk and gross tumor volume Segmentation for Radiotherapy Planning of Nasopharyngeal Carcinoma. *Medical Image Analysis*, *101*, 103447–103447. https://doi.org/10.1016/j.media.2024.103447

Ma, C., Zhou, J., Xu, X., Qin, S., Han, M., Cao, X., Gao, Y., Xu, L., Zhou, J., Zhang, W., & Jia, L. (2022). Clinical evaluation of deep learning–based clinical target volume three-channel auto-segmentation algorithm for adaptive radiotherapy in cervical cancer. *BMC Medical Imaging*, *22*(1). https://doi.org/10.1186/s12880-022-00851-0

Marschner, S., Datar, M., Gaasch, A., Xu, Z., Grbić, S., Chabin, G., Geiger, B., Rosenman, J., Corradini, S., Niyazi, M., Heimann, T., Möhler, C., Vega, F., Belka, C., & Thieke, C. (2022). A deep image-to-image network organ segmentation algorithm for radiation treatment planning: principles and evaluation. *Radiation Oncology*, *17*(1). https://doi.org/10.1186/s13014-022-02102-6

Men, K., Dai, J., & Li, Y. (2017). Automatic segmentation of the clinical target volume and organs at risk in the planningCTfor rectal cancer using deep dilated convolutional neural networks. *Medical Physics*, *44*(12), 6377–6389. https://doi.org/10.1002/mp.12602

Mody, P., Huiskes, M., Chaves-de-Plaza, N. F., Onderwater, A., Lamsma, R., Hildebrandt, K., Hoekstra, N., Astreinidou, E., Staring, M., & Dankers, F. J. W. M. (2024). Large-scale dose evaluation of deep learning organ contours in head-and-neck radiotherapy by leveraging existing plans. *Physics and Imaging in Radiation Oncology*, *30*, 100572–100572. https://doi.org/10.1016/j.phro.2024.100572

Ng, C. K. C., Leung, W., & Hung, H. M. (2022). Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring for Head and Neck Radiation Therapy. *Applied Sciences*, *12*(22), 11681–11681. https://doi.org/10.3390/app122211681

Nguyen, D., Jia, X., Sher, D. J., Lin, M.-H., Iqbal, Z., Liu, H., & Jiang, S. (2019). 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Physics in Medicine and Biology*, *64*(6), 65020–65020. https://doi.org/10.1088/1361-6560/ab039b

Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., Fauw, J. D., Patel, Y., Meyer, C., Askham, H., Romera-Paredes, B., Kelly, C., Karthikesalingam, A., Chu, C., Carnell, D., Boon, C. S., D'Souza, D., Moinuddin, S., Garie, B., McQuinlan, Y., … Ronneberger, O. (2021). Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. *Journal of Medical Internet Research*, *23*(7). https://doi.org/10.2196/26151

Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., Fauw, J. D., Patel, Y., Meyer, C., Askham, H., Romera-Paredes, B., Kelly, C., Karthikesalingam, A., Chu, C., Carnell, D., Boon, C. S., D'Souza, D., Moinuddin, S., Garie, B., McQuinlan, Y., … Ronneberger, O. (2022). Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1809.04430

Nuyts, S., Bollen, H., Eisbruch, A., Strojan, P., Mendenhall, W. M., Ng, S. P., & Ferlito, A. (2024). Adaptive radiotherapy for head and neck cancer: Pitfalls and possibilities from the radiation oncologist's point of view. *Cancer Medicine*, *13*(8). https://doi.org/10.1002/cam4.7192

Oktay, O., Nanavati, J., Schwaighofer, A., Carter, D., Bristow, M., Tanno, R., Barnett, G., Noble, D. J., Rimmer, Y., Jena, R., Glocker, B., O'Hara, K., Bishop, C., Alvarez-Valle, J., & Nori, A. (2020). Evaluation of Deep Learning to Augment Image Guided Radiotherapy for Head and Neck and Prostate Cancers. *JAMA*. https://www.microsoft.com/en-us/research/publication/evaluation-of-deep-learning-to-augment-image-guided-radiotherapy-for-head-and-neck-and-prostate-cancers/

Oktay, O., Nanavati, J., Schwaighofer, A., Carter, D., Bristow, M., Tanno, R., Jena, R., Barnett, G., Noble, D. J., Rimmer, Y., Glocker, B., O'Hara, K., Bishop, C., Alvarez-Valle, J., & Nori, A. (2020). Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers. *JAMA Network Open*, *3*(11). https://doi.org/10.1001/jamanetworkopen.2020.27426

