Title

Results of an AI-Based Image Review System to Detect Patient Misalignment Errors in a Multi-Institutional Database of CBCT-Guided Radiotherapy Treatments

Authors

Luximon, Dishane C
Neylon, Jack
Ritter, Timothy
et al.

Publication Date

2024-03-01

DOI

10.1016/j.ijrobp.2024.02.065

INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Results of an Artificial Intelligence–Based Image Review System to Detect Patient Misalignment Errors in a Multi-institutional Database of Cone Beam Computed Tomography–Guided Radiation Therapy

Dishane C. Luximon, MS,* Jack Neylon, PhD,* Timothy Ritter, PhD,† Nzhde Agazaryan, PhD,* John V. Hegde, MD,* Michael L. Steinberg, MD,* Daniel A. Low, PhD,* and James M. Lamb, PhD*

*Department of Radiation Oncology, David Geffen School of Medicine, University of California, Los Angeles, California; and
†Department of Medical Physics, Virginia Commonwealth University, Richmond, Virginia

Received Sep 12, 2023; Accepted for publication Feb 28, 2024

Purpose: Present knowledge of patient setup and alignment errors in image guided radiation therapy (IGRT) relies on voluntary reporting, which is thought to underestimate error frequencies. A manual retrospective patient-setup misalignment error search is infeasible owing to the bulk of cases to be reviewed. We applied a deep learning–based misalignment error detection algorithm (EDA) to perform a fully automated retrospective error search of clinical IGRT databases and determine an absolute gross patient misalignment error rate.

Methods and Materials: The EDA was developed to analyze the registration between planning scans and pretreatment cone beam computed tomography scans, outputting a misalignment score ranging from 0 (most unlikely) to 1 (most likely). The algorithm was trained using simulated translational errors on a data set obtained from 680 patients treated at 2 radiation therapy clinics between 2017 and 2022. A receiver operating characteristic analysis was performed to obtain target thresholds. DICOM Query and Retrieval software was integrated with the EDA to interact with the clinical database and fully automate data retrieval and analysis during a retrospective error search from 2016 to 2017 and from 2021 to 2022 for the 2 institutions, respectively. Registrations were flagged for human review using both a hard-thresholding method and a prediction trending analysis over each individual patient's treatment course. Flagged registrations were manually reviewed and categorized as errors (>1 cm misalignment at the target) or nonerrors.

Corresponding author: Dishane C. Luximon, MS; E-mail: dluximon@mednet.ucla.edu

Disclosures: J.N. reports consulting fees from ViewRay, Inc, and U.S. Patent 9,786,092: Physics–based High–Resolution Head and Neck Biomechanical Models (issued 2017). N.A. reports the following: clinical cooperation grant by Brainlab; travel payments for invited presentations made by Brainlab; payment for expert testimony (The Law Offices of Michael D. Gonzalez); and membership in a Novalis Certified Accreditation Expert Panel. J.V.H. reports the following: consulting fees from ViewRay, Inc; participation in the UCLA Data Safety Monitoring Board; departmental GIFL support/clinical study materials support by Soylent Nutrition, Inc; and payment for expert testimony by ViewRay, Inc. M.L.S. reports the following: consulting fees from ViewRay, Inc; payment for expert testimony by Blue Shield Policy Consulting; and participation as chair of the ACGME Radiation Oncology Residency Review Committee. D.A.L. reports the following: a grant from Varian MRA; consulting fees from ViewRay, Inc; honoraria for talks by ViewRay, Inc; support for conferences by ViewRay, Inc, and Varian; and consulting fees from ViewRay, Inc. D.A.L. is the founder of Pulmonum LLC. This research was supported by the Agency for Healthcare Research and Quality (AHRQ) under award number 1R01HS026486.

Data Sharing Statement: Research data are not available at this time.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2024.02.065.

Int J Radiation Oncol Biol Phys, Vol. 000, No. 00, pp. 1–10, 2024
0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2024.02.065

Results: A total of 17,612 registrations were analyzed by the EDA, resulting in 7.7% flagged events. Three previously reported errors were successfully flagged by the EDA, and 4 previously unreported vertebral body misalignment errors were discovered during case reviews. False positive cases often displayed substantial image artifacts, patient rotation, and soft tissue anatomy changes.

Conclusions: Our results validated the clinical utility of the EDA for bulk image reviews and highlighted the reliability and safety of IGRT, with an absolute gross patient misalignment error rate of 0.04% ± 0.02% per delivered fraction.

Introduction

In radiation oncology, primary treatment goals are the precise delivery of radiation such that the dose to the target is optimized for better tumor control, while the doses to the organs at risk are minimized to limit side effects. Studies have shown that incidents regarding incorrect dose delivery are still prevalent even with safety procedures and technologies, such as image guidance.[1-3] Between 2014 and March 2023, a total of 3730 therapeutic radiation incidents were reported to the Radiation Oncology Incident Learning System (RO-ILS) Portal, with 18.4% of those being identified as having severe or moderate severity scores.[1]

According to the American Association of Physicists in Medicine (AAPM) Task Group 100, the patient positioning step within the external beam radiation therapy (EBRT) workflow is a high-severity and high-risk failure mode, ranking in the top 20% most hazardous steps in the entire external radiation therapy workflow, following a failure modes and effects analysis.[4] Ezzell et al have shown that in a cohort of 336 critical events submitted to RO-ILS between 2014 and 2016, 34 such errors occurred because the wrong shift was performed at the treatment table, with 29 of them reaching the patient.[5]

With the widespread adoption of stereotactic radiosurgery/stereotactic body radiation therapy (SRS/SBRT) and the introduction of ultrahigh-dose-rate treatments such as FLASH, proper patient setups become even more critical, owing to their high dose per fraction. In a 2017 study of RO-ILS SRS/SBRT events, Hoopes et al found that one of the most common event types was the incorrect shift and alignment of the patient.[6] McGurk et al further reinforced this finding when they discovered, in a 2023 study involving 4 institutions in the United States, a patient who was wrongly aligned for 1 of 5 of their multilevel spine SBRT fractions.[7] Those findings have shown that efforts are needed to find and analyze cases of reported and unreported patient setup incidents so that the characteristics and causes of these events can be understood and the treatment workflow can be consolidated accordingly to enhance patient safety.

