(X)AI for treatment error identification in radiotherapy: on the challenge of explaining multidimensional AI decisions with high precision.

Tilburg University

Luca Heising

Maastro

Introduction

Artificial Intelligence (AI) for healthcare is becoming popular, although its implementation is lagging behind [1]. Recent studies found that clinicians' trust in AI is low [2] and that opaqueness of the AI models can be an obstacle for succesful implementation in clinic [3-5]. The trend of AI development is also eminent in radiotherapy (RT). For every step in the RT workflow, promising AI applications have been introduced (Figure 1). Increased transparency and explainability of the deep neural network models are needed to address issues such as, accountability, AI quality assurance (QA), responsibility, and trust.

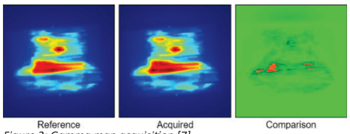


Figure 2: Gamma map acquisition [7].

One of the proposed AI applications for radiotherapy is a convolutional neural network (CNN) for identification of treatment error types in lung cancer patients [6]. During radiotherapy treatment, a radiation beam rotates around the patient. Opposite of the linac head (which provides the beam), an electronic portal imaging device (EPID) is located to capture the delivered dose. A gamma map is a comparison of the aquired dose and the planned dose (Figure 2) which is used for treatment delivery verification. As of today, clinical physicists are notified if the discrepancy between planned and delivered dose is higher than 10% after which they have to detect the error and error type manually. A CNN could aid the physicists in detecting and identifying errors, even below the critical threshold.

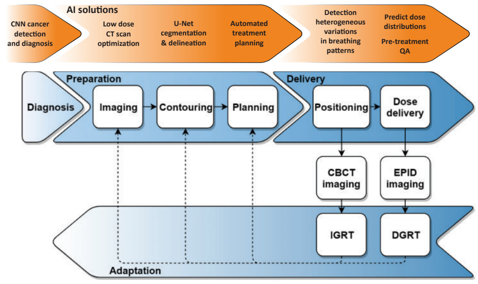


Figure 1: AI applications in the radiotherapy workflow. Addatition to original image by [7].

Background

For the identification of treatment errors, three CNNs based on a VGG-16 architecture were developed and trained to detect errors from multiple classes. The first CNN detects three general error groups: 1) anatomical changes, 2) positioning errors, and 3) mechanical errors. The second CNN model detects subclasses within these groups, and the last CNN also identifies the error magnitude.

The data

The input data consists of real data and synthetic data. Data collection is done for every angle of the linac head, resulting in a gamma map for every timepoint. An aggregation of these time points lead to an integrated 2D gamma map. The dose can also be reconstructed into the CT images, resulting in a 3D dose distribution, when this is done for every time point, we get 4D dose distributions (3D + time) (Figure 3).

Errors occur only infrequently, because of this, only scarce amount of data is available to train a CNN model on. To account for this, synthetic data was created by simulating errors in CT images. More information on data simulations can be found in [6].

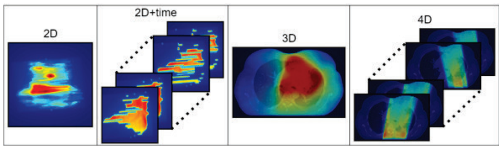
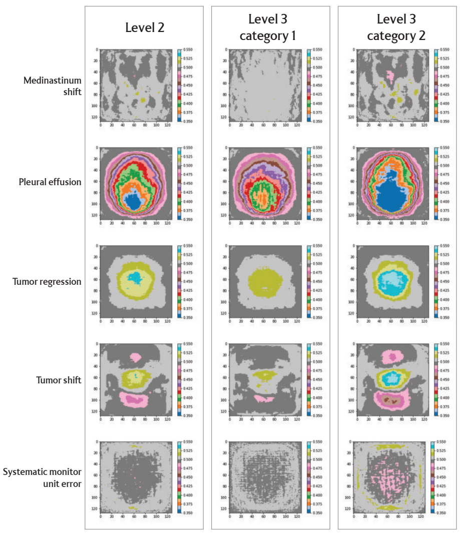


Figure 3: EPID data in four different dimensionalities [7].

Understanding the data



To fully understand the data, extensive data analyses are required. The average image of all images in one class for level 2, level 3 with magnitude category 1, and level 3 magnitude category 2 are shown. Here we can see patterns in the data for that specific error type.

