

УДК 54+004.27

## КОМПЬЮТЕРНАЯ РАЗРАБОТКА ЛЕКАРСТВ: ОТ ТРАДИЦИОННЫХ МЕТОДОВ МОДЕЛИРОВАНИЯ К ЯЗЫКОВЫМ МОДЕЛЯМ И КВАНТОВЫМ ВЫЧИСЛЕНИЯМ

<sup>+</sup> Адилова Ф.Т.<sup>1</sup>, Давронов Р.Р.<sup>1</sup>

<sup>1</sup> Институт математики им. В.И. Романовского АН Республики Узбекистан, Ташкент, Узбекистан

<sup>+</sup> fatadilova@matinst.uz

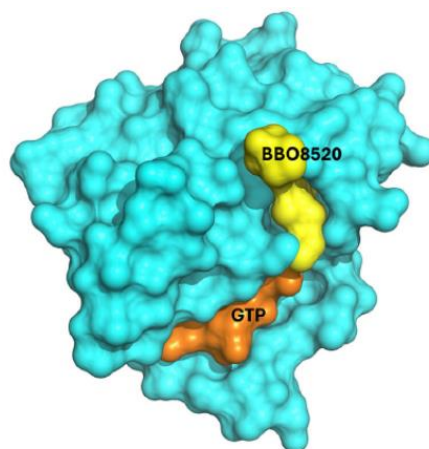
**Аннотация.** Разработка лекарств - центральная тема на стыке структурной биологии, биохимии и медицины, связанная со значительными проблемами, такими как высокая стоимость (миллиарды долларов), низкие показатели успеха (менее 10%) и чрезвычайно длительные циклы разработки (10-15 лет). Система автоматизированного поиска лекарств (Computer Aided Drug Design, CADD) демонстрирует огромные преимущества в решении этих задач и ускорении процесса, что делает ее незаменимым инструментом в фармацевтической промышленности и научных исследованиях. Недавняя разработка AlphaFold2 и AlphaFold3, - лауреатов Нобелевской премии 2024г., знаменует собой значительный прогресс в области CADD. Помимо AlphaFold, различные методы машинного обучения (ML) революционизируют различные этапы разработки лекарств - от виртуального скрининга до прогнозирующего моделирования взаимодействий между лекарством и таргетами лечения. Языковые модели, такие как GPT-модели, предлагают многообещающие приложения для выработки исследовательских гипотез и помощи в интерпретации сложных биологических данных. Квантовые вычисления обладают потенциалом для решения сложных задач молекулярного моделирования и оптимизации, которые в настоящее время неразрешимы для классических компьютеров, хотя их практическая реализация все еще находится на ранних стадиях. В данном аналитическом обзоре представлены новейшие достижения в этой области и оценены возможности, представляемые машинным обучением, языковыми моделями и квантовыми вычислениями в CADD.

**Ключевые слова:** компьютерная разработка лекарств, машинное обучение, глубокое обучение, квантовые вычисления, языковые модели, количественный анализ взаимосвязи структура-активность, поиск лекарств, молекулярное моделирование, предсказание структуры белка, виртуальный скрининг, AlphaFold2, AlphaFold3, гибридные квантово-классические алгоритмы, молекулярный докинг, неполные данные, молекулярный эмбединг, отпечатки Моргана.

### 1 ВВЕДЕНИЕ

В последние годы разработка современных лекарственных средств становится все более сложной задачей из-за лекарственной устойчивости, неудач клинических испытаний, необходимости комбинированной терапии и т.д. [1]. Сложность лечения многофакторных заболеваний, например, рака, подчеркивает настоятельную необходимость инновационных подходов в лечении и разработке лекарственных средств. Эти факторы делают традиционные методы проб и ошибок, основанные исключительно на экспериментах, все более непрактичными. Компьютерный виртуальный скрининг - высокоэффективный метод на ранних стадиях разработки лекарств значительно сокращает время и затраты. Общий процесс включает оптимизацию трехмерных структур белков-мишеней и низкомолекулярных лигандов, определение активных участков для стыковки и использование надежных функций оценки качества стыковки. Расширяющиеся базы данных о белковых структурах (рецепторах) и малых молекулах (лигандах) в сочетании с более точными методами стыковки и эффективными высокопроизводительными системами стыковки рецептор-лиганд [2] произвели революцию в этой области. В частности, методы стыковки на основе машинного обучения и предсказания структуры белка, подобные тем, которые были продемонстрированы в AlphaFold2 и 3, значительно повысили точность и эффективность этих процессов [3].

Одним из недавних заслуживающих внимания достижений является испытание на людях препарата от рака ВВ0-85207, разработанного Livermore Laboratory и BridgeBio,[4] которое получено с использованием высокопроизводительных вычислений (рис. 1). Обеспечивая быстрый анализ сложных биологических данных и моделирование молекулярных взаимодействий в беспрецедентных масштабах, суперкомпьютеры могут значительно сократить сроки от открытия до клинических испытаний.



**Рис. 1.** Компьютерное открытие препарата-кандидата от рака ВВ-8520 [4]. Наглядно показано взаимодействие между ВВ-8520 (желтый), природным субстратом гуанозинтрифосфатом (GTP; оранжевый) и белком KRAS (голубой)

Yang and Zhang в своем обзоре за 2022 г. подробно описали интеграцию методов машинного обучения, которые повысили точность и эффективность выявления перспективных препаратов-кандидатов [5]. В последнее время внимание привлекают подходы к глубокому обучению для прогнозирования структуры белка, которые произвели революцию в понимании сворачивания белка, достигнув беспрецедентной точности. Модификации AlphaFold2, такие как ParaFold, [6] AlphaFold-Multimer v.2.3 [7] и AlphaFold3 [8], уточнили эти прогнозы для решения конкретных задач в области структурной биологии. Методы ML способствовали быстрому созданию противовирусных препаратов во время пандемии коронавируса [9], и выявили потенциальные методы лечения с помощью компьютерного скрининга и прогнозирующего моделирования. Но сегодня появились новые технологии, языковые модели, (Natural Language Processing, NLP) и квантовые вычисления (Quantum Computing, QC), которые являются предметом данного аналитического обзора.

Обзор состоит из описания методов машинного и глубокого обучения в QSAR, методологии языковых моделей и квантовых вычислений в разработке лекарств. В заключительной части обзор приводятся доказательные результаты преимущества квантовых вычислений в области QSAR.

