

## ОТЗЫВ

официального оппонента на диссертацию

Ерещенко Алексея Владимировича

«Применение графовых нейронных сетей для анализа молекулярных структур»,  
представленную на соискание учёной степени кандидата технических наук по  
специальности 1.2.1 «Искусственный интеллект и машинное обучение»

### **Актуальность темы исследования**

Создание лекарственных средств является комплексным и дорогостоящим процессом. Методы компьютерного моделирования всё больше внедряются в разработку лекарственных молекул, позволяя сократить время и стоимость работ, особенно на ранних стадиях разработки. В частности, методы машинного обучения находят все большее применение в данной сфере, в том числе за счёт растущего объема доступных данных для обучения моделей. Отдельный интерес представляют модели, способные работать с 3D структурами белков в рамках задач структурно-ориентированного дизайна лекарств, а также с 3D представлениями малых молекул.

Графовые нейронные сети представляют собой перспективный класс нейросетевых архитектур для работы с молекулярными структурами, которые часто представлены именно в виде графов. За счёт гибкости, данная категория нейросетей способна учитывать структуру, связность, химический состав молекулы, а также её физико-химические дескрипторы и иные свойства. Разработанные в ходе данной работы алгоритмы основаны на данной архитектуре нейросетей.

Исследуемые в данной работе практические области применения представленных алгоритмов, а именно поиск сайтов связывания, оценка их биохимической активности, а также предсказание фармацевтических свойств малых молекул и первичный анализ их возможных мишеней, обладают большим потенциалом для внедрения алгоритмов с применением графовых нейронных сетей.

### **Научная новизна**

В диссертационной работе были получены следующие научные результаты.

1. Реализован новый многоэтапный алгоритм поиска объёмных сайтов связывания на основании графовых нейронных сетей, использующий алгоритмическое формирование пространства на поверхности целевой трёхмерной структуры белковой молекулы, классификацию данного пространства с использованием разработанной

графовой нейронной сети, кластеризацию размеченного пространства, а также ранжирование сформированных объектов с помощью оценок нейросетевых классификаторов. Алгоритм продемонстрировал более высокую точность по сравнению с другими современными подходами.

2. Предложен новый подход на основании графовых нейронных сетей для аннотации трёхмерного пространства поверхности белковой молекулы с использованием экспертно разработанных функциональных групп и реализованных нейросетевых классификаторов. Продемонстрировано преимущество данного подхода по сравнению с иными методами. Показано, что проведённая подобным образом разметка пространства может быть использована в качестве признакового описания другими моделями машинного обучения на основании графовых нейронных сетей, в частности для решения задач классификации сайтов связывания и предсказания аффинности малых молекул. Экспериментально продемонстрировано качество подобных моделей, которое сравнимо или превосходит аналоги.

3. Реализован новый алгоритм по предсказанию ингибирующей активности против заданной мишени с применением разработанной графовой нейронной сети и экспертного метода трёхмерного моделирования конформаций малых молекул. Экспериментально продемонстрировано преимущество алгоритма по сравнению с иными подходами машинного обучения.

### **Теоретическая значимость**

Теоретическая значимость данной работы заключается в проведённом исследовании применимости графового описания трёхмерных структур в области медицинской химии и разработки лекарственных средств. Кроме того, изучена применимость оценок, генерируемых разработанными в ходе диссертационной работы графовыми нейронными сетями аннотации трёхмерного пространства на поверхности белковой молекулы, для дополнения либо замещения признакового описания в целях обучения иных графовых нейросетевых моделей, решающих задачи, требующих знания о биохимических свойствах рассматриваемого пространства.

