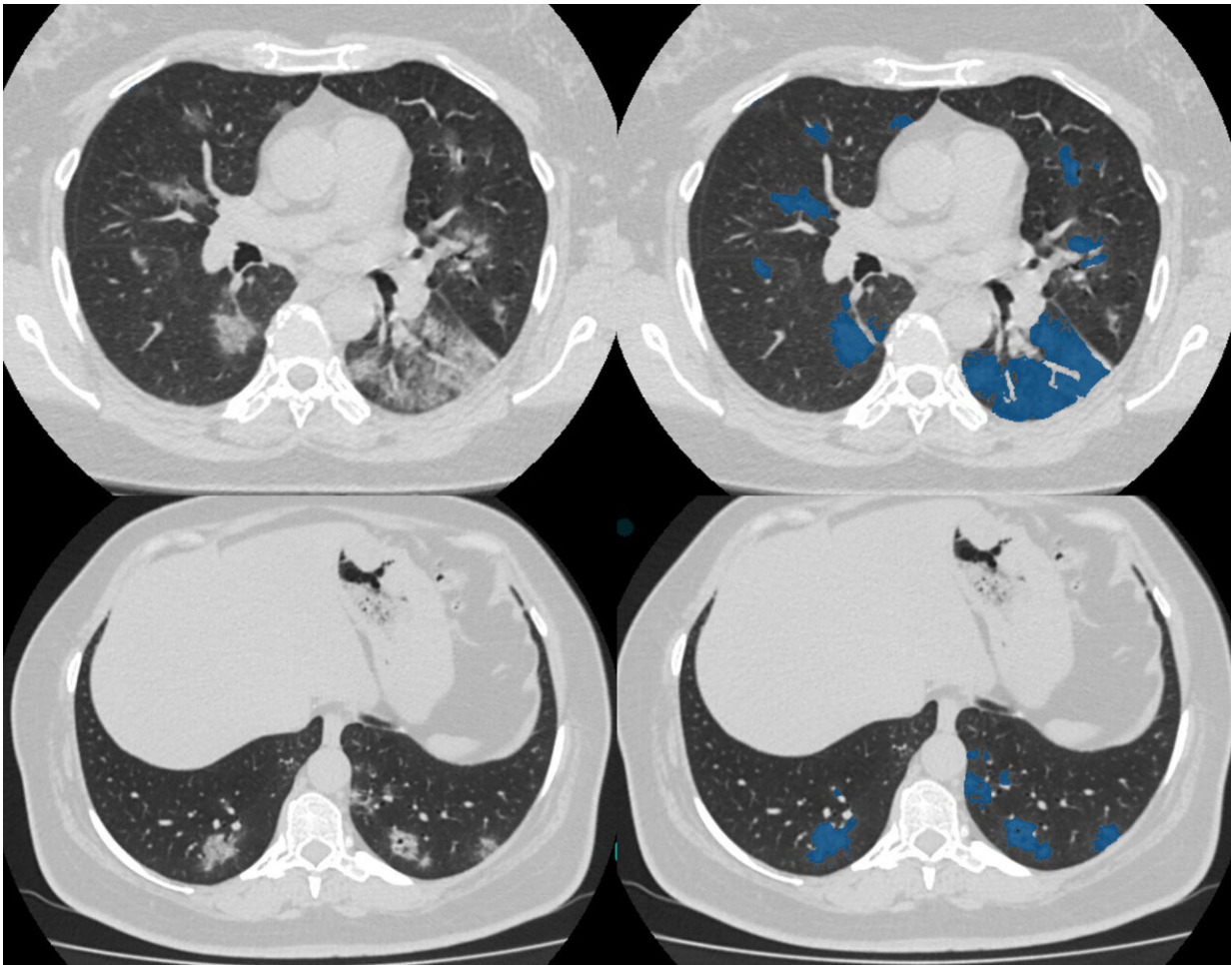# CT scan database of 1000 sets was created for teaching AI to diagnose COVID-19

May 11 2020

Center's experts marked explicitly 50 studies (5% of the total array), where pixels' zones of ground glass opacities and consolidations, specific for COVID-19 are indicated on each CT slice with lung tissue abnormalities. Credit: Moscow Diagnostics and Telemedicine Center

Scientists of the Diagnostics and Telemedicine Center report that they have expanded the original database of CT studies of patients with laboratory-confirmed COVID-19 infection by 20 times. It now contains more than 1,000 anonymized sets of chest CT scans. The studies were collected in Moscow from March 1 to April 25, 2020 using the Unified Radiological Information Service (URIS) involving the diagnostic equipment of 80 Moscow healthcare institutions.

The database has no analogs anywhere in the world. For example, the dataset collected at the University of San Diego has 349 CT scans (single) of 216 patients, while the dataset collected in Moscow contains three-dimensional CT studies. The set of RAIOSS & Livon Saúde's cases contains 10 CT scans so far. There are more than 70 scans in the constantly updated database of the Italian Radiological Society. The Radiological Society of North America's collection of new coronavirus infection cases is scattered and suitable only for familiarization. The British Society of Thoracic Radiology also has a database, but it also does not contain more than a hundred studies.

The number of cases is not the only fundamental difference between the Russian database and foreign ones. All CT studies in the Moscow dataset have a special marking. This marking is made according to the classification, reflecting manifestation of pathological abnormalities of COVID-19 in the lung tissue based on the chest computed tomography. It divides the studies into five large groups: from CT-0 (normal and absence of CT signs of viral pneumonia) to CT-4 (diffuse ground glass opacities, pulmonary parenchymal involvement more than 75%).

According to experts of the Diagnostics and Telemedicine Center, a database with CT scans converted into the research NIFTI format is intended for developing artificial intelligence algorithms. Holistic marking of cases is suitable for preparing automatic patient sorting systems. The marking of localizations (those areas of interest within

which artificial intelligence algorithms should detect pathology) can be used in training services to help radiologists by indicating suspicious sites on CT scans. Marking the pathology contouring can be used for automatic quantitative assessment of lung lesions, as well as for assessing dynamics between two CT studies of a patient.

In addition, the center's experts noted 50 studies (5% of the total array) indicating the pixel zones of ground glass opacities and consolidations specific for COVID-19 on each CT slice with lung tissue abnormalities. It is the most informative type of marking of CT scan images for artificial intelligence.

"The additional advantage of this dataset is that all CT scans included were performed in primary healthcare facilities for the adult population. Besides that, it has been posted in public domain, and thin CT slices of up to 1 mm have already been converted into the NIFTI format recognized among machine learning professionals, " said Sergey Morozov, chief regional radiology and instrumental diagnostics officer of the Moscow Department of Health, CEO of Diagnostics and Telemedicine Center, Moscow.

The creation of the Russian dataset of computer tomograms of patients with the signs of COVID -19 was part of a large Moscow experiment on the use of computer vision in radiation diagnostics, which started in February and will last until the end of this year. All detailed information can be found on the project website.

Provided by Moscow Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies

scan-database-ai-covid-.html