CENTER OF DIAGNOSTICS AND TELEMEDICINE

MosMedData: chest XR with and without pulmonary nodules type VII

About Dataset

> This dataset contains the results of chest XR with and without (normal) pulmonary nodules . The research data were collected in the radiology departments of the Moscow medical facilities between 16.11.2017 and 06.04.2022. Two radiologists reviewed the studies and verified the presence of pulmonary nodules. (The Fleischner Society glossary defines "pulmonary nodule" as a tumor of 6 to 30 mm in size, located in the lung parenchyma). To make the dataset more balanced, we added chest CT studies w/o pathological findings. An expert (a radiologist with more than 10 years of experience) did a final review of the dataset and verified the studies based on the CT results (within 2 weeks after RG). The final dataset contains 100 studies: 50 w/ pathological changes (i.e., lung nodules); the other 50 w/o pathological changes.

Disclaimer
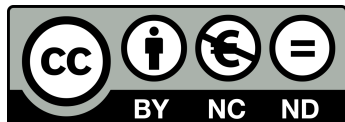
> Scope of use of this dataset:
>
> - development, training and validation of AI-powered services (incl. those based on the computer vision technologies) capable of detecting the pulmonary nodules;
> - rising awareness among health professionals and the general public.
>
> The license permits **free use (sharing)** of this dataset which includes making copies and sharing on any medium or in any format, **under the following conditions**:
>
> - the dataset must contain a declaration of original authorship which includes:
>   - contributors;
>   - affiliations;
>   - copyright notice;
>   - permanent link to the dataset.
> - link to the license.

The license **forbids** the following:

> - using this dataset for commercial purposes;
> - sharing a modified or transformed dataset or any new dataset built upon the dataset;
> - providing a fee-based access to the dataset;
> - preventing the sharing of the dataset through any technical methods.

General Information

Dataset Title

MosMedData: chest XR with and without pulmonary nodules type VII

Internal Code

LUNGNODULE

Annotation Class

C-1

Verification

Review by a specialist, investigation of another modality (CT)

Keywords

MosMedData, chest, XR, x-ray, radiology, pulmonary nodes, lung malignancies, cancer, lung

Language

English, Russian

Funding Sources

This dataset was prepared by a group of authors as a part of the medical research project (No. USIS (in the Unified State Information System for Accounting of Research, Development, and Technological Works): 122112400040-1) "Reference datasets for sustainable development of artificial intelligence-based diagnostics to minimize the long-term impact of the COVID-19 pandemic on the health of the Moscow population".

Dataset Version

1.0.0

Permanent link

https://mosmed.ai/datasets/

Release Date

26.07.2023

Affiliation and Contributors

Contributors

- Vasilev Y.A. [1]
- Vladzymyrskyy A.V. [1]
- Arzamasov K.M. [1]
- Busygina Y.S. [1]
- Pankratov A.V. [1]
- Shihmuradov D.U. [1]
- Kokina D.Ju. [1]
- Sahaschik U.A. [1]
- Stetsjuk L.D. [1]
- Novik V.P. [1]
- Chetverikov S.F. [1]
- Bobrovskaja T.M. [1]
- Semenov S.S. [1]

Affiliation

[1] Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department.

Data Structure

```
.
|-- dataset_registry.xlsx
|-- README_EN.md
|-- README_RU.md
|-- README_EN.pdf
|-- README_RU.pdf
`-- studies
    |--studyUID_X
    |       |-- seriesUID_X
    |       |        |-- UID_X.dcm
    |       |        |-- UID_X.dcm
    |       |        `-- ...
    |       |
    |       `-- seriesUID_X
    |               |-- UID_X.dcm
    |               |-- UID_X.dcm
    |               `-- ...
    |--studyUID_X
    |       |-- seriesUID_X
    |       |        |-- UID_X.dcm
    |       |        |-- UID_X.dcm
    |       |        `-- ...
    |       |
```

```
    |     `-- seriesUID_X
    |          |-- UID_X.dcm
    |          |-- UID_X.dcm
    |          `-- ...
    `--...
```

- README_EN.md and README_RU.md contain general information about the dataset in the Markdown format in English and Russian languages, respectively. README_EN.pdf and README_RU.pdf contain the same information in the PDF format.
- dataset_registry.xlsx contains a list of studies included in the dataset and the path to the corresponding dataset file.
- The Studies folder contains studyUID-X folders, each of which contains studies in DICOM format.

Data Overview

Property	Value
Number of studies, pcs.	100
Number of patients, ppl.	100
Distribution by sex,ppl (M/ F)	47/ 51
Distribution by age, years (min./ median/ max.)	18/ 58/ 87
Number of studies in each category, psc. (With pathology/ without pathological changes)	50/ 50

Sharing and Access Information

License

Copyright © 2020 Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department. This dataset is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) License. For more information see the LICENSE file or follow this link for more information.

Citation

Recommended citation:

Database registration certificate No. 2023622421 Russian Federation. MosMedData: chest XR with and without pulmonary nodules : No. 2023621436 : Appl. 05/19/2023 : publ. July 17, 2023 / Y. A. Vasiliev, A. V. Vladzimirskyy, K. M. Arzamasov [and others]; applicant: Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department.

Distribution

> It is prohibited to share this dataset without specifying:
>
> - contributors;
> - affiliations;
> - copyright notice;
> - permanent link to the dataset;
> - link to the license.