

UCLA

UCLA Previously Published Works

Title

Feasibility of a deep-learning based anatomical region labeling tool for Cone-Beam Computed Tomography scans in radiotherapy

Permalink

<https://escholarship.org/uc/item/0159j1xh>

Authors

Luximon, Dishane C

Neylon, John

Lamb, James M

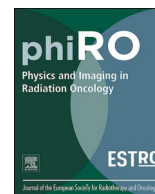
Publication Date

2023

DOI

10.1016/j.phro.2023.100427

Peer reviewed



Original Research Article

Feasibility of a deep-learning based anatomical region labeling tool for Cone-Beam Computed Tomography scans in radiotherapy

Dishane C Luximon^{*}, John Neylon, James M Lamb

Department of Radiation Oncology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA



ARTICLE INFO

Keywords:

Deep learning
Cone-beam computed tomography
Anatomy labeling
Radiotherapy

ABSTRACT

Background and purpose: Currently, there is no robust indicator within the Cone-Beam Computed Tomography (CBCT) DICOM headers as to which anatomical region is present on the scan. This can be a predicament to CBCT-based algorithms trained on specific body regions, such as auto-segmentation and radiomics tools used in the radiotherapy workflow. We propose an anatomical region labeling (ARL) algorithm to classify CBCT scans into four distinct regions: head & neck, thoracic-abdominal, pelvis, and extremity.

Materials and methods: Algorithm training and testing was performed on 3,802 CBCT scans from 596 patients treated at our radiotherapy center. The ARL model, which consists of a convolutional neural network, makes use of a single CBCT coronal slice to output a probability of occurrence for each of the four classes. ARL was evaluated on the test dataset composed of 1,090 scans and compared to a support vector machine (SVM) model. ARL was also used to label CBCT treatment scans for 22 consecutive days as part of a proof-of-concept implementation. A validation study was performed on the first 100 unique patient scans to evaluate the functionality of the tool in the clinical setting.

Results: ARL achieved an overall accuracy of 99.2% on the test dataset, outperforming the SVM (91.5% accuracy). Our validation study has shown strong agreement between the human annotations and ARL predictions, with accuracies of 99.0% for all four regions.

Conclusion: The high classification accuracy demonstrated by ARL suggests that it may be employed as a pre-processing step for site-specific, CBCT-based radiotherapy tools.

1. Introduction

Cone-beam computed tomography (CBCT) is commonly used for radiotherapy image guidance because it facilitates accurate and precise positioning and alignment of the patient. In real-time adaptive radiotherapy, the CBCT may also be used to adapt the treatment plan based on the new target location and size, and the position of organs at risk (OARs). In this case, the delineation of the target(s) and OARs may be required on the CBCT scan prior to the plan adjustment [1]. With the rise of machine learning and deep learning techniques in the field of medical image analysis, many algorithms are being developed to automate and expedite this delineation process [2–5]. Similarly, algorithms have been proposed for the detection of setup errors [6–7] and for early treatment response assessment using CBCT scans [8]. However, these algorithms are typically anatomical region-specific and assume the presence of the organs-of-interest irrespective of the body region inputted to the algorithm.

The recognition of the global body region may be useful as a pre-processing step for these tools, such that they are applied to body regions within their domain. However, this step is often neglected due to the assumption that the anatomy information is present on the Digital Imaging and Communications in Medicine (DICOM) headers. While a ‘Body Part Examined’ tag is indeed present in the DICOM headers of the planning Computed Tomography (CT), it has been shown that this information is not very reliable, with a mis-labeling rate of 15.3% [9]. Furthermore, these pre-defined labels are driven by the acquisition protocol. Due to the variability and differences among the patients’ anatomies, an imaging protocol for a different body region may be used by the clinical personnel in order to obtain better image quality. While the header can be adjusted following the CT acquisition, this is not commonly done in the clinic, which may lead to a wrong body region label [10]. Additionally, this ‘Body Part Examined’ tag may be completely absent in the CBCT DICOM headers, as is the case at our institution, highlighting the need for an automatic region-labeling

^{*} Corresponding author at: 765 Weyburn Place, Apt 301, Los Angeles, CA 90024, United States.

E-mail address: dluximon@mednet.ucla.edu (D.C. Luximon).

