

Разработка системы оценки персонализированного эффекта лечения на основе модели смеси экспертов с использованием расстояния Махаланобиса

С. Мхаммад, С. А. Молодяков, Л. В. Уткин

Аннотация— В последние годы в области персонализированной медицины появились практические разработки, основанные на использовании современных технологий и искусственного интеллекта. Большинство этих исследований было сосредоточено на использовании существующих алгоритмов и оценке их эффективности, в то время как некоторые пытались разработать новые методы. В данном исследовании предлагается новая структура для оценки персонализированного эффекта лечения с использованием модели «смеси экспертов» (Mixture of Experts) и сямской сети. Эффективность персонализированного лечения оценивается с помощью показателя условного среднего эффекта лечения. Для учета корреляции между признаками используется метрика расстояния Махаланобиса. Система была протестирована на наборе данных пациентов с рассеянным склерозом с целью совершенствования лечения и определения, какой метод лечения более эффективен для конкретных пациентов. Полученные результаты показывают, что система смогла предсказать более эффективный метод лечения с меньшей среднеквадратичной ошибкой по сравнению с другими моделями, применяемыми в данной области.

Ключевые слова — условный средний эффект лечения, большая языковая модель, смесь экспертов, Сямская сеть, расстояние Махаланобиса.

I. ВВЕДЕНИЕ

Индивидуализированное лечение направлено на адаптацию медицинских процедур к потребностям каждого пациента с учетом его индивидуальных особенностей. В работе [1] показано, что индивидуализированные методы лечения в отличие от усредненных методов демонстрируют значительно более высокие показатели эффективности от 48,7 % до 87 %, а также меньшее количество побочных реакций на лекарственные препараты или их полное отсутствие. К таким методам относятся антидепрессивная терапия с учетом данных фармакогеномики, индивидуализированная диетотерапия и целевые методы лечения рака.

Еще одним преимуществом применения

персонализированного лечения в медицине является улучшение качества лечения заболеваний. В онкологии, как указано в [2], персонализированные методы лечения, основанные на генетических мутациях и характеристиках новообразований, произвели революцию в лечении рака, что привело к повышению показателей выживаемости. Кроме того, персонализированные способы лечения позволяют снизить частоту побочных реакций на лекарственные препараты (Adverse Drug Reactions ADRs) на целых 30%, повышая безопасность пациентов и упрощая процесс лечения.

Существуют различные методы оценки для определения эффективности лечения. Одним из наиболее важных является условный средний эффект лечения (Conditional Average Treatment Effect CATE). В данной статье представлена новая архитектура, использующая фактор CATE для предоставления более эффективных персонализированных решений в лечении рассеянного склероза. С этой целью был предоставлен набор данных по рассеянному склерозу, содержащий сегментацию очагов, выполненную группой медицинских специалистов.

Поскольку данное заболевание подразделяется на несколько клинических подтипов в зависимости от хода течения, а методы лечения варьируются в зависимости от состояния пациентов, существуют также редкие формы, характеризующиеся различными неврологическими нарушениями, которые зачастую диагностируются неверно. Это привело к тому, что традиционные алгоритмы не способны определять наиболее подходящие методы лечения. Исследование было направлено на разработку модели оценки персонализированного эффекта лечения для улучшения результатов терапии и диагностики в этой области.

II. ОБЗОР ПУБЛИКАЦИЙ

В предыдущих исследованиях для повышения эффективности персонализированного лечения были использованы различные алгоритмы. Марковский процесс принятия решений (Markov Decision Process

Статья получена 11 апреля 2026.

Мхаммад Салма — Аспирантка Санкт-Петербургского политехнического университета Петра Великого (СПбПУ) email: mhammad.s@edu.spbstu.ru

Молодяков Сергей Александрович — д.т.н., профессор СПбПУ email: molodyakov_sa@spbstu.ru.

Уткин Лев Владимирович — д.т.н., профессор СПбПУ email: utkin_lv@spbstu.ru.

MDP) был представлен в [3]. MDP был использован для персонализированного планирования лечения сложных зубочелюстных деформаций путем формулирования лечения как последовательного процесса принятия решений, оптимизирующего одновременно несколько клинических целей в приложении многозадачного обучения с подкреплением (Multi-Task Reinforcement Learning MTRL) в здравоохранении. Основным недостатком MDP является то, что для точной оценки функций вероятности перехода и функций вознаграждения для каждого возможного действия требуется значительный объем данных.

В некоторых исследованиях [4] была сделана попытка использовать метод распределения ближайших соседей (Distributional nearest neighbors DNN) для точного исследования оценки и вывода персонального эффекта лечения (Heterogeneous Treatment Effects HTE) в общей непараметрической среде. Используя реальные данные, авторы изучили HTE курения на вес детей при рождении в зависимости от возраста матерей. Однако DNN сталкивается с проблемой «проклятия размерности», поскольку для поиска схожих объектов опирается на евклидово расстояние, но по мере роста числа ковариат (признаков) расстояния между точками становятся примерно равными. Это делает метод неинформативным по мере увеличения размера проблемы [5].

В других исследованиях, в которых использовался метод случайного леса в [6], оценивались индивидуальные эффекты лечения путем непосредственного моделирования реакции с использованием случайных лесов (Random Forests RF) в рамках контрфактуальной модели. Исследователи построили отдельные леса для каждой из групп лечения и заменили леса Бреймана синтетическими лесами.

Мета-модели (мета-Learners) специально разработаны для оценки CATE. Основные типы таких алгоритмов можно описать следующим образом: S-Learner (одномодельный), T-Learner (двухмодельный), X-Learner (Cross-Learner). В [7] были рассмотрены различные Мета-модели и выделены сильные и слабые стороны каждого из них в различных условиях, таких как обработка сложных функций CATE, несбалансированные группы или слабая гетерогенность. В [8] были обобщены Мета-модели для оценки CATE для каждого значения лечения, что позволило сделать вывод о влиянии лечения на результаты. Была продемонстрирована способность X-learner восстанавливать модель истинного состояния и подчеркнут выбор базового обучающего алгоритма по качеству оценки CATE.

