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[1] 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)[2]. 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

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

[1] 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.

[2] Постоянная ссылка: https://mosmed.ai/datasets/covid19_1110.

Рукопись получена: 12.10.2020 **Рукопись одобрена:** 11.12.2020 **Опубликована:** 21.12.2020

© Eco-Vector, 2020 The article can be used under the CC BY-NC-ND 4.0 license.

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

© Eco-Vector, 2020 The article can be used under the CC BY-NC-ND 4.0 license.

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）2。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

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

[1] 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。
[2] 永久链接：https://mosmed.ai/datasets/covid19_1110

收到：12.10.2020 接受：11.12.2020 发布时间：21.12.2020

© Eco-Vector, 2020 The article can be used under the CC BY-NC-ND 4.0 license.

ОБОСНОВАНИЕ

Во время пандемии COVID-19 в большинстве стран на структуры здравоохранения легло дополнительное тяжёлое бремя. Ситуация потребовала как никогда тщательного использования финансовых и людских ресурсов. К сожалению, проводимых в медицинских учреждениях профилактических мероприятий не всегда достаточно для того, чтобы избежать гибели медицинских работников. Особую озабоченность вызывает потеря обученных специалистов в области неотложной помощи, лучевой диагностики и других экстренных видах медицины. Компьютерная томография (КТ) считается ключевым инструментом диагностики и оценки прогрессирования COVID-19, проводится в амбулаторных условиях и предназначена пациентам с острыми респираторными симптомами, а также пациентам с установленным диагнозом и прогрессированием лёгких заболеваний, выполняющих терапию в домашних условиях под наблюдением врача с использованием телемедицинских технологий. В стационарных учреждениях КТ используется для первичной и дифференциальной диагностики, для оценки прогрессирования заболевания и определения дальнейшей тактики ведения пациента — в отделении интенсивной терапии или на дому под наблюдением врача медицинской организации первичного уровня [1–3].

Всё более широкое применение КТ ложится тяжёлым бременем на систему здравоохранения. Так, в Москве сеть муниципальных амбулаторных центров КТ проводит около 90 исследований на 1 компьютерный томограф в день (регистратор КТ провел 163 исследования в течение одного дня). Для стандартизации и упорядочения принятия клинических решений специалистами была разработана классификационная модель, которая наряду с другими симптомами оценивает степень тяжести аномалий в лёгочной ткани, наблюдаемых на снимках КТ органов грудной клетки (табл. 1). Так, определение степени поражения лёгочной паренхимы позволяет прогнозировать летальные исходы при COVID-19 [4].

Профессиональное выгорание и высокий риск смерти среди медицинских работников требуют автоматизации процесса анализа изображений, что позволит повысить производительность и минимизировать ошибки [5]. Предварительные данные свидетельствуют о том, что алгоритмы искусственного интеллекта обладают достаточной точностью для диагностики COVID-19 (чувствительность 90%, специфичность 96%, AUC (area under the curve, площадь под характериологической кривой) 0,96, общая точность 76,37–98,26) [6, 7].

Методика сканирования, реконструкции изображений и сохранения базы данных

КТ органов грудной клетки выполняли на 42 томографах одинаковой модели (Toshiba Aquilion 64; Canon Medical Systems, Япония). Все исследования выполнялись по стандартной методике и по стандартным протоколам, рекомендованным производителем (табл. 2).

Одно исследование относится к одному пациенту и включает одну трёхмерную реконструкцию. Критериями включения в исследование являлись факты обращения пациента в городскую поликлинику, на время эпидемии переоборудованную в амбулаторный центр КТ; консультация врача-терапевта с направлением на КТ органов грудной клетки за счёт средств фонда обязательного медицинского страхования.

К критериям невключения из исследования относились беременность и возраст младше 18 лет. Пациент с оксигенацией крови < 93%, выявленной перед КТ, снимался с исследования и направлялся на госпитализацию по службе скорой помощи.

Формирование датасета включало 5 этапов (рис. 1).

