

ISSN 1561-8323 (Print)
ISSN 2524-2431 (Online)

ИНФОРМАТИКА

INFORMATICS

УДК 004.9, 004.85, 004.89, 004.942, 579.6, 616-002.5, 615.281.9
<https://doi.org/10.29235/1561-8323-2025-695-5-367-375>

Поступило в редакцию 06.08.2025
Received 06.08.2025

Г. М. Башко¹, Ю. В. Корноушенко²,
член-корреспондент А. В. Тузиков³, А. М. Андрианов²

¹Белорусский государственный университет, Минск, Республика Беларусь

²Институт биоорганической химии Национальной академии наук Беларуси, Минск, Республика Беларусь

³Объединенный институт проблем информатики Национальной академии наук Беларуси,
Минск, Республика Беларусь

ПРОГНОСТИЧЕСКАЯ МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВИРТУАЛЬНОГО СКРИНИНГА ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ *MYCOBACTERIUM TUBERCULOSIS*

Аннотация. Разработана ансамблевая модель машинного обучения на основе методологии бустинга, использующая молекулярные дескрипторы в качестве входных данных для предсказания значений свободной энергии связывания малых молекул с большим микобактериальным мембранным белком 3 (MmpL3) *Mycobacterium tuberculosis* – транспортером миколовых кислот и липидов, критически важным для роста и жизнеспособности клеток. В результате тестирования этой модели на двух наборах структурно разнородных молекул с использованием метрик регрессионной оценки MAE, MSE, R^2 и R показано, что она сопоставима по предсказательной эффективности с оценочной функцией Vina программы молекулярного докинга QuickVina 2. При этом разработанная модель позволяет значительно ускорить процесс виртуального скрининга потенциальных лекарств, что является важным фактором при анализе молекулярных библиотек, включающих сотни тысяч и даже миллионы химических структур. В связи с этим предлагаемая модель может быть использована в качестве экспресс-метода для быстрого отбора в химических базах данных перспективных соединений с последующим предсказанием их положений в сайте связывания MmpL3 с помощью молекулярного докинга и исследованием стабильности комплексов лиганд/MmpL3 методами молекулярной динамики. Полученные результаты свидетельствуют о высокой эффективности разработанной модели и ее значительном потенциале для использования в виртуальном скрининге соединений-кандидатов, антибактериальное действие которых основано на ингибировании белка MmpL3 *Mycobacterium tuberculosis* – одной из приоритетных терапевтических мишеней для создания новых эффективных препаратов против лекарственно-устойчивого туберкулеза.

Ключевые слова: прогностические модели, машинное обучение, бустинг, молекулярный докинг, виртуальный скрининг, *Mycobacterium tuberculosis*, MmpL3, противотуберкулезные препараты

Для цитирования. Прогностическая модель машинного обучения для виртуального скрининга потенциальных ингибиторов *Mycobacterium tuberculosis* / Г. М. Башко, Ю. В. Корноушенко, А. В. Тузиков, А. М. Андрианов // Доклады Национальной академии наук Беларуси. – 2025. – Т. 69, № 5. – С. 367–375. <https://doi.org/10.29235/1561-8323-2025-69-5-367-375>

Georgy M. Bashko¹, Yuri V. Kornoushenko²,
Corresponding Member Alexander V. Tuzikov³, Alexander M. Andrianov²

¹Belarusian State University, Minsk, Republic of Belarus

²Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

³United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

A PREDICTIVE MACHINE LEARNING MODEL FOR VIRTUAL SCREENING OF POTENTIAL INHIBITORS AGAINST *MYCOBACTERIUM TUBERCULOSIS*

Abstract. A boosting-based ensemble machine learning model that utilizes molecular descriptors as input data has been developed to predict the values of binding free energy of small-molecule compounds to the MmpL3 of *Mycobacterium tuberculosis* (Mtb), an essential mycolic acid and lipid transporter required for growth and cell viability. Testing this model on two large sets of structurally heterogeneous molecules via regression evaluation metrics MAE, MSE, R^2 , and R showed that

it is comparable in predictive performance to the Vina scoring function of the QuickVina 2 molecular docking program, but allows one to significantly speed up the structure-based virtual screening, which is an important factor in the analysis of molecular libraries containing hundreds of thousands, and occasionally millions, of chemical structures. In this regard, the developed model can be used as an express method for the rapid selection of candidate compounds in chemical databases, followed by prediction of their poses in the MmpL3 binding site using molecular docking and a study of the stability of ligand/MmpL3 complexes via molecular dynamics methods. The results obtained demonstrate the high efficiency of the developed model and its significant potential for the use in virtual screening of candidate compounds with antibacterial action based on the inhibition of the MmpL3 protein of *Mycobacterium tuberculosis*, one of the priority therapeutic targets for the design of new effective therapeutics against drug-resistant tuberculosis.