С другой стороны, в работе [9] основное внимание уделялось подходам на основе T-Learner для оценки HTE. Было установлено, что T-Learner может быть подвержен переобучению, особенно при несбалансированности размеров выборок в группах, получавших и не получавших лечение, что приводит к «нестабильным» прогнозам эффекта. Кроме того, в нем были выявлены конкретные условия, при которых более простые модели прогнозирования результатов могут превосходить более сложные причинно-следственные модели в задачах, по своей сути являющихся причинно-

следственными. В результате исследование показывает, что каждый подход предпочтителен в разных ситуациях в зависимости от дисперсии оценок вероятности, возможности коррекции смещения и корреляции между результатами и эффектами лечения.

В работе [10] была предложена обучаемая регрессия Надарая-Уотсон (Nadaraya–Watson) для оценки CATE. Основная идея заключалась в обучении ядер регрессии Надарая-Уотсон с использованием нейронной сети с общим весом определенной формы для прогнозирования результатов пациентов из контрольной и лечебной групп. Авторы избегали использования стандартных ядер и обучали модель на контрольной группе, чтобы перенести знания о структуре вектора признаков из контрольной группы в группу лечения. Однако предложенная модель показала основное ограничение, связанное с необходимостью настройки гиперпараметров. Это требует много времени для обучения и валидации.

В таблице 1 приведены лучшие результаты нейронных алгоритмов, применяемых в данной области. В работе [11] авторы объединили оценку коэффициента склонности и прогнозирование исхода в одной сети. В то же время в работе [12] авторы использовали обучение представлению для минимизации расхождений между экспериментальной и контрольной группами.

ТАБЛИЦА I. ЛУЧШИЕ РЕЗУЛЬТАТЫ НЕЙРОННЫХ СЕТЕЙ В ПЕРСОНАЛИЗИРОВАННОМ ЛЕЧЕНИИ

Модель	Метрики		
	MSE	R ²	MAE
Random Forests [6]	0.12	0.82	0.18
Adapting Neural Nets [11]	0.15	0.79	0.22
Deep Counterfactual Networks CFRNet [12]	0.19	0.74	0.28
Deep Learning for Health GANITE [13]	0.23	0.68	0.32

В [14] были использованы большие языковые модели (Large Language Models LLM), для внебольничного лечения пациентов с раком груди в целях персонализации, эмоциональной поддержки и повышения точности. Несмотря на значительные преимущества LLM, они не продемонстрировали высокой эффективности по сравнению с предыдущими исследованиями. Это потребовало поиска более эффективных методов, которые бы использовали возможности LLM. Кроме того, в этой области не было исследований, представляющих использование смеси экспертов (Mixture-of-Experts MoE) с LLM.

III. РЕАЛИЗАЦИЯ СУЩЕСТВУЮЩИХ МЕТОДОВ

A. Формальная постановка задачи

Предлагаемая модель построена на основе некоторых математических формул и факторов, которые составляют основу персонализированного лечения. Наиболее важным из них является условный средний эффект лечения (CATE), который часто обозначается как $\tau(x)$ или $g(x)$. Формально CATE определяется как ожидаемая разница в результатах между группой, прошедшей лечение, и контрольной группой, с учетом набора

наблюдаемых ковариат, как показано ниже [10]:

$$CATE(x) = \tau(x) = E[\hat{Y}_1 - \hat{Y}_0 | X = x], \quad (1)$$

где:

- $\hat{Y}_1$ - результат при лечении (treatment);
- $\hat{Y}_0$ - результат без лечения (control);
- X - характеристики пациентов.

В то время как CATE представляет собой средний эффект для подгруппы, определяемой конкретными ковариатами, индивидуальный эффект лечения (Individual Treatment Effect ITE) относится к причинному эффекту для отдельного человека [15]. Поэтому его можно выразить следующим образом:

$$ITE(x) = \hat{Y}_1 - \hat{Y}_0, \quad (2)$$

В отличие от CATE, который учитывает неоднородность на основе индивидуальных характеристик. Средний эффект лечения (Average Treatment Effect ATE) предоставляет единую общую оценку эффекта лечения [16]. ATE отражает средний эффект лечения для всех пациентов, считается средним значением для ITE и может быть выражен следующим образом:

$$ATE(x) = E[\hat{Y}_1 - \hat{Y}_0]. \quad (3)$$

В. Использование модели LLM

Стандартная модель LLM состоит из нескольких слоев, предназначенных для обработки данных. Начинается она со слоев токенизации и встраивания, которые преобразуют исходные данные во встраивания. Затем блок Transformer, состоящий в основном из прямопроходных сетей (Feedforward Networks FFN), использует эти встраивания для обучения и нормализации. Наконец, выходной блок, декодирование и детокенизация служат для обработки результатов и их представления в удобном для пользователя виде.

В данной работе использовалась модель с открытым

исходным кодом Mistral-7B-Instruct-v0.3 в качестве ядра для преобразования данных в скрытое представление. Модель применялась в качестве средства извлечения признаков, а не в качестве прогнозирующей модели. Для каждого пациента сгенерированное клиническое описание кодировалось в 32-мерный плотный вектор вложения, отражающий скрытые семантические связи между симптомами, неврологическими данными и характеристиками заболевания. Эти вложения были объединены с исходными структурированными клиническими переменными из используемого набора данных для создания гибридного представления признаков. Объединенная матрица признаков впоследствии использовалась в качестве входных данных для последующих этапов работы сети прогнозирования эффекта лечения с целью оценки исхода и моделирования индивидуального эффекта лечения.

С. Использование смеси экспертов MoE

Стандартная структура модели «смеси экспертов» (MoE) состоит из слоя токенизации (Tokenization), слоя эмбединга (Embedding), блока трансформера (Transformer) с MoE, выходного блока, декодирования и детокенизации (Detokenization). Слои токенизации и эмбединга служат для преобразования исходных данных в эмбединги. В свою очередь, выходной блок, декодирование и детокенизация используются для обработки результатов и их представления в понятной для пользователя форме.

