Physics and Imaging in Radiation Oncology 37 (2026) 100907

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Tumor-conditioned inter-patient registration using planning computed tomography for voxel-based analysis to predict radiation pneumonitis in lung cancer patients

Chloe Min Seo Choi [a,d], Jue Jiang [a], Nikhil P. Mankuzhy [b], Nishant Nadkarni [a], Sudharsan Madhavan [a], Abraham J. Wu [b], Joseph O. Deasy [a], Maria Thor [a], Andreas Rimner [c], Harini Veeraraghavan [a,*]

[a] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, United States
[b] Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, United States
[c] Department of Radiation Oncology, University of Freiburg, Freiburg, Germany
[d] Department of Radiation Oncology, Yonsei Cancer Center, Heavy Ion Therapy Research Institute, Yonsei University College of Medicine, Seoul, the Republic of Korea

ARTICLE INFO

Keywords:
Lung cancer
Deep learning
Inter-patient registration
Voxel-based analysis
Radiation pneumonitis prediction

ABSTRACT

Background and purpose: Deformable image registration (DIR) for voxel-based analysis (VBA) can be challenging in patients with non-small cell lung cancer (NSCLC) due to large variations in tumor size and location. This study aimed to assess whether a tumor-preserving inter-patient DIR approach improves VBA-based prediction of radiation pneumonitis (RP).

Methods and materials: Three DIR methods were evaluated: deep learning-based Tumor-Aware Recurrent Registration (TRACER) and Patient-Specific Context and Shape (PACS), trained on a public dataset of 268 locally-advanced (LA) NSCLC patients, and iterative Symmetric Normalization (SyN). All methods were tested on 240 patients with LA-NSCLC. Geometric, dosimetric, and tumor preservation metrics were compared using the Wilcoxon signed-rank test. VBA was conducted with each DIR method to identify cohort-relevant regions (CRRs). Machine learning models incorporating clinical, dosimetric, and CRR dose features were used to predict grade 2 or higher RP.

Results: TRACER best preserved tumor volume (1.39 %) and organ doses (mean 0.08 Gy) compared with PACS and SyN ($p < 0.001$). PACS showed higher geometric but worse dose preservation accuracy than TRACER. All DIR-based VBA methods identified the right lung as the CRR associated with RP. TRACER-derived CRR had slightly higher RP predictive performance (AUC 0.78 vs PACS 0.73 vs SyN 0.71), and outperformed the MLD-based ML model (AUC = 0.78 vs 0.69, $p = 0.04$; specificity = 0.62 vs 0.48).

Conclusions: TRACER improved registration accuracy, with better tumor volume preservation and reduced OAR dose impact. Incorporating VBA-derived dose enhanced RP prediction accuracy compared with using MLD. CRRs identified through VBA were robust to the choice of DIR.

1. Introduction

Voxel-based analysis (VBA) is a method to study population-based radiation therapy (RT) dose–response relationships for radiation toxicities following RT [1–6]. VBA extracts spatially varying dose effects, revealing detailed and localized dose–response relationships that are often missed by traditional DVH-derived summary metrics (e.g., mean dose or V20). VBA requires accurate spatial normalization of patient scans to a common reference through inter-patient deformable image registration (DIR). Iterative DIR methods such as Symmetric Normalization (SyN) optimize spatial alignment between each moving and a reference image by maximizing image and geometric similarity [7,8] and have been used for VBA including lung cancer [9–12]. These methods provide good alignment but may require extensive pre-processing for each image pair. Recently, deep learning (DL)-based DIR methods have been developed that predict transformations with

* Corresponding author at: Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 321, East 61St, New York, NY 10065, United States.
E-mail address: veerarah@mskcc.org (H. Veeraraghavan).

https://doi.org/10.1016/j.phro.2026.100907
Received 9 July 2025; Received in revised form 15 January 2026; Accepted 16 January 2026
Available online 20 January 2026

minimal preprocessing, achieving accuracy comparable to or better than iterative approaches [13–16].

Inter-patient DIR of scans containing cancer presents unique challenges. For instance, tumors introduce considerable anatomic variations on an individual level that cannot be easily accounted for when aligning two different patient scans [17]. Previous methods sidestepped this issue by selecting reference patients with small tumors, masking out anatomically different regions (i.e., cost function masking), or focusing only on organs such as the heart that were unaffected by potential tumor-induced misregistration [18–20]. If required, poorly registered moving images are excluded from further analysis [10,21]. Depending on the analyzed endpoint, excluding poorly registered patients may limit findings to anatomically similar cases, reducing applicability to the broader population. Methods used in neuroimaging that exclude tumor masks or replace tumors with normal tissue intensities from contralateral voxels [22–26] are inapplicable to organs outside the brain due to limited left–right symmetry. Likewise, DL-based DIR methods for tracking longitudinal treatment changes in tumors [27–29] assume spatial correspondence of tumors, which is not possible for inter-patient DIR.

Not accounting for tumors in non-corresponding locations may lead to erosion of tumor volumes in the moving image, unnatural stretching of normal tissues in the reference image and significant dose discrepancies in the deformed dose for organs and tumors [30]. Our approach, called tumor-aware recurrent registration (TRACER), explicitly addresses the issue of non-corresponding tumors by using lesion masks to enforce preservation of tumor volume and topology following deformation. This process, which we refer to as tumor conditioning, helps maintain tumor integrity while preserving the topology of normal tissues surrounding tumors [30].

To the best of our knowledge, the impact of using tumor-preserving DIR on VBA has not been studied. Hence, we performed an end-to-end analysis with the following aims: (a) compare geometric and dosimetric accuracies of TRACER against two other DIR methods for inter-patient computed tomography (CT) registration, (b) study the impact of DIR on VBA to extract cohort-relevant region (CRR) to predict grade 2 or higher radiation pneumonitis (RP) in patients with locally advanced non-small cell lung cancer (LA-NSCLC) undergoing RT, and (c) evaluate the benefit to predictive accuracy of combining CRR dose with clinical and dosimetric variables to predict RP using machine learning (ML) methods.