The American College of Radiology American Society for Radiation Oncology Practice Parameter for Image-Guided Radiotherapy, the AAPM Task Group (TG) 275, and the Medical Physics Practice Guidelines (MPPG) 11.a all provide recommendations on how to mitigate alignment-based failure modes and promote incident learning as a way to reduce future events.[8-10] However, those reports note that many components of the current safety checks, including those involving patient setups, are heavily human reliant and therefore prone to be overlooked owing to the fast-paced working conditions in the clinic. For instance, McGurk et al have shown through a Human Factor Analysis and Classification System that 95.2% of 189 reported SBRT safeguard failures occurred due to human errors.[7]

As recommended by TG-275 and MPPG 11.a, the use of automation during these safety checks can act as a safety barrier and help in the analysis of bulk data for efficient incident learning by identifying error pathways that may not be easily detected by a human reviewer. Our group has previously developed deep learning–based algorithms for the detection of simulated gross misalignments for 2D planar x-ray image guidance[11] and cone beam computed tomography (CBCT) image guidance.[12] In prior studies, those algorithms were trained and tested on simulated patient misalignment errors, showing high sensitivity (>85.5%) in detecting the simulated misalignment errors for a fixed specificity of 95%. Furthermore, an automated pipeline using the CBCT-based patient misalignment detection algorithm was developed to facilitate offline image reviews and showed promising results when validated against expert medical physicists through a feasibility study.[13] In this study, we applied the previously developed artificial intelligence (AI)–based patient misalignment detection pipeline to perform a bulk retrospective patient setup error search on CBCT-guided radiation therapy treatments at 2 radiation therapy centers.

The primary goal of this study was to measure the rate of gross patient setup misalignment errors in CBCT-guided radiation therapy at 2 large academic centers. To the best of our knowledge, this is the first study to apply an AI-based image review system for a bulk retrospective patient misalignment error search on CBCT-guided radiation therapy treatments and to report previously unknown patient setup misalignments from the 2 institutions. The following points highlight the major contributions of this work:

- A manual retrospective patient misalignment search is infeasible owing to the large number of cases to be reviewed. By performing this AI-assisted image review, an absolute gross patient misalignment error rate in CBCT-guided radiation therapy at 2 radiation therapy facilities was determined, which is an important aspect of understanding radiation therapy safety.

- While previous studies have shown promising results for gross patient setup misalignment detection, the model validations were only performed on simulated errors,[12] and no gross patient misalignment error was found during the proof-of-concept implementation study.[13] This study focused on detecting and reporting real-world incidents, which is a fundamental step toward a robust validation of the real-world clinical performance of the tool, as suggested by the AAPM TG-273.[14]

Methods and Materials

Error detection algorithm

An error detection algorithm (EDA), composed of 3 distinct densely connected convolutional neural networks (CNNs), was developed to analyze the registration between the simulation computed tomography (simCT) scan and the setup CBCT scan. As the type of misalignment could vary by treatment site, based in part on the landmarks used during the registration process, each of the 3 CNNs in the EDA was designed to handle a specific body site and its corresponding error type.

In the thoracic-abdominal (TA) region, for example, vertebra identification errors can lead to an incorrect patient setup. This is a well-known risk in the EBRT process as reported by the French Nuclear Safety Authority, which encountered 40 events related to vertebral body misalignments between 2015 and 2017.[15] Hence, 1 of the 3 models in the EDA was trained to detect off-by-one vertebral body misalignments in the TA region. For the other regions, including head and neck (HN) and pelvis (PL), the risk of smaller, variable, 5- to 10-mm misalignments may occur owing to tumor growth or shrinkage, anatomy changes, and patient pose changes. However, depending on several factors, such as tumor location and size, the 5- to 10-mm misalignments may lead to variable treatment impact. To have an overall balance between clinical significance of the misalignment and a practical implementation, our definition of an error required a 10-mm misalignment, and the HN and PL models were trained accordingly.

As each CNN in the EDA was trained on a specific treatment region, it is essential to identify the treatment region present on each incoming scan such that it can be sent to the corresponding model. However, there is currently no robust indicator within the CBCT DICOM headers to identify the treatment region. Hence, a deep learning–based Anatomical Region Labeling (ARL) model[16] was trained to recognize the treatment region from a single coronal CBCT slice, such that the registered simCT-CBCT pair could be sent to the corresponding error detection pipeline and model, as shown in Figure 1A. During the validation phase, the ARL model achieved an overall accuracy of 99.1% in classifying CBCTs into the 3 distinct treatment regions (HN, TA, and PL).[16]

The error detection models in the EDA took as input 3 orthogonal planes (axial, coronal, and sagittal) from the registered setup CBCT and simCT scans. The 2D slice pairs were automatically extracted in each anatomic plane about a point within either the vertebral column (for the HN and TA regions) or the treatment isocenter (PL region). A 3-branch DenseNet architecture[17] was used for each error detection model, as shown in Figure 1B, such that each orthogonal plane was processed separately before going through a final dense layer and outputting a probability of misalignment between 0 (most unlikely) and 1 (most likely).

