

«УТВЕРЖДАЮ»

Директор Федерального
государственного учреждения
«Федеральный исследовательский центр
«Информатика и управление»
Российской академии наук»,



М.А. Посыпкин

« 03 » июля 20 25 г.

ЗАКЛЮЧЕНИЕ

Федерального государственного учреждения «Федеральный исследовательский центр
«Информатика и управление» Российской академии наук»

Диссертация Ерещенко Алексея Владимировича на тему: «Применение графовых нейронных сетей для анализа молекулярных структур» выполнена в отделе №27 (математическое моделирование гетерогенных систем) Федерального исследовательского центра «Информатика и управление» Российской академии наук.

Ерещенко Алексей Владимирович 1994 года рождения, гражданин России, в 2021 году окончил ФГБОУ Московский авиационный институт (национальный исследовательский университет) по специальности / направлению 02.04.02 «Фундаментальная информатика и информационные технологии», направленность (профиль) образовательной программы «Машинное обучение и управление большими данными».

На текущий момент является аспирантом очной формы обучения Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» по научной специальности 1.2.1 — «Искусственный интеллект и машинное обучение», дата окончания 5 октября 2025 года.

Тема диссертационного исследования была утверждена на заседании ученого совета, протокол № 3 от 30.09.2024 года.

В период подготовки диссертации соискатель ученой степени Ерещенко Алексей Владимирович работал в Федеральном государственном унитарном предприятии «Всероссийский научно-исследовательский институт автоматики им. Н.Л. Духова» (ФГУП «ВНИИА»), структурное подразделение «Лаборатория медицинской химии и хемоинформатики (0171.Л.2)», период работы с 16.08.2021 по 12.10.2021 в должности младшего научного сотрудника. В настоящее время работает в Федеральном государственном унитарном предприятии «Всероссийский научно-исследовательский институт автоматики им. Н.Л. Духова» (ФГУП «ВНИИА»), структурное подразделение «Лаборатория медицинской химии и хемоинформатики (0171.Л.2)», в должности научного сотрудника.

Справка о периоде обучения с результатами сдачи кандидатских экзаменов выдана в 2025 г. в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук».

Научный руководитель — д.ф.-м.н., проф. Ревизников Дмитрий Леонидович — работает в ФГБОУ «Московский авиационный институт (национальный исследовательский университет)» в должности профессора кафедры вычислительной математики и программирования, в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» в должности ведущего научного сотрудника отдела №27 ФИЦ ИУ РАН по совместительству.

По итогам обсуждения принято следующее заключение.

Актуальность темы

Разработка новых лекарственных молекул является высокотехнологичным, многоэтапным и дорогостоящим процессом. Чтобы сократить время и финансовые расходы на ранних этапах разработки, обычно используются различные методы компьютерного моделирования. Это позволяет проанализировать, например, вероятную зону связывания белковой молекулы с лигандом, оценить свойства и возможные механизмы связывания подобной области, выбрать наиболее перспективные молекулы для виртуального скрининга, предсказать их фармакокинетические параметры и оценить другие важные свойства. Ключевым элементом работы современных вычислительных методов является анализ молекулярных структур, как белковых молекул, так и малых молекул. Методы машинного обучения становятся все более популярным инструментом для решения задач разработки лекарственных средств и виртуального анализа биологических соединений, однако их качество прямо зависит от качества и количества доступных данных. В этой связи существует практический интерес в решениях с применением компьютерных технологий, ассистирующих экспертам на различных стадиях разработки лекарств, для поиска сайтов связывания, выявления свойств сайта связывания, оценки свойств малых молекул. Графовые нейронные сети (GNN) являются перспективным инструментарием для решения задач данного класса, обладая большой гибкостью возможного для применения математического аппарата за счет реализации различных способов агрегации данных на вершинах.

Обоснованность научных положений

Обоснованность научных положений подтверждена экспериментальной проверкой результатов предлагаемых методов на реальных исследованиях в рецензируемых научных изданиях и конференциях по машинному обучению, экспертной валидацией результатов работы алгоритмов, а также практическим внедрением разработанных алгоритмов по месту работы соискателя.

Личное участие соискателя ученой степени в получении результатов, изложенных в диссертации

Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад автора в работу. Подготовка к публикации полученных результатов проводилась совместно с соавторами, причём вклад диссертанта был определяющим. Все представленные в диссертации результаты получены лично автором.

Степень достоверности результатов проведенных соискателем ученой степени исследований

Достоверность результатов подтверждена экспериментальной проверкой результатов, полученных с использованием предлагаемых методов и алгоритмов на представительных наборах данных.