Сбор данных и формирование датасета

Сбор данных осуществлялся в период с 1 марта по 25 апреля 2020 г. в медицинских учреждениях городского здравоохранения г. Москвы, ведущих амбулаторный приём: ГБУЗ ГП № 19 ДЗМ, ГБУЗ ГП № 214 ДЗМ, ГБУЗ ГП № 52 ДЗМ, ГБУЗ ГП № 23 ДЗМ, ГБУЗ ГП № 6 ДЗМ, ГБУЗ ДЦ № 5 ДЗМ, ГБУЗ ГП № 3 ДЗМ, ГБУЗ ГП № 209 ДЗМ, ГБУЗ ГП № 9 ДЗМ, ГБУЗ ГП № 62 ДЗМ, ГБУЗ КДЦ № 4 ДЗМ, ГБУЗ ГП № 218 ДЗМ, ГБУЗ ГП № 175 ДЗМ, ГБУЗ ГП № 212 ДЗМ, ГБУЗ ГП № 170 ДЗМ, ГБУЗ ГП № 191 ДЗМ, ГБУЗ ГП № 8 ДЗМ, ГБУЗ ГКБ им. М.П. Кончаловского ДЗМ (амбулаторный и стационарный приём), ГБУЗ ГП № 195 ДЗМ, ГБУЗ ГП № 64 ДЗМ, ГБУЗ ГП № 134 ДЗМ, ГБУЗ ГП № 115 ДЗМ, ГБУЗ ДКЦ № 1 ДЗМ, ГБУЗ ГП № 67 ДЗМ, ГБУЗ КДП № 121 ДЗМ, ГБУЗ ГП № 36 ДЗМ, ГБУЗ ГП № 68 ДЗМ, ГБУЗ КДЦ № 2 ДЗМ, ГБУЗ ГП № 11 ДЗМ, ГБУЗ ГП № 180 ДЗМ, ГБУЗ ГП № 45 ДЗМ, ГБУЗ ГП № 5 ДЗМ, ГБУЗ ГП № 5 ДЗМ (Филиал № 1), ГБУЗ ГП № 2 ДЗМ, Филиал ГБУЗ МНПЦ борьбы с туберкулёзом ДЗМ по Юго-Восточному административному округу г. Москвы, ГБУЗ ГП № 46 ДЗМ, ГБУЗ ГП № 166 ДЗМ, Филиал ГБУЗ МНПЦ борьбы с туберкулёзом ДЗМ по Центральному и Западному административным округам г. Москвы, ГБУЗ ГП № 12 ДЗМ, ГБУЗ ГП № 220 ДЗМ, ГБУЗ ГП № 66 ДЗМ, ГБУЗ ДЦ № 3 ДЗМ.

В данном датасете (1110 исследований) содержатся анонимизированные КТ лёгких человека с признаками COVID-19 (КТ1–КТ4), а также без него (КТ0) (рис. 2). Характеристики выборки: 1110 человек, из них мужчин

Таблица 1. Классификация тяжести поражения лёгочной ткани при COVID-19 и правила маршрутизации

Степень тяжести	Категория по данным КТ	Клинические данные	Решение
КТ0	Нет признаков пневмонии (в том числе и COVID-19)	-	Информирование лечащего врача
Лёгкая (КТ1)	Зоны уплотнения по типу матового стекла с/без консолидации, ретикулярные изменения. Минимальный объём/распространённость. Вовлечение паренхимы легкого ≤ 25%	T <38,0 °C ЧДД <20/мин SpO_2 >95%	Динамическое наблюдение на дому с применением телемедицинских технологий (обязательный дистанционный контроль состояния здоровья)
Среднетяжёлая (КТ2)	Зоны уплотнения по типу матового стекла с/без консолидации, ретикулярные изменения. Средний объём/распространённость. Вовлечение паренхимы лёгкого 25–50%	T <38,5 °C ЧДД 20–30/мин SpO_2 95%	Динамическое наблюдение на дому врачом медицинской организации первичного уровня
Тяжёлая (КТ3)	Зоны уплотнения по типу матового стекла. Зоны консолидации, ретикулярные изменения. Значительный объём/распространённость. Вовлечение паренхимы лёгкого 50–75%. Увеличение объёма поражения до 50% за 24–48 ч на фоне дыхательных нарушений, если исследования выполняются в динамике	≥2 признаков на фоне лихорадки: T >38,5 °C ЧДД ≥30/мин SpO_2 ≤93% PaO_2 / FiO_2 ≤300 мм рт.ст.	Немедленная госпитализация в стационар, профилированный для оказания помощи пациентам с COVID-19. В условиях стационара: немедленный перевод в ОРИТ. Экстренная КТ (если не было)
Критическая (КТ4)	Диффузное уплотнение лёгочной ткани по типу матового стекла и консолидации в сочетании с ретикулярными изменениями. Гидроторакс (двусторонний, преобладает слева). Субтотальный объём/распространённость. Вовлечение паренхимы лёгкого ≥75%	Признаки шока, полиорганной недостаточности, дыхательная недостаточность	Оказание экстренной медицинской помощи. Немедленная госпитализация в стационар, профилированный для оказания помощи пациентам с COVID-19. В условиях стационара: немедленный перевод в ОРИТ. Экстренная КТ (если не было и позволяет состояние)

