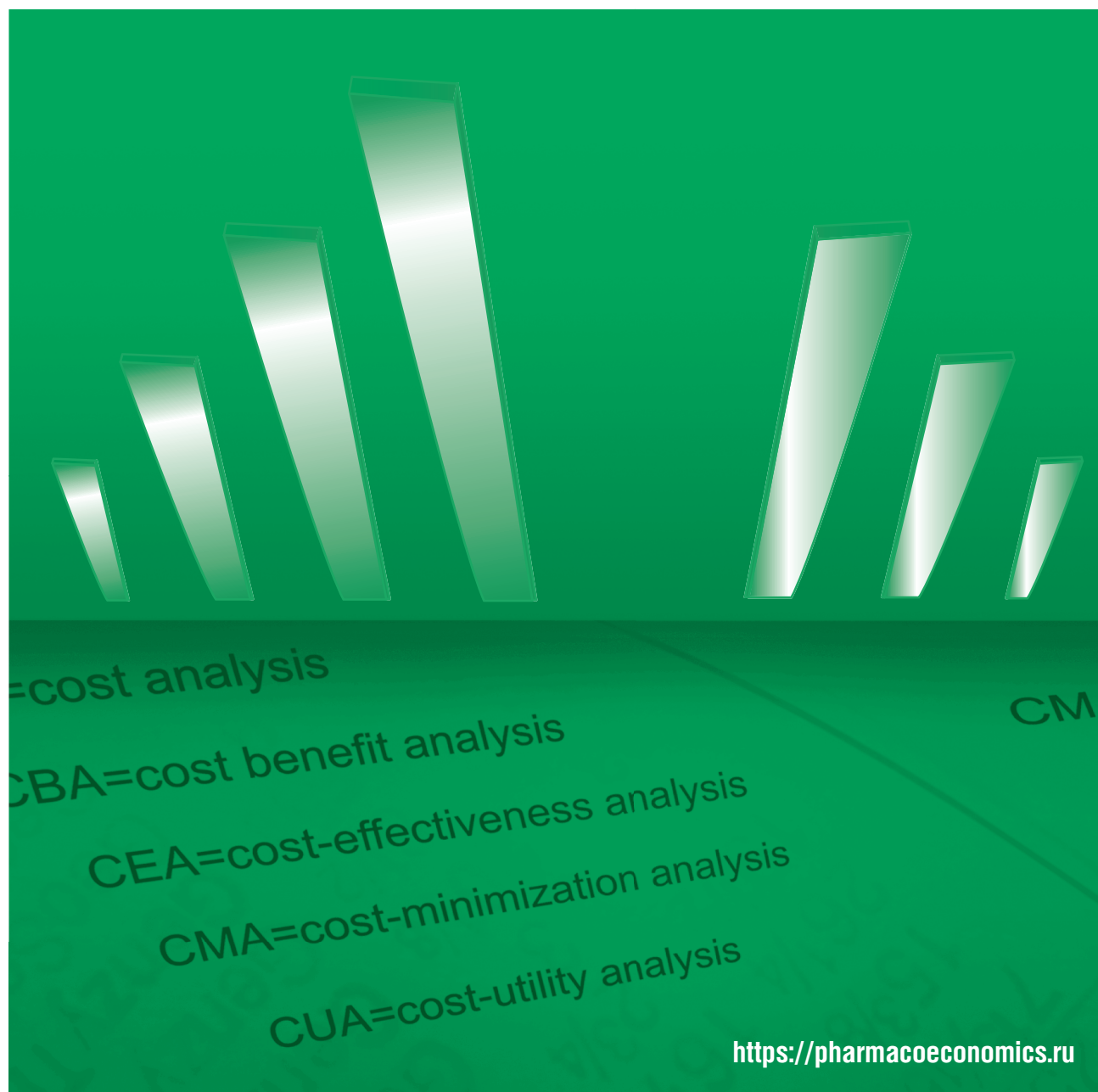


Фармакоэкономика

Современная фармакоэкономика и фармакоэпидемиология



FARMAKOEKONOMIKA
Modern Pharmacoeconomics and Pharmacoepidemiology

2024 Vol. 17 No. 4

№4
2024

Том 17



<https://doi.org/10.17749/2070-4909/farmakoeconomika.2024.283>

ISSN 2070-4909 (print)

ISSN 2070-4933 (online)

Сценарное моделирование процесса назначения лекарственных препаратов детям: применение методов машинного обучения

А.А. Кондрашов¹, М.М. Курашов², Е.Е. Лоскутова^{1,2}

¹ Федеральное государственное автономное образовательное учреждение высшего образования «Московский государственный университет им. М.В. Ломоносова» (Ленинские горы, д. 1, Москва 119991, Россия)

² Федеральное государственное автономное образовательное учреждение высшего образования «Российский университет дружбы народов им. Патриса Лумумбы» (ул. Миклухо-Маклая, д. 6, Москва 117198, Россия)

Для контактов: Александр Андреевич Кондрашов, e-mail: kondrasasha@mail.ru

РЕЗЮМЕ

Цель: определение наиболее подходящего метода машинного обучения для решения задачи назначения лекарственных препаратов (ЛП) детям в условиях медицинской организации, оценка его производительности и потенциала внедрения в системы сценарного моделирования структуры фармацевтической помощи.

Материал и методы. Использованы данные о назначениях ЛП детям из медицинских информационных систем клиник г. Москвы за период с января по декабрь 2023 г. включительно. Данные содержали информацию о пациентах, дате обращения, диагнозах, назначенных ЛП и специальности врача. Проведены предварительная обработка данных, извлечение дополнительных признаков и определение процесса как задачи многометочной классификации. Разработаны и валидированы модели следующих архитектур: полносвязная нейронная сеть (англ. fully connected neural network, FCNN), сверточная нейронная сеть (англ. convolutional neural network, CNN), модель обучения одного классификатора для каждого класса (англ. One-vs-Rest, OvR), градиентный бустинг деревьев решений (англ. eXtreme Gradient Boosting Classifier, XGBC) и случайный лес (англ. RandomForestClassifier, RFC). Оценка моделей проводилась с использованием площади под кривой (англ. area under curve, AUC) рабочей характеристики приемника (англ. receiver operating characteristic, ROC), F1-меры и собственной метрики точности.

Результаты. Модель XGBC показала наилучшие результаты по всем задачам и метрикам. После оптимизации модели и набора данных AUC ROC достигла 0,9993, F1-мера – 0,8318, собственная метрика точности – 0,8548. Модель эффективно предсказывает назначение аналогичных по фармакологическому действию ЛП, позволяя оценивать структуру фармацевтической помощи в рамках конкретного сценария. Оптимизация данных и модели повысила точность прогнозов до 85%.

Заключение. Модель XGBC является наиболее подходящей для решения задачи сценарного моделирования назначения ЛП. Выявленные проблемы с предсказанием похожих ЛП указывают на необходимость дальнейшего совершенствования модели и данных. Полученные результаты свидетельствуют о потенциале интеграции методов машинного обучения в системы сценарного моделирования фармацевтической помощи.

КЛЮЧЕВЫЕ СЛОВА

Назначение лекарственных препаратов, дети, машинное обучение, нейронные сети, многометочная классификация, сценарное моделирование, фармацевтическая помощь, лекарственное обеспечение.

Для цитирования

Кондрашов А.А., Курашов М.М., Лоскутова Е.Е. Сценарное моделирование процесса назначения лекарственных препаратов детям: применение методов машинного обучения. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология.* 2024; 17 (4): 421–431. <https://doi.org/10.17749/2070-4909/farmakoeconomika.2024.283>.

Scenario modeling of the drug prescription process for children: application of machine learning methods

A.A. Kondrashov¹, M.M. Kurashov², E.E. Loskutova^{1,2}

¹ Lomonosov Moscow State University (1 Leninskie Gory, Moscow 119991, Russia)

² Peoples' Friendship University of Russia named after Patrice Lumumba (6 Miklukho-Maklay Str., Moscow 117198, Russia)

Corresponding author: Alexander A. Kondrashov, e-mail: kondrasasha@mail.ru

SUMMARY

Objective: determining the most appropriate machine learning method to solve the problem of drug prescription for children, evaluating its performance and potential for implementation into scenario modeling systems of the pharmaceutical care structure.

