September 2025
Volume 14, Number 3

# Applied Radiation Oncology™



www.appliedradiationoncology.com

# Applied Radiation Oncology™

Anderson Publishing, Ltd.
180 Glenside Avenue,
Scotch Plains, NJ 07076
Tel: 908-301-1995
Fax: 908-301-1997
info@appliedradiology.com
ESSN: 2334-5446 (online)





**CIRCULATION, COVERAGE and ADVERTISING RATES:** Details regarding circulation, coverage, advertising rates, space sizes, and similar information are available to prospective advertisers. Closing date is 30 days preceding date of issue. View our media planner at appliedradiationoncology.com/advertise.


**EDITORIAL CONTRIBUTIONS:** *Applied Radiation Oncology* accepts clinical review articles, research papers, and case reports that pertain to radiation oncology and related oncologic imaging procedures that will be of interest to radiation oncologists. Author guidelines are available at https://appliedradiationoncology.com/Author-Guidelines. Every precaution is taken to ensure accuracy, but the publishers cannot accept responsibility for the correctness or accuracy of the information supplied or for any opinion expressed. Before using procedures or treatments discussed or suggested by authors, clinicians should evaluate their patients' conditions, compare the recommendations of other authorities, consider possible contraindications or dangers, and review applicable manufacturer's product information. Review articles and cases should be geared to the practitioner and should reflect practical everyday clinical applications rather than research activity. Articles and cases may pertain to clinical management, administration, fiscal, technical, and medico-legal issues. All review articles, research articles and case reports undergo a double-anonymized peer review process.


*Applied Radiation Oncology* Print ISSN: 3065-4467, Online ISSN: 2334-5446, USPS 25688 is published quarterly by Anderson Publishing, LTD at 180 Glenside Avenue, Scotch Plains, NJ 07076. Readers can renew or subscribe at appliedradiationoncology.com/subscribe.




# Editorial Advisory Board

Applied Radiation Oncology™

SEPTEMBER 2025, Vol 14, No 3

FOCUS: ARTIFICIAL INTELLIGENCE



*Applied Radiation Oncology* (ISSN: 2334-5446) is published quarterly by Anderson Publishing, Ltd.,180 Glenside Avenue, Scotch Plains, NJ 07076. Subscription is free of charge to all medical professionals. To update your subscription preferences, visit appliedradiationoncology.com/subscribe. Complaints concerning non-receipt of this journal should be made via email to our publisher, Kieran Anderson at kieran@appliedrradiology.com.

**Dr Suh** is the editor-in-chief of *Applied Radiation Oncology*, and professor and chairman, Department of Radiation Oncology, at the Taussig Cancer Institute, Rose Ella Burkhardt Brain Tumor and Neuro-Oncology Center, Cleveland Clinic, Cleveland, OH.

# Artificial Intelligence: Aiding Precision and Practice in Radiation Oncology

John H. Suh, MD, FASTRO, FACR

Artificial intelligence (AI) has quickly moved from promise to practice in radiation oncology. Once viewed as futuristic, AI now shapes daily workflows in imaging, treatment planning, adaptive therapy, and even education. The fall issue of *Applied Radiation Oncology* captures this pivotal moment, highlighting both clinical applications and questions about how best to integrate and leverage AI into practice and training.

The article, *Recent Trends of Artificial Intelligence in Radiation Oncology: A Narrative Review of Prospective Studies,* focuses on prostate and head and neck cancers, where AI is already being incorporated into guidelines and practice. In prostate cancer, AI-driven digital pathology and multimodal models are improving prognostication and guiding treatment selection—advances now reflected in the NCCN guidelines. In head and neck cancer, deep learning enhances the detection of extranodal extension and supports treatment de-escalation strategies, while machine learning assists in identifying high-risk patients. The review also explores the use of AI to personalize treatment, improve outcomes, and advance precision medicine in daily practice.

In *Artificial Intelligence-Assisted Peer Review in Radiation Oncology*, the authors discuss how AI can be used during the peer review process to assist in identifying patients at high risk for interruptions or changes to their treatment plans. Their retrospective study demonstrates that AI and machine learning can help expedite peer review and mitigate the need for modifications after the start of therapy.

While AI can assist in overcoming human challenges to collective decision-making, such as interobserver variability, it is not the only solution. Another research article, *A Head and Neck Contour Grading System Provides an Objective Assessment of Radiation Oncology Resident Contouring Skills*, highlights how a structured grading system used with a peer review process is enabling attendings to objectively track residents' skill development in contour grading of head and neck malignancies, which provides a unique opportunity to optimize training.

As AI-assisted processes continue to replace manual tasks such as contouring, residents and more experienced radiation oncologists grapple with the prospect of trading skill development for efficiency. In the latest Residence Voice, *Auto Contouring in Residency: Cutting Corners or Creating Confidence?* Dr Elizabeth Thompson raises an important question: Should residents lean on AI to accelerate contouring, or does manual practice remain essential for mastery? Her perspective reflects the tension that many programs now face, the exchange of experience for efficiency, which has implications for the training of future radiation oncologists.

With the continued adoption of AI throughout the medical system, our challenge will be how to best balance automation with expert oversight. The future of radiation oncology will depend on how well we integrate human expertise with algorithmic precision.

As always, I truly appreciate your support of *Applied Radiation Oncology*. I look forward to seeing you at the upcoming ASTRO meeting in San Francisco!

https://doi.org/10.37549/ARO-D-25-0051

# Recent Trends of Artificial Intelligence in Radiation Oncology: A Narrative Review of Prospective Studies

Jas Virk, MD;[1] Simeng Zhu, MD;[2] Austin J. Sim, MD, JD;[2] Sung Jun Ma, MD[2]*

## Abstract

In the past several decades, the delivery of radiation therapy has become increasingly intricate and precise. Such advancements were observed in conjunction with abundant multimodal data available for analysis; these include sophisticated diagnostic imaging, electronic health records, and digital pathology. The impact of artificial intelligence (AI) has become more prominent as numerous prior and ongoing prospective studies aim to integrate it into clinical care in radiation oncology. This review article provides an overview of such prospective studies and examines the role of AI in radiation therapy. By providing an understanding of recent trends in AI, we hope to contribute to improved patient outcomes and precision medicine in radiation oncology.

**Keywords:** AI, machine learning, deep learning, radiomics, large language model, multimodal

## Introduction

Radiation therapy has progressed significantly over the past decades through such advances as stereotactic body radiation therapy (SBRT) for lung cancer[1-3] and oligometastatic cancer,[4-6] proton therapy for leptomeningeal metastasis,[7] magnetic resonance imaging (MRI)-guided SBRT for prostate cancer,[8] MRI-guided adaptive radiation therapy for pancreatic cancer,[9] and adaptive radiation therapy for head and neck cancer.[10] In addition, precision medicine has evolved to improve patient selection for various treatment approaches, including prostate-specific membrane antigen (PSMA) positron emission tomography (PET) for prostate cancer, $^{11,12}$F-fluoromisonidazole PET for head and neck cancer, $^{13,14}$gallium DOTATATE PET for meningioma,[15,16] 21-gene recurrence scores for breast cancer,[17,18] osimertinib after definitive chemoradiation for stage III epidermal growth factor receptor (EGFR)-mutant non-small cell lung cancer,[19] and chimeric antigen receptor T-cell therapy.[12]

With such advancements in precision medicine, cancer genetics, and imaging modalities leading to abundant multimodal data available for health care professionals to interpret, artificial intelligence (AI) has emerged to leverage such data.[20] For example, AI-based algorithms have greatly improved early diagnosis of breast cancer,[21] pancreatic cancer,[22] lung cancer,[23] and skin cancer.[24] Furthermore, generative AI has been shown to answer questions with more empathy than humans[25] and to assist with medical documentation.[26] In radiation

**Affiliations:** [1]Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, NY. [2]Department of Radiation Oncology, The Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, The Ohio State University Comprehensive Cancer Center, Columbus, OH.
**Corresponding author:** *Sung Jun Ma, MD, Department of Radiation Oncology, The Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, The Ohio State University Comprehensive Cancer Center 460 W 10th Ave, Columbus, OH 43210. (SungJun.Ma@osumc.edu)
**Disclosures:** Simeng Zhu receives consulting fees from Radformation, grants from Varian Medical Systems and Google, and is the founder and chief executive officer of IntelliOnc outside the submitted work. Austin Sim receives research funding from Novartis. He also receives travel support from and serves as a nonvoting member of the ASTRO Board of Directors outside the submitted work.

https://doi.org/10.37549/ARO-D-25-0004

oncology, several AI-related studies have emerged to minimize unplanned hospitalization[27] and detect extranodal extension (ENE) in head and neck cancer.[28,29] Since then, numerous reviews have summarized the role of AI in radiation oncology.[30-33] However, none have focused on prospective studies incorporating AI into practice. In this review, we aimed to highlight the overview of recent trends in the application of AI in radiation oncology based on prior and ongoing prospective studies.

## Methods

To identify relevant prospective studies on AI trends in radiation oncology, a literature search was conducted of the following electronic databases: PubMed, Medline, and Google Scholar. The following keywords were used: "radiation," "radiation oncology," and "artificial intelligence." The search was limited to publications ranging from January 2002 to December 2024 and excluded retrospective studies, systematic reviews, case reports, conference abstracts, and expert opinion articles. Additional filters included utilizing only English language-written articles. Article titles and abstracts were then reviewed after initial screening, followed by full-text review prior to finalizing study inclusion.

Current clinical trials were searched utilizing the ClinicalTrials.gov website with the following keywords: "cancer," "artificial intelligence," and "radiation." Studies that were completed or active (recruiting or not) were included, while those that were suspended or withdrawn were excluded. Trials were further categorized based on type, with only interventional studies included with no specific date range. When evaluating prospective studies or clinical trials, two reviewers determined the eligibility of such studies for inclusion.

## Results

Of 4469 articles found through our literature search, 234 were initially identified as prospective studies. After reviewing abstracts and full texts to confirm their eligibility, 30 studies met our criteria, as shown in **Table 1**.

### AI in Prostate Cancer

AI has been investigated extensively to improve outcomes of patients with prostate cancer. In earlier years, because of substantial interobserver disagreements in Gleason grade among pathologists,[61,62] AI-assisted digital pathology algorithms based on whole-slide images of hematoxylin and eosin-stained tissues were developed to improve reproducibility in determining Gleason grade,[63] which were recognized by Food and Drug Administration and other regulatory agencies.[63]

Beyond assessment of Gleason grades, the role of digital pathology has been investigated in radiation oncology. Esteva et al. initially leveraged five NRG Oncology phase III randomized clinical trials (NRG/RTOG 9202, 9413, 9910, 0126, and 9408) that included patients with localized prostate cancer who received radiation with or without androgen-deprivation therapy (ADT).[56] Self-supervised, prognostic, and multimodal AI architecture was developed based on clinical variables (age, Gleason primary and secondary grades, T stage, and baseline PSA) from over 5600 patients and imaging features from over 16 000 histopathology slides.[56] Across all endpoints, AI outperformed the National Comprehensive Cancer Network (NCCN) risk-stratification tool by 9.2%-14.6% for relative improvements in area under the receiver operating characteristic curve (AUC).[56]

With its early success, digital pathology was further investigated for its predictive ability. Spratt et al. utilized four NRG Oncology phase III clinical trials (NRG/RTOG 9202, 9413, 9910, 0126) to develop a similar multimodal AI architecture and validated its performance on the NRG/RTOG 9408 dataset.[57] The primary objective of this study was to identify a subgroup of patients who might benefit from adding ADT to radiation.[57] The development cohort comprised over 2000 patients, with the majority having intermediate-risk prostate cancer, while the validation cohort consisted of over 1500 patients, with more than half having intermediate-risk prostate cancer.[57] Over a third of patients in the validation cohort were classified as predictive model-positive, demonstrating an absolute improvement of 10% by adding ADT for distant metastasis-free survival and prostate-cancer-specific survival at 15 years.[57] However, no differential treatment benefits were identified between predictive model subgroups for metastasis-free survival and overall survival.[57] Spratt et al. performed a separate analysis using six NRG Oncology clinical trials (NRG/RTOG 9202, 9408, 9413, 9910, 9902, 0521), validating the multimodal AI algorithm as prognostic for distant metastasis and prostate cancer-specific mortality among patients with high-risk prostate cancer.[59] Subsequently, the NCCN Guideline for prostate cancer included ArteraAI Prostate as the first AI-based tool with prognostic and predictive benefits from ADT

**Table 1. Prior prospective studies**

| AUTHORS | YEAR | DISEASE SITE | PROSPECTIVE DATA | DATA TYPES | MAIN FINDINGS |
|---|---|---|---|---|---|
| Zeleznik et al.[34] | 2021 | Breast | Not available | CT scan | With deep learning assistance, heart segmentation time was significantly reduced. Expert accuracy was comparable with deep learning-only segmentations. |
| Ma et al.[35] | 2023 | Breast | ClinicalTrials.gov ID: NCT05609058 | CT scan | Deep learning model identified the lead wire markers in the CT scan images, and the organ feature based on such markers was correlated with ipsilateral lung V20. |
| Dembrower et al.[14] | 2023 | Breast | ScreenTrustCAD | Mammogram | Replacing one radiologist with AI for independent assessment of screening mammograms was non-inferior for cancer detection compared with reading by two radiologists. |
| Preetha et al.[36] | 2021 | CNS | CORE, CENTRIC, EORTC 26101 | MRI scan | Synthetic postcontrast MRI scan based on pre-contrast MRI scanning using deep learning was feasible with no statistically significant difference in the contrast-enhancing tumor burden when compared to postcontrast MRI scanning. |
| Tsang et al.[37] | 2024 | CNS | Not available | CT scan | 94% of ML plans and 93% of manual plans were deemed to be clinically acceptable. ML plans were able to give 1 Gy less radiation to the normal brain than the manual plan. ML plans required 45 fewer minutes on average to create compared to manual plans. |
| George et al.[38] | 2024 | CNS | ClinicalTrials.gov ID: NCT02336165 | MRI scan | First on-treatment MRI features were correlated with overall and progression-free survival, while baseline MRI features were not. |
| Hong et al.[27] | 2020 | General | SHIELD-RT | Clinical variables | AI-based algorithm based on routine electronic health record data triaged patients and reduced acute care visits during treatments. |
| Friesner et al.[39] | 2022 | General | NCT02649569, NCT03102229, NCT03115398 | Daily step counts | Daily step counts using an ML model were correlated with hospitalizations. |
| Kehayias et al.[40] | 2024 | General | Not available | CT scan | The integration of Deep Learning On-Demand Assistant, an automated clinical platform to help with auto-segmentations and QA reporting using AI, into radiation oncology clinic workflow was feasible. |



https://doi.org/10.37549/ARO-D-25-0004

**Table 1. continued**

| AUTHORS | YEAR | DISEASE SITE | PROSPECTIVE DATA | DATA TYPES | MAIN FINDINGS |
|---|---|---|---|---|---|
| Natesan et al.[41] | 2024 | General | SHIELD-RT | Clinical variables | High-risk patients identified by the AI-based algorithm experienced lower total medical costs from twice-weekly evaluations. |
| Wang et al.[42] | 2022 | GI | RTOG 0822 | CT scan | AI-based algorithm using clinical variables, DVH, and radiomic features predicted pCR. |
| Wesdorp et al.[43] | 2023 | GI | CAIRO5 | CT scan | A DL autosegmentation model accurately segmented the liver and metastatic lesions. |
| Fremond et al.[44] | 2023 | GYN | PORTEC-1, PORTEC-2, PORTEC-3, TransPORTEC | Whole-slide images of H&E slides | A DL model predicted molecular classification. |
| Walker et al.[45] | 2014 | Head/Neck | Not available | CT scan | Autosegmentation of organs at risk reduced the amount of time needed for segmentation, but expert oversight is still required for accuracy. |
| Men et al.[46] | 2019 | Head/Neck | RTOG 0522 | CT scan | AI-based algorithm predicted the incidence of late xerostomia. |
| Sher et al.[47] | 2021 | Head/Neck | Not available | Radiation plans | AI-based decision support tool improved the dose metrics for organs at risk. |
| Osapoetra et al.[48] | 2021 | Head/Neck | ClinicalTrials.gov ID: NCT03908684 | Quantitative ultrasound | AI-based algorithm predicted treatment response of involved lymph nodes. |
| Mashayekhi et al.[49] | 2023 | Head/Neck | Not available | Radiation plans | AI-based decision support tool improved uniformity of practice. |
| Kann et al.[29] | 2023 | Head/Neck | ECOG/ACRIN 3311 | CT scan | AI-based algorithm predicted extranodal extension more effectively than did radiologists. |
| Sher et al.[50] | 2023 | Head/Neck | INRT-AIR | CT scan | AI-based algorithm identified involved or suspicious lymph nodes, and there was no solitary elective nodal recurrence at 2 years without elective nodal irradiation. |
| Nicolae et al.[51] | 2020 | Prostate | Not available | Ultrasound | AI-based radiation treatment planning reduced the time required for planning and was considered clinically acceptable. |
| McIntosh et al.[52] | 2021 | Prostate | Not available | Radiation plans | AI-based radiation treatment planning reduced the time required for planning and was considered clinically acceptable. |
| Sanders et al.[53] | 2022 | Prostate | Not available | MRI scan | Autosegmentation of prostate and organs at risk was considered clinically feasible. |