Under an institutional review board–approved protocol (#18-001430), registered simCT-CBCT image pairs were retrospectively collected from 680 patients treated between 2018 and 2022 at the University of California, Los Angeles Medical Center (UCLA) and between 2017 and 2019 at the Virginia Commonwealth University. The treatments at UCLA had been performed on the TrueBeam and Novalis Tx linear accelerator treatment machines (Varian Medical Systems), and the treatments at VCU were performed on Varian Trilogy and TrueBeam treatment machines. Those totaled to 7 C-arm linear accelerators, with varying couch features. For example, some machines had 6-degrees-of-freedom (6DOF) couches, while some did not have this feature. For the machines with a 6DOF couch feature, some used the feature for specialized cases only (eg, spine SBRT), while some used the feature for all cases. Hence, this added to the heterogeneity of the data set, making the algorithm more robust to machines with varying couch features.

It is noteworthy that the 2 institutions, in addition to being from 2 different states, are independent in terms of their clinical practices. However, a number of commonalities exist between the 2 institutions' IGRT protocols. For both institutions, whenever CBCT-guidance is prescribed, it is the final verification step before beam delivery, irrespective of additional image guidance. Following the CBCT, couch corrections are usually performed remotely, and in cases in which the couch is moved by more than 2 cm, it is standard procedure at both institutions to reimage the patient. In cases in which a patient was reimaged, only the final registration performed before beam delivery was used as the clinically performed registrations during model training and testing and during the retrospective error search.

During model training and testing, the clinically performed registrations were used as aligned (true-negative) cases. As gross misalignment was rare in the clinic, misalignment errors for each region were simulated for initial model training and testing. For TA scans, off-by-one vertebral body misalignments were simulated in both the superior and inferior directions. Those off-by-one vertebral body misalignments were manually performed to optimize the alignment at the adjacent vertebral body, including clinically reasonable rotations, hence mimicking this known failure mode and scenario. For HN and PL cases, 10-mm shifts were applied. For each individual CT-CBCT pair in the HN and PL data sets, the direction of the translation was randomly chosen and ranged from 1-dimensional (1 of cranial-caudal, posterior-