Keywords: predictive models, machine learning, boosting, molecular docking, virtual screening, *Mycobacterium tuberculosis*, MmpL3, anti-tuberculosis drugs

For citation. Bashko G. M., Kornoushenko Yu. V., Tuzikov A. V., Andrianov A. M. A predictive machine learning model for virtual screening of potential inhibitors against *Mycobacterium tuberculosis*. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2025, vol. 69, no. 5, pp. 367–375. <https://doi.org/10.29235/1561-8323-2025-69-5-367-375>

Введение. Создание новых лекарственных препаратов представляет собой сложный и трудоемкий процесс, включающий идентификацию низкомолекулярных соединений с высокой активностью, высокой селективностью, приемлемыми фармакокинетическими свойствами и низкой токсичностью с последующим их синтезом, характеристикой, скринингом и анализом на терапевтическую эффективность [1; 2]. Однако в настоящее время этот процесс претерпевает значительную трансформацию, что обусловлено быстрым развитием компьютерных технологий, применяемых не только для ускорения исследований, но и для оценки рисков и затрат в клинических испытаниях [3; 4]. При этом ключевую роль в данном революционном преобразовании играют модели машинного обучения, позволяющие значительно сократить время и финансовые затраты, необходимые для разработки новых терапевтических средств [5]. В последние годы эти модели широко используются в фармацевтических исследованиях для прогнозирования потенциальных терапевтических мишеней, перепрофилирования лекарств, разработки новых лекарств и прогнозирования активности, фармакокинетических и токсикологических свойств этих молекул на основе анализа их химических структур [5]. Модели машинного обучения могут предсказывать потенциальные взаимодействия между различными лекарствами, что имеет решающее значение для предотвращения побочных эффектов при одновременном приеме нескольких препаратов (<https://www.mrlcg.com/resources/blog/machine-learning-for-drug-discovery/>). Кроме того, применение этих моделей позволяет выявлять потенциальные биомаркеры, которые можно использовать для диагностики заболеваний, прогнозирования ответов на проведенную терапию и мониторинга результатов лечения пациентов, а также проводить анализ генетических данных для персонализации вариантов лечения с учетом индивидуальных характеристик пациента (<https://www.mrlcg.com/resources/blog/machine-learning-for-drug-discovery/>). Использование моделей машинного обучения в виртуальном скрининге больших химических библиотек показало их высокую эффективность для выявления потенциальных кандидатов в лекарственные средства, что позволяет значительно сократить количество соединений, которые необходимо синтезировать и тестировать на моделях *in vitro* и *in vivo* (<https://www.mrlcg.com/resources/blog/machine-learning-for-drug-discovery/>).

Туберкулез (ТБ), возбудителем которого являются микобактерии *Mycobacterium tuberculosis* (МБТ), входит в десятку основных причин смертности во всем мире, ежегодно унося около 1,3 миллиона жизней (<https://deathmeters.info>). Широкое использование противотуберкулезных препаратов, таких как рифампицин, изониазид и пиразинамид, снизило риск развития активного ТБ. Однако появление в последние годы штаммов МБТ с множественной и широкой лекарственной устойчивостью сделало существующие антибиотики неэффективными и создало новые угрозы для борьбы с этим тяжелым инфекционным заболеванием [6]. Недавно одобренные для применения в клинике антибактериальные препараты бедахинолин, претоманид, деламаид и линезолид продемонстрировали хороший потенциал для повышения эффективности лечения лекарственно-устойчивого ТБ (<https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024>). К сожалению, менее 20 % пациентов могут получать пользу от лечения этими новыми препаратами из-за их высокой стоимости, низкой доступности, недостаточной проверки эффективности и уже наблюдаемой в ряде случаев

клинической резистентности. Таким образом, по-прежнему существует острая необходимость в разработке новых противотуберкулезных агентов или более быстром и эффективном перепрофилировании уже одобренных для лечения ТБ препаратов.

Одним из перспективных подходов к решению этой проблемы является использование алгоритмов машинного обучения и молекулярного моделирования для виртуального скрининга баз данных химических соединений, с помощью которых уже был идентифицирован ряд ингибиторов МБТ, находящихся в настоящее время на различных стадиях медицинских испытаний [7–9]. Однако в настоящее время исследования туберкулеза, проводимые с привлечением технологий машинного обучения, в основном охватывают диагностику и результаты лечения этой болезни. Это является одним из пробелов, который необходимо устранить, чтобы иметь возможность ускорить процесс создания новых и эффективных препаратов для терапии устойчивых форм туберкулеза и облегчить тяжелое бремя этой инфекции. В связи с этим разработка и применение моделей машинного обучения для поиска потенциальных кандидатов в терапевтические агенты, в том числе новых эффективных ингибиторов лекарственно-устойчивых штаммов МБТ, имеют большое научное и практическое значение.

