Crowds Cure Cancer: Data collected at the RSNA 2017 annual meeting (Crowds-Cure-2017)

Description

Many Cancers routinely identified by imaging haven't yet benefited from recent advances in computer science. Approaches such as machine learning and deep learning can generate quantitative tumor 3D volumes, complex features and therapy-tracking temporal dynamics. However, cross-disciplinary researchers striving to develop new approaches often lack disease understanding or sufficient contacts within the medical community. Their research can greatly benefit from labeling and annotating basic information in the images such as tumor locations, which are obvious to radiologists.

Crowd-sourcing the creation of publicly-accessible reference data sets could address this challenge. In 2011 the National Cancer Institute funded development of The Cancer Imaging Archive (TCIA), a free and open-access database of medical images. However, most of these collections lack the labeling and annotations needed by image processing researchers for progress in deep learning and radiomics. As a result, TCIA has partnered with the Radiological Society of North America (RSNA) and numerous academic centers to harness the vast knowledge of RSNA meeting attendees to generate these tumor markups. Data sets annotated included CT scans from 352 subjects from the The Cancer Genome Atlas Lung Adenocarcinoma Collection (TCGA-LUAD), The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma Collection (TCGA-KIRC), The Cancer Genome Atlas Liver Hepatocellular Carcinoma Collection (TCGA-LIHC), and The Cancer Genome Atlas Ovarian Cancer Collection (TCGA-OV) collections on TCIA.

A full explanation of the project can be seen in the Detailed Description tab.

Data Access

Data Type	Download all or Query/Filter	License
Image Annotations (CSV, 925 kb)	*Download*	CC BY 3.0
DICOM-SR files see note (ZIP, 3.7 Mb)	*Download*	CC BY 3.0
Clinical Data snapshot see note (CSV, 53kb)	*Download*	CC BY 3.0

Collections Used in this Third Party Analysis

Below is a list of the Collections used in these analyses:

Source Data Type	Download all or Query/Filter	License
Corresponding original source Images from TCGA-LUAD, TCGA-KIRC, TCGA-LIHC, TCGA-OV (DICOM, 24.2 GB)	*Download* (Requires NBIA Data Retriever)	CC BY 3.0

- The Cancer Genome Atlas Lung Adenocarcinoma Collection (TCGA-LUAD)
- The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma Collection (TCGA-KIRC)
- The Cancer Genome Atlas Liver Hepatocellular Carcinoma Collection (TCGA-LIHC)
- The Cancer Genome Atlas Ovarian Cancer Collection (TCGA-OV)

Detailed Description

- **DICOM-SR note**: The conversion XSLT and Makefile depends on pixelmed.jar as a DICOM toolkit, and dicom3tools, dcsrdump and dciodvfy for validation.
- **Clinical data note:** Because all subjects were pulled from The Cancer Genome Atlas cohorts, clinical data was available through the NCI Genomic Data Commons. A CSV dump of that data is provided here for convenience.

Booth posters

1

BACKGROUND

Many cancers routinely identified by imaging haven't yet benefited from recent advances in computer science. Approaches such as machine learning and deep learning can generate quantitative tumor 3D volumes, complex features and therapy-tracking temporal dynamics. However, cross-disciplinary researchers striving to develop new approaches often lack disease understanding or sufficient contacts within the medical community. Their research can greatly benefit from labeling and annotating basic information in the images such as tumor location, which are obvious to radiologists.

"Crowdsourcing" the creation of publicly-accessible reference data sets could address this challenge. In 2011 the National Cancer Institute funded development of The Cancer Imaging Archive (TCIA), a free and open-access database of de-identified medical images. Many of its data-sets are cross-linked to extensive genetic/proteomic and clinical data gathered during the NIH's The Cancer Genome Atlas (TCGA) as well as other sources. However, most of these collections lack the labeling and annotations needed by computer image researchers for progress in deep learning and radiomics.

With this booth the NCI and RSNA seek to harness the vast knowledge of meeting attendees to generate these tumor markups. Data resulting from this experiment will be openly shared with the radiology and the computer science community.

HTTP://BIT.LY/CROWDCURE

HELP US LOCATE TUMORS IN CANCER IMAGING DATA

2

PARTICIPATE
IN THIS CROWD-SOURCING EXPERIMENT

Unlabeled data from The Cancer Imaging Archive has been loaded into a web-based annotation environment. It contains numerous subjects with proven cancer diagnoses spanning a variety of cancer types (brain, renal, lung, etc.). The tumor is usually fairly obvious to an experienced imager.

