2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Automated Identification of AI Contouring Outliers Using Template Shape-Based Analysis and Centroid Mapping

Libing Zhu[1], Jingwei Duan[2], Aimee Dubuque[1], Nathan Y. Yu[1], Randa Tao[1], Yi Rong[1], Quan Chen[1]
[1]Department of Radiation Oncology, Mayo Clinic Arizona,
[2]Department of Radiation Oncology, University of Texas MD Anderson Cancer Center

ABSTRACT

Background:
Deep learning–based auto-segmentation has become widely adopted in radiation therapy; however, automatically generated contours may still contain anatomically implausible errors, including substantial mislocalization or severe under- and over-segmentation. When these errors are overlooked, they can negatively affect treatment planning efficiency and patient safety.
Purpose:
To develop and validate a fully automated quality assurance framework capable of detecting major contouring errors in clinically approved organ-at-risk (OAR) contours without relying on case-specific reference contours.
Methods:
A retrospective dataset of 1,746 clinically approved cases was analyzed. Shape abnormalities were identified by registering each OAR to a template contour using coherent point drift (CPD) and evaluating surface deviations. Spatial inconsistencies were assessed by applying Random Sample Consensus (RANSAC) to OAR centroids. A contour was classified as a major outlier if either the bidirectional local surface distance or the centroid displacement exceeded predefined thresholds.
Results:
Shape analysis flagged 82 potential outlier cases, of which 53 were confirmed as true positives, including 17 thoracic, 4 abdominal, 29 genitourinary, and 3 head-and-neck cases. Centroid analysis identified 37 true-positive cases using a 2σ threshold.
Conclusions:
The proposed template-based framework combines shape and centroid analyses to automatically identify major contouring errors in clinically approved OAR contours without requiring patient-specific ground truth. This approach provides an efficient quality assurance strategy that can assist clinicians by highlighting suspicious contours for targeted review, with the potential to improve contouring reliability and patient safety.

Quan Chen
Mayo Clinic Arizona
5881 E Mayo Blvd, Phoenix, AZ 85054
Chen.quan@mayo.edu

INTRODUCTION

Accurate delineation of target volumes and organs-at-risk (OARs) is essential for safe and effective radiation therapy. Although deep learning–based auto-segmentation has substantially improved contouring efficiency and generally achieves high geometric accuracy, clinically significant failures can still occur [1].

Most existing contour quality assurance (QA) methods evaluate AI-generated contours by comparing them with case-specific reference contours using geometric similarity metrics. While effective for retrospective validation, these approaches are not applicable during routine clinical practice because reference contours are unavailable at the time of contour generation [2]. Existing commercial QA tools primarily detect technical issues, such as missing slices or disconnected components, but have limited ability to identify anatomically implausible contour shapes or major spatial mislocalization.

In this study, we developed a template-based contour QA framework that combines shape and centroid analyses to identify major contouring errors without requiring case-specific reference contours. The method was validated using 1,746 clinically approved cases spanning four anatomical regions. By identifying clinically meaningful errors that remained after routine clinical review, this work demonstrates the potential of automated QA to provide an additional safeguard for improving contour reliability in radiation therapy.

METHODS AND MATERIALS

A total of 1,746 clinically approved cases were retrospectively collected for this study, including 814 thoracic, 169 abdominal, 601 genitourinary (GU), and 162 head-and-neck (HN) cases. Across these cases, 88 OARs were analyzed.

Two complementary contour error detection (CED) methods were developed to identify abnormal OAR contours. The first was a shape-based approach that employed CPD registration to align each OAR with an organ-specific template shape. Following registration, bilateral local distance (BLD) was calculated between each contour and the template to characterize local geometric variation across the population. The second approach utilized centroid-based analysis to detect global anatomical inconsistencies. A whole-body centroid template was constructed from clinically approved reference cases, with each OAR represented by its three-dimensional centroid as shown in Figure 1.

