

Паспорт набора данных

1. Идентификатор набора данных.

990 суточных мониторингов артериального давления пациентов с артериальной гипертонией и без артериальной гипертонии, для классификации артериальной гипертонии

2. Версия набора данных.

1.2

3. Наименование набора данных.

Датасет по задаче классификации артериальной гипертонии на основании результатов суточного мониторинга артериального давления.

4. Описание набора данных.

4.1. Аннотация

В датасете будут представлены данные 990 протоколов суточного мониторинга артериального давления, предназначенных для обучения и тестирования моделей машинного обучения для классификации артериальной гипертонии. Данные будут разделены на основной (408 протоколов исследования) и тестовый (582 протоколов исследования) наборы.

4.2. Клиническая задача

Артериальная гипертензия является одним из наиболее распространённых заболеваний сердечно-сосудистой системы. Она выявляется у 30-40% взрослого населения и не менее чем у 60-70% лиц старше 60 лет. Обследование больных с артериальной гипертензией традиционно основывается на офисном измерении артериального давления во время визита к врачу или во время пребывания в стационаре. Однако, однократное определение

артериального давления даёт информацию лишь на отдельно взятый момент времени и не всегда отражает реальную клиническую картину. В настоящее время широко применяется суточное мониторирование артериального давления (СМАД), предоставляющее дополнительную информацию об уровне артериального давления вне врачебного кабинета. СМАД позволяет более точно определить уровень артериального давления в реальных условиях, как правило, типичных для пациента. В этих условиях, разработка систем поддержки принятия врачебного решения (на основе машинного обучения) в кардиологии и терапии, которые позволят классифицировать артериальную гипертонию, должно положительно влиять на качество своевременной и точной постановки диагноза, что является стратегически важной задачей для современного здравоохранения и РФ.

4.3. Назначение набора данных

Датасет предназначен для разработки ML-решений автоматического анализа цифровых данных СМАД для классификации артериальной гипертонии, что позволит упростить своевременную и точную постановку диагноза.

4.4. Нозологии

- I11 - Гипертензивная болезнь сердца [гипертоническая болезнь с преимущественным поражением сердца]
- I12 - Гипертензивная [гипертоническая] болезнь с преимущественным поражением почек
- I13 - Гипертензивная [гипертоническая] болезнь с преимущественным поражением сердца и почек

5. Владелец набора данных

Федеральное государственное автономное образовательное учреждение высшего образования Первый Московский государственный медицинский

университет имени И. М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет).

5.1. Контактные данные

119991, г. Москва, ул. Трубецкая, д. 8, стр. 2; Тел.: +7 (495) 609-14 00; Эл. почта.: rektorat@sechenov.ru;

5.2. Авторы

- Осадчук Михаил Алексеевич, д.м.н., профессор
- Васильева Инна Николаевна к.м.н., доцент
- Миронова Екатерина Дмитриевна, к.м.н., ассистент

6. Порядок предоставления доступа к набору данных

Данные протоколов СМАД будут взяты из электронного архива данных ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет).

7. Тэги

ДАТАСЕТ, ЕДИНАЯ ГОСУДАРСТВЕННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА В СФЕРЕ ЗДРАВООХРАНЕНИЯ, МЕДИЦИНСКАЯ ИНФОРМАЦИОННАЯ СИСТЕМА, РАЗМЕТКА ДАННЫХ, КАРДИОЛОГИЯ, СЕРДЕЧНО-СОСУДИСТЫЕ ЗАБОЛЕВАНИЯ, АРТЕРИАЛЬНАЯ ГИПЕРТНОНИЯ, АРТЕРИАЛЬНОЕ ДАВЛЕНИЕ, СУТОЧНОЕ МОНИТОРИРОВАНИЕ АРТЕРИАЛЬНОГО ДАВЛЕНИЯ, НЕЙРОННЫЕ СЕТИ, КЛАССИФИКАЦИЯ

8. Источник данных

Федеральное государственное автономное образовательное учреждение высшего образования Первый Московский государственный медицинский

университет имени И. М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет).