Find the slice that shows the maximum dimension of the tumor. With the linear pencil tool draw a line across the tumor's maximum diameter and save your work. Any individual case will probably take you less than a minute or two. It's just like a "RECIST" measurement commonly used in clinical trials.

Don't agonize about getting it perfect. Do as many or as few cases as you like – but of course more would always be welcome! The booth docent would enjoy hearing your comments.

Let us know if you want to stay in the loop! Markups from this booth are saved and will be made public to the broad image science community. Leave us your email address when you register to label cases if you'd like to be notified when the results are released. Or sign up for one of TCIA's social media options (LinkedIn, Twitter, Facebook) to get push notifications any time new data sets are added to the site.
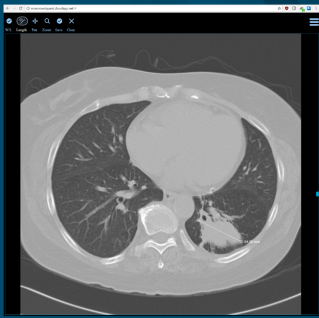
HTTP://BIT.LY/CROWDCURE

HELP US LOCATE TUMORS IN CANCER IMAGING DATA

3

DRAW YOUR MEASUREMENT

HTTP://BIT.LY/CROWDCURE

HELP US LOCATE TUMORS IN CANCER IMAGING DATA

4

THANKS FOR YOUR PARTICIPATION

As a token of appreciation for participation you will receive a badge ribbon. It may stimulate some further interesting conversations with your colleagues.

At any time you can explore the NCI Cancer Imaging Archive large image data collections and their linked clinical and genetic/proteomic metadata. We encourage you to check out The Cancer Imaging Archive (TCIA) to learn about the other available data on the site, and to stop by our hands-on workshop on Thursday from 2:30-4:00 in S401CD.

http://**www.cancerimagingarchive.net**

Exhibit Contributors:
Frederick National Laboratory for Cancer Research: Justin Kirby, Carl Jaffe, Brenda Fevrier-Sullivan and John Freymann
Massachusetts General Hospital: Artem Mamonov, Andrew Beers, Jayashree Kalpathy-Cramer
University of Arkansas for Medical Sciences: Lawrence Tarbox, Jeff Tobler, Fred Prior
Emory University: Ashish Sharma
Quantitative Image Informatics for Cancer Research (QIICR): Steve Pieper and the QIICR team
Open Health Imaging Foundation (OHIF): André Botelho Almeida, Erik Ziegler, Rob Lewis, Gordon Harris and the OHIF team
RSNA Radiology Informatics Committee

HTTP://BIT.LY/CROWDCURE

HELP US LOCATE TUMORS IN CANCER IMAGING DATA

Citations & Data Usage Policy

Users of this data must abide by the TCIA Data Usage Policy and the Creative Commons Attribution 3.0 Unported License under which it has been published. Attribution should include references to the following citations:

Data Citation

Kalpathy-Cramer, J., Beers, A., Mamonov, A., Ziegler, E., Lewis, R., Almeida, A. B., Harris, G., Pieper, S., Sharma, A., Tarbox, L., Tobler, J., Prior, F., Flanders, A., Dulkowski, J., Fevrier-Sullivan, B., Jaffe, C., Freymann, J., & Kirby, J. (2019). **Crowds Cure Cancer: Crowdsourced data collected at the RSNA 2017 annual meeting [Data set]**. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2018.OW73VLO2

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**. Journal of Digital Imaging, 26(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

Other Publications Using This Data

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

Version 1 (Current): 2018/05/17

Data Type	Download all or Query/Filter
Images (DICOM, 24.2 GB)	**⊙ Download** (Open this *.tcia manifest with NBIA Data Retriever)
Image Annotations (CSV)	**⊙ Download**
DICOM-SR files (ZIP) *	**⊙ Download**
Clinical Data (CSV) **	**⊙ Download**

* The conversion XSLT and Makefile depends on pixelmed.jar as a DICOM toolkit, and dicom3tools, dcsrdump and dciodvfy for validation.

** Because all subjects were pulled from The Cancer Genome Atlas cohorts, clinical data was available through the NCI Genomic Data Commons. A CSV dump of that data is provided here for convenience.