2. Material and methods

2.1. Overview of the study

This study consists of three components for end-to-end evaluation of DIR-based VBA. First, TRACER was compared against two published DIR methods, a DL-DIR method called patient-specific anatomic context (PACS) [31] and iterative DIR-based SyN [7], using an independent testing dataset to measure their geometric, dosimetric, and tumor preservation accuracy. Second, all DIR methods were used to extract VBA-based CRR. Third, the impact of CRR dose on the downstream RP prediction was evaluated by combining it with clinical and dosimetric predictors in ML models. The overall workflow is summarized in Fig. 1.

2.2. Dataset characteristics

This study used separate datasets for DL-DIR (TRACER and PACS) training and VBA testing. The iterative DIR method Symmetric Normalization (SyN) was applied only to the testing dataset, as it did not require training. The training dataset consisted of 268 free-breathing 3D planning CT scans and expert delineated gross tumor volume (GTV) of LA-NSCLC patients treated either RT alone or chemoradiation from a public dataset [30]. The VBA testing dataset included 240 patients with stage IIIA–IIIB LA-NSCLC, treated with chemotherapy and CT-based intensity-modulated RT (IMRT) to 60 Gy in 2 Gy fractions at Memorial Sloan Kettering Cancer Center (MSKCC) between 2004 and 2014. All patients underwent free-breathing 3D planning CT, which were used to evaluate DIR methods and perform RP prediction. Of the 240 patients, 29 developed grade 2 or higher RP based on CTCAE v4.03 [31]. Retrospective analysis was approved by the MSKCC institutional review board.

Expert-verified segmentation of GTV and organs at risk (OARs), including the heart, esophagus, and lungs, was available from a published study using the testing dataset [32]. Additional OARs including the inferior vena cava (IVC), left and right ventricles (LV, RV), left and right atria (LA, RA), pulmonary artery (PA), and superior vena cava (SVC), the proximal bronchial tree (PBT), and spinal cord (SC), were segmented using a bespoke AI model [33].

2.3. Analyzed methods

Three DIR methods, bespoke DL-DIR TRACER [30] (Supplementary Fig. S1a), PACS [31] (Supplementary Fig. S1b), and an iterative DIR method called SyN, available through the ANTs library [7], were analyzed. TRACER and PACS use the same architecture, consisting of 3D convolutional long short-term memory (CLSTM) blocks (Supplementary Fig. S1d) within their encoders, followed by convolutional block decoders (Supplementary Fig. S1e). CLSTM is used to progressively refine deformations as a sequence to capture large anatomical differences. Whereas PACS uses a pair of images as input, TRACER also takes the GTV masks in moving and target images as additional inputs for tumor conditioning. SyN performs iterative intensity-based registration using a multi-step optimization strategy (Supplementary Fig. S1c).

2.4. Data preprocessing

For training and testing, planning CT images were resampled to 1 × 1 × 3 mm resolution and cropped to include the lungs and chest with a five-slice margin in the z-axis. To remove the background and couch, a body mask was created for each CT slice by thresholding at −200 HU. Next, the largest connected component mask was extracted after morphological hole filling and erosion. The relevant image and GTV mask volumes of interest were extracted by cropping the body mask with a 30-pixel margin.

Consistent with prior studies, the planning CT scan of a patient with lung volume closest to the population median was chosen as the

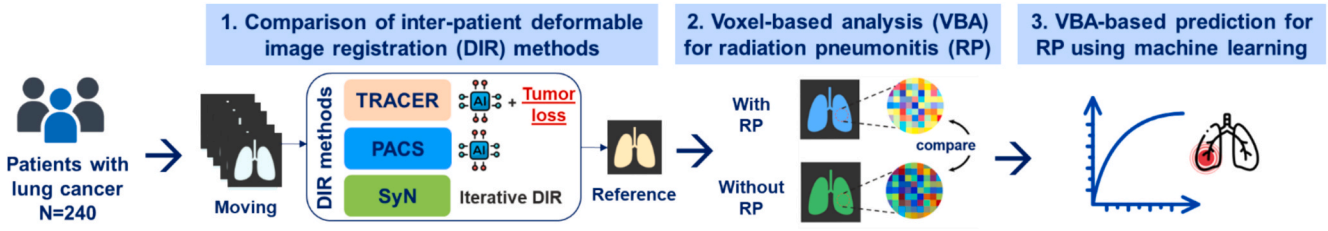

Fig. 1. Overview of the study workflow.

reference for spatial normalization of all patients [1,4]. All patient scans were rigidly aligned to this reference CT scan using SimpleElastix prior to performing DIR [34]. DL-DIR methods (TRACER and PACS) required additional preprocessing, including intensity clipping within −500 HU to 500 HU, followed by normalization to [0,1].

2.5. DIR models pipeline

2.5.1. Deep learning models (TRACER and PACS)

TRACER [30] and PACS [31] were trained from scratch using random pairs of images (35 K examples) from the training dataset with data augmentation (flipping, rotation, and intensity scaling) for up to 1500 epochs with a batch size of 4 on an Nvidia A100 GPU, Adam optimizer, and a learning rate of 1e-4. A grid search was conducted on the training set to optimize TRACER and PACS hyperparameters (smoothness range: 25–50, CLSTM steps: 4–8), which extracted optimal values (smoothness: 45, CLSTM steps: 8) and were used for testing [30,31]. Both models were optimized using intensity (mean square error), geometry (Dice similarity of OARs) and bi-directional tumor rigidity used selectively for TRACER for topology preservation.

