Automated Dose-Based Anatomic Region and Target Organ Classification of Radiotherapy Treatment for Big Data Applications

Justin B. Hink[1], Yasin Abdulkadir[1], Jack Neylon[1], Jim M. Lamb[1]

[1]Department of Radiation Oncology, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, California, USA

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

ABSTRACT

Background: Curation remains a significant barrier to using "big data" radiotherapy planning databases of 100,000+ patients. As the field moves toward training large-scale foundation models, anatomic site stratification becomes essential for downstream tasks. However, current methods rely on locally implemented plan labels or target nomenclature. This reliance on institutional metadata is highly unreliable for multi-institutional data, ultimately bottlenecking the utilization of massive radiotherapy archives.

Purpose: To address the need for robust data standardization, we developed software that automates labeling by inferring anatomic regions and target organs directly from dose-volume overlap with deep-learning segmentations, eliminating metadata reliance.

Methods: The software processes DICOM files in bulk, utilizing TotalSegmentator and an in-house breast model to segment 116 structures (organs, glands, and bones) categorized into six regions: Cranial, Head and Neck, Pelvis, Abdomen, Thorax, and Extremity. It converts the 85% and 50% isodose lines to binary structure masks to compute organ-specific dose-overlap metrics. Plans are assigned ranked regional labels and target organ labels based on these intersections. The regional algorithm was refined using 109 expert-labeled cases and validated on 100 consecutive clinical plans. The target organ algorithm was optimized by maximizing the Matthews correlation coefficient on a 358-case cohort and validated on the same 100-plan dataset.

Results: Regional labeling achieved 90.0% Exact Accuracy (matching all expert labels and order), 93.0% Top-2 Accuracy (matching the top two expert regions regardless of order), and 95.0% Top-1 Accuracy (matching the primary expert label). The target organ algorithm achieved a sensitivity of 1.0 for the brain, breast, lung, and vertebrae, with corresponding specificities of 0.99, 1.0, 0.99, and 0.98. For the prostate, the algorithm achieved a specificity of 1.0 but a comparatively lower sensitivity of 0.73. This lower sensitivity occurred because optimizing the Matthews correlation coefficient naturally converged on a high threshold, driven by the upstream TotalSegmentator model occasionally contouring the prostate bed as an intact prostate.

Conclusions: By resolving the fundamental limitations of non-standardized metadata, this algorithm provides a scalable, standardized foundation essential for curating large, multi-institutional datasets required for "big data" radiotherapy research and provides an important complement to text-based approaches.

CONTACT

Justin Hink
University of California – Los Angeles
jhink@mednet.ucla.edu
(520) 334-5909

INTRODUCTION

The Problem: Curating "big data" radiotherapy databases requires knowing each plan's treated anatomic region or target organ, but institutional metadata and text-based plan/target nomenclature are inconsistent across sites, unreliable even within a single institution, and effectively unusable for multi-institutional cohorts needed to train large-scale AI models.

Our Solution: A fully automated, metadata-independent algorithm that determines anatomic region and target organ directly from geometric dose-mask overlap with deep-learning organ segmentation (116 structures via TotalSegmentator + an in-house breast model), validated against expert labels on 100 consecutive clinical plans from our CT simulation schedule.

METHODS AND MATERIALS

Data Processing & Workflow (see Figure 1):

- **Autosegmentation:** Processes bulk DICOM files using TotalSegmentator and an in-house breast model to contour 116 structures.
- **Classifications:** Maps plans to six Anatomic Regions (Cranial, Head & Neck, Pelvis, Abdomen, Thorax, Extremity) and 13 Target Organs.
- **Intersection Analysis:** Converts the volume within 85% and 50% of the maximum dose into binary masks ($V_{85\%}$ and $V_{50\%}$) to compute physical dose overlaps.

