**Создать объединённую модель машинного обучения (нейронную сеть), которая по совокупности признаков из трёх медицинских источников данных предсказывает вероятность тяжёлого сердечно-сосудистого события (модель 1 предсказывает бинарную вероятность, то есть ток 0 или 1, модель 2 вероятность)**

Лян И. В.

Студент 4-го курса бакалавриата

факультета Передовая инженерная школа «Интеллектуальные системы тераностики»

Сеченовского Университета

**Введение**

Сердечно-сосудистые заболевания (ССЗ) остаются одной из ведущих причин смертности во всём мире. Раннее выявление и прогнозирование риска тяжёлых сердечно-сосудистых событий является ключевым фактором снижения смертности и осложнений. Современные методы машинного обучения позволяют использовать комплексные данные о пациенте, включая демографические, клинические и лабораторные показатели, для оценки вероятности развития ССЗ. Сердечная недостаточность (СН) определяется как клинический синдром с симптомами и/или признаками, вызванными структурной и/или функциональной аномалией сердца. Она остается ведущей глобальной причиной смертности, заболеваемости и низкого качества жизни (КЖ), требуя больших ресурсов и затрат на здравоохранение. Поэтому она является областью активных исследований. Острый инфаркт миокарда является одной из основных причин заболеваемости и смертности во всем мире. Клинические исследования показали, что тяжесть повреждения сердца после инфаркта миокарда имеет циркадный характер, с более крупными инфарктами и более худшими исходами у пациентов с утренними приступами. Медицинские работники могут более точно диагностировать заболевания сердца, применяя технологии машинного обучения. Ранняя и точная диагностика имеет решающее значение для снижения заболеваемости и смертности. Машинное обучение стало перспективным инструментом для прогнозирования и выявления различных заболеваний на ранних стадиях. В нескольких исследованиях изучалось применение машинного обучения для прогнозирования и диагностики сердечно-сосудистых заболеваний с использованием различных источников данных, таких как медицинские карты и электрокардиограммы (ЭКГ).

Целью данной работы является создание объединённой модели машинного обучения, которая на основе данных из трёх медицинских источников (кардиологический скрининг, клинические записи по сердечной недостаточности и дополнительные сердечно-сосудистые показатели) предсказывает вероятность тяжёлого сердечно-сосудистого события.

Задачи работы:

Провести подготовку и объединение данных из трёх источников в единый набор признаков.

Построить две модели на основе нейронных сетей:

Модель 1: бинарная классификация с выходом 0 или 1.

Модель 2: вероятностная классификация с выходом в диапазоне 0–1. Обучить и протестировать модели, оценить качество предсказаний и интерпретировать результаты.

**Описание моделей**

1. Подготовка данных

Данные были получены из трёх источников:

cardio\_train.csv — данные кардиологического скрининга.

heart\_failure\_clinical\_records\_dataset.csv — данные по пациентам с сердечной недостаточностью.

heart.csv — дополнительные клинические данные по сердечно-сосудистым событиям.

Для объединения:

Все признаки были приведены к общим именам (age, sex, cholesterol, ap\_hi, ap\_lo, creatinine, serum\_sodium, target).

Возраст преобразован в годы.

Пропущенные значения заменены медианой, а признаки стандартизированы с помощью StandardScaler.

Набор данных был разделён на train (72%), validation (8%) и test (20%) с сохранением пропорций классов (stratify=y).

2. Модель 1: бинарная классификация (0/1)

Архитектура:

Многослойный перцептрон (MLP) с двумя скрытыми слоями (64 и 32 нейрона).

Функция активации ReLU для скрытых слоёв.

Выходной слой: nn.Linear(n\_hidden2, n\_classes) с CrossEntropyLoss.

Особенности:

Выход модели — класс (0 или 1), определяемый максимальным значением logits.

Используется стохастический градиентный спуск (SGD) с learning rate 0.1.

Метрика качества: Accuracy на train/validation/test.

3. Модель 2: вероятностная классификация (0–1)

Архитектура:

Многослойный перцептрон (MLP) с двумя скрытым слоями (64 и 32 нейрона).

ReLU для скрытых слоёв.

Выходной слой: 1 нейрон с функцией Sigmoid (nn.Sigmoid()), что обеспечивает вероятность события в диапазоне 0–1.

Особенности:

Выход модели интерпретируется как вероятность тяжёлого сердечно-сосудистого события.

Потеря: BCELoss() (Binary Cross-Entropy).

Оптимизатор: Adam с lr=0.001.

Метрика: Accuracy с порогом 0.5, но для аналитики чаще используются реальные вероятности.

**Заключение**

Используя экспериментальные методы машинного обучения, основанные на высококачественных больших данных, мы показали, что подход машинного обучения с использованием глубинных нейронных сетей (DNN) может привести к более точному прогнозированию риска смертности и госпитализации пациентов с гипертонией, чем традиционный статистический подход. Необходимо проведение дальнейших исследований для выявления пациентов с высоким риском госпитализации или смерти от сердечно-сосудистых заболеваний после постановки диагноза гипертонии с помощью моделей прогнозирования, разработанных с использованием различных методов глубокого обучения, чтобы обеспечить персонализированное лечение и мониторинг путем распределения ресурсов для пациентов с наибольшим риском.

В результате работы были разработаны две нейронные сети для прогнозирования сердечно-сосудистых событий:

1. **Модель 1 (бинарная)** — определяет наличие события, выдавая метку 0 или 1.
2. **Модель 2 (вероятностная)** — выдаёт вероятность события, что позволяет проводить более гибкий анализ риска и ранжирование пациентов.

Такие модели могут быть использованы как инструмент раннего прогнозирования и поддержки принятия медицинских решений, особенно в задачах скрининга и оценки риска тяжёлых сердечно-сосудистых событий.

**Ссылки**

1. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9697313/?utm_source=chatgpt.com>
2. <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1583459/full?utm_source=chatgpt.com>
3. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12095075/>
4. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12015489/>
5. https://pmc.ncbi.nlm.nih.gov/articles/PMC11769673/