Основной блок трансформера включает: элемент маскирования (Masked Self-Attention), резидуальные связи, маршрутизатор и несколько параллельных «экспертов». Элемент маскирования (маскированное самовнимание) используется для построения контекста на основе каждого токена по отношению к другим токенам в последовательности [17]. Для этого используется обучаемая матрица весов W с тремя различными векторами: запрос (Query Q), ключ (Key K) и значение (Value V). Эти векторы определяются как:

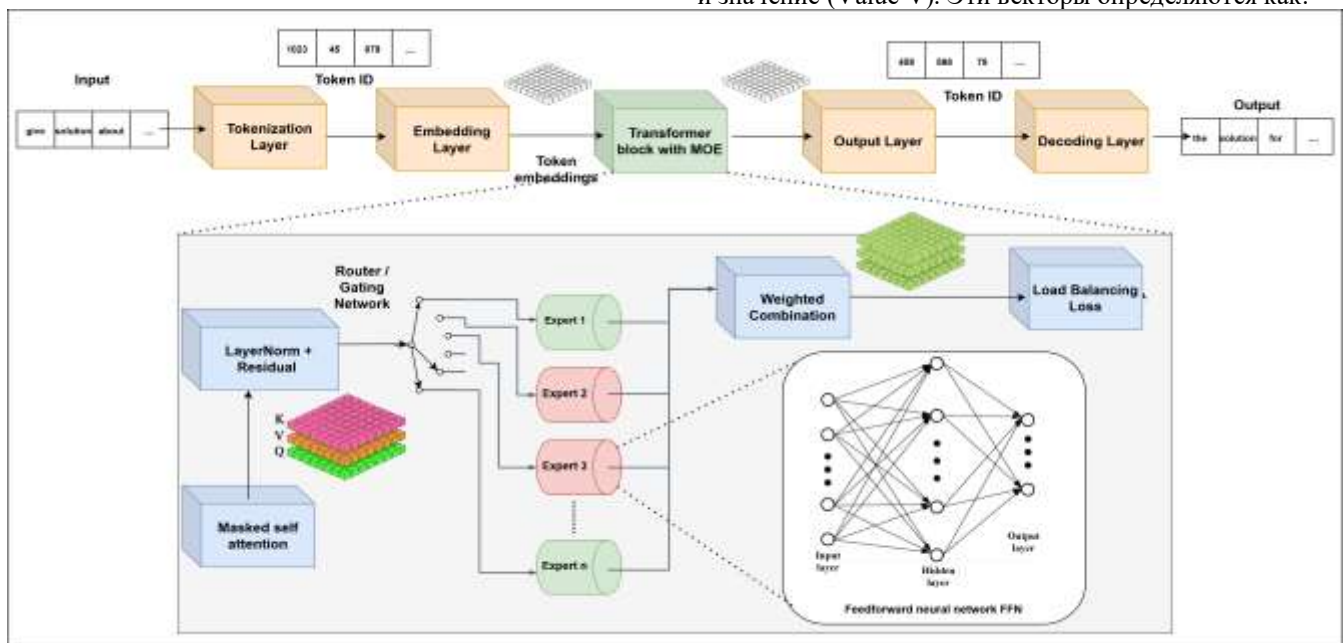


Рис. 1. Основная структура архитектуры MoE в слоях LLM.

$$Q = XW_Q, K = XW_K, V = XW_V, \quad (4)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (5)$$

Эксперты представляют собой отдельные нейронные сети, как правило, прямого распространения (Feed-Forward Networks FFN). FFN использует вектор W_1 , который увеличивает размерность до $4d$, и вектор W_2 , который уменьшает размерность обратно до d . Это можно описать следующим образом:

$$\text{Expert}_k(x) = \text{FFN}(x) = W_{2,k} \cdot \text{GELU}(W_{1,k}x), \quad (6)$$

Каждый эксперт специализируется на конкретном аспекте или подмножестве общей задачи или данных. Маршрутизатор анализирует входные данные и определяет, каких экспертов лучше всего использовать. В зависимости от (4) алгоритм MoE можно описать следующим образом:

$$\text{MoE}(x) = \sum_{k \in S(x)} g_k(x) \cdot \text{Expert}_k(x), \quad (7)$$

где:

- $g_k(x)$ – оценка отбора для эксперта k ;
- $S(x)$ – выбор лучших k экспертов;
- x – эмбединги.

Это означает, что для каждого входного вида данных будет использоваться лишь часть экспертов, что сокращает объем активных вычислений примерно на 30–40 % от общего объема. Основную структуру архитектуры MoE можно описать с помощью рис. 1. В данном исследовании MoE используется в новой структуре для оценки улучшения результатов лечения. Каждой группе схожих пациентов назначаются конкретные эксперты, что позволяет корректировать веса каждого эксперта с учетом особенностей конкретной группы. Это помогает адаптировать лечение в соответствии с конкретными экспертами.

D. Использование глубокой нейронной сети

Глубокие нейронные сети (Deep neural networks DNN) представляют собой ключевой компонент, по структуре близкий к человеческому мозгу. DNN состоят из взаимосвязанных слоёв узлов, предназначенных для обучения на основе данных [18]. Кроме того, в зависимости от требуемой задачи между входным и выходным слоями может быть несколько скрытых слоёв. Скрытые слои могут содержать LayerNorm для нормализации, активационную функцию GELU для плавной и нелинейной активации, а также слои Dropout для предотвращения переобучения. В предлагаемой модели архитектура глубокой нейронной сети определяется следующим уравнением:

$$\text{DNN}(x) = \text{LayerNorm}(W_2 \cdot \text{Dropout}(F(x)) + b_2), \quad (8)$$

$$F(x) = \text{GELU}(\text{LayerNorm}(W_1x + b_1)) \quad (9)$$

где:

- W_1 – матрица весов для экстрактора признаков;
- W_2 – матрица весов для комбинатора признаков;
- b_1, b_2 – корректировка вектора смещения.

Предложено использовать активационную функцию GELU для плавного и нелинейного выхода, слои Dropout для предотвращения переобучения, а также LayerNorm для нормализации. Описание слоев приведено в таблице 2. В предыдущих исследованиях глубокая нейронная сеть (DNN) продемонстрировала высокую способность выявлять сложные медицинские взаимосвязи. DNN позволяет моделировать взаимодействия между признаками и скрытыми закономерностями. В результате наилучшим доступным вариантом интеграции с MoE является использование DNN, поскольку его структура во многом напоминает структуру каждого эксперта.

С другой стороны, DNN требуют большого количества помеченных и чистых данных, которые трудно найти в медицинских данных. Еще одним недостатком, о котором можно упомянуть, является высокая вероятность риска переобучения. В данном исследовании предложено использовать сети в лечебной ветви, оставив разработку этого компонента и решение этих проблем для будущих исследований.