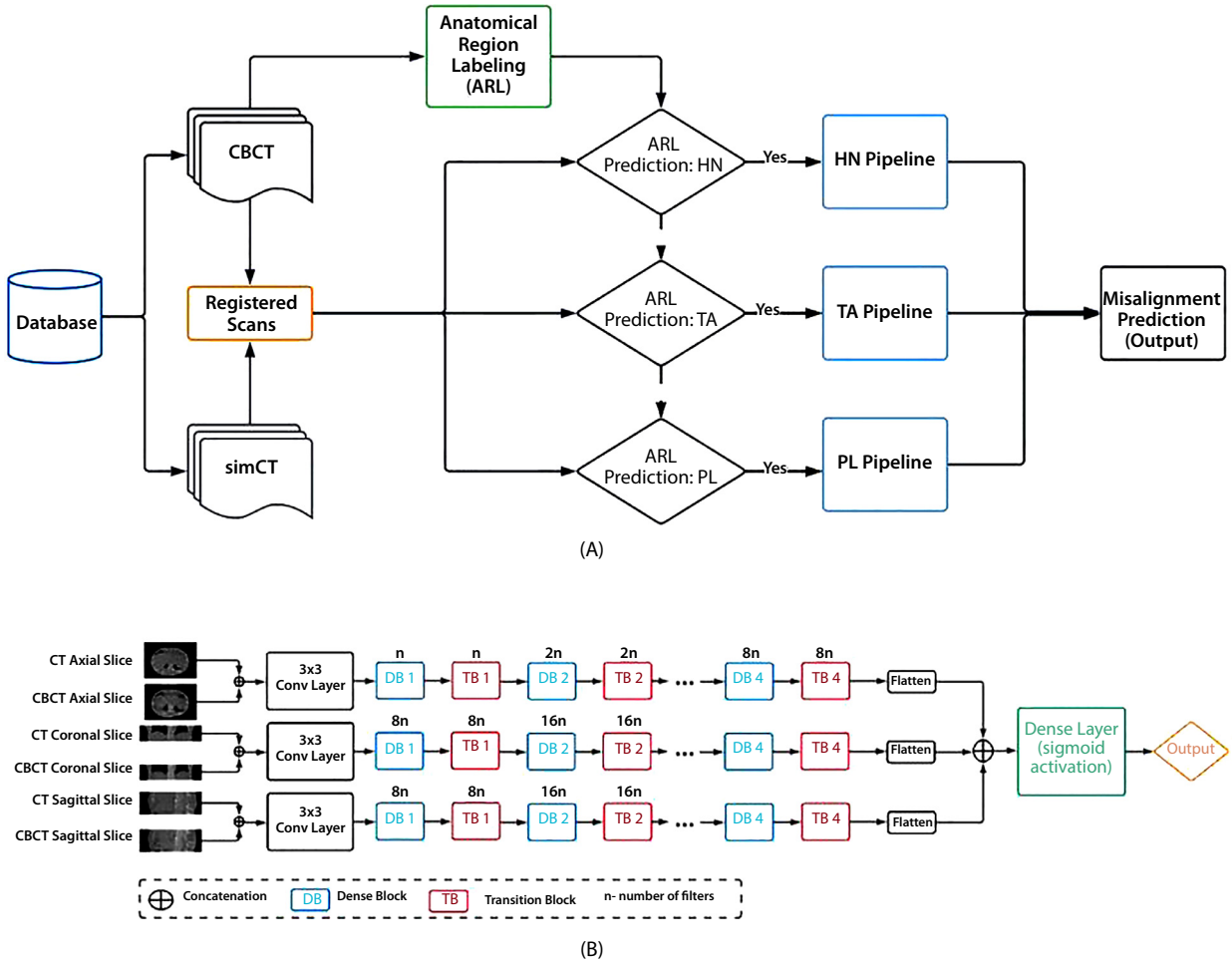

Fig. 1. (A) Illustration of the error detection algorithm used in this work. (B) The network architecture used in the error detection pipeline for each treatment site (n = 4). The Dense Block consists of 2 densely connected layers connected in a feed-forward mode (each composed of 2 convolutional layers, 2 batch normalization layers, 2 activation layers, and 1 dropout layer), and the Transition Block, 3 layers (batch normalization, convolutional, and max pooling). *Abbreviations:* HN = head and neck; PL = pelvis; TA = thoracic-abdominal.

anterior, or medial-lateral) to 3-dimensional shifts. Details regarding the training and validation data sets are shown in Table 1.

Following training, each model was validated on their respective test data set. Receiver operating characteristic (ROC) curves were subsequently produced, and the areas under the curves were calculated to assess the performance of each model. Furthermore, the thresholds leading to at least 90% (tr_{90}) and 99% (tr_{99}) sensitivities were found for each model, and the corresponding specificities were obtained.

Retrospective error search

In-house DICOM Query and Retrieval (DQR) software[13] was used to interact with the ARIA image management system to collect registrations performed between 2016 and 2017 at UCLA and between September 2021 and September 2022 at VCU. An example implementation of the DQR can be found on the referenced webpage.[18] Those images were automatically sent to the EDA for simCT-CBCT registration analysis. As described in the Error Detection Algorithm section, the EDA was trained separately on HN, TA, and PL images. Lower extremities (including glutes, thigh, knee, and calf) and upper extremities (including forearm, upper arm, and shoulder) were not used in training owing to their scarcity. In the retrospective search, those extremity cases were assigned to the HN, TA, or PL class by the ARL, which would then redirect the scans to the corresponding error detection model dealing with features most resembling the extremity site. Cases in which the imaged anatomy overlapped the HN, TA, and PL regions were assigned to the closest matching region by the ARL.

Registrations resulting in a score greater than their respective tr_{90} were automatically flagged for human review. Additionally, observing the trends in the misalignment probability prediction over the patient treatment course can also add value in detecting anomalies. Hence, registrations resulting in a considerable jump in misalignment score compared with the adjacent treatment fractions scores (ratio

Table 1 Description of the data set used to train and test the deep learning models in the error detection algorithm

				CBCT image pairs		
Treatment region	Simulated error type	Data set partition	Number of patients	Total	Aligned	Misaligned
Thoracic/abdominal	OVBM	Training (UCLA/VCU)	374 (304/70)	1887 (1677/210)	1139 (1069/70)	748 (608/140)
		Validation (UCLA?VCU)	39 (29/10)	186 (156/30)	108 (98/10)	78 (58/20)
		Testing (UCLA/VCU)	67 (47/20)	303 (243/60)	169 (149/20)	134 (94/40)
Head and neck	10-mm shift	Training (UCLA only)	60	912	456	456
		Validation (UCLA only)	10	76	38	38
		Testing (UCLA only)	30	354	177	177
Pelvis	10-mm shift	Training (UCLA only)	60	1600	800	800
		Validation (UCLA only)	10	262	131	131
		Testing (UCLA only)	30	796	398	398

The data set, composed of images from 2 radiation therapy sites (ie, UCLA and VCU), was partitioned into a training, validation, and testing set for each treatment site based on unique patient identifiers.

Abbreviations: CBCT = cone beam computed tomography; OVBM = off-by-one vertebral body misalignment.

of predictions, $>10^4$) were also identified; among those cases, registrations with a score in the range of $tr_{99} \leq$ score $< tr_{90}$ were flagged for review.

Flagged cases were subsequently reviewed by a human expert and classified as error (true-positive) or no-error (false-positive). The reviewing process and final judgment of each case were performed by 2 clinical medical physicists with more than 15 years of experience in the field of radiation oncology. During the case reviews, a true-positive event was one that considerably deviated from the correct simCT-CBCT registration at the target ($> \sim$1 cm), leading to a gross tumor volume (GTV) undercoverage, and subsequently, a considerable deviation in the prescribed dose to the target. This definition corresponds to a level of significance that is at least reportable to the institutions' incident learning systems. The gross patient misalignment error rate in CBCT-guided radiation therapy at the 2 institutions was subsequently calculated based on the number of true-positive events found. Given that the incidents that occurred are rare, independent, and discrete, with a very large number of registration analyses, the variability in the error rate was determined using a Poisson approximation to the binomial distribution.

Additionally, for each incident found, a dosimetric analysis was performed over the corresponding treatment course to understand the dosimetric impact of the misalignment at the clinical target volume (CTV). The D95 (dose covering 95% of the target volume) values were calculated for the single misdelivered treatment fractions, as well as for the whole treatment course (accumulated D95). Each D95 value was compared with the corresponding prescribed dose in this assessment. The dosimetric analysis was performed on MIM software (MIM Software Inc).