Примечание. КТ — компьютерная томография, КТ-1 – КТ-4 — степень поражения лёгких по результатам КТ, ЧДД — частота дыхательных движений, ОРИТ — отделение реанимации и интенсивной терапии, T — температура тела, PaO_2 — артериальное парциальное давление кислорода, FiO_2 — концентрация кислорода, SpO_2 — сатурация крови кислородом.

Таблица 2. Методика сканирования, реконструкции изображений и сохранения базы данных

Группа	Параметр	Значение и комментарий
Оборудование	КТ-сканер	Toshiba Aquilion 64 (Canon Medical Systems, Япония)
	Количество срезов	64
Пациент	Позиция пациента	Расположение груди в центре гентри. Высота стола и центровка отрегулированы таким образом, что средняя ключичная линия находится в изоцентре. Руки над головой. Инструкции по дыханию. Обучение пациентов и инструктаж по дыханию перед сканированием
	Одежда и инородные предметы	Все инородные предметы, которые могут быть удалены, убраны из зоны сканирования, включая украшения и цепочки на шее. Нижнее бельё допустимо

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

Группа	Параметр	Значение и комментарий
Пациент	Localizer / scout / томограмма*	Проводился на уровне грудной клетки на предмет ограничения сканирования диапазоном лёгких. Проводился для поиска дополнительных инородных предметов на уровне сканирования, которые могут ухудшить качество. Сканирование при задержке дыхания на глубине вдоха
	Диапазон сканирования	Весь объём лёгких, включая 5 см выше и 5 см ниже лёгких
	Фаза дыхания	КТ-сканирование при задержке дыхания на глубине вдоха
	Поле обзора Display Field of View (FOV)	Не менее 1 см от рёбер (от 350 до 500 мм). Молочные железы включались в область сканирования, однако могли быть частично исключены из поля обзора
Медицинский персонал	Рентгенолаборант	Находился в пультовой. С пациентом не контактировал. Очные контакты с укладчиком были минимизированы с целью безопасности
	Укладчик	Укладчик является медицинским сотрудником отдела лучевой диагностики, в виде дополнительного персонала переведён из маммографических рентгенолаборантов в кабинет КТ на время эпидемии согласно приказу Департамента здравоохранения Москвы. Находился в аппаратной (укладка пациента и поднятие со стола) и в коридоре (во время сканирования). С пациентом контактировал. Действовал согласно методическим рекомендациям [8, 9]
Протокол сканирования и реконструкции, просмотра и интерпретации	Наклон гентри	Нет
	Продолжительность сканирования	≤ 10 с (чаще всего 6 с)
	Внутривенное контрастирование	Отсутствовало
	Пероральное контрастирование	Отсутствовало
	Напряжение	120 кВ
	Сила тока	Система автоматической модуляции силы тока Sure exp.3D, встроенная в компьютерный томограф производителем. Система автоматически настраивала силу тока, добиваясь уровня шума 10 HU для срезов толщиной 5,0 мм в диапазоне 80–500 мА. XY модуляция включена
	Скорость ротации рентгеновской трубки	0,5 с
	Объёмный питч	95,0
	Процесс реконструкции	QDS+
	Количество реконструированных КТ-серий	2 (с лёгочным и мягкотканным кернелом[3])
	Кернел реконструкции для мягких тканей (отдельная КТ-серия)	FC07 или FC18
	Кернел реконструкции для лёгких (отдельная КТ-серия)	FC51
	Толщина срезов	1,0 мм (одинаковая для обоих кернел)
	Шаг между срезами	0,8 мм (одинаковая для обоих кернел)
	Итеративные реконструкции	AIDR 3D встречались всего в 5 томографах, у остальных отсутствовали алгоритмы итеративной реконструкции, поэтому использовалась FBP (обратная прямая проекция)

[3] Кернел — фильтр, используемый при реконструкции данных компьютерной томографии.