E. Использование сиамской сети

Сиамская сеть — это тип нейронной сети, который обычно используется для классификации. Но модель учится на сходстве между двумя входами, а не классифицирует входы. Эта архитектура состоит из двух или более идентичных подсетей с общими весами [19]. Каждая подсеть обучается скрытому представлению входного вектора. Затем выходы этих параллельных сетей сравниваются с помощью метрики расстояния. Этот процесс позволяет сети развивать знания о сходстве путем анализа различий между характеристиками пар входов и отображения их в многомерное пространство характеристик. Метрика расстояния может быть вычислена с помощью функции D:

$$D(z_1, z_2) = ||G_w(z_1) - G_w(z_2)||, \quad (10)$$

где:

- Z_1 – вектор эмбедингов первого входа;
- Z_2 – вектор эмбедингов второго входа;
- G_w – многослойная сеть с общими весами.

Основная проблема персонализированного лечения заключается в различиях между структурами данных контрольной и лечебной групп. Данные контрольной и лечебной групп для одного и того же пациента не всегда доступны. Предлагается использовать сиамские сети для компенсации недостатка данных путем сбора данных от пациентов с аналогичными параметрами и использования их для корректировки весов в процессе обучения.

F. Использование расстояния Махаланобиса

Расстояние Махаланобиса (Mahalanobis distance MD) измеряет расстояние от точки до центра пространства. Впервые оно было введено П. К. Махаланобисом в 1936 году [20]. MD зависит от ковариационной матрицы всех эмбедингов M (по всем пациентам X), чтобы учесть дисперсию каждой переменной и ковариацию между ними, а также скорректировать различия в масштабе и

зависимости между признаками. М можно найти следующим образом:

$$M = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T, \quad (11)$$

где:

- n – количество эмбедингов;
- x – эмбединг;
- μ – среднее значение эмбедингов.

Чтобы устранить влияние дисперсии и корреляции, в уравнении MD используется обратная матрица ковариации M^{-1} . Уравнение расстояния Махаланобиса можно записать следующим образом:

$$d_{MD}(x_i, x_j)^2 = (x_i - x_j)^T M^{-1} (x_i - x_j), \quad (12)$$

В отличие от евклидова расстояния (Euclidean distance), которое предполагает независимость осей, MD учитывает корреляции между переменными. Для некоррелирующих переменных MD равно евклидовому расстоянию. Однако, если переменные коррелируют, евклидово расстояние может давать вводящие в заблуждение результаты.

IV. АРХИТЕКТУРА РАЗРАБОТАННОЙ СИСТЕМЫ

В данной статье представлена новая архитектура, которая сочетает в себе возможности предыдущих моделей с модификациями для обработки индивидуальных данных и генерации решений, которые независимо поддерживают эффекты лечения, специфичные для каждого человека. В новой структуре

использованием MoE в качестве его части. В этом блоке анализ данных разделен на две параллельные ветви: контрольная ветвь для получения предполагаемого результата без лечения $\hat{Y}_0$ и ветвь лечения для получения предполагаемого результата при лечении $\hat{Y}_1$. Полная структура представлена на рис. 2.

ГБПЛ будет получать данные после этапов предварительной обработки в следующем формате: X — характеристики пациента и T — индикатор лечения. Вначале общий кодер используется для отображения необработанных, неоднородных медицинских характеристик в непрерывное латентное пространство. Это преобразование сжимает информацию и устраняет шум. h_i обозначает латентное представление пациента, обученное общим кодером, и может быть представлено следующим образом:

$$h_i = f_{\theta}(X_i), h_i \in R^k \quad (13)$$

где:

- θ – все настраиваемые параметры кодера f ;
- f_{θ} – нейронная сеть с параметрами θ ;
- k – размерность скрытого представления h_i .

На этапе кодирования вычисляется индекс сходства S_{ij} на основе признаков. Это весовой коэффициент, который оценивает степень взаимосвязи между двумя пациентами на основе их вложений и принимает значения в диапазоне от 0 до 1. Высокое значение указывает на слишком большое сходство пациентов. В данной структуре предложено использовать S для обеспечения того, чтобы у схожих пациентов наблюдались схожие

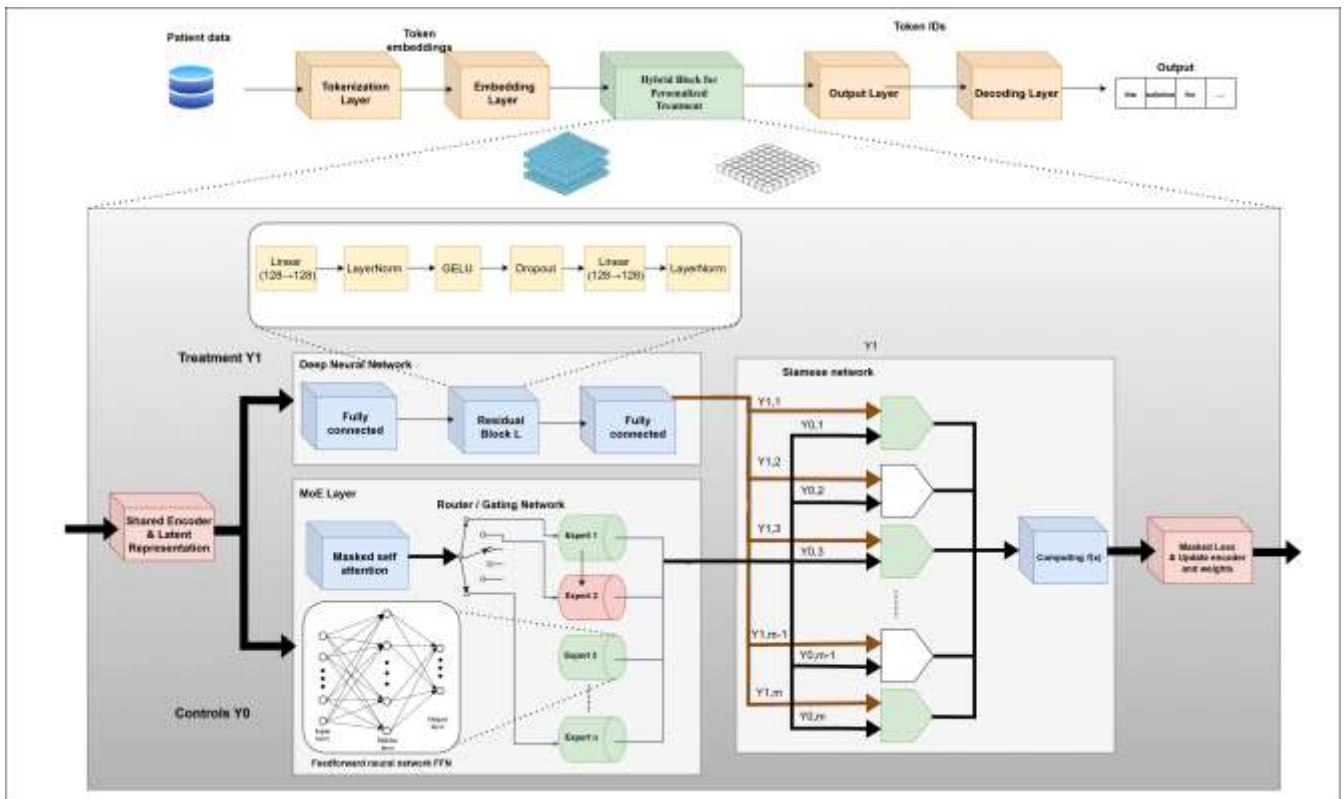


Рис. 2. Основная структура новой архитектуры.