В связи со стремительным распространением резистентных штаммов МБТ в последние годы большое внимание исследователей сфокусировано на эффлюксных помпах – больших микобактериальных мембранных белках (MmpL), которые способны транспортировать молекулы субстрата различных размеров из внутриклеточного пространства во внешнюю среду, обеспечивая один из основных механизмов формирования лекарственной устойчивости [10]. Среди эффлюксных помп МБТ следует особо выделить белок MmpL3, который отвечает за транспорт липидов, в основном миколовых кислот, необходимых для построения уникальной клеточной стенки микобактерий, и играет важную роль в патогенезе ТБ [10]. В связи с этим этот консервативный белок рассматривается в качестве одной из приоритетных мишеней для разработки новых эффективных ингибиторов лекарственно-устойчивых штаммов МБТ [10].

В настоящем исследовании разработана прогностическая модель машинного обучения, позволяющая предсказывать значения свободной энергии связывания малых молекул с белком MmpL3 и предназначенная для экспресс-анализа больших химических баз данных, направленного на идентификацию потенциальных ингибиторов этой важной для выживания МБТ молекулярной мишени.

Материалы и методы исследования. На начальном этапе исследований для обучения и тестирования моделей были сформированы два набора данных, соответствующих двум виртуальным библиотекам низкомолекулярных химических соединений. Первая библиотека (библиотека 1) содержала малые молекулы, структуры которых включали основные фармакофоры ингибиторов белка MmpL3 МБТ SQ109 [11] и AU1235 [12], описывающие совокупность структурно-функциональных свойств этих противотуберкулезных агентов, обеспечивающих эффективность их связывания с терапевтической мишенью. Для идентификации таких молекул осуществляли фармакофорный анализ химических баз данных веб-сервера Pharmit (<http://pharmit.csb.pitt.edu/>), позволяющего проводить интерактивное исследование химического пространства с целью поиска потенциальных лекарств на основе их структурно-функционального сходства с высокоаффинными лигандами белка-мишени. Фармакофорные модели ингибиторов SQ109 и AU1235 формировали с использованием программных средств Pharmit на основе анализа их кристаллических комплексов с белком MmpL3 *Mycobacterium smegmatis* (PDB ID 6AJG и 6AJH; <https://www.rcsb.org>), являющимся надежной моделью для изучения и разработки новых ингибиторов MmpL3 МБТ [13]. В процессе фармакофорного скрининга использовали ряд фильтров, накладывающих ограничения на физико-химические параметры молекул, связанные с «правилом пяти» Липинского [14], которые принято считать основными критериями эффективности потенциального лекарства при пероральном приеме. На заключительном этапе с использованием Python и программной библиотеки RDKit (<https://www.rdkit.org>) из полученного набора соединений удаляли невалидные молекулы и дубликаты. В результате библиотека 1 включала 46163 соединения, которые были разделены на обучающую и тестовую выборки в соотношении 80 % к 20 % от общего числа молекул соответственно. Библиотека 2 включала 26 000 биологически активных

молекул, в том числе одобренных для применения в клинике лекарств, экспериментальных препаратов и находящихся на различных стадиях испытаний соединений-кандидатов, собранных из баз данных DrugBank (<https://www.drugbank.com/datasets>), ZINC 15 [15] и Selleck Chemicals (<https://www.selleckchem.com>). Соединения из этой библиотеки были использованы для дополнительного тестирования разработанной прогностической модели машинного обучения.

После формирования библиотек 1 и 2 каждому соединению, включенному в их состав, было поставлено в соответствие 270 дескрипторов, рассчитанных в программных пакетах RDKit (85 дескрипторов) (<https://www.rdkit.org>) и Mordred (185 дескрипторов) (<https://github.com/mordred-descriptor/mordred>) и описывающих численными параметрами его молекулярную структуру. Кроме того, всем соединениям присваивали соответствующие значения свободной энергии связывания с белком MmpL3, предсказанные методами молекулярного докинга.

Молекулярный докинг выполняли с использованием кристаллической структуры белка MmpL3 *Mycobacterium smegmatis* в комплексе с ингибитором МБТ SQ109 (PDB ID: 6AJG; <http://www.rcsb.org/pdb/>). Подготовку белка и лигандов к расчетам осуществляли с помощью программного пакета MGLTools (<https://ccsb.scripps.edu/mgltools/>). Докинг соединений из библиотек 1 и 2 с MmpL3 проводили в программе QuickVina 2 (<https://github.com/QVina/qvina>) в приближении жесткого рецептора и гибких лигандов. Ячейка для докинга охватывала область белка MmpL3, включающую сайт связывания ингибитора SQ109 (PDB ID: 6AJG; <http://www.rcsb.org/pdb/>) и имела следующие параметры: $\Delta X = 20,358 \text{ \AA}$, $\Delta Y = 30,16 \text{ \AA}$, $\Delta Z = 28,652 \text{ \AA}$ с центром $X = 36,163 \text{ \AA}$, $Y = -0,678 \text{ \AA}$, $Z = -18,582 \text{ \AA}$. Параметр, характеризующий полноту поиска (охват конформационного пространства), был задан равным 100 (<https://github.com/QVina/qvina>).

