

# MosMedData: датасет 1110 компьютерных томографий органов грудной клетки, выполненных во время эпидемии COVID-19



© С.П. Морозов, А.Е. Андрейченко, И.А. Блохин, П.Б. Гележе, А.П. Гончар, А.Е. Николаев, Н.А. Павлов, В.Ю. Чернина, В.А. Гомболевский\*

ГБУЗ «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», Москва, Российская Федерация

В условиях пандемии COVID-19 и лавинообразного роста числа выполняемых компьютерных томографий (КТ) лёгких особое значение приобретают методы автоматизации процесса анализа изображений, использование которых позволит повысить производительность и минимизировать ошибки. Создание качественных наборов данных необходимо для развития технологий искусственного интеллекта. Алгоритмы искусственного интеллекта обладают достаточной точностью для диагностики COVID-19. Данный датасет<sup>1</sup> содержит как анонимизированные компьютерные томограммы (КТ) лёгких человека с признаками COVID-19, так и нормальные исследования грудной клетки. Некоторая часть исследований была размечена с использованием бинарных пиксельных масок представляющих интерес областей (например, зон консолидации и уплотнений по типу матового стекла). КТ-данные были получены в период с 1 марта 2020 г. по 25 апреля 2020 г. и предоставлены муниципальными больницами г. Москвы (Россия)<sup>2</sup>. Предлагаемый набор данных лицензирован Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0).

**Ключевые слова:** искусственный интеллект; COVID-19; машинное обучение; датасет; КТ; органы грудной клетки.

## Как цитировать

Морозов С.П., Андрейченко А.Е., Блохин И.А., Гележе П.Б., Гончар А.П., Николаев А.Е., Павлов Н.А., Чернина В.Ю., Гомболевский В.А. MosMedData: датасет 1110 компьютерных томографий органов грудной клетки, выполненных во время эпидемии COVID-19 // *Digital Diagnostics*. 2020;1(1):49–59. DOI: <https://doi.org/10.17816/DD46826>

<sup>1</sup> Data set — *набор данных (set of data)*, обработанная и структурированная информация, обычно в табличном виде, пригодная для статистического анализа, визуализации и обработки алгоритмами машинного обучения.

<sup>2</sup> Постоянная ссылка: [https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110).



# MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic

Sergey P. Morozov, Anna E. Andreychenko, Ivan A. Blokhin, Pavel B. Gelezhe, Anna P. Gonchar, Alexander E. Nikolaev, Nikolay A. Pavlov, Valeria Yu. Chernina, Victor A. Gombolevskiy\*

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Department of Health Care of Moscow, Moscow, Russian Federation

With the ongoing COVID-19 pandemic decreasing availability of polymerase chain reaction with reverse transcription and the snowballing growth of medical imaging, especially the number of chest computed tomography (CT) scans being performed, methods to augment and automate the image analysis, increasing productivity and minimizing human error are of particular importance. The creation of high-quality datasets is essential for the development and validation of artificial intelligence algorithms. Such technologies have sufficient accuracy in diagnosing COVID-19 in medical imaging. The presented large-scale dataset contains anonymized human CT scans with COVID-19 features as well as normal studies. Some studies were tagged by radiologists using binary pixel masks of regions of interest (e.g., characteristic areas of consolidation and ground-glass opacities). CT data were acquired between March 1, 2020, and April 25, 2020, and provided by municipal hospitals in Moscow, Russia. The presented dataset is licensed under Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0).

**Keywords:** artificial intelligence; COVID-19; machine learning; dataset; CT, chest.

## To cite this article

Morozov SP, Andreychenko AE, Blokhin IA, Gelezhe PB, Gonchar AP, Nikolaev AE, Pavlov NA, Chernina VYu, Gombolevskiy VA. MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic. *Digital Diagnostics*. 2020;1(1):49–59. DOI: <https://doi.org/10.17816/DD46826>

DOI: <https://doi.org/10.17816/DD46826>

Received: 12.10.2020

Accepted: 11.12.2020

Published: 21.12.2020



# MosMedData: COVID-19疫情期间进行的1110次胸部CT扫描数据集

Sergey P. Morozov, Anna E. Andreychenko, Ivan A. Blokhin, Pavel B. Gelezhe, Anna P. Gonchar, Alexander E. Nikolaev, Nikolay A. Pavlov, Valeria Yu. Chernina, Victor A. Gombolevskiy\*

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Department of Health Care of Moscow, Moscow, Russian Federation

在COVID-19大流行和雪崩式增加肺部计算机断层扫描的数量背景下，图像分析过程的自动化方法特别重要，使用这种方法将提高生产率并减少错误。高质量数据集的创建是人工智能技术发展的必要条件。人工智能算法对COVID-19的诊断具有足够的准确性。该数据集1包含有COVID-19征象的患者的匿名肺部CT图像和正常的胸部检查。一些研究使用感兴趣区域的二元像素遮罩进行标记（例如，肺结节整合和磨砂玻璃结节）。获取2020年3月1日至2020年4月25日期间的CT数据，提供给莫斯科市医院（俄罗斯）<sup>2</sup>。建议的数据集由Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported授权（CC BY-NC-ND 3.0）。

