

## ОТЗЫВ

официального оппонента Порозова Юрия Борисовича на диссертационную работу Столбова Леонида Алексеевича «Разработка подходов к виртуальному скринингу антивирусных соединений с учетом гетерогенности информации», представленной на соискание ученой степени кандидата биологических наук по специальности 1.5.8. – Математическая биология, биоинформатика

### **Актуальность исследования**

В настоящее время алгоритмы, технологии и прикладные решения поиска соединений с желаемой активностью, предваряющие испытания на животных и человеке, играют всё возрастающую роль. Ярким подтверждением этого является принятый в 2022-м году акт (закон) H.R.2565, который явно разрешает Управлению по контролю качества пищевых продуктов и лекарственных средств (FDA) использовать альтернативные по отношению к испытаниям на животных методы исследования эффективности и безопасности. Связано это, очевидно, с появлением математических моделей и методов, которые обладают требуемой точностью и обеспечивают повторяемость результатов. Нужно подчеркнуть, что объемы информации о химических соединениях и их биологических эффектах растут темпами, которые уже давно не позволяют вести разработку, только лишь полагаясь на мощности лабораторий. Более того, накопление (или генерация) этой информации заметно опережает наши возможности в её компьютерной обработке, фильтрации и анализе. Именно поэтому определенные ожидания учёные возлагают на появление доступных квантовых компьютеров и ПО для них. Но в то же время усложняются, становятся более «интеллектуальными» и алгоритмы и подходы для имеющихся вычислительных мощностей.

В представленной диссертации исследования проведены на данных о соединениях, активных в отношении ВИЧ и SARS-CoV-2. Излишне будет говорить о том, что оба эти вируса а) надолго вошли в современную действительность, б) оказывают значительное влияние на продолжительность жизни населения и экономику стран и в) все ещё не имеют дешевого и надежного ответа в плане терапии и полного излечения.

Диссертация Столбова Л.А. посвящена усовершенствованию методов виртуального скрининга соединений с учетом гетерогенности входных данных. Несомненно, работы в этом направлении, и представленная Леонидом Алексеевичем Столбовым диссертация, являются высоко актуальными в плане разработки новых математических подходов. Кроме того, актуальным является и объект исследования – потенциальные анти-ВИЧ и анти-SARS-CoV-2 соединения и базы данных.

## **Научная новизна и практическая значимость диссертации**

Автором диссертационной работы созданы методы и алгоритмы, позволяющие использовать гетерогенные исходные данные в задачах классификации и анализа взаимосвязей «структура-активность» - методы самосогласованной логистической и самосогласованной экстремальной классификации. Эти алгоритмы автор реализовал в программном коде и впервые применил их к оценке ингибирующей активности соединений в отношении интегразы, протеазы и обратной транскриптазы ВИЧ-1 и ферментов SARS-CoV-2. Проведя моделирование, Столбов Л.А. доказал, что разработанные им методы самосогласованной классификации обладают лучшими по сравнению с количественными методами показателями точности и предсказательной способности в условиях гетерогенности входных данных.

Разработанные автором теоретические и программные решения легли в основу открытого веб-сервиса AntiHIV Pred, который открывает неограниченному кругу пользователей возможности для прогноза анти-ВИЧ активности с использованием самых современных и наиболее точных методов.

### **Апробация работы**

Основные положения диссертации были представлены на российских и международных конференциях и симпозиумах: XXXVIII Symposium of Bioinformatics and Computer-Aided Drug Discovery (Virtual, 2022), XXVII Российский национальный конгресс «Человек и лекарство». (Москва, 2020), Third international School-Seminar «From Empirical to Predictive Chemistry» (Казань, 2018), XXV Российский национальный конгресс «Человек и лекарство» (Москва 2018), IX Международный конгресс «Биотехнология: состояние и перспективы развития. Науки о жизни» (Москва, 2019).

Кроме того, по материалам диссертации опубликована 21 работа в российских и международных научных изданиях, включая шесть статей в рецензируемых научных журналах, входящих в перечень, рекомендованный ВАК, и 14 публикаций в трудах конференций. Поскольку по результатам теоретических исследований были разработаны и программные решения, автором получено свидетельство Федеральной службы по интеллектуальной собственности о регистрации программы для ЭВМ Sarmath №: RU2022662930 от 07.07.2022 г.