Material and methods. The study was based on data on drug prescription for children from medical information systems of Moscow clinics for the period from January to December 2023 including information about patients, the date of treatment, diagnoses, prescribed medications and the doctor's specialty. Preliminary data processing enabled to extract additional features and define the process as a multi-label classification task. The following model architectures were developed and validated: fully connected neural network (FCNN), convolutional neural network (CNN), One-vs-Rest (OvR) classifier, eXtreme gradient boosting classifier (XGBC), and RandomForestClassifier (RFC). The models were evaluated using area under curve (AUC) of receiver operating characteristic (ROC), F1-measure metrics and Custom Accuracy metrics.

Results. The XGBC model showed the best results for all tasks and metrics. After optimizing the model and dataset, the AUC ROC reached 0.9993, the F1-measure was 0.8318, and Custom Accuracy metric was 0.8548. The model effectively predicted the prescription of drugs with similar pharmacological effects, allowing us to evaluate the structure of pharmaceutical care within a specific scenario. Optimization of the data and model increased the accuracy of predictions up to 85%.

Conclusion. The XGBC model proved to be the most appropriate for solving the problem of scenario modeling of drug prescription. The identified problems with predicting similar drugs validate the demand for further improvement of the model and data. Concurrently, the results obtained attest the potential of integrating machine learning methods into scenario modeling systems for pharmaceutical care.

KEYWORDS

Drug prescription, children, machine learning, neural networks, multi-label classification, scenario modeling, pharmaceutical care, drug provision.

For citation

Kondrashov A.A., Kurashov M.M., Loskutova E.E. Scenario modeling of the drug prescription process for children: application of machine learning methods. *FARMAKOEKONOMIKA. Sovremennaya farmakoeconomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology*. 2024; 17 (4): 421–431 (in Russ.). <https://doi.org/10.17749/2070-4909/farmakoeconomika.2024.283>.

ВВЕДЕНИЕ / INTRODUCTION

Сценарное моделирование (СМ) – инструмент стратегического планирования путем создания множества сценариев, описывающих возможное развитие событий при заданном наборе определяющих факторов: на основе больших данных, результатов проведенного анализа, предположений и экспертных оценок создаются модели, используемые для симуляции процессов и анализа их исходов [1]. В условиях цифровизации отраслевых процессов методы СМ нашли применение в различных областях, в здравоохранении они используются для моделирования эпидемиологических эффектов экономической политики и построения прогнозных моделей развития региональных систем здравоохранения [2, 3]. Одними из ключевых аспектов моделей СМ являются их гибкость, модульность и возможность интеграции с другими моделями, такими как системы агент-ориентированного и динамического моделирования, что позволяет достигать более точных конечных результатов [4].

Неотъемлемыми элементами разрабатываемых стратегий (по причине развития технологий и возможностей для оптимизации рутинных процессов) все чаще становятся модели машинного обучения. Например, современные интеллектуальные медицинские информационные системы (МИС) имеют модули распознавания медицинской документации для создания структурированных данных при помощи методов машинного обучения и нейронных сетей для обработки естественного языка. Подобный функционал способствует накоплению больших данных для дальнейших исследований и снижает нагрузку на медицинский персонал при осуществлении документооборота [5]. Анализ данных и внедрение технологий машинного обучения в МИС соответствуют общемировой тенденции по развитию персонифицированной медицины и имеют перспективы для использования в качестве инструмента

прогнозирования неблагоприятных медицинских событий в реальной клинической практике [6].

Отдельный пример использования СМ совместно с методами машинного обучения – моделирование и симуляция тестирования программного обеспечения [7]. Основное преимущество такой комбинации методов заключается в ее способности учитывать сложные и нелинейные зависимости, что позволяет более точно моделировать возможные исходы и их последствия, а также отражать поведение системы в целом. При разработке любых интеллектуальных систем (в особенности связанных с медициной) необходимо обеспечивать объяснимость и воспроизводимость лежащих в их «начинке» алгоритмов. U. Orji и E. Ukwandu в своем исследовании рассматривают применение алгоритмов машинного обучения для прогнозирования суммы расходов на медицинское страхование и подчеркивают важность их объяснимости [8].

Цель – определение наиболее эффективного метода машинного обучения для решения задачи назначения лекарственных препаратов (ЛП) детям в условиях медицинской организации, оценка его производительности и потенциала внедрения в системы СМ структуры фармацевтической помощи.

МАТЕРИАЛ И МЕТОДЫ / MATERIAL AND METHODS

Информационная база / Information base

Информационная база исследования включала данные о назначении ЛП врачами из МИС сети клиник г. Москвы, оказывающими амбулаторную помощь детям, за период с января по декабрь 2023 г. включительно. Данные содержали информацию о пациентах (включая пол, возраст), дате обращения к врачу, поставленных диагнозах, назначенных ЛП и специальности лечащего врача. Сведения о пациентах не содержали информацию, относящуюся к персональным данным: числовой идентификационный номер

Основные моменты

Что уже известно об этой теме?

- ▶ Применение методов машинного обучения (МО) и сценарного моделирования (СМ) в медицине и страховании позволяет разрабатывать эффективные и объяснимые алгоритмы симуляции процессов и анализа их исходов
- ▶ Разработка и интеграция моделей МО в современные медицинские информационные системы используются для оптимизации рутинных процессов медицинских организаций
- ▶ Комбинация СМ с МО позволяет более точно моделировать сложные и нелинейные зависимости, получать результаты, основанные на индивидуальных особенностях пациента и врача, что способствует персонализации подхода к оказанию медицинской и фармацевтической помощи детям

Что нового дает статья?

- ▶ Подробно описаны подходы к предварительной обработке данных о назначениях лекарственных препаратов (ЛП) детям, разработке дополнительных признаков на основе анализа данных информационной базы исследования, а также адаптации полученных значений в вид, удобный для интерпретации моделями МО
- ▶ Процесс назначения ЛП сформулирован как задача многометочной классификации моделей МО, в ходе математического обоснования которой предложена новая метрика оценки точности таких моделей, измеряющая долю истинных меток среди наиболее вероятных предсказаний
- ▶ Сравнительный анализ 25 моделей МО и нейронных сетей различной архитектуры и конфигурации гиперпараметров позволил определить модель градиентного бустинга деревьев решений (XGBC) как наиболее эффективную для многометочной классификации действующих веществ ЛП (точность 85%)

Как это может повлиять на клиническую практику в обозримом будущем?

- ▶ Комбинированная система СМ структуры фармацевтической помощи детям с помощью модели XGBC позволяет получать точные предсказания назначений ЛП, которые могут быть использованы для оценки стоимости терапии при разработке индивидуальных программ лекарственного страхования
- ▶ Полученная модель может быть применена для формирования ассортимента аптек готовых лекарственных форм, предоставляя перечень востребованных в детской практике ЛП для улучшения доступности фармацевтической помощи
- ▶ Интеграция модели в качестве системы поддержки клинических решений способствует персонализации и повышению качества медицинской помощи, позволяя врачу принимать во внимание прогнозируемые значения, основанные на математической интерпретации индивидуальных особенностей пациента

пациента – единственный способ определить принадлежность назначения ЛП.

Методы анализа / Methods of analysis

Все этапы предобработки данных, обучения и оценки модели, а также последующее прогнозирование результатов проводились в среде разработки Jupyter Notebook с использованием языка программирования Python. В процессе анализа использованы библиотеки: xgboost версии 1.6.0 для реализации моделей машинного обучения, keras версии 3.5.0 для реализации моделей глубокого обучения, pandas версии 2.2.2 для обработки данных и scikit-learn версии 1.5.1 как для реализации моделей, так и для оценки их качества. Результаты прогноза сравнивали с истинными данными о назначениях ЛП, а оценка точности модели проводилась с использованием площади под кривой (англ. area under curve,

Highlights

What is already known about the subject?

