

# An Open Library of CT Patient Projection Data

Baiyu Chen<sup>a</sup>, Shuai Leng<sup>a</sup>, Lifeng Yu<sup>a</sup>, David Holmes III<sup>b</sup>, Joel Fletcher<sup>a</sup>, Cynthia McCollough\*<sup>a</sup>

<sup>a</sup>Department of Radiology, Mayo Clinic, Rochester, MN 55905 USA

<sup>b</sup>Department of Physiology and Biomedical Engineering

## ABSTRACT

Lack of access to projection data from patient CT scans is a major limitation for development and validation of new reconstruction algorithms. To meet this critical need, we are building a library of CT patient projection data in an open and vendor-neutral format, DICOM-CT-PD, which is an extended DICOM format that contains sinogram data, acquisition geometry, patient information, and pathology identification. The library consists of scans of various types, including head scans, chest scans, abdomen scans, electrocardiogram (ECG)-gated scans, and dual-energy scans. For each scan, three types of data are provided, including DICOM-CT-PD projection data at various dose levels, reconstructed CT images, and a free-form text file. Several instructional documents are provided to help the users extract information from DICOM-CT-PD files, including a dictionary file for the DICOM-CT-PD format, a DICOM-CT-PD reader, and a user manual. Radiologist detection performance based on the reconstructed CT images is also provided. So far 328 head cases, 228 chest cases, and 228 abdomen cases have been collected for potential inclusion. The final library will include a selection of 50 head, chest, and abdomen scans each from at least two different manufacturers, and a few ECG-gated scans and dual-source, dual-energy scans. It will be freely available to academic researchers, and is expected to greatly facilitate the development and validation of CT reconstruction algorithms.

**Keywords:** Computed tomography (CT); patient data library; projection data; reconstruction; DICOM-CT-PD.

## 1. INTRODUCTION

Projection data from patient CT scans, especially those with known pathology, are essential to the development and validation of new reconstruction algorithms. However, the patient projection data collected

from commercial CT scanners are proprietary, which means researchers need research agreements with CT vendors to access the data; the projection data collected from commercial CT scanners are also vendor-specific, which means each CT vendor stores the projection data in its own format (the geometries used to store the projection images and the acquisition parameters, the unit used to store the acquisition parameters, and the precision and range of the numerical values might all differ from vendor to vendor).

To allow more researchers to access patient CT projection data with minimum efforts, a library of CT patient projection data is being built and will be freely available to academic researchers. The projection data in the library were decoded from commercial CT scans, and have been converted into an open and standard format.

## 2. METHODS

As shown in Figure 1, the patient library has three main components: scan data, instructional documents, and radiologists' diagnostic performance, as detailed below.

