Low Dose CT Image and Projection Data (LDCT-and-Projection-data)

Summary

Investigators at the Mayo Clinic, with funding from the National Institute of Biomedical Imaging and Bioengineering (EB 017095 and EB 017185), have built a library of CT patient projection data in an open and vendor-neutral format. This format, referred to as DICOM-CT-PD (1), is an extended DICOM format that contains CT projection data and acquisition geometry. The de-identified patient projection data in the library were decoded with help of the manufacturer and have been converted into an open standardized format.

Reconstructed images, patient age and gender, and pathology annotation are also provided for these de-identified data sets. The library consists of scans from various exam types, including non-contrast head CT scans acquired for acute cognitive or motor deficit, low-dose non-contrast chest scans acquired to screen high-risk patients for pulmonary nodules, and contrast-enhanced CT scans of the abdomen acquired to look for metastatic liver lesions.
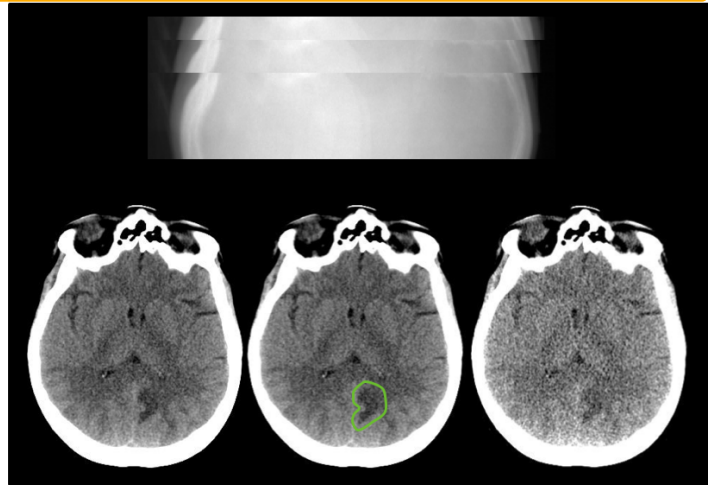
2016 Low Dose CT Grand Challenge

The 2016 Low Dose CT Grand Challenge, sponsored by the AAPM, NIBIB, and Mayo Clinic, used 30 contrast-enhanced abdominal CT patient scans, 10 for training and 20 for testing. Thirteen of the 20 testing datasets from the Grand Challenge were subsequently included in this larger collection of CT image and projection data (TCIA LDCT-and-Projection-data). Because of the frequency of requests received by Mayo and the AAPM for the complete 2016 Grand Challenge dataset, on September 21, 2021 all 30 cases were updated to use the same projection data format as used for the TCIA data library and made publicly available in a single location. Please refer to the READ ME file at that location for a mapping between the case ID numbers used in the 2016 Grand Challenge and the case ID numbers used in the TCIA library for the 13 cases that exist in both libraries.

Additional information about the 2016 Low Dose CT Grand Challenge can be found on the AAPM website and in the *Medical Physics* paper by McCollough et al.

Acknowledgements

This work would not have been possible without the support and efforts of many individuals and organizations.

- A complete list of acknowledgements can be found here.

Data Access

Data Access

Some data in this collection contains images that could potentially be used to reconstruct a human face. To safeguard the privacy of participants, users must sign and submit a TCIA Restricted License Agreement to help@cancerimagingarchive.net before accessing the data.

Data Type	Download all or Query/Filter	License
Images (DICOM, 1.2 TB)	*Download Search* (Requires NBIA Data Retriever.)	TCIA Restricted
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*	TCIA Restricted
DICOM-CT-PD User Manual Version 3 (.pdf, 1.32 MB)	*Download*	CC BY 3.0
Matlab DICOM-CTPD data dictionary (.txt, 104 kB)	*Download*	CC BY 3.0
Matlab DICOM-CTPD reader script (.txt, 4 kB)	*Download*	CC BY 3.0
Clinical Data (CSV, zip, 107 kB)	*Download*	CC BY 3.0

Click the Versions tab for more info about data releases.

Additional Resources

- The Helix2Fan: Helical to fan-beam CT geometry rebinning and differentiable reconstruction of DICOM-CT-PD projections repository on Github provides source code to load raw helical DICOM-CT-PD CT projections and rebin them to flat detector fan-beam geometry.