- ▶ The use of machine learning (ML) and scenario modeling (SM) methods in medicine and insurance enables to develop effective and explicable algorithms for simulating processes and analyzing their outcomes
- ▶ Developing and integrating ML models into modern medical information systems optimize routine processes in healthcare organizations
- ▶ Combining SM with ML enables more accurate modeling of complex dependencies and getting results based on individual patient and physician characteristics that promotes medical and pharmaceutical care for children

What are the new findings?

- ▶ The approaches to preliminary data processing on drug prescriptions for children, development of additional features based on the data analysis from the research information base, as well as adaptation of the obtained results into a form convenient for interpretation by ML models were described in detail
- ▶ Drug prescription process was formulated as a task of multi-label classification for ML models; their mathematical substantiation proposed a new metric to evaluate model accuracy, measuring the proportion of true labels among the most probable predictions
- ▶ Comparative analysis 25 ML models and neural networks allowed us to determine eXtreme gradient boosting classifier (XGBC) as a most effective model for multi-label classification of active drug substances, achieving 85% accuracy

How might it impact the clinical practice in the foreseeable future?

- ▶ Combined system of SM of pharmaceutical care for children with the XGBC model enables to get accurate predictions of drug prescriptions, which can be used for estimating pharmacotherapy costs upon developing individual drug insurance programs
- ▶ The obtained model can help to create an assortment of stock ready-made dosage forms for pharmacies, providing a list of drugs in demand in pediatric practice, and improving availability of pharmaceutical care
- ▶ Integrating the model as a clinical decision support system promotes personalization and improvement of medical care quality by enabling the physician to involve predicted values based on a mathematical interpretation of the patient individual characteristics

AUC) рабочей характеристики приемника (англ. receiver operating characteristic, ROC), F1-меры (англ. F1-score) и самостоятельно разработанной функции Custom Accuracy.

Предварительная обработка данных и определение дополнительных признаков

Исследуемые данные загружались из таких источников, как файлы в текстовом формате для представления табличных данных (англ. comma-separated values, CSV) и файлы Microsoft Excel (Microsoft, США), с использованием библиотеки pandas. В начале проведены предварительная обработка (приведение типов и удаление некорректных значений) и группировка данных по ряду характеристик, таких как идентификационный номер пациента, дата и время приема, специальность врача и тип заключения. Таким образом, таблица стала отражать информацию не по отдельным

назначениям одного ЛП пациенту, а по нескольким назначениям, сгруппированным в одну запись, которую можно определить как прием у врача определенной специальности.

В ходе дальнейшего анализа и обработки полученных данных были выведены дополнительные признаки:

- категориальное значение «Тип приема» (первичный или повторный) – если с момента последнего приема пациента прошло менее 30 дней (согласно времени действия направления на прием врача или исследование), прием обозначался как повторный;
- категориальное значение «Сезон» (времена года);
- число назначенных диагнозов в рамках одного приема – «Количество диагнозов»;
- общее число посещений пациентом клиники на момент приема – «Количество посещений».

Также проведено дополнение данных о назначенных ЛП информацией о действующих веществах ЛП и их фармакологических группах. Для адаптации категориальных данных в вид, удобный для восприятия моделями машинного обучения и нейронными сетями, использовался процесс быстрого кодирования (англ. one-hot encoding) категориальных переменных в бинарные векторы. Числовые признаки были нормализованы и стандартизованы методом StandartScaler библиотеки scikit-learn. Итоговая таблица содержала 134 951 запись о назначениях ЛП детям за 2023 г. и 4914 столбцов, относящихся к полученным признакам и целевым значениям. Структура полученного набора данных до применения нормализации отображена в **таблице 1**. Количество полученных столбцов категориальных данных определено после операций предварительной обработки и кодирования признаков.

Определение задачи

Пусть у нас есть набор данных, содержащий информацию о назначениях. Наша цель – предсказать список ЛП, которые следует назначить пациенту на основе его характеристик и диагнозов. Для начала введем следующие обозначения:

- $X \in R^{n \times d}$: матрица признаков, где n – количество пациентов, а d – количество признаков (например, возраст, пол, диагнозы и т.д.).
- $Y \in \{0,1\}^{n \times L}$: матрица меток, где L – количество различных ЛП, а $y_{ij}=1$ означает, что пациенту i был назначен препарат j .

Задачу можно представить как поиск функции $f: R^d \rightarrow [0,1]^L$, которая принимает на вход вектор признаков пациента x_i и возвращает вектор вероятностей $\hat{p}_i$ назначения каждого ЛП:

$$\hat{p}_i = f(x_i) = \{\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{iL}\},$$

где $\hat{p}_{ij}$ – вероятность назначения препарата j пациенту i .

Для обучения модели мы минимизируем функцию потерь, которая измеряет разницу между истинными метками Y и предсказанными вероятностями $\hat{P}$:

$$\mathcal{L}(Y, \hat{P}) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^L [y_{ij} \log(\hat{p}_{ij}) + (1 - y_{ij}) \log(1 - \hat{p}_{ij})],$$

где $\hat{p}_{ij}$ – предсказанная моделью вероятность того, что препарат j будет назначен пациенту i .

После того как модель предсказывает вероятности назначения препаратов, необходимо принять решение о том, какие препараты

Таблица 1. Структура полученного набора данных для обучения и оптимизации моделей

Table 1. Structure of the obtained data set for learning and optimizing models

№ / No.	Название / Name	Тип признака / Type of feature	Количество столбцов / Number of columns
Признаки / Features			
1	Код МКБ-10 / ICD-10 code	Категориальный / Categorical	2291
2	Диагноз / Diagnosis	Категориальный / Categorical	2426
3	Пол / Gender	Категориальный / Categorical	2
4	Специальность врача / Doctor's specialty	Категориальный / Categorical	40
5	Тип заключения / Type of conclusion	Категориальный / Categorical	26
6	Возраст (в виде десятичной дроби) / Age (as a decimal)	Количественный непрерывный / Quantitative continuous	1
7	Тип приема / Reception type	Категориальный / Categorical	2
8	Сезон / Season	Категориальный / Categorical	4
9	Количество диагнозов / Number of diagnoses	Количественный дискретный / Quantitative discrete	1
10	Количество посещений / Number of visits	Количественный дискретный / Quantitative discrete	1
Целевые значения / Targeted values			
1	Лекарственный препарат (торговое наименование) / Medication (trade name)	Категориальный / Categorical	953
2	Действующие вещества / Active substances	Категориальный / Categorical	611
3	Фармакологические группы / Pharmacological groups	Категориальный / Categorical	120

Примечание. МКБ-10 – Международная классификация болезней 10-го пересмотра.

Note. ICD-10 – International Classification of Diseases, 10th revision.

действительно назначить пациенту. Это делается с использованием пороговой функции:

$$\hat{y}_{ij} = \begin{cases} 1, & \text{если } \hat{p}_{ij} \geq \tau, \\ 0, & \text{если } \hat{p}_{ij} < \tau, \end{cases}$$

где τ – порог, который определяет, при какой вероятности назначение препарата считается оправданным. Обычно $\tau = 0,5$, но может варьироваться в зависимости от специфики задачи. В нашем случае τ может быть определен экспериментально в рамках оценки показателей каждой из моделей.

После обучения модели и формирования прогнозов необходимо оценить ее качество. В многометочной классификации используются следующие метрики.

Метрика AUC ROC измеряет, насколько хорошо модель различает разные классы. Если AUC ROC близка к 1, модель почти идеальна, если же ближе к 0,5, она не лучше случайного угадывания:

$$\text{AUC ROC} = \int_0^1 \text{TPR}(t) d\text{FPR}(t),$$

где $\text{TPR}(t)$ – доля предсказанных верно положительных примеров среди всех реальных положительных примеров (англ. True Positive Rate) при пороге t ; $\text{FPR}(t)$ – доля неверно предсказанных положительных примеров среди всех реальных отрицательных примеров (англ. False Positive Rate) при пороге t ; t – порог, который изменяется от 0 до 1, позволяя строить ROC-кривую.