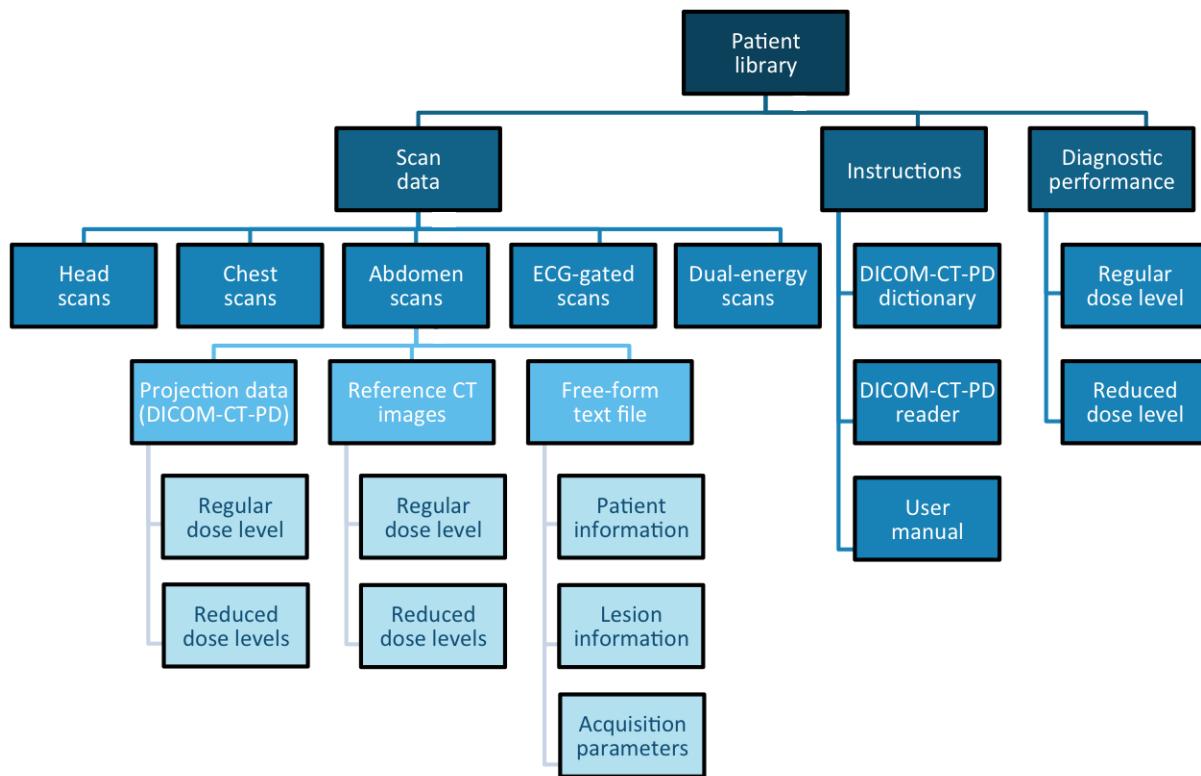


Figure 1: The structure of the patient library.

The scan data is composed of scans of five types, including routine non-contrast-enhanced head scans, low dose non-contrast-enhanced chest scans for lung nodule screening, routine contrast-enhanced abdomen scans, ECG-gated scans, and dual-energy scans. For each scan, three types of data are provided: projection data, reconstructed CT images, and a free-form text file. The projection data were acquired on third generation CT scanners from two major vendors (Somatom Definition Flash, Siemens Healthcare, Forchheim, Germany; and Discovery CT750 HD, GE Healthcare, Waukesha, WI). Because the commercial projection data were in a proprietary format and could not be accessed directly, they were decoded with the assistance of the vendors and converted into an open and vendor-neutral format, DICOM-CT-PD [1]. The DICOM-CT-PD format is an extended DICOM format, which stores the projection data as a DICOM image and stores other important information (acquisition geometry, patient information, and pathology identification) in a DICOM header with newly defined private tags. The accuracy and completeness of the DICOM-CT-PD format have been previously validated by off-line reconstructions. More information about the format can be found in [1]. In addition to projection data acquired at regular clinical dose levels, projection data at reduced dose levels are also provided, which were simulated by inserting noise into the regular dose projection data using a verified technique [2]. The reconstructed CT images are provided along with the projection data as a reference. For the projection data collected on Siemens scanners, all reconstructions were performed on the scanner console. For the projection data collected on GE scanners, the reconstructions of the regular dose level data were performed on the scanner console and the reconstructions of the reduced dose level data were performed off-line. The free-form text file stores the patient information, the pathology information, and most acquisition parameters. Although the same information has been provided in the header of the DICOM-CT-PD files, we purposely kept this redundancy for user convenience.

To help the users extract information from DICOM-CT-PD files, several instructional documents are available, including a dictionary file that describes all tags defined in DICOM-CT-PD, a reader (MATLAB script) that extracts information from DICOM-CT-PD files, and a user manual that explains the DICOM-CT-PD format.

Using the Siemens' reconstructed CT images, detection tasks were performed by radiologists at our institution. The diagnostic performance was characterized in the form of Jackknife free-response receiver operating characteristic (JAFROC) curves [3] for each radiation dose level and reconstruction setting. The results will be made available at the conclusion of a multi-reader, multi-case (MRMC) study that is underway.

### 3. RESULTS

The patient cases that have been collected so far for evaluation include: 328 routine non-contrast-enhanced head exams (164 positive cases and 164 normal cases), 228 low-dose chest exams for lung nodule screening (114 indeterminate cases and 114 normal cases), and 228 routine contrast-enhanced abdomen exams (114 cases with malignant liver lesions, 57 cases with benign liver lesions, and 57 normal cases). The final library will include a selection of 50 head, 50 chest, and 50 abdomen scans, from at least two CT manufacturers (Siemens Healthcare, Forchheim, Germany; and GE Healthcare, Waukesha, WI) and a few ECG-gated scans and dual-source, dual-energy scans. An example of the material available in the library is illustrated in Figure 2. An example of the range of pathologies in the abdominal patient data library is given in Figure 3.

