

# АЛГОРИТМ ML-МОДЕЛИ, ПРОГНОЗИРУЮЩЕЙ ДОКЛИНИЧЕСКУЮ ЭФФЕКТИВНОСТЬ СОЕДИНЕНИЙ НА ОСНОВЕ СУЩЕСТВУЮЩИХ БАЗ ДАННЫХ, В КАЧЕСТВЕ АЛЬТЕРНАТИВНОГО МЕТОДА ИСПЫТАНИЯ НА ЖИВОТНЫХ

Савельев М. Ю., Вычкин А. В., Андриенко Я. Е., Меркулова А. С., Селезнева А. А.  
Научные руководители: Соколов Ф.С., Царегородцев С.В.

ФГБОУ ВО Российский университет медицины  
Минздрава России

**Резюме.** Проект посвящен созданию алгоритма обучения Machine Learning (ML)-модели искусственного интеллекта (ИИ) с целью ускорения и облегчения проведения доклинических исследований (ДКИ). В рамках работы был выполнен анализ публикаций и исследований существующих алгоритмов и методов оценки эффективности и безопасности *in silico*, а также сопутствующих законодательных актов. Обучение в соответствии с алгоритмом и дальнейшее внедрение такой ML-модели, как альтернативного метода в ДКИ, поможет прогнозировать свойства и активность молекул, а также при необходимости, оптимизировать молекулы-кандидаты, сократить количество испытаний на животных и ускорить анализ данных экспериментов. Наш алгоритм в перспективе может стать фундаментом для инновационных методов разработки лекарств для стран ЕАЭС.

**Ключевые слова:** машинное обучение, фармакология, молекулярный докинг, доклинические исследования, испытания на животных, принцип 3R.

**Актуальность.** Искусственный интеллект (ИИ) активно влияет на многие аспекты разработки лекарственного средства, включая поиск и оптимизацию молекул, доклинические и клинические испытания, клиническую токсикологию. Внедрение ИИ в процесс разработки лекарственных препаратов на стадии доклинических исследований (ДКИ) позволяет с наиболее эффективно оценивать безопасность молекулы-мишени. Так, виртуальное моделирование сложных биологических систем с использованием технологий ML/AI, а также платформ Organ-on-Chip (OoC), 3D-культур клеток совершенствует и ускоряет процесс ДКИ, реализуя

принципы 3R (replacement, reduction, refine-ment/замена, сокращение, усовершенствование) в соответствии с актуальными этическими принципами. В рекомендациях коллегии Евразийской экономической комиссии от 14 ноября 2023 г. N 33 "О Руководстве по работе с лабораторными (экспериментальными) животными при проведении доклинических (неклинических) исследований" отдельное внимание посвящено выбору метода исследования безопасности, по возможности альтернативного животным моделям. Безусловно, полный отказ от тестирования на животных в пользу ИИ-моделей невозможен ввиду

ограничений в прогнозировании и трансляции результатов на биологические системы ИИ-моделями, однако сочетание таких методов является наиболее результативным подходом с экономической и этической точки зрения.

По данным обзора "Gangwal A. et al. (2025)" создание и применение локального ИИ позволяет не только уменьшить использование лабораторных животных на этапе *in vivo* и сократить время ДКИ за счет виртуального скрининга, но и повышает точность прогнозирования эффективности и токсичности молекул. Кроме того, создание собственной базы данных с дальнейшим обучением на ее основе ML-модели оптимизирует этапы R&D, ускоряет выход препаратов на рынок, обеспечивая таким образом технологический суверенитет и конкурентное преимущество.

**Цель:** Проектирование алгоритма ML-модели ИИ, которая в результате обучения станет основой для создания локальной базы данных, а также позволит ускорить этап ДКИ за счет сокращения использования животных, минимизации ошибок в поиске наиболее перспективной молекулы.

**Задачи:**

1. Поиск и систематизация исследований из открытых источников.

2. Сравнение эффективности иностранных и Российских платформ, использующихся на этапах ДКИ с целью прогнозирования вида биологической активности.

Создание алгоритма обучения перспективной ML-модели.

**Материалы и методы.**

Проводился критический анализ качества отобранных исследований с использованием стандартизированных инструментов оценки для машинного обучения искусственного интеллекта в прогнозировании доклинической эффективности соединений. Поиск литературы охватил период с января 2020 г. по ноябрь 2025 г. и включал базы PubMed/MEDLINE, PASSonline, Google Scholar, eLibrary.ru. Рассматривались оригинальные и обзорные статьи, посвященные доклиническим исследованиям, испытаниям на животных, машинному обучению, прогнозированию токсичности соединений. Для выявления релевантных публикаций использовались сочетания терминов и их аналогов на английском языке, по ключевым словам: «машинное обучение», «искусственный интеллект», «испытания на животных», «глубокое обучение» и др. Дополнительный поиск выполнялся по спискам литературы отобранных статей.