**Table 1. continued**

| AUTHORS | YEAR | DISEASE SITE | PROSPECTIVE DATA | DATA TYPES | MAIN FINDINGS |
|---|---|---|---|---|---|
| Thomas et al.[54] | 2022 | Prostate | ClinicalTrials.gov ID: NCT03238170 | Radiation plans | AI-based algorithm predicted those who would benefit from rectal spacer placement. |
| Johnsson et al.[55] | 2022 | Prostate | OSPREY | PSMA PET/CT | AI-based algorithm identified potential lesions and autosegmented organs. |
| Esteva et al.[56] | 2022 | Prostate | NRG/RTOG 9202, 9413, 9910, 0126 | Whole slide images of H&E slides | AI-based algorithm risk stratified and identified patients with poor prognoses. |
| Spratt et al.[57] | 2023 | Prostate | NRG/RTOG 9202, 9413, 9910, 0126, 9408 | Whole slide images of H&E slides | AI-based algorithm predicted patients who would benefit from androgen deprivation therapy. |
| Ross et al.[58] | 2024 | Prostate | NRG/RTOG 9902 | Whole slide images of H&E slides | AI-based algorithm risk stratified and identified patients with poor prognoses. |
| Spratt et al.[59] | 2024 | Prostate | NRG/RTOG 9202, 9408, 9413, 9910, 9902, 0521 | Whole slide images of H&E slides | AI-based algorithm risk stratified and identified patients with poor prognoses. |
| Wong et al.[60] | 2020 | Prostate/Head Neck/CNS | Not available | CT scan | AI-based algorithm reduced the time required for contouring and autosegmented at-risk organs and target volumes. |

*Abbreviations: AI, artificial intelligence; CT, computed tomography; CNS, central nervous system; DVH, dose volume histogram; H&E, hematoxylin and eosin; GI; gastrointestinal, GYN, gynaecological; MRI, magnetic resonance imaging; ML, machine learning; PSMA, prostate specific membrane antigen; pCR, pathologic complete response; QA, quality assurance.*

among patients with localized prostate cancer.[64]

## AI in Head and Neck Cancer

Other malignancies targeted by extensive research in AI are head and neck cancers, especially with respect to radiomics. For example, ENE is a known adverse feature associated with poor locoregional control.[65,66] However, ENE identification has been largely based on pathologic evaluation, since radiographic determination has been inconsistent.[67-69] As a result, 24%-31% of patients with p16+ head and neck cancer receive trimodality therapy.[70,71] To reduce this knowledge gap, Kann et al. developed a deep-learning (DL) algorithm based on 270 patients from a single institution with over 650 lymph nodes segmented.[72] The model predicted ENE and nodal metastasis with an AUC of 0.91 for both endpoints.[72] Based on such early success, Kann et al. utilized validation datasets of 82 patients with 130 lymph nodes segmented from Mount Sinai Hospital and 62 patients with 70 lymph nodes segmented from The Cancer Genome Atlas imaging data through The Cancer Imaging Archive.[28] The DL model predicted ENE with an AUC of 0.84-0.90 on these validation datasets, outperforming diagnostic radiologists and improving interobserver agreement among these radiologists.[28] Owing to the small sample size of p16-positive oropharyngeal cancer in these retrospective datasets,[28] further validation was performed using a multicenter phase II clinical trial, ECOG-ACRIN 3311.[29] The DL model was retrained using three retrospective datasets as mentioned previously, ultimately identifying 178 patients from ECOG-ACRIN 3311 with 313 manually segmented lymph nodes.[29] It had an AUC of 0.86 for the identification of ENE, outperforming four radiologists, with a limitation of node level segmentation required prior to independent testing.[29]

Another evolving paradigm for treatment de-escalation among patients with head and neck cancer is to reduce treatment volume. Several phase II clinical trials and a large retrospective study demonstrated the feasibility of reducing the dose of elective nodal irradiation to 30-40 Gy.[73-75] To omit elective nodal irradiation, colleagues from the University of Texas Southwestern Medical Center evaluated several DL models using 129 patients and over 700 lymph nodes segmented with AUC of 0.88-0.98,[76-78] comparable to the AUC of 0.91 from the study by Kann et al.[72] Subsequently,



https://doi.org/10.37549/ARO-D-25-0004

Sher et al. incorporated this model in the prospective phase II INRT-AIR trial.[50] Of 67 patients with nonmetastatic head and neck cancer who underwent definitive radiation or chemoradiation, an average of 31 lymph nodes per patient were evaluated by the DL model, determining that approximately 10% were involved.[50] At a median follow-up of 33 months, overall and progression-free survival at 2 years were favorable at 91% and 82%, respectively.[50] One patient with heavy marijuana use had an out-of-field elective nodal recurrence with concurrent distant metastasis, but the study otherwise found favorable quality of life outcomes with no solitary elective nodal failure.[50]

### AI in Supportive Care

In addition to improving oncologic outcomes, another area incorporating AI is the effort to reduce acute care visits, such as emergency department visits and unplanned hospitalizations. Predicting such events has been investigated among patients without a cancer diagnosis.[79-82]

In radiation oncology, Hong et al. initially developed a machine learning (ML) model based on nearly 7000 patients with over 8000 treatment courses at a single institution; this model included variables such as baseline demographics, disease and treatment characteristics, prior acute care visits, laboratory values, and recent vital signs.[83] Internal validation demonstrated an AUC of 0.80 for the ML model in predicting acute care visits.[83] Subsequently, Hong et al. performed the SHIELD-RT single-institution, prospective quality improvement study.[27] This model was utilized to identify high-risk patients, who were defined as having more than a 10% risk of acute care visits, and randomized them to twice-weekly on-treatment visits versus standard of care.[27] Of nearly 1000 treatment courses, 311 were evaluated as high-risk courses, with the majority of patients having gastrointestinal cancer or primary brain cancer.[27] The ML model had a favorable performance with an AUC of 0.82 for triaging patients to high- versus low-risk for acute care visits, and fewer than 3% of low-risk patients had acute care visits.[27] Twice-weekly evaluation led to a reduction from 22% to 12% of acute care visits during radiation therapy, the primary endpoint of this study.[27] Furthermore, a post-hoc economic analysis showed that such a reduction in acute care visits translated to lower health care costs.[41]

### Ongoing Clinical Trials

**Table 2** consists of a list of ongoing clinical trials that incorporate AI. In particular, a multimodal AI risk-stratification developed by Spratt et al.[57,59] has been incorporated into two such clinical trials. The HypoElect study (ClinicalTrials.gov ID: NCT06582446) is a single-arm phase II clinical trial that consists of patients with NCCN high-risk, multimodal AI high-risk prostate cancer and is evaluating the role of whole-pelvis radiation in five fractions with radiation dose escalation using brachytherapy and two years of ADT. The second study is the (ClinicalTrials.gov ID: NCT06772441), a single-arm, phase II HypoPro clinical trial comprising patients with NCCN high-risk, multimodal AI low-/intermediate-risk prostate cancer and is investigating SBRT in combination with brachytherapy and concurrent ADT. Additionally, while most ongoing clinical trials leverage AI for adaptive radiation therapy (**Table 2**), another noteworthy study is a randomized clinical trial by researchers at the University of Hong Kong (ClinicalTrials.gov ID: NCT06636188). It is the first prospective study incorporating a chatbot, Digi-Coach, to help reduce physical and psychological distress versus usual nursing care among patients with head and neck cancer.

### Limitations

Limitations of this study include its utilization only of prospective studies while excluding retrospective studies and other types of journal articles. The rationale for this decision is that several published reviews already incorporate retrospective studies to discuss the role of AI in radiation oncology.[30-33] As a result, however, bias may be introduced toward reporting studies from major cancer centers with access to experts with significant AI technical skills. Subsequently, results from these prospective studies may not be generalizable to or implemented in smaller community cancer centers without access to such AI expertise. For instance, significant barriers hindered implementation of the SHIELD-RT trial process; these included labor-intensive, manual verification of treatment course data for each eligible patient, generating and verifying AI predictions by multiple investigators for each enrolled patient, and manually deploying clinical alerts for treating physicians and enrolled patients to ensure that the intervention was completed on time per protocol.[84] In addition, discussion of commercially available technologies is beyond the scope of this review. These have been comprehensively discussed by NRG Oncology in its summary of the roles of commercial products in adaptive radiation, autosegmentation, treatment planning, and clinical trial development.[85-88] Lastly, despite our efforts to include prospective AI data, we may have inadvertently excluded other relevant studies from

**Table 2. Ongoing Prospective Studies**

| CLINICAL TRIAL | CLINICAL TRIALS.GOV ID | START DATE | ESTIMATED END DATE | STUDY DESIGN | ROLE OF AI | STATUS | DISEASE SITE |
|---|---|---|---|---|---|---|---|
| Artificial Intelligencel for Prostate Cancer Treatment Planning | NCT04441775 | 2020 | 2022 | Observational | Improve consistency and quality of radiation treatment plans. | Completed | Prostate |
| Two Studies for Patients With High Risk Prostate Cancer Testing Less Intense Treatment for Patients With a Low Gene Risk Score and Testing a More Intense Treatment for Patients With a High Gene Risk Score, The PREDICT-RT Trial | NCT04513717 | 2020 | 2033 | Interventional | Radiation therapy quality assurance using an AI algorithm. | Recruiting | Prostate |
| ARtificial Intelligence for Gross Tumor Volume Segmentation (ARGOS) | NCT05775068 | 2021 | 2024 | Observational | Autosegmentation of GTV on CT scan. | Active, not recruiting | Thoracic |
| Artificial Intelligence in Functional Imaging for Individualized Treatment of Head and Neck Squamous Cell Carcinoma Patients (KIVAL-KHT) | NCT05192655 | 2021 | 2026 | Observational | Analysis of diagnostic imaging and clinical and histopathological data to predict outcomes. | Recruiting | Head/Neck |
| AI for Head Neck Cancer Treated With Adaptive RadioTherapy (RadiomicART) | NCT05081531 | 2021 | 2024 | Interventional | Analysis of diagnostic imaging to predict outcomes and toxicities. | Recruiting | Head/Neck |
| PostRadiotherapy MRI-based AI System to Predict Radiation Proctitis for Pelvic Cancers | NCT04918992 | 2021 | TBD | Observational | Analysis of post-radiation MRI scan to predict proctitis. | Unknown status | General |
| Clinical Validation of AI-Assisted Radiotherapy Contouring Software for Thoracic Organs At Risk | NCT05787522 | 2022 | 2024 | Observational | Autosegmentation of organs at risk on CT scan. | Completed | Thoracic |
| Simulation-Free Hippocampal-Avoidance Whole Brain Radiotherapy Using Diagnostic MRI-Based and Cone Beam Computed Tomography-Guided On-Table Adaptive Planning in a Novel Ring Gantry Radiotherapy Device | NCT05096286 | 2022 | 2022 | Interventional | Simulation-free workflow using a semi-automated planning based on AI. | Completed | CNS |
| The Impact of Radiotherapy on Oligometastatic Cancer | NCT05933876 | 2022 | 2037 | Observational | Analysis of clinical data, medical images, and biological samples to predict who will benefit from radiation to oligometastatic sites. | Recruiting | General |



https://doi.org/10.37549/ARO-D-25-0004

**Table 2. continued**

| CLINICAL TRIAL | CLINICAL TRIALS.GOV ID | START DATE | ESTIMATED END DATE | STUDY DESIGN | ROLE OF AI | STATUS | DISEASE SITE |
|---|---|---|---|---|---|---|---|
| Intensive Symptom Surveillance Guided by Machine Learning-Directed Risk Stratification in Patients With Non-Metastatic Head and Neck Cancer, The INSIGHT Trial | NCT05338905 | 2022 | 2027 | Interventional | Analysis of clinical data to identify high-risk patients who will benefit from symptom surveillance | Recruiting | Head/Neck |
| Artificial Intelligence in CNS Radiation Oncology (AI-RAD) | NCT06036394 | 2023 | 2028 | Observational | Autosegmentation of tumor and organs at risk, use radiomics to predict toxicities and outcomes. | Active, not recruiting | CNS |
| Stereotactic Body Radiation Therapy Planning With Artificial Intelligence-Directed Dose Recommendation for Treatment of Primary or Metastatic Lung Tumors, RAD-AI Study | NCT05802186 | 2023 | 2026 | Interventional | AI to guide radiation dose for primary lung cancer and lung metastases. | Recruiting | Thoracic |
| Adaptive Radiation in Anal Cancer | NCT05838391 | 2023 | 2025 | Interventional | Adaptive radiation using AI. | Recruiting | GI |
| Randomized Evaluation of Machine Learning Assisted Radiation Treatment Planning versus Standard Radiation Treatment Planning | NCT05979883 | 2023 | 2026 | Interventional-Phase III | AI-assisted radiation treatment planning. | Recruiting | Head/Neck |
| MR-guidance in Chemoradiotherapy for Cervical Cancer (AIM-C1) | NCT06142760 | 2023 | 2026 | Interventional | Adaptive radiation using AI. | Recruiting | GU |
| Daily-Adaptive Stereotactic Body Radiation Therapy for Biochemically Recurrent, Radiologic Apparent Prostate Cancer After Radical Prostatectomy | NCT05946824 | 2023 | 2028 | Interventional-Phase II | Adaptive radiation using AI. | Recruiting | Prostate |
| Computed Tomography-Guided Stereotactic Adaptive Radiotherapy (CT-STAR) for the Treatment of Central and Ultra-Central Early-Stage Non-Small Cell Lung Cancer | NCT05785845 | 2023 | 2026 | Interventional | Adaptive radiation using AI. | Recruiting | Thoracic |
| A Chatbot to Reduce Physical and Psychological Distress of Patients With Head and Neck Cancer Undergoing Radiotherapy | NCT06636188 | 2024 | 2027 | Interventional | AI-based patient navigator chatbot to reduce physical and psychological distress. | Active, not recruiting | Head/Neck |

**Table 2. continued**

| CLINICAL TRIAL | CLINICAL TRIALS.GOV ID | START DATE | ESTIMATED END DATE | STUDY DESIGN | ROLE OF AI | STATUS | DISEASE SITE |
|---|---|---|---|---|---|---|---|
| Glioma Adaptive Radiotherapy With Development of an Artificial Intelligence Workflow (GLADIATOR) | NCT06492486 | 2024 | 2028 | Interventional-Phase II | Adaptive radiation using AI. | Not yet recruiting | CNS |
| AI as an Aid for Weekly Symptom Intake in Radiotherapy | NCT06525181 | 2024 | 2024 | Interventional | Medical documentation for on-treatment visits to improve accuracy and efficiency. | Not yet recruiting | General |
| A phase II Clinical Trial of Artificial Intelligence-assisted One-stop Radiotherapy for Breast Cancer After Breast-conserving Surgery (BC-AIO) | NCT06686459 | 2024 | 2027 | Interventional-Phase II | Autosegmentation and radiation treatment planning. | Not yet recruiting | Breast |
| Evaluation of a Novel Auto Segmentation Algorithm for Normal Structure Delineation in Radiation Treatment Planning | NCT06200116 | 2024 | 2026 | Observational | Autosegmentation. | Recruiting | General |
| Online Adaptive Radiotherapy for Nasopharyngeal Carcinoma (OART) | NCT06516133 | 2024 | 2030 | Phase III Clinical Trial | Adaptive radiation using AI. | Recruiting | Head/Neck |
| One Fraction Simulation-Free Treatment With CT-Guided Stereotactic Adaptive Radiotherapy for Patients With Oligometastatic and Primary Lung Tumors (ONE STOP) | NCT06236516 | 2024 | 2025 | Phase III Clinical Trial | Adaptive radiation using AI. | Recruiting | Thoracic |
| Artificial Intelligence to Personalize Prostate Cancer Treatment (the HypoElect Trial) (HypoElect) | NCT06582446 | 2024 | 2027 | Interventional-Phase II | Patient selection and risk stratification. | Recruiting | Prostate |
| Artificial Intelligence Driven Personalisation of Radiotherapy and Concomitant Androgen Deprivation Therapy for Prostate Cancer Patients (the HypoPro Trial) (HypoPro) | NCT06772441 | 2024 | 2027 | Interventional | Patient selection and risk stratification. | Recruiting | Prostate |
| RAdiotherapy With FDG-PET Guided Dose-PAINTing Compared With Standard Radiotherapy for Primary Head and Neck Cancer-3 (RADPAINT-3) | NCT06297902 | 2024 | 2030 | Interventional | Analysis of blood samples to predict tumor response and toxicities. | Recruiting | Head/Neck |



https://doi.org/10.37549/ARO-D-25-0004

**Table 2. continued**

| CLINICAL TRIAL | CLINICAL TRIALS.GOV ID | START DATE | ESTIMATED END DATE | STUDY DESIGN | ROLE OF AI | STATUS | DISEASE SITE |
|---|---|---|---|---|---|---|---|
| Artificial Intelligence-Guided Radiotherapy Planning for Glioblastoma (ARTPLAN-GLIO) | NCT06657027 | 2025 | 2027 | Observational | Analysis of MRI scans to evaluate the extent of tumor infiltration. | Not yet recruiting | CNS |
| Locally Optimised Contouring With AI Technology for Radiotherapy (LOCATOR) | NCT06546592 | 2025 | 2029 | Interventional | Autosegmentation. | Not yet recruiting | General |

*Abbreviations: AI, artificial intelligence; CT, computed tomography; GTV, gross tumor volume; MRI, magnetic resonance imaging.*

our review. Further studies are warranted to capture the growing complexity of AI and its impact in radiation oncology.