Группа	Параметр	Значение и комментарий
Протокол сканирования и реконструкции, просмотра и интерпретации	Для интерпретации КТ использовались	AGFA Enterprise 8.0, Vitrea FX
	Проекции максимальной интенсивности (maximum intensity projections, MIP), проекции минимальной интенсивности (minimum intensity projections, MinIP), мультипланарные реконструкции (multiplanar reconstructions, MPR)	Применялись
	Алгоритмы искусственного интеллекта	Применялись, но не для всех исследований. В случае применения алгоритмы машинного обучения демонстрировали врачу дополнительную серию исследования, на которой были представлены выборочные среды, на которых наличие предполагаемого поражения COVID-19 было ограничено прямоугольниками красного цвета, привлекая внимание врача. Кроме этого, была представлена суммационная трёхмерная реконструкция лёгких с отмеченными красным цветом областями поражения, которые выявил автоматический алгоритм. Количественной информации для оценки степени поражения лёгких представлено не было
	Время на финализацию протокола	От 10 мин до 3 ч. В редких случаях 24 ч
	Стандартизация протокола	Шаблон протокола был сформирован и регламентирован в методических рекомендациях, а также внедрен в Единый радиологический информационный сервис, в котором происходило формирование протокола врачом-рентгенологом
	Классификация поражения COVID-19	Использовалась классификация по шкале КТ-0 – КТ-4 (см. табл. 1)
	Второе мнение	Для всех КТ-исследований из поликлиник было получено второе мнение от экспертов ГБУЗ НПКЦ ДиТ ДЗМ
	Расчёт дозы лучевой нагрузки	Использовались данные DLP из автоматически создаваемой КТ-серии DoseReport. В РФ, согласно методическим указаниям (МУ 2.6.1.2944-11) «Контроль эффективных доз облучения пациентов при проведении медицинских рентгенологических исследований», для расчёта эффективной дозы (мЗв) необходимо произведение DLP и 0,017 (коэффициент при КТ грудной клетки)
База данных	Источник данных	Единый радиологический информационный сервис г. Москвы (AGFA Enterprise 8.0)
	Формат первоначального сбора данных	DICOM 3.0
	Плоскость	Аксиальная
	Толщина срезов	1,0 мм
	Шаг между срезами	9,0 мм (так как сохранён каждый 10-й срез)
	Формат сохранённой базы данных	NIfTI
	Программное обеспечение для аннотации в виде бинарных масок с выделением поражения лёгких	MedSeg® (© 2020 Artificial Intelligence AS)

Примечание. * Отсутствует в базе данных, но необходимо для формирования КТ-сканирования. КТ — компьютерная томография, КТ-0–КТ-4 — степень поражения лёгких по результатам КТ.