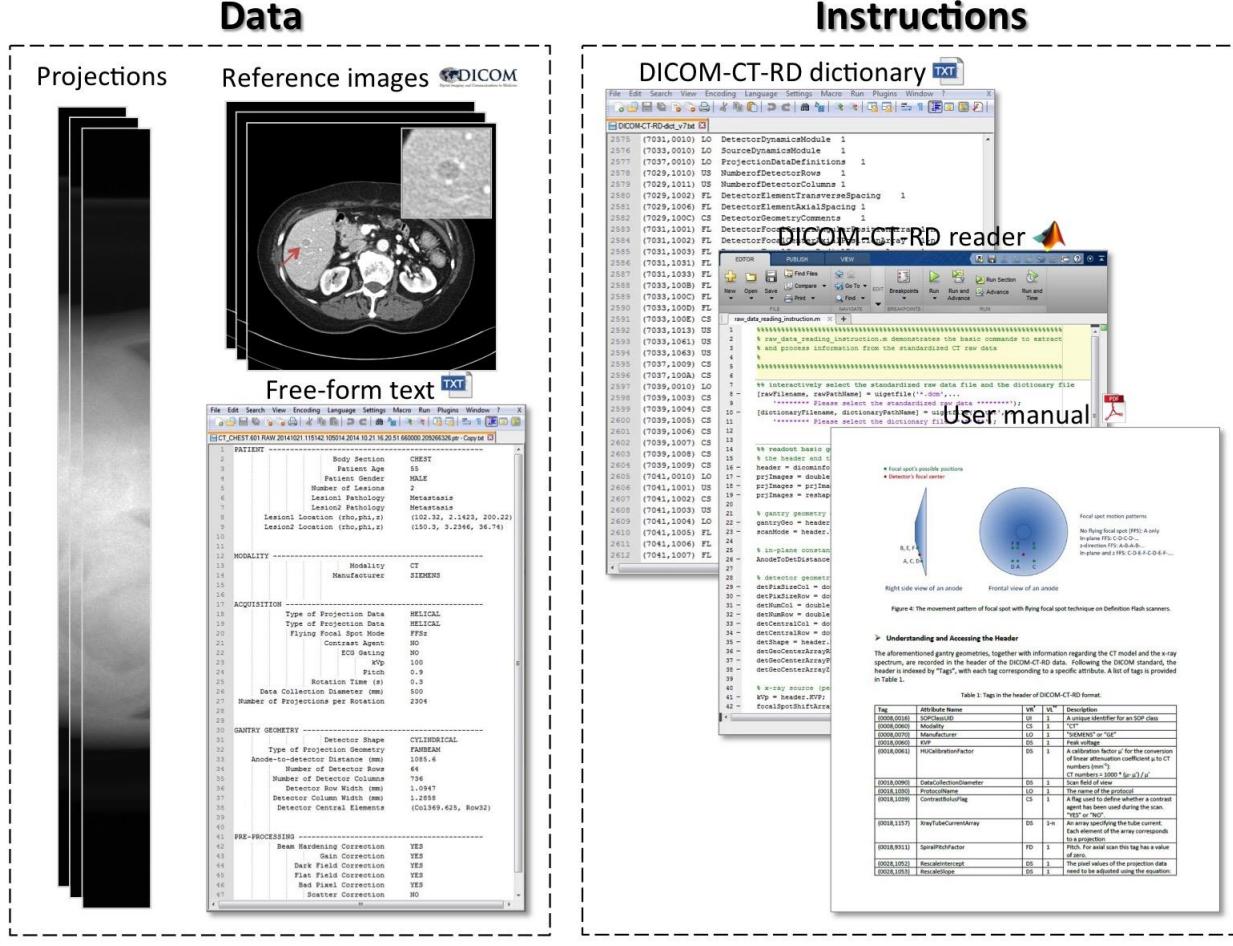


Figure 2: An abdominal case from the patient library. The reference images were reconstructed from commercial projection data using the scanner console.