<u>**Detailed Description**</u>
Detailed Description

Image Statistics	
Modalities	CT
Number of Participants	299
Number of Studies	597
Number of Series	1045
Number of Images	13,013,532
Images Size (TB)	1.2

For each patient CT scan, three types of data are provided: DICOM-CT-PD projection data, DICOM image data, and Excel clinical data reports. CT projection data are provided for both full and simulated lower dose levels and CT image data reconstructed using the commercial CT system are provided for the full dose projection data. For patients scanned on the SOMATOM Definition Flash CT scanner from Siemens Healthcare, CT image data reconstructed using the commercial CT system are also provided for the lower dose projection data. All CT images were reconstructed using a filtered back projection method. Several instructional documents are provided to help users extract needed information from the DICOM-CT-PD files, including a dictionary file for the DICOM-CT-PD format, a DICOM-CT-PD reader, and a user manual.

This collection comprises 99 head scans (labeled N for neuro), 100 chest scans (labeled C for chest), and 100 abdomen scans (labeled L for liver). Fifty cases for each scan type are from a SOMATOM Definition Flash CT scanner (Siemens Healthcare, Forchheim, Germany). Forty-nine head cases, 50 chest cases, and 50 abdomen cases are from a Lightspeed VCT CT scanner (GE Healthcare, Waukesha, WI). Together, these data will greatly facilitate the development and validation of new CT reconstruction and/or denoising algorithms, including those associated with machine learning or artificial intelligence.

Acquisition protocol

All CT scans were acquired at routine dose levels for the practice at which they were obtained using standard-clinical protocols for the anatomical region of interest. Each clinical case was processed to include a second projection dataset at a simulated lower dose level. Head and abdomen cases are provided at 25% of the routine dose and chest cases are provided at 10% of the routine dose.

[1]Additional information regarding the CT projection data format: Chen B, Duan X, Yu Z, Leng S, Yu L, McCollough CH. Technical Note: Development and validation of an open data format for CT projection data. Med Phys. 2015;42 (12):6964. (doi: https://doi.org/10.1118/1.4935406.)

<u>Citations & Data Usage Policy</u>

Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

> ⓘ **Data Citation**
>
> McCollough, C., Chen, B., Holmes III, D. R., Duan, X., Yu, Z., Yu, L., Leng, S., & Fletcher, J. (2020). Low Dose CT Image and Projection Data (LDCT-and-Projection-data) (Version 6) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/9NPB-2637

> ⓘ **Publication Citation**
>
> Moen, T. R., Chen, B., Holmes, D. R., III, Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J. G., & McCollough, C. H. (2020). Low dose CT image and projection dataset. Medical Physics. https://doi.org/10.1002/mp.14594

> ⓘ **Acknowledgement**
>
> Presentations and publications shall acknowledge grants EB017095 and EB017185 (Cynthia McCollough, PI) from the National Institute of Biomedical Imaging and Bioengineering.

> ⓘ **TCIA Citation**
>
> Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. In Journal of Digital Imaging (Vol. 26, Issue 6, pp. 1045–1057). Springer Science and Business Media LLC. https://doi.org/10.1007/s10278-013-9622-7

Other Publications Using This Data

TCIA maintains a list of publications which leverage TCIA data. If you have a manuscript you'd like to add please contact the TCIA Helpdesk.

<u>Versions</u>

Version 6 (Current): Updated 2023/04/03

Data Type	Download all or Query/Filter	License

Images (DICOM, 1.2 TB)	*Download Search* (Requires NBIA Data Retriever.)	TCIA Restricted
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*	TCIA Restricted
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*	CC BY 3.0
Matlab DICOM-CTPD data dictionary (.txt)	*Download*	CC BY 3.0
Matlab DICOM-CTPD reader script (.txt)	*Download*	CC BY 3.0
Clinical Data (CSV, zip)	*Download*	CC BY 3.0

Note the following corrections were made to the clinical data:

- Lesion Z locations for all GE cases now reflect slice number

Version 5: Updated 2022/12/15

Data Type	**Download all or Query/Filter**	**License**
Images (DICOM, 1.2 TB)	*Download Search* (Requires NBIA Data Retriever.)	TCIA Restricted
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*	TCIA Restricted
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*	CC BY 3.0
Matlab DICOM-CTPD data dictionary (.txt)	*Download*	CC BY 3.0
Matlab DICOM-CTPD reader script (.txt)	*Download*	CC BY 3.0
Clinical Data (CSV, zip)	*Download*	CC BY 3.0

Note: The following corrections were made to the clinical data:

- Incorrect Lesion coordinates

- In some places the x/y coordinates for lesions was incorrect
- In almost all Siemens datasets, the Z axis was flipped