## 2 МЕТОДОЛОГИЯ

### 2.1 Машинное и глубокое обучение в QSAR

Количественный анализ взаимосвязи структура-активность (QSAR) является ключевым вычислительным инструментом, устанавливающим количественные отношения между химическими структурами и соответствующими им биологическими активностями. В своё время QSAR произвела революцию в области создания лекарств, определяя разработку эффективных лекарственных средств и значительно сокращая время и затраты, связанные с традиционными методами проб и ошибок. Перечислим самые последние модели и алгоритмы машинного обучения в QSAR [10-18]. Отметим работу Kim et al. [19], где предложили и оценили расширенную модель Profile-QSAR (pQSAR) 2.0, которая объединяет регрессоры случайного леса (RFR) и частичные регрессоры наименьших квадратов (PLSR) для улучшения прогнозирующих характеристик регрессионных моделей QSAR. Общий вклад этих исследований демонстрирует постоянный прогресс и диверсификацию методов машинного и глубокого обучения в моделировании QSAR. Однако, остаётся ключевое структурное ограничение этих алгоритмов, а именно, последовательное применение линейных фильтров к локальным входным участкам. Стандартные подходы QSAR, основанные на молекулярных отпечатках, строках SMILES, графовых нейронных сетях и графовых трансформерах, не выполняют соответствующую операцию фильтрации на молекулярных графах. Именно отсутствием

последовательных локальных фильтров можно объяснить относительно низкую производительность даже мощных алгоритмов глубокого обучения в задачах QSAR. Для преодоления этого ограничения были предложены языковые модели и квантовые вычисления, которые являются сегодня передовыми технологиями в области искусственного интеллекта (ИИ).

## 2.2 Языковые модели в разработке лекарств

Генеративные модели, основанные на языковых структурах, используются для оптимизации последовательности белков и создания соединений, что делает их полезными инструментами для поиска лекарств. Например, модель языка белков ProGen2 была успешной в создании новых белковых последовательностей, имитирующих естественный процесс сворачивания [20]. ProtGPT2 может предсказывать структуру белка и улучшать функциональные свойства за счет создания оптимизированных последовательностей, что даёт возможность создания нового белкового дизайна, имеющего важное значение при разработке лекарств [21]. Многочисленные модели NLP были применены для решения биохимических задач, от прогнозирования химических реакций до понимания структуры белков и поддержки в разработке лекарств. Unsal et.al. проанализировали последние достижения в этой области [22]. Большие языковые модели общего назначения (large Language Model, LLM), такие как GPT3.5 и 4.0 (<https://openai.com/blog/chatgpt>) и Llama2,3 с открытым исходным кодом продемонстрировали большие возможности в различных областях [23, 24]. GPT 4.0, например, может служить виртуальным руководством при поиске лекарств, генерируя инновационные идеи и разрабатывая стратегии открытия новых лекарств после анализа важной информации из научной литературы. Lu and Zhang разработали T5Chem, - языковую модель глубокого обучения, основанную на платформе T5 (text-to-text transfer transformer, T5), которая предназначена для многозадачного прогнозирования реакций [25].

## 2.3 Квантовые вычисления для создания лекарств

В квантовых вычислениях используются квантовые схемы, которые оперируют квантовыми битами (кубитами) в качестве физических объектов для вычислений. Одной из ключевых особенностей квантовых компьютеров является их способность создавать квантовую запутанность между кубитами для описания очень сложных отношений и или закономерностей, скрытых в химических данных. Кратко опишем эти ключевые понятия квантовых вычислений (см. рис. 2А).

Кубит является квантовым аналогом классического бита. Классический бит имеет два состояния, то есть 0 и 1, в то время как кубит имеет непрерывные состояния, которые могут быть представлены сферой Блоха [26]. Экспоненциальная мощь кубитного представления является одним из источников квантовых преимуществ. Состояния кубитов контролируются и настраиваются квантовыми элементами, такими как элементы Адамара и Паули (Hadamard и Pauli), которые являются однокубитными элементами, поскольку в каждом из них задействован только один кубит. Квантовый вентиль обычно имеет один или несколько параметров, позволяющих точно изменять состояние кубита.

Основным элементом квантовых вычислений на прикладном уровне является квантовая схема. Квантовая схема состоит из кубитов и вентилях, которые могут реализовывать различные цели квантовой обработки данных. Учитывая вероятностную природу квантовых вычислений, необходимо выполнить множество вычислений или снимков одной и той же квантовой схемы с одинаковыми параметрами для измерения выходных данных до тех пор, пока не будет получено стабильное и надежное распределение вероятностей.

QSAR может быть предсказан с помощью квантового машинного обучения на основе параметризованных квантовых схем (Parameterized Quantum Circuit, PQC) [27]. Доступные параметры в этой квантовой схеме управляют углами поворота квантовых вентилях и обновляются путем минимизации функции затрат, рассчитанной классическим способом. С PQC можно создать гибридную квантово-классическую сеть. Известны несколько применений [27, 28], включая генерацию рисунков [29, 30], регрессию данных [31] и классификацию [32,33].

Благодаря большему гильбертову пространству, квантовые схемы часто превосходят классические модели - явление, известное как квантовое преимущество. Большинство исследований показывают, что гибридная модель отличается меньшей сложностью, более высокой эффективностью обучения [34, 35], большей выразительностью [36], устойчивостью к шуму и лучшей общей производительностью [35]. Квантовые преимущества гибридной модели также были обнаружены в исследовании [37], где лиганды, сгенерированные с помощью квантовых генеративных состязательных сетей (GAN) [38], демонстрируют лучшее сходство с лекарствами по сравнению с лигандами, сгенерированными с использованием классических GAN. При прогнозировании QSAR гибридная

квантово-классическая сверточная нейронная сеть демонстрирует сокращение времени процесса обучения на 40% при сохранении того же уровня производительности [39].

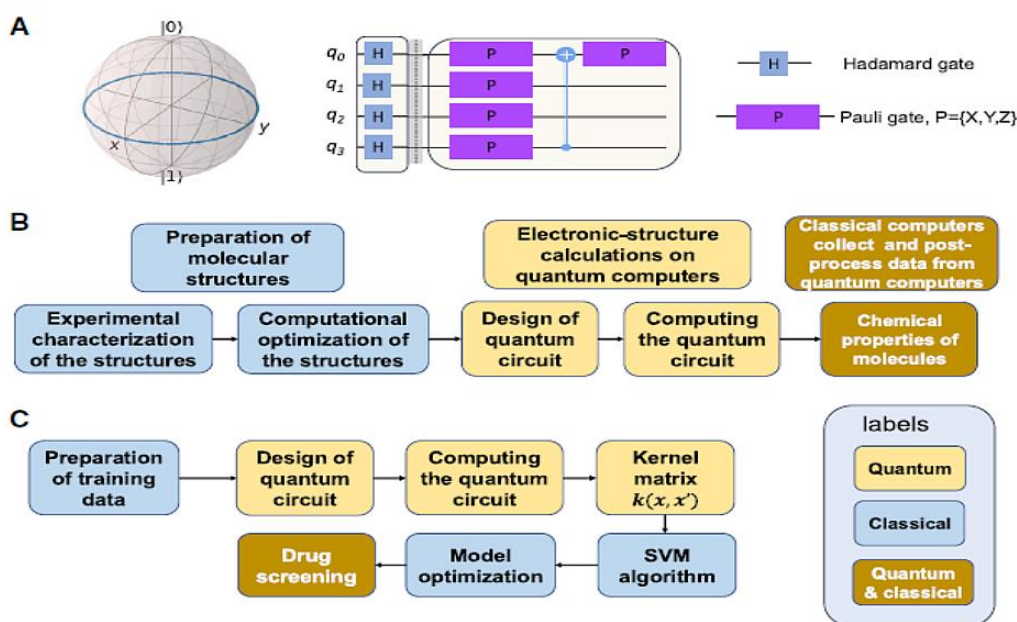
Однако данные низкого качества и ограниченной доступности в медицинских приложениях неизменно являются наиболее частыми и серьезными проблемами для машинного обучения [40]. При использовании различных методов эмбединга к данным и отбора признаков с помощью главных компонент (PCA), отмечается, что квантовый классификатор превосходит свой классический аналог, особенно когда учитывается всего несколько признаков [41].