<https://doi.org/10.1016/j.phro.2023.100427>

Received 12 December 2022; Received in revised form 21 February 2023; Accepted 28 February 2023

Available online 5 March 2023

2405-6316/© 2023 The Author(s). Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

algorithm to recognize the global patient anatomy and treatment region.

Several algorithms have recently been proposed for the classification of anatomical regions in CT and MRI scans [11–14]. Among those, Ouyang et al. [14] achieved the highest classification accuracy of 97.3% on their test dataset composed of 663 CT scans. These previous studies showcase the potential of deep learning techniques on such region labeling problem. Nevertheless, if these techniques are used as a pre-processing step for other clinical tools, which have their intrinsic error rate, it is imperative to minimize the pre-processing error rate as much as possible to improve the reliability of the labeling tool and reduce the overall algorithm's failure rate. Hence, it is vital to continuously identify and address limitations of such region labeling tools.

One common characteristic and limitation of the previous studies is that they have all been developed and tested on CT and MR images, which typically have improved image quality as compared to pre-treatment CBCT images [18–19]. Hence, classifying CBCT images may become a challenge as fewer useful features and more artifacts may be present on the CBCT scan for accurate region labeling. CBCT scans also have a small field-of-view (FOV), which is usually restricted to the treatment region only, making a consecutive body part recognition algorithm as in [14] more complicated.

To address this current limitation, we propose a CNN-based anatomical region labeling (ARL) tool which can classify a CBCT scan into four global regions, namely head & neck (HN), thoracic-abdominal (TA), pelvis (PL) and extremity (EX) using a single coronal slice from the CBCT volume. To the best of our knowledge, this will be the first region labeling algorithm built specifically for pre-treatment CBCT scans.

2. Materials and methods

2.1. Dataset for model training and testing

Under an IRB approved protocol (UID 18–001430), CBCTs were collected from 631 patients undergoing radiotherapy treatment at the University of California, Los Angeles Medical Center (UCLA) between January 2017 and April 2022. The dataset collection was performed using an in-house DICOM query and retrieval (DQR) application programming interface using the `pynetdicom`¹ Python package. The treatments at UCLA had been performed on three TrueBeams and one NovalisTx linear accelerator treatment machines (Varian Medical Systems, California, United States). CBCT scans were acquired using the on-board imager of each machine. For each CBCT, the corresponding planning CT, REG file and RTStruct file were also collected and used during the image pre-processing step in our implementation.

A visual inspection of the treatment isocenter was performed to sort the CBCT scans into four different global regions: head & neck (HN), thoracic-abdominal (TA), pelvis (PL), and extremity (EX). The C7 vertebral body was used as a limit to the HN region such that the CBCT scan only contained these two body parts, as shown in [Supplementary Figure S1](#). However, in the clinical setting, it is possible to have neck scans containing part of the thorax. For the first part of our experiment, which included model training and testing, these scans with substantial overlapping regions were withdrawn from our dataset to maintain the distinction between each category. Following the triage, 3802 CBCT scans from 596 patients remained, as described in [Supplementary Table S1](#). The limits of the TA region were the T1 vertebra and the L2 vertebra, avoiding the neck and pelvis regions. For the PL scans, the L3 and S2 vertebra were used as markers, avoiding the abdominal region and area below the pubic symphysis. Scans of the arms, legs and extremity of the shoulder were placed in the EX dataset.

Each of the four datasets was then separately and randomly split into a training, validation and test set using a 60:10:30 ratio. As scans from multiple treatment fractions were used in our study, the dataset split was

performed based on the patients' unique anonymized identifiers to avoid having scans from the same patient overlapping across the training, validation, and test sets.

2.2. Image Pre-processing

The pixel spacing of the CBCT scans ranged from 0.51 to 1.17 mm and the slice thickness from 1 to 2.5 mm. The CBCT scans were resampled based on their corresponding planning CT to produce uniform images with a voxel spacing of $1 \times 1 \times 1.5 \text{ mm}^3$. In our pipeline, this resampling, and volume matching was performed using the REG file present with the CT-CBCT pair. However, the CBCT resampling can be made independently from the planning CT and REG file for another application of the ARL. Furthermore, the treatment couch and immobilization devices were removed from the CBCT image using the body contour present in the RTStruct file. In the event that a body contour is not present in the RTStruct file, a thresholding method, including a morphological dilation followed by erosion, was used to extract the body contour from the CBCT. The dilation and erosion operations used 20×20 and 5×5 -pixel² rectangular structuring element, respectively.