2.5.2. Iterative model (SyN)

SyN was implemented using the AntsPy registration tool in Python (version 3.9.12). A multi-step strategy, adapted from a prior VBA study with parameter modifications (see Supplementary Table S1) was used [35]. First step used affine registration performed using the Mattes mutual information (MMI) metric (32 histogram bins, 30 % random sampling) with a multiresolution progressive down sampling and Gaussian smoothing. In the second step, DIR was applied with OAR mask guidance and optimized with both MMI and normalized cross-correlation. The third step used an aggressive local and fine-grained alignment approach called SyNAggro. Moving images were resampled using linear interpolation, and their corresponding masks were resampled using nearest-neighbor interpolation.

2.6. Metrics used for registration accuracy evaluation

Geometric accuracies were computed between aligned and reference patients' OARs using the Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff distance (HD95). Organ dose preservation accuracy compared the mean dose between the original moving and resampled moving images per organ. Dose preservation ratio (DPR) compared the relative difference in volumes receiving high dose between the deformed moving and original moving images as: DPR = $V60_{def}/V60_{m}$, where $V60_{def}$ and $V60_{m}$ indicate tumor volume receiving at least 60 Gy in the deformed moving and moving images, respectively. A DPR of 1 indicates accurate preservation; values >1.0 reflect overestimation of dose in the deformed image, whereas values <1.0 indicate underestimation. Tumor volume change (ΔT) was calculated as $(|V_{def} - V_m|/|V_m|) \times 100$, with V_{def} and V_m representing tumor volumes post- and pre-DIR, respectively. Statistical comparisons used a two-sided, paired Wilcoxon signed-rank test, with significance set at $p < 0.05$.

2.7. Voxel-based analysis

Following DIR, voxel-wise two-sample t-tests on spatially normalized dose maps were conducted for all three DIR methods to compare dose distributions between patients with and without RP. A step-down max-T permutation procedure with 1000 random label shuffles was used to account for multiple comparisons [6]. For each permutation, the maximum voxel-wise t-value across the lung was recorded to construct a null distribution, against which the observed voxel-wise t-statistics were compared. Voxels with permutation-corrected p-values < 0.05 were considered significantly associated with RP, defining the CRR [6].

2.8. Machine learning based RP prediction

A total of 16 clinical and dosimetric features, including age, sex, height, smoking status, histology, concurrent chemotherapy, Karnofsky Performance Status, mean dose within CRR (CRR dose), lung dose metrics (maximum, mean, and V20), mean GTV dose, esophageal dose (maximum and mean), and mean heart dose, were analyzed.

Firstly, the impact of DIR methods for predicting RP using the corresponding CRR dose was evaluated. ElasticNet logistic regression models (penalty='elasticnet' and saga solver in Scikit-learn [version 1.7.2]) were built using DIR-specific CRR dose, along with all features except mean lung dose (MLD). ElasticNet was selected as no hyperparameter tuning was needed and trained using stratified 10-fold cross-validation (CV).

Secondly, the predictive performance of the CRR dose compared with MLD was evaluated to assess the benefit of VBA. Using TRACER, two variants of prediction models were developed: one using the CRR dose and the other using the MLD, along with all other features. To assess the effect of ML models, ElasticNet was compared against random forest (RF) and XGBoost, using nested stratified 10-fold cross-validation and an inner 3-fold RandomizedSearchCV to select hyperparameters. Categorical features were one-hot encoded, and numerical features were standardized. Class imbalance was handled using a hybrid Synthetic Minority Oversampling with Edited Nearest Neighbors (SMOTEENN) sampling [36]. All preprocessing and class-imbalance handling were done within the training folds, and model performance was assessed on the outer test folds.

Feature importance within each model was evaluated and visualized using Shapley Additive Explanations (SHAP) [37] using Python (version 3.9.12). Models' accuracy was computed using the area under the ROC curve (AUC), sensitivity, and specificity.

2.9. Subanalysis to assess the robustness of models to patient characteristics

Normality of the data distributions was assessed before statistical analysis. The impact of anatomical differences on DSC for OARs (lung, heart, esophagus, PBT, and spinal cord) for individual patients due to size measured as body mass index (BMI) and gender was evaluated using Pearson correlation and Mann-Whitney U tests, respectively. Pearson correlations were classified as weak ($\rho < 0.3$), moderate ($0.3 < \rho < 0.5$), or strong ($\rho > 0.5$).

3. Results

3.1. Comparison of OAR accuracy

SyN and PACS caused tumor volume loss and tissue stretching, while TRACER preserved the tumor without distortion and avoided physically unrealistic stretching of normal tissues (Fig. 2). Importantly, TRACER preserved the isodose lines following resampling of the moving image, including for the tumor, and these closely matched those in the original moving image (Fig. 2c,d). SyN and PACS resulted in larger deviations of isodose lines, particularly for the tumor, as seen by larger differences in isodose line shapes in both representative cases.

PACS showed the highest deviations in organ-specific mean dose mapping (Fig. 3a), specifically for the PA, heart, esophagus, and lungs. In contrast, TRACER showed the lowest dose deviations, with a mean difference of 0.08 Gy across all organs and a maximum deviation of 0.66 Gy for the PA, significantly lower than PACS (2.3 Gy for the lung) and SyN (2.2 Gy for PBT). Although PACS outperformed TRACER and SyN for HD95 and DSC (Fig. 3b, 3c), it produced worse preservation of dose due to unbounded tissue warping that does not account for non-corresponding tumor locations in the image pairs.