### **Структура и содержание диссертации**

Диссертация построена по стандартной схеме, ее структура, объем и содержание

типичны для работ такого рода. В работе 132 страницы, включает 35 рисунков, 13 таблиц. В диссертации есть следующие разделы: список сокращений, введение, глава 1 «Обзор литературы», глава 2 «Материалы и методы», глава 3 «Результаты и обсуждение», заключение; выводы; список работ, опубликованных по теме диссертации, финансирование работы, благодарности, список литературы, приложения А-Д, в которых также имеются три таблицы.

Иллюстративный материал и цитируемая литература (169 источников за 47 лет) соответствуют задачам. Я должен отметить, что диссертация читается удивительно легко, она написана очень хорошим научным языком и довольно подробно. В разделе «Введение» очень чётко, убедительно и, вместе с тем, без избыточности, обоснованы состояние проблемы, актуальность, цель и задачи исследования. Они хорошо сформулированы и дополняют друг друга.

**Обзор литературы** весьма полон. Автор, безусловно, отлично ориентируется как в математической части темы своей работы, так и в её биологической части, что предполагает специальность «Математическая биология, биоинформатика». Сам обзор грамотно структурирован и сжато, но, тем не менее, вполне достаточно для понимания хода работы описывает современное состояние науки в области компьютерного конструирования лекарств в целом и, более подробно, в области QSAR моделирования и анализа в контексте виртуального скрининга. В обзоре также приведены основные мишени ВИЧ-1 и SARS-CoV-2, поскольку именно эти белки являются важнейшим компонентом дальнейших построений и выводов автора.

**Вторая глава** посвящена материалам и методам, которые автор использовал в работе. Подробно описано формирование обучающих и тестовых выборок, начиная от генерации искусственных данных пар «структура-активность» и до методов и подходов оценки собственно результатов работы предложенных автором методов SCLC и SCEL. Ядром этой главы, как мне видится, является раздел 2, в котором очень подробно описана логика развития метода самосогласованной регрессии (используемой для построения количественных зависимостей) в полноценные классификационные модели SCLC и SCEL. Автор приводит подробный ряд формул, который позволяет проследить логику, убедиться в истинности математических построений и получить представление о природе результатов, которые описываются в третьей главе.

**Третья глава** диссертации посвящена результатам собственных исследований и их

обсуждению. В первом подразделе этой главы автор убедительно показывает, что оба метода хорошо справляются с построением классификационных моделей для несбалансированных выборок. При этом не происходит снижения точности и прогностической способности модели. Считаю это важным достижением, так как до сих пор качество выборок являлось чуть ли не основным ограничивающим фактором, приводящим неискушённого исследователя к ошибкам интерпретации, а искушённого – к попыткам при помощи различных ухищрений привести выборку к приемлемому виду.

Второй подраздел третьей главы описывает полуавтоматические процедуры построения выборок. Это был важный этап работы, так как он позволил унифицировать, стандартизировать подготовку выборок соединений, ингибирующих белки ВИЧ-1 и SARS-Co-2. Хочу заметить, что такой подход свидетельствует о вдумчивом, ответственном отношении Столбова Л.А. к своему исследованию и пониманию им важности автоматизации этапов, которые совсем ещё недавно выполнялись вручную со всеми негативными возможностями человеческого влияния на данные.

Третий подраздел посвящён построению QSAR моделей для соединений, активных в отношении белков ВИЧ-1. Автор не просто построил прогностические модели для ингибиторов протеазы, интегразы и обратной транскриптазы ВИЧ-1, но и проанализировал значительные различия в качестве прогноза между протеазой ( $R_{\text{test}}^2=0.801$ ) и интегразой ( $R_{\text{test}}^2=0.678$ ) и обратной транскриптазой ( $R_{\text{test}}^2=0.502$ ). Автором был сделан справедливый вывод о влиянии на прогноз новых дескрипторов у соединений тестовой выборки.

Четвертый подраздел третьей главы описывает построение классификационных моделей для ингибиторов белков ВИЧ-1 и SARS-CoV-2 с использованием разработанных автором методов SCLC и SCEC и их сравнение с ранее известными SCR, SVM, PASS и ИНС. Результаты представлены в таблицах 5-7, на рисунках 23-25 и свидетельствуют о том, что SCEC для ингибиторов всех трёх ферментов превосходит остальные методы при меньшей размерности модели. Для белков SARS-CoV-2 данные, свидетельствующие к преимуществу SCEC, приведены в таблицах 8-10.