A coronal slice was then extracted from each CBCT present in our dataset and each image was labeled using their corresponding global region. The primary coronal slice was extracted by locating the CBCT slice with the highest mean Hounsfield Unit (HU). This slice-selection method was chosen such that the coronal slice would cover the whole extent of the patient scan while containing considerable bony structures (higher HU) which can be useful features in the recognition of the anatomical region.

For training purposes, two additional slices were extracted from the CBCT scans in the training and validation datasets, with each slice being 10 pixels away from the primary coronal slice location; one being 10 pixels in the anterior direction and the other being 10 pixels in the posterior direction. The extraction of these two extra slices were performed as an augmentation method during the model training due to the inter-patient variability in anatomy which can be present on the primary coronal slice. The slices were then cropped about the center of the patient body to reduce empty spaces around the body and obtain 150×400 pixel² images, as shown in [Fig. 1](#), and used as input to our ARL model.

2.3. Anatomical region labeling (ARL) model

The ARL model used the Dense-Net architecture as shown in [Supplementary Figure S2](#). The ARL model makes use of densely contracting paths to capture contextual information from the CBCT coronal image before outputting a probability of occurrence for each of the four classes. The Dense Block in our architecture constitutes of two densely connected layers, each comprising of seven layers. The two densely connected layers in the Dense Block were connected to each other in a feed-forward mode to maximize feature reuse, which been shown to be computationally efficient, hence allowing a deeper network [20].

2.4. Training configuration

The ARL model was implemented using Tensorflow 2.2 with Keras backend. Model training was performed using the Adam Optimizer [15], with a starting learning rate of 2×10^{-5} . During training, the model was evaluated on the validation dataset after each epoch, and a learning rate reducer (0.75) was applied if the validation loss did not decrease for 15 consecutive epochs. To avoid overfitting the model on the training dataset, an early stopping method was applied such that training would stop if the validation accuracy did not improve for 50 consecutive epochs, or for a maximum of 400 training epochs. For comparison purposes, we also trained a Support Vector Machine (SVM) [22], and saved the parameters which produced the highest accuracy on the validation set.

¹ <https://pydicom.github.io/pynetdicom/stable/#>.

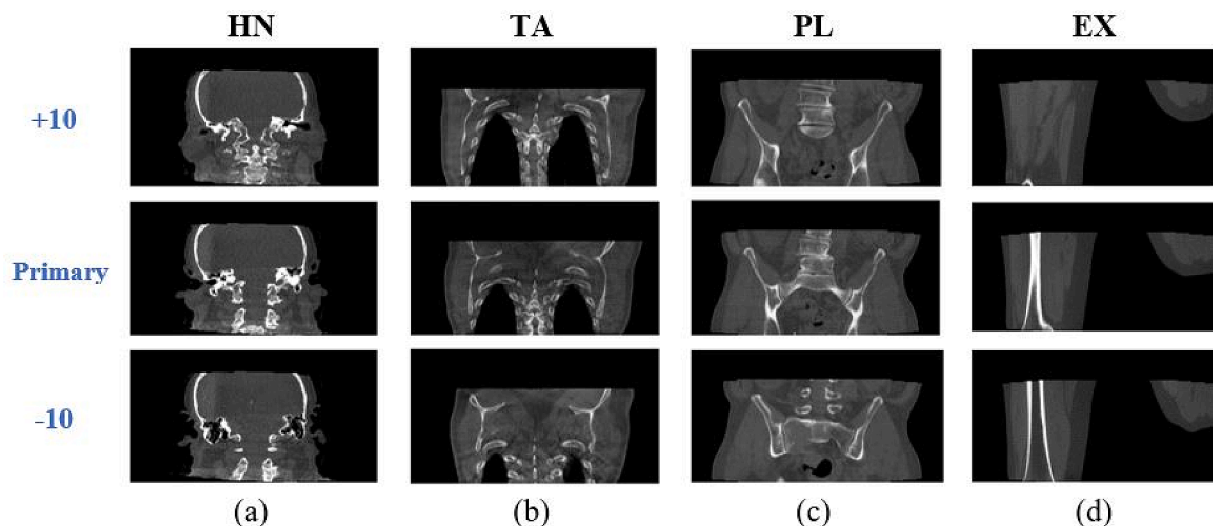


Fig. 1. Twelve coronal slices which were used as input to the ARL model during algorithm training. Each column (a-d) shows the three slices extracted from four different CBCT scans, one from each anatomical region. The first row represents the slices extracted 10 pixels away from the primary coronal slice location in the anterior direction. The second row show the slices which are extracted at the primary coronal slice location, and the third row represents the slices extracted 10 pixels away from the primary coronal slice location in the posterior direction. HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity.

