ESTRO Mobility Grant (TTG) Report

FAIR Quantitative Imaging Infrastructure for Deep Neural Networks

Date of visit: 27 – 28 January 2019
Host institute: Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA

Background

Quantitative, or artificial intelligence (AI)-assisted, prediction of local control and survival from pre-treatment radiological imaging has hitherto untapped potential to guide clinical risk estimation in this regard. Deep artificial neural networks (DNNs) [1] have been applied to the problem of predicting long-term outcomes using a combination of clinical and imaging data. The ground-breaking capabilities of DNNs to classify patterns are acquired through its self-evolving learning strategy using vast volumes of data, and these have shown near-expert levels of performance.

Aim of the Visit

The primary aim of the visit was to define a systematic methodology to make vast amounts of radiology images and tumour delineations, treatments and outcomes Findable, Accessible, Interoperable and Reusable (FAIR) [2] for the purpose of training generic DNNs.

Materials and Method

We developed a FAIR quantitative imaging analysis workflow (FAIR-QIAW) that could convert digital imaging and communications in medicine (DICOM) imaging data to FAIR quantitative imaging data. In total, 612 patients from four cohorts that were available on XNAT (https://xnat.bmia.nl) were used in the study (**Figure 1**).

The diagram of the conversion procedures is shown in **Figure 2**. First, we created a JavaScript object notation (JSON) file that consisted of the metadata of the region-of-interest segmentation. Then, the DICOM image and radiotherapy structure set (RTSTRUCT) were converted to image volume and binary mask in nearly raw raster data (NRRD) format. At the same level, we created a DICOM segmentation object (DICOM-SEG [3]) using the DICOM for quantitative imaging (DCMQI) toolbox [4]. DICOM-SEG is the standard way to encode segmentations defined as labelled image voxels. Next, the paths of binary mask and image were stored in a CSV table locally. Then, the deep learning model developed in [5], which is available on ModelHub [6], was used to extract deep learning-based features, and PyRadiomics [7], a radiomics extractor, was used to extract radiomic features. In another branch, PyRadiomics-dcm was used to compute DICOM structured reporting (DICOM-SR [3]).

Results

The proposed FAIR-QIAW program automatically processed the data of all 612 patients. The results for each patient consisted of deep learning-based features, radiomic features, a DICOM-SEG object and a DICOM-SR object. The DICOM-SEG files of several datasets that were used are now available on The Cancer Imaging Archive (TCIA) (https://www.cancerimagingarchive.net). These are: the reference image database to evaluate therapy response (RIDER); the LUNG1 dataset of 422 non-small-cell lung cancer (NSCLC) patients; the Interobserver dataset also of NSCLC patients, and head and neck 1 (HN1).

Future work

We have developed a workflow to generate FAIR imaging data automatically directly from DICOM data. The future work will mainly involve application of the proposed workflow to perform a real study, for instance development of a lung organ segmentation model via federated deep learning on DICOM-SEG generated by FAIR-QIAW.

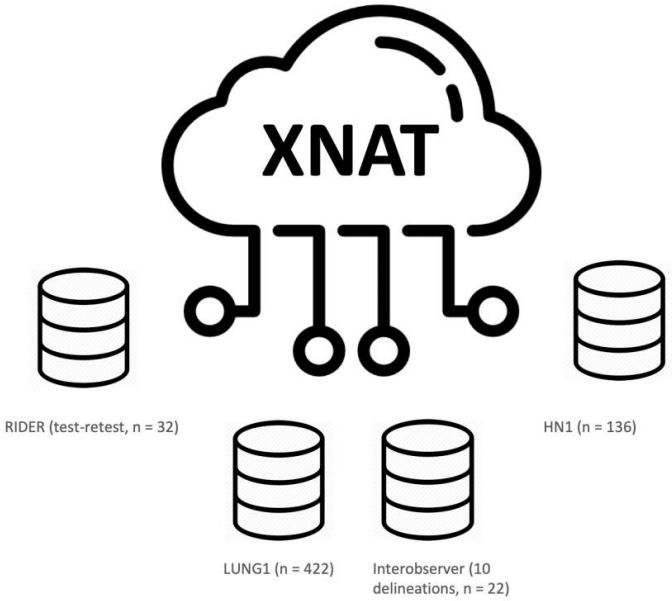

Figure 1: The diagram of data used in this study. All data (RIDER, LUNG1, Interobserver, and HN1) are available on XNAT, the open source project at the Washington University School of Medicine.