Geometric Metrics:

Classification utilizes three volumetric metrics between the autosegmented organ (*O*) and the dose volume (*D*):

- **Dose-Overlap Score:** Fraction of the dose volume occupied by the segmented organ. $Score_{dose} = \frac{V(O \cap D)}{V(D)}$
- **Structure-Overlap Score:** Fraction of the total organ volume occupied by the dose. $Score_{struct} = \frac{V(O \cap D)}{V(O)}$
- **RMS Score:** Balances dose concentration against total organ involvement. $Score_{RMS} = \sqrt{\frac{Score_{dose}^2 + Score_{struct}^2}{2}}$

Classification Algorithms

- **Anatomic Region:**
  - **Primary Logic**: Ranks regions by the highest single-organ overlap where the $V_{85\%}$ $Score_{dose} > 0.001$.
  - **Special Conditions**: Cranial is assigned if Brain Dice > 0.4, and if no organs meet the primary threshold, Head & Neck is assigned if the $V_{85\%}$ Center of Mass is superior to the C7 vertebrae.
  - **Secondary Logic**: Ranks regions by the highest single-organ overlap where the $V_{50\%}$ $Score_{dose} > 0.001$.
  - **Fallback Logic**: If still unlabeled, assign the label of the organ axially closest to the $V_{85\%}$ Center of Mass.
- **Target Organ:**
  - Evaluates 13 target sites using the $Score_{RMS}$ metric at both the $V_{85\%}$ and $V_{50\%}$ levels. Applies optimized, organ-specific thresholds found by maximizing the Matthews correlation coefficient on the development cohorts.

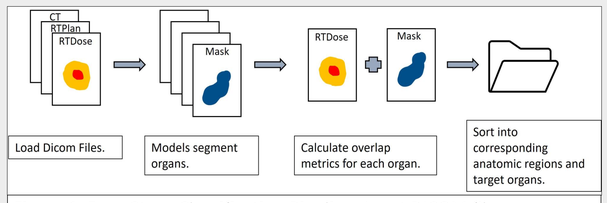

Figure 1. Four-Stage Classification Pipeline. Input DICOM files are segmented via deep learning, followed by geometric dose-mask intersection analysis to automatically sort treatment plans into specific anatomic regions and target organs.

RESULTS

Overall Results:

Plan-Level Regional Accuracy: Achieved 90.0% Exact, 93.0% Top-2, and 95.0% Top-1 Accuracy (Figure 2).
Region-Specific Metrics: Precision ranged from 0.80 (Extremity) to 1.00 (Cranial, Head & Neck), with Sensitivity >0.97 across all six anatomic regions (Figure 4).
Target Organ Sensitivity: Achieved perfect (1.0) sensitivity for high-volume sites including brain, breast, lung, vertebrae (Table 1).
Plan-Level Target Organ Performance: Target organ classification yielded a sensitivity of 0.85, specificity of 0.92, precision of 0.91, and F1-Score of 0.88 (Figure 3).

Figure 2. End-to-End Anatomic Region Performance. Exact Accuracy (90%) demands perfect structural and sequential matching against the manual expert consensus baseline on a per-plan basis. Top-1 (95%) and Top-2 (93.0%) metrics reflect the algorithm's performance in identifying the high-dose treatment regions. Center points represent mean performance values; whiskers denote asymmetric 95% confidence intervals.

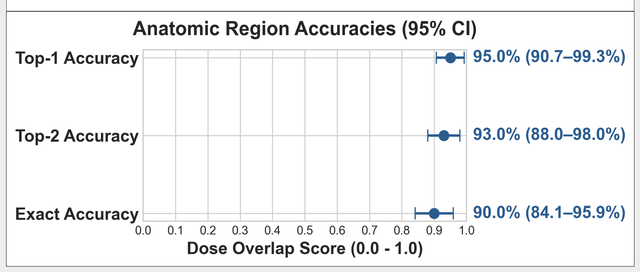

Figure 3. Plan-Level Target Organ Classification Metrics. Performance is assessed using a strict plan-level constraint where any single missed or extraneous predicted structure constitutes a global network failure (n = 41 true positives, 49 true negatives, 4 false positives, 7 false negatives). Points represent the calculated metrics (Sensitivity: 0.85, Specificity: 0.92, Precision: 0.91, F1-Score: 0.88); whiskers denote asymmetric 95% confidence intervals.

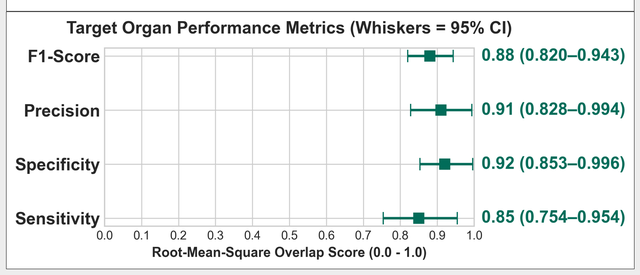

Figure 4. Anatomic Region Multi-Metric Profile. Heatmap of five performance metrics (Top-1 Accuracy, Top-2 Accuracy, Sensitivity, Precision, and F1-Score) for each of the six anatomic regions across all 100 validation plans. Cell shading reflects metric value (scale at right). Extremity shows the lowest precision (0.80) of any region, while Cranial and Head & Neck achieve near-perfect performance across all five metrics.