В пятом подразделе диссертации Столбов Л.А. описывает веб-сервис для прогноза активности на основе структурной формулы соединения для мишеней ВИЧ-1.

В разделе «Заключение» автор кратко обобщает проблематику, логику проведения исследования и полученные результаты.

**Выводы** работы полностью вытекают из полученных результатов и согласуются с поставленными задачами исследования. Несомненно, они имеют научно-практическую значимость и их достоверность сомнению не подлежит.

## Автореферат

Автореферат в сжатой форме полностью отражает смысл диссертации и даёт исчерпывающее представление о проблематике работы, о методах и полученных результатах. Как и сама диссертация, автореферат оформлен в соответствии с предъявляемыми к таким документам требованиями.

## Вопросы и замечания

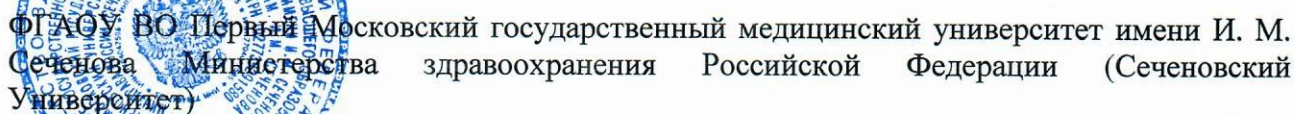
Несмотря на то, что диссертационная работа Л.А. Столбова легко читается, хорошо проиллюстрирована, а выводы логично следуют из результатов, достоверность которых не вызывает сомнений, позволю себе высказать некоторые замечания, задать вопросы, которые, однако, не являются принципиальными.

1. На с.25 автор пишет: «Так, количественные значения могут быть классифицированы, **невзирая** (выделено мной) на различные экспериментальные условия и наличие ошибок измерения. Сама классификация при этом происходит по заранее принятому порогу активности [77, 78]». Как в таком случае избежать ошибок классификации при неверных\намеренно сфальсифицированных измерениях или сфабрикованных количественных оценках?
2. Иногда в тексте встречается ссылка на приложение с цифрой (например, на с. 64. (пример кода для основных методов дан в Приложении 4), а иногда с буквой (с. 72). При этом все приложения имеют буквенную сортировку.
3. Раздел главы 2, в котором описывается используемое программное обеспечение, хотелось бы видеть более подробным. Правильно ли я понимаю, что разработанные модули и программный код, которые работают в веб-интерфейсе AntiHIV Pred, не являются доступными в репозитории открытого кода GitHub, а значит, недоступны для развития\модификации или встраивания в иные программные продукты по лицензии GNU GPL/MIT?
4. В качестве конкурирующих методов автором применены, в частности, искусственные нейронные сети. При этом каких-либо подробностей об их архитектуре, реализации, параметрам, автор не приводит. При этом утверждение на с. 85 «В случае SVM и ИНС автоматического отбора значимых переменных не происходит, поэтому они были отобраны с использованием SCEC» верно только лишь для той ИНС, которую использовал автор в работе для сравнения, и характеристики\авторство\ссылки на результаты тестирования которой не приведены.

- Следует подчеркнуть, что вопросы и замечания относятся преимущественно к тексту диссертационной работы и не затрагивают принципиальных выводов и результатов работы, которые несомненно вносят существенный вклад в современные представления и методологию дизайна лекарств, основанного на структуре лиганда как важнейшем компоненте виртуального скрининга.

Исходя из вышеизложенного можно утверждать, что диссертационная работа Столбова Леонида Алексеевича «Разработка подходов к виртуальному скринингу противовирусных соединений с учетом гетерогенности информации», представленная на соискание ученой степени кандидата биологических наук по специальности 1.5.8. – Математическая биология, биоинформатика, является законченным научно-квалификационным исследованием и соответствует требованиям п. 9 положения «О присуждении ученых степеней», утвержденного постановлением Правительства РФ N 842 от 24.09.2013г (в редакции постановления Правительства РФ №101 от 26.01.2023). Работа соответствует всем требованиям, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8. – Математическая биология, биоинформатика.

Порозов Юрий Борисович



Эл. почта: [yuri.porozov@gmail.com](mailto:yuri.porozov@gmail.com).