Within the date range of this retrospective study, 3 patient setup incidents involving CBCT guidance were known at the 2 institutions. These known setup incidents were the result of off-by-one vertebral body misalignments and had been submitted to RO-ILS as part of the institutions' quality and safety protocols. To validate the real-world clinical performance of EDA, the true-positive cases found following human review were compared with the known incidents.

Results

After evaluating the models on their respective test data set containing the simulated errors, target thresholds (tr_{90} and tr_{99}) were found using an ROC analysis. For each target

Table 2 Description of the performance and target thresholds of each model using a receiver operating characteristic analysis

Model	Error type	AUC	tr_{90}	Specificity @ tr_{90}	Sensitivity @ tr_{90}	tr_{99}	Specificity @ tr_{99}	Sensitivity @ tr_{99}
Thoracic/abdominal	OVBM	99.4%	0.812	98.8%	95.0%	0.001	84.6%	99.2%
Head and neck	10-mm shift	99.6%	0.990	98.9%	90.0%	0.050	98.9%	100.0%
Pelvis	10-mm shift	99.2%	0.920	95.0%	97.0%	0.360	90.5%	100.0%

Abbreviations: AUC = area under the receiver operating characteristic curve; OVBM = off-by-one vertebral body misalignment; tr_{90} = threshold resulting in sensitivity ≥90%; tr_{99} = threshold resulting in sensitivity ≥99%.

Table 3 Summary of the artificial intelligence–assisted retrospective patient setup error search performed at the 2 radiation therapy sites

Radiation therapy site	Date range	Treatment region	Number of registrations processed by EDA	Number of registrations flagged for human review		
				Via thresholding	Via trending analysis	Total
UCLA	January 2016 to December 2017	HN	4583 (480*)	167	104	271 (117*)
		TA	3252 (580*)	250	199	449 (192*)
		PL	3912 (741*)	258	50	308 (161*)
VCU	September 2021 to September 2022	HN	1897 (172*)	12	7	19 (13*)
		TA	1860 (283*)	86	87	173 (102*)
		PL	2108 (158*)	123	19	142 (46*)
Total			17,612 (2414)	896	466	1362 (631)

Abbreviations: EDA = error detection algorithm; HN = head and neck; PL = pelvis; TA = thoracic-abdominal.
* Number of patients.

threshold, the sensitivity and specificity were obtained and are reported in Table 2.

During the retrospective study, 11,747 and 5865 registrations (from 1801 and 613 patients, respectively) were processed by the EDA from UCLA and VCU, respectively. Using the hard-thresholding method and the trending analysis method described in the Retrospective Error Search section, 1028 and 334 events (from 470 and 161 patients, respectively) were flagged from UCLA and VCU, respectively. Further details regarding the number of the flagged cases are shown in Table 3. Compared with performing a fully manual retrospective review, requiring 880 hours of human effort (assuming an average of 3 minutes per fraction), the AI-aided review only required an average of 68 hours of human effort, hence being considerably less laborious.

Among the 1362 flagged events, 7 off-by-one vertebral body misalignment incidents (true-positives) were found, as shown in Figure 2, translating to an absolute gross patient misalignment error rate of 0.04% ± 0.02% in CBCT-guided radiation therapy. Three of the 7 events (Fig. 2E-G) were the known incidents, and the other 4 events (Fig. 2A-D) were previously unreported incidents. Of those 7 cases, 5 were caught via the hard-thresholding method (Fig. 2A-D, G), and 2 were caught via the trending analysis (Fig. 2E, F).

Those 7 misdelivered fractions were from 4 individual treatment courses, with the magnitude of misalignment ranging from 1.85 cm to 2.5 cm at the clinical target volume (CTV). For each incident, the percent dose deviation of the CTV resulting from the misalignment (compared with the prescribed dose per fraction) ranged from 44% to 99%. Further details about the treatments and delivered doses (including accumulated doses) are presented in Table 4 and Appendix E1.

The rest of the flagged registrations from the 2 institutions, while often demonstrating some imperfections such as patient rotations, patient weight loss, soft tissue differences (caused by tumor growth/shrinkage or deformable organs) and substantial CBCT image artifacts, were found to not be severe enough to be reportable and be labeled as incidents. Some of those false-positive cases are highlighted in Figure E1, showing the EDA's potential for flagging cases containing not only systematic shifts from the registration, but also other clinically relevant imperfections on the registered scans.

As for the out-of-domain scans (ie, extremity cases), it was found that they constituted less than 3% of the whole data set analyzed in this study. From those cases, it was observed that lower extremity scans (including glutes, thigh, knee, and calf treatments), were generally sent to the PL model (75.7%, 22.1%, and 2.2% of the scans were sent to the PL, TA, and HN pipelines, respectively). For the upper extremity scans (including forearm, upper arm, and shoulder), 37.5%, 42.5%, and 20% of the scans were sent to the PL, TA, and HN pipelines, respectively.

Discussion

In this work, a deep learning–based patient setup EDA was used to aid in a bulk retrospective incident search for the CBCT-guided radiation therapy treatments performed between 2016 and 2017 at UCLA and between September 2021 and September 2022 at VCU. Initial model training and testing were performed on a separate data set composed of simulated errors (true-positive) and clinically performed registrations (true-negatives). An ROC analysis was performed to obtain target thresholds, which were used to flag cases for human review. A hard-thresholding method and a model prediction trending analysis of individual patients' treatment courses were used to identify cases for review.