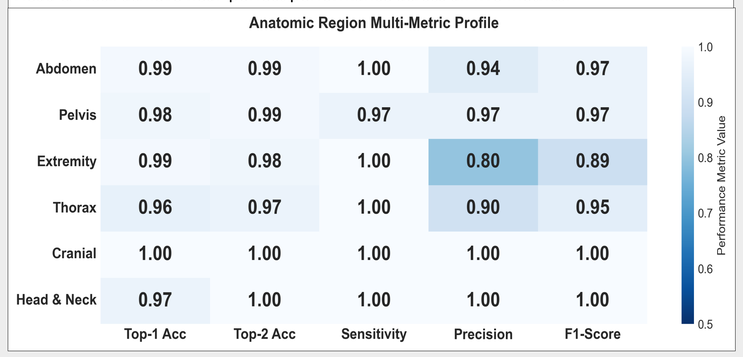

Table 1. Organ-level performance of the automated target organ classification pipeline on the 100 consecutive treatment plan validation dataset.

- **Liver:** No cases present in this sample.
- **Prostate Bed / Rectum:** Unsupported by the current algorithm but retained to demonstrate true real-world failure rates (precision undefined).

	TP	TN	FN	FP	Sensitivity	Specificity	Precision
Brain	10	89	0	1	1	0.99	0.91
Breast Left	3	97	0	0	1	1	1
Breast Right	3	97	0	0	1	1	1
Lung Left	3	97	0	0	1	1	1
Lung Right	4	95	0	1	1	0.99	0.8
Prostate	8	89	3	0	0.73	1	1
Prostate Bed	0	99	1	0	0	1	-
Rectum	0	97	3	0	0	1	-
Liver	0	100	0	0	-	1	-
Vertebrae_C	3	97	0	0	1	1	1
Vertebrae_T	4	95	0	1	1	0.99	0.8
Vertebrae_L	3	97	0	0	1	1	1
Vertebrae_S	2	97	0	1	1	0.99	0.67

DISCUSSION

Success & Significance:

- **Proof of concept:** Objective, geometric dose-volume metrics can achieve reliable radiotherapy plan curation.
- **Regional accuracy:** 95.0% Top-1 Accuracy shows strong clinical alignment.
- **Target organ accuracy:** Perfect sensitivity for high-volume sites (Brain, Lung, and Breast).
- **Clinical utility:** Automation and metadata-independence allows for efficient database organization and curation, accelerating cohort identification and large-scale AI training.

Key Challenges:

- **Segmentation-dependent:** Performance is capped by the upstream deep-learning segmentations available.
- **Protocol mismatch:** Framework counts small dose coverage as a hit; experts apply a 5 mm penetration rule and ignore grazes, a source of small but critical disagreements.

Future steps:

- **Better segmentation:** Improving upstream autosegmentation to close current gaps.
- **Broader validation:** Test on a larger, multi-institutional dataset, including liver cases.
- **Refine fallback logic:** Use a dose-weighted Center of Mass for asymmetric-dose edge cases.

CONCLUSIONS

The Solution: A fully automated, metadata-independent labeling framework utilizing objective dose-volume metrics and deep-learning autosegmentation.

The Results: 95.0% Top-1 regional accuracy and perfect (1.0) sensitivity for high-volume targets (brain, breast, lung); plan-level performance of target organs was 0.85 sensitivity / 0.92 specificity across 100 validation plans.

The Limitation: Accuracy is capped by upstream segmentations, most notably the rectum and post-operative prostate bed.

The Impact: Provides a baseline for curating massive multi-institutional datasets, enabling large-scale deep-learning models and "big data" radiotherapy research.

REFERENCES

Wasserthal J, Breit H-C, Meyer MT, et al. doi:10.48550/arXiv.2208.05868
Walter A, Hoegen-Saßmannshausen P, Stanic G, et al. *Cancers*. 2024;16(2):415.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com