2.5. Evaluation metrics and quality control

After the trained ARL model and SVM was applied to the test dataset, the true positive (tp), false positive (fp), false negative (fn), and true negative (tn) counts were obtained for each anatomical region. Subsequently, the four metrics shown in Equations 1–4 were used to evaluate and compare the performance of our models.

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$F\text{-1 Score} = \frac{2tp}{2tp + fp + fn} \quad (2)$$

$$\text{Precision} = \frac{tp}{tp + fp} \quad (3)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (4)$$

To obtain visual explanations of the model's prediction, the Gradient-weighted Class Activation Mapping (Grad-CAM) [16] was implemented. This Grad-CAM method uses the gradients from the final convolutional layer from the ARL model to produce a heat map describing the regions which contributed the most to the activation of the predicted anatomical region.

2.6. Clinical implementation and validation

Using our in-house DQR system to interface with the ARIA system (Varian Medical Systems, Palo Alto, CA), the ARL was implemented at our clinic to automatically classify incoming CBCT data on a daily basis for 22 consecutive days between August and September 2022 as part of a pilot process for an automated weekly chart check image analysis [17]. For validation purposes, the predictions for the first 100 unique patients were compared to a human perspective. Without any information about the predictions of the ARL, each of the 100 unique scans was visually analyzed and labeled by a human observer to obtain the ground truth label.

However, in contrast to the dataset used during model training and testing this validation dataset did not exclude scans containing overlapping regions, such as neck and thorax, or abdomen and pelvis. Hence, the ground truth labels were obtained by identifying the dominant

region (i.e. the region encompassing the majority of the CBCT scan). Furthermore, the other less pronounced region(s), if present, was noted as a 'less-pronounced region'. For example, for a neck treatment scan containing mostly the neck and part of the thorax, the region would be labeled as HN, with the 'less-pronounced region(s)' being TA. Following the human annotations, the predictions from the ARL were compared with their respective ground truth labels and the model performance was evaluated.

3. Results

3.1. Model training and evaluation

During the algorithm training, the ARL model achieved convergence after 49 epochs with training and validation accuracies of 99.8% and 99.3%, respectively. Following the testing phase, the ARL model resulted in 9 misclassifications out of the 1,090 test cases, for an overall accuracy of 99.2%. Selected true-positives and misclassifications are shown in Fig. 2 and Fig. 3, respectively.

For the SVM, a polynomial kernel was found to produce the best fit, with training and validation accuracies of 96.0%. Following testing, the SVM obtained an overall accuracy of 91.5%. Using Student paired t-tests, results from the ARL model and SVM were found to be statistically significant ($p\text{-value} < 0.0001$). The detailed results obtained from the ARL model and SVM are reported in Table 1.

3.2. Validation of the proof-of-concept implementation

During 22 consecutive treatment days between August and September 2022, 798 patient scans were processed and classified by the ARL algorithm. The validation dataset was composed of the first 100 unique patient scans, which were labeled by a human observer, and described in Supplementary Table S2.

The ARL prediction for each of the 100 cases was compared to its respective ground truth label (dominant region), and the results of this validation study are reported in Table 2. Out of the 100 individual cases, two cases had an ARL prediction-ground truth mismatch. However, it was found that each of these two cases had overlapping regions present on the CBCT scan, and the ARL prediction matched with the referenced less-pronounced regions, as shown in Fig. 4.