введен Гибридный Блок Персонализированного Лечение (ГБПЛ, Hybrid Block for Personalized Treatment) с

результаты лечения. Для каждой пары пациентов i и j индекс сходства рассчитывался с использованием

расстояния Махаланобиса следующим образом:

$$S_{ij} = \exp\left(-\frac{d_{MD}(h_i, h_j)^2}{2\sigma^2}\right), \quad (14)$$

где:

- d_{MD} – расстояние Махаланобиса в латентном пространстве из (12);
- σ – параметр, определяющий степень «локальности» сходства.

В контрольной ветви для разделения пациентов на группы используется архитектура MoE с сетью гейтинга. Основное уравнение $\hat{Y}_0$ приведено в (7). С другой стороны, Ветвь лечения вводится с использованием глубокой нейронной сети с множеством скрытых слоёв. Эта ветвь обучается взаимодействиям между признаками и лечением, а также результатам вмешательства. Основное уравнение $\hat{Y}_1$ приведено в (8). Скрытое представление h_i используется в обеих ветвях следующим образом:

$$\hat{Y}_1 = DNN(h_i), \hat{Y}_0 = MoE(h_i) \quad (15)$$

На основе (1) для CATE индивидуальный эффект лечения для пациента можно вычислить с помощью описанных характеристик. Однако эти расчеты по-прежнему считаются неудовлетворительными, поскольку они могут привести к очень разным прогнозируемым эффектам лечения для двух очень похожих пациентов. Это приведет к нестабильности и высокой дисперсии результатов. Для регулирования этого процесса было предложено использовать сиамскую сеть.

Полную структуру гибридного персонализированного блока можно увидеть на рис. 2. В таблице 2 приводится полное описание слоев, используемых в новой архитектуре.

ТАБЛИЦА II. Слои нейронных сетей в новой архитектуре

Компонент	Характеристики			
	Роль в системе	Ввод / Вывод	Функция активации	Гипер-параметры
Fully Connected	Извлечение характеристик из данных о пациентах	$R^{d_{in}} \rightarrow R^{d_{out}}$	ReLU, GELU	Units ¹ : 32–512, Dropout rate: 0.1–0.5
Linear (128→128)	Преобразование признаков без изменения измерения	$R^{128} \rightarrow R^{128}$	$y = Wx + b$	W : Weight matrix b: Bias vector
LayerNorm	Стабилизировать и нормализовать тренировки	Одинаковые размеры	Linear (scaling + shift)	Epsilon (σ): 1e-5
Dropout	Регуляризация и предотвращение переобучения	Одинаковые размеры	-	Dropout rate: 0.1–0.5
Shared Encoder	Изучить скрытое представление пациентов	$X_i \rightarrow h_i \in R^k$	ReLU / GELU	2–4 layers, units: 64–256
Control Branch (MoE)	Прогнозировать результат без лечения $\hat{Y}_0$	$h_i \rightarrow \hat{Y}_0$	Expert activations	Experts: 2–8; layers: 1–3 per expert; gating network: softmax
Treatment Branch (DNN)	Прогнозировать результат при лечении $\hat{Y}_1$	$h_i \rightarrow \hat{Y}_1$	GELU (hidden), linear (output)	2–4 hidden layers; Dropout: 0.1–0.3; units: 64–256
Siamese network	Выявить сходства между пациентами	Эмбединги пар h_i, h_j	cosine similarity	similarity index S_{ij} : 0–1
ReLU (Rectified)	Нелинейная активация	Скалярное или векторное	$f(x) = \max(0, x)$	-

Компонент	Характеристики			
	Роль в системе	Ввод / Вывод	Функция активации	Гипер-параметры
Linear Unit				
GELU (Gaussian Error Linear Unit)	Плавная нелинейная активация	Скалярное или векторное	$f(x) = x * \phi(x)^2$	-
Softmax	Масштабирование или ограничение	Скалярное или векторное	распределение вероятностей	-

1. Units: размеры

2. $\phi(x)$: Функция накопительного распределения Гаусса

V. МЕТОДОЛОГИЯ

A. Характеристика набора данных

Эксперименты проводились с использованием набора данных MPT головного мозга при рассеянном склерозе. Он содержит файлы данных MPT 60 пациентов с рассеянным склерозом, включающие результаты ручной сегментации очагов поражения, оценки по шкале EDSS и метаданные о пациентах. Файлы данных MPT представлены в формате NIfTI (Neuroimaging Informatics Technology Initiative). Набор данных включает исходные MPT-изображения и предварительно обработанные файлы сегментации очагов, которые содержат информацию о расположении очагов и оценку прогрессирования заболевания. Эти очаги были сегментированы и проверены тремя радиологами и одним неврологом.

В данной работе мы не уделяем особого внимания обработке изображений, а тестируем нашу систему на предварительно обработанных файлах и данных из таблиц. Вопрос обработки и обучения на MPT-изображениях мы оставили для будущих исследований. Метаданные о пациентах содержат следующие столбцы: пол, возраст, возраст начала заболевания, EDSS, типы лекарств, ведущий симптом, сопутствующие заболевания (да = аномалия и нет = норма) и аномалии при неврологическом обследовании (1 = ненормально и 0 = нормально).

B. Используемые гиперпараметры

Исследованные гиперпараметры предлагаемой модели представлены в таблице 3. Наиболее важными из них являются σ (сигма), определяющая степень сходства двух пациентов, и ϵ (эпсилон), определяющая, насколько модель отдаёт приоритет гладкости кривой CATE по сравнению с прогнозированием фактических исходов.