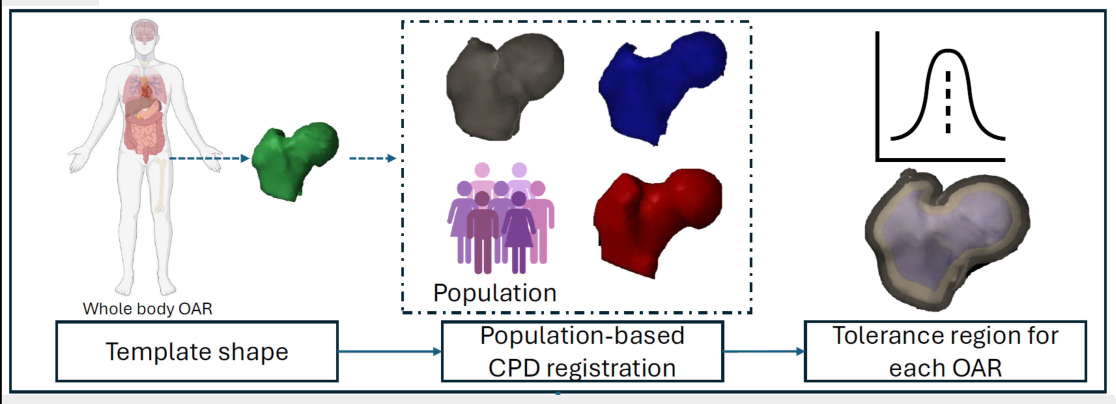

Figure 1. Workflow of clinical contouring error detection using template shaped-based alignment.

RESULTS

Among 1,746 clinically approved cases, the proposed contour error detection framework identified numerous clinically significant contouring errors that had been missed during routine human review in Table 1. At the selected 2% shape-based threshold, 82 cases were flagged, of which 64 represented previously undetected true-positive contouring errors, corresponding to positive predictive values of 73.7%, 77.8%, 96.7%, and 61.1% for thoracic, abdominal, GU, and head-and-neck structures, respectively.

Centroid-based analysis complemented shape-based detection by identifying contour localization and labeling errors that could not be recognized from shape alone, such as left-right organ swaps and organ mislocalization. True-positive detections increased as the centroid threshold was relaxed but plateaued at approximately 2.2σ, indicating that lower thresholds primarily increased false positives.

Table 1. Clinical contouring error detection results using organ centroid mapping.

Threshold (σ)	Flagged case	TP	FP	PPV	FDR
2	196	37	159	18.9%	81.1%
2.2	123	37	86	30.1%	69.9%
2.5	65	31	34	47.7%	52.3%
3	32	22	10	68.8%	31.3%
3.5	26	20	6	76.9%	23.1%
4	22	17	5	77.3%	22.7%