С целью выбора наилучшего решения задачи для обучения моделей использовали пять алгоритмов: LightGBM – классификатор повышения градиента, использующий древовидные алгоритмы обучения (<https://lightgbm.readthedocs.io/en/stable/>) (модель 1); MLPRegressor – многослойная нейронная сеть по типу перцептрона (https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html) (модель 2); TabNetRegressor – высокопроизводительная каноническая архитектура глубокого обучения на основе табличных данных (<https://pytorch-tabnet/>) (модель 3) и ExtraTreesRegressor – вариация случайного леса с дополнительной рандомизацией в процессе построения деревьев (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html>) (модель 4), а также ансамбль из трех базовых моделей LGBMRegressor, XGBRegressor и CatBoostRegressor в схеме голосования VotingRegressor (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html>) (модель 5).

Обучение и тестирование моделей проводили с помощью вычислительных ресурсов сервиса Google Colab – интерактивной облачной среды для работы с кодом на языке Python (<https://colab.google>).

Для оценки производительности моделей использовали следующие метрики качества – среднюю абсолютную ошибку (MAE, Mean Absolute Error), среднеквадратичную ошибку (MSE, Mean Squared Error), коэффициент детерминации R^2 и коэффициент корреляции Пирсона R , которые рассчитывали по следующим формулам:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |p_i - y_i|,$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2,$$

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}},$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}},$$

где y_i – значения свободной энергии связывания, рассчитанные методом молекулярного докинга; p_i – значения свободной энергии связывания, предсказанные моделью машинного обучения; N – число значений; $\bar{x}$ и $\bar{y}$ – средние значения; RSS – сумма квадратов разницы между фактическими и предсказанными линейной регрессией значениями зависимой переменной; TSS – общая дисперсия.

Результаты и их обсуждение. В таблице приведены данные о предсказательной эффективности моделей 1–5, полученные в результате их апробации на тестовой выборке набора данных библиотеки 1 и представленные в величинах метрик MAE, MSE, R^2 и R . Из анализа значений этих метрик следует, что наибольшую эффективность демонстрирует ансамблевая модель 5, обученная с помощью алгоритмов градиентного бустинга LGBM, XGBoost и CatBoost в схеме голосования VotingRegressor. Эта модель показывает хорошее соответствие между предсказанными и целевыми значениями свободной энергии образования комплексов лиганд/MmpL3 (рис. 1), о чем свидетельствуют величины метрик MAE, MSE, R^2 и R , равные соответственно 0,235, 0,118 ккал/моль, 0,923 и 0,961 (таблица).

Величины метрик MAE, MSE, R^2 и R , полученные для моделей 1–5 в результате их апробации на соединениях из тестовой выборки набора данных библиотеки 1

The values of the MAE, MSE, R^2 , and R metrics obtained for models 1–5 as a result of their testing on the compounds from the test sample of library 1 data set

Метрика Metric	Модель 1 Model 1 (LGBM)	Модель 2 Model 2 (MLPRegressor)	Модель 3 Model 3 (TabNetRegressor)	Модель 4 Model 4 (eTRegressor)	Модель 5 Model 5 (LGBM, XGBR, CatBoost)
MAE	0,792	0,884	0,761	0,415	0,235
MSE	1,520	0,936	0,925	0,528	0,118
R^2	0,522	0,256	0,204	0,654	0,923
R	0,728	0,517	0,628	0,810	0,961