**关键词：**人工智能；COVID-19；机器学习；数据集；CT；胸部器官

## 引用本文：

Morozov SP, Andreychenko AE, Blokhin IA, Gelezhe PB, Gonchar AP, Nikolaev AE, Pavlov NA, Chernina VYu, Gombolevskiy VA. MosMedData: COVID-19疫情期间进行的1110次胸部CT扫描数据集. *Digital Diagnostics*. 2020;1(1):49-59. DOI: <https://doi.org/10.17816/DD46826>

<sup>1</sup> Data set是数据集 (set of data)，经过处理的结构化信息，通常以表格形式，适用于机器学习算法的统计分析、可视化和处理。

<sup>2</sup> 永久链接: [https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110)



# 论证

在大多数国家COVID-19大流行期间，卫生机构承受了额外的沉重负担。这种情况需要比以往更加谨慎地使用财政和人力资源。不幸的是，医疗机构采取的预防措施并不总是足以避免医务人员的死亡。特别令人关注的是在紧急护理、放射学和其他紧急医学方面训练有素的专业人员的流失。计算机断层扫描(CT)被认为是诊断和评估COVID-19进展的关键工具，是在门诊基础上进行的，适用于有急性呼吸系统症状的患者，以及诊断明确、病情轻微的患者，在医生的监督下，利用远程医疗技术在家中进行治疗。在住院医疗机构中，CT用于初级和鉴别诊断、评估疾病进展和确定进一步的患者管理策略——在重症监护病房或在基层医疗机构医生的监督下在家[1-3]。

## 扫描、图像重建和数据库保存技术

CT的使用越来越多，给医疗保健系统带来了沉重的负担。例如，在莫斯科，一个市级门诊CT中心网络每天为每台计算机断层扫描机进行90次检查(CT记录器每天进行163次检查)。为了标准化和简化专家的临床决策，开发了一种分类模型，该模型与其他症状一起评估胸部CT扫描所观察到的肺组织异常的严重程度(表1)。因此，确定肺实质的损害程度可以预测COVID-19的死亡[4]。

医疗保健专业人员的职业倦怠和高死亡风险要求图像分析过程的自动化，这将提高生产率并将错误降到最低[5]。初步数据显示，人工智能算法诊断COVID-19具有足够的准确性(敏感性90%，特异性96%，AUC达到了0.96，总准确度76.37-98.26)[6,7]。

对同一模型的42张X线胸片进行胸部CT扫描(Toshiba Aquilion 64; Canon Medical Systems, 日本)。所有的研究都是根据标准方法学和制造商推荐的标准方案进行的(表2)。

一项研究涉及一名患者，包括一项3D重建。纳入研究的标准是患者在城市综合诊所的治疗事实，该诊所在疫情时被转换为门诊CT中心；由全科医生会诊，并由强制医疗保险基金支付费用，转诊做胸部器官的CT扫描。

排除标准为怀孕和年龄在18岁以下。CT前血氧检测低于93%的患者退出研究，转到救护车服务住院。

数据集包括5个阶段(图1)：

## 数据收集

数据收集是在2020年3月1日至4月25日期间在莫斯科市卫生保健中心的医疗机构进行的。这些医疗机构进行门诊预约：国家预算保健机构莫斯科市第19卫生局城市综合诊所、国家预算保健机构莫斯科市第214卫生局城市综合诊所、国家预算保健机构莫斯科市第52卫生局城市综合诊所、国家预算保健机构莫斯科市第23卫生局城市综合诊所、国家预算保健机构莫斯科市第6卫生局城市综合诊所、国家预算保健机构莫斯科市第5卫生局城市综合诊所、国家预算保健机构莫斯科市第3卫生局城市综合诊所、国家预算保健机构莫斯科市第209卫生局城市综合诊所、国家预算保健机构莫斯科市第9卫生局城市综合诊所、国家预算保健机构莫斯科市第62卫生局城市综合诊所、国家预算保健机构莫斯科市第4卫生局城市综合诊所、国家预算保健机构莫斯科市第218卫生局城市综合诊所、国家预算保健机构莫斯科市第175卫生局城市综合诊所、国家预算保健机构莫斯科市第212卫生局城市综合诊所、国家预算保健机构莫斯科市第170卫生局城市综合诊所、国家预算保健机构莫斯科市第191卫生局城市综合诊所、国家预算保健机构莫斯科市第8卫生局城市综合诊所、莫斯科市以M.P. Konchalovsky卫生保健部命名的国家卫生保健城市临床医院预算机构(门诊和住院接待)、国家预算保健机构莫斯科市第195卫生局城市综合诊所、国家预算保健机构莫斯科市第64卫生局城市综合诊所、国家预算保健机构莫斯科市第134卫生局城市综合诊所、国家预算保健机构莫斯科市第115卫生局城市综合诊所、国家预算医疗机构莫斯科市第1卫生部诊断临床中心、国家预算保健机构莫斯科市第67卫生局城市综合诊所、国家预算保健机构莫斯科市第121卫生局城市综合诊所、国家预算保健机构莫斯科市第36卫生局城市综合诊所、国家预算保健机构莫斯科市第68卫生局城市综合诊所、国家预算保健机构莫斯科市第2卫生部咨询和诊断中心、国家预算保健机构莫斯科市第11卫生局城市综合诊所、国家预算保健机构莫斯科市第180卫生局城市综合诊所、国家预算保健机构莫斯科市第45卫生局城市综合诊所、国家预算保健机构莫斯科市第5卫生局城市综合诊所、国家预算保健机构莫斯科市第5卫生局城市综合诊所(第一医院分支机构)、国家预算保健机构莫斯科市第2卫生局城市综合诊所、莫斯科东南行政区卫生部国家预算机构莫斯科结核病控制科学