## 2.4 Квантовая модель QSAR при неполных данных

Низкое качество и ограниченная доступность данных до сих пор являются наиболее распространенными и критичными проблемами для приложений машинного и глубокого обучения в QSAR. В этом разделе обзора обсудим эффективность классического и квантового классификаторов в предсказании QSAR и преимущества квантового классификатора в обобщающей способности в условиях ограниченной доступности данных и ограниченного числа признаков.

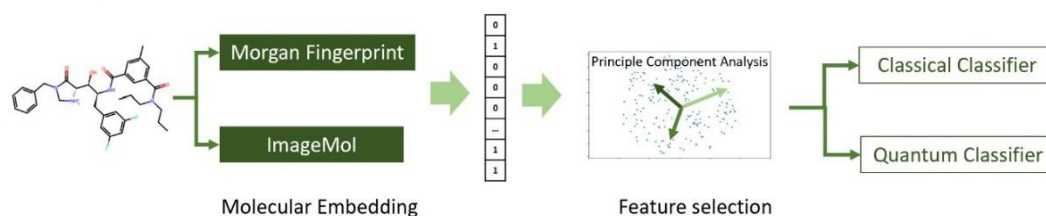
## 3 МАТЕРИАЛЫ И МЕТОДЫ

На рис.3 показан рабочий процесс. В предлагаемой в [42] квантовой модели QSAR входные молекулы сначала кодируются с помощью отпечатков Моргана [43] и эмбединга ImageMol [44]. Затем методом главных компонент (PCA) уменьшается размерность и имитируется неполнота данных. В этих экспериментах для классификации используются  $2^n$  признаков, где  $n$  - количество кубитов. Затем сокращенный эмбединг передается в классические или квантовые классификаторы для сравнения их производительности.



**Рис. 2.** Квантовые вычисления для разработки лекарств (А). Слева направо основными понятиями квантовых вычислений являются сфера Блоха [26], которая представляет все возможные состояния кубита, квантовой схемы и типичных квантовых элементов (элемент Адамара и элемент Паули). (В) Рабочий процесс молекулярного моделирования или стыковки молекул в квантовых вычислениях. Это может повысить точность расчетов аффинности связывания. (С) Пример рабочего процесса для квантового машинного обучения. Здесь квантовая схема может использоваться в качестве ядра в составе гибридного классического и квантового алгоритмов. Он также может быть реализован независимо, например, в квантовых нейронных сетях [42]

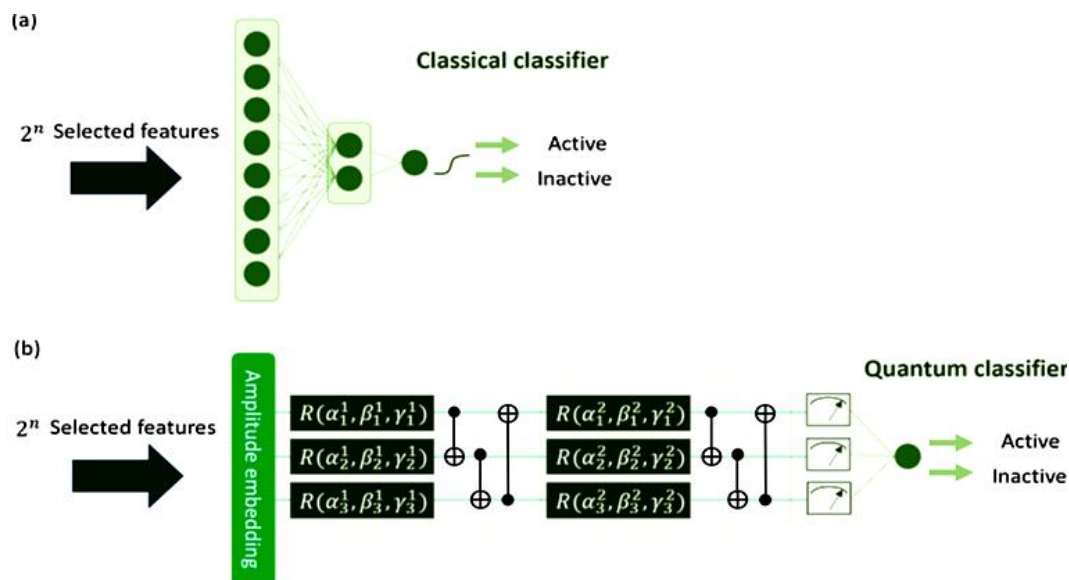
Применялись 2 различных типа молекулярного эмбединга: молекулярный эмбединг, основанный на локальной структуре [43], и молекулярный эмбединг, основанный на изображении [44]. В данной работе для извлечения отпечатка Моргана для каждой молекулы авторы используют RDKit [45] с радиусом по умолчанию и 512 битами.



**Рис. 3.** Workflow: структура ингибитора (SMILES) внедряется с помощью Morgan fingerprint или Image Mol с последующим выбором признаков.  $2^n$  выбранных признаков заносятся в классическую и квантовую классификации.  $n$  - количество кубитов в квантовом классификаторе [42]

Фреймворк Image Mol сочетает в себе методы обработки изображений с обширными знаниями в области молекулярной химии для извлечения мельчайших молекулярных особенностей на уровне пикселей с помощью визуального компьютерного подхода. ImageMol использует молекулярные изображения, т.е. соединения представлены в виде изображений, что помогает достичь более высокой точности прогнозов при относительно меньших вычислительных затратах. ImageMol использует предобучение, извлекая структурную информацию из набора данных, содержащего 10 миллионов соединений, и охватывающего различные биологические активности, что позволяет использовать фреймворк для решения различных биологических задач по созданию лекарств. <https://hub.docker.com/r/pykao/imagemol>.

В качестве классического классификатора рассмотрена архитектура трехслойного персептрона размером  $N \times 2 \times 1$ , показанная на рис.4(a), а число обучаемых параметров равно  $2(N + 1)$ , где  $N$  - число признаков. После измерения всех кубитов функция регрессии делает бинарный прогноз с нулевым порогом. Количество обучаемых параметров равно  $7n$ , 80% и 20% всех данных задаются для данных обучения и тестирования соответственно. Точность классификаторов определяется путем: (i) определения наилучшей производительности каждого обучения за 100 эпох, (ii) вычисления среднего значения наилучшей точности за 20-кратное обучение, (iii) повторной выборки обучающего и тестового набора с последующим (i) и (ii) повторением 5 раз и вычислением среднего значения.