## Conclusion

Radiation oncology is poised to be influenced substantially by AI in the coming decades. Emerging AI tools will streamline radiation treatment planning and adaptive radiation, guide treatment recommendations by improving patient selection based on digital pathology and radiomics, and tailor supportive care to reduce acute care visits. As a result, such efforts will translate to further progress in radiation oncology and patient outcomes.

## References

1) Bezjak A, Paulus R, Gaspar LE, et al. Safety and efficacy of a five-fraction stereotactic body radiotherapy schedule for centrally located non-small-cell lung cancer: NRG oncology/RTOG 0813 trial. *J Clin Oncol*. 2019;37(15):1316-1325. doi:10.1200/JCO.18.00622

2) Singh AK, Gomez-Suescun JA, Stephans KL, et al. One versus three fractions of stereotactic body radiation therapy for peripheral stage I to II non-small cell lung cancer: a randomized, multi-institution, phase 2 trial. *Int J Radiat Oncol Biol Phys*. 2019;105(4):752-759. doi:10.1016/j.ijrobp.2019.08.019

3) Videtic GMM, Hu C, Singh AK, et al. A randomized phase 2 study comparing 2 stereotactic body radiation therapy schedules for medically inoperable patients with stage I peripheral non-small cell lung cancer: NRG oncology RTOG 0915 (NCCTG N0927). *Int J Radiat Oncol Biol Phys*. 2015;93(4):757-764. doi:10.1016/j.ijrobp.2015.07.2260

4) Deek MP, Van der Eecken K, Sutera P, et al. Long-term outcomes and genetic predictors of response to metastasis-directed therapy versus observation in oligometastatic prostate cancer: analysis of STOMP and ORIOLE trials. *J Clin Oncol*. 2022;40(29):3377-3382. doi:10.1200/JCO.22.00644

5) Palma DA, Olson R, Harrow S, et al. Stereotactic ablative radiotherapy for the comprehensive treatment of oligometastatic cancers: long-term results of the SABR-COMET phase II randomized trial. *J Clin Oncol*. 2020;38(25):2830-2838. doi:10.1200/JCO.20.00818

6) Tsai CJ, Yang JT, Shaverdian N, et al. Standard-of-care systemic therapy with or without stereotactic body radiotherapy in patients with oligoprogressive breast cancer or non-small-cell lung cancer (Consolidative Use of Radiotherapy to Block [CURB] oligoprogression): an open-label, randomised, controlled, phase 2 study. *Lancet*. 2024;403(10422):171-182. doi:10.1016/S0140-6736(23)01857-3

7) Yang JT, Wijetunga NA, Pentsova E, et al. Randomized phase II trial of proton craniospinal irradiation versus photon involved-field radiotherapy for patients with solid tumor leptomeningeal metastasis. *J Clin Oncol*. 2022;40(33):3858-3867. doi:10.1200/JCO.22.01148

8) Kishan AU, Ma TM, Lamb JM, et al. Magnetic resonance imaging-guided vs computed tomography-guided stereotactic body radiotherapy for prostate cancer: the MIRAGE randomized clinical trial. *JAMA Oncol*. 2023;9(3):365-373. doi:10.1001/jamaoncol.2022.6558

9) Chuong MD, Lee P, Low DA, et al. Stereotactic MR-guided on-table adaptive radiation therapy (SMART) for borderline resectable and locally advanced pancreatic cancer: a multi-center, open-label phase 2 study. *Radiother Oncol*. 2024;191:110064. doi:10.1016/j.radonc.2023.110064

10) Castelli J, Thariat J, Benezery K, et al. Weekly adaptive radiotherapy vs standard intensity-modulated radiotherapy for improving salivary function in patients with head and neck cancer: a phase 3 randomized clinical trial. *JAMA Oncol*. 2023;9(8):1056-1064. doi:10.1001/jamaoncol.2023.1352

11) Hofman MS, Lawrentschuk N, Francis RJ, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (propsma): a prospective, randomised, multicentre study. *Lancet*. 2020;395(10231):1208-1216. doi:10.1016/S0140-6736(20)30314-7

12) June CH, Sadelain M. Chimeric antigen receptor therapy. *N Engl J Med*. 2018;379(1):64-73. doi:10.1056/NEJMra1706169

13) Lee NY, Sherman EJ, Schöder H, et al. Hypoxia-directed treatment of human papillomavirus-related oropharyngeal carcinoma. *J Clin Oncol*. 2024;42(8):940-950. doi:10.1200/JCO.23.01308

14) Dembrower K, Crippa A, Colón E, et al. Artificial intelligence for breast cancer detection in screening mammography in sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health*. 2023;5(10):e703-e711. doi:10.1016/S2589-7500(23)00153-X

15) Perlow HK, Nalin AP, Handley D, et al. A prospective registry study of $^{68}$Ga-DOTATATE PET/CT incorporation into treatment planning of intracranial meningiomas. *Int J Radiat Oncol Biol Phys*. 2024;118(4):979-985. doi:10.1016/j.ijrobp.2023.10.014

16) Perlow HK, Raleigh DR, Wang TJC, et al. Consensus radiation treatment planning guidelines using (68)Ga-DOTATATE PET/CT for resected meningiomas. *Int J Radiat Oncol Biol Phys*. 2025;122(1):150-158. doi:10.1016/j.ijrobp.2024.12.003

17) Kalinsky K, Barlow WE, Gralow JR, et al. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. *N Engl J Med*. 2021;385(25):2336-2347. doi:10.1056/NEJMoa2108873

18) Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*. 2018;379(2):111-121. doi:10.1056/NEJMoa1804710

19) Lu S, Kato T, Dong X, et al. osimertinib after chemoradiotherapy in stage III *EGFR* -mutated NSCLC . *N Engl J Med*. 2024;391(7):585-597. doi:10.1056/NEJMoa2402614

20) Howell MD, Corrado GS, DeSalvo KB. Three epochs of artificial intelligence in health care. *JAMA*. 2024;331(3):242-244. doi:10.1001/jama.2023.25057

21) McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. doi:10.1038/s41586-019-1799-6

22) Cao K, Xia Y, Yao J, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat Med*. 2023;29(12):3033-3043. doi:10.1038/s41591-023-02640-w

23) Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25(6):954-961. doi:10.1038/s41591-019-0447-x

24) Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med*. 2020;26(6):900-908. doi:10.1038/s41591-020-0842-3

25) Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838

26) Nayak A, Alkaitis MS, Nayak K, et al. Comparison of history of present illness summaries generated by a chatbot and senior internal medicine residents. *JAMA Intern Med*. 2023;183(9):1026-1027. doi:10.1001/jamainternmed.2023.2561

27) Hong JC, Eclov NCW, Dalal NH, et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): a prospective randomized study of machine learning-directed clinical evaluations during radiation and chemoradiation. *J Clin Oncol*. 2020;38(31):3652-3661. doi:10.1200/JCO.20.01688

28) Kann BH, Hicks DF, Payabvash S, et al. Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. *J Clin Oncol*. 2020;38(12):1304-1311. doi:10.1200/JCO.19.02031

29) Kann BH, Likitlersuang J, Bontempi D, et al. Screening for extranodal extension in HPV-associated oropharyngeal carcinoma: evaluation of a CT-based deep learning algorithm in patient data from a multicentre, randomised de-escalation trial. *Lancet Digit Health*. 2023;5(6):e360-e369. doi:10.1016/S2589-7500(23)00046-8

30) Huynh E, Hosny A, Guthier C, et al. Artificial intelligence in radiation oncology. *Nat Rev Clin Oncol*. 2020;17(12):771-781. doi:10.1038/s41571-020-0417-8

31) Kalsi S, French H, Chhaya S, et al. The evolving role of artificial intelligence in radiotherapy treatment planning-a literature review. *Clin Oncol (R Coll Radiol)*. 2024;36(10):596-605. doi:10.1016/j.clon.2024.06.005

32) Kawamura M, Kamomae T, Yanagawa M, et al. Revolutionizing radiation therapy: the role of AI in clinical practice. *J Radiat Res*. 2024;65(1):1-9. doi:10.1093/jrr/rrad090

33) Parkinson C, Matthams C, Foley K, Spezi E. Artificial intelligence in radiation oncology: a review of its current status and potential application for the radiotherapy workforce. *Radiography (Lond)*. 2021;27 Suppl 1:S63-S68. doi:10.1016/j.radi.2021.07.012

34) Zeleznik R, Weiss J, Taron J, et al. Deep-learning system to improve the quality and efficiency of volumetric heart segmentation for breast cancer. *NPJ Digit Med*. 2021;4(1):43. doi:10.1038/s41746-021-00416-5

35) Ma L, Yang Y, Ma J, et al. Correlation between AI-based CT organ features and normal lung dose in adjuvant radiotherapy following breast-conserving surgery: a multicenter prospective study. *BMC Cancer*. 23(1):1085. doi:10.1186/s12885-023-11554-2

36) Jayachandran Preetha C, Meredig H, Brugnara G, et al. Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. *Lancet Digit Health*. 2021;3(12):e784-e794. doi:10.1016/S2589-7500(21)00205-3

37) Tsang DS, Tsui G, Santiago AT, et al. A prospective study of machine learning-assisted radiation therapy planning for patients receiving 54 Gy to the brain. *Int J Radiat Oncol Biol Phys*. 2024;119(5):1429-1436. doi:10.1016/j.ijrobp.2024.02.022

38) George E, Flagg E, Chang K, et al. Radiomics-based machine learning for outcome prediction in a multicenter phase II study of programmed death-ligand 1 inhibition immunotherapy for glioblastoma. *AJNR Am J Neuroradiol*. 2022;43(5):675-681. doi:10.3174/ajnr.A7488

39) Friesner ID, Feng J, Kalnicki S, et al. Machine learning-based prediction of hospitalization during chemoradiotherapy with daily step counts. *JAMA Oncol*. 2024;10(5):642-647. doi:10.1001/jamaoncol.2024.0014

40) Kehayias CE, Yan Y, Bontempi D, et al. Prospective deployment of an automated implementation solution for artificial intelligence translation to clinical radiation oncology. *Front Oncol*. 2023;13:1305511. doi:10.3389/fonc.2023.1305511

41) Natesan D, Eisenstein EL, Thomas SM, et al. Health care cost reductions with machine learning-directed evaluations during radiation therapy—an economic analysis of a randomized controlled study. *NEJM AI*. 2024;1(4):AIoa2300118. doi:10.1056/aioa2300118

42) Wang D, Lee SH, Geng H, et al. Interpretable machine learning for predicting pathologic complete response in patients treated with chemoradiation therapy for rectal adenocarcinoma. *Front Artif Intell*. 2022;5:1059033. doi:10.3389/frai.2022.1059033

43) Wesdorp NJ, Zeeuw JM, Postma SCJ, et al. Deep learning models for automatic tumor segmentation and total tumor volume assessment in patients with colorectal liver metastases. *Eur Radiol Exp*. 2023;7(1):75. doi:10.1186/s41747-023-00383-4

44) Fremond S, Andani S, Barkey Wolf J, et al. Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. *Lancet Digit Health*. 2023;5(2):e71-e82. doi:10.1016/S2589-7500(22)00210-2

45) Walker GV, Awan M, Tao R, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol*. 2014;112(3):321-325. doi:10.1016/j.radonc.2014.08.028

46) Men K, Geng H, Zhong H, et al. A deep learning model for predicting xerostomia due to radiation therapy for head and neck squamous cell carcinoma in the RTOG 0522 clinical trial. *Int J Radiat Oncol Biol Phys*. 2019;105(2):440-447. doi:10.1016/j.ijrobp.2019.06.009

47) Sher DJ, Godley A, Park Y, et al. Prospective study of artificial intelligence-based decision support to improve head and neck radiotherapy plan quality. *Clin Transl Radiat Oncol*. 2021;29:65-70. doi:10.1016/j.ctro.2021.05.006

48) Osapoetra LO, Dasgupta A, DiCenzo D, et al. Assessment of clinical radiosensitivity in patients with head-neck squamous cell carcinoma from pre-treatment quantitative ultrasound radiomics. *Sci Rep*. 2021;11(1):6117. doi:10.1038/s41598-021-85221-6

49) Mashayekhi M, McBeth R, Nguyen D, et al. Artificial intelligence guided physician directive improves head and neck planning quality and practice uniformity: a prospective study. *Clin Transl Radiat Oncol*. 2023;40:100616. doi:10.1016/j.ctro.2023.100616

50) Sher DJ, Moon DH, Vo D, et al. Efficacy and quality-of-life following involved nodal radiotherapy for head and neck squamous cell carcinoma: the INRT-AIR phase II clinical trial. *Clin Cancer Res*. 2023;29(17):3284-3291. doi:10.1158/1078-0432.CCR-23-0334

51) Nicolae A, Semple M, Lu L, et al. Conventional vs machine learning-based treatment planning in prostate brachytherapy: results of a phase I randomized controlled trial. *Brachytherapy*. 2020;19(4):470-476. doi:10.1016/j.brachy.2020.03.004

52) McIntosh C, Conroy L, Tjong MC, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med*. 2021;27(6):999-1005. doi:10.1038/s41591-021-01359-w

53) Sanders JW, Kudchadker RJ, Tang C, et al. Prospective evaluation of prostate and organs at risk segmentation software for MRI-based prostate radiation therapy. *Radiol Artif Intell*. 2022;4(2):e210151. doi:10.1148/ryai.210151

54) Thomas C, Dregely I, Oksuz I, et al. Neural network dose prediction for rectal spacer stratification in dose-escalated prostate radiotherapy. *Med Phys*. 2022;49(4):2172-2182. doi:10.1002/mp.15575

55) Johnsson K, Brynolfsson J, Sahlstedt H, et al. Analytical performance of apromise: automated anatomic contextualization, detection, and quantification of [(18)F]DCFPyL (PSMA) imaging for standardized reporting. *Eur J Nucl Med Mol Imaging*. 2022;49(3):1041-1051. doi:10.1007/s00259-021-05497-8

56) Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med*. 2022;5(1):71. doi:10.1038/s41746-022-00613-w

57) Spratt DE, Tang S, Sun Y, et al. Artificial intelligence predictive model for hormone therapy use in prostate cancer. *NEJM Evid*. 2023;2(8):EVIDoa2300023. doi:10.1056/EVIDoa2300023

58) Ross AE, Zhang J, Huang H-C, et al. External validation of a digital pathology-based multimodal artificial intelligence architecture in the NRG/RTOG 9902 phase 3 trial. *Eur Urol Oncol*. 2024;7(5):1024-1033. doi:10.1016/j.euo.2024.01.004

59) Spratt DE, Liu VYT, Jia AY, et al. Meta-analysis of individual patient-level data for a multimodal artificial intelligence biomarker in high-risk prostate cancer: results from six NRG/RTOG phase 3 randomized trials. *Eur Urol*. 2024;86(4):369-371. doi:10.1016/j.eururo.2024.06.019

60) Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol*. 2020;144:152-158. doi:10.1016/j.radonc.2019.10.019

61) Glaessgen A, Hamberg H, Pihl C-G, et al. Interobserver reproducibility of percent gleason grade 4/5 in total prostatectomy specimens. *J Urol*. 2002;168(5):2006-2010. doi:10.1016/S0022-5347(05)64283-8

62) Sadimin ET, Khani F, Diolombi M, Meliti A, Epstein JI. Interobserver reproducibility of percent gleason pattern 4 in prostatic adenocarcinoma on prostate biopsies. *Am J Surg Pathol*. 2016;40(12):1686-1692. doi:10.1097/PAS.0000000000000714