表 1 COVID-19肺组织损伤严重程度的分类及处理规则

严重程度	CT分类	临床数据	决定
Zero KT-0	无肺炎症状 (包括COVID-19)	-	通知主治医师
轻度 (CT-1)	磨砂玻璃压实区有/无固结区, 铺路石征。 最小数量/流行率。 肺实质受累 $\leq$ 25%	$T < 38.0\text{ }^{\circ}\text{C}$ 呼吸速率 $< 20$ /分钟 $\text{SpO}_2 > 95\%$	利用远程医疗技术在家中 进行动态监测 (强制远程 监测健康状况)
中-重度 (CT-2)	磨砂玻璃压实区有/无固结区, 铺路石征。 平均数量/流行率。 肺实质受累25-50%	$T < 38.5\text{ }^{\circ}\text{C}$ 呼吸速率20-30/分钟 $\text{SpO}_2 \geq 95\%$	由初级保健组织的一名 医生在家中 进行动态监测
重度 (CT-3)	磨砂玻璃密封区。 固结带, 铺路石征。 显著的数量/流行率。 肺实质受累50-75%。 在呼吸系统疾病的背景下, 如果研究是动态进行的, 病 变体积在24-48小时内增加 50%	$\geq$ 发烧背景上的2个症 状: $T > 38.5\text{ }^{\circ}\text{C}$ $\text{RR} \geq 30$ /分钟 $\text{SpO}_2 \leq 93\%$ $\text{PaO}_2 / \text{FiO}_2 \leq 300\text{ mmHg}$ (1 mmHg=0.133 kPa)	立即入住指定的收治 COVID-19患者的住院 设施。 在医院环境下: 立即 转到ICU。 紧急CT (如果之前没 有进行过)
极重度 (CT-4)	磨砂玻璃型肺组织弥漫 性压实, 固结区伴铺路 石征。 胸水 (双侧, 以左侧为 主)。 小计数量/流行率。 肺实质受累 $\geq$ 75%	休克, 多器官衰竭, 呼吸 衰竭	提供紧急医疗救护。立 即入住指定的收治 COVID-19患者的住院 设施。 在医院环境下: 立即 转到ICU。 急诊CT (如果没之前 没有进行过, 而且病 人的情况满意)

注: CT — 计算机断层摄影; CT1-CT4 — 基于CT结果的肺损伤程度; RR—呼吸速率; ICU — 重症监护室; T—体温;  $\text{PaO}_2$  — 动脉血氧分压;  $\text{FiO}_2$  — 氧浓度。

表2扫描、图像重建和数据库保存技术

组类型	变数	意义和评论
设备	CT扫描仪	Toshiba Aquilion 64 (Canon Medical Systems, 日本)
	切片数量	64
病患	患者位置	胸部的位在龙门的中心。 调整工作台高度和定心, 使锁骨中线在等心。 手举过头顶。 呼吸指令。 病人教育和预扫描呼吸指导
	衣物及异物	所有可以移除的异物都要从扫描区域移除, 包括珠宝和项链。 内衣是可以接受的
	Localizer / scout / 射线照片*	为了将扫描范围限制在肺的范围内, 我们在胸部进行了扫描。 用于在扫描级别搜索可能降低质量的额外异物。 吸气深度屏气扫描
	扫描范围	全肺体积, 包括肺上5cm和肺下5cm
	呼吸阶段	CT扫描, 同时屏气吸气深度
	视野 Display Field of View (FOV)	离肋骨不少于1厘米 (从350毫米至500毫米)。 乳腺包括在扫描区域, 但可以部分排除在视野之外