**Рис. 4.** (a) Классический многослойный персептрон (MLP) имеет входной слой с  $2^n$  нейронами, скрытый слой с двумя узлами и выходной узел. (b) Квантовый классификатор состоит из слоя амплитудного эмбединга для кодирования входных данных размерностью  $2^n$ , двух сильно переплетающихся слоев, за которыми следуют измерения для всех кубитов. На выходе результаты измерений объединяются в выходные данные [42]

## 4 РЕЗУЛЬТАТЫ

Вычислительные эксперименты в данном исследовании должны были ответить на вопрос: как влияет количество признаков и методы эмбединга на классические и квантовые классификаторы? Преимуществом квантового классификатора считается его надежность. Оказалось, что в режиме

малоразмерных признаков производительность квантового классификатора снижается при сокращении объема обучающих данных.

Но каково универсальное квантовое преимущество для различных наборов данных? Квантовые классификаторы превосходят классические, если количество выбранных признаков невелико, и при этом для оценки характеристик моделей классических и квантовых классификаторов используется точность.

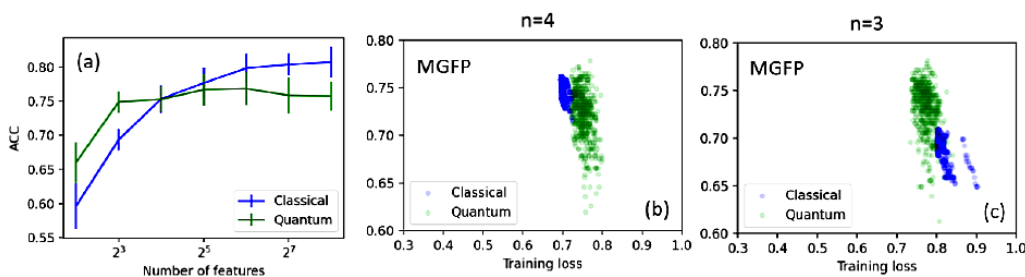
Для тестирования производительности классификаторов отобраны  $2n$  лучших признаков. Классический классификатор работает лучше, чем квантовый при огромном числе признаков. Однако, если число выбранных признаков небольшое, квантовый классификатор превосходит классический. Из табл.1 можно видеть, что классический классификатор работает лучше на 5%, когда данные, внедренные MGFP, содержат  $2^n$  ( $n = 8$ ) признаков. С другой стороны, квантовый классификатор превосходит классический при  $n = 3$ , что приводит к заметному повышению точности на 8%.

**Таблица 1.** Сравнение характеристик классического и квантового классификаторов для данных с различным количеством признаков и двух методов эмбединга на наборе данных VACE

n	MGFP		IMG MOL	
	Classical	Quantum	Classical	Quantum
2	0.60 ±0.03	0.66 ±0.03	0.60 ±0.03	0.62 ±0.02
3	0.69 ±0.02	0.75 ±0.02	0.65 ±0.04	0.70 ±0.02
4	0.75 ±0.02	0.75 ±0.02	0.66 ±0.02	0.72 ±0.02
8	0.80 ±0.02	0.76 ±0.02	0.80 ±0.02	0.74 ±0.02

Из таблицы 1 следует, что квантовое преимущество, обнаруживаемое в низкоразмерном классификаторе, проявляется в различных методах эмбединга. Также видно, что производительность квантового классификатора выше, когда в обоих методах эмбединга выбрано небольшое количество признаков. Однако, осталось неясным, какой метод эмбединга подходит для конкретного классификатора. Поэтому показаны результаты классификации объектов, - по эмбедингу MGFP на наборе данных VACE.

Чтобы получить более глубокое представление об этом квантовом преимуществе, систематически сравнивается производительность классических и квантовых классификаторов, используя признаки MGFP. Точность классических и квантовых классификаторов для различных значений характеристик представлена на рис. 5(a).



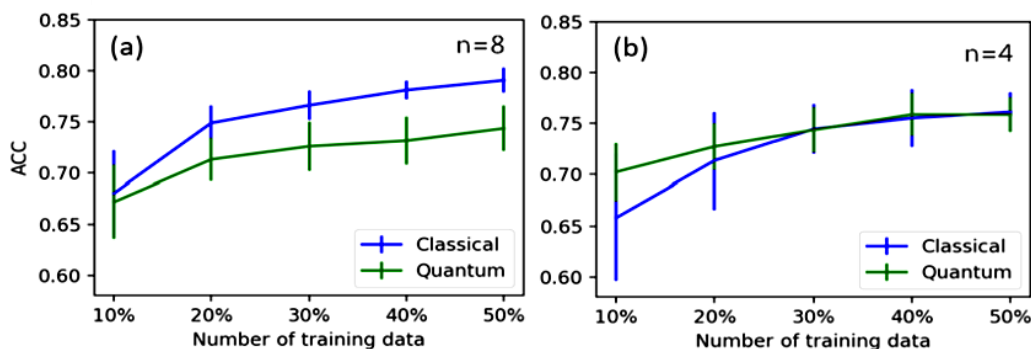
**Рис. 5.** (а) Тенденция к повышению точности при изменении количества признаков. Изменение точности сильно зависит от количества признаков в классическом классификаторе. (б) и (в) показывают взаимосвязь между потерями при обучении и точностью тестирования для  $n = 4$  и  $n = 3$  соответственно. Потери при обучении и точности тестирования показывают, что квантовый классификатор обладает большей способностью к обобщению по сравнению с классическим классификатором [42]

На рис. 5(б) видно, что классические классификаторы демонстрируют меньшие потери при обучении, но сравнимую точность при тестировании с квантовыми классификаторами. Оказалось, что квантовому классификатору требуется всего 82% обучаемых параметров классического классификатора для достижения аналогичной точности. Более того, при  $n = 3$  (рис. 5(с)), классические и квантовые классификаторы показывают сходные потери при обучении, но более высокую точность на тестировании в квантовых классификаторах. Это говорит о том, что квантовые классификаторы обладают большей способностью к обобщению при малоразмерных признаках.

Возникает вопрос про объем обучающих выборок, которых в разработке лекарств обычно не хватает. Ответ, полученный авторами, говорит в пользу квантовых компьютеров.