### **Практическая значимость**

Практическая значимость данной работы заключается в успешном внедрении разработанных моделей и алгоритмов в производственное применение. В частности, насколько я знаю, алгоритм поиска сайтов связывания белковых молекул, алгоритм аннотации пространства сайтов связывания, модели классификации сайтов связывания и

предсказания аффинности малых молекул были внедрены в разрабатываемой программной платформе во ФГУП "ВНИИА". Помимо этого, реализованный алгоритм предсказания ингибирующей активности против заданных мишеней был применен для разметки открытой библиотеки из более чем 80 000 соединений, не имеющих известного профиля активности. Разработанные нейросетевые модели и алгоритмы с их использованием имеют широкий спектр применения в сфере компьютеризированной разработки лекарственных средств.

### **Достоверность и обоснованность результатов**

Достоверность и обоснованность результатов подтверждается непротиворечивостью полученных результатов исследованиям и известным фактам в рассматриваемой области, валидацией алгоритмов экспертами предметной области на представительных наборах данных, апробацией на научных конференциях, публикациями результатов в четырех рецензируемых тематических научных изданиях, из которых три входят в Q1 Scopus, а также успешным внедрением разработанных алгоритмов в платформу для создания лекарственных молекул.

### **Содержание работы**

Диссертационная работа состоит из Введения, трёх глав, Заключения и Списка литературы.

Во Введении обоснованы актуальность темы исследования, научная новизна и практическая значимость полученных результатов.

В первой главе рассматривается задача поиска сайтов связывания на поверхности трёхмерных белковых структур. Проводится обзор данной области, описывается предлагаемая методология решения задачи. Описывается подготовка обучающих данных, разработка графовой нейронной сети, демонстрируется реализованный алгоритм. Проводится сравнение созданного алгоритма с существующими решениями, применяемыми для данного рода задач, на экспертно подобранной независимой выборке данных.

Во второй главе рассматривается задача разметки пространства связывания белка на основании его трёхмерной структуры, в соответствии с экспертно разработанными типами. Проводится обзор предметной области и существующих решений. Описывается разработанный алгоритм, подготовка обучающих данных. Проводится валидация алгоритма на независимой выборке данных, а также его сравнение с существующими методами. Демонстрируется применение разметки пространства, проводимой подобным

образом, для обучения других моделей машинного обучения, на примере графовых нейронных сетей для классификации сайтов связывания и предсказания аффинности. Разработанные модели также сравниваются с аналогами.

В третьей главе рассматривается задача предсказания свойств малых молекул без использования информации о белковом окружении. Описываются разработанные в ходе диссертационной работы алгоритмы предсказания наличия ингибирующей активности против заданных мишеней, а также предсказания растворимости. Приводится сравнение созданного подхода на основе архитектуры графовых нейронных сетей с другими методами машинного обучения.

**Замечания по диссертации** носят преимущественно рекомендательный характер и не влияют на общую положительную оценку работы.

1. В главе 1 ограничением разработанного алгоритма поиска сайтов связывания указана скорость расчетов. Было бы целесообразно привести сравнение скорости работы алгоритма с рассмотренными в работе существующими аналогами.
2. В главе 2 сравнение точности модели классификации сайтов связывания с существующими аналогами было проведено лишь по некоторым из предсказываемых классов. Проведение сравнительного анализа точности работы алгоритма в рамках всех рассматриваемых классов позволило бы более полно оценить разработанный метод.
3. В главе 3 в рамках исследования предсказания растворимости малой молекулы с применением графовых нейронных сетей, в ходе проведенного сравнения с существующими подходами, не был рассмотрен один из популярных методов, а именно использование глубокой нейронной сети, обученной на табличном представлении данных в виде крупного набора физико-химических дескрипторов. Принимая во внимание то, что данный метод был рассмотрен в ходе сравнительного анализа, проведенного в рамках решения задачи предсказания наличия ингибирующей активности, было бы целесообразно применить его и к задаче предсказания растворимости.

### **Заключение**

Диссертационная работа «Применение графовых нейронных сетей для анализа молекулярных структур» Ерещенко Алексея Владимировича является завершенной научно-квалификационной работой, посвященной актуальному научному направлению.