**Результаты и их обсуждение.** В результате проведенной оценки и анализа литературных источников был разработан алгоритм самообучения локальной ML-модели в доклинических исследованиях (рис.1), который в формате замкнутой системы способен к накоплению данных. Обучение данной платформы может осуществляться на открытых базах данных, таких как DrugBank, ChEMBL, IT Microcosm, PubCHEM,

Chebi. В дальнейшем при помощи веб-систем и информационно-вычислительных платформ, например, SwissADME и Way2Drug возможно прогнозирование свойств молекул, включая их биологическую активность, ограниченные фармакокинетические параметры, а также специфическую токсичность по отношению к конкретным органам-мишеням. Эти данные служат основой для выбора молекул и проведения их дальнейшего тестирования. Так, можно не только осуществлять поиск наиболее перспективных молекул-кандидатов, но и при необходимости модифицировать их с помощью метода прогнозирования QSAR и процесса оптимизации молекулы “Hit to Lead”, учитывая липофильность, стерические и электронные эффекты соединения.

Следующий этап алгоритма подразумевает обработку результатов тестов *in vitro* (МТТ-тест, Ревазуриновый тест) и данных, полученных с использованием Organ-on-Chip и на 3D-культурах, которые представляют собой альтернативу исследования безопасности на животных моделях, и способны имитировать тканеспецифичные функции биологического организма. Описанные виртуальные органные модели, контролируемые ИИ, способны воспроизводить физиологические особенности, автономно поддерживать гомеостаз, учитывать видовую принадлежность, тем самым повышая эффективность, воспроизводимость исследований и минимизируя необходимое количество животных в дальнейших

исследованиях *in vivo*. OoC-платформа для воспроизведения различных функций тканей - мощный инструмент для изучения межорганных взаимодействий, эффектов лекарств и механизмов заболеваний. Этот мультиорганный подход устраняет разрыв между исследованиями *in vitro* и биологией человека, способствуя прогрессу в персонализированной медицине и тестировании лекарств. Результаты исследований *in silico* и обязательно проводимых *in vivo* в дальнейшем подвергаются анализу ML-моделью, а на основании полученной информации происходит обучение модели и формирование собственной базы данных.

В рамках проведённого анализа осуществлено сравнение функциональных возможностей зарубежных платформ (баз данных, цифровых ресурсов и веб-сервисов), применяемых на этапах доклинических исследований для прогнозирования цито-, гепато-, нефро-, и кардиотоксичности, канцерогенности и иных видов токсического действия, с соответствующими инструментами, разработанными и используемыми в странах СНГ. Зарубежные предсказательные системы, включая AnimalGAN (FDA), ADMETlab 2.0 и DDI, согласно данным литературы, в большинстве случаев демонстрируют высокие значения прогностической точности, достигающие в ряде исследований 92–100%.

Аналогичные программно-аналитические решения, доступные на территории СНГ (PASS Online,

ADVER-Pred, CLC-Pred, MetaTox и другие сервисы платформы Way2Drug), характеризуются сопоставимо высокими показателями эффективности. Вместе с тем их прямое сравнение с зарубежными моделями представляется методологически затруднительным вследствие гетерогенности целевых конечных точек, различий в структуре обучающих выборок, используемых метрик качества, подходов к классификации и особенностей интерпретации выходных данных.

Так, точность прогнозирования канцерогенности отечественной нейросетевой моделью ROSC-Pred составляет около 95,2%, что соответствует уровню современных зарубежных решений. Модуль CLC-Pred демонстрирует точность порядка 92% при предсказании цитотоксичности, а средняя точность определения потенциальной биоактивности по данным PASS Targets достигает 96%. Платформы, учитывающие метаболическую трансформацию веществ (MetaTox, RA, SOMP), обеспечивают значения точности в диапазоне 90–99%, что также коррелирует с мировыми показателями.

Полученные результаты позволяют заключить, что российские базы данных и веб-сервисы, представленные на территории СНГ, для прогнозирования токсических эффектов обладают существенным потенциалом и способны удовлетворять специфические потребности фармацевтического сектора. Однако ограниченный доступ

к закрытым зарубежным источникам данных, требования к защите информации и ряд технологических ограничений подчеркивают необходимость формирования полнофункциональной локальной базы данных, которая может служить фундаментом для последующего обучения и валидации отечественных моделей машинного обучения.

