**Создать объединённую модель машинного обучения (нейронную сеть), которая по совокупности признаков из трёх медицинских источников данных предсказывает вероятность тяжёлого сердечно-сосудистого события**

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). Клинические и визуализационные данные собираются из многочисленных источников данных во время встреч с пациентами и часто анализируются независимо друг от друга. Однако такие модели не используют продольную и дополнительную информацию из различных потоков данных. В целом, ИИ позволяет разрабатывать модели, которые объединяют мультимодальные данные из больших популяций для прогнозирования исходов сердечно-сосудистых заболеваний. Первоначальные исследования, интегрирующие как клинические, так и визуализационные характеристики с использованием подхода слияния на основе ИИ, показали превосходную прогностическую эффективность по сравнению с традиционными оценками для прогнозирования основных неблагоприятных сердечных событий.

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 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. Авторы Seung-Jae Lee, Sung-Ho Lee, Hyo-In Choi, Jong-Young Lee, Yong-Whi Jeong, Dae-Ryong Kang, Ki-Chul Sung.   
   **Deep Learning Improves Prediction of Cardiovascular Disease-Related Mortality and Admission in Patients with Hypertension: Analysis of the Korean National Health Information Database.** <https://pmc.ncbi.nlm.nih.gov/articles/PMC9697313/>   
   (доступно по состоянию на 26.10.2025)
2. Авторы Marly van Assen, Amara Tariq, Alexander C Razavi, Carl Yang, Imon Banerjee, Carlo N De Cecco.  
    **Fusion Modeling; Combining Clinical and Imaging Data to Advance Cardiac Care**.   
   <https://pmc.ncbi.nlm.nih.gov/articles/PMC10754220>   
   (доступно по состоянию на 26.10.2025)
3. Авторы [Wei Ruan](https://pubmed.ncbi.nlm.nih.gov/?term=%22Ruan%20W%22%5BAuthor%5D), [Tao Li](https://pubmed.ncbi.nlm.nih.gov/?term=%22Li%20T%22%5BAuthor%5D), [In Hyuk Bang](https://pubmed.ncbi.nlm.nih.gov/?term=%22Bang%20IH%22%5BAuthor%5D), [Jaewoong Lee](https://pubmed.ncbi.nlm.nih.gov/?term=%22Lee%20J%22%5BAuthor%5D), [Wankun Deng](https://pubmed.ncbi.nlm.nih.gov/?term=%22Deng%20W%22%5BAuthor%5D), [Xinxin Ma](https://pubmed.ncbi.nlm.nih.gov/?term=%22Ma%20X%22%5BAuthor%5D), [Cong Luo](https://pubmed.ncbi.nlm.nih.gov/?term=%22Luo%20C%22%5BAuthor%5D), [Fang Du](https://pubmed.ncbi.nlm.nih.gov/?term=%22Du%20F%22%5BAuthor%5D), [Seung-Hee Yoo](https://pubmed.ncbi.nlm.nih.gov/?term=%22Yoo%20SH%22%5BAuthor%5D), [Boyun Kim](https://pubmed.ncbi.nlm.nih.gov/?term=%22Kim%20B%22%5BAuthor%5D), [Jiwen Li](https://pubmed.ncbi.nlm.nih.gov/?term=%22Li%20J%22%5BAuthor%5D), [Xiaoyi Yuan](https://pubmed.ncbi.nlm.nih.gov/?term=%22Yuan%20X%22%5BAuthor%5D), [Katherine Figarella](https://pubmed.ncbi.nlm.nih.gov/?term=%22Figarella%20K%22%5BAuthor%5D), [Yu A An](https://pubmed.ncbi.nlm.nih.gov/?term=%22An%20YA%22%5BAuthor%5D), [Yin-Ying Wang](https://pubmed.ncbi.nlm.nih.gov/?term=%22Wang%20YY%22%5BAuthor%5D), [Yafen Liang](https://pubmed.ncbi.nlm.nih.gov/?term=%22Liang%20Y%22%5BAuthor%5D), [Matthew DeBerge](https://pubmed.ncbi.nlm.nih.gov/?term=%22DeBerge%20M%22%5BAuthor%5D), [Dongze