

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ФАРМАКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ: ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ В СОЗДАНИИ ЛЕКАРСТВЕННЫХ СРЕДСТВ

Максим Алексеевич Перфильев¹✉, Павел Михайлович Васильев²,
Дмитрий Васильевич Мальцев³, Мария Олеговна Скрипка⁴

^{1,2,3,4} Волгоградский государственный медицинский университет,
кафедра фармакологии и биоинформатики, Волгоград, Россия

^{1,2} Лаборатория информационных технологий в фармакологии и компьютерного моделирования
лекарств, Научный центр инновационных лекарственных средств, Волгоград, Россия

^{3,4} Лаборатория экспериментальной фармакологии, Волгоградский медицинский научный центр,
Волгоград, Россия

¹✉ maxim.firu@yandex.com, Elibrary AuthorID: 1131051, <https://orcid.org/0000-0002-5326-3299>

² pvassiliev@mail.ru, Elibrary AuthorID: 101402, <https://orcid.org/0000-0002-8188-5052>

³ maltsevdmitriy@rambler.ru, Elibrary AuthorID: 628229, <https://orcid.org/0000-0002-2005-6621>

⁴ marusyaskripka@mail.ru, Elibrary AuthorID: 1045082, <https://orcid.org/0000-0002-4173-7143>

Аннотация. Проведен анализ публикаций по компьютерному поиску лекарственных соединений в открытых базах данных научной литературы. Обзор посвящен современным методам доэкспериментальных исследований фармакологически активных соединений, актуальным вопросам применения искусственных нейронных сетей. Представлены новые направления в исследованиях *in silico*. Показаны приоритетные направления в этой области.

Ключевые слова: лекарственные вещества, скрининг *in silico*, машинное обучение, искусственные нейронные сети, квантовые вычисления

Благодарности: Обзор выполнен в рамках государственного задания Министерства здравоохранения Российской Федерации № 121060700050-2 «Разработка методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей».

Для цитирования: Современные технологии фармакологических исследований: искусственные нейронные сети в создании лекарственных средств / М. А. Перфильев, П. М. Васильев, Д. В. Мальцев, М. О. Скрипка // Лекарственный вестник. 2022. Т. 23, № 1. С. 22–26.

В настоящее время для поиска фармакологически активных соединений широко применяются различные подходы *in silico*. Наиболее распространенными среди них являются методы виртуального скрининга с помощью молекулярного моделирования, например, докинг и молекулярная динамика (МД). Активно используются алгоритмы классификации и регрессии деревом решений (Classification and Regression Tree – CART). Эти методы позволяют быстро и достаточно точно найти соединения-лидеры из числа новых химических структур, кроме того, они успешно применяются с целью перефилирования лекарств [1].

Иные подходы заложены в таких методах машинного обучения, как искусственные нейронные сети, которые позволяют сделать процесс поиска биологически активных веществ еще более эффективным [2]. Методы искусственного интеллекта, как и широко применяемые классические методы компьютерного скрининга, также можно использовать в направленном поиске новых соединений с требуемым видом фармакологической активности или пойти путем перефилирования, достигая при этом значительно большей точности прогноза.

Технологии искусственных нейронных сетей используются в трех основных областях

вычислительной фармакологии: предсказание взаимодействия лекарство-мишень; создание новых молекул; предсказание свойств абсорбции, распределения, метаболизма, экскреции и токсичности химических соединений (ADMET) [3].

Подобная популярность технологий искусственных нейронных сетей обуславливается несколькими факторами, а именно:

1) бионической аналогией с естественными способами обработки информации живыми организмами, в частности, напрямую имитирующих когнитивные функции человека;

2) эвристическим принципом поиска решения, что позволяет оперировать малыми объемами входных данных;

3) универсальным подходом к построению моделей, применимым для обработки данных любой структуры и сложности;

4) высокой эффективностью работы с большими данными.

За последние 5 лет направление доэкспериментальных *in silico* исследований значительно продвинулось вперед. В 2021 г. был предложен целый ряд принципиально новых подходов в области искусственного интеллекта (ИИ) при создании лекарств.

Для написания обзора был проведен анализ публикаций из открытых баз данных научной литературы Science Direct, PubMed, Google Scholar, а также из профильных сетевых изданий, посвященных методам компьютерного поиска лекарственных соединений.

AlphaFold2

Предсказание трехмерной структуры белка по первичной аминокислотной последовательности долгое время оставалось достаточно сложно решаемой задачей. Использование для этой цели нейронной сети AlphaFold2 показало, что искусственный интеллект может весьма точно предсказывать белковые 3D-структуры, обеспечивая понимание функций человеческого протеома.