Чтобы изучить влияние объема обучающих данных на производительность, исходные обучающие данные разделили на подгруппы в 10%, 20%, 30%, 40%, и 50% от исходного объема, и подготовили классификаторы. На рис. 6 показано, что производительность обоих классификаторов улучшается по мере увеличения объема обучающих выборок. Хотя этот параметр существенно влияет

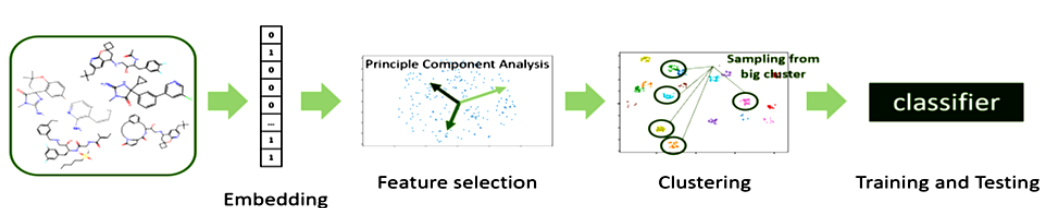
на точность, классический классификатор неизменно превосходит квантовый при работе с объектами большой размерности (рис.6(a)).



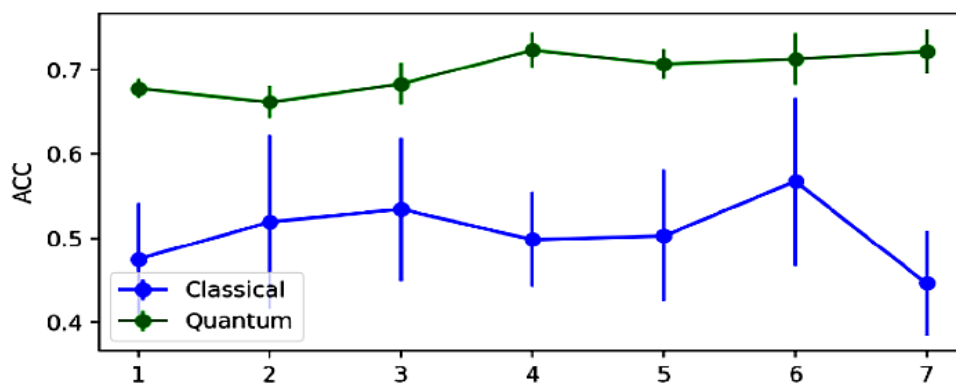
**Рис. 6.** (а) и (б) представлены зависимости точности от количества обучающих данных при  $n = 8$  и  $n = 4$  соответственно. На рисунке показано, что при  $n = 4$  квантовое преимущество достигается при небольшом количестве режимов обработки данных, в то время как классический классификатор по-прежнему превосходит квантовый при  $n = 8$  [42]

Однако эта зависимость приводит к снижению точности классического классификатора при работе с данными, представленными малоразмерными признаками, в режиме небольших обучающих данных (рис. 6(б)).

Исследовалась проблема масштабируемости классических/квантовых классификаторов в режимах малоразмерных признаков. Можно точно контролировать свойства обучающих данных и гарантировать, что модель видит похожие признаки, выбрав обучающую выборку из кластеризованных данных. Обучающие данные группируются, после чего из кластера, размер которого превышает 20, извлекается обучающая выборка. Количество обучающих выборок из разных кластеров варьируется от 1 до 7 (рис.7). Из рис.8 видно, что точность тестирования квантового классификатора выше даже при небольшом количестве обучающих данных, что подразумевает лучшую способность квантового классификатора к обобщению для режима малоразмерных признаков.

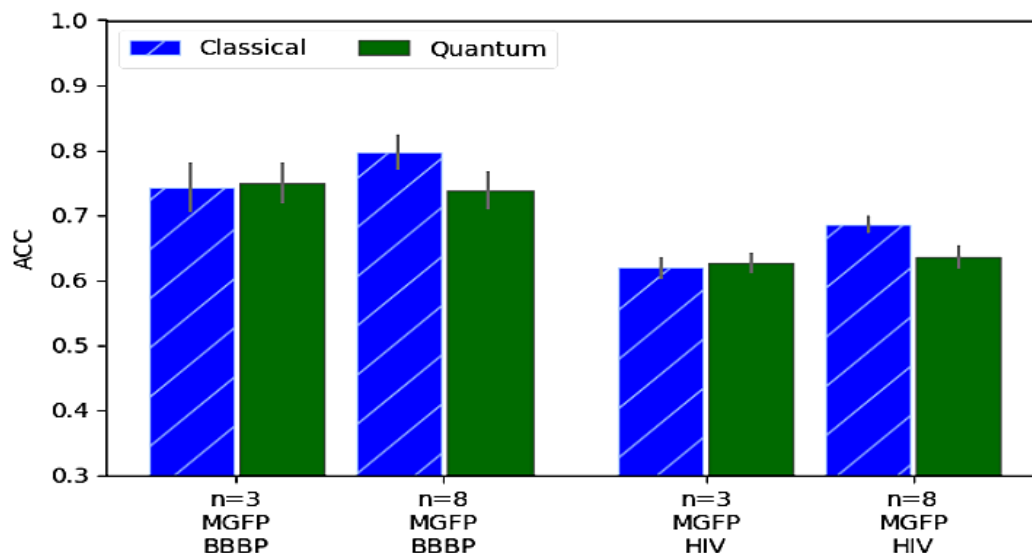


**Рис. 7.** Чтобы проверить обобщающую способность классических и квантовых моделей, была проведена кластеризация 29 молекул после MGFP эмбединга, а в качестве обучающих данных были выбраны от 1 до 7 точек данных из каждого более крупного кластера

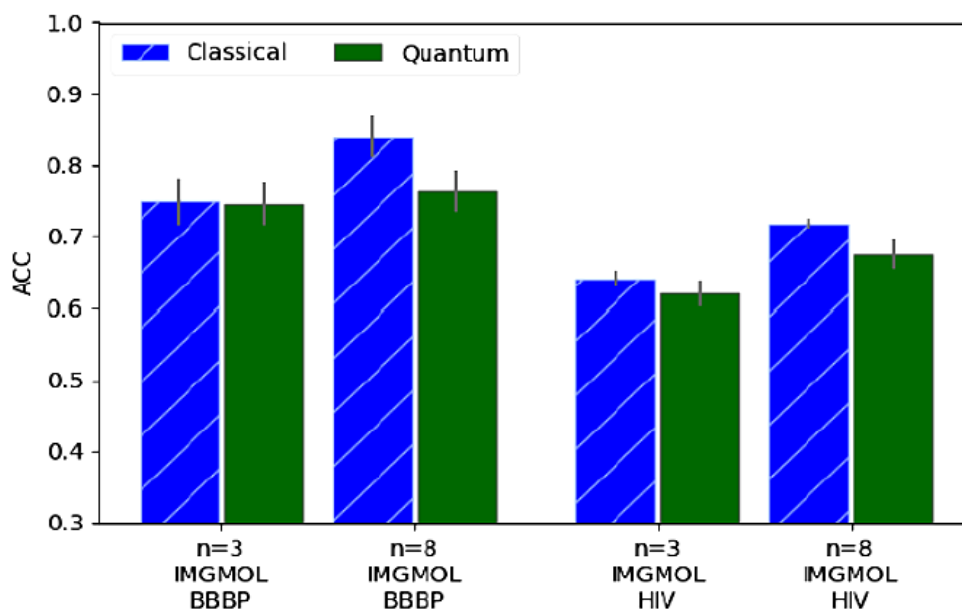


**Рис. 8.** Сравнение точности тестирования, когда для обучения используется разное количество выборок, отобранных из каждого кластера. В сценарии с малоразмерными признаками,  $n = 3$ , точность тестирования в квантовом классификаторе значительно выше