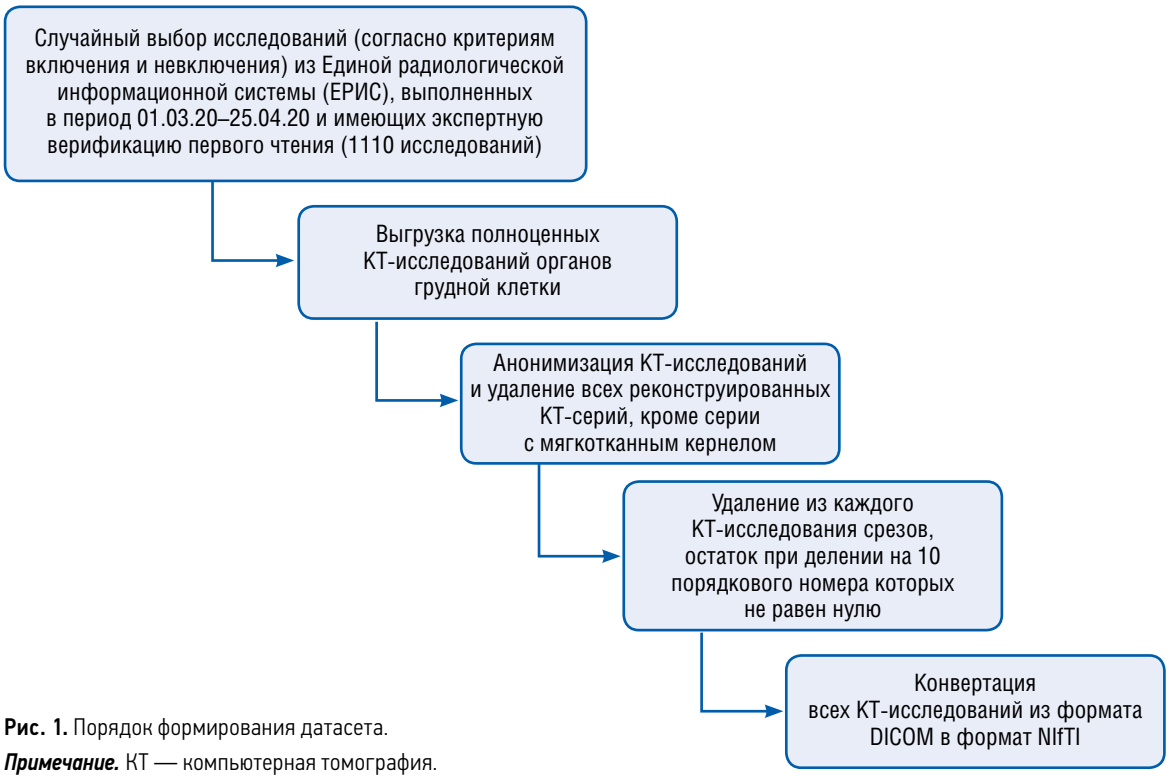

Рис. 1. Порядок формирования датасета.
Примечание. КТ — компьютерная томография.

42%, женщин 56%, прочих/неизвестных 2%; возраст от 18 до 97 лет, медианный возраст 47 лет.

На первом этапе все исследования (n = 1110) были распределены по 5 категориям в соответствии с классификацией (см. табл. 1). Количество случаев по категориям: КТ-0 — 254 (22,8%), КТ-1 — 684 (61,6%), КТ-2 — 125 (11,3%), КТ-3 — 45 (4,1%), КТ-4 — 2 (0,2%). Каждое исследование было сохранено в формате NIfTI и заархивировано в Gzip. В ходе этого процесса только каждое 10-е изображение (Instance) сохранялось в итоговом файле исследования.

Небольшая часть исследований (n = 50) была размечена специалистами Научно-практического клинического центра диагностики и телемедицинских технологий Департамента здравоохранения Москвы (ГБУЗ НПКЦ ДиТ ДЗМ). Во время разметки для каждого из снимков были выбраны положительные (белые) пиксели на соответствующей бинарной пиксельной маске. Полученные маски были сохранены в формате NIfTI, а затем преобразованы в архивы Gzip. Для создания бинарных масок использовалось программное обеспечение (ПО) для аннотирования MedSeg® (© 2020 Artificial Intelligence AS).

В данном ПО проводилась разметка только изменений, характерных для COVID-19, включая изменения по типу матового стекла, консолидации.