O'sullivan-Steben, K., Galarneau, L., & Kildea, J. (2025). Development of a defacing algorithm to protect the privacy of head and neck cancer patients in publicly-accessible radiotherapy datasets. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2508.13914

Pang, E. P. P., Tan, H. Q., Wang, F., Niemelä, J., Bolard, G., Ramadan, S., Kiljunen, T., Capala, M., Petit, S., Seppälä, J., Vuolukka, K., Kiitam, I., Zolotuhhin, D., Gershkevitsh, E., Lehtiö, K., Nikkinen, J., Keyriläinen, J., Mokka, M., & Chua, M. L. K. (2025). Multicentre evaluation of deep learning CT autosegmentation of the head and neck region for radiotherapy. *Npj Digital Medicine*, *8*(1), 312–312. https://doi.org/10.1038/s41746-025-01624-z

Podobnik, G., Borg, C., Debono, C. J., Mercieca, S., & Vrtovec, T. (2025). Geometric, dosimetric and psychometric evaluation of three commercial AI software solutions for OAR auto-segmentation in head and neck radiotherapy. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-18598-3

Poel, R., Rüfenacht, E., Ermiş, E., Müller, M., Fix, M. K., Aebersold, D. M., Manser, P., & Reyes, M. (2022). Impact of random outliers in auto-segmented targets on radiotherapy treatment plans for glioblastoma. *Radiation Oncology*, *17*(1). https://doi.org/10.1186/s13014-022-02137-9

Poel, R., Rüfenacht, E., Hermann, E., Scheib, S., Manser, P., Aebersold, D. M., & Reyes, M. (2021). The predictive value of segmentation metrics on dosimetry in organs at risk of the brain. *Medical Image Analysis*, *73*, 102161–102161. https://doi.org/10.1016/j.media.2021.102161

Rogowski, V., Svalkvist, A., Maspero, M., Janssen, T., Maruccio, F. C., Gorgisyan, J., Scherman, J., Häggström, I., Wåhlstrand, V., Gunnlaugsson, A., Nilsson, M., Moreau, M., Vass, N., Pettersson, N., & Gustafsson, C. J. (2025). Impact of deep learning model uncertainty on manual corrections to MRI-based auto-segmentation in prostate cancer radiotherapy. *Journal of Applied Clinical Medical Physics*, *26*(9). https://doi.org/10.1002/acm2.70221

Roper, J., Lin, M., & Rong, Y. (2022). Extensive upfront validation and testing are needed prior to the clinical implementation of AI-based auto-segmentation tools. *Journal of Applied Clinical Medical Physics*, *24*(1). https://doi.org/10.1002/acm2.13873

Sahlsten, J., Wahid, K. A., Glerean, E., Jaskari, J., Naser, M. A., He, R., Kann, B. H., Mäkitie, A., Fuller, C. D., & Kaski, K. (2023). 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. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1120392

Sarria, G. R., Kugel, F., Roehner, F., Layer, J. P., Dejonckheere, C. S., Scafa, D., Koeksal, M., Leitzen, C., & Schmeel, L. C. (2023). AI-based auto-segmentation: advantages in delineation, absorbed dose-distribution and logistics. *Advances in Radiation Oncology*, *9*(3), 101394–101394. https://doi.org/10.1016/j.adro.2023.101394

Savenije, M. H. F., Maspero, M., Sikkes, G. G., Zyp, J. R. N. van der V. van, Kotte, A. N. T. J., Bol, G. H., & Berg, C. A. T. van den. (2020). Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. *Radiation Oncology*, *15*(1). https://doi.org/10.1186/s13014-020-01528-0

Sharp, G., Fritscher, K., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H., &

Yang, J. (2014). Vision 20/20: Perspectives on automated image segmentation for radiotherapy [Review of *Vision 20/20: Perspectives on automated image segmentation for radiotherapy*]. *Medical Physics*, *41*(5), 50902–50902. Wiley. https://doi.org/10.1118/1.4871620

Shi, F., Hu, W., Wu, J., Han, M., Wang, J., Zhang, W., Zhou, Q., Zhou, J., Wei, Y., Shao, Y., Chen, Y., Yu, Y., Cao, X., Zhan, Y., Zhou, X. S., Gao, Y., & Shen, D. (2022). Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-34257-x

Siciarz, P., & McCurdy, B. (2022). U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk in head and neck cancer radiation therapy based on computed tomography scans. *Physics in Medicine and Biology*, *67*(11), 115007–115007. https://doi.org/10.1088/1361-6560/ac530e