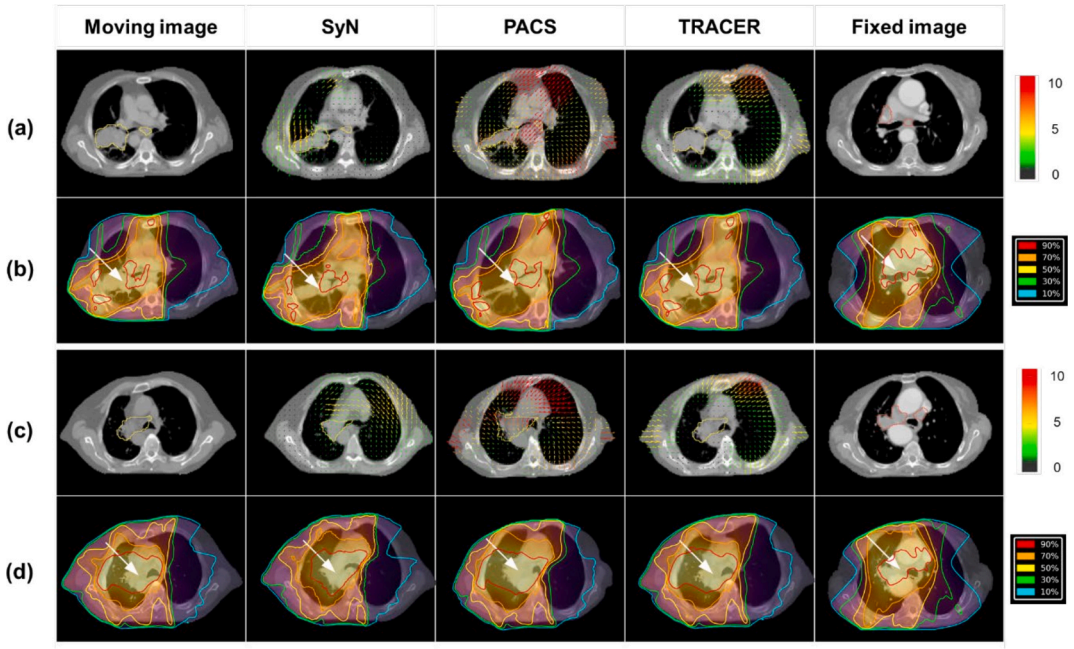

Fig. 2. Two examples illustrating the rationale for tumor conditioning with corresponding dose maps. (a) Moving image with a right lung tumor extending to the hilum, SyN-, PACS-, and TRACER-generated deformed images, and the fixed image. (b) Original and deformed isodose lines for the case in (a). (c) Moving image with a centrally located tumor, SyN-, PACS-, and TRACER-generated deformed images, and the fixed image. (d) Original and deformed isodose lines for the case in (c). Tumors on moving and fixed images are overlaid in yellow and red, respectively; Colored arrows indicate the 3D deformation field generated from deformable image registration; white arrows mark tumor locations; and contour lines represent isodose distributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Comparison of tumor volume and dose preservation accuracy

TRACER produced the best dose preservation with a DPR of 0.97, compared with PACS (1.08; $p < 0.001$) and SyN (0.78; $p < 0.001$). The population DPR is visualized in Fig. 3d, which also shows the least variability of TRACER compared with PACS and SyN. TRACER also produced the best tumor volume preservation, with the smallest ΔT (1.39 %) compared to PACS and SyN (26.83 % and 29.04 %, respectively; both $p < 0.001$) (Fig. 3e).

3.3. VBA for predicting grade 2 or higher RP

For all three DIRs, the VBA-derived CRRs were located in the middle and lower lobes of the right lung (Fig. 4). When comparing the predictive performance of the CRR dose derived from each DIR method, TRACER achieved a slightly higher RP predictive performance (0.78) compared with PACS (0.73) and SyN (0.71), with comparable sensitivity and specificity (Table 1). Shapley Additive Explanation (SHAP) analysis ranked CRR as the second most important feature for TRACER and PACS and the sixth for SyN (Supplementary Fig. S2).

Next, the predictive performance of the CRR dose versus MLD across different ML models is summarized in Table 2. Models incorporating CRR consistently outperformed those using MLD, with AUCs ranging from 0.74 to 0.78 versus 0.64 to 0.69, respectively. ElasticNet achieved the highest accuracy in both CRR- and MLD-based predictions.

3.4. Robustness to differences in patient characteristics

Model performance was consistent across patient subgroups. Accuracy was similar for both sexes (Supplementary Fig. S3a). Weak correlations between DSC and BMI were noted for all OARs, with modest correlations for the PBT and SC using TRACER and SyN (Supplementary Fig. S3b).

4. Discussion

This study compared tumor-conditioned TRACER with two other inter-patient DIR methods and used them in VBA to identify the CRR for predicting grade 2 or higher RP in patients with large tumor variability. The novelty of our work lies in applying a topology-preserving DIR model to VBA for the first time and demonstrating the end-to-end impact on downstream prediction modeling. We found that TRACER best preserved tumor and led to more accurate VBA-based predictions compared with other DIR methods.

Building on previous DIR studies, we investigated how tumor-conditioned TRACER affects VBA-derived predictions of RP. Monti et al. assessed the impact of elastic registration on VBA by comparing Demons and B-spline registrations for late lung toxicity after thoracic irradiation [38]. They found both methods produced similar overall patterns of dose-toxicity associations, despite different geometric accuracies. Our study extends this work in two ways: first, by examining DL-based DIR approaches, including tumor-conditioned, topology-preserving DIR (TRACER) and PACS alongside iterative DIR; and second, by demonstrating that the choice of DIR impacts downstream predictive performance for RP.

Our results highlighted that preventing non-physical deformation of high-dose regions following inter-patient DIR, resulted in better dose preservation. On the other hand, for the VBA, only a slight improvement in accuracy was noted for RP prediction with TRACER. All three methods identified similar CRR regions and produced comparable specificity and sensitivity for RP prediction. In addition, TRACER-VBA was more accurate than a published study that applied a normal tissue complication probability model to the same test cohort (AUC = 0.78 vs 0.73, respectively) [39]. On the other hand, neither PACS (AUC = 0.73) nor SyN (AUC = 0.71) outperformed the published study.