Научная новизна работы

Разработан алгоритм с применением графовых нейронных сетей для разметки объемных сайтов связывания. Разработан алгоритм с применением графовых нейронных сетей для аннотации пространства сайта связывания белка. Продемонстрирована применимость оценок обученных аннотационных моделей для обучения нейросетей, решающих иные задачи. Применены графовые нейронные сети на основании трехмерного фармакофорного представления малой молекулы для предсказания наличия ингибирующей активности.

Теоретическая значимость

Теоретическая значимость работы заключается в разработке алгоритмов машинного обучения на основе графовых нейронных сетей для решения задач разметки сайтов связывания, аннотации пространства сайта связывания, предсказания ингибирующей активности.

Представлен алгоритм с применением графовых нейронных сетей и графового представления данных для работы с трехмерной структурой белка и пространства на его поверхности, с созданием, по результату работы алгоритма, объемного сайта связывания. Изложена разработанная соответствующая архитектура нейронной сети, проведено обучение

моделей и их тестирование. Рассмотрены сфера применения алгоритма, существующие аналоги. Изучено и продемонстрировано преимущество разработанного алгоритма по сравнению с аналогами.

Представлен алгоритм для аннотации сайта связывания белка с применением графовых нейронных сетей и графового представления трехмерной структуры белка. Изложена разработанная соответствующая архитектура, проведено обучение моделей и их тестирование. Рассмотрены сфера применения алгоритма, существующие аналоги. Изучено и продемонстрировано их преимущество по сравнению с аналогами.

Представлено применение оценок, сгенерированных обученными аннотационными нейросетями, для обучения других графовых нейронных сетей, решающих иные задачи; в данной работе продемонстрированы модели, классифицирующие сайты связывания, и модель, предсказывающая аффинность малых молекул. Экспериментально показано, что созданные подобным образом модели сравнимы или превосходят существующие аналоги.

Изложен подход по предсказанию вероятности наличия ингибирующей активности против заданной мишени (на примере 75 белковых мишеней) с применением графовой нейронной сети, обученной на ансамбле трехмерного фармакофорного представления малой молекулы, показано преимущество подхода по сравнению с иными подходами машинного обучения.

Констатируется применимость графовых нейронных сетей в рассмотренной предметной области, критически проанализированы конкретные практические сферы применения, а также преимущества представленных алгоритмов в сравнении с существующими решениями.

Практическая значимость результатов проведенных соискателем ученой степени исследований

Практическая значимость заключается в том, что модели аннотации и поиска сайтов связывания, а также созданные на основе моделей аннотации модели классификации сайтов связывания и предсказания аффинности малых молекулы были внедрены в Федеральном государственном унитарном предприятии «Всероссийский научно-исследовательский институт автоматики им. Н.Л. Духова» (ФГУП «ВНИИА»). В частности, результаты исследования были использованы для создания алгоритма и программного обеспечения по поиску лиганд-белковых сайтов связывания SiteRadar. По разработанному коду было зарегистрировано авторское право, по алгоритму SiteRadar была опубликована научная статья. Кроме того, результаты диссертационного исследования были применены для разработки алгоритма и программного обеспечения по разметке лиганд-белковых сайтов связывания. По данному алгоритму был получен патент на изобретение, была опубликована научная статья. Также результаты диссертационного исследования используются в разрабатываемой платформе по дизайну лекарственных соединений. Разработанная модель предсказания ингибирующей активности против заданных мишеней была применена для разметки открытой библиотеки из более чем 80 000 соединений, не имеющих известного профиля активности.

Апробация работы

11-я Московская конференция по вычислительной молекулярной биологии (MCCMB) 3-6 августа 2023 года, г. Москва, тема доклада: «SiteRadar: Utilizing Graph Machine Learning for Precise Mapping of Protein-Ligand-Binding Sites»

XIII International Conference on Chemistry for Young Scientists "MENDELEEV 2024", г. Санкт-Петербург, 2-6 сентября 2024 года, тема доклада: «KINASE SELECTIVITY PROFILING USING 3D PHARMACOPHORES AND GRAPH NEURAL NETWORKS»

XXX Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2024), онлайн-конференция, 16-18 сентября 2024, тема доклада: «LIGAND-PROTEIN BINDING SITE ANNOTATION USING GRAPH NEURAL NETWORKS»

Всероссийский форум молодых исследователей ХимБиоSeasons, г. Калининград, 14-18 апрель 2025 года, тема доклада: «Генерация псевдолигандов с помощью графовых нейронных сетей для решения задач структурной биологии»

Полнота изложения материалов диссертации в публикациях

Основные результаты диссертационного исследования опубликованы в 4 работах общим объемом 64 п.л.; личный вклад автора составляет 40 п.л.