Система представляет собой комплекс из нескольких различных по архитектуре нейросетей, в основе которых лежит выравнивание множественных последовательностей первичной аминокислотной цепи. AlphaFold2 быстро стала весьма популярным инструментом одной из основных баз данных по белкам UniProt.

22 июля 2021 г. исходный код AlphaFold2 был открыт для свободного доступа [4].

BiteNet

Ученые из группы iMolecule Сколковского института науки и технологий по научным и инженерным вычислительным технологиям в задачах с большими массивами данных (CDISE) разработали алгоритм машинного обучения для поиска в белках сайтов связывания с лигандами – потенциальных лекарственных мишеней.

Система BiteNet анализирует 3D-белковые структуры и находит оптимальные места для присоединения малых молекул. Метод, представленный разработчиками, основан на применении к белковым структурам, которые представлены в виде трехмерных изображений, алгоритмов глубокого обучения и компьютерного зрения. Технология позволяет детектировать сайты, которые скрыты внутри экспериментальных моделей или формируются несколькими белковыми субъединицами [5].

Grokking

Удобство и универсальность технологии искусственных нейронных сетей основывается на принципах моделирования, которые были сформулированы в 1958 г. А. Н. Колмогоровым [6]. Вне зависимости от сложности модели при достаточно большом времени обучения возможно аппроксимировать любую функцию как суперпозицию нескольких простых элементарных функций. Недавняя работа ученых компании Google вновь подтвердила эти положения [7].

Обучение нейронных сетей происходит эпохами, в ходе которых сети могут запоминать объекты обучающей выборки. В этом случае распознающая способность сети, направленная на классификацию исходных данных в процессе обучения, весьма высока, но точность прогностической способности на тестовой выборке оказывается значительно более низкой. Частым решением становится компромисс между результатами распознающей и прогностической способностей. Проблема переобучения наиболее актуальна, когда для моделирования были подобраны относительно небольшие объемы данных, а архитектура нейронной сети, наоборот, достаточно сложная.

Сотрудники научного подразделения компании Google установили, что если продолжить

обучение после того, как наступает переобучение, то модель все-таки может выйти на нормальную предсказательную способность.

При 10^6 эпох, значительно позже стадии переобучения, веса межнейронных связей модели перенастраиваются таким образом, что значение ошибки на обучающей и на контрольных выборках стремится к 0. Этот эффект и назвали гроккингем.

ISM001-05

30 ноября 2021 г. компания Insilico Medicine сообщила, что найденное с помощью ИИ соединение, средство для лечения идиопатического легочного фиброза, является первым подобным веществом, которое проходит клинические испытания на людях [8]. Соединение с шифром ISM001-055 было разработано с использованием ряда утилит и программ машинного обучения. Основным инструментом является платформа, которая состоит из трех ключевых компонентов: механизма обнаружения потенциальных биомшеней и специального анализа данных (PandaOmics), механизма молекулярного проектирования *de novo* (Chemistry42) и механизма прогнозирования результатов клинических испытаний (InClinico). Разработка помогла сократить время от обнаружения мишени до выдвижения кандидата на доклинические исследования с 3–6 лет до 18 месяцев, при бюджете в 2,6 млн дол., что в 170 раз меньше, чем обычно. В ходе длительных исследований было выявлено новое соединение и новый целевой белок – мишень для лечения идиопатического легочного фиброза. Высокоинтегрированная архитектура позволила интегрировать биологические и химические методы в плавный, хорошо организованный исследовательский рабочий процесс. Целевая оценка новизны и ассоциации с заболеванием оценивалась комплексом алгоритмов обработки естественного языка – Natural Language Processing (NLP), который анализирует миллионы файлов, включая патенты, исследовательские публикации, гранты и базы данных клинических испытаний.

У данного подхода есть оппоненты, подвергающие сомнению результаты работы, которые утверждают, что найденная новая мишень не является «новой», хотя она и впервые попадает в сферу клинических испытаний.

Результаты использования NLP для выбора целевой биомшени также нередко подвергаются сомнению, поскольку часто уже имеются достоверные публикации, подтверждающие значимость выявленной мишени.

Кроме того, в подобных статьях часто рассматриваются химические соединения, релевантные данным биомшеням. Нейронные сети не оставляют такие структуры без внимания, что ведет к созданию с их использованием молекулы, во многом схожей или прямо копирующей ту, что уже была опубликована. Однако все исследователи соглашаются, что если учесть объемы информации, проанализированные в подобных работах, то становится ясно – решить подобные задачи можно только с применением технологий машинного обучения [9].