63) Zhu L, Pan J, Mou W, et al. Harnessing artificial intelligence for prostate cancer management. *Cell Rep Med*. 2024;5(4):101506. doi:10.1016/j.xcrm.2024.101506

64) Network NCC. Prostate cancer (Version 1.2025). Accessed January 22, 2025. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf

65) Bernier J, Domenge C, Ozsahin M, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med*. 2004;350(19):1945-1952. doi:10.1056/NEJMoa032641

66) Cooper JS, Pajak TF, Forastiere AA, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. *N Engl J Med*. 2004;350(19):1937-1944. doi:10.1056/NEJMoa032646

67) Chai RL, Rath TJ, Johnson JT, et al. Accuracy of computed tomography in the prediction of extracapsular spread of lymph node metastases in squamous cell carcinoma of the head and neck. *JAMA Otolaryngol Head Neck Surg*. 2013;139(11):1187-1194. doi:10.1001/jamaoto.2013.4491

68) Maxwell JH, Rath TJ, Byrd JK, et al. Accuracy of computed tomography to predict extracapsular spread in p16-positive squamous cell carcinoma. *Laryngoscope*. 2015;125(7):1613-1618. doi:10.1002/lary.25140

69) Url C, Schartinger VH, Riechelmann H, et al. Radiological detection of extracapsular spread in head and neck squamous cell carcinoma (HNSCC) cervical metastases. *Eur J Radiol*. 2013;82(10):1783-1787. doi:10.1016/j.ejrad.2013.04.024

70) Nichols AC, Theurer J, Prisman E, et al. Randomized trial of radiotherapy versus transoral robotic surgery for oropharyngeal squamous cell carcinoma: long-term results of the ORATOR trial. *J Clin Oncol*. 2022;40(8):866-875. doi:10.1200/JCO.21.01961

71) Ferris RL, Flamand Y, Weinstein GS, et al. Phase II randomized trial of transoral surgery and low-dose intensity modulated radiation therapy in resectable p16+ locally advanced oropharynx cancer: an ECOG-ACRIN cancer research group trial (E3311). *J Clin Oncol*. 2022;40(2):138-149. doi:10.1200/JCO.21.01752

72) Kann BH, Aneja S, Loganadane GV, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. *Sci Rep*. 2018;8(1):14036. doi:10.1038/s41598-018-32441-y

73) Maguire PD, Neal CR, Hardy SM, Schreiber AM. Single-arm phase 2 trial of elective nodal dose reduction for patients with locoregionally advanced squamous cell carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys*. 2018;100(5):1210-1216. doi:10.1016/j.ijrobp.2017.12.277

74) Sher DJ, Pham N-L, Shah JL, et al. Prospective phase 2 study of radiation therapy dose and volume de-escalation for elective neck treatment of oropharyngeal and laryngeal cancer. *Int J Radiat Oncol Biol Phys*. 2021;109(4):932-940. doi:10.1016/j.ijrobp.2020.09.063

75) Tsai CJ, McBride SM, Riaz N, et al. Evaluation of substantial reduction in elective radiotherapy dose and field in patients with human papillomavirus-associated oropharyngeal carcinoma treated with definitive chemoradiotherapy. *JAMA Oncol*. 2022;8(3):364-372. doi:10.1001/jamaoncol.2021.6416

76) Chen L, Dohopolski M, Zhou Z, et al. Attention guided lymph node malignancy prediction in head and neck cancer. *Int J Radiat Oncol Biol Phys*. 2021;110(4):1171-1179. doi:10.1016/j.ijrobp.2021.02.004

77) Chen L, Zhou Z, Sher D, et al. Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. *Phys Med Biol*. 64(7):075011. doi:10.1088/1361-6560/ab083a

78) Dohopolski M, Chen L, Sher D, Wang J. Predicting lymph node metastasis in patients with oropharyngeal cancer by using a convolutional neural network with associated epistemic and aleatoric uncertainty. *Phys Med Biol*. 2020;65(22):225002. doi:10.1088/1361-6560/abb71c

79) Donzé JD, Williams MV, Robinson EJ, et al. International validity of the HOSPITAL score to predict 30-day potentially avoidable HOSPITAL readmissions. *JAMA Intern Med*. 2016;176(4):496-502. doi:10.1001/jamainternmed.2015.8462

80) Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol*. 2017;2(2):204-209. doi:10.1001/jamacardio.2016.3956

81) Nguyen OK, Makam AN, Clark C, et al. Predicting all-cause readmissions using electronic health record data from the entire hospitalization: model development and comparison. *J Hosp Med*. 2016;11(7):473-480. doi:10.1002/jhm.2568

82) Tevis SE, Weber SM, Kent KC, Kennedy GD. Nomogram to predict postoperative readmission in patients who undergo general surgery. *JAMA Surg*. 2015;150(6):505-510. doi:10.1001/jamasurg.2014.4043

83) Hong JC, Niedzwiecki D, Palta M, Tenenbaum JD. Predicting emergency visits and hospital admissions during radiation and chemoradiation: an internally validated pretreatment machine learning algorithm. *JCO Clin Cancer Inform*. 2018;2:1-11. doi:10.1200/CCI.18.00037

84) Hong JC, Eclov NCW, Stephens SJ, Mowery YM, Palta M. Implementation of machine learning in the clinic: challenges and lessons in prospective deployment from the system for high intensity evaluation during radiation therapy (SHIELD-RT) randomized controlled study. *BMC Bioinformatics*. 2022;23(suppl 12):408. doi:10.1186/s12859-022-04940-3

85) Glide-Hurst CK, Lee P, Yock AD, et al. Adaptive radiation therapy (ART) strategies and technical considerations: a state of the ART review from NRG oncology. *Int J Radiat Oncol Biol Phys*. 2021;109(4):1054-1075. doi:10.1016/j.ijrobp.2020.10.021

86) Jia X, Rong Y, Wu Q, et al. NRG oncology assessment of artificial intelligence for automatic treatment planning in radiation therapy clinical trials: present and future. *Int J Radiat Oncol Biol Phys*. 2025. doi:10.1016/j.ijrobp.2025.03.045

87) Rong Y, Chen Q, Fu Y, et al. NRG oncology assessment of artificial intelligence deep learning-based auto-segmentation for radiation therapy: current developments, clinical considerations, and future directions. *Int J Radiat Oncol Biol Phys*. 2024;119(1):261-280. doi:10.1016/j.ijrobp.2023.10.033

88) Xiao Y, Benedict S, Cui Y, et al. Embracing the future of clinical trials in radiation therapy: an NRG oncology CIRO technology retreat whitepaper on pioneering technologies and AI-driven solutions. *Int J Radiat Oncol Biol Phys*. 2025;122(2):443-457. doi:10.1016/j.ijrobp.2025.01.006

RESEARCH ARTICLE

# Artificial Intelligence-Assisted Peer Review in Radiation Oncology

Renee F. Cattell, PhD;[1]* Jinkoo Kim, PhD; Ewa Zabrocka, MD;[1,2] Xin Qian, PhD; Brian O'Grady, BA;[1] Stephanie Butler, BS; Todd Yoder, MS, CMD;[1,3] Kartik Mani, MD, PhD; Mark Ashamalla, MD; Samuel Ryu, MD

## Abstract

**Objective** Peer review is an essential part of the patient treatment process that examines and, where necessary, recommends revisions to clinical data, therapeutic parameters, and potential alternative approaches to treatment. Our hypothesis is that artificial intelligence (AI) and machine language technologies can enhance peer-review efficacy by screening cases for potential treatment interruptions caused by re-planning and treatment cessation.

**Materials and Methods** Fifty-five features of clinical and therapeutic parameters from 3881 radiotherapy patients (7142 plans) treated from 2014 to 2021 were used as input for two AI models: a multivariable least absolute shrinkage and selection operator (LASSO) logistic regression model and a pattern recognition feed-forward neural network (NN). The dataset was split into 70% training and 30% testing, with the training set divided into five groups for cross-validation. Analysis was performed on the full cohort and on subsets based on treatment site. Performance metrics of accuracy, sensitivity, and specificity were calculated.

**Results** Overall, 8.1% of all cases had treatment interruptions, most commonly in the head and neck region compared to other sites (19% vs 6%-9%, $P<.01$). For the LASSO model, test set sensitivity, specificity, and accuracy ranged from 37%-70%, 59%-78%, and 60%-76%, respectively, with higher specificity than sensitivity for site subsets. For the NN model, test set sensitivity, specificity, and accuracy ranged from 41%-68%, 53%-79%, and 53%-78%, respectively. Both models demonstrated the highest accuracy in the brain subset. For the full cohort, NN accuracy (58%) was similar to LASSO (60%). The largest accuracy differences between LASSO and NN were in the lung/breast/chest (LASSO: 71% vs NN: 57%) and spine/extremity (LASSO: 66% vs NN: 54%) subsets.

**Conclusion** Our results provide proof-of-concept that AI- and ML-based technologies have potential as screening tools to aid peer review in radiation oncology. Early identification of patients at risk for radiation therapy interruptions using these tools could translate into higher treatment completion rates. The study is being continued to include more clinical features and to optimize model hyperparameters.

**Keywords:** peer review, artificial intelligence, machine learning, radiation oncology

**Affiliations:** [1]Department of Radiation Oncology, Stony Brook University Hospital, Stony Brook, NY. [2]Department of Radiation Oncology, Anchorage and Valley Radiation Therapy Centers, Anchorage, AK. [3]Department of Radiation Oncology, NYU Langone Health, New York, NY.
**Corresponding author:** *Renee F. Cattell, PhD, Stony Brook University Hospital, 101 Nicolls Rd, Stony Brook, NY 11794. (renee.cattell@stonybrookmedicine.edu)
**Disclosures:** The authors have no conflicts of interest to disclose. None of the authors received outside funding for the production of this original manuscript and no part of this article has been previously published elsewhere.
**Prior Publication/Presentation:** Some of the material in this article was published as an abstract in the following supplement for the ASTRO 2022 Annual Meeting: Cattell R, Ashamalla M, Kim J, et al. Artificial intelligence–assisted peer review in radiation oncology. *Int J Radiat Oncol Biol Phys*. 2022;114(3)(suppl):e471. doi:10.1016/j.ijrobp.2022.07.1726.

https://doi.org/10.37549/ARO-D-24-00025

## Introduction

Peer review in radiation oncology is essential for the safe and efficient delivery of radiation treatments;[1] however, little guidance and limited research exist regarding the frequency, mechanisms, and metrics of peer review from professional organizations.[1-4] Peer review addresses patient characteristics and clinical oncological information, planned radiotherapeutic parameters, and potential treatment-related variations. In a report series commissioned by the American Society for Radiation Oncology, seven main items are currently examined by peer reviewers.[1] These are (1) the decision to include radiation as part of treatment, (2) the general radiation treatment approach, (3) the target definition, (4) normal tissue image segmentation, (5) the planning directive, (6) technical plan quality, and (7) treatment delivery.

Many of these factors are indicative of cases requiring re-plans or cases that may have treatment interruptions. Although peer reviews should be conducted prior to the start of treatment to ensure safety and quality, they are typically performed either immediately before or after treatment has begun. Thus, there is minimal time to adjust therapy based on the recommendations of peer reviewers. Plan changes based on peer review discussions are not uncommon. Hoopes et al reported that 90% of physicians have changed their radiation plans because of peer review.[3] Other studies have shown that up to 10% of cases have been recommended for plan modifications based on peer review.[1,3] Studies have also demonstrated that prolonged radiation treatment times, for various reasons, correlated with reduced local disease control and worse overall survival.[5-7] Therefore, timely, efficient peer review may greatly improve treatment quality and, ultimately, enhance patient safety.

However, peer review has some practical difficulties, especially when the process must be coordinated among multiple facilities. While a fully integrated approach holds promise for improving the quality, safety, and value of cancer care, in reality, the process is often disjointed owing to differing provider schedules, caseloads, and expertise. These opportunities to optimize peer review are further highlighted by the fact that different approaches are taken by academic centers and community cancer centers operating within the same network.[8] Practically speaking, however, caseload, time, department resources, staffing, and increasing complexity in treatment techniques limit thorough peer review.[9] Thus, detecting cases that may experience treatment interruptions due to undesirable effects is not easy.

A pan-Canadian survey by Caissie et al demonstrated that barriers to peer review, including time constraints (27%) and radiation oncologist availability (34%), caused half of all programs surveyed to conduct peer review after the start of treatment.[10] Therefore, automating peer review, even if only partially, may significantly streamline the process.

Artificial intelligence (AI) tools can potentially be used to assist in peer review. Studies have used machine learning (ML) techniques to identify high-risk patients for detailed clinical evaluation during radiation and chemoradiation.[11] To identify potential treatment interruptions or unusual side effects resulting from suboptimal treatment plans, we hypothesize that complex, nonlinear relationships exist between different variables, including clinical characteristics and plan parameters. Our approach was to use AI and ML to uncover complex interactions among variables and to supplement clinical knowledge with treatment and plan parameters. By using AI, our intent was to screen treatment plans for clinical and radiotherapeutic factors that could lead to treatment interruptions or toxicity and submit the plans for more detailed inspection to help expedite peer review.

## Methods and Materials

### Data Collection

This retrospective study was approved by the Institutional Review Board. Clinical data and radiation therapy plan parameters from 3881 patient records (7142 plans) were retrospectively extracted from electronic medical records and treatment planning systems from January 2014 to March 2021 for patients older than 18 years. The study inclusion criteria consisted of patients who received external beam photon radiation therapy with either curative or palliative intent for initial treatment and/or re-treatment. This study excluded patients who underwent brachytherapy and electron therapy. Included patients were further divided into subsets based on bodily treatment sites (**Table 1**). Comparisons were made using two-tailed *t* tests with unequal variance.

The aim of this study was to identify patients likely to experience treatment interruptions (eg, changes in target volume, changes in prescription dose) or complications (eg, toxicity). As a surrogate for this outcome, plans with treatment interruptions (remaining fractions) were designated as abnormal cases,


https://doi.org/10.37549/ARO-D-24-00025

**Table 1. Distribution of Plans With and Without Treatment Interruptions Across Separate Subset Groups. "All" Indicates the Full Cohort Before Separated into Subsets Based on Treatment Site**

| TREATMENT SITE | TOTAL NUMBER OF PLANS | PLANS WITH TREATMENT INTERRUPTIONS | PLANS WITHOUT TREATMENT INTERRUPTIONS |
|---|---|---|---|
| All | 7142 | 579 (8.1%) | 6563 (91.9%) |
| Lung, breast, and chest | 3329 | 228 (6.8%) | 3101 (93.2%) |
| Pelvis and prostate | 1077 | 90 (8.4%) | 987 (91.6%) |
| Spine and extremity | 983 | 92 (9.4%) | 891 (90.6%) |
| Brain | 1285 | 81 (6.3%) | 1204 (93.7%) |
| Head and neck | 468 | 88 (18.8%) | 380 (81.2%) |

whereas radiation therapy plans with no remaining fractions (ie, no interruptions) were considered to be normal cases. We defined remaining fractions as any plan that was not completed; we did not separate re-planned cases versus discontinued cases.

Most interruptions are the result of toxicity or treatment response based on radiosensitivity and tumor histology. This surrogate, although imperfect, was easily translated into an instance that could be extracted automatically from the patient records. We note that the specific reason(s) for treatment interruption were not included in this study, as our intention was to identify potential cases and prevent treatment modification. Various clinical factors and therapeutic technical parameters were collected for logistic regression analysis and developing neural network (NN) models. The input factors are listed in **Table 2**.

### Predictive Modeling

The two classification models used in this study were a multivariable logistic regression with least absolute shrinkage and selection operator (LASSO) and a pattern recognition feed forward NN. For LASSO, the maximum number of non-zero coefficients was set to 10. The NN model had one hidden layer with 10 neurons, and the weights were initialized with random seeds. The scaled conjugate gradient algorithm was used for training, and the mean absolute error was the cost function. Regularization was performed using error weights to prevent overfitting of the NN model due to the unbalanced dataset; entry samples with treatment interruptions had two times the weight of those without treatment interruption for training.

For both techniques, the dataset was first split into a 70% training set and a 30% testing set. The training set was normalized with z-score normalization, with the same center and standard deviation normalization applied to the testing set. The training set was further divided into five folds for cross-validation. Within each cross-validation, the minority class was oversampled using adaptive synthetic sampling.[12] The validation fold was not oversampled. The testing set was not oversampled or included in the training of the algorithms. The number of input features to each model was 55. MATLAB (2022a, The MathWorks, Inc., Natick Massachusetts, United States) was used for development, training, and evaluation of both models.