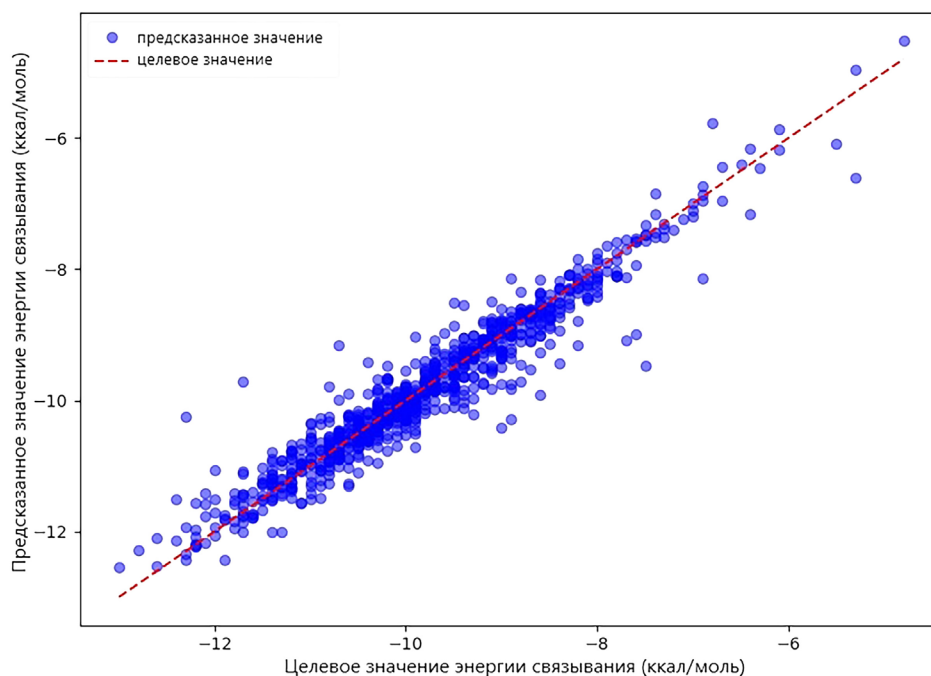


Рис. 1. График взаимосвязи между предсказанными (модель 5) и целевыми (оценочная функция Vina) значениями энергии связывания белка MmpL3 МБТ с соединениями из тестовой выборки набора данных библиотеки 1

Fig. 1. Plot of the relationship between the predicted (model 5) and target (Vina scoring function) values of binding free energies of the Mtb MmpL3 to the compounds from the test sample of library 1 data set

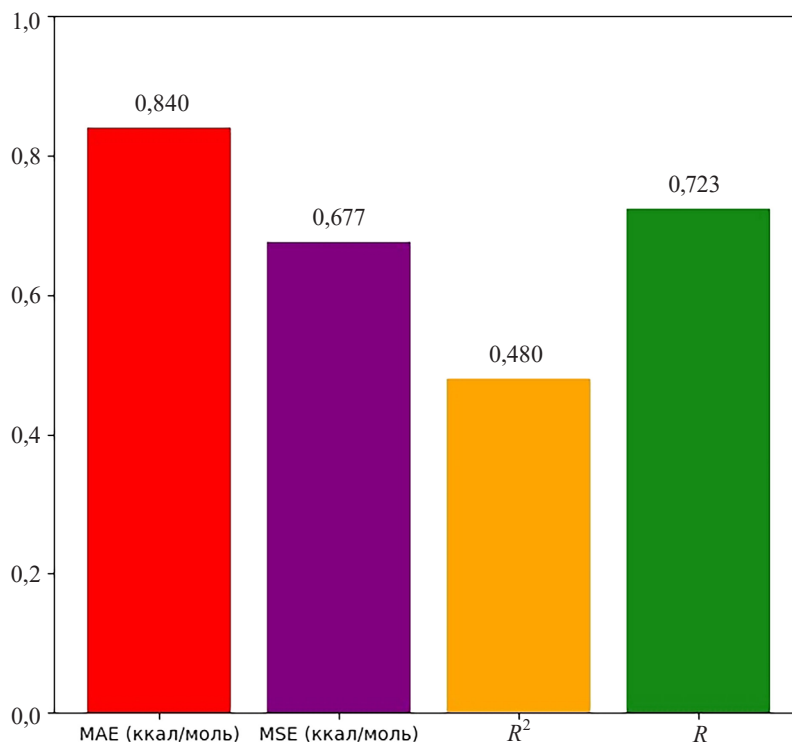


Рис. 2. Результаты тестирования модели 5 на наборе данных библиотеки 2, представленные в величинах метрик MAE, MSE, R^2 и R

Fig. 2. The results of testing model 5 on the library 2 data set, presented in terms of the MAE, MSE, R^2 , and R metrics

Как было отмечено выше, при формировании наборов данных библиотек 1 и 2 молекулярным дескрипторам каждого соединения присваивалось соответствующее значение энергии связывания с белком MmpL3, рассчитанное с помощью классической оценочной функции Vina программы QuickVina 2 (<https://github.com/QVina/qvina>) – ускоренной версии AutoDock Vina, широко используемой для виртуального скрининга потенциальных лекарств. Однако при анализе данных таблицы следует учитывать, что оценочные функции молекулярного докинга, в том числе программы QuickVina 2, основаны на различных приближениях, вносящих определенные, а в отдельных случаях и весьма значительные, погрешности в рассчитанные значения энергии межмолекулярных взаимодействий. Тем не менее, значения MAE, MSE, R^2 и R , полученные для модели 5 (таблица), позволяют сделать вывод о том, что достигнутая с ее помощью точность предсказания энергии связывания для комплексов лиганд/MmpL3 как минимум сопоставима с точностью, которую могут обеспечить программы докинга, использующие оценочную функцию Vina. Этот вывод подтверждают результаты дополнительного тестирования модели 5 на наборе данных библиотеки 2, содержащем молекулярные дескрипторы 26 000 биоактивных молекул из баз данных DrugBank, ZINC15 и Selleck Chemicals. Величины метрик MAE (0,677 ккал/моль), MSE (0,840 ккал/моль), R^2 (0,48) и R (0,723), вычисленные для этой ансамблевой модели (рис. 2), свидетельствуют о высокой корреляции между предсказанными и целевыми значениями энергии межмолекулярных взаимодействий.