Zhang](https://pubmed.ncbi.nlm.nih.gov/?term=%22Zhang%20D%22%5BAuthor%5D), [Zhen Zhou](https://pubmed.ncbi.nlm.nih.gov/?term=%22Zhou%20Z%22%5BAuthor%5D), [Yanyu Wang](https://pubmed.ncbi.nlm.nih.gov/?term=%22Wang%20Y%22%5BAuthor%5D), [Joshua M Gorham](https://pubmed.ncbi.nlm.nih.gov/?term=%22Gorham%20JM%22%5BAuthor%5D), [Jonathan G Seidman](https://pubmed.ncbi.nlm.nih.gov/?term=%22Seidman%20JG%22%5BAuthor%5D), [Christine E Seidman](https://pubmed.ncbi.nlm.nih.gov/?term=%22Seidman%20CE%22%5BAuthor%5D), [Sary F Aranki](https://pubmed.ncbi.nlm.nih.gov/?term=%22Aranki%20SF%22%5BAuthor%5D), [Ragini Nair](https://pubmed.ncbi.nlm.nih.gov/?term=%22Nair%20R%22%5BAuthor%5D), [Lei Li](https://pubmed.ncbi.nlm.nih.gov/?term=%22Li%20L%22%5BAuthor%5D), [Jagat Narula](https://pubmed.ncbi.nlm.nih.gov/?term=%22Narula%20J%22%5BAuthor%5D), [Zhongming Zhao](https://pubmed.ncbi.nlm.nih.gov/?term=%22Zhao%20Z%22%5BAuthor%5D), [Alemayehu A Gorfe](https://pubmed.ncbi.nlm.nih.gov/?term=%22Gorfe%20AA%22%5BAuthor%5D), [Jochen D Muehlschlegel](https://pubmed.ncbi.nlm.nih.gov/?term=%22Muehlschlegel%20JD%22%5BAuthor%5D), [Kuang-Lei Tsai](https://pubmed.ncbi.nlm.nih.gov/?term=%22Tsai%20KL%22%5BAuthor%5D), [Holger K Eltzschig](https://pubmed.ncbi.nlm.nih.gov/?term=%22Eltzschig%20HK%22%5BAuthor%5D).   
   **BMAL1–HIF2A heterodimer modulates circadian variations of myocardial injury.** <https://pmc.ncbi.nlm.nih.gov/articles/PMC12095075/>   
   (доступно по состоянию на 26.10.2025)
4. Авторы [Shahid Mohammad Ganie](https://pubmed.ncbi.nlm.nih.gov/?term=%22Ganie%20SM%22%5BAuthor%5D), [Pijush Kanti Dutta Pramanik](https://pubmed.ncbi.nlm.nih.gov/?term=%22Pramanik%20PKD%22%5BAuthor%5D), [Zhongming Zhao](https://pubmed.ncbi.nlm.nih.gov/?term=%22Zhao%20Z%22%5BAuthor%5D).   
   **Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets**. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12015489/>   
   (доступно по состоянию на 26.10.2025)
5. Авторы [Alberto Beghini](https://pubmed.ncbi.nlm.nih.gov/?term=%22Beghini%20A%22%5BAuthor%5D), [Antonio Maria Sammartino](https://pubmed.ncbi.nlm.nih.gov/?term=%22Sammartino%20AM%22%5BAuthor%5D), [Zoltán Papp](https://pubmed.ncbi.nlm.nih.gov/?term=%22Papp%20Z%22%5BAuthor%5D), [Stephan von Haehling](https://pubmed.ncbi.nlm.nih.gov/?term=%22von%20Haehling%20S%22%5BAuthor%5D), [Jan Biegus](https://pubmed.ncbi.nlm.nih.gov/?term=%22Biegus%20J%22%5BAuthor%5D), [Piotr Ponikowski](https://pubmed.ncbi.nlm.nih.gov/?term=%22Ponikowski%20P%22%5BAuthor%5D), [Marianna Adamo](https://pubmed.ncbi.nlm.nih.gov/?term=%22Adamo%20M%22%5BAuthor%5D), [Luigi Falco](https://pubmed.ncbi.nlm.nih.gov/?term=%22Falco%20L%22%5BAuthor%5D), [Carlo Mario Lombardi](https://pubmed.ncbi.nlm.nih.gov/?term=%22Lombardi%20CM%22%5BAuthor%5D), [Matteo Pagnesi](https://pubmed.ncbi.nlm.nih.gov/?term=%22Pagnesi%20M%22%5BAuthor%5D), [Gianluigi Savarese](https://pubmed.ncbi.nlm.nih.gov/?term=%22Savarese%20G%22%5BAuthor%5D), [Marco Metra](https://pubmed.ncbi.nlm.nih.gov/?term=%22Metra%20M%22%5BAuthor%5D), [Daniela Tomasoni](https://pubmed.ncbi.nlm.nih.gov/?term=%22Tomasoni%20D%22%5BAuthor%5D).   
   **2024 update in heart failure**.  
   https://pmc.ncbi.nlm.nih.gov/articles/PMC11769673/  
   (доступно по состоянию на 26.10.2025)