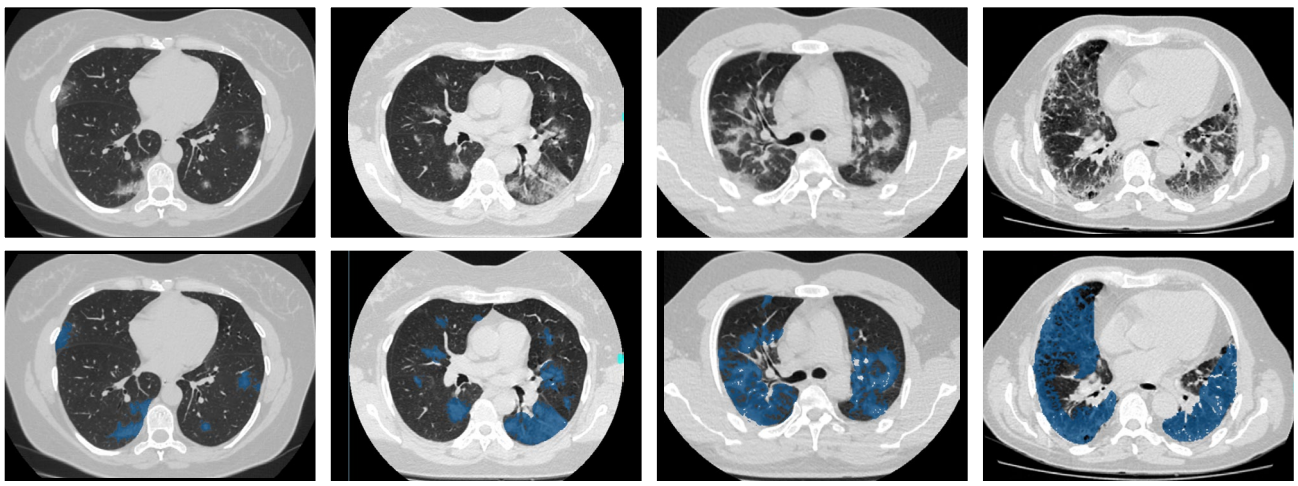

Рис. 2. Примеры разметки компьютерных томограмм органов грудной клетки пациентов с различной степенью тяжести COVID-19.
Примечание. Слева направо верхний ряд: аксиальные срезы компьютерных томограмм (КТ) пациентов с COVID-19 от слабой (КТ-1) до критической (КТ-4) степени тяжести. Слева направо нижний ряд: те же данные КТ после разметки.

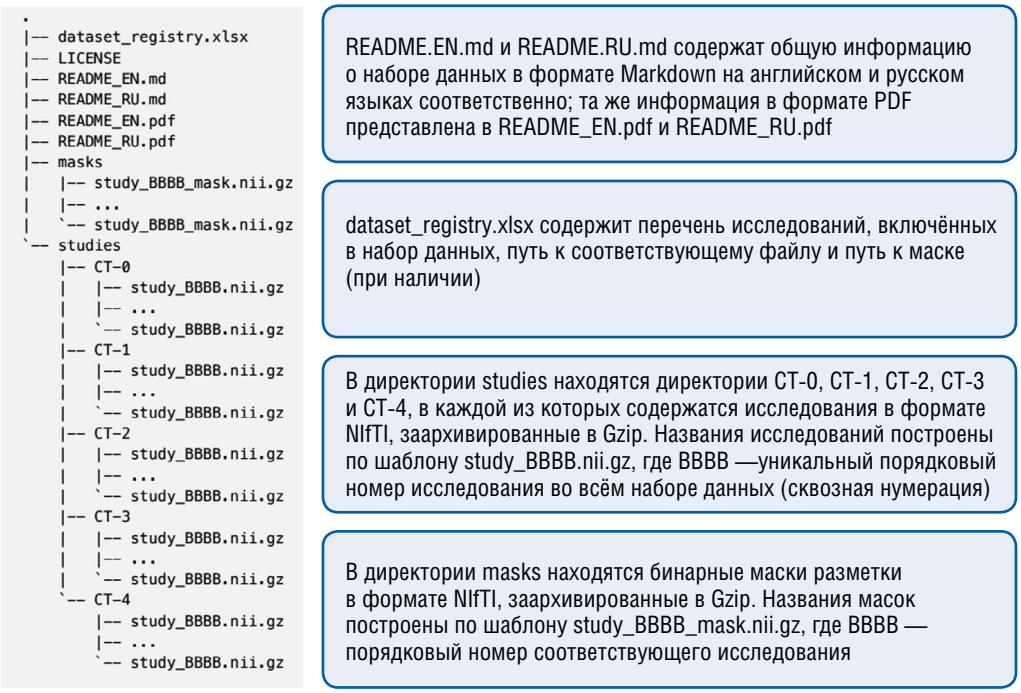

Рис. 3. Структура хранения данных в датасете.

Средняя плотность маски для разметки составляет от -700 HU до -130 HU, однако она могла отличаться в зависимости от глубины вдоха. Исключением разметки служили крупные сосуды и бронхи, визуально неизменённая лёгочная паренхима, двигательные артефакты (дыхательные за счёт кашля и дыхательной недостаточности), гравитационные изменения (если их можно было достоверно дифференцировать), кальцинаты, плевральный выпот.

Все исследования, включенные в датасет, имели два совпадающих независимых чтения: врачом-рентгенологом по месту проведения исследования и экспертом ГБУЗ НПКЦ ДиТ ДЗМ. В дополнение к этому, разметка 50 исследований с помощью бинарных масок проводилась независимыми специалистами ГБУЗ НПКЦ ДиТ ДЗМ с использованием внешнего программного обеспечения MedSeg® (© 2020 Artificial Intelligence AS).