Дисбаланс классов является одной из особенностей данной задачи, из-за чего для оценки качества моделей следует использовать микроусредненный (англ. micro-averaged) показатель AUC ROC, который будет отражать общее качество классификации.

Метрика F1-score помогает понять, насколько хорошо модель находит все нужные примеры (англ. recall) и насколько точно она это делает (англ. precision):

$$\text{F1-score} = \frac{1}{n} \sum_{i=1}^n \frac{2 \times \text{precision}_i \times \text{recall}_i}{\text{precision}_i + \text{recall}_i},$$

$$\text{где } \text{precision}_i = \frac{|\hat{y}_i \cap y_i|}{|\hat{y}_i|}, \text{ recall}_i = \frac{|\hat{y}_i \cap y_i|}{|y_i|}.$$

В случае задачи назначения ЛП следует использовать взвешенную F1-меру (англ. weighted F1-score) для учета важности каждого класса пропорционально частоте его присутствия в данных.

Метрика точности по совпадению предсказаний (англ. Custom Accuracy, CA) разработана под конкретную задачу:

$$\text{CA} = \frac{1}{n} \sum_{j=1}^n \frac{|\text{ArgTop}_{|\text{ind}_j|}(y_{\text{pred}_j}) \cap \{i \mid y_{\text{true}_{ji}} = 1\}|}{|\{i \mid y_{\text{true}_{ji}} = 1\}|},$$

где y_{pred_j} – вектор предсказанных вероятностей для образца j ; $\text{ind}_j = \{i \mid y_{\text{true}_{ji}} = 1\}$ – множество индексов истинных меток для образца j ; $\text{ArgTop}_k(y_{\text{pred}_j})$ – множество индексов топ- k наибольших значений вектора y_{pred_j} , где $k = |\text{ind}_j|$.

Пояснения:

– пересечение множеств $\text{ArgTop}_{|\text{ind}_j|}(y_{\text{pred}_j}) \cap \{i \mid y_{\text{true}_{ji}} = 1\}$ определяет количество правильных предсказаний среди топ- k вероятностей;

– деление на размер множества истинных меток $|\{i \mid y_{\text{true}_{ji}} = 1\}|$ нормализует значение точности для каждого образца;

– суммирование по всем образцам и деление на их количество $1/n \sum_{j=1}^n$ дает среднюю точность по всему набору данных.

Для минимизации функции потерь и оптимизации параметров модели используются методы градиентного спуска, такие как стохастический градиентный спуск (англ. stochastic gradient descent, SGD) или его вариации. В процессе оптимизации модель корректирует свои параметры θ , чтобы минимизировать функцию потерь $\mathcal{L}(Y, \hat{P})$:

$$\theta = \theta - \eta \nabla_{\theta} \mathcal{L}(Y, \hat{P}),$$

где η – скорость обучения (англ. learning rate), а $\nabla_{\theta} \mathcal{L}$ – градиент функции потерь по параметрам модели.

Таким образом, процесс назначения ЛП пациенту можно рассматривать как задачу многометочной классификации, где модель обучается на данных пациентов и их характеристиках, предсказывает вероятности назначения ЛП, а затем принимает окончательное решение о назначении на основе порогового значения. Оценка качества модели проводится с использованием таких метрик, как разработанная под контекст данной задачи метрика точности Custom Accuracy, F1-score и AUC ROC, что позволяет сделать вывод об эффективности ее работы.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ / RESULTS AND DISCUSSION

Разработка моделей / Model development

В ходе анализа научной литературы были определены алгоритмы машинного обучения, которые могут быть потенциально эффективны для решения задачи назначения ЛП. В итоге выбраны и адаптированы описанные ниже модели.

Полносвязная нейронная сеть (англ. fully connected neural network, FCNN) библиотеки keras

Архитектура: три скрытых слоя с 256, 128 и 64 нейронами с функцией активации ReLU, выходной многометочный слой с функцией активации sigmoid. Компилировалась с оптимизатором Adam и функцией потерь binary_crossentropy для независимых бинарных классификаций каждой метки. Подобные методы математического моделирования часто используются в биомедицинских исследованиях [9].

Сверточная нейронная сеть (англ. convolutional neural network, CNN) библиотеки keras

Архитектура: сверточный слой с 64 фильтрами, функцией активации ReLU, за которым следует слой субдисcretизации (MaxPooling) для уменьшения размерности, преобразование в одномерный вектор с помощью слоя Flatten, который передается в два полносвязных слоя с 128 и 64 нейронами (также с активацией ReLU), выходной многометочный слой и оптимизатор аналогично FCNN.

Модель обучения одного классификатора для каждого класса (англ. One-vs-Rest, OvR) библиотеки scikit-learn

Архитектура: базовый классификатор – логистическая регрессия, максимальное количество итераций – 1000. Модель OvR использовали для обучения моделей неинвазивной диагностики АКТГ-зависимого¹ эндогенного гиперкортицизма на основе клинических данных [10].

¹ АКТГ – адренокортикотропный гормон.

Модель градиентного бустинга деревьев решений (англ. eXtreme Gradient Boosting Classifier, XGBC) библиотеки xgboost²

Архитектура: модель стандартной реализации библиотеки xgboost, где количество деревьев в ансамбле – 100 (гиперпараметр n_estimators), максимальная глубина каждого дерева – 6 (гиперпараметр max_depth), гиперпараметр скорости обучения (learning_rate) – 0,3. Метрика оценки качества – кросс-энтропия (logloss).

Модель ансамбля из 20 деревьев решений (англ. RandomForestClassifier, RFC) библиотеки scikit-learn

Архитектура: модель стандартной реализации библиотеки scikit-learn, представляющая собой ансамбль из 20 деревьев решений (гиперпараметр n_estimators), где каждое дерево обучается на случайной подвыборке данных и признаков, гиперпараметр максимальной глубины деревьев не задан, количество признаков (гиперпараметр max_features) определяется автоматически. Модели градиентного бустинга и случайного леса использовали для определения структуры фармакотерапии при гипертонической болезни [11].

Построенные модели были обучены на тренировочной выборке размером 121 455 значений, что составляет 90% от общих данных, где оставшиеся 10% составляет валидационная выборка. В качестве целевых определялись значения множеств классов, принадлежащих таким признакам ЛП, как торговые наименования, действующие вещества и фармакологические группы.

Анализ производительности / Performance analysis

В таблице 2 представлены результаты сравнительного анализа производительности моделей по ключевым метрикам в рамках решения разных задач.

Модель XGBC демонстрирует наилучшие результаты во всех задачах и по всем метрикам. Согласно значению метрики Custom Accuracy, XGBC с вероятностью 58% предсказывает все фармакологические группы ЛП, необходимые для лечения конкретного пациента. Метрика F1-score показывает значительное снижение по мере увеличения сложности задачи, особенно в определении торговых наименований: высокий показатель AUC ROC при значениях немногим выше 0,5 (50%). F1-score и Custom Accuracy указывают на проблему дисбаланса классов в имеющихся данных, которую для успешного решения задачи назначения ЛП необходимо доработать путем балансировки классов, улучшения параметров моделей многометочной классификации или оптимизации обучающего набора данных.

