A Comparison of Convolutional Neural Networks and Logistic Regression for the Detection of Vertebral Body Misalignments During Radiation Therapy

R. Petragallo [1], J. Lamb[1]

(1) University of California, Los Angeles

JULY 12–16 2020 VIRTUAL
Improving Health QUALITY. Increasing Global IMPACT
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

OBJECTIVES

Despite advances in modern image guidance technology, alignment to the wrong vertebral level remains a rare but serious error in radiotherapy delivery. Our work aims to develop a software failsafe system to mitigate the human failure error mode. Here we compare image classification results from two convolutional neural network (CNN)-based approaches and a logistic regression model for detection of translational shifts of one vertebral body away from the target.

BACKGROUND

Many radiation oncology patients at our clinic receive daily image guidance prior to each individual fraction of radiation. The current clinical approach relies on a radiation therapist comparing these two images and making a judgement on whether the patient must be repositioned before delivering the radiation. While clinically significant errors in radiation oncology are rare, when they do occur they have the potential to cause serious harm or death to the patient. Misalignment by one vertebral body is just one example of a positioning error that can be difficult to catch manually, since adjacent vertebrae appear similar on imaging of bony anatomy. In this work we show that our purpose-built deep learning model can be trained to detect errors in patient positioning by one vertebral body with higher accuracy than both a transfer learning-based approach and traditional logistic regression model. This lays the groundwork for our eventual goal of incorporating this technology into the clinical workflow as a software failsafe on human error.

METHODS

Vertebral Body Images: A dataset was developed using x-ray and digitally reconstructed radiograph (DRR) image pairs from 71 consecutive thoracic spine patients treated at our institution using a stereoscopic onboard image guidance system. The unshifted DRRs were obtained by searching UCLA's electronic medical record for day-of x-ray imaging taken after the patient had been moved into the final treatment position by the radiation therapist (**Figure 1**).

Dataset generation: Synthetic translational errors were then created by shifting each original DRR by one vertebral body in both directions along the spinal column for every x-ray/ DRR image pair. The shifts were acquired by calculating the cross-correlation coefficient of the shifted DRR and unshifted x-ray image pair and aligning to a local maximum value corresponding to a shift of one vertebral body. Translational shifts of +/- one vertebral body were manually selected from a set of automatically generated shifted cross-correlation local maxima (**Figure 2**). The new DRRs were then cropped to ensure uniform field size between shifted and non-shifted images. The final dataset was composed of 1,980 x-ray/ DRR image pairs and was divided into training and test sets using an 80/20 split. Each image from the stereoscopic image guidance system was treated independently.

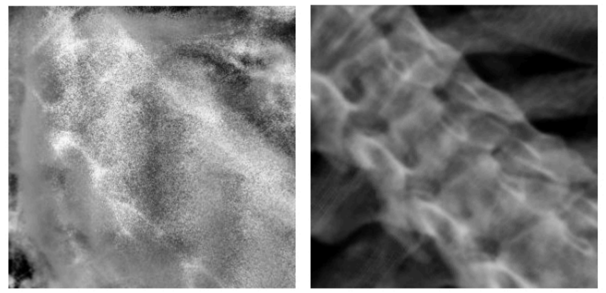

Figure 1: x-ray (left) and DRR (right) image pair from an example thoracic spinal patient.

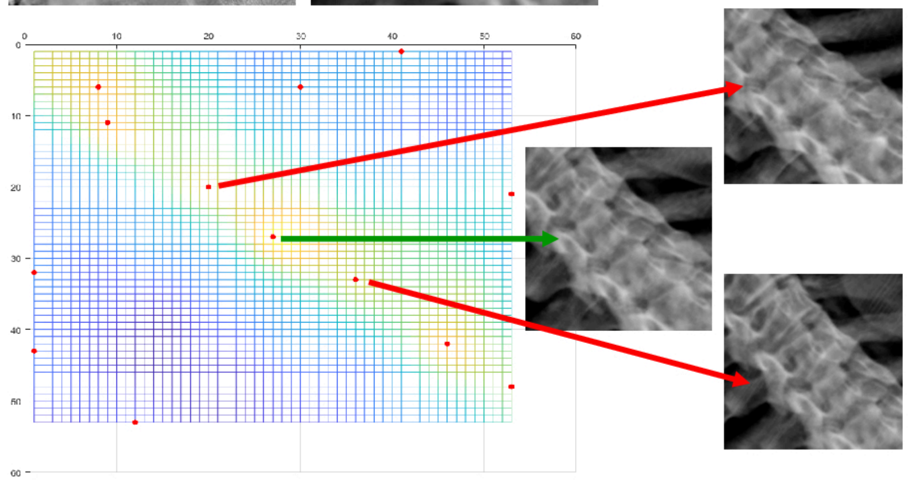

Figure 2: synthetic DRRs obtained by shifting the correctly aligned DRR (green) in both directions along the spinal column and realigning to a local maximum of the cross-correlation coefficient.

Neural Networks: A neural network previously trained on simple translational shifts of a fixed 1 cm was used to evaluate the potential for transfer learning in the specific application of detecting shifts of one vertebral body. For this CNN, the original dataset consisted of 57,036 x-ray/ DRR image pairs from a variety of anatomical sites. In half of the dataset the DRR was shifted by a fixed 1 cm in one of eight set translational directions. The pre-trained network was then applied to the smaller vertebral body training dataset previously described using a 75/25 training/validation split. The top layer of the network was trainable, whereas the weights in all preceding layers were frozen. Our purpose-built CNN was not trained on the translational images used for transfer learning, but instead was trained from scratch on the vertebral body image dataset using the same training/validation split as that in the transfer learning-based CNN.

