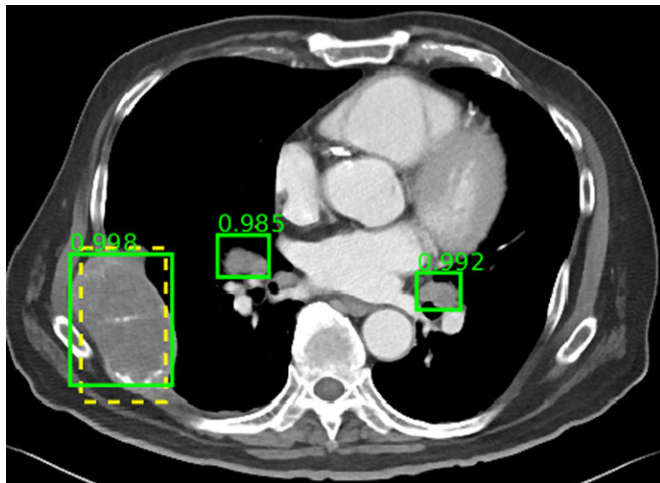


# SPIE journal announces public access to largest multi-lesion medical imaging dataset

20 July 2018



The ground-truth and two enlarged lymph nodes are correctly detected, even though the lymph nodes are not annotated in the dataset. Credit: @SPIE

A paper published today in the *Journal of Medical Imaging* - "DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning,—announced the open availability of the largest CT lesion-image database accessible to the public. Such data are the foundations for the training sets of machine-learning algorithms; until now, large-scale annotated radiological image datasets, essential for the development of deep learning approaches, have not been publicly available.

DeepLesion, developed by a team from the National Institutes of Health Clinical Center, was developed by mining historical medical data from their own Picture Archiving and Communication System. This new [dataset](#) has tremendous potential to jump-start the field of computer-aided detection (CADe) and diagnosis (CADx).

The [database](#) includes multiple lesion types, including kidney lesions, bone lesions, lung

nodules, and enlarged lymph nodes. The lack of a multi-category lesion dataset to date has been a major roadblock to development of more universal CADe frameworks capable of detecting multiple lesion types. A multi-category lesion dataset could even enable development of CADx systems that automate radiological diagnosis.

The database is built using the annotations—bookmarks—of clinically meaningful findings in medical images from the image archive. After analyzing the characteristics of these bookmarks—which take different forms, including arrows, lines, ellipses, segmentation, and text—the team harvested and sorted those bookmarks to create the DeepLesion database.

Whereas the field of computer vision has access to the robust ImageNet3 dataset, which contains millions of images, the medical imaging field has not had access to the same quantity of data. Most publicly available medical image datasets contain just tens or hundreds of cases. With over 32,000 annotated lesions from over 10,000 case studies, the DeepLesion dataset is now the largest publicly available medical image dataset.

"We hope the dataset will benefit the medical imaging area just as ImageNet benefited the computer vision area," says Ke Yan, the lead author on the paper and a postdoctoral fellow in the laboratory of senior author Ronald Summers, MD, Ph.D.

In addition to building the database, the team also developed a universal lesion detector based on the database. The researchers note that lesion detection is a time-consuming task for radiologists, but a key part of diagnosis. This detector may be able to serve as an initial screening tool for radiologists or other specialist CADe systems in the future.

In addition to lesion detection, the DeepLesion

database may also be used to classify lesions, retrieve [lesions](#) based on query strings, or predict lesion growth in new cases based on existing patterns in the database. The database can be downloaded at <https://nihcc.box.com/v/DeepLesion>.

Future work will include extending the database to other image modalities, like MR, including data from multiple hospitals, and improving the detection accuracy of the detector algorithm.

**More information:** Ke Yan et al, DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning, *Journal of Medical Imaging* (2018). DOI: [10.1117/1.JMI.5.3.036501](https://doi.org/10.1117/1.JMI.5.3.036501)

Provided by SPIE

APA citation: SPIE journal announces public access to largest multi-lesion medical imaging dataset (2018, July 20) retrieved 20 September 2021 from <https://medicalxpress.com/news/2018-07-spie-journal-access-largest-multi-lesion.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*