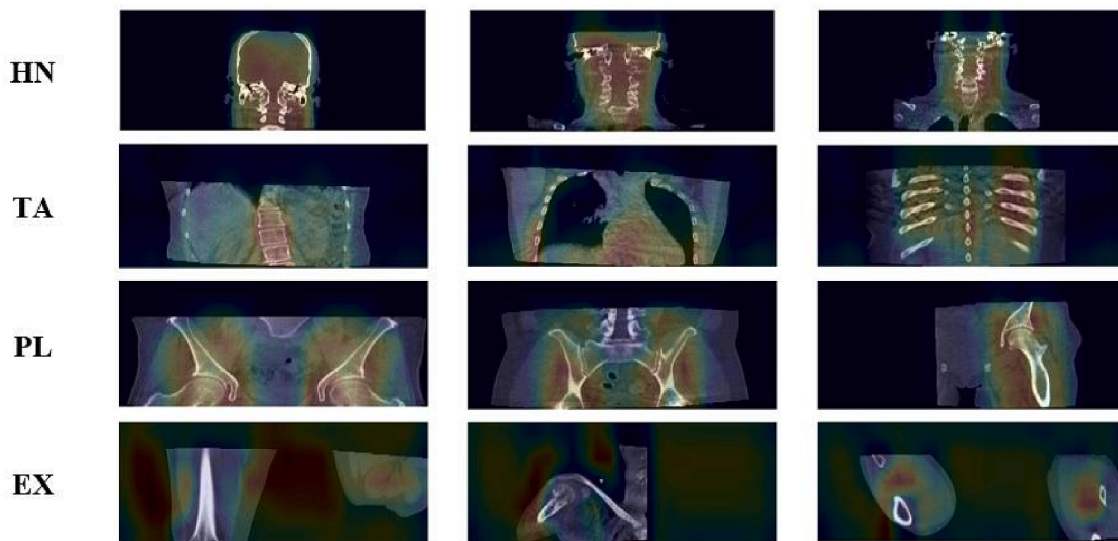


Fig. 2. 12 selected CBCT slices (from unique patients) which were inputted to the ARL model and resulted in true-positives. The Grad-CAM activation heat map is overlaid on the CBCT image to display the regions which had the greatest weight in the prediction. The red areas mean the region contributed more to the prediction. HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

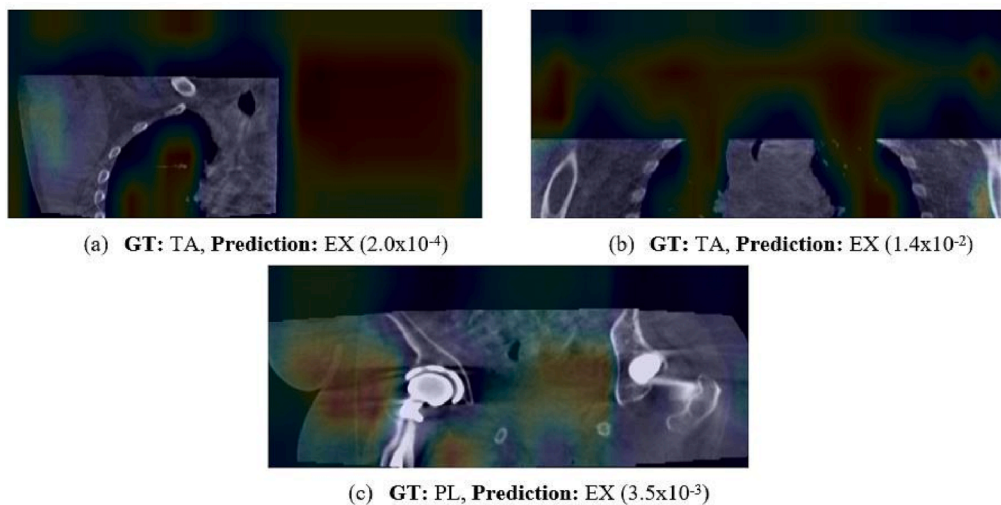


Fig. 3. Coronal slices of three selected misclassified cases, with their corresponding activation map overlaid on top. The red area signify higher weight in the model decision for the predicted area. The output probability of the model class prediction is also shown for each case. GT: Ground Truth; HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Performance of the Anatomical Region Labeling (ARL) model and the Support Vector Machine (SVM) on the 1,090 test cases. The results are shown for each three global regions separately. Bold texts represent the better result between the two models.

	HN		TA		PL		EX	
	ARL	SVM	ARL	SVM	ARL	SVM	ARL	SVM
Accuracy	99.9%	97.9%	99.4%	92.8%	99.6%	95.3%	99.4%	97.0%
F-1 Score	99.8%	96.4%	98.9%	86.2%	99.4%	93.1%	97.1%	85.3%
Precision	100.0%	96.3%	99.0%	94.3%	100.0%	90.3%	94.3%	76.8%
Recall	99.7%	96.6%	98.7%	79.4%	98.9%	96.1%	100.0%	96.0%

HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity, ARL: anatomical region labeling model, SVM: support vector machine.