- Incorrect UID

  - Correction has been made to case L210 UID

Version 4: Updated 2022/03/31

Data Type	Download all or Query/Filter
Images (DICOM, 1.2 TB)	*Download Search* (Requires NBIA Data Retriever.)
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*
Matlab DICOM-CTPD data dictionary (.txt)	*Download*
Matlab DICOM-CTPD reader script (.txt)	*Download*
Clinical Data (CSV, zip)	*Download*

Update on the GE data

The following corrections/changes have been made to the GE projection data:

- The value in DICOM tag (0028,1052) RescaleIntercept has been re-calculated and updated.
- DICOM tag (7033,1065) PhotonStatistics, which describes the incident x-ray beam profile after passing through the bowtie filter, has been added for every projection. The beam profile is characterized in terms of noise equivalent number of incident photons (i.e., noise equivalent quanta)..
- DICOM tag (0018,1151) XRayTubeCurrent has been updated for the chest and abdomen cases, which use tube current modulation. The value reflects the tube current for each projection view. With tube current modulation on, this value will vary across different project views. Tube current modulation is off for the head exams.
- It was discovered that the projection data for chest case C197 was a duplication of case C009. Therefore, C197 has been omitted from the dataset, resulting in a total of 49 chest cases.
- The projection data for the sequential head scans are contained in one directory for each dose level for each patient. DICOM tag (0020, 0012) Acquisition Number, which is a number identifying the single continuous gathering of data over a period of time (i.e., one scan), can be used to sort the data based on each sequential acquisition within the exam. A MATLAB code that will sort the data in this manner has been provided here. The .m files must be placed in the same directory containing the projection files to correctly identify and sort them.

The previous version of the GE data has been replaced with the new data (Version 4); click here to download a spreadsheet that maps the old UIDs to the new UIDs. You can access the updated series by downloading this manifest. Please contact the TCIA Helpdesk at help@cancerimagingarchive.net with any questions.

Version 3: Updated 2022/01/26

Data Type	Download all or Query/Filter
Images (DICOM, 953 GB)	*Download Search* (Requires NBIA Data Retriever.)
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*
Matlab DICOM-CTPD data dictionary (.txt)	*Download*
Matlab DICOM-CTPD reader script (.txt)	*Download*
Clinical Data (CSV, zip)	*Download*

A fractional shift in pixel positions was identified between the full dose and low dose Siemens' patient cases included in Version 1 and 2. In the DICOM tag called "image position patient", the full dose positions are written with seven digits after the decimal while the low dose positions have only two digits after the decimal and, in some cases, there is an offset close to 0.5 mm. This occurred due to having the full dose (original) data reconstructed on the scanner at the time of the patient exam and the low dose data reconstructed at a later time (after noise was inserted in the projection data) using Siemens' off-line reconstruction tools. Note that this shift is a consequence of a difference in the reconstruction tools used. **The projection data are not affected.** The shift primarily impacted those using the full dose and low dose images to train machine learning algorithms.

The Siemens image data have been reprocessed so that the pixel shift issue has been addressed. The DICOM tag "image patient position" is now identical for the full dose and low dose images. The previous version of the data has been replaced with the new data (Version 3); click here to download a spreadsheet that maps the old UIDs to the new UIDs. You can access the updated series by downloading this manifest.

Version 2: Updated 2020/08/11

Data Type	Download all or Query/Filter
Images (DICOM, 953 GB)	*Search*
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*
Matlab DICOM-CTPD data dictionary (.txt)	*Download*

Matlab DICOM-CTPD reader script (.txt)	*Download*
Clinical Data (CSV, zip)	*Download*

Important information about the GE patient cases.

An incorrect value was found in the RescaleIntercept DICOM tag (0028,1052) in the GE data. Additionally, we have been asked by users to determine and add PhotonStatistics values to DICOM tag (7033,1065). Hence, access to the GE data (149 cases) has been temporarily suspended while we address these issues.

- If you have already downloaded the GE data, you will want to download the updated data when it is again available.
- The GE data are readily recognized, as they have only 3 series per patient case (Siemens data have 4).
- The GE data can also be identified using the DICOM tag (0008,0070).
- If you have already started a project with the GE data, please contact the Mayo team to discuss the issues in greater detail to discern if they impact your work. Please send your inquiry to CTCIC@mayo.edu and the team will follow up with you.

Version 1: Updated 2020/04/22

Data Type	Download all or Query/Filter
Images (DICOM, 1.23 TB)	*Search*
Images (DICOM, 2.0 GB) Phantom Object Only	*Download*
DICOM-CT-PD User Manual Version 3 (.pdf)	*Download*
Matlab DICOM-CTPD data dictionary (.txt)	*Download*
Matlab DICOM-CTPD reader script (.txt)	*Download*
Clinical Data (CSV, zip)	*Download*