Анализируя результаты тестирования, важно отметить, что модель 5 была обучена на выборке соединений, содержащих основные фармакофоры ингибиторов МБТ SQ109 и AU1235, а ее тестирование проводилось на двух выборках, одна из которых была представлена соединениями с такими же фармакофорами, а другая – структурно разнородными соединениями различных классов. Однако в обоих случаях эта модель продемонстрировала низкие значения средней абсолютной ошибки и среднеквадратичной ошибки и высокие величины коэффициента корреляции Пирсона (таблица и рис. 2). Это указывает на ее способность предсказывать

с хорошей точностью свободную энергию связывания белка MmpL3 не только с соединениями, в структурах которых присутствуют фармакофоры SQ109 или AU1235, но и с биоактивными молекулами, принадлежащими различным фармакологическим группам. При этом разработанная модель позволяет значительно ускорить процесс виртуального скрининга потенциальных соединений-кандидатов по сравнению с методами молекулярного докинга, что является важным фактором при анализе больших молекулярных библиотек, включающих сотни тысяч и даже миллионы химических структур. В частности, для проведения молекулярного докинга белка MmpL3 с 26 000 соединений из библиотеки 1 потребовалось около 110 ч непрерывной работы суперкомпьютера «СКИФ-ГЕО-ЦОД» ОИПИ НАН Беларуси, тогда как расчеты молекулярных дескрипторов этих соединений вместе с предсказаниями значений свободной энергии образования комплексов лиганд/MmpL3 заняли менее 1 ч.

Таким образом, результаты тестирования модели 5 на двух тестовых выборках свидетельствуют о ее высокой предсказательной эффективности и значительном потенциале для использования в виртуальном скрининге соединений-кандидатов, антибактериальное действие которых основано на ингибировании белка-транспортера MmpL3 МБТ – одной из приоритетных терапевтических мишеней для создания новых эффективных препаратов против лекарственно-устойчивого туберкулеза.

Заключение. В настоящей работе рассмотрены пять моделей машинного обучения, использующих в качестве входных данных дескрипторы химических соединений для предсказания свободной энергии связывания этих молекул с мембранным белком MmpL3 МБТ. В результате их тестирования на двух наборах структурно разнородных молекул с использованием метрик регрессионной оценки MAE, MSE, R^2 и R отобрана модель 5, объединившая алгоритмы градиентного бустинга LGBM, XGBoost и CatBoost. Установлено, что точность предсказания энергии межмолекулярных взаимодействий с помощью этой модели близка к точности, обеспечиваемой оценочной функцией Vina популярных программ молекулярного докинга семейства Vina, таких как AutoDock Vina, QuickVina, Vina-GPU и некоторых других. Разработанная модель может быть использована в качестве экспресс-метода для быстрого отбора в химических базах данных перспективных соединений с последующим предсказанием методами молекулярного докинга их ориентаций в сайте связывания белка MmpL3 и исследованием устойчивости комплексов лиганд/MmpL3 методами молекулярной динамики.

Благодарности. Работа выполнена при поддержке грантов БРФФИ (проект Ф24-КИТГ-016), Международного научно-технического центра (МНТЦ, проект PR150) и Консорциума и Портала программы «Лекарственно-устойчивый туберкулез» (<https://tbportals.niaid.nih.gov>).

Acknowledgments. This work was supported by grants from the BRFFR (project Ф24-КИТГ-016), the International Scientific and Technical Center (ISTC, project PR150) and the Consortium and the Drug Resistant Tuberculosis Portal Program (<https://tbportals.niaid.nih.gov>).

Список использованных источников

1. The stages of drug discovery and development process / A. B. Deore, J. R. Dhumane, R. Wagh, R. Sonawane // Asian Journal of Pharmaceutical Research and Development. – 2019. – Vol. 7, N 6. – P. 62–67. <https://doi.org/10.22270/ajprd.v7i6.616>
2. Berdigaliyev, N. An overview of drug discovery and development / N. Berdigaliyev, M. Aljofan // Future Medicinal Chemistry. – 2020. – Vol. 12, N 10. – P. 939–947. <https://doi.org/10.4155/fmc-2019-0307>
3. Roney, M. The importance of in-silico studies in drug discovery / M. Roney, M. F. F. Mohd Aluwi // Intelligent Pharmacy. – 2024. – Vol. 2, N 4. – P. 578–579. <https://doi.org/10.1016/j.ipha.2024.01.010>
4. *In silico* methods and tools for drug discovery / B. Shaker, S. Ahmad, J. Lee [et al.] // Computers in Biology and Medicine. – 2021. – Vol. 137. – Art. 104851. <https://doi.org/10.1016/j.compbiomed.2021.104851>
5. Advances in *de novo* drug design: from conventional to machine learning methods / V. D. Mouchlis, A. Afantitis, A. Serra [et al.] // International Journal of Molecular Sciences. – 2021. – Vol. 22, N 4. – Art. 1676. <https://doi.org/10.3390/ijms22041676>
6. Khawbung, J. L. Drug resistant tuberculosis: a review / J. L. Khawbung, D. Nath, S. Chakraborty // Comparative Immunology, Microbiology and Infectious Diseases. – 2021. – Vol. 74. – Art. 101574. <https://doi.org/10.1016/j.cimid.2020.101574>
7. A deep learning approach to antibiotic discovery / J. M. Stokes, K. Yang, K. Swanson [et al.] // Cell. – 2020. – Vol. 180, N 4. – P. 688–702. <https://doi.org/10.1016/j.cell.2020.01.021>
8. Identification of new *Mycobacterium tuberculosis* proteasome inhibitors using a knowledge-based computational screening approach / T. M. Almeleebia, M. A. Shahrani, M. Y. Alshahrani [et al.] // Molecules. – 2021. – Vol. 26, N 8. – Art. 2326. <https://doi.org/10.3390/molecules26082326>