Оптимизация набора данных / Data set optimization

P. Silva et al. [12] в своем исследовании использования методов машинного обучения для определения и прогнозирования назначений ЛП разделили процесс назначения ЛП на сценарии и ограничили количество категорий ЛП до 40. Авторы уделили значительное внимание разработке и отбору признаков, которые могут улучшить точность предсказаний моделей: в рамках каждой из записей были созданы отдельные столбцы в виде

Таблица 2. Сравнительный анализ производительности моделей по ключевым метрикам до оптимизации

Table 2. Comparative analysis of the efficiency of models by key metrics before optimization

Метрика / Metrics	Модель / Model				
	CNN	FCNN	OvR	RFC	XGBC
Задача: определение фармакологических групп / Task: detection of pharmacological groups					
AUC ROC	0,9722	0,9719	0,9632	0,9208	0,9759
F1-score	0,5360	0,5270	0,5090	0,5329	0,5411
Custom Accuracy	0,5675	0,5574	0,5577	0,5522	0,5808
Задача: определение действующих веществ / Task: detection of active substances					
AUC ROC	0,9567	0,9552	0,9483	0,9051	0,9621
F1-score	0,4890	0,4785	0,4622	0,4829	0,4958
Custom Accuracy	0,5113	0,5024	0,5057	0,4952	0,5204
Задача: определение торговых наименований / Task: detection of trade names					
AUC ROC	0,9402	0,9388	0,9301	0,8907	0,9504
F1-score	0,4327	0,4253	0,4105	0,4283	0,4389
Custom Accuracy	0,4708	0,4633	0,4677	0,4521	0,4809

Примечание. CNN (англ. convolutional neural network) – сверточная нейронная сеть; FCNN (англ. fully connected neural network) – полносвязная нейронная сеть; OvR – (англ. One-vs-Rest) – модель обучения одного классификатора для каждого класса; RFC (англ. RandomForestClassifier) – модель случайного леса; XGBC – (англ. eXtreme Gradient Boosting Classifier) – модель градиентного бустинга деревьев решений; AUC ROC – площадь под кривой рабочей характеристики приемника; F1-score – метрика, позволяющая определить, насколько хорошо модель находит все нужные примеры (англ. recall) и насколько точно она это делает (англ. precision); Custom Accuracy – метрика точности по совпадению предсказаний (разработана авторами).

Note. CNN – convolutional neural network; FCNN – fully connected neural network; OvR (One-vs-Rest) – a model for training one classifier for each class; RFC (RandomForestClassifier) – a random forest model; XGBC (eXtreme Gradient Boosting Classifier) – a model for gradient boosting of decision trees; AUC ROC – area under the receiver operating characteristic curve; F1-score – a metric to determine how well the model finds all the necessary examples (recall) and how accurately it works (precision); Custom Accuracy – a metric of accuracy based on the coincidence of predictions (developed by the authors).

² xgboost – это название библиотеки языка программирования Python. Если в коде указано XGBoost (вместо xgboost), интерпретатор кода выдаст ошибку, поэтому в тексте указываем xgboost.

трех наиболее часто назначаемых ЛП при заболевании, а также определена иерархическая структура заболеваний по кодам Международной классификации болезней 10-го пересмотра с анализом наиболее часто назначаемых препаратов [12]. Подобные оптимизации являются одним из способов помочь моделям машинного обучения справиться с дисбалансом классов и выявить новые скрытые зависимости между признаками для повышения предсказательной способности.

Наше исследование назначения ЛП детям не имеет ограничений по выборкам и сценариям: количество категорий целевых признаков, описанных в таблице 1, не ограничено. Один из способов помочь модели сузить круг возможных вариантов категорий для предсказания – определить спектр фармакологических групп ЛП, назначаемых врачами в качестве фармакотерапии поставленного пациенту диагноза и дополнить им тренировочную выборку. Информация из полученных 120 столбцов категорий фармакологических групп ЛП, назначаемых пациенту, основана на анализе больших данных информационной базы исследования и отражает особенности поведения врачей при определении структуры фармацевтической помощи в рамках конкретного диагноза. Подобное решение позволяет модели сместить фокус с общей картины наиболее релевантных ЛП и сконцентрироваться на других зависимостях.

Оценка полученных результатов / Evaluation of obtained results

В таблице 3 показаны результаты сравнительного анализа производительности оптимизированных моделей по ключевым метрикам в рамках решения разных задач, где целевыми значениями являются действующие вещества и торговые наименования ЛП.

Оптимизация позволила добиться значительного улучшения предсказательной способности: модель XGBC демонстрирует лучшие результаты по каждой из ключевых метрик. Предпринятые меры по дополнению тренировочного набора данных новыми признаками с целью снижения влияния дисбаланса классов на предсказательную способность модели оказались эффективными, однако не являются полноценным решением данной проблемы.

Точность моделей, определяющих действующее вещество, выше, чем у моделей, определяющих конкретный ЛП, по нескольким причинам. Во-первых, определение действующего вещества – более общая задача: меньшее количество действующих веществ по сравнению с конкретными ЛП упрощает их классификацию. Моделям легче найти различия между несколькими действующими веществами, чем между большим количеством различных препаратов, которые могут иметь схожие составы или базироваться на тех же биологически активных соединениях. Во-вторых, количество обучающих данных для каждого класса (в данном случае для каждого действующего вещества) может быть более сбалансированным по сравнению с ЛП. Проблема дисбаланса классов может возникнуть, когда одни классы значительно более многочисленны, чем другие. В случае конкретных ЛП этот дисбаланс выражен сильнее, что затрудняет точную классификацию, особенно для менее распространенных препаратов. Кроме того, множество ЛП могут содержать одинаковые действующие вещества, что создает перекрытие между классами и усложняет задачу для модели, пытающейся различить конкретные препараты. Это приводит к тому, что модели, обученные для предсказания торговых наименований ЛП, часто сталкиваются с большей неопределенностью и, как следствие, с меньшей точностью.

Характеристика полученного решения / Characteristics of resulting solution

В результате проведенного исследования была выделена модель XGBC как наилучший способ решения задачи многометочной классификации ЛП. Полученная модель XGBC для предсказания действующих веществ была оптимизирована при помощи настройки гиперпараметров: количество деревьев (гиперпараметр `n_estimators`) увеличено до 500, а глубина деревьев (гиперпараметр `max_depth`) – до 10; для уменьшения риска переобучения при тренировке модели использовалась фиксированная доля данных (гиперпараметр `subsample=0,9`), а также ранняя остановка процесса обучения (гиперпараметр `early_stopping_rounds=10`), если улучшения ее показателей не наблюдаются в течение 10 итераций.

Таблица 3. Сравнительный анализ производительности моделей по ключевым метрикам после оптимизации

Table 3. Comparative analysis of the efficiency of models by key metrics after optimization

Метрика / Metrics	Модель / Model				
	CNN	FCNN	OvR	RFC	XGBC
Задача: определение действующих веществ / Task: detection of active substances					
AUC ROC	0,9953	0,9947	0,9894	0,9735	0,9981
F1-score	0,6954	0,6816	0,6266	0,6954	0,7431
Custom Accuracy	0,7330	0,7135	0,7362	0,7374	0,7741
Задача: определение торговых наименований / Task: detection of trade names					
AUC ROC	0,9948	0,9937	0,9876	0,9627	0,9977
F1-score	0,6267	0,6163	0,5505	0,6308	0,6843
Custom Accuracy	0,6693	0,6491	0,6743	0,6769	0,7174

Примечание. CNN (англ. convolutional neural network) – сверточная нейронная сеть; FCNN (англ. fully connected neural network) – полносвязная нейронная сеть; OvR – (англ. One-vs-Rest) – модель обучения одного классификатора для каждого класса; RFC (англ. RandomForestClassifier) – модель случайного леса; XGBC – (англ. eXtreme Gradient Boosting Classifier) – модель градиентного бустинга деревьев решений; AUC ROC – площадь под кривой рабочей характеристики приемника; F1-score – метрика, позволяющая определить, насколько хорошо модель находит все нужные примеры (англ. recall) и насколько точно она это делает (англ. precision); Custom Accuracy – метрика точности по совпадению предсказаний (разработана авторами).

Note. CNN – convolutional neural network; FCNN – fully connected neural network; OvR (One-vs-Rest) – a model for training one classifier for each class; RFC (RandomForestClassifier) – a random forest model; XGBC (eXtreme Gradient Boosting Classifier) – a model for gradient boosting of decision trees; AUC ROC – area under the receiver operating characteristic curve; F1-score – a metric to determine how well the model finds all the necessary examples (recall) and how accurately it works (precision); Custom Accuracy – a metric of accuracy based on the coincidence of predictions (developed by the authors).