Консорциумы в сфере разработки лекарств

Современные фармакологические исследования остаются очень затратными в плане материальных, финансовых и человеческих ресурсов и затрат времени. Обозначенные проблемы могут быть нивелированы путем использования высокопроизводительных вычислительных систем. Однако только наиболее крупные фирмы могут позволить себе создавать такие компьютеры и поддерживать их функционирование, что часто осуществляется путем создания консорциумов.

Проекты использования суперкомпьютеров в нейросетевом моделировании, посвященные наукам о жизни, выполняются, например, партнерством AstraZeneca, GlaxoSmithKline (GSK), фондом National Health Service Foundation Trust, Королевским колледжем в Лондоне и Oxford Nanopore Technologies. Они сосредоточились на улучшении модели, которая ускорит процесс создания лекарств за счет прогнозирования уровня активности, молекулярной оптимизации и создания молекул *de novo*. GSK и NVIDIA намерены разрабатывать более эффективные соединения, объединяя генетику, геномику и машинное обучение с использованием передовой технологии искусственного интеллекта [10].

Другим крупным проектом является сотрудничество компаний Roche, Genentech и Recursion, которые запустили совместную работу по поиску лекарств с помощью искусственного интеллекта. Roche, пионер в области

фармацевтики и диагностики, и ее дочерняя биотехнологическая компания Genentech 12 декабря 2021 г. начали кооперацию с Recursion, биотехнологической компанией, специализирующейся на проведении клинических испытаний лекарственных препаратов, которая применяет машинное обучение для создания лекарств.

Кооперация позволит использовать разработки Recursion для ускорения выявления новых целевых рецепторов, а также исследований в области неврологии и онкологии [11].

Молекулярная динамика

Метод молекулярной динамики анализирует, как молекулы ведут себя и взаимодействуют друг с другом в течение определенного временного промежутка, в том числе с использованием их полноатомного представления. При разработке лекарств наиболее часто МД применяется для оценки взаимодействия между белком и лигандом, а также для оценки стабильности их связывания. МД является достаточно достоверной и надежной, но весьма трудоемкой и долгой процедурой. Методы машинного обучения способны ускорить процесс симуляции с помощью МД. В работе, посвященной данному вопросу, было выполнено МД-моделирование с расчетом свободных энергий для последующего обучения сверточной искусственной нейронной сети [12]. Исследователи обнаружили, что обученные нейросети предсказывали значения свободных энергий с высокой точностью, сопоставимой с результатами полученными методом МД-моделирования. Это исследование показывает, что нейронные сети могут ускорить процесс МД, однако для этого требуются большие обучающие выборки.

CNN – SMILES

Первоначальный подход при использовании методов глубокого обучения заключался в использовании обработанных человеком исходных данных, таких как молекулярные дескрипторы. К настоящему времени все больше исследований нацеливаются в первую очередь на представление «необработанных» химических данных. В одной из таких работ описывается использование SMILES [13] – однострочного текстового формата, позволяющего взаимоднозначно представлять структуры химических соединений [14].

Собственные разработки

Некорректно считать, что без привлечения мощных суперкомпьютеров или какой-либо иной помощи от крупнейших технологических компаний невозможно использовать методы искусственного интеллекта.

Многообразие подходов в построении различных архитектур нейронных сетей позволяет применять их в разных по сложности задачах. Так, при наличии нескольких биомшеней, влияющих на конкретный фармакологический ответ, более целесообразным будет использовать для комплексного описания биологической системы полносвязные нейронные сети [15]. Такие архитектуры используются для создания моделей с ассоциативной памятью. Суть такой задачи сводится к изучению с помощью сети объекта, данные о котором могут быть искажены. После ввода входных данных, параметры сети пересчитываются и по результатам вычислений определяют, соответствует ли анализируемый объект ранее сформированному образу.

В рамках общей постановки задачи поиска зависимостей «структура – активность» весьма целесообразно проводить вариативное представление химической структуры фармакологически активных соединений. К примеру, здесь могут быть использованы упомянутые ранее фрагментные дескрипторы. В лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств НЦИЛС ВолгГМУ и на кафедре фармакологии и биоинформатики ВолгГМУ, кроме фрагментных, применяются и другие дескрипторы, например, квантово-химические параметры и параметры аффинности. При подготовке обучающих и прогнозных выборок применяются различные способы свертки входных данных. На этой основе разрабатываются качественно новые архитектуры консенсусных ансамблевых нейросетевых моделей с различными типами представления входных данных [16]. Таким образом, в настоящее время исследователи имеют возможность при использовании методов искусственного интеллекта выбирать самые разные варианты их использования, что делает технологию искусственных нейронных сетей весьма перспективной, быстрым и качественным подходом для фармакологических исследований.