8.1. Глубина набора данных

В набор будут включены данные протоколов СМАД из архива ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет), проведенные до 2023 года.

8.2. Объем набора данных

Планируется произвести ретроспективный набор 990 протоколов СМАД, из них 465 исследований в обучающем наборе данных и 525 исследований в тестовом наборе.

8.3. Критерии отбора клинических параметров и пациентов им соответствующих, а также виды лабораторно-инструментальных исследований.

Для анализа будут отобраны протоколы СМАД, проведенные в ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет).

Критерии включения:

Протоколы СМАД исследований пациентов в возрасте от 18 до 99 лет

8.4. Особенности подготовки исследований

Данные СМАД исследований обезличиваются в процессе выгрузки из архива ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет).

8.5. Поло-возрастные характеристики набора данных

Обучающий набор СМАД:

- Планируется включить в базу данных 212 протоколов СМАД пациентов мужского пола и 196 протоколов СМАД пациентов женского пола.
- Минимальный возраст 18 лет, максимальный 99 лет.

Тестовый набор СМАД

- Планируется включить в базу данных 287 протоколов СМАД пациентов мужского пола и 295 протоколов СМАД пациентов женского пола.
- Минимальный возраст 18 лет, максимальный 99 лет.

8.6. Характеристики оборудования

СМАД планируется проводить осциллометрическим методом с использованием аппарата «BP-lab» («Петр Телегин», Нижний Новгород) с анализом среднесуточных значений систолического (САД), диастолического (ДАД), пульсового (ПАД) и вариабельность артериального давления за все временные периоды суток; показатели «нагрузки давлением», оцениваемые по индексу времени (ИВ), индексу площади (ИП).

8.7. Дата публикации набора данных

30.09.2023

8.8. Дата обновления набора данных

Не обновлялся

9. Модели машинного обучения

Нейронные сети, градиентный бустинг, случайный лес и другие классифицирующие модели машинного обучения

10. Расположение набора данных

Будет уточнен

11. Структура набора данных

11.1. Схема структуры директорий и файлов

```
----- train_set
      |----- Labels.xlsx
----- test_set
      |----- Labels.xlsx
```

11.2. Описание файлов в корневой директории

Датасет разбит на обучающую (train) и тестовую выборки (test). Данные разбиты по соответствующим папкам. Файл содержит данные в текстовом формате с разбиением на строки (по исследованиям) и столбцы. Набор столбцов соответствует параметрам, перечисленным в разделе «12. Обзор разметки данных».

12. Обзор разметки данных

12.1. Особенности формата разметки

Номера протоколов СМАД исследований включенных в датасет соответствуют номерам исследований в архиве ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет).

12.2. Классы разметки

Результатом разметки будет является таблица, содержащая информацию о наличии/отсутствии артериальной гипертензии в анамнезе, в случае ее наличия будут представлены ее классификационные данные, а также половозрастные характеристики СМАД исследований пациентов и метainформацию о записи СМАД.

Таблица 1. Размеченные параметры

Параметр	Описание	Значение
Eid	идентификатор записи	Число
male	мужской пол	Yes/no
female	женский пол	Yes/no
Age	возраст	Число
АН, arterial hypertension	Артериальная гипертензия в анамнезе	Yes/no