9. Machine learning-enabled virtual screening indicates the anti-tuberculosis activity of aldoxorubicin and quarfloxin with verification by molecular docking, molecular dynamics simulations, and biological evaluations / S. Zheng, Ya. Gu, Yu. Gu [et al.] // *Briefings in Bioinformatics*. – 2024. – Vol. 26, N 1. – Art. bbae696. <https://doi.org/10.1093/bib/bbae696>
10. Перспективы и препятствия для клинического применения ингибиторов эффлюксных помп *Mycobacterium tuberculosis* / И. Г. Фелькер, Е. И. Гордеева, Н. В. Ставицкая [и др.] // *Биологические мембраны*. – 2021. – Т. 38, № 5. – С. 317–339.
11. MmpL3 inhibition as a promising approach to develop novel therapies against tuberculosis: a spotlight on SQ109, clinical studies, and patents literature / M. Imran, M. K. Arora, A. Chaudhary [et al.] // *Biomedicines*. – 2022. – Vol. 10, N 11. – Art. 2793. <https://doi.org/10.3390/biomedicines10112793>
12. Specifically targeting Mtb cell-wall and TMM transporter: the development of MmpL3 inhibitors / Q. Luo, H. Duan, H. Yan [et al.] // *Current Protein and Peptide Science*. – 2021. – Vol. 22, N 4. – P. 290–303. <https://doi.org/10.2174/138920372666210421105733>
13. *Mycobacterium smegmatis*: the vanguard of mycobacterial research / I. L. Sparks, K. M. Derbyshire, W. R. Jr. Jacobs, Ya. S. Morita // *Journal of Bacteriology*. – 2023. – Vol. 205, N 1. – Art. e00337-22. <https://doi.org/10.1128/jb.00337-22>
14. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings / C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney // *Advanced Drug Delivery Reviews*. – 2001. – Vol. 46, N 1–3. – P. 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
15. Sterling, T. ZINC 15 – ligand discovery for everyone / T. Sterling, J. J. Irwin // *Journal of Chemical Information and Modeling*. – 2015. – Vol. 55, N 11. – P. 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>