组类型	变数	意义和评论
医务人员	X射线实验室助理	在控制室里未与患者接触。为了安全起见，与训练员面对面的接触已经减少到了最低限度
	处理人	处理人是辐射诊断部的一名医务人员，在疫情期间，根据莫斯科卫生部的命令，他从乳房x光摄影技术人员转到CT室。在控制室（放置患者并从手术台上抬起）和在走廊（扫描期间）。有人与病人有过接触。按照指导方针行事[8,9]
扫描和重建方案, 检查和解释	龙门倾斜	无
	扫描时间	≤10秒（通常为6秒）
	静脉注射对比	缺席
	口服对比	缺席
	电压	120 kV
	电流强度	Sure exp.3d自动电流调制系统内置在CT扫描仪的制造商。系统自动调整安培数，以达到噪音水平为10胡5.0毫米切片在80-500毫安范围。 XY调制在
	X射线管转速	0.5 s
	体积间距	95.0
	侦察的过程	QDS+
	重建CT序列的数目	2（含肺和软组织内核） <sup>3</sup>
	软组织重建核（单独CT系列）	FC07或FC18
	肺重建核（单独CT系列）	FC51
	切片厚度	1.0 mm（两个内核相同）
	切片厚度间的步骤	0.8 mm（两个内核相同）
迭代重建	AIDR 3D仅在5个层析图中发现，其余缺乏迭代重建算法，故采用FBP (filter back projection)	
对于CT解释，我们使用	AGFA Enterprise 8.0 Vitrea FX	
扫描和重建方案, 检查和解释	最大强度投影 (maximum intensity projections, MIP)、最小强度投影 (minimum intensity projections, MinIP)、多平面重建 (multiplanar reconstructions, MPR)	使用
人工智能算法	并不是对所有研究使用。 在应用时，机器学习算法向医生展示了一系列额外的研究，这些研究提供了样本环境，其中疑似COVID-19病变的存在仅限于红色矩形，吸引了医生的注意。此外，一个总结三维重建的肺病变区域标记为红色，并通过自动算法识别。没有定量的信息被提供来评估肺损伤的程度。	

<sup>3</sup> Kernel是一种用于CT数据重建的滤波器



组类型	变数	意义和评论
扫描和重建方案, 检查和解释	敲定协议的时间	10分钟到3小时在极少数情况下, 24小时
	协议标准化	该方案模板在方法学建议中形成和规范, 也被引入到统一放射信息服务中, 该方案是由放射科医生形成的
	COVID-19病变分类	根据CT0-CT4标度使用分类(见表1)
	第二个意见	对于来自综合诊所的90%的CT检查, 莫斯科卫生部科学和实用诊断和远程医疗技术临床中心的专家提供了第二种意见
数据库	辐射剂量计算	我们使用自动生成的DoseReport CT系列的DLP数据。在俄罗斯联邦, 根据方法学指南(MU 2.6.1.2944-11)为了计算有效剂量(mSv), 需要计算DLP和0.017(胸部CT扫描系数)的乘积
	收集研究	统一放射信息服务, 包括AGFA Enterprise 8.0
	初始数据收集格式	DICOM 3.0
	平面	轴平面
	切片厚度	1.0 mm
	切片厚度间的步骤	8.0 mm (每10切片保存一次)
	保存数据库格式	NIfTI
	肺病灶高亮的二元掩膜注释软件	MedSeg® (© 2020 Artificial Intelligence AS)

注: \*不在数据库中, 但对形成CT扫描是必要的。CT—计算机断层摄影, CT1-CT4—根据CT结果判断肺损伤程度。

和实践中心的分支机构、国家预算保健机构莫斯科市第46卫生局城市综合诊所、国家预算保健机构莫斯科市第166卫生局城市综合诊所、莫斯科中部和西部行政区卫生部莫斯科结核病控制科学和实践中心国家预算机构分支机构、国家预算保健机构莫斯科市第12卫生局城市综合