ТАБЛИЦА III. ИССЛЕДУЕМЫЕ ГИПЕРПАРАМЕТРЫ ПРЕДЛАГАЕМОЙ МОДЕЛИ

Параметр	Диапазон	Описание
Sigma σ	0.5 - 2.0	Область сходства
Epsilon ϵ	0.1 - 1.0	Вес сглаживания CATE
EMBEDDIN_DIM	32 - 128	Размер выученного представления пациента
NUM_EXPERTS	2 - 8	Количество экспертов в ветке управления «Смесь экспертов»
DROPOUT	0.1-0.5	Регуляризация для уменьшения переобучения

Параметр	Диапазон	Описание
BATCH_SIZE	4 - 32	Количество пациентов на одно обновление программы обучения
LEARNING_RATE	1e-5 - 1e-3	Шаг оптимизатора.

С. Способы оценки качества

В данной статье мы рассмотрели несколько показателей эффективности, которые можно разделить на две категории: метрики оценки и функции потерь при обучении. Метрики оценки используются для определения качества ожидаемых результатов лечения пациентов и для оценки качества прогнозируемого эффекта лечения. Используемые метрики можно охарактеризовать следующим образом:

- 1) Среднеквадратичная ошибка (Mean Squared Error MSE) — это среднеквадратичная ошибка между фактическим и прогнозируемым результатом. Наилучшим показателем качества считается ситуация, когда MSE минимальна и близка к 0. Уравнение можно записать следующим образом:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (16)$$

где:

- y_i – реальные результаты;
 - $\hat{y}_i$ – прогнозируемые результаты;
 - N – общее количество исходов.
- 2) Коэффициент детерминации R^2 — показывает, насколько модель объясняет вариацию целевой переменной. Его значение может варьироваться от 0 до 1, где 1 означает идеальное предсказание. Уравнение можно записать следующим образом:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (17)$$

где:

- $\bar{y}$ – среднее значение фактических результатов;

Однако функции потерь используются для управления процессом оптимизации при обучении модели. Используемые функции потерь можно описать следующим образом:

- 1) Значения CATE для схожих пациентов L_{CATE} — исходя из того, что у схожих пациентов должны наблюдаться схожие результаты лечения, мы определили, что гладкость CATE для группы схожих пациентов можно определить следующим образом:

$$L_{CATE} = \sum_{(i,j) \in M} S_{ij} [\tau(x_i) - \tau(x_j)]^2, \quad (18)$$

где:

- S – индекс сходства на основе признаков из (14);
 - $\tau(x_i)$ – факторы CATE из (1).
- 2) Фактическая потеря $L_{Factual}$ — эта потеря оценивает, насколько точно модель предсказывает наблюдаемые результаты. Она считается усовершенствованным вариантом MSE, применяемым исключительно к наблюдаемым

результатам. Она гарантирует, что сеть соответствует реальным данным:

$$L_{Factual} = \sum_i [T_i (\hat{Y}_{1,i} - Y_{1,i})^2 + (1 - T_i) (\hat{Y}_{0,i} - Y_{0,i})^2], \quad (19)$$

- 3) Total Loss (L_{total}): для обеспечения баланса между точностью прогнозирования и плавностью изменения эффектов лечения при обучении контрольной и экспериментальной групп:

$$L_{total} = L_{Factual} + \varepsilon L_{CATE}, \quad (20)$$

где:

- ε – гиперпараметр, определяющий степень сглаживания, из таблицы 3.

VI. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

А. Сравнительный анализ моделей

Чтобы оценить качество предложенной модели, были рассчитаны значения предложенных метрик в процессе обучения модели. Затем проведено сравнение с лучшими результатами, полученными в предыдущих исследованиях в этой области. Мы не рассматриваем возможность применения модели в других областях, учитывая различия в способах обработки данных при персонализации лечения по сравнению с другими задачами классификации и прогнозирования. Результаты, приведенные в таблице 4, показывают, что предложенная модель продемонстрировала лучшие результаты по показателю MSE и показала очень близкие результаты по двум другим метрикам.

ТАБЛИЦА IV. СРАВНЕНИЕ НАИЛУЧШИХ РЕЗУЛЬТАТОВ ПРЕДЫДУЩЕЙ МОДЕЛИ И ПРЕДЛАГАЕМОЙ МОДЕЛИ

Метрика	Результаты	
	предыдущие	Полученные
MSE	0.12	0.10243
MAE	0.18	0.15613
R ²	0.82	0.8145

В. Анализ качества прогнозирования

Для оценки качества прогнозирования мы сосредоточились на значениях показателей $L_{Factual}$, L_{CATE} и L_{total} . Показатель $L_{Factual}$ продемонстрировал относительно низкое значение (0,201). Это свидетельствует о приемлемом уровне точности модели при прогнозировании наблюдаемых результатов по сравнению с фактическими значениями. С другой стороны, L_{CATE} указывает на согласованность ожидаемых эффектов лечения среди схожих пациентов. Он дал значение (0,0924), что подразумевает более согласованные и предсказуемые эффекты. Наконец, L_{total} представляет собой сумму предыдущих показателей (0,2934), определяя общее качество путем объединения обоих.

С. Анализ важности признаков

Анализ значимости признаков показал, что несколько измерений вложений, полученных на основе языковых моделей, оказались одними из наиболее сильных

предикторов результата. Наиболее влиятельной переменной оказалось измерение вложения llm_17 (важность = 0,671), за которым следовали llm_5 (0,323), llm_31 (0,189) и llm_28 (0,167). Среди традиционных клинических переменных наиболее важными предикторами были «Психический статус», «Нистагм» и «Вовлечение ствола мозга». Важность характеристик модели (20 ведущих) можно увидеть в рис. 3.

Выраженность клинических признаков, имеющих неврологическое значение, подтверждает биологическую обоснованность модели. Кроме того, преобладание множественных измерений встраивания свидетельствует о том, что полученные с помощью Mistral представления пациентов отражают клинически значимую информацию, выходящую за рамки одних лишь структурированных переменных.

Сравнение фактических и прогнозируемых распределений исходов показало значительное совпадение, что свидетельствует о том, что модель успешно воспроизвела общую картину исходов. Однако прогнозируемое распределение характеризовалось меньшей дисперсией и более сконцентрированным пиком по сравнению с наблюдаемым распределением, что указывает на тенденцию к регрессии к среднему значению и снижению чувствительности к экстремальным значениям исходов.