СПИСОК ИСТОЧНИКОВ

1. *Vanhaelen Q.* Computational Methods for Drug Repurposing // Humana Press. 2019. P. 343.
2. Deep learning enables rapid identification of potent DDR1 kinase inhibitors / A. Zhavoronkov [et al.] // Nature Biotechnology. 2019. Vol. 37. P. 1038–1040.
3. *Batool M., Ahmad B., Choi S.* A Structure-Based Drug Discovery Paradigm // International Journal of Molecular Sciences. 2019. Vol. 20. Iss. 11.
4. *Tunyasuvunakool K., Adler J., Wu Z.* Highly accurate protein structure prediction for the human proteome // Nature. 2021. Vol. 596. P. 590–596.
5. *Kozlovskii I., Popov P.* Spatiotemporal identification of druggable binding sites using deep learning // Communications Biology. 2020. Vol. 3, no 618.
6. *Колмогоров А. Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного // Докл. АН СССР. 1958. Т. 114, № 5. С. 953–956.
7. Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets / A. Power [et al.] // Materials of Math-AI ICLR 2021 Workshop. 2021. URL: https://mathai-iclr.github.io/papers/papers/MATHAI_29_paper.pdf
8. A breakthrough milestone in AI-powered drug discovery reached // insilico.com. 2021. URL: <https://insilico.com/blog/pcc> (дата обращения: 24.01.22).
9. *Zhavoronkov A.* Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry // Molecular Pharmaceutics. 2018. Vol. 15, no. 10. P. 4311–4313.
10. *Ciborowski J.* NVIDIA Launches UK's Most Powerful Supercomputer, for Research in AI and Healthcare // NvidiaNews. 2021. URL: <https://nvidianews.nvidia.com/news/nvidia-launches-uks-most-powerful-supercomputer-for-research-in-ai-and-healthcare> (дата обращения: 24.01.22).
11. *Chino Y.* Roche, Genentech, Recursion Launch Up-to-\$12B AI Drug Discovery Effort // Genetic Engineering & Biotechnology News. 2021. URL: <https://www.genengnews.com/artificial-intelligence/roche-genentech-recursion-launch-up-to-12b-ai-drug-discovery-effort/> (дата обращения: 24.01.22).
12. Predicting small molecule transfer free energies by combining molecular dynamics simulations and deep learning / W. F. Drew Bennett [et al.] // Journal of Chemical Information and Modeling. 2020. Vol. 60, no. 11. P. 5375–5381.
13. *Weininger D.* SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules // Journal of Chemical Information and Computer Sciences. 1988. Vol. 28, no. 1. P. 31–36.
14. Artificial intelligence to deep learning: machine intelligence approach for drug discovery / R. Gupta [et al.] // Molecular Diversity. 2021. Vol. 25. P. 1315–1360.
15. Корреляционная сеть ингибиторов киназ, регулирующих снижение LPS-интоксикации / М. А. Перфильев, П. М. Васильев, Д. А. Бабков [и др.] // Сборник научных трудов XXVII симпозиума «Биоинформатика и компьютерное конструирование лекарств». Институт биомедицинской химии имени В. Н. Ореховича. М., 2021. С. 65.
16. Консенсусная ансамблевая нейросетевая мультитаргетная модель RAGE-ингибирующей активности химических соединений / П. М. Васильев [и др.] // Биомедицинская химия. 2021. Т. 67. Вып. 3. С. 268–277.

Информация об авторах

Перфильев М. А. – аспирант кафедры фармакологии и биоинформатики ВолгГМУ, Волгоград, Россия; старший лаборант лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств Научного центра инновационных лекарственных средств, Волгоград, Россия

Васильев П. М. – доктор биологических наук, старший научный сотрудник НИИ фармакологии, заведующий лабораторией информационных технологий в фармакологии и компьютерного моделирования лекарств Научного центра инновационных лекарственных средств, Волгоград, Россия; профессор кафедры фармакологии и биоинформатики ВолгГМУ, Волгоград, Россия

Мальцев Д. В. – кандидат биологических наук, доцент кафедры фармакологии и биоинформатики ВолгГМУ, Волгоград, Россия; научный сотрудник лаборатории экспериментальной медицины ВМНЦ, Волгоград, Россия

Скрипка М. О. – ассистент кафедры фармакологии и биоинформатики ВолгГМУ, Волгоград, Россия; младший научный сотрудник лаборатории экспериментальной медицины ВМНЦ, Волгоград, Россия

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 10.02.2022; одобрена после рецензирования 14.02.2022; принята к публикации 18.02.2022