Another key strength of this study is its ability to handle inter-patient DIR in patients with large, heterogeneous tumors. Monti et al. [38] studied lung and heart toxicities in Hodgkin lymphoma patients without visible lung tumors. McWilliam et al. included tumors up to 34 cc but

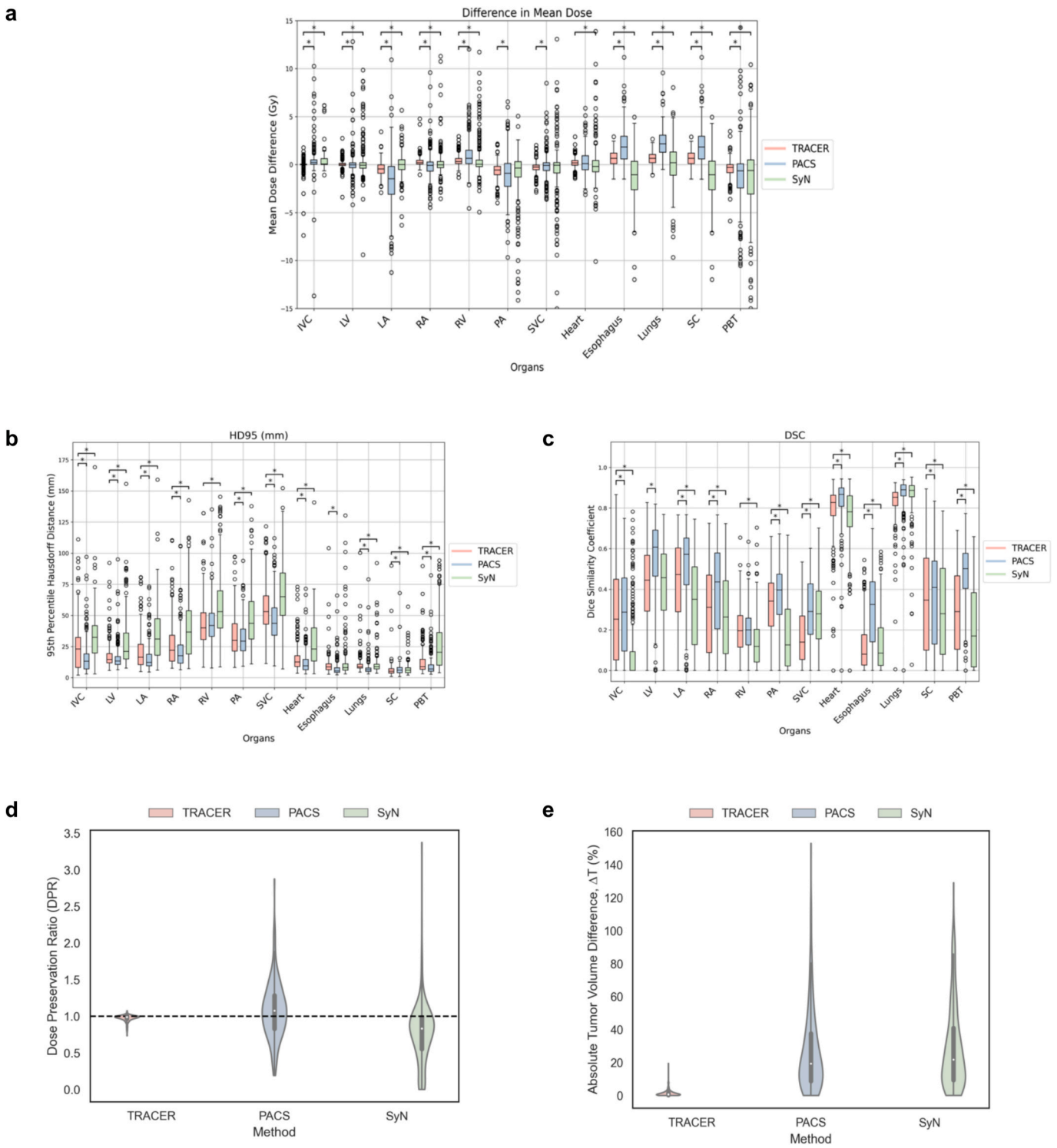

Fig. 3. Comparison of registration accuracies: (a) difference in mean OAR dose, (b) HD95, (c) DSC, (d) dose preservation ratio (DPR) at 60 Gy across methods, the dotted line (DPR = 1) represents optimal preservation of the high-dose region, (e) absolute tumor volume difference, where larger values indicate greater changes in tumor volume post-registration. Asterisks (*) indicate a statistically significant difference, with a p-value ≤ 0.05.

focused on cardiac toxicity, where lung misalignments had little effect [20]. Visser et al. [40] addressed DIR errors in glioblastoma by selecting reference images with spatially similar pathology, which is impractical for VBA. In contrast, TRACER was applicable to stage III lung cancer patients with large tumor variability and preserved tumor volume and high-dose geometry more accurately than the other DIR methods. The poorer tumor preservation observed with PACS, despite a similar DL backbone, highlights the importance of enforcing topology constraints. The strengths and weaknesses of the analyzed DIR methods are summarized in Supplementary Table S2.