Оптимизированная модель XGBC имела следующие значения метрик оценки качества: AUC ROC 0,9993; F1-score 0,8318; Custom Accurasy 0,8548.

Для валидации модели и получения наглядного представления о ее предсказательной способности при помощи инструментов библиотеки matplotlib языка программирования Python была построена матрица ошибок (рис. 1), которая показывает распределение истинных и ошибочных классификаций, что помогает выявить возможные проблемные зоны и направления для дальнейшего улучшения модели. Для построения графика было выбрано 20 наиболее часто назначаемых действующих веществ, что по суммарному количеству назначений составляет более 51% от всех назначаемых ЛП.

Идеальная модель имеет матрицу ошибок, где по диагонали от верхнего левого до правого нижнего угла (в клетках, где значения категорий истинных и предсказанных значений совпадают) есть

значения, а вне этой диагонали отображаются нули. По соответствию значений категорий на матрице можно провести анализ предсказательной способности модели и сделать выводы о ключевых закономерностях ее работы.

Для еще более детального изучения прогностической способности полученной модели описаны сценарии назначения ЛП для сравнения истинных и предсказываемых значений. В таблице 4 представлена информация о некоторых из них. По этим данным можно выделить ряд проблем и сложностей как у модели, так и у решаемой ею задачи:

– в первом сценарии модель посчитала нужным добавить к предсказанным значениям такое действующее вещество, как сахаромикеты буларди (*Saccharomyces boulardii* – вид дрожжевых грибов, способствующих нормализации микрофлоры кишечника), дополнительно к уже назначенному пробиотическому комплексу;

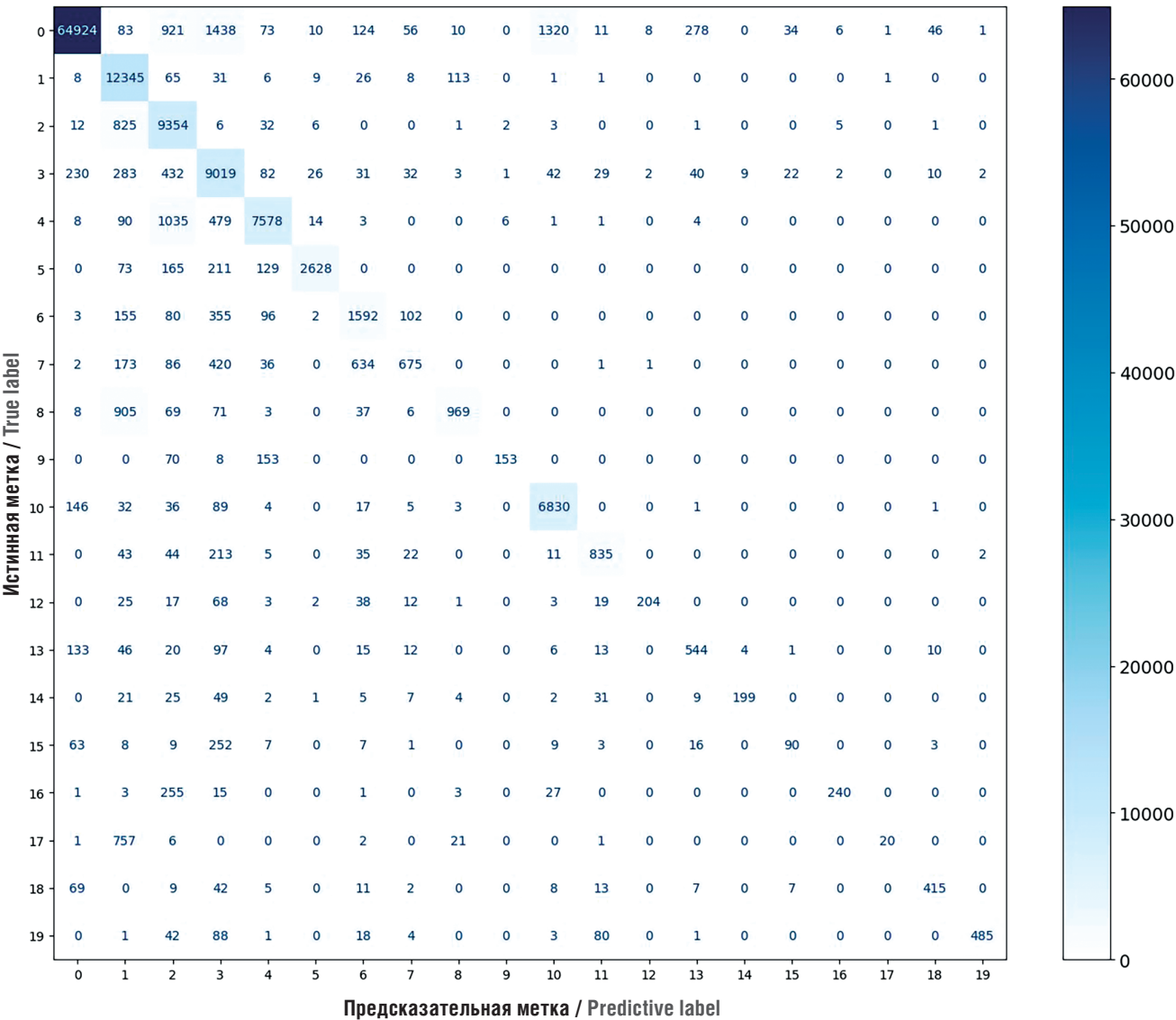


Рисунок 1. Матрица ошибок оптимизированной модели XGBC для наиболее релевантных действующих веществ лекарственных препаратов. 0 – морская вода; 1 – мометазон; 2 – цетиризин; 3 – мирамистин; 4 – ибупрофен; 5 – гомеопатический комплекс; 6 – оксиметазолин; 7 – ксилметазолин; 8 – будесонид; 9 – парацетамол; 10 – холекальциферол; 11 – неомицин, полимиксин, дексаматезон; 12 – феназон, лидокаин; 13 – амоксициллин, клавулановая кислота; 14 – ацетилцистеин, туаминогептан; 15 – лизоцим, пиридоксин; 16 – диметинден; 17 – флутиказон; 18 – растительные экстракты; 19 – тобрамицин, дексаметазон

Figure 1. Error matrix of the optimized XGBC model for the most relevant active ingredients of drugs. 0 – sea water; 1 – mometasone; 2 – cetirizine; 3 – miramistin; 4 – ibuprofen; 5 – homeopathic complex; 6 – oxymetazoline; 7 – xylometazoline; 8 – budesonide; 9 – paracetamol; 10 – cholecalciferol; 11 – neomycin, polymyxin, dexamethasone; 12 – phenazone, lidocaine; 13 – amoxicillin, clavulanic acid; 14 – acetylcysteine, tuaminoheptane; 15 – lysozyme, pyridoxine; 16 – dimethindene; 17 – fluticasone; 18 – plant extracts; 19 – tobramycin, dexamethasone

