

Паспорт набора данных

1. Идентификатор набора данных

Будет уточнен

2. Версия набора данных

1.0

3. Наименование набора данных

Датасет «StopCOVID» по задаче изучение постковидного синдрома у пациентов, ранее госпитализированных по поводу новой коронавирусной инфекции в Университетские клинические больницы №1, 2, 3, 4 ФГАОУ ВО Первого Московского Государственного Медицинского Университета им. И.М. Сеченова.

4. Описание набора данных

4.1. Аннотация

Для создания датасета использовались данные из электронных медицинских карт пациентов, госпитализированных с подозрением на новую коронавирусную инфекцию, а также результаты телефонного опроса, проведенного согласно протоколу сбора информации пациентов, перенесших новую коронавирусную инфекцию. Все данные собирались согласно протоколам, разработанным Международным консорциумом по тяжелым острым респираторным и новым инфекциям (ISARIC).

4.2. Клиническая задача

В последние годы мир столкнулся с пандемией, вызванной новым коронавирусом (COVID-19). Помимо активной фазы заболевания, многие пациенты, переболевшие COVID-19, сталкиваются с долгосрочными последствиями, которые называют "постковидным синдромом". Эти последствия могут включать в себя усталость, дыхательные проблемы, боли в суставах, неврологические симптомы и многие другие.

Цель данной клинической задачи - глубоко изучить постковидный синдром, определить его характерные признаки, продолжительность, факторы риска и возможные методы лечения. Для этого необходимо создать датасет, который будет содержать данные о пациентах, переболевших COVID-19, их клинические проявления, длительность симптомов, проведенное лечение и его результаты.

Изучение постковидного синдрома может способствовать более глубокому пониманию особенностей течения данного заболевания, а также определению фенотипа пациентов, наиболее подверженных риску развития долгосрочных последствий в различных возрастных группах. В перспективе такие исследования должны положительно повлиять на подход к ведению пациентов, перенесших новую коронавирусную инфекцию, что приведет к улучшению качества их жизни.

4.3. Назначение набора данных

Датасет предназначен для анализа распространенности симптомов постковидного синдрома у пациентов, ранее госпитализированных по поводу новой коронавирусной инфекции.

Датасет разбит на два набора:

1. Демографические, клиничко-лабораторные, инструментальные параметры пациентов, а также

- результаты телефонного опроса во время их нахождения в стационаре с подозрением на новую коронавирусную инфекцию;
2. Результаты опроса пациентов спустя 6 и 12 месяцев после перенесения новой коронавирусной инфекции.

4.4. Нозологии

COVID-19:

- U07.1 – COVID-19, вирус идентифицирован
- U07.2 – COVID-19, вирус не идентифицирован
- U09.9 – Состояние после COVID-19 неуточненное

5. Владелец набора данных

ФГАОУ ВО Первый Московский Государственный Медицинский Университет им. И.М. Сеченова (Сеченовский университет)

5.1. Контактные данные

Профессор, Мунблит Даниил Борисович, munblit_d_b@staff.sechenov.ru

5.2 Авторы

Профессор Мунблит Даниил Борисович, д.м.н., академик РАН Авдеев Сергей Николаевич, Неклюдов Никита Андреевич, Бугаева Полина Евгеньевна, Кислова Мария Игоревна, Листовская Екатерина Максимовна, Галаяутдинова Айсылу Наиловна, Трефилова Анастасия Алексеевна, Беляев Владимир Евгеньевич, д.х.н. Тимашев Петр Сергеевич, Политов Михаил Евгеньевич, Яворовский Андрей Георгиевич, к.м.н. Капустина Валентина Андреевна, д.м.н. Фомин Виктор Викторович, д.м.н. Свистунов Андрей Алексеевич, к.м.н. Бутнару Денис Викторович, д.м.н. Глыбочко Петр Витальевич, Бобкова Полина Александровна, Яковлева Екатерина Игоревна, Андреева Маргарита Константиновна, Эль-Тараби Ясмин Ахмед Али, Румянцев Михаил Анатольевич, Байрашевская Анастасия Васильевна, к.м.н. Буланов Николай Михайлович

6. Порядок предоставления доступа к набору данных

Будет уточнен.