The findings of the current study suggest several potential applications in the RT planning context. Previous studies have suggested that population-based spatial information could be incorporated into inverse IMRT optimization to guide stricter clinical objectives by avoiding CRRs and thereby improving OAR sparing [3,4,41]. Another application could be patient monitoring in adaptive RT settings, where population-level

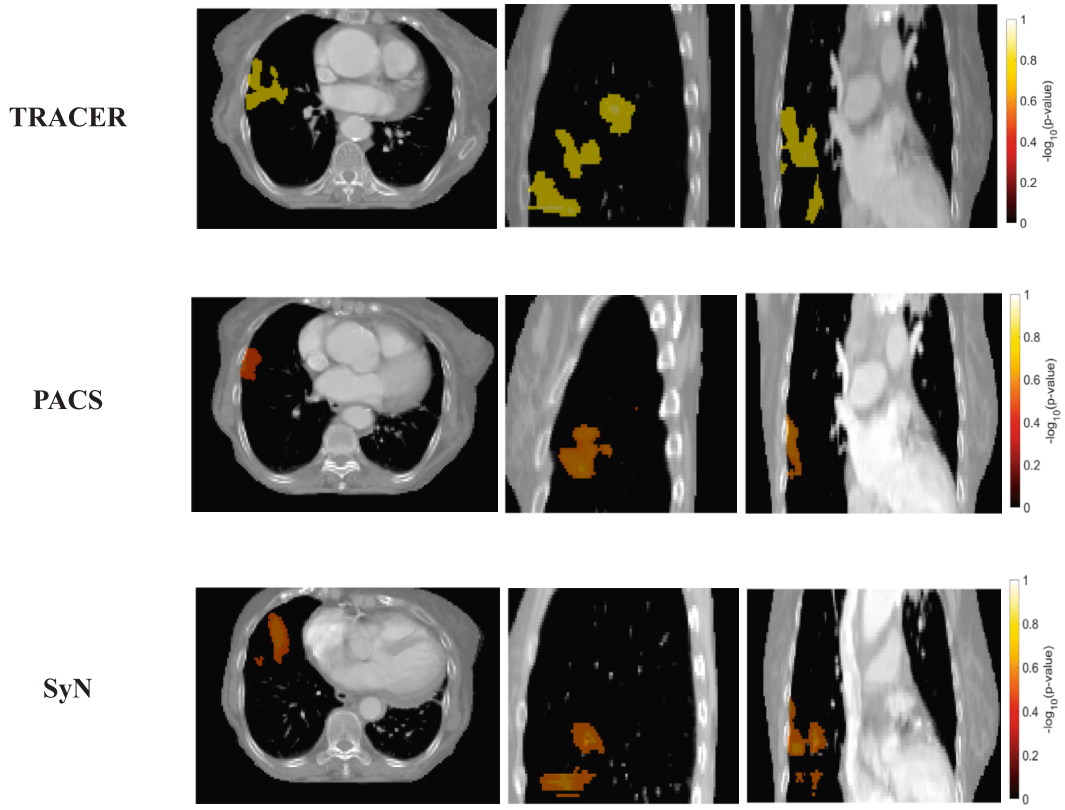

Fig. 4. Cohort relevant region for radiation pneumonitis identified through voxel based analysis of the three deformable image registration methods (TRACER, PACS and SyN), shown in axial, coronal, and sagittal views.

Table 1
Performance of ElasticNet-based prediction models using cohort-relevant region dose derived from each deformable image registration (DIR) methods. The 95% confidence intervals are given in brackets.

DIR method	AUC	Specificity	Sensitivity
SyN	0.71 ± 0.21 (0.55–0.87)	0.57 ± 0.18 (0.44–0.71)	0.72 ± 0.30 (0.49–0.94)
PACS	0.73 ± 0.19 (0.59–0.87)	0.58 ± 0.15 (0.47–0.69)	0.63 ± 0.27 (0.43–0.83)
TRACER	0.78 ± 0.18 (0.64–0.92)	0.62 ± 0.13 (0.52–0.71)	0.73 ± 0.32 (0.49–0.97)

CRRs may help identify patients who could benefit from replanning [42].

This study has a few limitations. First, our analysis did not account for intra-patient motion because the analyzed scans were 3D free-breathing scans. Second, we used a previously published single-institution dataset to evaluate the feasibility of predicting RP using an ML approach with VBA. Hence, the TRACER-identified CRR is specific to this cohort. Identifying biological correlates and radiosensitive regions will require larger, multi-institutional cohorts, which is beyond the scope of this study.

In conclusion, this study showed that incorporating tumor-conditioned DIR (TRACER) into inter-patient VBA enhanced tumor preservation. All evaluated DIR methods consistently identified the right lung as the CRR, demonstrating robustness to the choice of registration method. Using the CRR-based dose feature improved accuracy over the conventional whole-lung mean dose. These findings indicate the potential utility of TRACER for VBA in heterogeneous thoracic cohorts.

Data statement

Research data are not available at this time.

Ethical approval

This retrospective analysis was approved by the MSKCC institutional review board.

Table 2
Performance of machine learning based prediction models using cohort-relevant region (CRR) dose derived from TRACER (CRR dose model) vs. mean lung dose (MLD model). All other clinical and dosimetric features remained the same. Asterisk indicates a statistically significant difference. The 95% confidence intervals are given in brackets.

	Model type	AUC	Specificity	Sensitivity
MLD model	Random forest	0.64 ± 0.17 (0.51–0.77)	0.72 ± 0.11(0.63–0.80)	0.40 ± 0.28(0.19–0.62)
	XGBoost	0.66 ± 0.19(0.52–0.80)	0.73 ± 0.10(0.65–0.80)	0.45 ± 0.37(0.17–0.73)
	ElasticNet	0.69 ± 0.15(0.57–0.80)	0.48 ± 0.09(0.41–0.54)	0.73 ± 0.32(0.49–0.97)
CRR model	Random forest	0.74 ± 0.17(0.61–0.87)	0.86 ± 0.10(0.79–0.93)	0.40 ± 0.32(0.16–0.64)
	XGBoost	0.76 ± 0.14(0.65–0.87)	0.84 ± 0.09(0.78–0.91)	0.55 ± 0.29(0.33–0.77)
	ElasticNet	0.78 ± 0.18*(0.64–0.92)	0.62 ± 0.13(0.52–0.71)	0.73 ± 0.32(0.49–0.97)