#### **Выводы:**

1. Внедрение в практику разработанного нами алгоритма обучения ML-модели способно повысить эффективность, снизить стоимость и время проведения этапов ДКИ, благодаря использованию веб-сервисов и ИИ-платформ. Как на этапе прогнозирования перспективы исследования молекулы-мишени, так и на этапе оценки ее безопасности, что дает возможность минимизировать использование животных моделей и соответствует современным этическим стандартам ЕАЭС и принципу 3R.

2. Российские и зарубежные платформы демонстрируют сопоставимую эффективность в прогнозировании молекул-кандидатов на доклинических этапах исследований, однако их прямое сравнение затруднено различиями в методологии и доступности данных.

3. Созданная в результате обучения ML-модели ИИ база данных позволяет обеспечить условия локального технологического суверенитета и информационной безопасности для фармкомпаний стран ЕАЭС

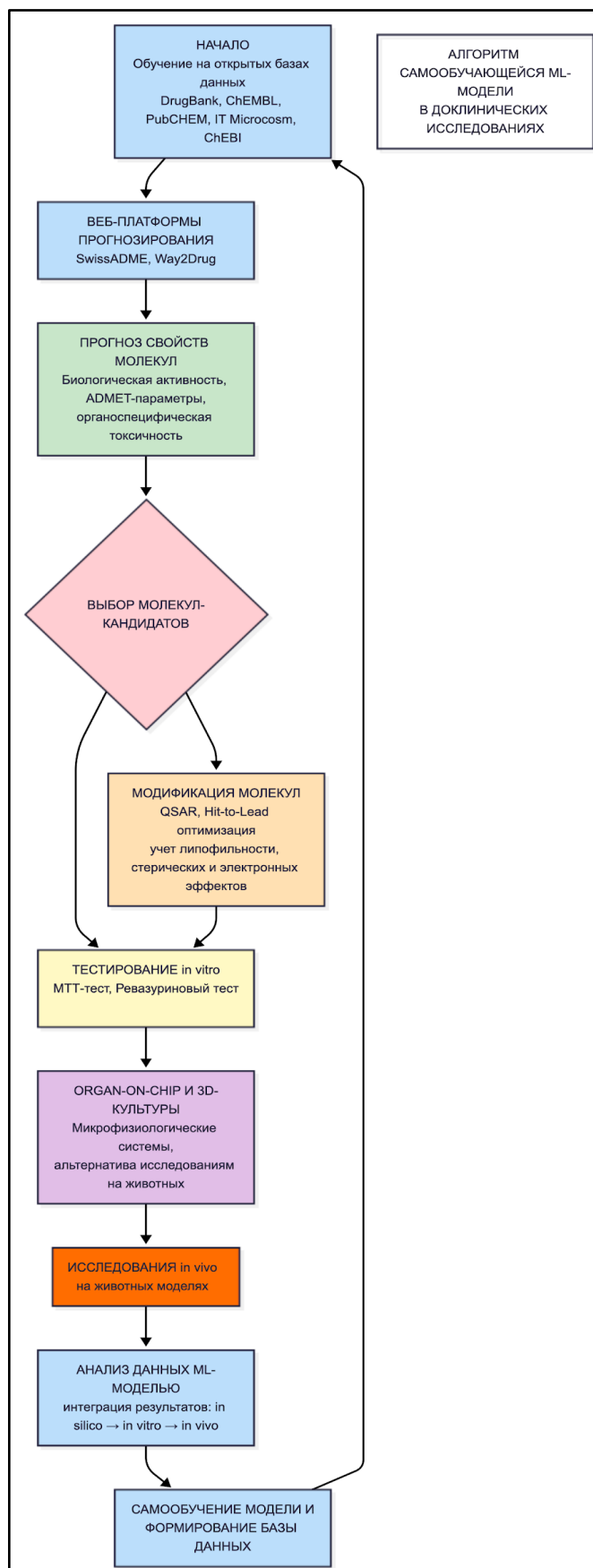


Рис. 1 – Алгоритм самообучающейся ML-модели в доклинических исследованиях

## Литература

1. Sangsoo Lim, Sangseon Lee, Yinhua Piao, MinGyu Choi, Dongmin Bang, Jeonghyeon Gu, Sun Kim, On modeling and utilizing chemical compound information with deep learning technologies: A task-oriented approach, *Computational and Structural Biotechnology Journal*, Volume 20, 2022, Pages 4288-4304, ISSN 2001-0370, <https://doi.org/10.1016/j.csbj.2022.07.049>.
2. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, Wu C, Lu A, Chen X, Hou T, Cao D. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W5-W14. doi: 10.1093/nar/gkab255. PMID: 33893803; PMCID: PMC8262709.
3. Chen Y, Yu X, Li W, Tang Y, Liu G. In silico prediction of hERG blockers using machine learning and deep learning approaches. *J Appl Toxicol.* 2023 Oct;43(10):1462-1475. doi: 10.1002/jat.4477. Epub 2023 May 6. PMID: 37093028.
4. Chen, X., Roberts, R., Liu, Z. et al. A generative adversarial network model alternative to animal studies for clinical pathology assessment. *Nat Commun* 14, 7141(2023). <https://doi.org/10.1038/s41467-023-42933-9>.
5. Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 89 «Об утверждении Правил проведения исследований биологических лекарственных средств Евразийского экономического союза». – [Электронный ресурс]. – Режим доступа: <https://docs.eaeunion.org/documents/306/2596/> (дата обращения: 16.10.2025).
6. Руководство для промышленности по валидации биоаналитических методов [Guidance for Industry: Bioanalytical Method Validation]. — Мин-во здравоохранения и соц. служб США, Управление по пищ. и медпродуктам (FDA), Центр по оценке и исслед. лекарств (CDER). — Вашингтон, округ Колумбия : Тип. Правительства США, 2018. — [Электронный ресурс]. — Режим доступа: <https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf> (дата обращения: 16.10.2025).