诊所、国家预算保健机构莫斯科市第220卫生局城市综合诊所、国家预算保健机构莫斯科市第66卫生局城市综合诊所、莫斯科第3保健部国家预算保健机构诊断中心。

本数据集(1110项研究)包含有COVID-19体征(CT1-CT4)和无体征(CT0)的人类肺部匿名

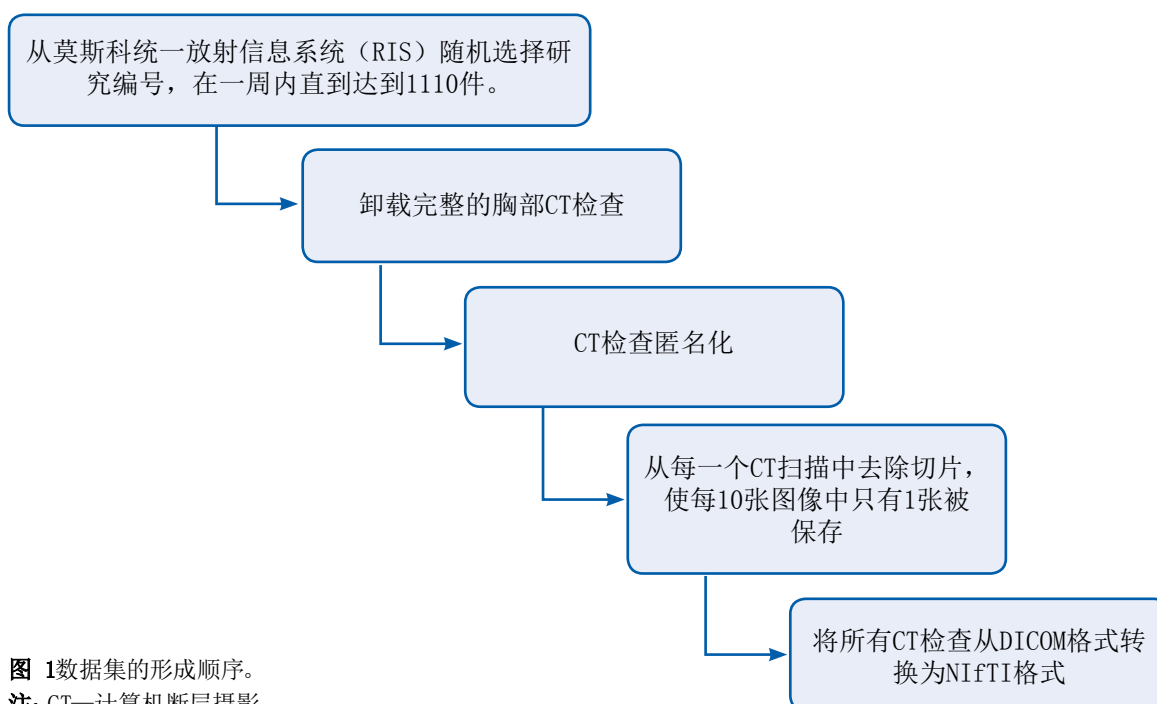


图 1数据集的形成顺序。

注: CT—计算机断层摄影



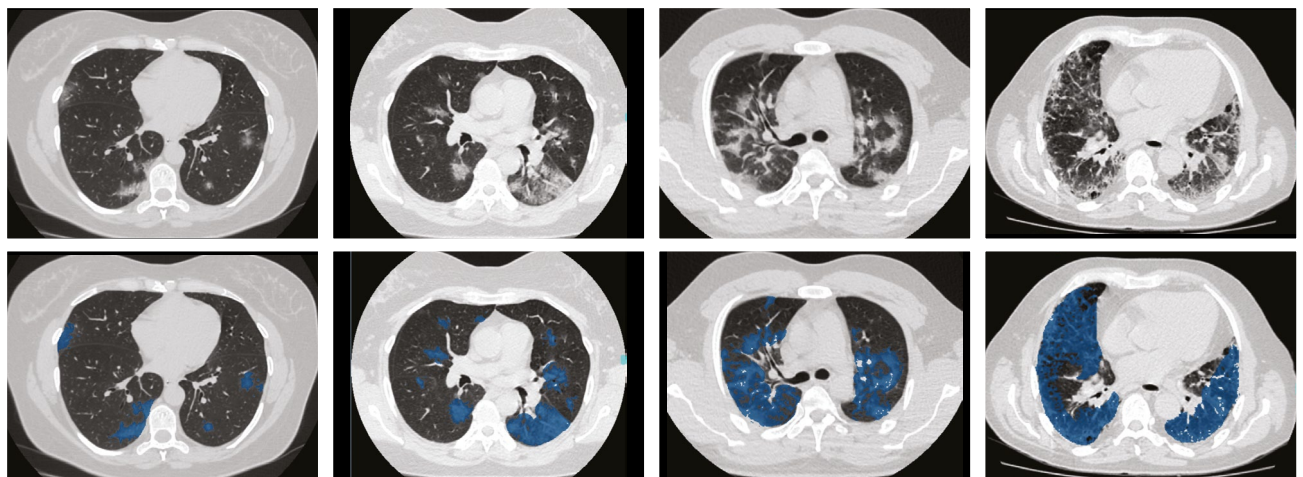
**Table 3:** List of medical organizations where CT data was collected

Municipal Hospital (MH) № 19 Department of Health Care of Moscow	MH № 214	MH № 52
MH № 23	MH № 6	Diagnostic Center № 5
MH № 3	MH № 209	MH № 9
MH № 62	Diagnostic Center № 4	MH № 218
MH № 175	MH № 212	MH № 170
MH № 191	MH № 8	M. P. Conchalovsky hospital (outpatient and in-patient care)
MH № 195	MH № 64	MH № 134
MH № 115	Pediatric Diagnostic Center № 1	MH № 67
Diagnostic Center № 121	MH № 36	MH № 68
Diagnostic Center № 2	MH № 11	MH № 180
MH № 45	MH № 5	MH № 5
MH № 2	Moscow Research and Practical Center for Tuberculosis Control of the South-East Moscow District	MH № 46
MH № 166	Moscow Research and Practical Center for Tuberculosis Control of the Central and West Moscow Districts	MH № 12
MH № 220	MH № 66	Diagnostic Center № 3

CT扫描(图2)。样本特征:1110人,其中42%是男性,56%是女性,2%是其他/未知;年龄为18—97岁,中位年龄为47岁。

在第一阶段,所有研究( $n=1110$ )按分类分为5类(见表1)。按类别划分的个案数目:CT-0—254

例(22.8%)、CT-1—684例(61.6%)、CT-2—125例(11.3%)、CT-3—45例(4.1%)、CT-4—2例(0.2%)。每项研究均以NifTI格式保存,并以Gzip格式存档。在这个过程中,只有每10个图像(Instance)被保存在最终的研究文件中。