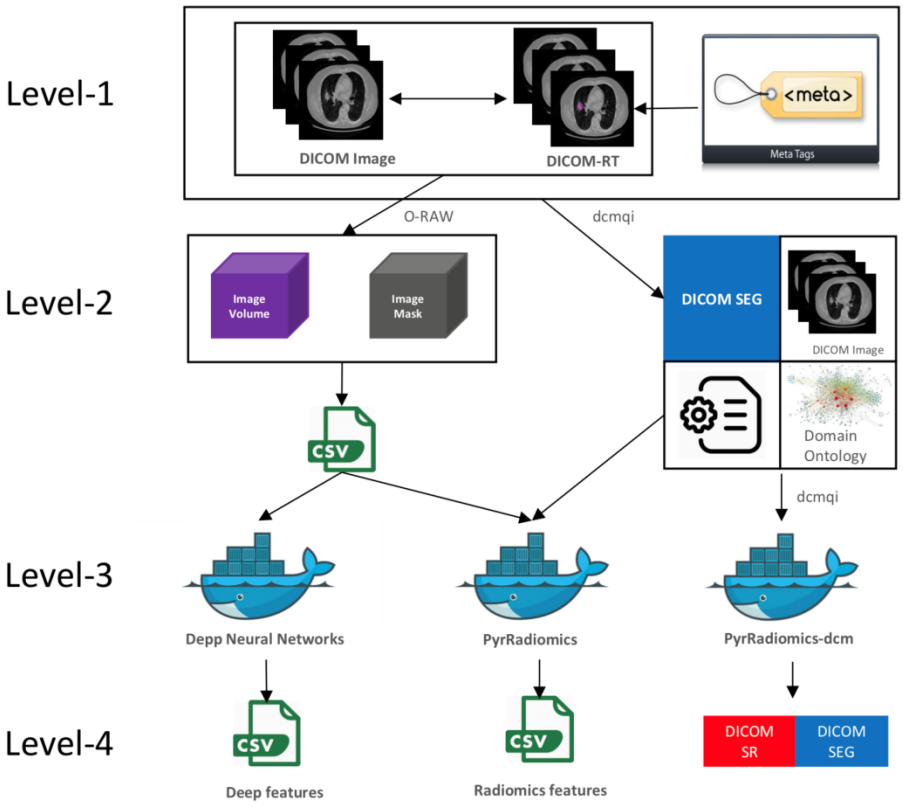

Figure 2: Four levels of the processing in FAIR-QIAW.

References

1. LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.
2. Wilkinson, M.D., et al., The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 2016. 3.
3. Fedorov, A., et al., Standardized representation of the LIDC annotations using DICOM. 2019, PeerJ Preprints.
4. Herz, C., et al., DCMQI: an open source library for standardized communication of quantitative image analysis results using DICOM. Cancer research, 2017. 77(21): p. e87-e90.
5. Hosny, A., et al., Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS medicine, 2018. 15(11): p. e1002711.
6. Hosny, A., et al., ModelHub. AI: Dissemination Platform for Deep Learning Models. arXiv preprint arXiv:1911.13218, 2019.
7. van Griethuysen, J.J., et al., Computational Radiomics System to Decode the Radiographic Phenotype. Cancer research, 2017. 77(21): p. e104-e107.

Zhenwei Shi
Department of Radiation Oncology (MAASTRO)
GROW School for Oncology & Developmental Biology
Maastricht University Medical centre
Maastricht, the Netherlands.
Email: zhenwei.shi@maastro.nl