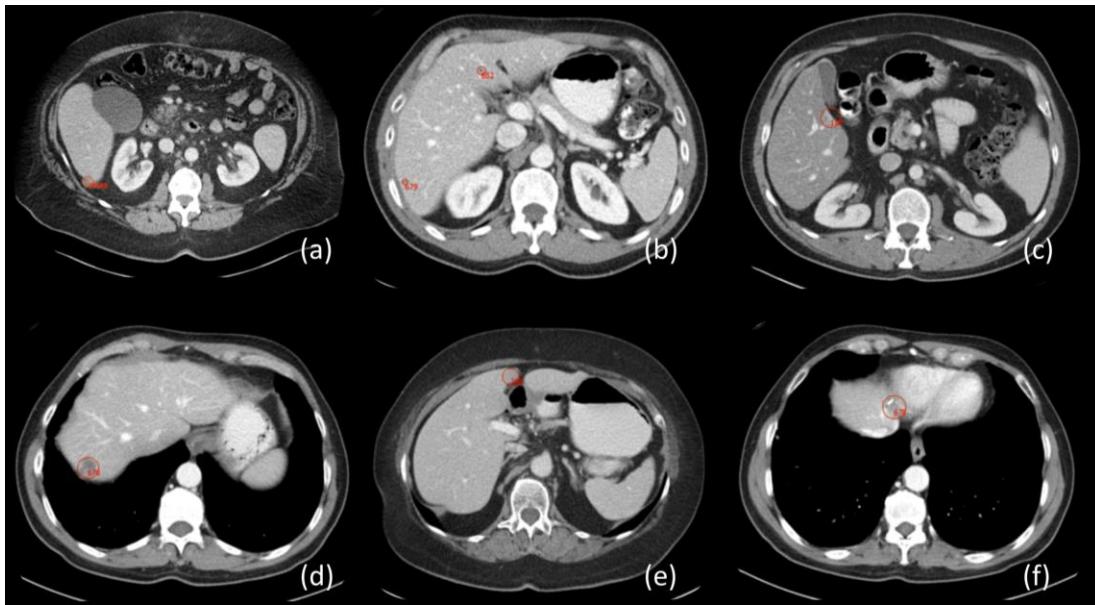


Figure 3: An example of the range of pathologies in the abdominal patient data library, including (a) metastasis, (b) benign cyst, (c) focal fat/ fatty sparing, (d) hemangioma, (e) vascular-perfusion defect, and (f) post op/post RFA defect.

#### 4. DISCUSSION

To our knowledge, this is the first library of patient CT projection data that is freely-available to researchers. Because the projection data are in an open and vendor-neutral format, they can be accessed without having research agreements with the CT manufacturers. Because the projection data are acquired with patients instead of phantoms, they have anatomical complexity at the clinical level.

The library is expected to greatly facilitate the development and validation of reconstruction algorithms. For example, the library can be used to benchmark reconstruction algorithms against each other. A subset of the patient library (30 reduced-dose abdomen cases) has recently been used to conduct a Low Dose CT Grand Challenge [4], in which CT reconstruction researchers from over 20 countries participated and reconstructed the same projection dataset using their noise-reducing or reconstruction algorithms. The reconstructed images will be viewed by radiologists and compared in terms of their diagnostic accuracy, and the winners will be invited to present their winning-algorithms at the 2016 AAPM annual meeting.

## **ACKNOWLEDGEMENT**

This work was funded by National Institute of Biomedical Imaging and Bioengineering (U01 EB017185). The authors would like to thank Dr. Karl Stierstorfer and Dr. Jiang Hsieh for their help in decoding commercial CT raw data. The authors would also like to thank Dr. Gregory J. Michalak and Ms. Alice E. Huang for their help in the patient data collection.

## **REFERENCE**

- [1] B. Chen, X. Duan, Z. Yu at al., "Technical Note: Development and validation of an open data format for CT projection data", *Medical Physics*, 42(12), 6964 (2015).
- [2] L. Yu, M. Shiung, D. Jondal et al., "Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols," *Journal of computer assisted tomography*, 36(4), 477-87 (2012).
- [3] D.P. Chakraborty, K. S. Berbaum, "Jackknife free-response ROC methodology", in *Medical Imaging 2004*, Vol. 5372 (SPIE, Bellingham, WA 2004), pp.144-153.
- [4] <http://www.aapm.org/GrandChallenge/LowDoseCT>