### Predictive Performance Evaluation

The full cohort and subsets were analyzed based on treatment site. The optimal operating point threshold of the receiver-operating characteristic curve was determined only from the training set. This same threshold was applied to the validation and testing sets. Sensitivity, specificity, and accuracy were calculated. The model with the greatest validation accuracy across the five-fold cross-validation was used to predict the independent testing set to assess predictive performance. Comparisons were made using two-tailed paired *t* tests.

## RESULTS

### Description of Cohort

A total of 3881 patients with 7142 plans made up the full cohort. Of the cohort, 58.1% were between 50 and 74 years, 33.1% were >75 years, and 8.8% were between 19 and 49 years. There were more females than males (59.6% vs 40.4%). The most common treatment sites were the breasts (30.2%), brain (18.0%), and lungs (13.0%). Most patients were treated with primary definitive intent (74.6%) versus treatment of metastatic disease (25.4%). **Table 1** summarizes the study cohort. As shown, 8.1% of all plans experienced treatment interruptions. The head-and-neck subset had the largest percentage of treatment interruptions (18.8%), while the other subsets ranged from 6.3% to 9.4% ($P < .01$).

### Testing Set Breakdown

The model with the highest accuracy from the five-fold cross-validation was selected as the

**Table 2. Input Features for the LASSO and NN Models**

| CLINICAL FEATURES | PLAN PARAMETER FEATURES |
|---|---|
| Patient sex (male, female) | Number of beams |
| Patient age group in years ( ≥75, 50-74,19-49) | Type of plan (IMRT, RapidArc, other) |
| Site | Monitor units (total, average, minimum, maximum) |
| Primary or metastasis | Source to skin distance (SSD; average, minimum, maximum) |
| Inpatient status | Table tolerance |
| Personal history of cancer | Gantry angle (minimum, maximum) |
| Family history of cancer | Bolus |
| Tobacco use | Gating |
| Smoker | Collimator angle (average, minimum, maximum) |
| Second hand smoke | Couch lateral (maximum) |
| Obesity | Couch longitudinal (maximum) |
| Alcohol | Couch vertical (maximum) |
| Human papillomavirus (HPV) | Couch angle (average, minimum, maximum) |
| Immunosuppression | Field size (average, minimum, maximum) |
| Immunodeficiency | Isocenter (*X*, *Y*, *Z*) |
| Neutropenia | Maximum beam energy |
| Anemia | Dose per fraction |
| Human immunodeficiency virus (HIV) | |
| Hepatitis | |
| Heart disease | |
| Diabetes | |
| Hypertension | |
| Hyperlipidemia | |
| Failure to thrive | |

*Abbreviations: LASSO, least absolute shrinkage and selection operator; NN, neural network; IMRT, intensity-modulated radiation therapy.*

predictive model on an independent testing set. The breakdown of plans with and without treatment interruptions in the selected folds are shown in **Table 3**. If the same fold was selected for the LASSO and NN models, the split was the same for training and validation sets. If a different fold was selected, the number of positive samples may be slightly different, owing to randomization during the splitting process. For the testing set, the split was the same for LASSO and NN because this was a hold-out set and not involved in the training process.

## LASSO Performance

The performance metrics for the LASSO model are shown in **Figure 1** and Supplemental Table A (**www.appliedradiationoncology.com**). For the training set, the average sensitivity, specificity, and accuracy ranged from 69% to 92%, 52% to 67%, and 67% to 79%, respectively. Sensitivity was significantly higher than specificity for the full cohort, the spine/extremity subset, and the brain subset ($P<.05$). For the validation set, the average sensitivity, specificity, and accuracy ranged from 42% to 82%, 49% to 64%, and 51% to 64%, respectively. Sensitivity remained significantly higher than specificity only for the full cohort ($P<.05$).

For the independent testing set, the sensitivity, specificity, and accuracy ranged from 37% to 70%, 59% to 78%, and 60% to 76%, respectively. Except for the full cohort, specificity was higher than sensitivity. The brain subset had the highest accuracy (76.1%). Overall, the higher specificity than sensitivity on the independent testing set indicated that the model was better able to predict true negatives (those without treatment interruptions) than true positives (those with treatment interruptions).

For the LASSO model, **Table 4** shows the features selected by the algorithm for the prediction, which included the components of clinical features and plan parameters.

## Neural Network Performance

The performance metrics for the NN model are shown in **Figure 2** and Supplemental Table B (**www.appliedradiationoncology.com**). For the training set, the average sensitivity, specificity, and accuracy ranged from 79% to 99%, 34% to 68%, and 59% to 83%, respectively. Sensitivity was significantly higher than specificity ($P<.05$) for all subsets except the pelvis/prostate subset ($P=.15$). For the validation set, average sensitivity, specificity, and accuracy ranged from 53% to 77%, 35% to 68%, and 38% to 67%, respectively. We observed a similar trend in the validation set as in the training set; sensitivity in the validation set was generally higher than specificity,


https://doi.org/10.37549/ARO-D-24-00025

**Table 3. Distribution of Plans With and Without Treatment Interruptions Across Separate Subset Groups from the Selected Cross-Validation Fold With the Highest Validation Set Sensitivity. Numbers Shown Indicate the Number of Plans With Treatment Interruptions Divided by the Total Number of Plans. "All" Indicates the Full Cohort Before Separated into Subsets Based on Treatment Site**

| | TRAINING SET | | VALIDATION SET | | TESTING SET | |
|---|---|---|---|---|---|---|
| | LASSO | NN | LASSO | NN | LASSO | NN |
| All | 315/3999 (7.9%) | 297/3999 (7.4%) | 76/1001 (7.6%) | 94/1001 (9.4%) | 188/2142 (8.8%) | 188/2142 (8.8%) |
| Lung, breast, and chest | 129/1864 (6.9%) | 141/1864 (7.6%) | 42/467 (9.0%) | 30/467 (6.4%) | 57/998 (5.7%) | 57/998 (5.7%) |
| Pelvis and prostate | 50/603 (8.3%) | 51/604 (8.4%) | 13/152 (8.6%) | 12/151 (7.9%) | 27/322 (8.4%) | 27/322 (8.4%) |
| Spine and extremity | 56/550 (10.2%) | 56/550 (10.2%) | 11/139 (7.9%) | 11/139 (7.9%) | 25/294 (8.5%) | 25/294 (8.5%) |
| Brain | 32/719 (4.5%) | 32/719 (4.5%) | 13/181 (7.2%) | 13/181 (7.2%) | 36/385 (9.4%) | 36/385 (9.4%) |
| Head and neck | 47/262 (17.9%) | 50/262 (19.1%) | 14/67 (20.9%) | 11/67 (16.4%) | 27/139 (19.4%) | 27/139 (19.4%) |

*Abbreviations: LASSO, least absolute shrinkage and selection operator; NN, neural network*

although it was only statistically significant in the spine/extremity subset ($P<.05$). For the testing set, the sensitivity, specificity, and accuracy ranged from 41% to 68%, 53% to 79%, and 53% to 78%, respectively. Overall, accuracy was highest in the brain subset (78%).

## DISCUSSION

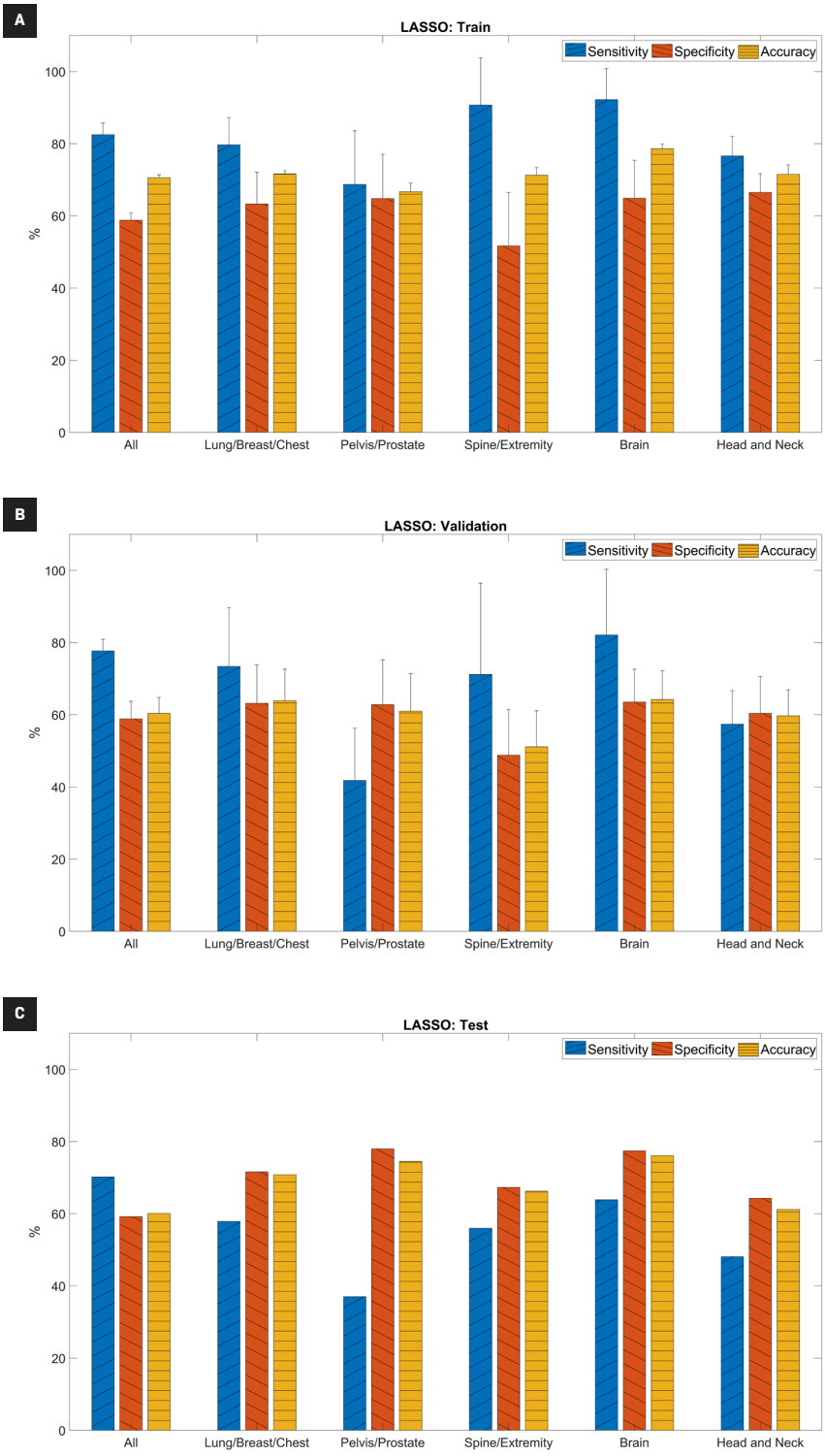
These study results provide proof of concept that AI can be a reliable screening tool in the peer review process to help identify cases early on that may cause treatment interruption or major changes in treatment course. We found that logistic regression and NN-based models had some predictive power in recognizing cases that would experience treatment interruptions. This could be useful in identifying cases that will require a replan for various reasons, such as a patient who experiences toxicity or has a dramatic change in target volume. Cases identified as "high risk" for interruptions could be highlighted in peer review for a more in-depth evaluation.

Our study demonstrates the potential of AI in radiation oncology peer review to prospectively identify treatment interruption. A simple example of one potential software display in a peer-review setting is shown in **Figure 3**. Each patient could be assigned a "score" associated with treatment interruption risk that could help prioritize cases for discussion. The software might also display the clinical or plan features flagged for that specific case. Links directly to the plan and other relevant clinical documents could facilitate quick reference during discussions.

In radiation oncology, every patient receives a personalized plan chosen by their physician that best fits their characteristics, such as their diagnosis and performance status. However, for reasons that are sometimes unknown, a patient may require a change or break in treatment, which can be detrimental to their oncologic outcome. In patients with head-and-neck cancer, the hazard rate of death increased 4.2% for each additional day needed to finish radiation therapy beyond 8 weeks.[13] Even small disruptions in radiation therapy can have negative consequences in gynecologic patients. Lanciano et al reported a 7.7% reduction in 4-year survival when the radiation therapy course was >10 weeks compared with 8.0-9.9 weeks.[7]

The outcome analyzed in our study was based on whether or not a plan had remaining fractions. Although remaining fractions are not always indicative of treatment toxicity, they could be a major contributor. Several studies have looked at the ability of ML models to predict toxicity using clinical and dose factors and/or radiomic features.[14,15] Reddy et al used random forest, gradient-boosted decision tree, and logistic regression models with input of clinical and treatment variables to predict breast cancer treatment toxicity and achieved an area under the curve (AUC) ranging from 0.56 to 0.85.[16] Das et al created a fusion of different non-linear multivariate models (decision trees, NNs, support vector machines, and self-organizing maps) with input of dose and non-dose patient variables to predict

**Figure 1.** Predictive performance of the least absolute shrinkage and selection operator (LASSO) model for (A) training, (B) validation, and (C) testing datasets. For the training and validation datasets, it is the average across five-fold cross-validation. "All" indicates the full cohort before separation into subsets based on treatment site and error bars indicate standard deviation.





https://doi.org/10.37549/ARO-D-24-00025

**Table 4. Features Selected by Logistic Regression Least Absolute Shrinkage and Selection Operator (LASSO) Model for Each Subset Group. "All" Indicates the Full Cohort Before Separation into Subsets Based on Treatment Site. Isocenter Location *X*, *Y*, and *Z* Refer to Lateral, Anterior/Posterior, and Superior/Inferior Directions, Respectively**

| SITE | CLINICAL FEATURES | PLAN FEATURES |
|---|---|---|
| All | • Sex<br>• Inpatient status<br>• Personal history of cancer | • Monitor units (max)<br>• Source to skin distance (max)<br>• Gating<br>• Collimator rotation (max)<br>• Field size (average)<br>• Field size (max)<br>• Dose per fraction |
| Lung, breast and chest | • Inpatient status<br>• Personal history of cancer | • Type of plan<br>• Monitor units (average)<br>• Source to skin distance (average)<br>• Source to skin distance (max)<br>• Bolus<br>• Gating<br>• Field size (min)<br>• Field size (max) |
| Pelvis and prostate | • Anemia<br>• Diabetes | • Type of plan<br>• Couch longitudinal (max)<br>• Field size (max)<br>• Dose per fraction |
| Spine and extremity | • Patient age group<br>• Inpatient status<br>• Tobacco | • Monitor units (total)<br>• Source to skin distance (min)<br>• Couch longitudinal (max)<br>• Field size (max)<br>• Isocenter (*X*)<br>• Energy (max)<br>• Dose per fraction |
| Brain | • Inpatient status<br>• Immunosuppression<br>• Anemia | • Monitor units (average)<br>• Tolerance table<br>• Gantry angle (max)<br>• Bolus<br>• Couch vertical (max)<br>• Isocenter (*Z*)<br>• Dose per fraction |
| Head and neck | • Patient age group<br>• Inpatient status<br>• Tobacco<br>• Obesity<br>• Hypertension | • Monitor units (average)<br>• Monitor units (min)<br>• Tolerance table<br>• Field size (average) |

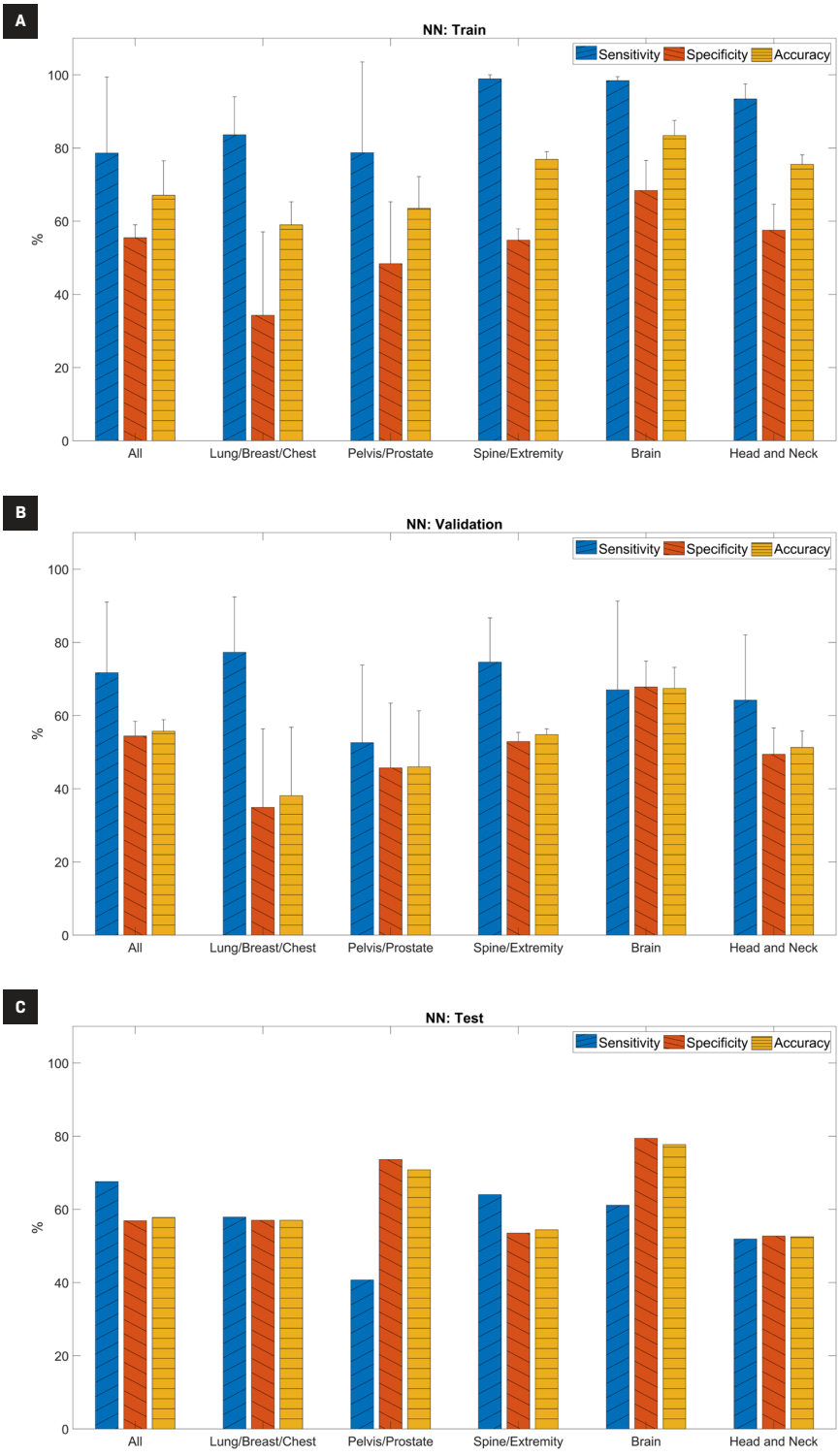
radiation-induced pneumonitis in lung cancer patients with an AUC of 0.79.[17]

The test set accuracy in our study ranged from 57% to 71% for the lung/breast/chest subset. We grouped lung and breast into the same subset due to data size limitations; thus, the predictive accuracy of our models may be improved by further separation.