4. Discussion

The ARL model presented in this study has shown high classification ability for each of the four global regions (HN, TA, PL and EX), with

accuracies of 99.9%, 99.4%, 99.6%, and 99.4% respectively, outperforming the SVM model in all four regions. As compared to other CNN-based anatomy recognition algorithms developed by Roth et al. and Ouyang et al., which achieved the highest reported accuracies in the

Table 2

Performance of the Anatomical Region Labeling (ARL) model on the 100 cases used for clinical validation. The results are shown for each three global regions separately.

	HN	TA	PL	EX
Accuracy	99.0%	99.0%	99.0%	99.0%
F-1 Score	98.8%	98.2%	98.3%	66.7%
Precision	97.6%	100.0%	96.7%	100.0%
Recall	100.0%	96.4%	100.0%	50.0%

HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity.

literature (94.1% and 97.3%, respectively) [12,14], our ARL resulted in a better performance with an overall accuracy of 99.2%. However, a direct comparison between those methods is not the primary aim of this study as different imaging modalities, number of classes, and imaging planes have been used in each method. Nevertheless, the high classification accuracy produced by the ARL model demonstrates the feasibility of applying such deep learning tool to pre-treatment CBCT scans to identify the global anatomical region.

Fig. 2 shows the input coronal slices of 12 true-positive cases with the Grad-CAM activation heat map of the ARL model overlaid on the CBCT slice. It can be observed that the regions which activated the model are in the vicinity of the craniovertebral junction for HN cases, the spine, abdominal organs and the ribs for the TA cases, and the pelvic bones for PL cases. As for the extremity cases, the model was activated by the empty regions around the patient anatomy. While this may not be the most logical and robust way of identifying extremity cases to a human observer, this feature is a characteristic of most extremity scans. However, it must be noted that the amount of extremity cases in the training dataset was limited, which may be the source of the decrease in performance for EX classification.

Out of the 1,090 scans, 9 scans were wrongly classified by the ARL model. Fig. 3 illustrates some misclassified cases, with their corresponding activation maps overlaid on top. It can be observed from Fig. 3 (a) and 3(b) that the limited FOV resulted in a wrong classification of the thorax as an extremity due to the empty spaces around the patient. On Fig. 3(c), the presence of metal artifacts may have been the cause of the misclassification as shown by the heatmap. A potential solution would be to use activation gates [21] within the ARL model such that it focuses on targeted regions instead of irrelevant regions on the image.

Nevertheless our proof-of-concept implementation and validation study have shown that the ARL predictions correlate with the human observer annotations, with accuracies of 99.0% for all four global regions. Out of the 100 cases, two cases had an ARL prediction-dominant

region mismatch, as shown in Fig. 4. However, it can be observed that the ARL prediction was still consistent with the overlapping region present on the scan in both cases. The results of this validation study hence reinforces the relevance and ability of the ARL tool to label CBCT images from daily treatments.

To be more robust to the entire patient population, the current algorithm could be further refined to accommodate for outlier cases, such as extremity treatment scans. However, these types of CBCT scans are seen more sporadically in the clinical setting due to the rare occurrence of soft tissue sarcomas [23], leading to too few cases for optimal model training or refinement. Furthermore, the ARL was trained and tested on only a single institution's data. To validate and improve the generalizability of the ARL on other facilities' datasets, a multi-institutional study needs to be performed, which will be part of our future studies.

Another limitation of the current ARL is that it uses a single 2D coronal slice which contains limited anatomical information as compared to the whole 3D image. A 3D Dense-Net [24] may potentially improve the performance of the ARL by obtaining more useful features as compared to the current 2D model [25]. However, training a 3D CNN is computationally expensive and the inference time of the tool will be higher with our current system. With the increased availability of high performance Graphics Processing Units, this 3D method may be feasible in the future.

In this work, a CNN-based Anatomical Region Labeling (ARL) tool was developed to classify pre-treatment CBCT scans into four regions, namely head & neck, thoracic-abdominal, pelvis, and extremity. Our results have shown strong agreement between the model predictions and human annotations for all four regions, confirming the strong performance of the model. The ARL algorithm may be employed in the clinical setting as a pre-processing step for radiotherapy tools which have been developed for pre-treatment CBCTs containing specific anatomical regions, such as auto-segmentation algorithms, patient setup error detection algorithms, and radiomics tools for early treatment response assessment. Furthermore, the tool may be used as a quality assurance check by comparing the model's prediction to the treatment site to avoid wrong-site radiotherapy treatment.

