

ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата химических наук Хрисанфова Михаила Дмитриевича
на тему: «Поиск ошибок в хроматографических базах данных и
моделирование свойств идентифицируемых молекул
методами машинного обучения»
по специальности 1.4.2. Аналитическая химия

Представленное диссертационное исследование посвящено разработке открытых вычислительных инструментов для автоматизированного поиска ошибок в хроматографических базах данных, стандартизации масс-спектрометрических данных с точки зрения величин m/z , прогнозирования масс-спектров электронной ионизации по структуре, и прогнозированию структурных фрагментов по масс-спектру. Эти задачи являются, безусловно, весьма актуальными в аналитической химии, поскольку стремительный рост объемов данных, получаемых от современных приборов, требует применения автоматизированных подходов к работе, которые при этом не компрометировали бы точность результатов анализа. Исследование полностью соответствует профилю специальности 1.4.2. Аналитическая химия.

Научная новизна диссертационной работы определяется несколькими обстоятельствами. Во-первых, впервые предложен и теоретически обоснован ансамблевый подход к выявлению ошибок в базах данных по хроматографическим индексам удерживания, основанный на механизме голосования нескольких моделей машинного обучения и демонстрирующий статистически значимое превосходство над другими способами поиска подозрительных данных. Подход всесторонне изучен, предложены способы оценки эффективности поиска. Во-вторых, систематизированы и экспериментально сопоставлены алгоритмы округления значений m/z , реализуемые в распространенных хромато-масс-спектрометрических

программных пакетах; предложен оптимизированный вариант, минимизирующий влияние инструментальных погрешностей на итоговую идентификацию. В-третьих, разработаны два варианта полносвязных нейросетевых моделей (EI2FP:Lite и EI2FP:Full) для прогноза молекулярных отпечатков на основе масс-спектров электронной ионизации, которые обеспечивают многократное ускорение вычислений при одновременном повышении показателей точности, правильности и полноты относительно разработанной ранее модели DeepEI. В-четвертых, предложена усовершенствованная нейросетевая архитектура для генерации масс-спектров по молекулярной структуре NEIMS-PyTorch, демонстрирующая превосходство над известным алгоритмом NEIMS. Отдельно стоит отметить, что все необходимое для самостоятельного запуска модели NEIMS-PyTorch размещено в открытых репозиториях и доступно для использования всем заинтересованным, что, несомненно, повышает значимость исследования.

Полученные результаты обладают выраженной прикладной направленностью и представляются готовыми к внедрению в аналитическую практику лабораторий хромато-масс-спектрометрического анализа, что определяет **практическую значимость** работы. Разработанные модели и библиотеки Python предоставляют исследователям готовые инструменты для автоматизированной очистки и верификации данных из референтных баз (NIST, METLIN и др.); для стандартизированной обработки масс-спектрометрических данных; для быстрого прогнозирования молекулярных отпечатков и масс-спектров электронной ионизации. Предложенные модели и инструменты расширяют возможности проведения нецелевого скрининга на основе хромато-масс-спектрометрических подходов.

Достоверность выводов исследования подтверждается многоуровневой валидацией: на реальных базах данных (NIST 17 RI, METLIN SMRT) и на контролируемых синтетических выборках с целенаправленным внесением артефактов известной природы. Эффективность разработанных методов количественно оценена посредством общепринятых статистических

метрик (точность, полнота, правильность). Полная открытость исходного кода, предобученных весов, конфигурационных файлов гарантируют возможность независимой верификации, воспроизведения и использования результатов научным сообществом.

Диссертационная работа имеет традиционную структуру, состоит из введения, обзора литературы, методов и алгоритмов, обсуждения результатов (две главы), заключения, выводов и списка литературы. Работа изложена на 110 страницах машинописного текста, содержит 39 рисунков, 12 таблиц и список литературы из 158 наименований. Не смотря на относительную компактность текста, он вполне информативен и достаточно детализирован. По теме диссертации опубликовано 4 статьи в рецензируемых научных изданиях, а также сделано множество докладов на всероссийских и международных конференциях.

В целом, работа производит очень благоприятное впечатление, написана хорошим грамотным языком, материал изложен в логической последовательности, хорошо структурирован. Все положения, выносимые на защиту, подкреплены фактическим материалом исследования и не вызывают сомнений.

После знакомства с текстом диссертации замечаний принципиального характера у меня не возникло, однако, есть ряд вопросов уточняющего характера и комментариев.

1. В контексте аналитической химии термин «предсказание» звучит неестественно, привнося некий оттенок мистики. Гораздо более корректным и научно обоснованным мне видится использование термина «прогноз» (или «оценивание»). Разумеется, устоявшаяся в сообществе машинного обучения калька с английского predict прочно вошла в обиход. Однако, никакой официально утвержденной терминологии для применения машинного обучения в аналитической химии пока нет, сейчас как раз подходящий момент, чтобы задать правильный тон и договориться о корректной номенклатуре.

2. Есть замечания по поводу баланса дисциплин: за плотными слоями терминов из области машинного обучения в работе часто теряется собственно аналитическая химия. Во многих случаях для приведенных на рисунках и в таблицах результатов моделирования не указывается размерность величин, динамический диапазон целевых переменных для прогноза и т.п. В итоге, читатель остается один на один с горами цифр, которые без привязки к физико-химическому смыслу происходящего не позволяют разделить с автором восторг открытий.

3. Дискуссионный вопрос по поводу поиска «ошибок» в базах данных – как отличить объект с аномальными значениями от ошибки? Всегда ли выбросы в таких базах можно трактовать, как ошибки?

4. На мой взгляд, представление подхода к поиску структур с использованием молекулярных отпечатков сильно бы украсила более показательная валидация. Два примера соединений из Таблицы 12, безусловно, впечатляют, но выглядят не вполне достаточными. Возможно, вопрос стоило бы поставить несколько иначе: если взять 10 000 реальных неизвестных спектров, как часто истинная структура будет входить в лучшие 10 или лучшие 5 кандидатов?

5. Остался не вполне понятным вопрос, для чего использовать в ансамбле несколько моделей схожей архитектуры, если показано, что модели разного типа в ансамбле работают лучше. Почему проверка подхода на разных базах (NIST и METLIN) выполнялась на основе разных моделей? Мне представляется логичным тестировать на реальных базах уже одну отлаженную модель самого высокого качества.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.4.2. Аналитическая химия (по химическим

наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова. Диссертационное исследование оформлено согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Хрисанфов Михаил Дмитриевич заслуживает присуждения ученой степени кандидата химических наук по специальности 1.4.2. Аналитическая химия.

Официальный оппонент:

доктор химических наук,

профессор кафедры аналитической химии Института химии СПбГУ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»,

Институт химии,

кафедра аналитической химии

Кирсанов Дмитрий Олегович


подпись


15 июля 2026 г.

Контактные данные:

Специальность, по которой официальным оппонентом защищена диссертация:

02.00.02 – Аналитическая химия

Адрес места работы:

198504, г. Санкт-Петербург, Петергоф, Университетский пр. 26

Институт химии СПбГУ, кафедра аналитической химии

Подпись сотрудника Института химии СПбГУ

Кирсанова Д.О. удостоверяю: 