**图 2** 不同严重程度COVID-19患者胸部器官计算机断层标记示例

**注:** 上一行,从左至右: COVID-19患者的CT轴向切片,从轻度(CT-1)到极重度(CT-4)。从左到右,下一行:打标后相同的CT数据。



研究的一小部分 (n=50) 由 Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health 的专家进行标记。在标记过程中, 对于每幅图像, 在对应的二值像素掩模上选择正 (白) 像素。结果蒙版被保存为 NIfTI 格式, 然后转换为 Gzip 档案。MedSeg® 注释软件用于创建二进制掩码 (© 2020 Artificial Intelligence AS)。本软件仅标记 COVID-19 典型的变化, 包括磨玻璃类型、实变、小血管和细支气管的变化。条纹口罩的平均密度为 -700 至 -130 HU, 但可以根据灵感的深度而有所不同。大血管和支气管、可见不变的肺实质、运动伪影 (咳嗽和呼吸衰竭引起的呼吸伪影)、重力变化 (如果可以可靠鉴别)、钙化、胸腔积液作为标记的例外。

在数据库建立之前, 所有胸部 CT 扫描都由 Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health 的放射科医师独立审核。根据公认的 CT0-CT4 分类, 将该意见作为评估 COVID-19 肺损伤严重程度的最后意见。这些数据以结构化形式存放在统一的放射信息系统 (RIS) 中, 形成最终的评估结果表。因此, 所有研究都由至少两名专家进行评估。此外, 50 项研究由 3 位

专家进行评估, 并使用外部 MedSeg 软件进行注释。

该数据集用于人工智能 (计算机视觉) 算法的训练、校准和独立评估 [10]。为了帮助抗击 COVID-19, 人工智能 (计算机视觉) 算法将:

- 1) 在门诊检查患者, 以确保其快速和一致的路径选择 (包括基于 CT0-CT4 标准);
- 2) 优先考虑有 COVID-19 迹象的研究;
- 3) 通过对多项研究的比较, 对异常变化进行快速、定性的评估;
- 4) 最小化错误和遗漏异常的风险。

目前, 有广泛的 COVID-19 公开数据集 [11, 12]。然而, 这不应该被视为一个障碍, 因为人工智能算法的发展需要大量高质量的临床信息, 这些信息能够代表真实的患者群体。此外, 人工智能算法需要使用在训练和校准阶段没有使用的新数据集进行验证。可用的开源数据越多, 对开发人员就越有利。可用的数据集相对较小, 很少包含额外的信息, 如标签和/或感兴趣区域 (ROI) 的二进制掩码。