Open-Source Code Access.

The python scripts for the DQR and ARL algorithm are available on the following website: https://github.com/dcluximon/ARL_repo.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

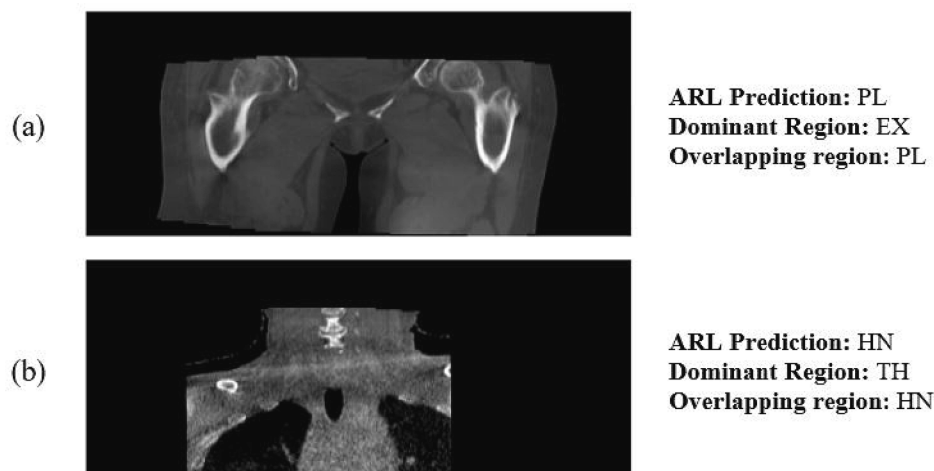


Fig. 4. Coronal slices of the two misclassified cases in the clinical validation, with the ARL prediction and human annotations (dominant region and overlapping region) reported. HN: Head & Neck, TA: Thoracic-abdominal, PL: Pelvis, EX: Extremity.

the work reported in this paper.

Acknowledgement

The research reported in this study was supported by the Agency for Healthcare Research and Quality (AHRQ) under award number 1R01HS026486.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.phro.2023.100427>.

