

Паспорт набора данных

1. Идентификатор набора данных.

Будет уточнен.

2. Версия набора данных.

1.0

3. Наименование набора данных.

Датасет по задаче характеристика профиля ферритина у мужчин и женщин в московском регионе.

4. Описание набора данных.

4.1. Аннотация

База данных содержит данные о уровне ферритина, взятых у мужчин и женщин разного возраста, в том числе у женщин в разные периоды до-, репродуктивного и постменопаузального возраста, составив в общем 1987 человек. База данных предназначена для выявления профилей уровня ферритина в рамках как острых воспалительных патологических процессов, так и хронической патологии, ряде дифференциальной диагностики, в зависимости от возрастного диапазона у мужчин и женщин и при разных репродуктивных периодах у женщин. а также как фактора прогноза для многих хронических заболеваний, для других научных исследований, в том числе касающихся изучения постпандемического периода COVID-19 и для современных подходов обработки больших данных (машинное обучение, ИИ).

4.2. Клиническая задача

Сбор данных по возрасту полу и уровню ферритина и статистическая обработка результатов с расчетом средних величин возраста и уровня ферритина в группе мужчин и женщин, а также расчет средних показателей уровня ферритина в разных возрастных подгруппах мужчин и женщин и выявление отличий в ранжированных возрастных подгруппах уровня

ферритина представляет научный интерес, позволяющий проспективно сформировать выводы и рекомендации проведенного ферритинового мониторинга в различных сферах исследовательской деятельности. (для выявления профилей уровня ферритина в рамках как острых воспалительных патологических процессов, так и хронической патологии, ряде дифференциальной диагностики, в зависимости от возрастного диапазона у мужчин и женщин и при разных репродуктивных периодах у женщин. а также как фактора прогноза для многих хронических заболеваний, для других научных исследований, в том числе касающихся изучения постпандемического периода COVID-19 и для современных подходов обработки больших данных (машинное обучение, ИИ).

4.3. Назначение набора данных

Датасет предназначен для выявления особенностей уровня ферритина у обследуемых в зависимости от возрастного диапазона у мужчин и женщин и при разных репродуктивных периодах у женщин.

4.4. Нозологии

Сплошной скрининг лиц (популяционное скринирование), обратившихся в медицинское учреждение вне конкретных нозоформ.

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. Б. Пироговская д. 6, стр. 1; Тел.: +7 916 038 28 33; Эл. почта.: gadaev_i_yu@staff.sechenov.ru

5.2. Авторы

- Гадаев Игорь Юрьевич
- Гадаева Ирина Викторовна
- Кузнецов Владимир Михайлович
- Леоненко Валентина Васильевна

6. Порядок предоставления доступа к набору данных

Будет уточнен

7. Тэги

ДАТАСЕТ, ЕДИНАЯ ГОСУДАРСТВЕННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА В СФЕРЕ ЗДРАВООХРАНЕНИЯ, МЕДИЦИНСКАЯ ИНФОРМАЦИОННАЯ СИСТЕМА, уровень и профиль ферритина, возрастные категории, репродуктивные периоды женщин.

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

За период с января 2023 года по февраль 2024 года (1 год 1 месяц)

8.2. Объем набора данных

Произведен набор данных по уровню ферритина в лаборатории «Dialab» у мужчин (n=267) и женщин (n=1720)- всего (n=**1987**), соответственно.

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

Сплошной скрининг лиц (популяционное исследование), обратившихся в медицинское учреждение вне конкретных нозоформ с последующим формированием возрастных групп мужчин и женщин, а также подгрупп женщин по репродуктивным возрастам. Проведены лабораторные исследования уровня ферритина в лаборатории «Dialab».

Критерии включения:

Возрастной диапазон всех обследуемых (мужчины и женщины) составил от 1 до 96 лет.

8.4. Особенности подготовки исследований

Данные пациентов 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

Возрастной диапазон всех обследуемых (мужчины и женщины) составил от 1 до 96 лет. Средний возраст всех обследуемых (мужчины и женщины) составил $39,4 \pm 13,7$ лет

8.6. Характеристики оборудования

Биохимический анализатор лаборатории «диалаб»

8.7. Дата публикации набора данных

Будет уточнена

8.8. Дата обновления набора данных

Без обновления

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

Тип реализующей ЭВМ: IBM PC-совместимый ПК

СУБД: Microsoft Excel 97-2003

ОС: Windows

Объем базы: 1,51 МБ

11.2. Описание файлов в корневой директории

Сводная таблица в формате Excel БД состоит из 9 столбцов

1. Столбец А- нумерация взятой пробы крови
2. Столбец В- дата взятия образца крови
3. Столбец С-номер каждой отдельно взятой пробы
4. Столбец Е/В-дата рождения
5. Столбец G-возраст обследуемого
6. Столбец Н-пол обследуемого
7. Столбец IJ-значение ферритина, выраженного в (мкг/л)
8. Столбец К-заказчик выполнения исследования

12. Обзор разметки данных

12.1. Особенности формата разметки

Номера пациентов, 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 половозрастные характеристики

Таблица 1. Размеченные параметры

- 1Столбец А- нумерация взятой пробы крови
- 2Столбец В- дата взятия образца крови
- 3Столбец С-номер каждой отдельно взятой пробы
- 4Столбец Е/В-дата рождения
- 5Столбец G-возраст обследуемого
- 6Столбец Н-пол обследуемого
- 7Столбец IJ-значение ферритина, выраженного в (мкг/л)
- 8Столбец К-заказчик выполнения исследования

12.3. Принципы разметки и верификации

Верификация датасета производится по следующему алгоритму:

- 1.Отбирались данные в заданный промежуток времени на регулярной основе о содержании уровня ферритина.

2.Врачами-экспертами просматривались отобранные записи, исключая из них анамнестические данные и нерелевантные события;

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

4.Данные записей вносились в датасет.

12.4. Статистика использования лейблов и классов

По итогу набора данных планируется произвести статистический анализ по урону ферритина в зависимости от возрастного диапазона у мужчин и женщин и при разных репродуктивных периодах у женщин.

13. Правила использования и распространения

13.1. Копирайт

ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России
(Сеченовский Университет)