Handcrafted versus deep learning radiomics for prediction of cancer therapy response

See Articles page e136

This online publication has been corrected. The corrected version first appeared at thelancet.com/digital-health on July 1, 2019

In *The Lancet Digital Health*, Bin Lou and colleagues[1] apply deep learning methods to analyse pre-treatment CT scans in a retrospective cohort study of 944 patients (849 in the internal study cohort and 95 in the independent validation cohort) treated with stereotactic body radiation therapy, a form of high-dose, pinpoint radiation therapy for lung tumours. The study presents a novel analysis by integrating traditional radiomics features through multi-task learning, applying a time-based survival analysis, and incorporating new deep learning methods including a three-dimensional (3D) convolutional neural network to analyse lung tumours before treatment. The authors input pre-therapy lung CT images into Deep Profiler, a multi-task deep neural network that has radiomics incorporated into the training signal. They combined these data with clinical variables to derive *i*Gray, an individualised radiation dose that estimates the probability of treatment failure to be below 5%. Models that included Deep Profiler and clinical variables predicted treatment failures with a concordance index of 0·72 (95% CI 0·67–0·77), a significant improvement compared with traditional radiomics (p<0·0001) or clinical variables (p<0·0001) alone. The potential clinical applications of such models include identifying tumours at the highest risk of resistance to radiation therapy, and personalised dosing of radiation therapy to maximise likelihood of tumour control.

This study is representative of a major turning point in the underlying radiomics methodologies used in treatment response prediction and prognosis, specifically in radiation therapy with broader implications across other cancer therapies. Traditional radiomics makes use of handcrafted features and has been studied extensively as an imaging biomarker to predict cancer outcomes and responses to therapy.[2,3] The handcrafted radiomics approach involves manual segmentation of the region of interest (eg, the tumour) on medical imaging, and extraction of thousands of human-defined and curated quantitative features from the region of interest, which describe tumour shape and texture among other characteristics. In the final step, the approach involves application of machine learning methods to identify the imaging features that are associated with a given clinical endpoint. However, the human-derived nature of traditional radiomics methods has been criticised for introducing a source of human bias into the process;[4] there have been concerns of reproducibility[5] due to the intra-reader and inter-reader variability that results from the reliance on manual segmentation of the tumour, and due to variation in imaging and pre-processing techniques for feature extraction. Moreover, the value of traditional radiomics has recently come under question with the advent of deep learning methods and consequent proof-of-principle applications in predicting cancer outcomes.[6,7] For many of the deep learning radiomics applications, region of interest definition is based on a single point placement within the tumour volume, essentially replacing full tumour segmentations with approximate localisation and minimising the need for human input. Additionally, deep learning methods allow for automated learning of relevant radiographic features without the need for previous definition by researchers. In turn, these abstract representations have enabled a larger learning capacity, boosting generalisability and accuracy while reducing potential bias.[8]

Some key caveats remain for clinical use of the deep learning model proposed by Lou and colleagues.[1] Firstly, the radiation dose delivered via stereotactic body radiation therapy for lung cancers represents the upper limit of what can be safely delivered to treat cancer in the human body with current technological capabilities. Of note, other tumours are often treated at substantially lower biological doses, and this study does not capture that range of radiation dose and tumour response curves. Secondly, radiation regimens used for stereotactic body radiation therapy are typically achievable only for localised (eg, stage I lung cancers) and small tumours (e.g. <5 cm diameter), and thus these dose predictions are not easily generalisable to more advanced tumours. Lastly, the model is built on a relatively rare event (3-year cumulative incidence of local failure was 13·5% in the overall population) which is an advantage to patients because it means stereotactic body radiation therapy works well, but a

disadvantage for predictive model building because of the increased risk of over-fitting.

In this study, the authors chose to identify handcrafted radiomics features as ground truth while comparing them to features identified by deep learning methods. The level of agreement between these two sets of features was then used as a cost function to train and optimise the predictive model. This method was understandably chosen as a means to provide a connection to the previous traditional radiomics landscape and greater interpretability. However, we believe that deep learning can emerge as an independent methodology that does not need to rely on handcrafted radiomics to move forward. Combining traditional radiomic features into deep learning models risks incorporating the aforementioned known human biases into the model. Additionally, a combined approach does not address the interpretability problem since even most mathematically-derived handcrafted features capture uninterpretable imaging characteristics that cannot be discerned by the human eye. Nevertheless, the challenges of traditional radiomics approaches such as lack of reproducibility and interpretability as well as over-fitting on small datasets will only be amplified in deep learning-driven prediction models of cancer outcome. Fortunately, interpretability of features learned through neural networks is an active area of research,[9] while sharing and transparency initiatives are paving the way for larger curated cancer imaging repositories.[10]

Deep learning may also allow the decoding of new insights from cancer images and non-intuitive information that is uncharted thus far. We look with great interest at the saliency mapping in figure 5 of the Article, which identifies the regions of the CT scan in and around the tumour that are most associated with the predicted outcome of local tumour recurrence. Our group identified similar peri-tumoural localisation when performing activation mapping for a 3D convolutional neural network trained for a prognostication task in non-small lung cancer patients, which suggests potentially important imaging characteristics at the cancer-normal tissue interface.[7] Although these findings are preliminary and qualitative in nature, future work to understand the biology of this interface in relation to cancer therapy response prediction, and perhaps more importantly applying deep learning radiomics to target localised cancer therapies such as radiation therapy and surgery, represent a truly exciting new frontier of cancer care.

*Ahmed Hosny, Hugo J Aerts, *Raymond H Mak*
Department of Radiation Oncology (AH, HJA, RHM) and Department of Radiology (HJA), Brigham and Women's Hospital, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02115, USA
rmak@partners.org

HJA reports personal fees from Genospace and personal fees from Sphera, outside the submitted work. RHM reports personal fees from AstraZeneca and personal fees from NewRT, outside the submitted work. AH declares no competing interests.

1 Lou B, Doken S, Zhuang T, et al. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. *Lancet Digital Health* 2019; **1:** e136–47.
2 Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 2015; **114:** 345–50.
3 Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5:** 4006.
4 Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One* 2015; **10:** e0124165.
5 Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys* 2018; **102:** 1143–58.
6 Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019; **25:** 3266–75.
7 Hosny A, Parmar C, Coroller TP, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med* 2018; **15:** e1002711.
8 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.
9 Chakraborty S, Tomsett R, Raghavendra R, et al. Interpretability of deep learning models: a survey of results. 2017 IEEE SmartWorld: ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); San Francisco Bay Area, CA, USA; Aug 4–8, 2017.
10 Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013; **26:** 1045–57.