Similar to their study on breast cancer patients, Reddy et al used random forest, gradient-boosted decision tree, and logistic regression models with clinical and treatment parameters to predict head-and-neck cancer treatment toxicity and achieved an AUC of 0.64-0.76 in their validation set.[18] Jiang et al used three supervised learning methods (ridge logistic regression, lasso logistic regression, and random forest) to predict xerostomia, resulting in an AUC of 0.7.[19]

Our head-and-neck subset came in lower at 53%-61% test set accuracy. We did not directly analyze toxicity because we predicted whether there were remaining fractions in each plan. Remaining fractions could also be caused by re-planning owing to tumor shrinkage; this may explain the reduced predictive value of our study compared to others looking solely at toxicity. Carrara et al applied the artificial NN approach with five input variables relating to patient dose, history, and therapy to predict toxicity after high-dose prostate cancer radiation therapy, achieving an AUC of 0.78.[20] Pella et al used support vector machines and neural network-based algorithms to predict acute toxicity of the bladder and rectum due to prostate irradiation, resulting in overall accuracy similar in both models at an AUC of 0.7.[21] The accuracy of our testing set in the prostate/pelvis subset, at 71%-75%, is similar.

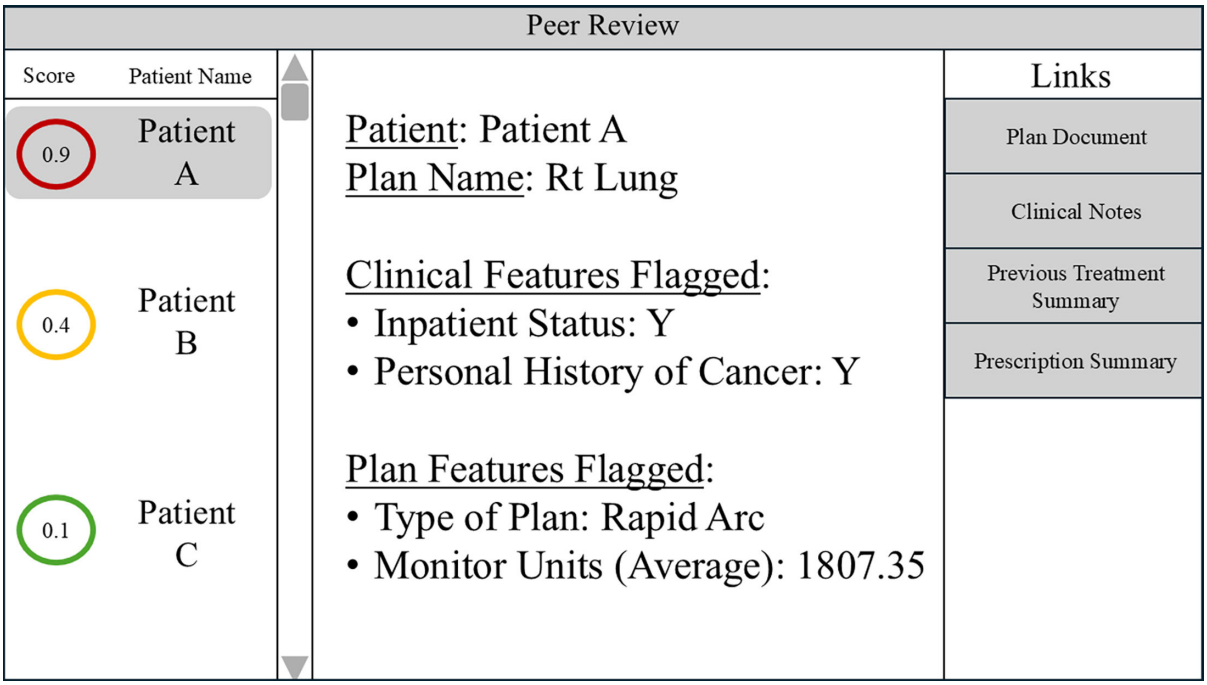
**Figure 2.** Predictive performance of the neural network (NN) model for (A) training set, (B) validation set, and (C) testing set. For the training and validation sets, it is the average across five-fold cross-validation. "All" indicates the full cohort before separation into subsets based on treatment site and error bars indicate standard deviation.





https://doi.org/10.37549/ARO-D-24-00025

**Figure 3.** An example of an application interface that can be displayed during peer review. The application incorporates the risk of treatment interruption score, highlighting features flagged for review and quick access links to patient data.




Although toxicity is one cause of treatment interruption, physician input/suggestions can also cause interruptions if the case is not presented before treatment and/or some information is not assessed during peer review. The traditional approach to peer review is based on "chart rounds," where the physicians, physicists, dosimetrists, and therapists review details of each case (eg, clinical history, treatment technique, prescription dose, treatment plan, and patient setup). The average amount of time spent on each patient during peer review was reported to range between 1 and 4 minutes.[2,22]

Therefore, reviewing the more technical aspects of treatment delivery (eg, monitor units and couch/collimator/gantry parameters), which may provide additional information, may not be feasible owing to time constraints. As demonstrated by our study, subtle, complex relationships within/between clinical data and plan parameters may influence successful treatment delivery. AI and ML have the potential to identify and analyze these relationships for peer review.[23] Similar to the goals of many AI-driven studies, ours is not to replace humans with AI but instead to offer providers additional tools to enhance the process. These can supplement clinical intuition and deepen our understanding of factors that previously might go unanticipated.

Since time is a limited resource, AI/ML can also help to identify and prioritize cases that require more time for discussion. Ultimately, these technologies could improve patient safety and treatment outcomes.

Our study offers a concept that can be used to better identify charts requiring more in-depth review. Many software tools for peer review or chart checking take "rule-based" approaches, meaning that the parameter being flagged will need to be predefined with a range or value.[24] Azmandian et al used clustering techniques for outlier detection based on treatment parameters in four-field box prostate plans; this study helped to detect plan abnormalities without using the rule-based approach.[25] Kalet et al developed a Bayesian network model using clinical and plan parameters to detect errors in radiation therapy plans. Their model utilized a clinical layer (eg, morphology or tumor type), a prescription layer (eg, total dose, dose per fraction, technique), and a treatment layer (eg, monitor unit per fraction, number of beams, beam energy). Their study achieved an AUC of 0.88, 0.98, and 0.89 for the lung, brain, and breast cancer error detection networks, respectively.[26] Similar to our study, Kalet et al found the highest testing set accuracy in the brain subset.

Luk et al also used a Bayesian model incorporating prescription, plan, setup, and diagnostic parameters to detect chart review

errors. The AUC for this study ranged from 0.82 to 0.89.[27] Our testing set accuracy is lower than these other studies, possibly because our cohort included not only "abnormal" cases from a chart-checking perspective but also cases with toxicity-related interruptions.

Although the sensitivity and specificity of our models are relatively low, there are multiple ways in which we can improve these metrics of our study. First, while our study specified remaining fractions as a surrogate for cases likely to experience treatment interruption or complications, remaining fractions can result from reasons unrelated to treatment. Our definition of treatment interruptions as remaining fractions includes patients whose treatments were re-planned and those who discontinued treatment. Separating these cases into two cohorts could improve model predictability.

Second, we grouped subsets based on treatment site, but further dividing them into more focused groups may improve model predictability, eg, separating pelvic plans based on whether they include pelvic lymph nodes or separating spine plans based on vertebral level. Overall, model performance and rigor are expected to improve with increased curation of the dataset and additional clinical factors, including dosimetric plan and structure set data. Additionally, introducing socioeconomic and personal factors, which are commonly seen as causes of treatment interruptions, could improve our models.

Future study directions can also include optimizing hyperparameters and layer structures for the NN. Alternative techniques for data re-balancing or using convolutional NNs for deep learning could also be investigated. A weakness of this study is that it does not include brachytherapy, electron therapy, other treatment sites, or pediatric populations. To address this issue, we are continuing the study with brachytherapy, electrons, and pediatric populations, as well as including additional, more robust parameters.

A known limitation of NNs is their difficulty in identifying the single feature that contributes most to the model, given the complexity of their relationships and the numerous weights/biases assigned to them. Future studies can explore a technique to uncover those features selected by the NN deemed to be most important to the predictive task. At the current stage of our model, which demonstrates moderate predictive performance, making conclusions about or connecting a specific result and/or feature to a reason for a predicted interruption is difficult. Improvements in model performance should enable exploration of more advanced, explainable-AI techniques.

Another current focus of AI and ML is on treatment planning.[23] For example, AI is being used to adapt treatment plans in real time to match the day-to-day variations in patient anatomy, thereby reducing interruptions caused by re-simulation and/or discussion manual re-planning.[28] As the technology for adaptive planning becomes more widely available, our model will likely need additional training to keep up with technological advancements. Future studies may focus on sites where adaptive planning is often beneficial or necessary, such as the head and neck owing to tumor progression or treatment response. The ability to predict cases requiring adaptive planning due to tumor change can allow the care team to anticipate and minimize treatment breaks.

## CONCLUSIONS

Our study demonstrated the ability of AI and ML models to predict major changes in patient treatment, including re-planning and radiation therapy cessation. The findings point to the promising capability of AI and ML to augment peer review and encourage further studies in this aspect of radiation oncology.

## References

1) Marks LB, Adams RD, Pawlicki T, et al. Enhancing the role of case-oriented peer review to improve quality and safety in radiation oncology: executive summary. *Pract Radiat Oncol.* 2013;3(3):149-156. doi:10.1016/j.prro.2012.11.010

2) Martin-Garcia E, Celada-Álvarez F, Pérez-Calatayud MJ, et al. 100% peer review in radiation oncology: is it feasible? *Clin Transl Oncol.* 2020;22(12):2341-2349. doi:10.1007/s12094-020-02394-8

3) Hoopes DJ, Johnstone PA, Chapin PS, et al. Practice patterns for peer review in radiation oncology. *Pract Radiat Oncol.* 2015;5(1):32-38. doi:10.1016/j.prro.2014.04.004

4) Duggar WN, Bhandari R, Yang CC, Vijayakumar S. Group consensus peer review in radiation oncology: commitment to quality. *Radiat Oncol.* 2018;13(1):55. doi:10.1186/s13014-018-1006-1

5) Shaikh T, Handorf EA, Murphy CT, et al. The impact of radiation treatment time on survival in patients with head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2016;96(5):967-975. doi:10.1016/j.ijrobp.2016.08.046

6) Dahlke S, Steinmann D, Christiansen H, et al. Impact of time factors on outcome in patients with head and neck cancer treated with definitive radio(chemo)therapy. *In Vivo.* 2017;31(5):949-955. doi:10.21873/invivo.11152

7) Lanciano RM, Pajak TF, Martz K, Hanks GE. The influence of treatment time on outcome for squamous cell cancer of the uterine cervix treated with radiation: a patterns-of-care study. *Int J Radiat Oncol.* 1993;25(3):391-397. doi:10.1016/0360-3016(93)90058-4

8) Thaker NG, Sturdevant L, Jhingran A, et al. Assessing the quality of a radiation oncology case-based, peer-review program in an integrated academic and community cancer center network. *J Oncol Pract*. 2016;12(4):e476-86. doi:10.1200/JOP.2015.005983

9) Vijayakumar S, Duggar WN, Packianathan S, Morris B, Yang CC. Chasing zero harm in radiation oncology: using pre-treatment peer review. *Front Oncol*. 2019;9:302. doi:10.3389/fonc.2019.00302

10) Caissie A, Rouette J, Jugpal P, et al. A pan-canadian survey of peer review practices in radiation oncology. *Pract Radiat Oncol*. 2016;6(5):342-351. doi:10.1016/j.prro.2016.01.014

11) Hong JC, Eclov NCW, Dalal NH, et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): a prospective randomized study of machine learning-directed clinical evaluations during radiation and chemoradiation. *J Clin Oncol*. 2020;38(31):3652-3661. doi:10.1200/JCO.20.01688

12) He HB, Bai Y, Garcia EA, Li ST. *ADASYN: adaptive synthetic sampling approach for imbalanced learning*. IEEE IJCNN; 2008:1322-1328. doi:10.1109/IJCNN.2008.4633969

13) Sher DJ, Posner MR, Tishler RB, et al. Relationship between radiation treatment time and overall survival after induction chemotherapy for locally advanced head-and-neck carcinoma: a subset analysis of tax 324. *Int J Radiat Oncol*. 2011;81(5):e813-e818. doi:10.1016/j.ijrobp.2010.12.005

14) Isaksson LJ, Pepa M, Zaffaroni M, et al. Machine learning-based models for prediction of toxicity outcomes in radiotherapy. *Front Oncol*. 2020;10:790. doi:10.3389/fonc.2020.00790

15) Zhang B, Shi H, Wang H. Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. *J Multidiscip Healthc*. 2023;16:1779-1791. doi:10.2147/JMDH.S410301

16) Reddy J, Lindsay WD, Berlind CG, Ahern CA, Smith BD. Applying a machine learning approach to predict acute toxicities during radiation for breast cancer patients. *Int J Radiat Oncol*. 2018;102(3):S59. doi:10.1016/j.ijrobp.2018.06.167

17) Das SK, Chen SF, Deasy JO, et al. Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction. *Med Phys*. 2008;35(11):5098-5109. doi:10.1118/1.2996012

18) Reddy JP, Lindsay WD, Berlind CG, et al. Applying a machine learning approach to predict acute radiation toxicities for head and neck cancer patients (vol 105, pg S69, 2019). *Int J Radiat Oncol Jan*. 106(1):223-223. doi:10.1016/j.ijrobp.2019.10.013

19) Jiang W, Lakshminarayanan P, Hui X, et al. Machine learning methods uncover radiomorphologic dose patterns in salivary glands that predict xerostomia in patients with head and neck cancer. *Adv Radiat Oncol*. 2019;4(2):401-412. doi:10.1016/j.adro.2018.11.008

20) Carrara M, Massari E, Cicchetti A, et al. Development of a ready-to-use graphical tool based on artificial neural network classification: application for the prediction of late fecal incontinence after prostate cancer radiation therapy. *Int J Radiat Oncol*. 2018;102(5):1533-1542. doi:10.1016/j.ijrobp.2018.07.2014

21) Pella A, Cambria R, Riboldi M, et al. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. *Med Phys*. 2011;38(6):2859-2867. doi:10.1118/1.3582947

22) Lawrence YR, Whiton MA, Symon Z, et al. Quality assurance peer review chart rounds in 2011: a survey of academic institutions in the united states. *Int J Radiat Oncol Biol Phys*. 2012;84(3):590-595. doi:10.1016/j.ijrobp.2012.01.029

23) Kiser KJ, Fuller CD, Reed VK. Artificial intelligence in radiation oncology treatment planning: a brief overview. *J Med Artif Intell*. 2019;2:9-9. doi:10.21037/jmai.2019.04.02

24) Luk SMH, Ford EC, Phillips MH, Kalet AM. Improving the quality of care in radiation oncology using artificial intelligence. *Clin Oncol*. 2022;34(2):89-98. doi:10.1016/j.clon.2021.11.011

25) Azmandian F, Kaeli D, Dy JG, et al. Towards the development of an error checker for radiotherapy treatment plans: a preliminary study. *Phys Med Biol*. 2007;52(21):6511-6524. doi:10.1088/0031-9155/52/21/012

26) Kalet AM, Gennari JH, Ford EC, Phillips MH. Bayesian network models for error detection in radiotherapy plans. *Phys Med Biol*. 2015;60(7):2735-2749. doi:10.1088/0031-9155/60/7/2735

27) Luk SMH, Meyer J, Young LA, et al. Characterization of a bayesian network-based radiotherapy plan verification model. *Med Phys*. 2019;46(5):2006-2014. doi:10.1002/mp.13515

28) Wu QJ, Li T, Wu Q, Yin FF. Adaptive radiation therapy: technical components and clinical applications. *Cancer J*. 2011;17(3):182-189. doi:10.1097/PPO.0b013e31821da9d8

# A Head and Neck Contour Grading System Provides an Objective Assessment of Radiation Oncology Resident Contouring Skills

Astha Rohit, BS; Diego A.S. Toesca, MD; Justin D. Gagneur, MA; Samir H. Patel, MD; Jean-Claude M. Rwigema, MD; Lisa McGee, MD*

## ABSTRACT

**Purpose:** This study evaluated the effectiveness of a peer-review process to objectively assess the skills of radiation oncology (RO) resident physicians (RPs) in contour grading of head and neck (HN) malignancies.