7. Тэги

ДАТАСЕТ, ЕДИНАЯ ГОСУДАРСТВЕННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА В СФЕРЕ ЗДРАВООХРАНЕНИЯ, МЕДИЦИНСКАЯ ИНФОРМАЦИОННАЯ СИСТЕМА, КОГОРТА, COVID-19, SARS-CoV-2, ПОСТКОВИДНЫЙ СИНДРОМ, СОСТОЯНИЕ ПОСЛЕ COVID-19

8. Источник данных

ФГАОУ ВО Первый Московский Государственный Медицинский Университет им. И.М. Сеченова (Сеченовский университет)

8.1 Глубина набора данных

Выгрузка из медицинской информационной системы Университетских клинических больниц №1,2,3,4 ФГАОУ ВО Первого Московского Государственного Медицинского Университета им. И.М. Сеченова в

период апрель – июль 2020 года с последующим опросом пациентов спустя 6,12 месяцев после выписки.

8.2 Объем набора данных

В состав первой части датасет (острая фаза COVID-19) включена информация по 5060 пациентам. Представлены результаты ПЦР-диагностики на новую коронавирусную инфекцию, данные по симптомам, сопутствующим заболеваниям, в т.ч., полученные путем телефонного опроса, лабораторной диагностике, компьютерной томографии органов грудной клетки, поддерживающей терапии во время госпитализации.

В состав второй части датасета (катамнез через 6 и 12 месяцев после выписки) включены данные опроса 2649 и 1986 пациентов, соответственно. Представлена информация по симптомам постковидного синдрома, сопутствующим заболеваниям, острым состояниям с момента выписки, а также результаты опроса при использовании валидированных шкал (например, EQ5D5L, PHQ4).

Результаты представлены в виде *xlsx*-файлов. Описание каждого параметра представлено в словаре данных острой фазы и катамнеза. Общий объем датасета 8,34 Мб.

8.3. Критерии включения пациентов и исследований

Для формирования датасета были отобраны все пациенты старше 18 лет, госпитализированные в Университетские клинические больницы №1,2,3,4 ФГАОУ ВО Первого Московского Государственного Медицинского Университета им. И.М. Сеченова, с подозрением на наличие или наличием лабораторно подтвержденной инфекции COVID-19, в период с апреля по июль 2020г.

Критерием исключения являлся отказ пациента от участия в исследовании.

8.4. Особенности подготовки исследований

Для сбора данных был использован протокол, разработанный Международным консорциумом по тяжелым острым респираторным и вновь возникающим инфекциям (ISARIC) при поддержке Всемирной Организации Здравоохранения (ВОЗ). В дополнение к вопросам в стандартной опросника ISARIC, анкета также была дополнена предложениями рабочих групп клиницистов, вовлеченных в лечение пациентов поступающих в специализированные коронавирусные отделения университетских клиник.

8.5. Поло-возрастные характеристики набора данных

Все пациенты старше 18 лет были отобраны при формировании датасета без дополнительных ограничений.

8.6. Характеристики оборудования

Дополнительное оборудование не применялось.

8.7. Дата публикации набора данных

Будет уточнено

8.8. Дата обновления набора данных

Будет уточнено

9. Модели машинного обучения

Не применимо

10. Расположение набора данных

Будет уточнено

11 Структура набора данных

11.1. Схема структуры директорий и файлов

```
..... Acute
      | Part-1-Baseline.xlsx
      | Part-2-Lab.xlsx
      | Supportive-Care.xlsx
      | Symptoms-Comorbidities.xlsx
      | CT-results.xlsx
      | Data-Dictionary-Acute.xlsx

..... Follow-up
      | Follow-up-6-Months-Responses.xlsx
      | Follow-up-12-Months-Responses.xlsx
      | Data-Dictionary-Follow-up.xlsx
```

11.2. Описание структуры и принципа наименования файлов и директорий

Директория «*Acute*» содержит информацию о пациентах в период их нахождения в стационаре с подозрением на новую коронавирусную инфекцию. Директория «*Follow-up*» содержит результаты опроса пациентов через 6 и 12 месяцев после выписки.

Каждая директория содержит набор файлов, содержащих деидентифицированные данные пациентов, а также словарь данных. Все файлы представлены в формате *xlsx*.

12. Обзор разметки данных

Не применимо

13. Правила использования и распространения

13.1. Копирайт

ФГАОУ ВО Первый Московский Государственный Медицинский Университет им. И.М. Сеченова
(Сеченовский университет)