Logistic Regression: We developed a logistic regression model in order to compare our deep learning-based approach to traditional machine learning techniques. This model used a combination of pixel-wise cross correlation coefficients, cross-correlation coefficients of down-sampled images, and a gradient-based cross-correlation coefficient as independent variables to classify x-ray/DRR image pairs.

Model Performance Evaluation: All three models (transfer learning-based CNN, purpose-built CNN, and logistic regression) were trained using the same 80% of the final image dataset. Following training of all three models, classification accuracies were evaluated and compared using the remaining 20% of test images reserved for this purpose.

RESULTS

Results: When the purpose-built CNN was used to classify the previously unseen test image pairs, the resulting receiver operating characteristic area under the curve (AUC) was 0.972 (**Figure 3**). For comparison, the transfer learning and logistic regression models tested on the same test image dataset obtained AUCs of 0.876 and 0.801, respectively. With the specificity fixed at 99%, the purpose-built CNN achieved a sensitivity of 64.5% in correctly classifying translational shifts of one vertebral body as compared to a sensitivity of 32.3% for the transfer learning model and 23.7% for the logistic regression model.

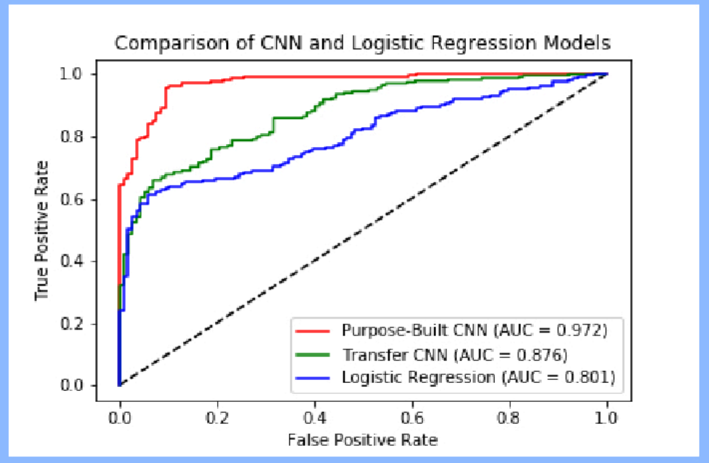

Figure 3: comparison of classification performance in correctly identifying shifts of one vertebral body between two CNN and Logistic Regression models.

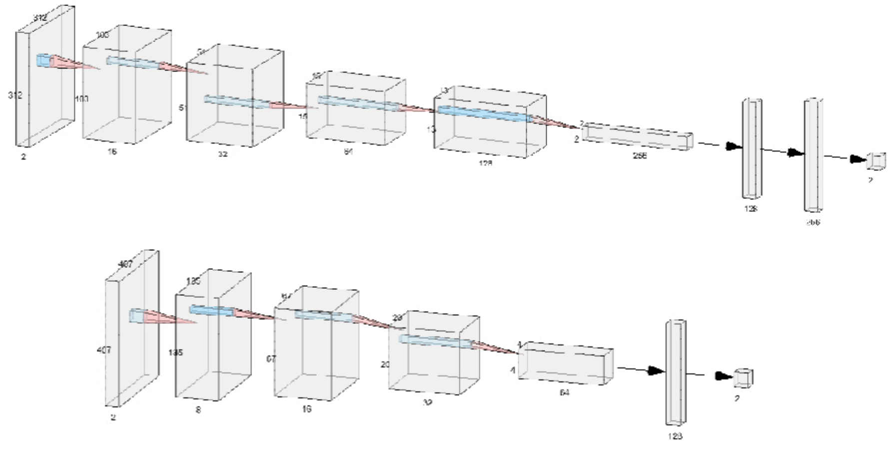

Figure 4: purpose-built CNN (top) and transfer learning (bottom) neural network architectures for binary x-ray/DRR image pair classification

CONCLUSIONS

Initial results of our research are promising and suggest that further development and refinement of our purpose-built deep learning-based approach could have promising clinical applications in the future. Relatively few patient images from stereoscopic x-ray guided thoracic spine treatments are available at our institution, leading us to initially hypothesize that transfer learning utilizing a CNN originally trained on a more general set of synthetic translational errors might be more effective than a purpose-built CNN. However, according to our results, a purpose-built CNN performs better than both transfer learning and logistic regression even for a relatively small training dataset of approximately 1,500 image pairs. In summary, we have developed a purpose-built deep learning-based algorithm which successfully detects shifts of one vertebral body with a higher degree of accuracy than that achieved by either a transfer learning-based CNN approach or a standard logistic regression model.

FUTURE WORK

While the stereoscopic imaging system generates a set of two x-ray images, for the purposes of this work we considered each of these images independently. Future work will focus on methods for treating the two x-ray images and the associated DRRs as a single image set rather than our current method of x-ray/DRR image pairs. Additionally, while our work to date has shown promise as a software failsafe method to mitigate human error, our current sensitivity must be improved before our results can be translated into the clinical workflow. Based on the current patient load at our institution, a specificity of 99% and sensitivity of 95% would correspond to approximately one false positive flag of patient misalignment per week. For our algorithm to be useful, it must not impose an additional burden on physicians manually reviewing false positives. Future work in this aspect will include expanding our image dataset to include x-ray/ DRR image pairs from more recent years. We will also focus on algorithm improvements to increase the classification power of the model to a point where it is clinically meaningful.

ACKNOWLEDGEMENTS

This work supported in part by AHRQ R01 HS026486.

Network rendering using http://alexlenail.me/NN-SVG/LeNet.html.

UCLA RADIATION ONCOLOGY
Division of Medical Physics

Email: RPetragallo@mednet.ucla.edu