This information can be fruitful for developers and clinicians as errors are uncommon. Class averages can guide them in their error identification. Additionally, a thorough data description can prevent incorrect use of the AI model.

For example, in the case of tumor regression, the tumorsize has decreased. As tumor tissue has a higher density than normal tissue, an overdose can be seen on the gamma map. This is visible in the class average as a spot with a value of > 0.5 is shown, where 0.5 is the cutoff point for desired dose and overdose (> 0.5) or underdose (< 0.5).

Challenges

In radiotherapy, **high precision** is required. Rigor and easily interpretable explanations are needed as current XAI methods are inadequate.

Clinicians only have **limited time**. AI are often compared to clinicians, although a hybrid approach is desired to aid the clinician.

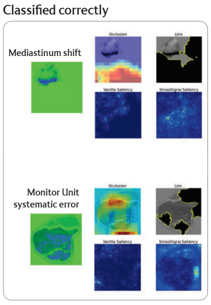
Developers and clinicians have different goals. Providing useful explanations for both is challenge, and **close collaboration** is required.

Incorrect predictions can be disastrous; **verification of the AI** is required (e.g., with physics-informed verification methods).

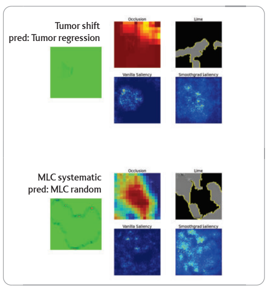
The **complexity and multidimensionality** of the EPID images are challenging for XAI. A full understanding of the input data is required.

Understanding the model

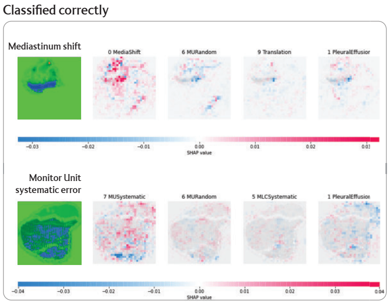
Classified correctly



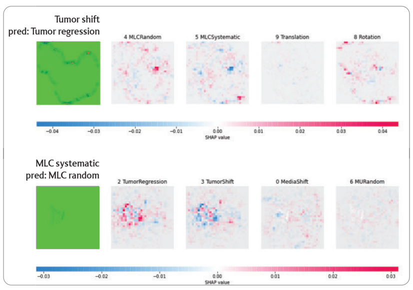
Misclassified



Classified correctly



Misclassified



The input data for this CNN is inherently complex and difficult to interpret. This raises the challenge of finding suitable XAI methods for clear interpretation of the CNN model. Both correctly classified images and misclassified images show anomalies in the feature importance maps.

References

[1] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," Nat. Med., vol. 25, no. 1, pp. 30–36, Jan. 2019, doi: 10.1038/s41591-018-0307-0.

[2] X. Liu et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," Lancet Digit. Health, vol. 1, no. 6, pp. e271–e297, Oct. 2019, doi: 10.1016/S2589-7500(19)30123-2.

[3] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," J. Am. Med. Inform. Assoc., vol. 27, no. 3, pp. 491–497, Mar. 2020, doi: 10.1093/jamia/ocz192.

[4] J. Shaw, F. Rudzicz, T. Jamieson, and A. Goldfarb, "Artificial Intelligence and the Implementation Challenge," J. Med. Internet Res., vol. 21, no. 7, p. e13659, Jul. 2019, doi: 10.2196/13659.

[5] B. H. Kann, A. Hosny, and H. J. W. L. Aerts, "Artificial intelligence for clinical oncology," Cancer Cell, vol. 39, no. 7, pp. 916–927, Jul. 2021, doi: 10.1016/j.ccell.2021.04.002.

[6] C. J. A. Wolfs, R. A. M. Canters, and F. Verhaegen, "Identification of treatment error types for lung cancer patients using convolutional neural networks and EPID dosimetry," Radiother. Oncol., vol. 153, pp. 243–249, Dec. 2020, doi: 10.1016/j.radonc.2020.09.048.

[7] C. (Julie A. Wolfs, "Quantitative methods for improved error detection in dose-guided radiotherapy," Doctoral Thesis, Ipskamp Printing BV, Maastricht, 2020. doi: 10.26481/dis.20200925cw.