CRediT authorship contribution statement

Chloe Min Seo Choi: Methodology, Investigation, Software, Formal analysis, Writing – original draft. **Jue Jiang:** Software, Formal analysis, Writing – review & editing. **Nikhil P. Mankuzhy:** Investigation, Writing – review & editing. **Nishant Nadkarni:** Software. **Sudharsan Madhavan:** Investigation, Writing – review & editing. **Abraham J. Wu:** Supervision, Writing – review & editing. **Joseph O. Deasy:** Conceptualization. **Maria Thor:** Data curation, Writing – review & editing. **Andreas Rimner:** Supervision, Writing – review & editing. **Harini Veeraraghavan:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: A.R. received consulting fees from AstraZeneca, Merck, and MoreHealth; honoraria from Boehringer Ingelheim; support for attending meetings from AstraZeneca; served on the data safety monitoring board for Merck; and holds positions as Vice President of ITMIG, member of the Board of Directors at IMIG, and oral board examiner for the ABR. All other authors have no disclosures to share..

Acknowledgement

This research was partially supported by the NCI R01CA258821.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2026.100907.

References

[1] Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello MC, et al. A voxel-based approach to explore local dose differences associated with radiation-induced lung damage. Int Radiat Oncol Biol Phys 2016;96:127–33. https://doi.org/10.1016/j.ijrobp.2016.04.033.
[2] Monti S, Paganelli C, Buizza G, Preda L, Valvo F, Baroni G, et al. A novel framework for spatial normalization of dose distributions in voxel-based analyses of brain irradiation outcomes. Phys Med 2020;69:164–9. https://doi.org/10.1016/j.ejmp.2019.12.017.
[3] Acosta O, Drean G, Ospina JD, Simon A, Haigron P, Lafond C, et al. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. Phys Med Biol 2013;58:2581. https://doi.org/10.1088/0031-9155/58/8/2581.
[4] Mylona E, Acosta O, Lizee T, Lafond C, Crehange G, Magné N, et al. Voxel-based analysis for identification of urethrovesical subregions predicting urinary toxicity after prostate cancer radiation therapy. Int Radiat Oncol Biol Phys 2019;104:343–54. https://doi.org/10.1016/j.ijrobp.2019.01.088.
[5] Beasley W, Thor M, McWilliam A, Green A, Mackay R, Slevin N, et al. Image-based data mining to probe dosimetric correlates of radiation-induced trismus. Int J Radiat Oncol Biol Phys 2018;102:1330–8. https://doi.org/10.1016/j.ijrobp.2018.05.054.
[6] Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: a methodological cookbook. Phys Med 2020;69:192–204. https://doi.org/10.1016/j.ejmp.2019.12.013.
[7] Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 2008;12:26–41. https://doi.org/10.1016/j.media.2007.06.004.
[8] Klein S, Staring M, Pluim JPW. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Trans Image Process 2007;16:2879–90. https://doi.org/10.1109/TIP.2007.909412.
[9] Klein A, Ghosh SS, Avants B, Yeo BTT, Fischl B, Ardekani B, et al. Evaluation of volume-based and surface-based brain image registration methods. Neuroimage 2010;51:214–20. https://doi.org/10.1016/j.neuroimage.2010.01.091.
[10] Monti S, Palma G, Caroprese M, Cella L. External validation of dose patterns and dosimetric predictors for radiation-induced esophagitis in non-small cell lung cancer. Int Radiat Oncol Biol Phys 2025. https://doi.org/10.1016/j.ijrobp.2025.07.1445.
[11] Palma G, Monti S, Thor M, Rimner A, Deasy JO, Cella L. Spatial signature of dose patterns associated with acute radiation-induced lung damage in lung cancer patients treated with stereotactic body radiation therapy. Phys Med Biol 2019;64:155006. https://doi.org/10.1088/1361-6560/ab2e16.
[12] Cella L, Monti S, Xu T, Liuzzi R, Stanzione A, Durante M, et al. Probing thoracic dose patterns associated to pericardial effusion and mortality in patients treated with photons and protons for locally advanced non-small-cell lung cancer. Radiother Oncol 2021;160:148–58. https://doi.org/10.1016/j.radonc.2021.04.025.
[13] Ramadan H, El Bourakadi D, Yahyaouy A, Tairi H. Medical image registration in the era of transformers: a recent review. Inf Med Unlocked 2024;49:101540. https://doi.org/10.1016/j.imu.2024.101540.
[14] Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Anal 2018;49:1–13. https://doi.org/10.1016/j.media.2018.07.002.
[15] Haskins G, Kruger U, Yan P. Deep Learning in Medical Image Registration: A Survey 2020. https://doi.org/10.48550/arXiv.1903.02026.
[16] Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: Transformer for unsupervised medical image registration. Med Image Anal 2022;82:102615. https://doi.org/10.1016/j.media.2022.102615.
[17] Gil N, Lipton ML, Fleysher R. Registration quality filtering improves robustness of voxel-wise analyses to the choice of brain template. Neuroimage 2021;227:117657. https://doi.org/10.1016/j.neuroimage.2020.117657.
[18] Craddock M, Nestle U, Koenig J, Schimek-Jasch T, Kremp S, Lenz S, et al. Cardiac function modifies the impact of heart base dose on survival: a voxel-wise analysis of patients with lung cancer from the PET-plan trial. J Thorac Oncol 2023;18:57–66. https://doi.org/10.1016/j.jtho.2022.09.004.
[19] Jönsson H, Ekström S, Strand R, Pedersen MA, Molin D, Ahlström H, et al. An image registration method for voxel-wise analysis of whole-body oncological PET-CT. Sci Rep 2022;12:18768. https://doi.org/10.1038/s41598-022-23361-z.
[20] McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. Eur J Cancer 2017;85:106–13. https://doi.org/10.1016/j.ejca.2017.07.053.
[21] Mylona E, Ebert M, Kennedy A, Joseph D, Denham J, Steigler A, et al. Rectal and urethro-vesical subregions for toxicity prediction after prostate cancer radiation therapy: validation of voxel-based models in an independent population. Int Radiat Oncol Biol Phys 2020;108:1189–95. https://doi.org/10.1016/j.ijrobp.2020.07.019.
[22] Brett M, Leff AP, Rorden C, Ashburner J. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 2001;14:486–500. https://doi.org/10.1006/nimg.2001.0845.
[23] Ghaleh M, Lacey EH, Fama ME, Anbari Z, DeMarco AT, Turkeltaub PE. Dissociable mechanisms of verbal working memory revealed through multivariate lesion mapping. Cereb Cortex 2020;30:2542–54. https://doi.org/10.1093/cercor/bhz259.
[24] Jühling D, Rajashekar D, Cheng B, Hilgetag CC, Forkert ND, Werner R. Spatial normalization for voxel-based lesion symptom mapping: impact of registration approaches. Front Neurosci 2024;18. https://doi.org/10.3389/fnins.2024.1296357.
[25] Nachev P, Coulthard E, Jäger HR, Kennard C, Husain M. Enantiomorphic normalization of focally lesioned brains. Neuroimage 2008;39:1215–26. https://doi.org/10.1016/j.neuroimage.2007.10.002.
[26] Hwang K, Shine JM, Bruss J, Tranel D, Boes A. Neuropsychological evidence of multi-domain network hubs in the human thalamus. Elife 2021;10:e69480. https://doi.org/10.7554/eLife.69480.
[27] Hu X, Yang J, Yang J. A CNN-based approach for lung 3D-CT registration. IEEE Access 2020;8:192835–43. https://doi.org/10.1109/ACCESS.2020.3032612.
[28] Ding Y, Feng H, Yang Y, Holmes J, Liu Z, Liu D, et al. Deep-learning based fast and accurate 3D CT deformable image registration in lung cancer. Med Phys 2023;50:6864–80. https://doi.org/10.1002/mp.16548.
[29] Zheng Y, Jiang S, Yang Z. Deformable registration of chest CT images using a 3D convolutional neural network based on unsupervised learning. J Appl Clin Med Phys 2021;22:22–35. https://doi.org/10.1002/acm2.13392.
[30] Jiang J, Choi CMS, Thor M, Deasy JO, Veeraraghavan H. Tumor aware recurrent inter-patient deformable image registration of computed tomography scans with lung cancer. Med Phys 2025;52:938–50. https://doi.org/10.1002/mp.17536.
[31] Jiang J, Veeraraghavan H. One shot PACS: Patient specific Anatomic Context and Shape prior aware recurrent registration-segmentation of longitudinal thoracic cone beam CTs. IEEE Trans Med Imaging 2022;41:2021–32.
[32] Hotca A, Thor M, Deasy JO, Rimner A. Dose to the cardio-pulmonary system and treatment-induced electrocardiogram abnormalities in locally advanced non-small cell lung cancer. Clin Transl Radiat Oncol 2019;19:96–102. https://doi.org/10.1016/j.ctro.2019.09.003.
[33] Jiang J, Hu Y-C, Liu C-J, Halpenny D, Hellmann MD, Deasy JO, et al. Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. IEEE Trans Med Imaging 2019;38:134–44. https://doi.org/10.1109/TMI.2018.2857800.
[34] Marstal K, Berendsen F, Staring M, Klein S. SimpleElastix: A User-Friendly, Multi-Lingual Library for Medical Image Registration, 2016, p. 134–42.
[35] Madhavan S, Gamez M, Garces YI, Lester SC, Ma DJ, Mundy DW, et al. Impact of radiation dose distribution on nutritional supplementation needs in head and neck cancer radiotherapy: a voxel-based machine learning approach. Front Oncol 2024;14:1346797. https://doi.org/10.3389/fonc.2024.1346797.
[36] Yang F, Wang K, Sun L, Zhai M, Song J, Wang H. A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest

neighbor for missed abortion diagnosis. BMC Med Inf Decis Making 2022;22:344. https://doi.org/10.1186/s12911-022-02075-2.
[37] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc.; 2017.
[38] Monti S, Pacelli R, Cella L, Palma G. Inter-patient image registration algorithms to disentangle regional dose bioeffects. Sci Rep 2018;8:4915. https://doi.org/10.1038/s41598-018-23327-0.
[39] Thor M, Deasy J, Iyer A, Bendau E, Fontanella A, Apte A, et al. Toward personalized dose-prescription in locally advanced non-small cell lung cancer: Validation of published normal tissue complication probability models. Radiother Oncol 2019;138:45–51. https://doi.org/10.1016/j.radonc.2019.05.011.
[40] Visser M, Petr J, Müller DMJ, Eijgelaar RS, Hendriks EJ, Witte M, et al. Accurate MR image registration to anatomical reference space for diffuse glioma. Front Neurosci 2020;14. https://doi.org/10.3389/fnins.2020.00585.
[41] Wang Q, Stanforth A, LePain WA, Gelover E, Chen H, Zhu M, et al. Applying voxel-based analysis to oropharyngeal cancer proton therapy patients: a correlation study on radiation-induced acute dysphagia. Med Phys 2025;52:e70110. https://doi.org/10.1002/mp.70110.
[42] Shelley LEA, Sutcliffe MPF, Thomas SJ, Noble DJ, Romanchikova M, Harrison K, et al. Associations between voxel-level accumulated dose and rectal toxicity in prostate radiotherapy. Phys Imaging Radiat Oncol 2020;14:87–94. https://doi.org/10.1016/j.phro.2020.05.006.