Таблица 4. Сценарии назначения лекарственных препаратов и предсказанные значения

Table 4. Drug prescription scenarios and predicted values

№ / No.	Набор меток / Set of labels	Истинные значения / True values	Предсказанные значения / Predicted values
1	Девочка, 12 лет, осень, врач педиатр (оториноларинголог), повторный прием, 3 визита, диагнозы: вирусная инфекция неуточненная, функциональное нарушение кишечника неуточненное / Girl, 12 years old, autumn, pediatrician (otolaryngologist), follow-up appointment, 3 visits, diagnoses: unspecified viral infection, unspecified functional bowel disorder	Алюминия фосфат, пробиотический комплекс / Aluminum phosphate, probiotic complex	Алюминия фосфат, пробиотический комплекс, сахаромикеты буларди / Aluminum phosphate, probiotic complex, saccharomyces boulardii
2	Мальчик, 15 лет, осень, врач педиатр, первичный прием, 15 визитов, диагноз: функциональное нарушение кишечника неуточненное / Boy, 15 years old, autumn, pediatrician, initial appointment, 15 visits, diagnosis: functional bowel disorder, unspecified	Нифуроксазид, полиметилсилоксан , пробиотический комплекс / Nifuroxazide, polymethylsiloxane , probiotic complex	Диоктаэдрический смектит , нифуроксазид, пробиотический комплекс / Diocahedral smectite , nifuroxazide, probiotic complex
3	Девочка, 12 лет, осень, врач педиатр, первичный прием, 4 визита, диагноз: вирусная инфекция неуточненная / Girl, 12 years old, autumn, pediatrician, initial appointment, 4 visits, diagnosis: unspecified viral infection	Осельтамивир, оксиметазолин / Oseltamivir, oxymetazoline	Ксилометазолин , осельтамивир / Xylometazoline , oseltamivir
4	Девочка, 7 лет, осень, врач педиатр (аллерголог-иммунолог), первичный прием, 15 визитов, диагноз: острый ларинготрахеит / Girl, 7 years old, autumn, pediatrician (allergist-immunologist), initial appointment, 15 visits, diagnosis: acute laryngotracheitis	Будесонид, ипратропия бромид, фенотерол, монтелукаст / Budesonide, ipratropium bromide, fenoterol, montelukast	Будесонид, ипратропия бромид, фенотерол, монтелукаст / Budesonide, ipratropium bromide, fenoterol, montelukast

Примечание. Полу жирным шрифтом выделены истинные и предсказанные значения, различающиеся в рамках одного сценария.

Note. True and predicted values, which differ within a scenario are highlighted in bold.

– во втором сценарии модель заменила адсорбент полиметилсилоксан на диоктаэдрический смектит – его аналог по фармакологической группе;

– в третьем сценарии действующее вещество оксиметазолин было заменено на ксилометазолин (эти вещества входят в одну фармакологическую группу и практически аналогичны по фармакологическому действию);

– четвертый сценарий – хороший пример полного совпадения истинных и предсказанных значений.

Поведение модели, наблюдаемое в приведенных сценариях, может быть обусловлено несколькими факторами. Во-первых, модель может склоняться к предсказанию дополнительных препаратов, которые считаются одинаково релевантными в контексте определенного диагноза или анамнеза пациента, что приводит к включению в прогноз лишних веществ как в первом, так и в третьем сценариях. Во-вторых, такие замены, как в сценарии с полиметилсилоксаном и диоктаэдрическим смектитом, могут свидетельствовать о недостаточном (согласно расчетам внутри обученного алгоритма модели) различии между схожими препаратами, что может быть следствием их близости по фармакологической группе. Проблему третьего сценария с действующими веществами оксиметазолин и ксилометазолин можно также наблюдать на матрице ошибок (см. рис. 1). Для работы с матрицей ошибок важно обращать внимание на количество правильных и неправильных предсказаний для каждого класса. Анализируя диагональные элементы матрицы (где совпадают истинные и предсказанные метки), можно оценить точность модели для каждого отдельного класса, а несоответствия в остальных ячейках помогут выявить классы, которые модель путает чаще всего. На полученной матрице ошибок можно заметить определенные особенности в предсказаниях модели для меток «ксилотетазолин» и «оксиметазолин»: модель имеет тенденцию путать эти два пре-

парата, что выражается в перекрестных ошибках между метками 6 и 7, – оба препарата похожи по своему фармакологическому действию, из-за чего модель не всегда способна различить их в контексте задачи, и это приводит к неточностям в предсказаниях.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Сценарное моделирование – эффективный инструмент прогнозирования различных сценариев, нашедший свое применение, в частности, в сфере здравоохранения. В ходе исследования рассмотрена задача многометочной классификации в контексте назначения ЛП.

Проведен сравнительный анализ моделей XGBC, CNN, FCNN, OvR и RFC с акцентом на их способность классифицировать ЛП по таким признакам, как действующие вещества, фармакологические группы и торговые наименования. Оценка значений метрик AUC ROC, F1-score и Custom Accuracy полученных моделей определила модель XGBC как наилучший алгоритм решения задачи многометочной классификации каждого из целевых признаков. Целевой признак «Фармакологические группы» послужил элементом оптимизации обучающей выборки для помощи моделям в определении истинных меток торговых наименований ЛП и их действующих веществ. Метод доработки признаков как один из способов «помочь» моделям в определении истинных значений ЛП дал результат и значительно повысил их предсказательную способность. Последующая оптимизация гиперпараметров позволила довести прогностическую способность и точность модели XGBC до 85%, однако изучение результатов работы модели выявило ряд проблем ошибочного предсказания похожих по фармакологическому действию ЛП. Обнаруженная проблема свидетельствует о необходимости дальнейших оптимизаций как путем доработки самого технического задания решаемой задачи, так и с помощью

введения и разработки новых признаков в тренировочной выборке. Решение проблемы дисбаланса классов моделей многометочной классификации особенно актуально при прогнозировании торговых наименований назначаемых ЛП.

Тем не менее характеристики полученной модели XGBC позволяют рассматривать ее как основного кандидата для создания прототипа системы СМ: модель способна предсказывать если не

истинные, то похожие по группе и действию вещества и ЛП, что позволяет оценить предполагаемую структуру фармацевтической помощи в конкретном сценарии. Таким образом, предложенный в данном исследовании подход имеет значительный потенциал для моделирования достоверных сценариев и использования полученных результатов в системах СМ с целью улучшения и оптимизации медицинской и фармацевтической помощи детям.

ИНФОРМАЦИЯ О СТАТЬЕ	ARTICLE INFORMATION
Поступила: 24.10.2024 В доработанном виде: 18.11.2024 Принята к печати: 29.11.2024 Опубликована онлайн: 03.12.2024	Received: 24.10.2024 Revision received: 18.11.2024 Accepted: 29.11.2024 Published online: 03.12.2024
Вклад авторов	Authors' contribution
Кондрашов А.А. – концепция и дизайн исследования, анализ данных и разработка моделей, написание текста; Курашов М.М. – сбор и обработка информационной базы исследования, анализ данных и разработка моделей, редактирование текста; Лоскутова Е.Е. – написание и редактирование текста. Все авторы прочитали и утвердили окончательный вариант рукописи	Kondrashov A.A. – study concept and design, data analysis and model development, text writing; Kurashov M.M. – collection and processing of study information base, data analysis and model development, text editing; Loskutova E.E. – text writing and editing. All authors have read and approved the final version of the manuscript
Конфликт интересов	Conflict of interests
Авторы заявляют об отсутствии конфликта интересов	The authors declare no conflict of interests
Финансирование	Funding
Авторы заявляют об отсутствии финансовой поддержки	The authors declare no funding
Этические аспекты	Ethics declarations
Сведения о пациентах не содержат информацию, относящуюся к персональным данным: числовой идентификационный номер пациента – единственный способ определить принадлежность назначения лекарственного препарата. Исследование проводилось в соответствии с принципами Хельсинкской декларации Всемирной медицинской ассоциации	The patient information did not contain information related to personal data: the numerical identification number of the patient was the only way to determine the affiliation of drug prescription. The study was conducted in accordance with the principles of the Helsinki Declaration of the World Medical Association
Раскрытие данных	Data sharing
Первичные данные могут быть предоставлены по обоснованному запросу автору, отвечающему за корреспонденцию	Raw data could be provided upon reasonable request to the corresponding author
Комментарий издателя	Publisher's note
Содержащиеся в этой публикации утверждения, мнения и данные были созданы ее авторами, а не издательством ИРБИС (ООО «ИРБИС»). Издательство ИРБИС снимает с себя ответственность за любой ущерб, нанесенный людям или имуществу в результате использования любых идей, методов, инструкций или препаратов, упомянутых в публикации	The statements, opinions, and data contained in this publication were generated by the authors and not by IRBIS Publishing (IRBIS LLC). IRBIS Publishing disclaims any responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred in the content
Права и полномочия	Rights and permissions
ООО «ИРБИС» обладает исключительными правами на эту статью по Договору с автором (авторами) или другим правообладателем (правообладателями). Использование статьи регулируется исключительно условиями Договора и действующим законодательством	IRBIS LLC holds exclusive rights to this paper under a publishing agreement with the author(s) or other rightsholder(s). Usage of this paper is solely governed by the terms of the publishing agreement and applicable law