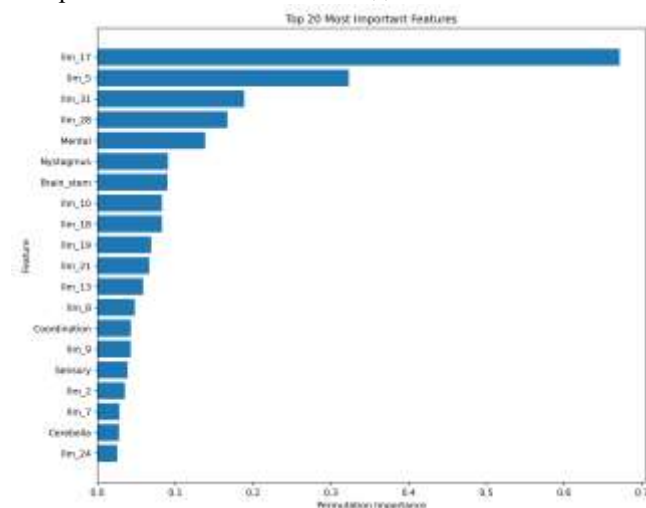


Рис. 3. Важность признаков модели (топ 20).

VII. ЗАКЛЮЧЕНИЕ

Существующие нейронные алгоритмы имеют недостатки, которые ограничивают их эффективность при использовании для персонализации лечения. Для точной оценки лечения каждого пациента требуется обучение на большом объеме данных. С другой стороны, обработка больших данных приводит к таким проблемам, как «проклятие размерности», из-за чего пространство решений становится слишком обширным, чтобы точно оценить все возможные правила принятия решений. Кроме того, большинство существующих моделей не учитывают данные о контроле и лечении и практически не способны давать прогнозы по персонализированному лечению.

В данной статье предложено использовать новую

архитектуру для подбора конкретных экспертов на основе параметров пациента и расчета показателя CATE с целью его применения при прогнозировании эффектов лечения для каждого пациента. Предлагаемая структура учитывает как контрольные, так и лечебные аспекты. Кроме того, она была модифицирована с учетом факторов сходства пациентов для улучшения контроля транзакций и обеспечения плавного и клинически согласованного применения CATE. Для расчета индекса сходства используется расстояние Махаланобиса. Для обучения и тестирования использовался набор данных MPT головного мозга при рассеянном склерозе, чтобы предсказать наиболее эффективное лечение в зависимости от ситуации и характеристик пациента.

Модель демонстрирует хорошие результаты по используемым метрикам. По показателю MSE результат оказался лучше, чем в предыдущих работах, примерно на 10 %. С другой стороны, показатель L_{total} не превысил 2.94 %, что свидетельствует о стабильности ожидаемых эффектов лечения у схожих пациентов. Кроме того, анализ признаков показал, что представления, полученные с помощью модели Mistral, оказывают значительное влияние, позволяя модели улавливать клинически значимую информацию, выходящую за рамки одних лишь структурных переменных. Это подчеркивает важность использования большой языковой модели в качестве ядра предлагаемой системы и открывает широкие перспективы на будущее с использованием более эффективных моделей.

Данная архитектура может быть использована в будущем для улучшения лечения многих заболеваний, таких как различные виды рака, сердечно-сосудистые заболевания и нейродегенеративные заболевания.

Работа выполнена при финансовой поддержке Российского Научного Фонда (проект No 25-11-00021).

БИБЛИОГРАФИЯ

- [1] Maltsev D., Tyravska Y., Shpryakh Y., Kaminsky V., Merzliak S. Individualised Medicine: Application of a Personalised Approach in Clinical Practice // *Journal of Pioneering Medical Sciences*. – 2025. – Т. 14, № 6. – С. 7–19. – DOI: <https://doi.org/10.47310/jpms2025140602>
- [2] Akomolafe O. O., Okoli A. O., Merotiwon D. O. A Systematic Review of Personalized Medicine's Role in Improving Cancer Therapy and Outcomes // *Int. J. Pharma Growth Res. Rev.* – 2025. – Т. 2. – С. 24–29.
- [3] Li Z., Wang L. Multi-task reinforcement learning and explainable AI-driven platform for personalized planning and clinical decision support in orthodontic-orthognathic treatment // *Scientific Reports*. – 2025. – Т. 15, № 1. – DOI: <https://doi.org/10.1038/s41598-025-09236-z>
- [4] Fan Y., Lv J., Wang J. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference // *SSRN Electronic Journal*. – 2018. – DOI: <https://doi.org/10.2139/ssrn.3238897>
- [5] Alagoz O., Hsu H., Schaefer A. J., Roberts M. S. Markov decision processes: A tool for sequential decision making under uncertainty // *Medical Decision Making*. – 2010. – Т. 30, № 4. – С. 474–483. – DOI: <https://doi.org/10.1177/0272989X09353194>
- [6] Athey S., Tibshirani J., Wager S. Generalized random forests. – 2019
- [7] Caron A., Baio G., Manolopoulou I. Estimating individual treatment effects using non-parametric regression models: A review // *Journal of the Royal Statistical Society. Series A: Statistics in Society*. – 2022. – Т. 185, № 3. – С. 1115–1149. – DOI: <https://doi.org/10.1111/rssa.12824>
- [8] Acharki N. et al. Heterogeneous Treatment Effects Estimation: When Machine Learning meets multiple treatments regime // *53èmes Journées de Statistique*. – 2022