Авторы протестировали универсальность этого существенного преимущества на различных наборах открытых данных, таких как HIV и BBVP. Чтобы устранить проблемы с дисбалансом данных, провели недостаточную выборку данных перед процедурой классификации. Как показано на рис. 9, классические классификаторы демонстрируют лучшую производительность в обоих наборах данных при  $n = 8$ , внедренных с помощью MGFP. Однако, когда данные внедряются с использованием IMG MOL, квантовые преимущества становятся менее очевидными из-за существенного улучшения производительности классических классификаторов (рис. 10). Из рис. 9 и рис. 10 видно, что производительность классических классификаторов зависит от методов эмбединга и количества признаков.



**Рис. 9.** Характеристики классического и квантового классификаторов: гистограмма показывает точность прогнозирования при внедрении MGFP как для  $n = 3$ , так и для  $n = 8$ . Из рисунка видно, что квантовый классификатор работает немного лучше, чем классический, когда  $n$  мало. ((Наборы данных о ГЭБ и ВИЧ внедряются с помощью MGFP) [42])



**Рис. 10.** Характеристики классических/квантовых классификаторов. Гистограмма точности прогнозирования при внедрении IMG MOL при  $n = 3$  и  $n = 8$ . Данные, внедренные с помощью IMG MOL, улучшают характеристики не только классических, но и квантовых классификаторов. Это улучшение сводит на нет едва заметные квантовые преимущества, показанные на рисунке 9[42]

## 5 ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

### 6.1. Status-Quo квантовых моделей

К числу важных относится заявление Google о достижении квантового превосходства с помощью своего 53-кубитного квантового компьютера [46]. После создания Питером Шором новаторского алгоритма разложения больших чисел на множители было разработано множество новых алгоритмов [47]. Среди них особое место заняли квантовые алгоритмы ML, которые часто используют квантовые схемы в качестве математических ядер [48] для имитации сложных функций, или используют квантовые схемы для представления архитектуры нейронных сетей с регулируемыми параметрами модели [49.]

Например, взаимодействие между белком и лигандом в основе своей является квантово-механическим, поэтому квантовые симуляторы или квантовые компьютеры обладают потенциалом для более эффективного моделирования этих взаимодействий, прокладывая путь к будущей точной и надежной разработке лекарств [50, 51]. В [52] использовали VQE- вариационный квантовый преобразователь собственных значений (variational quantum eigensolver, VQE) для оптимизации малых молекул и квантовых магнитов. Квантовые компьютеры также использовались для оптимизации процессов молекулярной стыковки и прогнозирования средства связывания лигандов с более высокой точностью [53-56].

### 6.2. Возможности применения, ограничения и сравнение с классическими методами

Квантовые вычисления могут повысить точность молекулярной геометрии или трехмерного расположения атомов, повысить конформационную гибкость лекарств и рецепторов при молекулярной стыковке и определить оптимальные участки для стыковки. Эти три важных физических параметра создают проблемы для традиционных методов на разных уровнях. Это повысит точность определения сродства стыковки, однако остаётся проблема шума среднего масштаба (noisy intermediate-scale quantum NISQ) [57, 58], и доступных высококачественных кубитов, что ограничивают биофизические возможности квантовых компьютеров. Квантовый классификатор превосходит классический при работе с неполными данными о признаках, и это преимущество сохраняется для различных методов эмбединга. Квантовый классификатор работает лучше, особенно когда количество доступных признаков невелико. Чтобы получить более глубокое представление об этом количественном преимуществе, систематически исследовали влияние количества признаков на классификаторы. Обнаружили, что классический классификатор более чувствителен к изменениям в количестве признаков, в то время как квантовый классификатор демонстрирует надежность. Надежность квантового классификатора также проявляется в его чувствительности к количеству обучающих выборок. Это приводит к более медленному снижению точности тестирования в зависимости от количества обучающих выборок и подчеркивает преимущество квантового классификатора в сценариях с небольшим количеством признаков. Улучшенная производительность квантового классификатора может быть объяснена ограничением, налагаемым размером Вапника-Червоненкиса (VC) на ошибку обобщения. Известно, что на оценку ошибки обобщения влияет корень квадратный из размерности VC, деленный на размер выборки. Интересно зависимость от данных: квантовое преимущество проявляется не во всех тестируемых наборах данных. Рассмотренные схемные/ нейронные слои являются общими и могут быть применены к различным другим структурам моделей. Понимая неотъемлемые свойства этих классификаторов в различных сценариях, можно эффективно использовать схемные модули для реализации крупномасштабных моделей квантового машинного обучения.

## 6 ЗАКЛЮЧЕНИЕ

Таким образом, квантовые вычисления могут оптимизировать молекулярное моделирование и повысить точность разработки лекарств за счет эффективной обработки сложных квантово-механических взаимодействий между атомами в молекулах. По мере развития этой технологии она может произвести революцию в фармацевтических исследованиях за счет сокращения вычислительных затрат и повышения точности молекулярного докинга и других важных процессов. Движущей силой этого прогресса является стратегическое сотрудничество между лидерами отрасли и передовыми подразделениями по QC [59]. Примером является сотрудничество между Boehringer Ingelheim [60] и Rigetti Computing [61], которое дало заметное 20-кратное повышение растворимости существующей молекулы лекарственного средства. Merck [62] в партнерстве с Zapata Computing [63] фокусируется на сложной динамике белково-лигандных взаимодействий. Количественное модели-

рование позволило добиться двукратного ускорения при моделировании этих взаимодействий, существенно ускорив поиск лекарств. Vertex Pharmaceuticals [64] и QuantumScape [65] являются пионерами в использовании квантового моделирования для разработки новых антибиотиков, нацеленных на конкретные уязвимые места бактерий. Exscientia и Sumitomo Dainippon Pharma используют квантовое моделирование в поиске материалов для систем доставки лекарств. AstraZeneca и PsiQuantum, используя квантовые алгоритмы, удалось повысить точность предсказания препарат-белковых взаимодействий на 10%, за счёт оптимизации подбора кандидата и разработки путей. Хотя возможности огромны и многообещающи, квантовые преимущества не всегда гарантированы, и нам все еще необходимо исследовать новые квантовые алгоритмы и разрабатывать квантовые схемы, которые наилучшим образом используют квантовую запутанность для достижения более высокой производительности, чем классические. Квантовые вычисления при разработке лекарств используются исключительно *in silico*, что означает, что они включают в себя только виртуальное моделирование, а не физические эксперименты или клинические испытания. В результате квантовые вычисления способствуют повышению биобезопасности, сводя к минимуму необходимость в тестировании на животных и людях, а также помогают снизить потребление энергии и воздействие на окружающую среду, способствуя устойчивому развитию, которое становится все более важной целью современного общества.