**Methods:** Target volumes from consecutive patients diagnosed with primary HN malignancies, treated in a single institution, were contoured by RPs during HN service rotations and were formally peer-reviewed by a minimum of 2 HN RO attendings and assigned a grade as follows: R0 (no change recommended); R1 (minor revision recommended, not clinically significant); and R2 (major revision recommended, deemed clinically significant). Progression of residents' HN contouring skills was assessed in accordance with their postgraduate year (PGY) in training.

**Results:** Formal contour peer review was performed for 218 patients with HN cancer contoured by 6 RO RPs from 2018 to 2024. Of those cases, 48 (22%) were contoured by PGY2 RPs, 98 (45%) by PGY3 RPs, 40 (18%) by PGY4 RPs, and 32 (15%) by PGY5 RPs. There was an objective improvement in contour grades and a reduced need for target volume modifications through the progression of academic years, with a mean score of 1.43 (SD = 0.71; CI = 0.2) for PGY2 trainees; 0.99 (SD = 0.81; CI = 0.16) for PGY3 trainees; 0.93 (SD = 0.92; CI = 0.29) for PGY4 trainees; and 0.69 (SD = 0.64; CI = 0.23) for PGY5 trainees. Improvement in scores was consistent among all RO RPs, with absolute mean improvements of –0.2 (RP #1), –0.32 (RP #2), –0.82 (RP #3), –0.4 (RP #4), –1.33 (RP #5), and –0.56 (RP#6).

**Conclusions:** Incorporating a formal HN contouring peer-review process and contour grade assignment into routine clinical evaluation of RO RPs provides an objective metric of their HN contour quality progression throughout training, by PGY. This tool can be used as an added, objective assessment of RO resident competency in contour evaluation.

**Keywords:** contouring peer review, target volume, target delineation, resident physician education, assessment, objective

## Introduction

The existing framework for monitoring the skill progression of radiation oncology (RO) resident physicians (RPs) throughout their postgraduate training has historically relied on subjective measures. Most institutions in the United States use the Accreditation Council for Graduate Medical Education (ACGME) milestones to evaluate RPs throughout their postgraduate training to ensure necessary competencies are met prior to graduation.[1,2] The ACGME Milestone 2.0 version outlines 6 key competencies:

**Affiliations:** Department of Radiation Oncology, Mayo Clinic, Phoenix, AZ.
**Corresponding author:** *Lisa McGee, MD, 5881 E Mayo Blvd, Phoenix, AZ 85054. (McGee.Lisa@mayo.edu)
**Disclosures:** Dr. Samir Patel has received a grant from Galera Therapeutics. All other authors declare that they have no conflicts of interests related to this publication.
**Data sharing statement:** All data created and analyzed in this series are included in this publication.

https://doi.org/10.37549/ARO-D-25-0010

Patient Care; Medical Knowledge, Practice-Based Learning and Improvement; Interpersonal and Communication Skills; Professionalism; and System-Based Practice. Within these competencies, there are subcompetencies, such as contouring and target delineation skill, that are graded on a scale from 1 to 5.[2] Every 6 months, programs' Clinical Competency Committees evaluate the RPs and determine their appropriate skill levels in all pertinent categories with the expectation that all RPs will reach level 4, at the minimum, in all sectors prior to their graduation.[2]

Due to the inherent subjectivity of the goals and the diversity of the field, the assessment of RPs' proficiency is largely left to the discretion of the evaluator.[1] Currently, the most overt objective measure appears to be the number of treatment cases performed or observed by the RP, with a minimum mandatory case attendance per disease site.[3] Despite recent advancements, there remains an imbalance of subjective and objective metrics, obscuring standardization and comparison among programs. Certain aspects of RO training are disproportionately affected, such as simulations, contouring, planning, treatment setup, and procedure proficiency, which are heavily skill-based and crucial to effective treatment.[1] These variations in training and monitoring practices make it difficult to gauge an RP's readiness for independent practice on a national scale.

In radiation therapy planning, the process of contouring involves delineating the target volume as well as avoidance structures and organs at risk, in order to maximize local tumor control and minimize treatment toxicity.[4] Contouring is a key skill for RPs to master as appropriate target volume delineation largely dictates the quality of a radiation therapy treatment plan and, ultimately, the oncological outcome for patients. Suboptimal contouring skills could therefore result in tumor recurrences or undesired side effects with the potential to significantly impact patient quality of life.[5-10]

Cancer of the head and neck (HN) is historically difficult to treat, exemplified by one of the highest inter-physician variabilities in contouring for the disease site.[5,6] A multitude of factors—including a complex local anatomy, high number of relatively small and sensitive anatomical structures in close proximity to one another, and anatomical variations between patients—can skew the standardization of contours in patients with HN cancer and contribute to one of the highest observed toxicities among cancer disease subsites commonly treated in RO.[11,12]

In an effort to improve patient outcomes by reducing interobserver variability, many institutions have implemented a peer-review process to optimize plan quality and patient safety.[9,13,14] Multiple professional organizations, including the American Society for Radiation Oncology (ASTRO), American College of Radiology, and Royal Australian and New Zealand College of Radiologists have recommended the use of a multiphysician team to review radiation therapy (RT) plans, provide feedback to treating physicians, and cultivate a safe environment that encourages performance improvement in areas such as contours and treatment planning.[9,13,15,16] These processes, and the aspects of RT planning that they evaluate, vary widely across institutions. ASTRO emphasizes the impact of contouring on patient outcomes, assigning the highest priority to target volume delineation in the peer-review process.[13] Repeatedly, institutional studies focused on contouring in HN cancers have shown that modifications resulting from peer review directly impact patient prognosis by reducing RT-induced toxicities and improving survival rates.[7,9,13,14,17]

The peer-review process fosters a collaborative, educational environment in which learning occurs through both participation and observation. It allows RPs to actively participate in treatment plan evaluations by asking questions and offering or receiving feedback, all of which can improve their confidence and skills while enhancing their understanding and implementation of institutional guidelines.[9,14]

In this study, we report the outcomes of a prospective, formal, HN contour grading process that was developed during peer-review sessions at our institution. The protocol was used to objectively assess competency improvement in target volume delineation by RO RPs throughout their postgraduate training.

## Materials and Methods

### Study Design and Procedures

Following Institutional Review Board approval, a protocol was implemented to assess the quality of contouring completed by RO RPs for patients diagnosed with primary HN cancers. From 2018 to 2024, RO RPs on HN service used auto-contouring technology to establish target volumes for patients requiring RT treatment. HN contours were completed on a treatment planning CT scan, aided by the incorporation of information from diagnostic imaging (PET and/or MRI) fused to the treatment planning CT, as well as information from clinical examination (e.g., flexible laryngoscopy exam, operative notes and surgical pathology when appropriate). Unless clinically contraindicated, contrast was used for CT simulations.

Once their contours were completed, the RO RP cases were formally presented for peer review to a minimum of 2 RO HN attending physicians (APs) who provided appraisal consensus, grading, and feedback. Sessions were held weekly following HN tumor board review, with additional ad hoc sessions scheduled as needed. Target contours were reviewed by all available HN APs, with mandatory review by the treating HN oncologist and a minimum of one additional HN AP. During the peer-review process, contour edit recommendations and feedback were provided verbally to the RO RPs for each individual case. Initial contours were assigned a grade as follows: R0 (no change recommended); R1 (minor revision recommended, not clinically significant);

and R2 (major revision recommended, deemed clinically significant). An R1 grade reflected the need for stylistic changes to match the contours of the AP, rather than a decrement in RP skill set, while an R2 grade indicated shortcomings in the target contours that could negatively impact patient outcomes, such as the omitting gross residual disease or inaccurate coverage of the postoperative tumor bed. The grades were recorded in a peer-review task area of the patient's electronic medical record ARIA (Varian Medical Systems, Inc., Palo Alto, CA), alongside the initials of the resident and all peer-review physicians. After the recommended contouring edits were completed by the RO RP and re-reviewed by the AP responsible for the case, contours were sent to dosimetry to initiate treatment planning.

Because the contour grade assigned to each case was determined by a collective consensus among HN RO APs, inter-reviewer variability was not assessed.

### Statistical Analysis

The progression of HN target contouring skills among RO RPs was assessed throughout their postgraduate year (PGY) training. Mean contour grades were calculated for each RP and compared throughout PGY progression and residency, with the expectation of improved scores by the end of training. Mean contour grades were also calculated by PGY cohort, with a similar expectation of improvement in the later years of training. Mean contour grades were compared across PGY levels to evaluate correlation. Confidence intervals (CIs) were calculated using the Student *t* test. Statistical analyses were performed using Prism version 10 (GraphPad Software, Boston, MA).

## Results

Over the course of this study, 218 HN cancer patient targets were contoured by 6 RO RPs and then formally peer reviewed. Among the patient population, 26% were females and 74% were males, with a median age of 67 years. The 5 most common HN tumor sites were oropharynx (35%), cutaneous (18%), oral cavity (12%), salivary glands (8%), and larynx/hypopharynx (6%). Patient characteristics are detailed in **Table 1**.

Of the 218 cases included, 22% (48) were contoured by PGY2 RPs, 45% (98) by PGY3 RPs, 18% (40) by PGY4 RPs, and 15% (32) by PGY5 RPs. An objective improvement in contour grades was observed across advancing training years, with lower scores (trending towards zero) indicating less need for target volume edits or modifications. The mean contour grades for RPs (**Figure 1**) were, for PGY2s, 1.43 (SD = 0.71; CI = 0.2); for PGY3s, 0.99 (SD = 0.81; CI = 0.16); for PGY4s, 0.93 (SD = 0.92; CI = 0.29); and for PGY5s, 0.69 (SD = 0.64; CI = 0.23).

Subsequently, the study assessed the mean contour grade for each RP throughout training (**Figure 2**): resident 1 (1.05, PGY3; 0.5, PGY4; 0.88, PGY5); resident 2 (1.07, PGY3; 1.13, PGY4; 0.75, PGY5); resident 3, (1.38, PGY2; 0.47, PGY3; 0.67, PGY4; 0.55, PGY5); resident 4 (1.30, PGY2; 1.25, PGY3; 1.14, PGY4; 0.70, PGY5); resident 5 (2.00, PGY2; 1.43, PGY3; 0.67, PGY4); and resident 6 (1.06, PGY2; 0.5, PGY3 [currently in training]). Overall, we observed a consistent improvement in contour grades for each RP, with an absolute mean improvement of –0.2 for resident 1, –0.32 for resident 2, –0.82 for resident 3, –0.4 for resident 4, –1.33 for resident 5, and –0.56 for resident 6 across years of training (**Figure 2**).

**Table 1. Patient Characteristics**

| CHARACTERISTICS | N (%) OR MEDIAN (RANGE) |
|---|---|
| Gender | |
| Female | 50 (26%) |
| Male | 162 (74) |
| Age | 67 (29-92) |
| Primary site | |
| Oropharynx | 76 (35) |
| Cutaneous | 40 (18) |
| Oral cavity | 25 (12) |
| Salivary glands | 17 (8) |
| Larynx/hypopharynx | 14 (6) |
| Other | 46 (21) |
| Intent | |
| Curative | 209 (96) |
| Palliative | 9 (4) |
| T stage | |
| T0 | 20 (9) |
| T1 | 30 (14) |
| T2 | 33 (15) |
| T3 | 49 (23) |
| T4 | 73 (33) |
| NA | 13 (6) |
| N stage | |
| N0 | 73 (33) |
| N+ | 134 (62) |
| NA | 11 (5) |
| M stage | |
| M0 | 202 (93) |
| M1 | 5 (2) |
| NA | 11 (5) |
| Radiation Therapy (RT) modality | |
| Proton | 122 (56) |
| Intensity Modulated Radiotherapy (IMRT) | 95 (44) |
| 3 Dimensional Conformal Radiation Therapy (3D-CRT) | 1 (0) |
| RT total dose | 60 (28-74.4) |
| RT n of fractions | 30 (12-37) |

## Discussion

### Practice Brings Improvement

This single-institution, prospective study demonstrates that target contour grading is an effective tool for evaluating the progress of RPs' competency in contouring throughout training. A steady reduction in the frequency of R2 contour grades was observed throughout postgraduate training from PGY2 to PGY5, represented by lower mean scores. Meanwhile, an absolute


https://doi.org/10.37549/ARO-D-25-0010

improvement in the frequency of individual R0 grades over residency training was observed based on the consensus of expert peer reviewers. This affirms the expectation that as RO RPs progress through training, their contours that would be considered unacceptable for use in treatment decrease, while the number of cases done accurately, requiring no additional modifications, increases over time. Additionally, by PGY4 and PGY5, the frequency of R0 versus R2 grades among residents was consistent with those observed among faculty, supporting the readiness of the RPs for independent HN contouring by graduation.

For some RPs, we observed a nonlinear progression of mean contour grades throughout their training. This trend might be attributed to several factors, including individual learning curve progression; varying levels of experience due to uneven case distribution over the PGY; heterogeneity in case complexity, which was not controlled for according to the RP's PGY; and the well-documented subjectivity associated with volume delineation in HN cancers. Despite this variation, all RPs demonstrated improvement in contouring skills when comparing the beginning of training to the end. This suggests that the institution's peer-review process, and factors such as feedback, documentation, and accountability, have a significant, positive impact on the RP's development of independent contouring skills. Overall, we demonstrate that contour grading allows evaluators and learners to effectively document progression of skills throughout training.