7. Руководство по принципам регуляторного признания подходов к тестированию, основанных на концепции 3R (замена, сокращение, усовершенствование) [Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches][Электронный ресурс] : принято : 21 марта 2015 г. : пересмотр. : 15 дек. 2016 г. / Европ. агентство по мед. препаратам (ЕМА), Комитет по мед. продуктам для человека (CHMP), Комитет по мед. продуктам для вет. применения (CVMP). — Режим доступа: [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches_en.pdf) (дата обращения: 15.10.2025).
8. Chen, X., Roberts, R., Liu, Z. et al. A generative adversarial network model alternative to animal studies for clinical pathology assessment. *Nat Commun* 14, 7141 (2023). <https://doi.org/10.1038/s41467-023-42933-9>
9. Коллегия ЕЭК. Решение от 26 ноября 2019 г. № 202 «Об утверждении Руководства по доклиническим исследованиям безопасности в целях проведения клинических исследований и регистрации лекарственных препаратов». – [Электронный ресурс]. – Режим доступа: <https://docs.eaeunion.org/documents/335/4808/> (дата обращения: 20.10.2025).
10. Международный совет по гармонизации технических требований к регистрации лекарственных средств для человека (ICH). Руководство M3(R2) по доклиническим исследованиям безопасности для проведения клинических исследований лекарственных средств и регистрации фармацевтических препаратов, декабрь 2009 г. [Электронный ресурс]. – Режим доступа: [https://database.ich.org/sites/default/files/M3\\_R2\\_Guideline.pdf](https://database.ich.org/sites/default/files/M3_R2_Guideline.pdf) (дата обращения: 20.10.2025).
11. Gangwal A., Lavecchia A. Artificial intelligence in preclinical research: enhancing digital twins and organ-on-chip to reduce animal testing // *Drug Discovery Today*. 2025. № 5 (30). С. 104360. <https://doi.org/10.1016/j.drudis.2025.104360>

# ALGORITHM OF AN ML-MODEL FOR PREDICTING PRECLINICAL EFFICACY OF COMPOUNDS BASED ON EXISTING DATABASES, AS AN ALTERNATIVE TO ANIMAL TESTING

*Saveliev M. U., Vychkin A. V., Andrienko Y. E., Merkulova A. S., Selezneva A. A.*

*Tutors: Sokolov Ph. S., Tsaregorodtsev S. V.*

*FSBEI HE "Russian University Of Medicine" Of The Ministry Of Health  
Of The Russian Federation*

**Resume.** The project is dedicated to the development of a training algorithm for a Machine Learning (ML) model of Artificial Intelligence (AI), aimed at accelerating and facilitating pre-clinical research. The work involved an analysis of publications and studies on existing in silico algorithms and methods for assessing efficacy and safety, as well as relevant legislative acts. Training according to this algorithm and the subsequent implementation of such an ML-model as an alternative method in pre-clinical studies will help predict molecular properties and activity, optimize candidate molecules if necessary, reduce the number of animal trials, and speed up the analysis of experimental data. Our algorithm has the potential to become the foundation for innovative drug development methods in the EAEU countries.

**Keywords:** machine learning, pharmacology, drug design, molecular docking, animal testing, 3Rs principle.