Our results showed that CBCT-guided radiation therapy is indeed a very reliable and safe treatment modality, with an absolute gross patient misalignment error rate of 0.04% ± 0.02% at the 2 institutions. Of the 17,612 registrations analyzed by the EDA, 1362 cases were flagged and

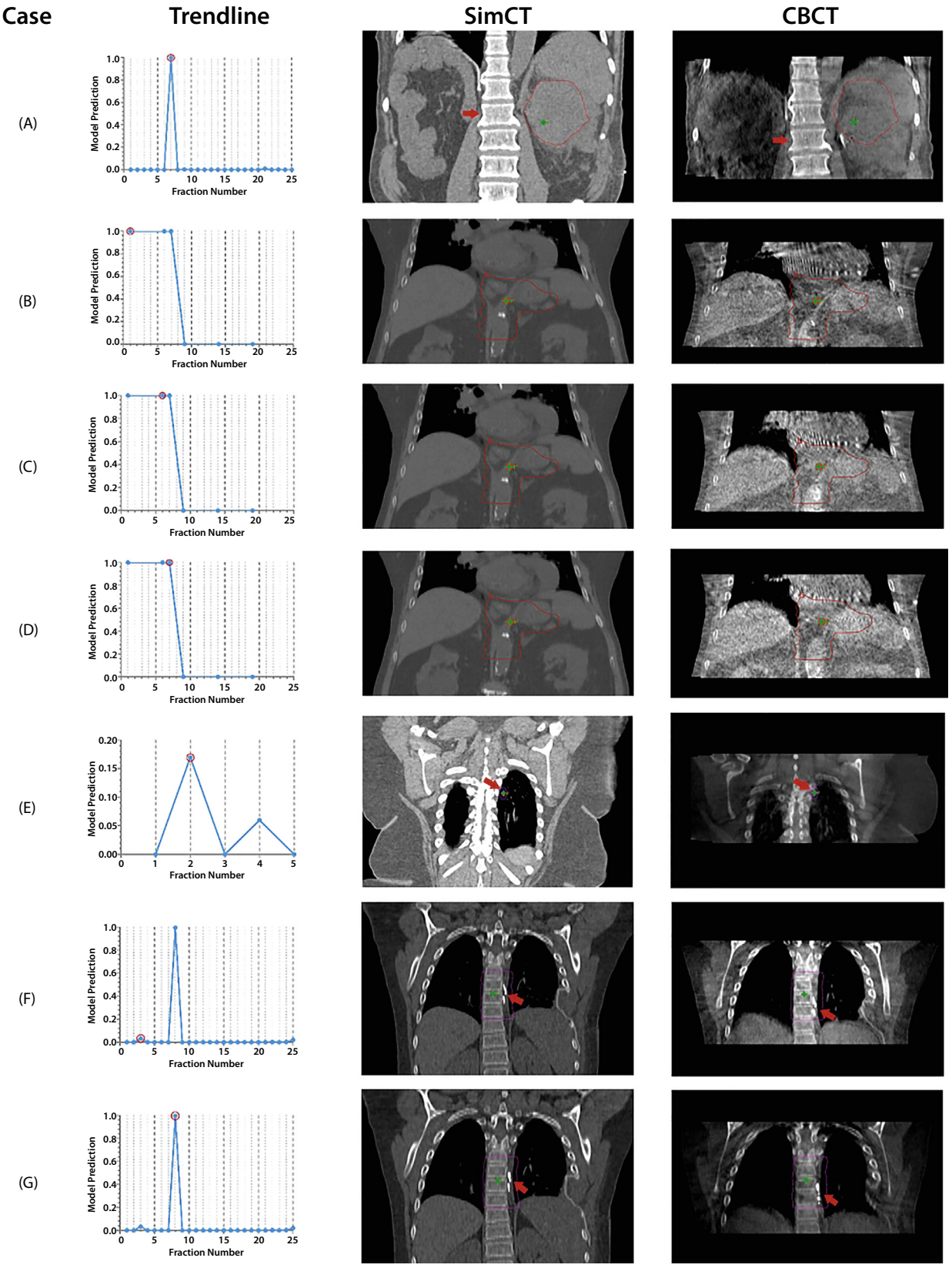

Fig. 2. The 7 incidents (A-G) found during the AI-assisted retrospective error search. For each case, the trend in the model predictions over the treatment course is shown, with each blue dot representing a treatment fraction and the incident circled in red. Additionally, selected coronal planes of the simCT and CBCT (at the corresponding slice location) are displayed for each incident. The contours present on the simCT and CBCT images represent the planning target volume used during treatment, the green star represents the treatment isocenter, and the red arrows highlight landmarks that reveal the misalignments (if present). *Abbreviations:* CBCT = cone beam computed tomography; simCT = simulation computed tomography.

Table 4 Summary of the dosimetric analysis performed on the treatments where patient misalignment incidents were found during the retrospective study

Treatment description	Number of fractions with patient misalignment	Magnitude of misalignment	Was incident(s) reported or known?	D95 for the CTV, single misaligned fraction	Accumulated D95 for the CTV, with patient misalignment(s)
Abdomen IMRT, with a dose prescription of 87.5 Gy over 25 fractions (3.5 Gy per fraction)	1	2.4 cm	No	1.35 Gy	86.5 Gy
Stomach IMRT, with a dose prescription of 45 Gy over 25 fractions (1.8 Gy per fraction)	3	2.2-2.5 cm	No	1.00 Gy	36.78 Gy*
Lung SBRT, with a dose prescription of 50 Gy over 4 fractions (12.5 Gy per fraction)	1	2.1 cm	Yes	0.13 Gy	37.50 Gy
Spine IMRT, with a dose prescription of 45 Gy in 25 fractions (1.8 Gy per fraction)	2	1.85-2.1 cm	Yes	0.22 Gy	37.0 Gy

Abbreviations: CTV = clinical target volume; D95 = dose covering 95% of the target volume; IMRT = intensity modulated radiation therapy; SBRT = stereotactic body radiation therapy.