References

- [1] Posiewnik M, Piotrowski T. A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer. *Phys Med* 2019;59:13–21. <https://doi.org/10.1016/j.ejmp.2019.02.014>.
- [2] Fu Y, Lei Y, Wang T, Tian S, Patel P, Jani AB, et al. Pelvic multi-organ segmentation on cone-beam CT for prostate adaptive radiotherapy. *Med Phys* 2020;47(8):3415–22. <https://doi.org/10.1002/mp.14196>.
- [3] Dai X, Lei Y, Wang T, Dhabaan AH, McDonald M, Beitler JJ, et al. Head-and-neck organs-at-risk auto-delineation using dual pyramid networks for CBCT-guided adaptive radiotherapy. *Phys Med Biol* 2021;66(4):045021. <https://doi.org/10.1088/1361-6560/abd953>.
- [4] Dai X, Lei Y, Wynne J, Janopaul-Naylor J, Wang T, Roper J, et al. Synthetic CT-aided multiorgan segmentation for CBCT-guided adaptive pancreatic radiotherapy. *Med Phys* 2021;48(11):7063–73. <https://doi.org/10.1002/mp.15264>.
- [5] Moazzezi M, Rose B, Kisling K, Moore KL, Ray X. Prospects for daily online adaptive radiotherapy via ethos for prostate cancer patients without nodal involvement using unedited CBCT auto-segmentation. *J Appl Clin Med Phys* 2021;22(10):82–93. <https://doi.org/10.1002/acm2.13399>.
- [6] Jani SS, Low DA, Lamb JM. Automatic detection of patient identification and positioning errors in radiation therapy treatment using 3-dimensional setup images. *Pract Radiat Oncol* 2015;5(5):304–11. <https://doi.org/10.1016/j.prro.2015.06.004>.
- [7] Luximon DC, Ritter T, Fields E, Neylon J, Petragallo R, Abdulkadir Y, et al. Development and interinstitutional validation of an automatic vertebral-body misalignment error detector for cone-beam CT-guided radiotherapy. *Med Phys* 2022;49(10):6410–23. <https://doi.org/10.1002/mp.15927>.
- [8] Shi L, Rong Y, Daly M, Dyer B, Benedict S, Qiu J, et al. Cone-beam computed tomography-based delta-radiomics for early response assessment in radiotherapy for locally advanced lung cancer. *Phys Med Biol* 2020;65(1):015009. <https://doi.org/10.1088/1361-6560/ab3247>.
- [9] Gueld MO, Kohnen M, Keyzers D, Schubert H, Wein BB, Bredno J, et al. Quality of DICOM header information for image categorization in medical imaging 2002. *SPIE* 2002;4685:280–7. <https://doi.org/10.1117/12.467017>.
- [10] Samara ET, Fitousi N, Bosmans H. Quality assurance of dose management systems. *Phys Med* 2022;99:10–5. <https://doi.org/10.1016/j.ejmp.2022.05.002>.
- [11] Wada Y, Morishita J, Yoon Y, Okumura M, Ikeda N. A simple method for the automatic classification of body parts and detection of implanted metal using postmortem computed tomography scout view. *Radiol Phys Technol* 2020;13:378–84. <https://doi.org/10.1007/s12194-020-00581-4>.
- [12] Roth HR, Lee CT, Shin HC, Seff A, Kim L, Yao J, et al. Anatomy-specific classification of medical images using deep convolutional nets. *IEEE* 2015:101–4. <https://doi.org/10.1109/ISBI.2015.7163826>.
- [13] Lee H, Huang C, Yune S, Tajmir SH, Kim M, Do S. Machine friendly machine learning: interpretation of computed tomography without image reconstruction. *Sci Rep* 2019;9(1):15540. <https://doi.org/10.1038/s41598-019-51779-5>.
- [14] Ouyang Z, Zhang P, Pan W, Li Q. Deep learning-based body part recognition algorithm for three-dimensional medical images. *Med Phys* 2022;49(5):3067–79. <https://doi.org/10.1002/mp.15536>.
- [15] Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* 2014. <https://doi.org/10.48550/arXiv.1412.6980>.
- [16] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *IEEE* 2017:618–26. <https://doi.org/10.1007/s11263-019-01228-7>.
- [17] Neylon J, Luximon DC, Ritter T, Lamb JM. Proof-of-Concept Study of Artificial Intelligence-Assisted Review of CBCT Image Guidance. Unpublished results; in review at *J Appl Clin Med Phys*. 2023.
- [18] Lechuga L, Weidlich GA. Cone beam CT vs. fan beam CT: a comparison of image quality and dose delivered between two differing CT imaging modalities. *Cureus* 2016;8(9). <https://doi.org/10.7759/cureus.778>.
- [19] Dubec M, Brown S, Chuter R, Hales R, Whiteside L, Rodgers J, et al. MRI and CBCT for lymph node identification and registration in patients with NSCLC undergoing radical radiotherapy. *Radiother Oncol* 2021;159:112–8. <https://doi.org/10.1016/j.radonc.2021.03.015>.
- [20] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *IEEE* 2017:4700–5478. <https://doi.org/10.48550/arXiv.1608.06993>.
- [21] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. *Med Image Anal* 2019;53:197–207. <https://doi.org/10.1016/j.media.2019.01.012>.
- [22] Noble WS. What is a support vector machine? *Nat biotechnol* 2006;24(12):1565–7. <https://doi.org/10.1038/nbt1206-1565>.
- [23] Lahat G, Lazar A, Lev D. Sarcoma epidemiology and etiology: potential environmental and genetic factors. *Surg Clin N Am* 2008;88(3):451–81. <https://doi.org/10.1016/j.suc.2008.03.006>.
- [24] Uemura T, Näppi JJ, Hironaka T, Kim H, Yoshida H. Comparative performance of 3D-DenseNet, 3D-ResNet, and 3D-VGG models in polyp detection for CT colonography. *SPIE* 2020;11314:736–41. <https://doi.org/10.1117/12.2549103>.
- [25] Yu J, Yang B, Wang J, Leader J, Wilson D, Pu J. 2D CNN versus 3D CNN for false-positive reduction in lung cancer screening. *J of Med Imaging* 2020;7(5):051202. <https://doi.org/10.1117/1.jmi.7.5.051202>.