References

1. Deore A. B., Dhumane J. R., Wagh R., Sonawane R. The stages of drug discovery and development process. *Asian Journal of Pharmaceutical Research and Development*, 2019, vol. 7, no. 6, pp. 62–67. <https://doi.org/10.22270/ajprd.v7i6.616>
2. Berdigaliyev M., Aljofan N. An overview of drug discovery and development. *Future Medicinal Chemistry*, 2020, vol. 12, no. 10, pp. 939–947. <https://doi.org/10.4155/fmc-2019-0307>
3. Roney M., Mohd Aluwi M. F. F. The importance of in-silico studies in drug discovery. *Intelligent Pharmacy*, 2024, vol. 2, no. 4, pp. 578–579. <https://doi.org/10.1016/j.ipha.2024.01.010>
4. Shaker B., Ahmad S., Lee J., Jung C., Na D. *In silico* methods and tools for drug discovery. *Computers in Biology and Medicine*, 2021, vol. 137, art. 104851. <https://doi.org/10.1016/j.compbiomed.2021.104851>
5. Mouchlis V. D., Afantitis A., Serra A., Fratello M., Papadimitriou A. G., Aidinis V., Lynch I., Greco D., Melagraki G. Advances in *de novo* drug design: from conventional to machine learning methods. *International Journal of Molecular Sciences*, 2021, vol. 22, no. 4, art. 1676. <https://doi.org/10.3390/ijms22041676>
6. Khawbung J. L., Nath D., Chakraborty S. Drug resistant tuberculosis: a review. *Comparative Immunology, Microbiology and Infectious Diseases*, 2021, vol. 74, art. 101574. <https://doi.org/10.1016/j.cimid.2020.101574>
7. Stokes J. M., Yang K., Swanson K., Jin W., Cubillos-Ruiz A., Donghia N. M., MacNair C. R., French S., Carfrae L. A., Bloom-Ackermann Z., Tran V. M., Chiappino-Pepe A., Badran A. H., Andrews I. W., Chory E. J., Church G. M., Brown E. D., Jaakkola T. S., Barzilay R., Collins J. J. A deep learning approach to antibiotic discovery. *Cell*, 2020, vol. 180, no. 4, pp. 688–702. <https://doi.org/10.1016/j.cell.2020.01.021>
8. Almeleebia T. M., Shahrani M. A., Alshahrani M. Y., Ahmad I., Alkahtani A. M., Alam M. J., Kausar M. A., Saeed A., Saeed M., Iram S. Identification of new *Mycobacterium tuberculosis* proteasome inhibitors using a knowledge-based computational screening approach. *Molecules*, 2021, vol. 26, no. 8, art. 2326. <https://doi.org/10.3390/molecules26082326>
9. Zheng S., Gu Ya., Gu Yu., Zhao Y., Li L., Wang M., Jiang R., Yu X., Chen T., Li J. Machine learning-enabled virtual screening indicates the anti-tuberculosis activity of aldoxorubicin and quarfloxin with verification by molecular docking, molecular dynamics simulations, and biological evaluations. *Briefings in Bioinformatics*, 2024, vol. 26, no. 1, art. bbae696. <https://doi.org/10.1093/bib/bbae696>
10. Felker I. G., Gordеева E. I., Stavitskaya N. V., Pershina V. A., Batyrshina Ya. R. Prospects and obstacles for clinical of the inhibitors of *Mycobacterium tuberculosis* efflux pumps. *Biologicheskie membrany = Biological Membranes*, 2021, vol. 38, no. 5, pp. 317–339 (in Russian).
11. Imran M., Arora M. K., Chaudhary A., Khan S. A., Kamal M., Alshammari M. M., Alharbi R. M., Althomali N. A., Alzimam I. M., Alshammari A. A., Alharbi B. H., Alshengeti A., Alsaleh A. A., Alqahtani S. A., Rabaan A. A. MmpL3 inhibition as a promising approach to develop novel therapies against tuberculosis: a spotlight on SQ109, clinical studies, and patents literature. *Biomedicines*, 2022, vol. 10, no. 11, art. 2793. <https://doi.org/10.3390/biomedicines10112793>
12. Luo Q., Duan H., Yan H., Liu X., Peng L., Hu Y., Liu W., Liang L., Shi H., Zhao G., Hu J. Specifically targeting Mtb cell-wall and TMM transporter: the development of MmpL3 inhibitors. *Current Protein and Peptide Science*, 2021, vol. 22, no. 4, pp. 290–303. <https://doi.org/10.2174/1389203722666210421105733>
13. Sparks I. L., Derbyshire K. M., Jacobs Jr. W. R., Morita Y. S. *Mycobacterium smegmatis*: the vanguard of mycobacterial research. *Journal of Bacteriology*, 2023, vol. 205, no. 1, art. e00337-22. <https://doi.org/10.1128/jb.00337-22>
14. Lipinski C. A., Lombardo F., Dominy B. W., Feeney P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 2001, vol. 46, no. 1–3, pp. 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
15. Sterling T., Irwin J. J. ZINC 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 2015, vol. 55, no. 11, pp. 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>

Информация об авторах

Башко Георгий Мстиславович – студент. Белорусский государственный университет (пр-т Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: grgbashko@gmail.com.

Корноушенко Юрий Валерьевич – канд. хим. наук, ст. науч. сотрудник. Институт биоорганической химии НАН Беларуси (ул. Купревича, 5/2, 220084, Минск, Республика Беларусь). E-mail: yurakorval@gmail.com.

Тузиков Александр Васильевич – член-корреспондент, д-р физ.-мат. наук, профессор, заведующий лабораторией. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: tuzikov@newman.bas-net.by.

Андрианов Александр Михайлович – д-р хим. наук, профессор, гл. науч. сотрудник. Институт биоорганической химии НАН Беларуси (ул. Купревича, 5/2, 220084, Минск, Республика Беларусь). E-mail: alexande.andriano@yandex.ru

Information about the authors

Bashko Georgy M. – Student. Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: grgbashko@gmail.com.

Kornoushenko Yuri V. – Ph. D. (Chemistry), Senior Researcher. Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus (5/2, Kuprevich Str., 220084, Minsk, Republic of Belarus). E-mail: yurakorval@gmail.com.

Tuzikov Alexander V. – Corresponding Member, D. Sc. (Physics and Mathematics), Professor, Head of the Laboratory. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: tuzikov@newman.bas-net.by.

Andrianov Alexander M. – D. Sc. (Chemistry), Professor, Chief Researcher. Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus (5/2, Kuprevich Str., 220084, Minsk, Republic of Belarus). E-mail: alexande.andriano@yandex.ru.