Набор данных предназначен для обучения, калибровки и независимой оценки алгоритмов искусственного интеллекта (компьютерного зрения) [10]. В помощь в борьбе с COVID-19 алгоритмы искусственного интеллекта (компьютерного зрения) позволят:

1) обследовать пациентов в амбулаторных учреждениях для их быстрой и последовательной маршрутизации (в т.ч. на основе критериев КТ0–КТ4);
2) приоритизировать исследования, содержащие признаки COVID-19, в рабочем списке;
3) провести быструю и качественную оценку аномальных изменений путём сравнения нескольких исследований;
4) минимизировать риск ошибок и пропущенных аномалий.

В настоящее время существует широкий спектр общедоступных наборов данных COVID-19 [11, 12]. Однако это не должно рассматриваться как препятствие, так как разработка алгоритмов искусственного интеллекта требует больших объёмов качественной клинической информации, репрезентативной для реальных популяций пациентов. Кроме того, алгоритмы искусственного интеллекта должны быть проверены с использованием новых наборов данных, которые не использовались на этапах обучения и калибровки. Чем больше данных имеется в открытых источниках, тем более высококачественные алгоритмы искусственного интеллекта могут создавать разработчики. Имеющиеся наборы данных относительно малы и редко содержат дополнительную информацию, например теги и/или бинарные маски для интересующих регионов (ROI).

Как использовать датасет

Постоянная ссылка: https://mosmed.ai/datasets/covid19_1110. Этот набор данных лицензирован Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0); рис. 3.

ДОПОЛНИТЕЛЬНО

Источник финансирования. Исследование и публикация статьи осуществлены на личные средства авторского коллектива.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Участие авторов: С.П. Морозов — концепция исследования, утверждение финальной версии рукописи; А.Е. Андрейченко — дизайн статьи, формирование набора данных; И.А. Блохин — разметка данных, редактирование текста рукописи; П.Б. Гележе — поиск публикаций по теме статьи, разметка данных; А.П. Гончар — разметка данных, экспертная оценка информации; А.Е. Николаев — разметка данных, экспертная оценка информации; Н.А. Павлов — написание статьи, формирование набора данных; В.Ю. Чернина — разметка данных, написание рукописи; В.А. Гомболевский — разметка данных, написание рукописи, утверждение финальной версии статьи, согласование правок. Все авторы внесли существенный вклад в проведение исследования и подготовку статьи, прочли и одобрили финальную версию до публикации.

Благодарности. Авторы выражают благодарность всем врачам медицинских организаций Департамента здравоохранения Москвы, борющимся с эпидемией.

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., Vladzymyrskyy 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. Verson 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 “Profilaktika, diagnostika i lechenie novoi koronavirusnoi infektsii (COVID-19). Ver-

siya 9" (utv. Ministerstvom zdravookhraneniya 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, Vladzymyrskyy AV, Klyashtornyy VG, et al. Clinical acceptance of software based on artificial intelligence technologies (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. Verson 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

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

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

Блохин Иван Андреевич;
ORCID: http://orcid.org/0000-0002-2681-9378;
eLibrary SPIN: 3306-1387; e-mail: 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

Гончар Анна Павловна;
ORCID: http://orcid.org/0000-0001-5161-6540;
eLibrary SPIN: 3513-9531; e-mail: 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

Павлов Николай Александрович;
ORCID: https://orcid.org/0000-0002-4309-1868;
eLibrary SPIN: 9960-4160; e-mail: 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

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

Sergey P. Morozov, MD, PhD, Professor;
ORCID: http://orcid.org/0000-0001-6545-6170;
eLibrary SPIN: 8542-1720; e-mail: 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

Ivan A. Blokhin, MD;
ORCID: http://orcid.org/0000-0002-2681-9378;
eLibrary SPIN: 3306-1387; e-mail: 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

Anna P. Gonchar, MD;
ORCID: http://orcid.org/0000-0001-5161-6540;
eLibrary SPIN: 3513-9531 e-mail: 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

Nikolay A. Pavlov, MD, MPA;
ORCID: https://orcid.org/0000-0002-4309-1868;
eLibrary SPIN: 9960-4160; e-mail: 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