SAH, stage of arterial hypertension	Стадия артериальной гипертензии по данным анамнеза	Число
DAH, degree of arterial hypertension	Степень артериальной гипертензии по данным анамнеза	Число
Daily SBP, mm Hg	САД сут., мм рт	Число
Daytime SBP, mm Hg	САД дневное, мм рт.	Число
Night SBP, mm Hg	САД ночное, мм рт. ст.	Число
Daily DBP, mm Hg	ДАД сут., мм рт. ст.	Число
Daytime DBP, mm Hg	ДАД дневное, мм.рт.ст.	Число
Night DBP, mm Hg	ДАД ночное, мм рт. ст.	Число
DPBP, Daytime pulse blood pressure, mm Hg	Пульсовое АД (сут.), мм рт. ст.	Число
DABP, Daily Average blood pressure per day, mm Hg	Среднее АД сут., мм рт. ст	Число
DBP, Daytime Average blood pressure per day, mm Hg	СрАД сут., дневное, мм рт. ст.	Число
Night Average blood pressure per day, mm Hg	СрАД сут., ночное, мм рт. ст.	Число
Daily HR, bpm	ЧСС, уд/мин сут	Число
Daytime HR, bpm	ЧСС, уд в мин день	Число
Night HR, bpm	ЧСС, уд/мин ночь	Число
Daily SBP TI, systolic blood pressure time index, %	ИВСАД сут., %	Число
Daytime SBP TI, systolic blood pressure time index, %	ИВСАД, % день	Число
Night SBP TI, systolic blood pressure time index, %	ИВСАД % ночь	Число
Daily SBP Area Index, c.u	Индекс площади, ИП САД сут., у.е	Число
Daytime SBP Area Index, c.u	Индекс площади САД, ИП САД, у.е день	Число

Night SBP Area Index, c.u	Индекс площади САД, ИП САД, у.е ночь	Число
Daytime DBP TI, diastolic blood pressure time index, %	ИБ ДАД, % день	Число
Daily DBP Time Index (Daily diastolic blood pressure time index), %	Индекс времени ДАД, ИВДАД сут., %	Число
Night DBP Time Index (Night diastolic blood pressure time index), %	Индекс времени ДАД, ИВДАД ночь, %	Число
Daily SBP Hyperbaric Index, %	Гипербарический Индекс, ГБИ САД сут., %	Число
Daytime SBP Hyperbaric Index, %	Гипербарический Индекс, ГБИ САД день, %	Число
Night SBP Hyperbaric Index, %	Гипербарический Индекс, ГБИ САД ночь, %	Число
SBP MBPS, morning systolic blood pressure surge, mm Hg	Величина Утреннего Подъема САД, мм рт. ст.	Число
DBP MBPS, morning diastolic blood pressure surge, mm Hg	Величина Утреннего Подъема ДАД, мм рт. ст.	Число
MRS SBP, Morning Rise Speed of systolic blood pressure, mm Hg.	Скорость Утреннего Подъема САД, мм рт. ст.	Число
MRS DBP, Morning Rise Speed of diastolic blood pressure, mm Hg.	Скорость Утреннего Подъема ДАД, мм рт. ст.	Число
SBP, mean daily variability, mm Hg	Вар САД сут., мм рт. ст.	Число
DBP, mean daily variability, mm Hg	Вариабельность ДАД сут., мм рт. ст.	Число
Heart Rate daily variability, bpm	Вариабельность ЧСС сут., ударов мин.	Число

12.3. Принципы разметки и верификации

Верификация датасета производится по следующему алгоритму:

1. При помощи использования регулярных выражений в исходных данных обнаруживаются записи, содержащие упоминание каждого из выявляемых

критических событий в соответствии со словарем, содержащим все варианты медицинской идентификации критических событий;

2. Врачи-эксперты просматривают отобранные записи, исключая из них анамнестические данные и нерелевантные события;

3. Перекрестная проверка другим экспертом проводится для подтверждения наличия в электронной медицинской карте пациента записей о соответствующем состоянии. Установленный в медицинской карте диагноз считается достоверным;

4. В случае прохождения всех этапов верификации метаданные записей вносятся в датасет.

12.4. Статистика использования лейблов и классов

По итогу набора данных планируется произвести статистический анализ по классификации артериальной гипертонии.

13. Правила использования и распространения

13.1. Копирайт