* Patient did not complete treatment course (21/25 fractions delivered).

investigated by human experts. Seven incidents were found during the case reviews. The 3 human-reported errors which occurred at the 2 institutions were detected during this study, validating the real-world performance of the EDA in detecting gross patient setup misalignment incidents. Four additional misalignment errors found during the case reviews were previously unreported, which highlights the utility of automation in incident detection and learning within EBRT.

Following a dosimetric analysis of the incidents, it was shown that the dosimetric impact resulting from the patient misalignments was quite significant, with the dose deviations at the CTV ranging from 44% to 99% less than the prescribed doses per fraction. It can also be observed, from Figure 2B to 2D, that the off-by-one vertebral body misalignment led to a considerable dose to the heart, which could have caused serious side effects to the patient. Those observations highlight the severe harm such patient misalignment may cause to the patient, and hence, the need to minimize this failure mode within the EBRT domain. It is also alarming that, in some of the cases, the error reached the patient even though 2 imaging modalities were used for patient alignment (Appendix E1). Additionally, for the SBRT incidents, the registrations were reviewed by at least 2 individuals (including the physician as per the institution's procedure), and the error still reached the patient.

The remaining flagged registrations, while often demonstrating some imperfections, were found to not be severe enough to be considered incidents. Those included registrations showing some patient rotation, tumor shrinkage, patient weight loss or substantial bowel/bladder differences, which often resulted in an imperfect overall alignment or potentially an increase in dose to adjacent organs at risk. In some cases, those images were flagged by the treating physician for follow-up remediation (eg, improve bladder filling at the next fraction). There were also instances where materials to boost surface dose (eg, skin bolus) or shield organs-at-risk (eg, clam-shell scrotal shield) were not included in the simCT images but were inserted during treatment, resulting in a mismatch with the CBCT, which included such material. Additionally, cases containing substantial image artifacts and treatments of the extremities (eg, leg, arm, and shoulder) were regularly flagged for review. Registrations involving palliative care patients were also occasionally flagged for review and often showed imperfections in the alignment, most probably due to difficulties in setting up the patient for treatment. However, all of those cases were judged to have reasonable PTV alignment as per the institutional practices (as defined by each expert reviewer, based on their own institution's set of practices and policies) and were labeled as false positives.

Among those, there were also 3 cases (from the same 5-fraction liver treatment course) where the vertebral body alignment of the patient was off by 1, but the alignment of the PTV was found to be reasonable. This could have represented a difference in breathing phase in the right lung between the planning CT scan and the treatment, with a collapsed left lung occurring between simulation and treatment as a contributory factor. One such example is shown and described in Figure E1A.

One current limitation of the EDA is that it uses select 2-dimensional slices from the whole 3-dimensional scans during the registration analysis. The choice to use 2-dimensional slices instead of the whole 3-dimensional scans means that the tool could be easily implemented on current computer systems in the clinic, without having a sizeable memory requirement. While a 3-dimensional model would have been able to capture more features, it is currently deemed

impractical owing to its memory requirements. However, with the rise in computation technologies and easier access to high-end graphics processing units, the 3-dimensional EDA could be a more effective and practical approach in the future, compared with the 2-dimensional EDA.

As seen during the image review, another limitation of the EDA was its application to images with characteristics that were not well represented in the training data set. Such cases include significantly limited field of view CBCT scans (eg, when less than half of the thorax is present for a shoulder treatment) and extremity scans. During initial model training and testing, extremity images were not included in the data sets, as too few of those cases were present for optimal model training. In addition to being scarce, their error modes were more complicated to simulate, as misalignments might only occur along the axis of the extremity site, such as the arm or leg. A larger training data set containing more cases with extremities and additional error modes (such as combinations of rotations and translations) could benefit the EDA and reduce the false-positive rate, thereby improving the generalizability and robustness of the algorithm.

Furthermore, a heuristic model prediction-ratio approach was also applied during the trending analysis to target the 99% sensitivity threshold while minimizing excess false-positives. While this approach proved useful in the identification of errors during our retrospective study (2 of the 7 incidents detected), it may not be robust to all outliers. For example, it will fail in single-fraction treatment cases and in anomalous cases where the model prediction-ratio criteria are not met. Future studies will include ways to more robustly analyze the trends in the model predictions through AI-based clustering methods,[19,20] which may help to identify deviations in image alignment compared with patients with similar disease and treatment sites. However, the results obtained in this study demonstrate that the current trending analysis approach is acceptable and complementary to the hard-thresholding method for patient setup error detection.

While this study highlights the high reliability and low patient setup incidence rate in CBCT-guided radiation therapy, it also exposes some safety gaps present within the current workflow, with 4 of the 7 incidents observed in this study going under the radar of both the safety and quality assurance checks at 2 adequately resourced radiation therapy centers. Additionally, the risk of similar unreported or undetected incidents may be higher in underresourced radiation therapy centers, where the lack of resources may translate to a decrease in safeguards.[21] Nevertheless, the results obtained from this study still emphasize the high reliability and safety of CBCT-guided radiation therapy, and the current safety practices present throughout the workflow should be commended.

