

# MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19



РАДИОЛОГИЯ МОСКВЫ  
ДИАГНОСТИКА БУДУЩЕГО

Данный набор данных (датасет) содержит результаты компьютерной томографии органов грудной клетки с рентгенологическими признаками вирусной пневмонии (COVID-19), а также без признаков (норма). Для некоторых исследований представлена разметка областей интереса (зон уплотнений по типу «матового стекла» и консолидации) в виде бинарной пиксельной маски. Данные исследований были собраны в отделениях лучевой диагностики лечебных учреждений города Москвы в период с 01.03.2020 по 25.04.2020.

## DISCLAIMER

Набор данных предназначен для следующих целей:

- разработка, дообучение и тестирование программных продуктов (использующих в том числе методы компьютерного зрения), выявляющих признаки, характерные для коронавирусной инфекции (COVID-19);
- информирование медицинского сообщества и общественности в целом.

Лицензия **позволяет свободно делиться (обмениваться)** набором данных, то есть копировать и распространять материал на любом носителе и в любом формате, **при обязательном соблюдении следующих условий:**

- указано авторство, а именно:
  - авторы;
  - их организации;
  - правообладатель (копирайт);
  - постоянная ссылка на оригинальный набор данных.
- указана ссылка на лицензию.

Лицензия **запрещает**, в том числе:

- использовать набор данных в коммерческих целях;
- распространять переработанный, преобразованный набор данных или новые наборы данных, созданные на основе этого набора;
- накладывать ограничения поверх существующих ограничений, указанных в лицензии, например:
  - предоставлять платный доступ к набору данных,
  - искусственно сдерживать распространение набора данных техническими методами.



## Общая информация

### Название набора данных

MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19

## Внутренний код

COVID19\_1110

## Классы разметки

2-C, 2-A

## Ключевые слова

компьютерная томография, КТ, органы дыхательной системы, вирусная, инфекция, легкие, грудная клетка, COVID-19

## Язык

Английский, русский

## Финансирование

Внутреннее финансирование

## Версия набора данных

1.0

## Постоянная ссылка

[https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110)

## Дата публикации

28.04.2020

## Аффилиация и авторы

---

### Авторы

- Морозов Сергей Павлович<sup>1</sup>
- Андрейченко Анна Евгеньевна<sup>1</sup>
- Блохин Иван Андреевич<sup>1</sup>
- Владзимирский Антон Вячеславович<sup>1</sup>
- Гележе Павел Борисович<sup>1</sup>
- Гомболевский Виктор Александрович<sup>1</sup>
- Гончар Анна Павловна<sup>1</sup>
- Ледихова Наталья Владимировна<sup>1</sup>
- Павлов Николай Александрович<sup>1</sup> ([n.pavlov@npcmr.ru](mailto:n.pavlov@npcmr.ru))
- Чернина Валерия Юрьевна<sup>1</sup>

### Аффилиация

1. Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»

## Структура набора данных

```
.
|-- dataset_registry.xlsx
|-- LICENSE
|-- README_EN.md
|-- README_RU.md
|-- README_EN.pdf
|-- README_RU.pdf
|-- masks
|   |-- study_BBBB_mask.nii.gz
|   |-- ...
|   `-- study_BBBB_mask.nii.gz
`-- studies
    |-- CT-0
    |   |-- study_BBBB.nii.gz
    |   |-- ...
    |   `-- study_BBBB.nii.gz
    |-- CT-1
    |   |-- study_BBBB.nii.gz
    |   |-- ...
    |   `-- study_BBBB.nii.gz
    |-- CT-2
    |   |-- study_BBBB.nii.gz
    |   |-- ...
    |   `-- study_BBBB.nii.gz
    |-- CT-3
    |   |-- study_BBBB.nii.gz
    |   |-- ...
    |   `-- study_BBBB.nii.gz
    `-- CT-4
        |-- study_BBBB.nii.gz
        |-- ...
        `-- study_BBBB.nii.gz
```

- README\_EN.md и README\_RU.md содержат общую информацию о наборе данных в формате Markdown на английском и русском языках соответственно; та же информация в формате PDF представлена в README\_EN.pdf и README\_RU.pdf.
- LICENSE содержит описание лицензии Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) License.
- dataset\_registry.xlsx содержит перечень исследований, включенных в набор данных, путь к соответствующему файлу и путь к маске (при наличии).
- В директории studies находятся директории CT-0, CT-1, CT-2, CT-3 и CT-4, в каждой из которых содержатся исследования в формате NIfTI, заархивированные в Gzip. Названия исследований построены по шаблону study\_BBBB.nii.gz, где BBBB – уникальный порядковый номер исследования во всем наборе данных (сквозная нумерация).
- В директории masks находятся бинарные маски разметки в формате NIfTI, заархивированные в Gzip. Названия масок построены по шаблону study\_BBBB\_mask.nii.gz, где BBBB – порядковый номер соответствующего исследования.