ЛИТЕРАТУРА

1. Романов И.А. Машинное обучение как конкурентное преимущество предприятия. *Московский экономический журнал*. 2022; 7 (3): 42. https://doi.org/10.55186/2413046X_2022_7_3_141.
2. Ксенофонтов Д.М. Сценарное моделирование эпидемиологических эффектов экономической политики. *Научные труды: Институт народнохозяйственного прогнозирования РАН*. 2020; 18: 542–65. <https://doi.org/10.47711/2076-318-2020-542-565>.
3. Цацулин А.Н., Цацулин Б.А. Сценарный подход к построению прогнозных моделей развития региональных систем здравоохранения. *Научно-технические ведомости СПбГПУ. Экономические науки*. 2021; 14 (2): 115–36 (на англ. яз.). <https://doi.org/10.18721/JE.14208>.
4. Аксенова Е.С., Евдокимов Д.С., Катасонова К.А. Усовершенствованная агент-ориентированная модель с функционалом сценарного моделирования и свойствами цифрового двойника для прогно-

зирования социо-эпидемиолого-экономических процессов в регионах России. *Искусственные общества*. 2023; 18 (4). <https://doi.org/10.18254/S207751800028782-9>.
5. Комков А.А., Мазяев В.П., Рязанова С.В. и др. Применение программы интеллектуальной аналитики текста с бумажного носителя и сегментации по заданным параметрам в клинической практике. *Кардиоваскулярная терапия и профилактика*. 2023; 21 (12): 3458. <https://doi.org/10.15829/1728-8800-2022-3458>.
6. Гусев А.В., Новицкий Р.Э., Ившин А.А., Алексеев А.А. Машинное обучение на лабораторных данных для прогнозирования заболеваний. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология*. 2021; 14 (4): 581–92. <https://doi.org/10.17749/2070-4909/farmakoeconomika.2021.115>.
7. Коледачкин А.А. Использование моделирования и симуляций

в тестировании: перспективы с применением ИИ. *Вестник науки*. 2024; 5 (9): 513–40.

8. Orji U., Ukwandu E. Machine learning for an explainable cost prediction of medical insurance. *Machine Learn App*. 2024; 15: 100516. <https://doi.org/10.1016/j.mlwa.2023.100516>.

9. Наркевич А.Н., Виноградов К.А., Параскевопуло К.М., Гржибовский А.М. Интеллектуальные методы анализа данных в биомедицинских исследованиях: нейронные сети. *Экология человека*. 2021; 28 (4): 55–64. <https://doi.org/10.33396/1728-0869-2021-4-55-64>.

10. Голоунина О.О., Белая Ж.Е., Воронов К.А. и др. Применение методов машинного обучения в дифференциальной диагностике

АКТГ-зависимого эндогенного гиперкортицизма. *Проблемы эндокринологии*. 2024; 70 (1): 18–29. <https://doi.org/10.14341/probl13342>.

11. Фирюлина М.А., Каширина И.Л., Гафанович Е.Я. Применение методов машинного обучения при назначении терапии гипертонической болезни. *Моделирование, оптимизация и информационные технологии*. 2020; 8 (4): 4. <https://doi.org/10.26102/2310-6018/2020.31.4.025>.

12. Silva P., Rivolli A., Rocha P., et al. Machine learning for drugs prescription. In: Yin H., Camacho D., Novais P., Tallón-Ballesteros A. (Eds.) *Intelligent Data Engineering and Automated Learning – IDEAL 2018*. Part I. Springer; 2018: 548–55. https://doi.org/10.1007/978-3-030-03493-1_57.

REFERENCES

1. Romanov I.A. Machine learning as a competitive advantage of the company. *Moscow Economic Journal*. 2022; 7 (3): 42 (in Russ.). https://doi.org/10.55186/2413046X_2022_7_3_141.

2. Ksenofontov D.M. Scenario modeling of the economic policy epidemiological effects. *Scientific Papers: Institute of National Economic Forecasting of RAS*. 2020; 18: 542–65 (in Russ.). <https://doi.org/10.47711/2076-318-2020-542-565>.

3. Tsatsulin A.N., Tsatsulin B.A. Scenario approach to building predictive models for the development of regional health systems. *St. Petersburg State Polytechnical University Journal. Economics*. 2021; 14 (2): 115–36. <https://doi.org/10.18721/JE.14208>.

4. Aksenova E.S., Evdokimov D.S., Katasonova K.A. An improved agent-based model with scenario modeling functional and digital twin properties for forecasting socioepidemiological-economic processes in the regions of Russia. *Artificial Societies*. 2023; 18 (4) (in Russ.). <https://doi.org/10.18254/S207751800028782-9>.

5. Komkov A.A., Mazaev V.P., Ryazanova S.V., et al. Application of the program for artificial intelligence analytics of paper text and segmentation by specified parameters in clinical practice. *Cardiovascular Therapy and Prevention*. 2023; 21 (12): 3458 (in Russ.). <https://doi.org/10.15829/1728-8800-2022-3458>.

6. Gusev A.V., Novitskiy R.E., Ivshin A.A., Alekseev A.A. Machine learning based on laboratory data for disease prediction. *FARMAKOEKONOMIKA. Sovremennaya farmakoeconomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and*

Pharmacoepidemiology. 2021; 14 (4): 581–92 (in Russ.). <https://doi.org/10.17749/2070-4909/farmakoeconomika.2021.115>.

7. Koledachkin A.A. Using simulation and simulation in testing: prospects with use of AI. *Vesnyk nauki / Bulletin of Science*. 2024; 5 (9): 513–40 (in Russ.).

8. Orji U., Ukwandu E. Machine learning for an explainable cost prediction of medical insurance. *Machine Learn App*. 2024; 15: 100516. <https://doi.org/10.1016/j.mlwa.2023.100516>.

9. Narkevich A.N., Vinogradov K.A., Paraskevopulo K.M., Grijbovski A.M. Intelligent data analysis in biomedical research: artificial neural networks. *Ekologiya cheloveka / Human Ecology*. 2021; 28 (4): 55–64 (in Russ.). <https://doi.org/10.33396/1728-0869-2021-4-55-64>.

10. Golounina O.O., Belaya Zh.E., Voronov K.A., et al. Machine learning methods in differential diagnosis of ACTH-dependent hypercortisolism. *Problems of Endocrinology*. 2024; 70 (1): 18–29 (in Russ.). <https://doi.org/10.14341/probl13342>.

11. Firyulina M.A., Kashirina I.L., Gafanovich E.Y. Using of machine learning methods in prescribing hypertension therapy. *Modeling, Optimization and Information Technology*. 2020; 8 (4): 4 (in Russ.). <https://doi.org/10.26102/2310-6018/2020.31.4.025>.

12. Silva P., Rivolli A., Rocha P., et al. Machine learning for drugs prescription. In: Yin H., Camacho D., Novais P., Tallón-Ballesteros A. (Eds.) *Intelligent Data Engineering and Automated Learning – IDEAL 2018*. Part I. Springer; 2018: 548–55. https://doi.org/10.1007/978-3-030-03493-1_57.

Сведения об авторах / About the authors

Кондрашов Александр Андреевич / Alexander A. Kondrashov – ORCID: <https://orcid.org/0009-0004-0601-3205>. eLibrary SPIN-code: 1222-7345.

Курашов Максим Михайлович, к.фарм.н., доцент / Maxim M. Kurashov, PhD, Assoc. Prof. – ORCID: <https://orcid.org/0000-0002-0349-905X>. WoS ResearcherID: ISS-9102-2023. Scopus Author ID: 57209803706. eLibrary SPIN-code: 8741-3469.

Лоскутова Екатерина Ефимовна, д.фарм.н., проф. / Ekaterina E. Loskutova, Dr. Sci. Pharm., Prof. – ORCID: <https://orcid.org/0000-0002-1514-0941>. eLibrary SPIN-code: 1133-0394.