Conclusion

In EBRT, gross patient setup errors are infrequent but are considered "never-events" owing to their severe consequences. This study employed a deep learning–based patient setup EDA to facilitate a comprehensive retrospective analysis of all CBCT-guided radiation therapy treatments administered at UCLA between 2016 and 2017 and at VCU between 2021 and 2022. Out of the 17,612 CBCT registrations (from 2414 patients) assessed by EDA, 1362 (from 631 patients) were flagged as potential incidents and subsequently reviewed. Following a thorough investigation of the flagged cases, 7 incidents involving patient setup errors were identified. All 3 reported errors that occurred at the 2 institutions were found during this study, validating the real-world performance of the EDA in detecting gross patient setup misalignments. The other 4 incidents observed during the case reviews were found to be previously unreported errors. While the results obtained highlight the reliability and safety of CBCT-guided radiation therapy (with an absolute gross patient misalignment incidence rate of 0.04% ± 0.02% at the 2 institutions), the incidents that occurred also expose safety gaps still present within the patient alignment process.

References

1. ASTRO. RO-ILS aggregate report Q1 2023. Accessed June 20, 2023. https://www.astro.org/Patient-Care-and-Research/Patient-Safety/RO-ILS/RO-ILS-Education.
2. Doroszczuk B, Bardet MC, Covard F, Javay O. ASN Report on the state of nuclear safety and radiation protection in France in 2021 + abstracts (INIS-FR—22-0646). Accessed September 4, 2024. https://www.french-nuclear-safety.fr/asn-informs/publications/asn-s-annual-reports/asn-report-on-the-state-of-nuclear-safety-and-radiation-protection-in-france-in-2021.
3. Smith S, Wallis A, King O, et al. Quality management in radiation therapy: A 15 year review of incident reporting in two integrated cancer centres. *Tech Innov Patient Support Radiat Oncol* 2020;14:15-20.
4. Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. *Med Phys* 2016;43:4209-4262.
5. Ezzell G, Chera B, Dicker A, et al. Common error pathways seen in the RO-ILS data that demonstrate opportunities for improving treatment safety. *Pract Radiat Oncol* 2018;8:123-132.
6. Hoopes DJ, Ford EC, Ezzell GA, Dicker AP, Chera BS, Potters L. Incident learning for stereotactic radiation therapy from RO-ILS: Radiation oncology incident learning system. *Int J Radiat Oncol Biol Phys* 2017;99:S46-S47.
7. McGurk R, Naheedy KW, Kosak T, et al. Multi-institutional stereotactic body radiation therapy incident learning: Evaluation of safety barriers using a human factors analysis and classification system. *J Patient Saf* 2023;19:e18-e24.
8. Luh JY, Albuquerque KV, Cheng C, et al. ACR–ASTRO practice parameter for image-guided radiation therapy (IGRT). *Am J Clin Oncol* 2020;43:459-468.
9. Ford E, Conroy L, Dong L, et al. Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275. *Med Phys* 2020;47:e236-e272.
10. Xia P, Sintay BJ, Colussi VC, et al. Medical physics practice guideline (MPPG) 11.a: Plan and chart review in external beam radiotherapy and brachytherapy. *J Appl Clin Med Phys* 2021;22:4-19.
11. Petragallo R, Bertram P, Halvorsen P, Iftimia I, Low DA, Morin O, Narayanasamy G, Saenz DL, Sukumar KN, Valdes G, Weinstein L. Development and multi–institutional validation of a convolutional

neural network to detect vertebral body mis−alignments in 2D x−ray setup images. *Med Phys* 2023;50:2662-2671.

12. Luximon DC, Ritter T, Fields E, Neylon J, Petragallo R, Abdulkadir Y, Charters J, Low DA, Lamb JM. Development and interinstitutional validation of an automatic vertebral−body misalignment error detector for cone−beam CT−guided radiotherapy. *Med Phys* 2022;49:6410-6423.
13. Neylon J, Luximon DC, Ritter T, Lamb JM. Proof−of−concept study of artificial intelligence−assisted review of CBCT image guidance. *J Appl Clin Med Phys* 2023;24 e14016.
14. Hadjiiski L, Cha K, Chan HP, et al. AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging. *Med Phys* 2023;50:e1-e24.
15. French Nuclear Safety Authority. Patient safety: Paving the way for progress. Patient repositioning imaging: Vertebra identification error. Accessed December 8, 2023. https://www.french-nuclear-safety.fr/Media/Files/00-Publications/Patient-safety-12.-Patient-repositioning-imaging-vertebra-identification-error.
16. Luximon DC, Neylon J, Lamb JM. Feasibility of a deep-learning based anatomical region labeling tool for Cone-Beam Computed Tomography scans in radiotherapy. *Phys Imaging Radiat Oncol* 2023;25:100427.
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Conference on Computer Vision and Pattern Recognition (CVPR) and IEEE; 2017:4700-4708.
18. ROSAML Reserach Group. Accessed April 9, 2024. www.https://rosaml.net/.
19. Karim MR, Beyan O, Zappa A, et al. Deep learning-based clustering approaches for bioinformatics. *Brief Bioinform* 2021;22:393-415.
20. Arunkumar N, Mohammed MA, Abd Ghani MK, et al. K-means clustering and neural network for object detecting and identifying abnormality of brain tumor. *Soft Computing* 2019;23:9083-9096.
21. Agarwal JP, Krishnatry R, Panda G, et al. An audit for radiotherapy planning and treatment errors from a low−middle-income country centre. *Clin Oncol* 2019;31:e67-e74.