## ЛИТЕРАТУРА

- [1] Doytchinova, I. (2022). Drug design—past, present, future. *Molecules* 27, 1496.
- [2] Gorgulla, C., Boeszoermyeni, A., Wang, Z.-F., Fischer, P.D., Coote, P.W., Padmanabha Das, K.M., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 580, 663–668.
- [3] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zi'dek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature* 596, 583
- [4] Livermore Lab and BridgeBio Announce Human Trials for HPC-Discovered Cancer Drug, <https://www.llnl.gov/article/51336/llnl-bridgebio-announcetrials-supercomputing-discovered-cancer-drug> 2024 Accessed: 2024-06-07.
- [5] Yang, C., Chen, E.A., and Zhang, Y. (2022). Protein–ligand docking in the machine-learning era. *Molecules* 27, 4568
- [6] B. Zhong, X. Su, M. Wen, S. Zuo, L. Hong, J. Lin, Parafold: Paralleling alphafold for large-scale predictions, in: International Conference on High Performance Computing in Asia-Pacific Region Workshops, pp. 1–9.
- [7] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A.W., Green, T., Zi'dek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with alphafold-multimer. Preprint at bioRxiv. <https://doi.org/10.1101/2021.10.04.463034>
- [8] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* 630, 493–500.
- [9] Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., Collins, J., Diez-Cecilia, E., Kelly, B., Goodarzi, H., and Yuan, J.S. (2020). Artificial intelligence for covid-19 drug discovery and vaccine development. *Front. Artif. Intell.* 3, 65.
- [10] Muratov EN, Bajorath J, Sheridan RP et al (2020) QSAR without borders. *Chem Soc Rev* 49:3525–3564. <https://doi.org/10.1039/D0CS00098A>
- [11] Siramshetty VB, Nguyen D-T, Martinez NJ et al (2020) Critical assessment of artificial intelligence methods for prediction of hERG channel inhibition in the “Big Data” Era. *J Chem Inf Model* 60:6007–6019. <https://doi.org/10.1021/acs.jcim.0c00884>
- [12] Jain S, Siramshetty VB, Alves VM et al (2021) Large-scale modeling of multispecies acute toxicity end points using consensus of multitask deep learning methods. *J Chem Inf Model* 61:653–663. <https://doi.org/10.1021/acs.jcim.0c01164>
- [13] Casanova-Alvarez O, Morales-Helguera A, Cabrera-Púrez MB et al (2021) A novel automated framework for QSAR modeling of highly imbalanced leishmania high-throughput screening data. *J Chem Inf Model* 61:3213–3231. <https://doi.org/10.1021/acs.jcim.0c01439>
- [14] Neves BJ, Braga RC, Melo-Filho CC et al (2018) QSAR-based virtual screening: advances and applications in drug discovery. *Front Pharmacol* 9

- [15] Spiegel J, Senderowitz H (2020) Evaluation of QSAR equations for virtual screening. *Int J Mol Sci* 21:7828. <https://doi.org/10.3390/ijms21217828>
- [16] Matveieva M, Polishchuk P (2021) Benchmarks for interpretation of QSAR models. *J Cheminformatics* 13:41. <https://doi.org/10.1186/s13321-021-00519-x>
- [17] Wellnitz J, Martin H-J, Anwar Hossain M et al (2024) STOPLIGHT: a hit scoring calculator. *J Chem Inf Model* 64:4387–4391. <https://doi.org/10.1021/acs.jcim.4c00412>
- [18] Pourbasheer, E., Aalizadeh, R. & Ganjali, M. R. Qsar study of ck2 inhibitors by ga-mlr and ga-svm methods. *Arab. J.Chem.* 12, 2141–2149 (2019).
- [19] Kim, B. C., Joe, D., Woo, Y., Kim, Y. & Yoon, G. Extension of pqsar: Ensemble model generated by random forest and partial least squares regressions. *IEEE Access* 8, 180087–180099 (2020).
- [20] Nijkamp, E., Ruffolo, J.A., Weinstein, E.N., Naik, N., and Madani, A. (2023). Progen2: exploring the boundaries of protein language models. *Cell Syst.*14, 968–978.e3.
- [21] Ferruz, N., Schmidt, S., and Hockner, B. (2022). Protgpt2 is a deep unsupervised language model for protein design. *Nat. Commun.* 13, 4348.
- [22] Unsal, S., Atas, H., Albayrak, M., Turhan, K., Acar, A.C., and Dogan, T. (2022). Learning functional properties of proteins with language models. *Nat. Mach. Intell.* 4, 227–245.
- [23] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. (2023). Llama 2: Open foundation and fine-tuned chat models. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2307.09288>
- [24] Srivastava, A., Rastogi, A., Rao, A., Shoeb, A.A.M., Abid, A., Fisch, A., Brown, A.R., Santoro, A., Gupta, A., Garriga-Alonso, A., et al. (2022). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2206.04615>
- [25] Lu, J., and Zhang, Y. (2022). Unified deep learning model for multitask reaction predictions with explanation. *J. Chem. Inf. Model.* 62, 1376–1387
- [26] Bloch, F. (1946). Nuclear induction. *Phys. Rev.* 70, 460–474.
- [27] Benedetti, M., Lloyd, E., Sack, S. & Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Sci. Technol.* 4, 043001 (2019).
- [28] Biamonte, J. et al. Quantum machine learning. *Nature* 549, 195–202 (2017).
- [29] Tsang, S. L., West, M. T., Erfani, S. M. & Usman, M. Hybrid quantum-classical generative adversarial network for high resolution image generation. arXiv preprint arXiv:2212.11614 (2022).
- [30] Huang, H.-L. et al. Experimental quantum generative adversarial networks for image generation. *Phys. Rev. Appl.* 16, 024051 (2021).
- [31] Suzuki, T. & Katouda, M. Predicting toxicity by quantum machine learning. *J. Phys. Commun.* 4, 125012 (2020).
- [32] Wu, S. L. et al. Application of quantum machine learning using the quantum kernel algorithm on high energy physics analysis at the LHC. *Phys. Rev. Res.* 3, 033221 (2021).
- [33] Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* 567, 209–212 (2019).
- [34] Batra, K. et al. Quantum machine learning algorithms for drug discovery applications. *J. chemical information modeling* 61, 2641–2647 (2021).
- [35] Liu, Y., Arunachalam, S. & Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* 17, 1013–1017 (2021).
- [36] Du, Y., Hsieh, M.-H., Liu, T. & Tao, D. Expressive power of parametrized quantum circuits. *Phys. Rev. Res.* 2, 033125 (2020).
- [37] Cao, Y., Romero, J. & Aspuru-Guzik, A. Potential of quantum computing for drug discovery. *IBM J. Res. Dev.* 62, 6–1 (2018).
- [38] Kao, P.-Y. et al. Exploring the advantages of quantum generative adversarial networks in generative chemistry. *J. Chem. Inf. Model.* (2023).
- [39] Domingo, L., Djukic, M., Johnson, C. & Borondo, F. Hybrid quantum-classical convolutional neural networks to improve molecular protein binding affinity predictions. arXiv preprint arXiv:2301.06331 (2023).
- [40] Hekler, E.B. et al. Why we need a small data paradigm. *BMC medicine* 17, 1–9 (2019).
- [41] Jolliffe, I.T. *Principal component analysis for special types of data* (Springer, 2002).
- [42] Wei-Yin Chiang, Po-Yu Kao, Tzu-Lan Yeh, Ya-Chu Yang, Yen-Chu Lin, Alex Zhavoronkov Enhancing Drug Discovery: Quantum Machine Learning for QSAR Prediction with Incomplete Data arXiv:2501.13395v1 [quant-ph] 23 Jan 2025