- [9] Fernández-Loría C., Provost F. Causal classification: Treatment effect estimation vs. outcome prediction //Journal of Machine Learning Research. – 2022. – T. 23. – № 59. – C. 1-35.
- [10] Konstantinov A., Kirpichenko S., Utkin L. Heterogeneous Treatment Effect with Trained Kernels of the Nadaraya–Watson Regression // Algorithms. – 2023. – T. 16, № 5. – DOI: <https://doi.org/10.3390/a16050226>
- [11] Shi C., Blei D., Veitch V. Adapting neural networks for the estimation of treatment effects //Advances in neural information processing systems. – 2019. – T. 32.
- [12] Alaa A. M., Weisz M., Van Der Schaar M. Deep counterfactual networks with propensity-dropout //arXiv preprint arXiv:1706.05966. – 2017.
- [13] Alaa A. M., Van Der Schaar M. Bayesian inference of individualized treatment effects using multi-task gaussian processes //Advances in neural information processing systems. – 2017. – T. 30.
- [14] Wang Q. et al. Large language models could be applied in personalized out-of-hospital management for breast cancer: a prospective randomized single blind study //Scientific Reports. – 2025. – T. 15. – № 1. – C. 33589.
- [15] Vegetabile B. G. On the distinction between" conditional average treatment effects"(cate) and" individual treatment effects"(ite) under ignorability assumptions //arXiv preprint arXiv:2108.04939. – 2021
- [16] Wolf J. M. tehtuner: An R package to fit and tune models for the conditional average treatment effect // Journal of Open Source Software. – 2023. – T. 8, № 86. – C. 5453. – DOI: <https://doi.org/10.21105/joss.05453>
- [17] Kobayashi G. et al. Incorporating residual and normalization layers into analysis of masked language models //Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. – 2021. – C. 4547-4568.
- [18] Lu Y. et al. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations //International conference on machine learning. – PMLR, 2018. – C. 3276-3285.
- [19] Xu L. Configuration Strategy and Application of Siemens PLC Control Network //Learning & Education. – 2021. – T. 10. – № 3. – C. 40-43.
- [20] Bellet A., Habrard A., Sebban M. A survey on metric learning for feature vectors and structured data //arXiv preprint arXiv:1306.6709. – 2013.

Development of a system for evaluating the personalized effect of treatment based on an expert mixture model using the Mahalanobis distance

S. Mhammad, S. A. Molodyakov, L.V. Utkin

Abstract— In recent years, practical developments based on modern technologies and artificial intelligence have emerged in the field of personalized medicine. Most of these studies focused on using existing algorithms, and evaluating their effectiveness, while a few tried to develop new methods. In this study, a new structure for evaluating the personalized effect of treatment is proposed using Mixture of Experts model and a Siamese network. The effectiveness of personalized treatment is assessed using the conditional average treatment effect metric. To account for the correlation between features, the Mahalanobis distance metric is used. The system was tested on a dataset of patients with multiple sclerosis to improve treatment and determine which treatment method is more effective for specific patients. The results show that the system was able to predict a more effective treatment method with a smaller mean squared error compared to other models used in this field.

Keywords— Conditional Average Treatment Effect, Large Language Models, Mixture-of-Experts MoE, Siamese network, Mahalanobis Distance.

REFERENCES

- [1] Maltsev D., Tyravska Y., Shpryakh Y., Kaminsky V., Merzliak S. Individualised Medicine: Application of a Personalised Approach in Clinical Practice // Journal of Pioneering Medical Sciences. – 2025. – T. 14, № 6. – C. 7–19. – DOI: <https://doi.org/10.47310/jpms2025140602>
- [2] Akomolafe O. O., Okoli A. O., Merotiwon D. O. A Systematic Review of Personalized Medicine's Role in Improving Cancer Therapy and Outcomes // Int. J. Pharma Growth Res. Rev. – 2025. – T. 2. – C. 24–29.
- [3] Li Z., Wang L. Multi-task reinforcement learning and explainable AI-driven platform for personalized planning and clinical decision support in orthodontic-orthognathic treatment // Scientific Reports. – 2025. – T. 15, № 1. – DOI: <https://doi.org/10.1038/s41598-025-09236-z>
- [4] Fan Y., Lv J., Wang J. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference // SSRN Electronic Journal. – 2018. – DOI: <https://doi.org/10.2139/ssrn.3238897>
- [5] Alagoz O., Hsu H., Schaefer A. J., Roberts M. S. Markov decision processes: A tool for sequential decision making under uncertainty // Medical Decision Making. – 2010. – T. 30, № 4. – C. 474–483. – DOI: <https://doi.org/10.1177/0272989X09353194>
- [6] Athey S., Tibshirani J., Wager S. Generalized random forests. – 2019
- [7] Caron A., Baio G., Manolopoulou I. Estimating individual treatment effects using non-parametric regression models: A review // Journal of the Royal Statistical Society. Series A: Statistics in Society. – 2022. – T. 185, № 3. – C. 1115–1149. – DOI: <https://doi.org/10.1111/rssa.12824>
- [8] Acharki N. et al. Heterogeneous Treatment Effects Estimation: When Machine Learning meets multiple treatments regime // 53èmes Journées de Statistique. – 2022
- [9] Fernández-Loría C., Provost F. Causal classification: Treatment effect estimation vs. outcome prediction // Journal of Machine Learning Research. – 2022. – T. 23. – № 59. – C. 1–35.
- [10] Konstantinov A., Kirpichenko S., Utkin L. Heterogeneous Treatment Effect with Trained Kernels of the Nadaraya–Watson Regression // Algorithms. – 2023. – T. 16, № 5. – DOI: <https://doi.org/10.3390/a16050226>
- [11] Shi C., Blei D., Veitch V. Adapting neural networks for the estimation of treatment effects // Advances in neural information processing systems. – 2019. – T. 32.
- [12] Alaa A. M., Weisz M., Van Der Schaar M. Deep counterfactual networks with propensity-dropout // arXiv preprint arXiv:1706.05966. – 2017.
- [13] Alaa A. M., Van Der Schaar M. Bayesian inference of individualized treatment effects using multi-task gaussian processes // Advances in neural information processing systems. – 2017. – T. 30.
- [14] Wang Q. et al. Large language models could be applied in personalized out-of-hospital management for breast cancer: a prospective randomized single blind study // Scientific Reports. – 2025. – T. 15. – № 1. – C. 33589.
- [15] Vegetabile B. G. On the distinction between "conditional average treatment effects"(cate) and "individual treatment effects"(ite) under ignorability assumptions // arXiv preprint arXiv:2108.04939. – 2021
- [16] Wolf J. M. tehtuner: An R package to fit and tune models for the conditional average treatment effect // Journal of Open Source Software. – 2023. – T. 8, № 86. – C. 5453. – DOI: <https://doi.org/10.21105/joss.05453>
- [17] Kobayashi G. et al. Incorporating residual and normalization layers into analysis of masked language models // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. – 2021. – C. 4547–4568.
- [18] Lu Y. et al. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations // International conference on machine learning. – PMLR, 2018. – C. 3276–3285.
- [19] Xu L. Configuration Strategy and Application of Siemens PLC Control Network // Learning & Education. – 2021. – T. 10. – № 3. – C. 40–43.
- [20] Bellet A., Habrard A., Sebban M. A survey on metric learning for feature vectors and structured data // arXiv preprint arXiv:1306.6709. – 2013.