In our institution's timeline of clinical rotations for RO residents, there is a significant increase in the caseload of HN cancer patients between PGY2 and PGY3. In this study, 22% of the cases were contoured by PGY2s and 45% by PGY3s. The reason for this discrepancy is that RPs in the PGY2 HN rotation see both patients with HN cancer and with breast cancer, while those in the PGY3 HN rotation see primarily HN patients. Concurrent to this caseload increase, we observed the largest difference in mean contour grade between PGY2 and PGY3, which suggests that increased contour practice is a significant driver in contouring skill evolution. By PGY3, individuals were more likely to receive a score of R0 or R1 than R2, a trend that is consistent with similar studies revealing a correlation between a physician's contouring skills and level of expertise[7,14]

**Figure 1.** Contour grading change over the training period. PGY, postgraduate year.

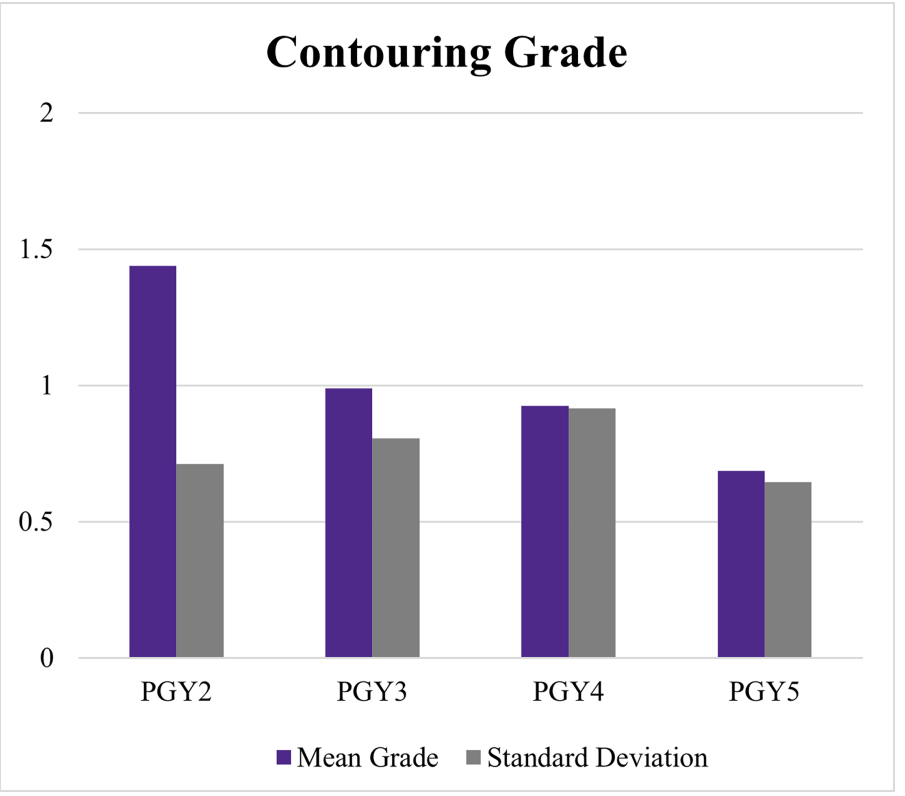
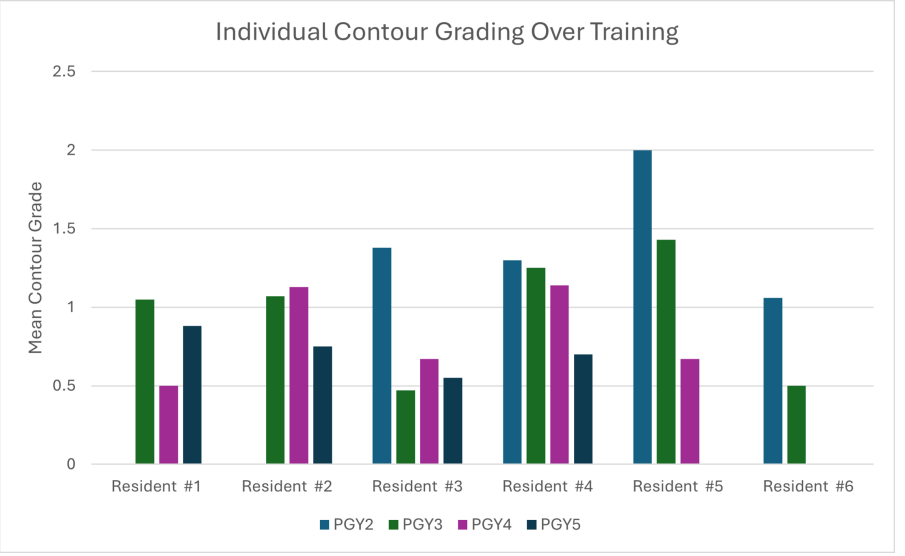



**Figure 2.** Individual contour grading change over the training period. PGY, postgrad uate year.



### Physician Collaboration Improves Treatment Planning

Contour grading, when used in conjunction with peer review, allows for RO RPs to collaborate with more experienced physicians in HN radiation treatment planning (RTP) and evaluation. The process facilitates discussions between APs and RPs that not only address contour improvement skills, but also areas of controversy within RO contouring practice, and different approaches to contouring and treatment based on a patient's anatomy and case specifics. These discussions culminate in consensus among RO APs, which minimizes interobserver variability in practice and decreases the frequency of systemic errors, such as dose delivery or geographic misses, which are known to compromise local control and increase morbidity.[6] This has high educational benefit for RO RPs by presenting more covert viewpoints and emerging considerations, ensuring a well-rounded approach to RTP.

A cross-sectional analysis of recent RO residency graduates in 2016-2017 observed that an increased caseload and independent treatment planning during residency correlated with greater confidence and comfort during independent clinical practice.[18] This is particularly relevant given the findings of a recent needs assessment conducted at 14 ACGME-accredited RO residency programs. In this study, 56% of RPs reported inadequate exposure to RTP, and 54% expressed a lack of confidence in independently evaluating RTP. Additionally, 47% indicated that their education in this area was insufficient, while 97% of all respondents believed that a structured RTP review process could improve RP competency in plan evaluation.[19] The inclusion of contour grading within a similar peer-review framework may serve as a standardized approach to addressing this educational gap, and for better preparing RPs for the transition to independent practice, both in terms of technical skills and mental readiness.

### Opportunities for Objective Performance Measurement

To our knowledge, this is the first prospective study to report the utility of target contour grading as a longitudinal, objective assessment of contouring skill progression in RO RPs. It was designed to address the current lack of objective metrics within the national ACGME RO RP evaluation framework.

ACGME has been transparent with its aspirations to follow the model of graduate education by moving toward a competency-based system of evaluation for residency programs. Its Milestones 1.0, which outlines 6 key competencies and additional, disease-site-specific subcompetencies, has been widely criticized for being difficult to implement consistently, ambiguous differentiation between levels of progression, and prioritization of competencies over key clinical skills. Milestones 2.0, the revised framework released in July 2022,[2] addresses some of these criticisms with the inclusion of an implementation guide and primary goals that focus on clinical skill presentation at each level. While improved, the framework fails to provide objective metrics for key skills such as target volume delineation, making standardization and nationwide comparison of RO residency programs challenging.[1,3] The primary goals remain largely subjective, and the accompanying implementation guides are rarely referenced in RO due to the diversity of the field its cases.

Under the standard process, faculty members evaluate the performance of RPs every 6 months. Without a national, standardized method for evaluating individual cases, this process often results in a generalized, subjective assessment of an RP's abilities rather than a clear, objective measure of skill improvement.

Contour grading offers a solution to this gap by enabling case-by-case scoring that allows institutions to objectively quantify an RP's progression over time. Incorporating objective contour grading into the standardized ACGME RO RP assessment would help ensure that residents demonstrate measurable proficiency in contouring prior to graduation, thereby preparing them for independent clinical practice.

### Study Limitations

There are some limitations to our study. It was restricted to a single RO department, with a small sample size of 6 RPs, which could potentially limit the applicability of these findings to all RO departments. Additionally, due to the timeframe restriction of reporting, the data do not take into account all of the years of training for each of the 6 included RPs. Longer follow-up would have ensured that more residents who had completed all 4 years of training (PGY2-PGY5) would have been included. Another limitation is the subjectivity of applying a grade to individual cases. Capturing the specific recommendations for contour amendment unique to each case could elucidate systematic errors that could be addressed with curriculum modification.

Incorporating a formal, consensus-driven RTP review process may further strengthen HN peer review and enhance the educational experience for RO RPs. Future research can expand upon these findings to optimize the use of HN peer review as an educational tool, guiding the development of targeted training strategies, educational resources, and objective assessment methods for RO RPs nationwide.

## Conclusion

This study demonstrates that the incorporation of a formal HN contouring peer-review process and RO RP target contour grade assignment into routine clinical practice is feasible and practical. The peer-review process can be used to objectively monitor RO RP contour competency progression and can enhance the existing framework of ACGME milestones.


https://doi.org/10.37549/ARO-D-25-0010

## References

1) Rajeev-Kumar G, Manjunath R, Hasan Y. Radiation oncology resident evaluations: current methods and resident perceptions. *Adv Radiat Oncol*. 2023;8(5):101230. doi:10.1016/j.adro.2023.101230

2) Thomas HR, Braunstein SE, Golden DW, et al. Milestones 2.0: refining the competency-based assessment of radiation oncology medical residents. *Int J Radiat Oncol Biol Phys*. 2023;115(1):12-18. doi:10.1016/j.ijrobp.2022.07.1838

3) Nissen C, Ying J, Kalantari F, et al. A prospective study measuring resident and faculty contour concordance: a potential tool for quantitative assessment of residents' performance in contouring and target delineation in radiation oncology residency. *J Am Coll Radiol*. 2024;21(3):464-472. doi:10.1016/j.jacr.2023.08.049

4) American Society for Radiation Oncology (ASTRO). RO-ILS case study 10: contour delineation: a crucial treatment planning process. Accessed January 3, 2025. https://www.astro.org/practice-support/quality-and-safety/ro-ils/case-studies/case-study-10

5) Abugideiri M, Schreibmann E, Switchenko J, et al. Prospective international pilot study evaluating the efficacy of a self-guided contouring teaching module with integrated feedback for transitioning from 2D to 3D treatment planning. *J Glob Oncol*. 2019;5(5):1-16. doi:10.1200/JGO.18.00224

6) Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? *Radiol Oncol*. 2016;50(3):254-262. doi:10.1515/raon-2016-0023

7) Gogineni E, Schaefer D, Ewing A, et al. Systematic implementation of effective quality assurance processes for the assessment of radiation target volumes in head and neck cancer. *Pract Radiat Oncol*. 2024;14(3):e205-e213. doi:10.1016/j.prro.2023.12.012

8) Bekelman JE, Wolden S, Lee N. Head-and-neck target delineation among radiation oncology residents after a teaching intervention: a prospective, blinded pilot study. *Int J Radiat Oncol Biol Phys*. 2009;73(2):416-423. doi:10.1016/j.ijrobp.2008.04.028

9) Martin-Garcia E, Celada-Álvarez F, Pérez-Calatayud MJ, et al. 100% peer review in radiation oncology: is it feasible? *Clin Transl Oncol*. 2020;22(12):2341-2349. doi:10.1007/s12094-020-02394-8

10) Gwynne S, Spezi E, Sebag-Montefiore D, et al. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. *Br J Radiol*. 2013;86(1024):20120398. doi:10.1259/bjr.20120398

11) Nielsen CP, Lorenzen EL, Jensen K, et al. Interobserver variation in organs at risk contouring in head and neck cancer according to the DAHANCA guidelines. *Radiother Oncol*. 2024;197:110337. doi:10.1016/j.radonc.2024.110337

12) Nuyts S. Defining the target for radiotherapy of head and neck cancer. *Cancer Imaging*. 2007;7 Spec No A(Special issue A):S50-S55. doi:10.1102/1470-7330.2007.9009

13) Marks LB, Adams RD, Pawlicki T, et al. Enhancing the role of case-oriented peer review to improve quality and safety in radiation oncology: executive summary. *Pract Radiat Oncol*. 2013;3(3):149-156. doi:10.1016/j.prro.2012.11.010

14) Ballo MT, Chronowski GM, Schlembach PJ, et al. Prospective peer review quality assurance for outpatient radiation therapy. *Pract Radiat Oncol*. 2014;4(5):279-284. doi:10.1016/j.prro.2013.11.004

15) Hoopes DJ, Johnstone PA, Chapin PS, et al. Practice patterns for peer review in radiation oncology. *Pract Radiat Oncol*. 2015;5(1):32-38. doi:10.1016/j.prro.2014.04.004

16) Duggar WN, Bhandari R, Yang CC, Vijayakumar S. Group consensus peer review in radiation oncology: commitment to quality. *Radiat Oncol*. 2018;13(1):55. doi:10.1186/s13014-018-1006-1

17) Cooper BT, Goenka A, Sine K, et al. Development of a comprehensive, contour-based, peer review workflow at a community proton center. *Int J Part Ther*. 2020;7(1):34-40. doi:10.14338/IJPT-19-00059.1

18) Brower JV, Chen S, Ritter A, et al. Comfort level of US radiation oncology graduates: assessment of transition to independent clinical practice. *J Cancer Educ*. 2021;36(2):278-283. doi:10.1007/s13187-019-01625-z

19) Boyd GH, Vanbenthuysen L, Jimenez RB. Radiation treatment plan evaluation education in residency: a needs assessment. *Int J Radiat Oncol Biol Phys*. 2023;117(2):e504-e505. doi:10.1016/j.ijrobp.2023.06.1753

# Auto-Contouring in Residency: Cutting Corners or Creating Confidence?

Elizabeth Thompson, MD

In the field of radiation oncology, the use of artificial intelligence (AI) for auto-contours is quickly becoming mainstream in settings from academics to private practice. A vast array of auto-contouring technologies is available, from cloud-based to deep-learning platforms.

Most contouring software programs that radiation oncologists use have options built into their capabilities that include auto-contouring organs at risk (OAR) or recognizing and defining target volumes. One survey found that about half of radiation oncologists are using auto-contouring for OAR at least half of the time, while only about 3% use it for target volumes.[1]

There are many benefits of auto-contouring. One study found that it improves efficiency by reducing overall contouring time by as much as 20% to 40%, allowing physicians to spend more time on other important clinical tasks.[2] Additionally, auto-contouring could help standardize contouring throughout an institution, and potentially throughout the field of radiation oncology, which could benefit future research efforts.

The drawbacks of auto-contouring mainly include issues with accuracy, as most auto-contours still require editing by a physician, which is time-consuming.

How does this affect radiation oncology residents and our education?

A survey from one institution suggests that residents and faculty disagree on whether auto-contouring affects the understanding of OARs.[3] In this example, faculty reported that some of the auto-contours they reviewed were incorrect, and they were unclear as to whether residents were aware of the mistakes. Additionally, they felt that the lack of repetition of OAR contouring hindered residents' review of CT anatomy.

Residents in the survey believed that auto-contouring improved their understanding of anatomy and positively impacted their education.[3] They noted that although the auto-contours were not perfect and required review, the time saved on manual contouring improved their quality of life and their ability to spend time on other educational activities.

However, both groups agreed that auto-contouring had a positive impact on clinic workflow and the overall education of residents.[3]

Anecdotally, the use of auto-contouring varies by residency program, with some allowing unrestricted use and others prohibiting it entirely. Some programs require PGY2 residents to contour manually, in the hope that they will master the technique before using auto-contouring as an aid to increase speed and efficiency in practice.

My institution has no specific rules against using auto-contouring, but my attendings and some co-residents recommended that I freehand the OAR contours each time I started a new disease site for the first few plans, then compare my results with the AI-generated contours. I followed this advice, and I think it was an effective way to learn anatomy without having to spend hours manually contouring, for example, every slice of the lung.

I have found there is a mixed use of auto-contouring among my co-residents. It is frequently

Dr Elizabeth Thompson, MD, is a PGY2 Radiation Oncology Resident, University of Florida, Gainesville, FL

**Disclosures:** The author has no conflicts of interest to disclose. None of the authors received outside funding for the production of this original manuscript and no part of this article has been previously published elsewhere.
**Corresponding author:** Elizabeth Thompson, MD, University of Florida, Gainesville, FL. (elthompson513@gmail.com)

https://doi.org/10.37549/ARO-D-25-00015

used to contour large, solid organs such as the liver and lungs, but manual contouring is standard practice for organs like the bowel and esophagus, where auto-contouring is not as reliable.

For some cases, we use an MR-based linac that does not have auto-contouring software. When we are planning cases on that machine, it is necessary to manually draw all OARs—thus, the knowledge and understanding of organs without the help of AI is still essential.

I view auto-contouring as a tool to provide guidance when treating a new anatomical area, to be used in conjunction with traditional contouring atlases. I find that after a few cases that have similar OARs, auto-contouring frees up time to tackle general treatment paradigms or research more complex cases.

Yet I do wonder whether I have become too reliant on these tools, and if I am truly confident identifying all critical structures without AI assistance. This conflict underscores a larger point: while AI is an incredibly useful aid, it is no substitute for a solid understanding of anatomy. It is still essential to study guidelines, review literature, and know nodal station boundaries thoroughly.

Auto-contouring is here to stay, and it will likely continue to evolve and improve. Used thoughtfully, it is a powerful educational and clinical tool. But as residents, we must ensure that it complements, not compromises, our training.

## References

1) Zhang YH, Cha E, Lynch K, et al. Attitudes and access to resources and strategies to improve quality of radiotherapy among US radiation oncologists: a mixed methods study. *J Med Imaging Radiat Oncol*. 2022;66(7):993-1002. doi:10.1111/1754-9485.13423

2) Baroudi H, Brock KK, Cao W, et al. Automated contouring and planning in radiation therapy: what is "clinically acceptable"? *Diagnostics (Basel)*. 2023;13(4):667. doi:10.3390/diagnostics13040667

3) Qian AS, Phuong C, Ni L, et al. Artificial intelligence auto contouring and its impact on resident education. *Int J Radiat Oncol Biol Phys*. 2024;119(4):e13. doi:10.1016/j.ijrobp.2024.04.054

AppliedRadiationOncology

www.appliedradiationoncology.com