## Обзор данных

Параметр	Значение
Количество исследований, ед.	1110
Количество пациентов, чел.	1110
Распределение по полу, % (М/ Ж/ др.)	42/ 56/ 2
Распределение по возрасту, лет (мин./ медиана/ макс.)	18/ 47/ 97

Параметр	Значение
Количество бинарных пиксельных масок разметки класса А, ед.	50
Распределение по классам (разметка класса С), ед. (КТ-0/ КТ-1/ КТ-2/ КТ-3/ КТ-4)	254/ 684/ 125/ 45/ 2

## Особенности подготовки исследований

- Одно исследование относится к одному пациенту.
- Каждое исследование включает одну трехмерную реконструкцию в мягкотканном режиме.

```
SeriesDescription LIKE '%BODY%'
```

- При преобразовании формата DICOM в NIfTI в серии сохранено каждое 10-е изображение.

```
InstanceNumber % 10 = 0
```

## Принцип разметки класса С

Исследования разделены на [5 категорий](#)<sup>1</sup>:

- **КТ-0** (директория `/studies/CT-0`): норма и отсутствие КТ-признаков вирусной пневмонии.
- **КТ-1** (директория `/studies/CT-1`): зоны уплотнения по типу «матового стекла». Вовлечение паренхимы легкого  $\leq 25\%$ .
- **КТ-2** (директория `/studies/CT-2`): зоны уплотнения по типу «матового стекла». Вовлечение паренхимы легкого  $= 25-50\%$ .
- **КТ-3** (директория `/studies/CT-3`): зоны уплотнения по типу «матового стекла» и консолидации. Вовлечение паренхимы легкого  $= 50-75\%$ .
- **КТ-4** (директория `/studies/CT-4`): Диффузное уплотнение легочной ткани по типу «матового стекла» и консолидации в сочетании с ретикулярными изменениями. Вовлечение паренхимы легкого  $\geq 75\%$ .

1. Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов : препринт № ЦДТ – 2020 – II. Версия 2 от 17.04.2020 / сост. С. П. Морозов, Д. Н. Проценко, С. В. Сметанина [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 65. – М. : ГБУЗ «НПКЦ ДиТ ДЗМ», 2020. – 78 с.

## Принцип разметки класса А

Для ограниченного количества исследований в наборе данных (50 исследований) экспертами НПКЦ ДиТ ДЗМ созданы бинарные пиксельные маски областей интереса (зон уплотнений по типу «матового стекла» и консолидации) (директория `/masks/`). Маски сохранены в формате NIfTI и заархивированы в Gzip. Маски имеют те же координаты, что и соответствующее исследование.

При разметке использовалось программное обеспечение [MedSeg](#) (© 2020 Artificial Intelligence AS).

## Правила использования и распространения

### Лицензия

Copyright © 2020 Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы».

Набор данных доступен под лицензией Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) License. За подробной информацией обратитесь к файлу [LICENSE](#) или перейдите по [ссылке](#).

## Цитирование

Рекомендованная форма для цитирования на английском языке:

Morozov, S., Andreychenko, A., Blokhin, I., Vladzimirskyy, A., Gelezhe, P., Gombolevskiy, V., Gonchar, A., Ledikhova, N., Pavlov, N., Chernina, V. MosMedData: Chest CT Scans with COVID-19 Related Findings, 2020, v. 1.0, [https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110)

Рекомендованная форма для цитирования на русском языке:

Морозов С. П., Андрейченко А. Е., Блохин И. А., Владзимирский А. В., Гележе П. Б., Гомболевский В. А., Гончар А. П., Ледихова Н. В., Павлов Н. А., Чернина В. Ю. MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19, 2020 г., версия 1.0, [https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110)

## Распространение

Данный датасет не должен распространяться без указания:

- авторов;
- аффилиаций;
- правообладателя (копирайта);
- постоянной ссылки на оригинальный набор данных;
- ссылки на лицензию.