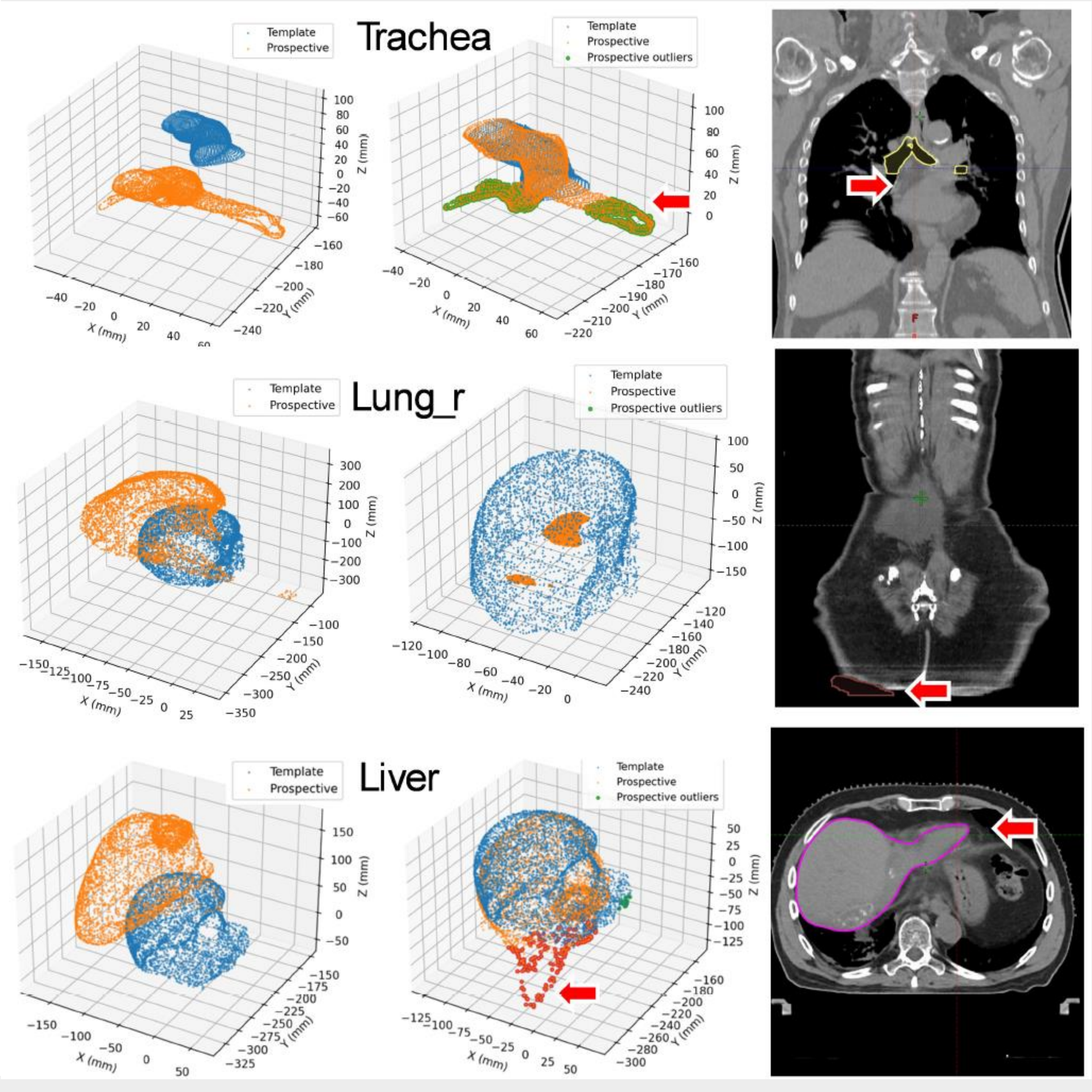

Figure 2. Examples of detected major errors for clinically approved contours

DISCUSSION

This study demonstrates that combining shape- and centroid-based outlier detection provides an effective automated quality assurance approach for identifying major contouring errors in clinically approved radiotherapy contours. The two methods complement each other by detecting different types of failures, including abnormal shapes, organ mislocalization, and left-right mislabeling. Importantly, the framework identified numerous clinically significant errors that had been missed during routine human review, highlighting that residual contouring errors remain in real-world clinical practice despite existing review processes. Rather than replacing human review, the proposed method serves as an additional safety layer to identify uncommon but potentially consequential contouring failures before treatment planning. Threshold selection represents a practical trade-off between detection sensitivity and reviewer workload rather than a formal ROC-based optimization. Lowering the detection threshold increased the number of true-positive detections but also substantially increased false positive cases.

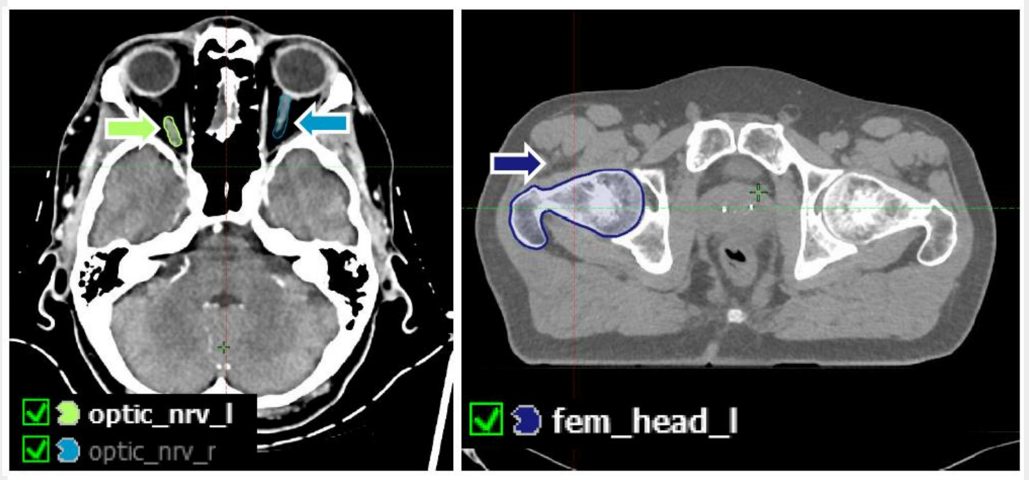

Figure 3. Examples of detected major errors for clinically approved contours using centroids

CONCLUSIONS

This study presents a template-based contour quality assurance framework that integrates shape and centroid analyses to identify clinically significant contouring errors without requiring reference contours for individual patients. The two approaches complement one another by detecting different types of failures, including abnormal organ geometry, spatial mislocalization, and laterality errors. Evaluation on clinically approved datasets demonstrated that the framework can identify major contouring errors that remained undetected during routine clinical review. These findings suggest that the proposed method has the potential to serve as an independent quality assurance layer, improving the safety and reliability of AI-assisted contouring in the radiotherapy workflow.

REFERENCES

1. Ford E, Conroy L, Dong L, de Los Santos LF, Greener A, Gwe-Ya Kim G, et al. Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275. Medical Physics 2020;47:e236-e72.
2. Rong Y, Chen Q, Fu Y, Yang X, Al-Hallaq HA, Wu QJ, 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;119:261-80.