

**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**

Диссертационная работа Хрисанфова Михаила Дмитриевича посвящена повышению надежности и качества обработки хроматомасс-спектральных данных за счет применения методов машинного и глубокого обучения. Работа объединяет два взаимодополняющих направления: очистку экспериментальных хромато-масс-спектрометрических баз данных от ошибочных записей и совершенствование предсказательных моделей масс-спектральных характеристик, что делает диссертацию цельным и логически выстроенным исследованием.

Актуальность темы не вызывает сомнений. Широчайшее применение газовой и жидкостной хроматографии в сочетании с масс-спектрометрией в исследовательской и аналитической практике требует все более совершенных инструментов обработки данных. Автор совершенно справедливо отмечает, что даже широко используемые и признанные «эталонными» базы данных, такие как NIST RI и METLIN SMRT, могут содержать ошибочные записи, а их выявление вручную практически невозможно из-за огромного объема данных. Предложенный в диссертации подход к автоматизированному поиску ошибок, основанный на механизме голосования независимых предсказательных моделей, представляет собой оригинальное и

эффективное решение этой проблемы, обладающее высокой научной новизной и очевидной практической значимостью.

Научная новизна работы подтверждается рядом убедительных результатов. Автором впервые предложен и апробирован «ансамблевый» подход к обнаружению потенциально ошибочных записей в хроматографических базах данных, эффективность которого продемонстрирована на специально сгенерированных синтетических наборах данных с контролируемым числом и типом ошибок. Методологическая безупречность и изящность данного подхода позволила диссертанту объективно оценить его работоспособность при отсутствии «эталонной разметки» реальных данных. Практическим результатом применения метода стало обнаружение 2093 потенциально ошибочных записей в NIST 17 RI и 1544 записей в METLIN SMRT, что представляет собой значимый вклад в повышение качества этих широко используемых мировых баз данных.

Не менее важны результаты, полученные в области обработки масс-спектральных данных. Предложенный автором алгоритм округления значений m/z с плавающей запятой до целочисленных для спектров низкого разрешения нашел реальное практическое применение. Он был включен разработчиками в широко распространяемый программный пакет OpenChrom, что наглядно свидетельствует о его востребованности и качестве. Усовершенствованные архитектуры нейросетевых моделей NEIMS-PyTorch и EI2FP:Lite/EI2FP:Full позволили заметно повысить точность предсказания масс-спектров электронной ионизации по структуре молекулы (доля точно предсказанных спектров выросла с 5% до 28%),

а также существенно (более чем в 30 раз) ускорить обучение и предсказание “молекулярных отпечатков пальцев” по сравнению с известным аналогом DeepEI при одновременном улучшении метрик точности, правильности и полноты.

Достоверность полученных результатов обеспечена корректным применением широко признанных статистических метрик и подходов машинного обучения, использованием стандартных для области библиотек языка Python, а также тщательно продуманной методологией валидации на синтетических наборах данных с контролируемыми ошибками. Результаты работы прошли серьезную апробацию: они представлены на 14 российских и международных конференциях и опубликованы в 4 статьях в рецензируемых научных журналах, рекомендованных для защиты в диссертационном совете МГУ, что полностью соответствует требованиям, предъявляемым к диссертационным работам.

Диссертация написана ясным и логичным языком, хорошо структурирована и отлично иллюстрирована (39 рисунков, 12 таблиц). Обзор литературы демонстрирует глубокое знание автором предметной области и критическое осмысление существующих подходов к предсказанию хроматографических характеристик и масс-спектров, а также к проблеме очистки химических баз данных. Все заявленные автором алгоритмы и подходы машинного обучения реализованы им лично на языке программирования Python и выложены в открытый доступ вместе с подробными инструкциями, что заметно повышает практическую ценность работы и обеспечивает возможность воспроизведения результатов сторонними исследователями.

В качестве отдельных замечаний, не влияющих на общую высокую оценку работы, можно отметить, что некоторые разделы, посвященные сравнению архитектур моделей, было бы полезно дополнить более развернутым обсуждением ограничений предложенных подходов при их переносе на другие типы хроматографических условий и приборные платформы. Также было бы интересно увидеть более подробный анализ природы найденных ошибочных записей в NIST RI и METLIN SMRT с точки зрения их вероятного происхождения.

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 полностью заслуживает присуждения ученой степени кандидата химических наук по специальности 1.4.2. Аналитическая химия.

Официальный оппонент:

доктор химических наук,

начальник лаборатории биохимии

Федерального государственного бюджетного учреждения

«Федеральный научный центр физической культуры и спорта»

Вирюс Эдуард Даниэлевич

28.07.2026

Контактные данные:

Специальность, по которой официальным оппонентом

защищена диссертация:

02.00.02 – Аналитическая химия

Адрес места работы:

105005, Москва, Елизаветинский переулок, д.10, стр. 1

Федеральное государственное бюджетное учреждение

«Федеральный научный центр физической культуры и спорта»

Подпись Вирюса Эдуарда Дани

Начальник отдела кадров и дел

правового и документационног

. Леонтьева