Evteev, S. A., Ereshchenko, A. V., & Ivanenkov, Y. A. SiteRadar: Utilizing Graph Machine Learning for Precise Mapping of Protein-Ligand-Binding Sites // Journal of chemical information and modeling. 2023. 63(4), 1124–1132. DOI: 10.1021/acs.jcim.2c01413

Ivanenkov, Y., Evteev, S., Malyshev, A., Terentiev, V., Bezrukov, D., Ereshchenko, A., Korzhenevskaya, A., Zagribelnyy, B., Shegai, P., Kaprin, A. AlphaFold for a medicinal chemist: tool or toy? // Russ. Chem. Rev. 2024. 93 (3), RCR5107, DOI: 10.59761/RCR5107

Evteev, S., Ereshchenko, A., et al. Skittles: GNN-Assisted Pseudo-Ligands Generation and Its Application for Binding Sites Classification and Affinity Prediction // Proteins. 2025. DOI:10.1002/prot.26816

Ereshchenko A.V. Applying machine learning for solubility prediction: comparing different representations of molecular data // Modelling and Data Analysis. 2025. Vol. 15, no. 1, pp. 35–50. DOI: 10.17759/mda.2025150103

Публикации полностью соответствуют теме диссертационного исследования и раскрывают её основные положения.

Ценность научных работ соискателя ученой степени

Ценность научных работ соискателя заключается в том, что:

предложен алгоритм поиска сайтов связывания белковых молекул с применением графовых нейронных сетей и графового представления трехмерной структуры белка, продемонстрирована более высокая точность поиска сайтов связывания данного алгоритма по сравнению с известными алгоритмами, применяемыми в данной области;

предложен алгоритм аннотации сайтов связывания белковых молекул с применением графовых нейронных сетей и графового представления трехмерной структуры белка, продемонстрирована более высокая точность аннотации данного алгоритма по сравнению с распространенными алгоритмами, применяемыми в данной области;

продемонстрирована применимость оценок, генерируемых предложенным алгоритмом, в качестве признакового описания пространства, показано, что модели, обученные на подобном описании, демонстрируют сравнимую или превосходящую точность по сравнению с существующими аналогами;

предложены алгоритмы предсказания свойств малых молекул, а именно ингибирующую активность против тех или иных мишеней, а также растворимость, с применением графовых нейронных сетей и графового представления молекулы;

продемонстрировано преимущество графовой нейронной сети по сравнению с известными алгоритмами машинного обучения в задаче предсказания наличия ингибирующей активности против заданной мишени.

Диссертация Еreshchenko Алексея Владимировича на тему: «Применение графовых нейронных сетей для анализа молекулярных структур»

– это законченная научно-квалификационная работа, которая соответствует:

требованиям пунктов 9, 10, 14 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, а также Паспорту научной специальности 1.2.1 — «Искусственный интеллект и машинное обучение» в частности, по следующим пунктам:

4. Разработка методов, алгоритмов и создание систем искусственного интеллекта и машинного обучения для обработки и анализа текстов на естественном языке, для изображений, речи, биомедицины и других специальных видов данных.

5. Методы и технологии поиска, приобретения и использования знаний и закономерностей, в том числе — эмпирических, в системах искусственного интеллекта. Исследования в области совместного применения методов машинного обучения и классического математического моделирования. Методы и средства использования экспертных знаний.

7. Разработка специализированного математического, алгоритмического и

программного обеспечения систем искусственного интеллекта и машинного обучения. Методы и средства взаимодействия систем искусственного интеллекта с другими системами и человеком-оператором.

Диссертация Ерещенко Алексея Владимировича на тему: «Применение графовых нейронных сетей для анализа молекулярных структур» рекомендуется к защите на соискание ученой степени кандидата технических наук по специальности 1.2.1 — «Искусственный интеллект и машинное обучение»

Заключение принято на заседании отдела №27 (математическое моделирование гетерогенных систем) ФИЦ ИУ РАН «01» июля 2025 г., протокол № 1.

Присутствовало на заседании 15 человек.

Результаты голосования: «за» – 15 человек, «против» – 0 человек, «воздержалось» – 0 человек.

Председательствующий на заседании:

Старший научный сотрудник ФИЦ ИУ РАН,
доктор физико-математических наук

Секретарь заседания:

Научный сотрудник ФИЦ ИУ РАН,
кандидат физико-математических наук



Морозов А.Ю.

Колбин И.С.