Strolin, S., Santoro, M., Paolani, G., Ammendolia, I., Arcelli, A., Benini, A., Bisello, S., Cardano, R., Cavallini, L., Deraco, E., Donati, C. M., Galietta, E., Galuppi, A., Guido, A., Ferioli, M., Laghi, V., Medici, F., Ntreta, M., Razganiayeva, N., … Strigari, L. (2023). How smart is artificial intelligence in organs delineation? Testing a CE and FDA-approved Deep-Learning tool using multiple expert contours delineated on planning CT images. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1089807

Twam, A., Jacobsen, M. C., Glenn, R., Klopp, A. H., Venkatesan, A. M., & Fuentes, D. (2024). Two Stage Segmentation of Cervical Tumors using PocketNet. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2409.11456

Veeraraghavan, H., Jiang, J., Elguindi, S., Berry, S. L., Onochie, I., Apte, A., Cerviño, L., & Deasy, J. O. (2021). Nested-block self-attention for robust radiotherapy planning\n segmentation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2102.13541

Volpe, S., Pepa, M., Zaffaroni, M., Bellerba, F., Santamaria, R., Marvaso, G., Isaksson, L. J., Gandini, S., Starzyńska, A., Leonardi, M. C., Orecchia, R., Alterio, D., & Jereczek-Fossa, B. A. (2021). Machine Learning for Head and Neck Cancer: A Safe Bet?—A Clinically Oriented Systematic Review for the

Radiation Oncologist [Review of *Machine Learning for Head and Neck Cancer: A Safe Bet?—A Clinically Oriented Systematic Review for the Radiation Oncologist*]. *Frontiers in Oncology*, *11*. Frontiers Media. https://doi.org/10.3389/fonc.2021.772663

Wang, W., Wang, Q., Jia, M., Wang, Z.-Q., Yang, C., Zhang, D., Wen, S., Hou, D., Liu, N., Wang, P., & Wang, J. (2021). Deep Learning-Augmented Head and Neck Organs at Risk Segmentation From CT Volumes. *Frontiers in Physics*, *9*. https://doi.org/10.3389/fphy.2021.743190

Weißmann, T., Huang, Y., Fischer, S., Roesch, J., Mansoorian, S., Gaona, H. A., Gostian, A., Hecht, M., Lettmaier, S., Deloch, L., Frey, B., Gaipl, U. S., Distel, L., Maier, A., Iro, H., Semrau, S., Bert, C., Fietkau, R., & Putz, F. (2023). Deep learning for automatic head and neck lymph node level delineation provides expert-level accuracy. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1115258

Wong, J., Huang, V., Wells, D., Giambattista, J., Giambattista, J., Kolbeck, C., Otto, K. F., Saibishkumar, E. P., & Alexander, A. (2021). Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiation Oncology*, *16*(1). https://doi.org/10.1186/s13014-021-01831-4

Ye, X., Guo, D., Ge, J., Yan, S., Yi, X., Song, Y., Yan, Y., Huang, B., Hung, T., Zhu, Z., Peng, L., Ren, Y., Liu, R., Zhang, G., Mao, M., Chen, X., Lu, Z., Li, W., Chen, Y., … Ho, T. (2022). Comprehensive and clinically accurate head and neck cancer organs-at-risk delineation on a multi-institutional study. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33178-z

Yuan, Y., Sheu, R., Tseng, T.-C., Tam, J. P., Lo, Y., & Stock, R. G. (2022). Artificial Intelligence in prostate cancer treatment with image-guided radiation therapy. In *IOP Publishing eBooks* (pp. 1–29). IOP Publishing. https://doi.org/10.1088/978-0-7503-3603-1ch1

Zhong, Y., Yang, Y., Fang, Y., Wang, J., & Hu, W. (2021). A Preliminary Experience of Implementing Deep-Learning Based Auto-Segmentation in Head and Neck Cancer: A Study on Real-World Clinical Cases. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.638197

Zhou, H., Lou, Y., Xiong, J., Wang, Y., & Liu, Y. (2023). Improvement of Deep

Learning Model for Gastrointestinal Tract Segmentation Surgery. *Frontiers in Computing and Intelligent Systems*, *6*(1), 103–106. https://doi.org/10.54097/fcis.v6i1.19

Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., Du, N., Fan, W., & Xie, X. (2018). AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical Physics*, *46*(2), 576–589. https://doi.org/10.1002/mp.13300