### 如何使用数据集

永久链接: [https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110)  
[https://mosmed.ai/datasets/covid19\\_1110](https://mosmed.ai/datasets/covid19_1110) 此数据集已获得许可 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)。

```

├── 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 和 README\_RU.md 包含 Markdown 数据集的一般信息, 分别为英语和俄语; 同样的信息以 PDF 格式显示在 README\_EN.pdf 和 README\_RU.pdf 中

dataset\_registry.xlsx 包含数据集中包含的研究列表, 对应文件的路径和掩码的路径 (如果有)

研究目录包含目录 CT-0, CT-1, CT-2, CT-3 和 CT-4, 其中每一个包含研究 NIfTI 格式存档在 Gzip。研究名称基于 study\_BBBB.nii.gz 模板, 其中 BBBB 是整个数据集中研究的唯一序列号 (顺序编号)

Masks 目录包含 NIfTI 格式的二进制标记掩码, 归档到 Gzip 中。蒙版的名称基于 study\_BBBB\_mask.nii.gz 模板, 其中 BBBB 是对应研究性的著作的序号

图 3 数据集中的数据存储结构。



## 附加信息

**资金来源。**这篇文章的研究和发表是由作者团队的个人费用进行的。

**利益冲突。**作者声明，没有明显的和潜在的利益冲突相关的发表这篇文章。

**作者的贡献:** S.P. Morozov—负责研究概念，批准最终版本的手稿；A.E. Andreychenko—负责文章设计，数据集形成；I.A. Blokhin—负责数据标记，手稿文本编辑；P.B. Gelezhe—负责搜索出

刊物上的主题文章，数据标记；A.P. Gonchar—扶额数据标记，信息专家评估；A.E. Nikolaev—负责数据标记，信息的专家评估；N.A. Pavlov—负责写文章，形成数据集；V. Yu. Chernina—负责数据标记，手稿书写；V.A. Gombolevsky—负责数据标记，批准最终版本的手稿。所有作者都对文章的研究和准备做出了重大贡献，在发表前阅读并批准了最终版本。

**谢意的表示。**提交人向正在抗击这一流行病的莫斯科卫生部医疗机构的所有医生表示感谢。

## СПИСОК ЛИТЕРАТУРЫ

1. Ai T., Yang Z., Hou H., et al. Correlation of chest CT and RT-PCR testing in Coronavirus Disease 2019 (COVID19) in China: a report of 1014 cases // *Radiology*. 2020. Vol. 296, N 2. E32–E40. doi: 10.1148/radiol.2020200642
2. Handbook of COVID-19 Prevention and Treatment. Ed. by T. Liang. Zhejiang University School of Medicine, 2020. 68 p.
3. Huang Z., Zhao S., Li Z., et al. The battle against Coronavirus Disease 2019 (COVID-19): emergency management and infection control in a Radiology Department. *J Am Coll Radiol*. 2020. Vol. 17, N 6. P. 710–716. doi: 10.1016/j.jacr.2020.03.011
4. Морозов С.П., Гомболевский В.А., Чернина В.Ю. и др. Прогнозирование летальных исходов при COVID-19 по данным компьютерной томографии органов грудной клетки // *Туберкулез и болезни легких*. 2020. Т. 98, № 6. С. 7–14. doi: 10.21292/2075-1230-2020-98-6-7-14
5. Morozov S., Guseva E., Ledikhova N., et al. Telemedicine-based system for quality management and peer review in radiology // *Insights Imaging*. 2018. Vol. 9, N 3. P. 337–341. doi: 10.1007/s13244-018-0629-y
6. Li L., Qin L., Xu Z., et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy // *Radiology*. 2020. Vol. 296, N 2. E65–E71. doi: 10.1148/radiol.2020200905
7. Ucar F., Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images // *Med Hypotheses*. 2020. Vol. 140. P. 109761. doi: 10.1016/j.mehy.2020.109761
8. Временные методические рекомендации «Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19). Версия 9» (утв. Министерством здравоохранения РФ 26 октября 2020). Режим доступа: <https://base.garant.ru/74810808/>. Дата обращения: 12.10.2020.
9. Морозов С.П., Проценко Д.Н., Сметанина С.В. и др. Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов: методические рекомендации. Серия «Лучшие практики лучевой и инструментальной диагностики». Вып. 65. Москва, 2020.
10. Morozov S.P., Vladzymirskyy A.V., Klyashtornyy V.G., et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Series «Best practices in medical imaging». Moscow, 2019. Issue 57.
11. Cohen J.P., Morrison P., Dao L. COVID-19 Image Data Collection [Internet]. 2020 [дата обращения: 25.03.2020]. Доступ по ссылке: <https://arxiv.org/abs/2003.11597>
12. Jun M., Cheng G., Yixin W., et al. COVID-19 CT lung and infection segmentation dataset. Version 1.0. 2020. doi: 10.5281/zenodo.3757476

## REFERENCES

1. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in Coronavirus Disease 2019 (COVID19) in China: a report of 1014 cases. *Radiology*. 2020;296(2):E32–E40. doi: 10.1148/radiol.2020200642
2. *Handbook of COVID-19 Prevention and Treatment*. Ed. by T. Liang. Zhejiang University School of Medicine; 2020. 68 p.
3. Huang Z, Zhao S, Li Z, et al. The battle against Coronavirus Disease 2019 (COVID-19): emergency management and infection control in a Radiology Department. *J Am Coll Radiol*. 2020;17(6):710–716. doi: 10.1016/j.jacr.2020.03.011
4. Morozov SP, Gombolevskiy VA, Chernina VY, et al. Prediction of lethal outcomes in COVID-19 cases based on the results chest computed tomography. *Tuberculosis and Lung Diseases*. 2020;98(6):7–14. (In Russ.) doi: 10.21292/2075-1230-2020-98-6-7-14
5. Morozov S, Guseva E, Ledikhova N, et al. Telemedicine-based system for quality management and peer review in radiology. *Insights Imaging*. 2018;9(3):337–341. doi: 10.1007/s13244-018-0629-y
6. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology*. 2020;296(2):E65–E71. doi: 10.1148/radiol.2020200905
7. Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses*. 2020;140:109761. doi: 10.1016/j.mehy.2020.109761
8. Vremennye metodicheskie rekomendatsii “Proflaktika, diagnostika i lechenie novoi koronavirusnoi infektsii (COVID-19). Ver-



siya 9" (utv. Ministerstvom zdavookhraneniya RF 26 oktyabrya 2020). Available from: <https://base.garant.ru/74810808/>

**9.** Morozov SP, Protsenko DN, Smetanina SV, editors. Radiation diagnostics of coronavirus disease (COVID-19): organization, methodology, interpretation of results: guidelines. Series "Best practices of radiation and instrumental diagnostics". Issue 65. Moscow; 2020.

**10.** Morozov SP, Vladzmyrskyy AV, Klyashtorny VG, et al. Clinical acceptance of software based on artificial intelligence tech-

nologies (radiology). Series "Best practices in medical imaging". Moscow; 2019. Issue 57.

**11.** Cohen JP, Morrison P, Dao L. COVID-19 Image Data Collection [Internet]. 2020 [cited 2020 Mar 25]. Available from: <https://arxiv.org/abs/2003.11597>

**12.** Jun M, Cheng G, Yixin W, et al. COVID-19 CT lung and infection segmentation dataset. Version 1.0. 2020. doi: 10.5281/zenodo.3757476

## ОБ АВТОРАХ

**\*Гомболевский Виктор Александрович**, к.м.н.;  
адрес: Россия, 127051, Москва, ул. Петровка, д. 24/1;  
ORCID: <https://orcid.org/0000-0003-1816-1315>;  
eLibrary SPIN: 6810-3279; e-mail: [g\\_victor@mail.ru](mailto:g_victor@mail.ru)

**Морозов Сергей Павлович**, д-р мед. наук, профессор;  
ORCID: <http://orcid.org/0000-0001-6545-6170>;  
eLibrary SPIN: 8542-1720; e-mail: [morozov@npcmr.ru](mailto:morozov@npcmr.ru)

**Андрейченко Анна Евгеньевна**, к.ф.-м.н.;  
ORCID: <https://orcid.org/0000-0001-6359-0763>;  
eLibrary SPIN: 6625-4186; e-mail: [a.andreychenko@npcmr.ru](mailto:a.andreychenko@npcmr.ru)

**Блохин Иван Андреевич**;  
ORCID: <http://orcid.org/0000-0002-2681-9378>;  
eLibrary SPIN: 3306-1387; e-mail: [i.blokhin@npcmr.ru](mailto:i.blokhin@npcmr.ru)

**Гележе Павел Борисович**, к.м.н.;  
ORCID: <https://orcid.org/0000-0003-1072-2202>;  
eLibrary SPIN: 4841-3234; e-mail: [gelezhe.pavel@gmail.com](mailto:gelezhe.pavel@gmail.com)

**Гончар Анна Павловна**;  
ORCID: <http://orcid.org/0000-0001-5161-6540>;  
eLibrary SPIN: 3513-9531; e-mail: [a.gonchar@npcmr.ru](mailto:a.gonchar@npcmr.ru)

**Николаев Александр Евгеньевич**;  
ORCID: <http://orcid.org/0000-0001-5151-4579>;  
eLibrary SPIN: 1320-1651; e-mail: [a.e.nikolaev@yandex.ru](mailto:a.e.nikolaev@yandex.ru)

**Павлов Николай Александрович**;  
ORCID: <https://orcid.org/0000-0002-4309-1868>;  
eLibrary SPIN: 9960-4160; e-mail: [n.pavlov@npcmr.ru](mailto:n.pavlov@npcmr.ru)

**Чернина Валерия Юрьевна**;  
ORCID: <http://orcid.org/0000-0002-0302-293X>;  
eLibrary SPIN: 8896-8051; e-mail: [v.chernina@npcmr.ru](mailto:v.chernina@npcmr.ru)

## AUTHORS INFO

**\*Victor A. Gombolevskiy**, MD, PhD, MPH;  
address: Petrovka 24/1, Moscow, Russia, 127051;  
ORCID: <https://orcid.org/0000-0003-1816-1315>;  
eLibrary SPIN: 6810-3279; e-mail: [g\\_victor@mail.ru](mailto:g_victor@mail.ru)

**Sergey P. Morozov**, MD, PhD, Professor;  
ORCID: <http://orcid.org/0000-0001-6545-6170>;  
eLibrary SPIN: 8542-1720; e-mail: [morozov@npcmr.ru](mailto:morozov@npcmr.ru)

**Anna E. Andreychenko**, MD;  
ORCID: <https://orcid.org/0000-0001-6359-0763>;  
eLibrary SPIN: 6625-4186; e-mail: [a.andreychenko@npcmr.ru](mailto:a.andreychenko@npcmr.ru)

**Ivan A. Blokhin**, MD;  
ORCID: <http://orcid.org/0000-0002-2681-9378>;  
eLibrary SPIN: 3306-1387; e-mail: [i.blokhin@npcmr.ru](mailto:i.blokhin@npcmr.ru)

**Pavel B. Gelezhe**, MD, PhD;  
ORCID: <https://orcid.org/0000-0003-1072-2202>;  
eLibrary SPIN: 4841-3234; e-mail: [gelezhe.pavel@gmail.com](mailto:gelezhe.pavel@gmail.com)

**Anna P. Gonchar**, MD;  
ORCID: <http://orcid.org/0000-0001-5161-6540>;  
eLibrary SPIN: 3513-9531 e-mail: [a.gonchar@npcmr.ru](mailto:a.gonchar@npcmr.ru)

**Alexander E. Nikolaev**, MD;  
ORCID: <http://orcid.org/0000-0001-5151-4579>;  
eLibrary SPIN: 1320-1651; e-mail: [a.e.nikolaev@yandex.ru](mailto:a.e.nikolaev@yandex.ru)

**Nikolay A. Pavlov**, MD, MPA;  
ORCID: <https://orcid.org/0000-0002-4309-1868>;  
eLibrary SPIN: 9960-4160; e-mail: [n.pavlov@npcmr.ru](mailto:n.pavlov@npcmr.ru)

**Valeria Yu. Chernina**, MD;  
ORCID: <http://orcid.org/0000-0002-0302-293X>;  
eLibrary SPIN: 8896-8051; e-mail: [v.chernina@npcmr.ru](mailto:v.chernina@npcmr.ru)