- [43] Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. chemical information modeling* 50, 742–754 (2010).
- [44] Zeng, X. *et al.* Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat. Mach. Intell.* 4, 1004–1016 (2022).
- [45] Landrum, G. *et al.* rdkit/rdkit: 2022\_09\_1b1 (q3 2022) release, DOI: 10.5281/zenodo.7179566 (2022).
- [46] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G.S.L., Buell, D.A., *et al.* (2019). Quantum supremacy using a programmable superconducting processor. *Nature* 574, 505–510.
- [47] Shor, P.W. (1995). Scheme for reducing decoherence in quantum computer memory. *Phys. Rev.* 52, R2493–R2496.
- [48] Grimsley, H.R., Economou, S.E., Barnes, E., and Mayhall, N.J. (2019). An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* 10, 3007.
- [49] Cong, I., Choi, S., and Lukin, M.D. (2019). Quantum convolutional neural networks. *Nat. Phys.* 15, 1273–1278.
- [50] Cao, Y., Romero, J., and Aspuru-Guzik, A. (2018). Potential of quantum computing for drug discovery. *IBM J. Res. Dev.* 62, 1–6.
- [51] Outeiral, C., Strahm, M., Shi, J., Morris, G.M., Benjamin, S.C., and Deane, C.M. (2021). The prospects of quantum computing in computational molecular biology. *WIREs Comput. Mol. Sci.* 11, e1481.
- [52] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., and Gambetta, J.M. (2017). Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* 549, 242–246.
- [53] Li, Y., Cui, X., Xiong, Z., Liu, B., Wang, B.-Y., Shu, R., Qiao, N., and Yung, M.-H. (2024). Quantum molecular docking with quantum-inspired algorithm. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2404.08265>
- [54] Shu, R., Liu, B., Xiong, Z., Cui, X., Li, Y., Cui, W., Yung, M.-H., and Qiao, N. (2024). Quantum-inspired machine learning for molecular docking. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2308.04098>
- [55] Ding, Q.-M., Huang, Y.-M., and Yuan, X. (2024). Molecular docking via quantum approximate optimization algorithm. *Phys. Rev. Appl.* 21, 034036.
- [56] Santagati, R., Aspuru-Guzik, A., Babbush, R., Degroote, M., Gonzalez, L., Kyoseva, E., Moll, N., Oppel, M., Parrish, R.M., Rubin, N.C., *et al.* (2024). Drug design on quantum computers. *Nat. Phys.* 20, 549–557.
- [57] Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum* 2, 79.
- [58] Bharti, K., Cervera-Lierta, A., Kyaw, T.H., Haug, T., Alperin-Lea, S., Anand, A., Degroote, M., Heimonen, H., Kottmann, J.S., Menke, T., *et al.* (2022). Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* 94, 015004.
- [59] “Boehringer Ingelheim - Life Forward,” Jan. 2024, <https://www.boehringer-ingenelheim.com>
- [60] G. Kumar, S. Yadav, A. Aukherjee, V. Hassija, M. Guizani. Recent Advances in Quantum Computing for Drug Discovery and Development DOI: 10.1109/ACCESS.2024.3376408
- [61] “Quantum Computing,” Jan. 2024, <https://www.rigetti.com>
- [62] “Merck | Home,” Jan. 2024, [Online; accessed 31. Jan. 2024]. [Online] Available: <https://www.merck.com>
- [63] Home, Jan. 2024, <https://zapata.ai>
- [64] Vertex Pharmaceuticals | Home <https://www.vrtx.com>
- [65] Building the Best Solid-State Battery | Quantum Scape, Oct. 2023, <https://www.quantumscape.com>

Поступила в редакцию 30.05.2025

**Цитирование:** Адилова А.Т., Давронов Р.Р. (2025). Компьютерная разработка лекарств: от традиционных методов моделирования к языковым моделям и квантовым вычислениям. *Международный журнал теоретических и прикладных вопросов цифровых технологий*, 8(3), –С. 110-122. <https://doi.org/10.62132/ijdt.v8i3.294>.

## COMPUTER DRUG DEVELOPMENT: FROM TRADITIONAL MODELING METHODS TO LANGUAGE MODELS AND QUANTUM COMPUTING

*Adilova F.T.<sup>1</sup>, Davronov R.R.<sup>1</sup>*

<sup>1</sup> V.I. Romanovsky Institute of Mathematics of the Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan

**Abstract.** Drug development is a central topic at the intersection of structural biology, biochemistry, and medicine, associated with significant challenges such as high cost (billions of dollars), low success rates (less than 10%), and extremely long development cycles (10-15 years). The Computer Aided Drug Design (CADD) system demonstrates tremendous advantages in solving these tasks and speeding up the process, making it an indispensable tool in the pharmaceutical industry and scientific research. The recent development of AlphaFold2 and AlphaFold3, the 2024 Nobel Prize winners, marks significant progress in the field of CADD. In addition to AlphaFold, various machine learning (ML) methods are revolutionizing various stages of drug development, from virtual screening to predictive modeling of drug interactions and treatment targets. Language models such as GPT models offer promising applications for developing research hypotheses and helping to interpret complex biological data. Quantum computing has the potential to solve complex molecular modeling and optimization problems that are currently unsolvable for classical computers, although their practical implementation is still in its early stages. This analytical review presents the latest developments in this field and evaluates the possibilities presented by machine learning, language models and quantum computing in CADD.

**Keywords:** Computer-Aided Drug Design (CADD), Machine Learning (ML), Deep Learning (DL), Quantum Computing (QC), Language Models (LM), Quantitative Structure-Activity Relationship (QSAR), Drug Discovery, Molecular Modeling, Protein Structure Prediction, Virtual Screening, AlphaFold2, AlphaFold3, Hybrid Quantum-Classical Algorithms, Molecular Docking, Incomplete Data, Molecular Embedding, Morgan Fingerprint, ImageMol.