

## ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ МАСС-СПЕКТРОМЕТРИИ С ПРЯМОЙ ИОНИЗАЦИЕЙ В НЕЙРОХИРУРГИИ

Д. С. Заворотнюк<sup>1</sup>✉, А. А. Сорокин<sup>1</sup>, Д. С. Бормотов<sup>1</sup>, В. А. Елиферов<sup>1</sup>, К. В. Бочаров<sup>2</sup>, С. И. Пеков<sup>1,3,4</sup>, И. А. Попов<sup>1,4</sup>

<sup>1</sup> Московский физико-технический институт, Москва, Россия

<sup>2</sup> Федеральный исследовательский центр химической физики имени Н. Н. Семенова Российской академии наук, Москва, Россия

<sup>3</sup> Сколковский институт науки и технологий, Москва, Россия

<sup>4</sup> Сибирский государственный медицинский университет, Томск, Россия

Радикальное удаление опухоли до сих пор остается наиболее эффективным методом лечения онкологических заболеваний головного мозга. Задачи интраоперационного мониторинга на сегодняшний день решают с помощью позитронно-эмиссионной томографии, магнитно-резонансной томографии и гистохимического анализа, однако они требуют применения дорогостоящего оборудования высококвалифицированным персоналом, поэтому до сих пор не получили широкого распространения. В качестве альтернативы возможно применение методов масс-спектрометрии без пробоподготовки с последующим анализом масс-спектрометрических данных методами машинного обучения. Так как для масс-спектрометрии без пробоподготовки характерны более богатые и разнообразные по количеству пиков спектры, ее применение требует специальной предварительной обработки экспериментальных данных. Целью исследования было разработать методы определения оптимальных значений параметров предварительной обработки данных масс-спектрометрии без пробоподготовки. В работе представлены два таких метода, а также приведены конкретные значения параметров для данных, полученных с помощью масс-спектрометра Thermo LTQ XL Orbitrap ETD.

**Ключевые слова:** масс-спектрометрия, прямая ионизация, анализ данных, предварительная обработка данных

**Финансирование:** работа выполнена в рамках государственного задания Министерства науки и высшего образования (соглашение № 075-03-2022-107, проект № 0714-2020-0006). Исследование выполнено с использованием оборудования ЦКП ФИЦ ХФ им. Н. Н. Семенова РАН.

**Вклад авторов:** Д. С. Заворотнюк — анализ и интерпретация данных, создание программного обеспечения, написание и редактирование рукописи; А. А. Сорокин — планирование исследования, анализ и интерпретация данных, редактирование рукописи; Д. С. Бормотов — сбор и интерпретация данных, написание рукописи; В. А. Елиферов — финансовое обеспечение эксперимента; К. В. Бочаров — сбор данных; С. И. Пеков — планирование исследования, анализ и интерпретация данных, редактирование рукописи. И. А. Попов — руководство проектом, обеспечение финансирования.

**Соблюдение этических стандартов:** исследование одобрено этическим комитетом НМИЦН имени Н. Н. Бурденко (протоколы № 40 от 12 апреля 2016 г. и № 131 от 17 июля 2018 г.), проведено в соответствии с принципами Хельсинкской декларации (2000 г.) и ее последующих пересмотров. Все пациенты подписали добровольное информированное согласие на участие в исследовании и использование биоматериалов в исследовательских целях.

✉ **Для корреспонденции:** Денис Сергеевич Заворотнюк  
Институтский переулок, д. 9, 141701, г. Долгопрудный, Московская область; denis.zavorotnyuk@gmail.com

**Статья получена:** 19.12.2023 **Статья принята к печати:** 03.03.2024 **Опубликована онлайн:** 27.04.2024

**DOI:** 10.24075/vrgmu.2024.013

## DETERMINING OPTIMAL AMBIENT IONIZATION MASS SPECTROMETRY DATA PRE-PROCESSING PARAMETERS IN NEUROSURGERY

Zavorotnyuk DS<sup>1</sup>✉, Sorokin AA<sup>1</sup>, Bormotov DS<sup>1</sup>, Eliferov VA<sup>1</sup>, Bocharov KV<sup>2</sup>, Pekov SI<sup>1,3,4</sup>, Popov IA<sup>1,4</sup>

<sup>1</sup> Moscow Institute of Physics and Technology, Moscow, Russia

<sup>2</sup> Semenov Federal Research Center for Chemical Physics of the Russian Academy of Sciences, Moscow, Russia

<sup>3</sup> Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>4</sup> Siberian State Medical University, Tomsk, Russia

Radical tumor resection is still the most effective treatment method for brain tumors. The problems of intraoperative monitoring are currently solved using positron emission tomography, magnetic resonance imaging, and histochemical analysis, however, these require using expensive equipment by highly qualified personnel and are therefore still not widely available. As an alternative, it is possible to use mass spectrometry methods without sample preparation and then the analysis of mass spectrometry data involving the use of machine learning methods. The spectra that are more rich and diverse in terms of peak number are typical for mass spectrometry without sample preparation, therefore the use of this method requires specific pre-processing of experimental data. The study was aimed to develop the methods to determine the optimal parameter values for pre-processing of the data acquired by ambient ionization mass spectrometry. The paper presents two such methods and provides specific parameter values for the data acquired using the Thermo LTQ XL Orbitrap ETD mass spectrometer.

**Keywords:** mass spectrometry, ambient ionization, data analysis, data preprocessing

**Funding:** the study was performed within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (agreement № 075-03-2022-107, project № 0714-2020-0006). The study involved the use of equipment of the Semenov Federal Research Center for Chemical Physics RAS.

**Author contribution:** Zavorotnyuk DS — data acquisition and interpretation, software development, manuscript writing and editing; Sorokin AA — study planning, data analysis and interpretation, manuscript editing; Bormotov DS — data acquisition and interpretation, manuscript writing; Eliferov VA — financial support of the experiment; Bocharov KV — data acquisition; Pekov SI — study planning, data analysis and interpretation, manuscript draft writing and manuscript text finalization; Popov IA — project management, financial support.

**Compliance with ethical standards:** the study was approved by the Ethics Committee of the Burdenko Research Institute of Neurosurgery (protocols № 40 dated 12 April 2016 and № 131 dated 17 July 2018) and conducted in accordance with the principles of the Declaration of Helsinki (2000) and its subsequent revisions. All patients submitted the informed consent to study participation and the use of biomaterial for scientific purposes.

✉ **Correspondence should be addressed:** Denis S. Zavorotnyuk  
Institutskiy per., 9, str. 7, Dolgoprudny, Moscow Region, 141701; denis.zavorotnyuk@gmail.com

**Received:** 19.12.2023 **Accepted:** 03.03.2024 **Published online:** 27.04.2024

**DOI:** 10.24075/brsmu.2024.013

Масс-спектрометрия с прямой ионизацией — один из перспективных методов повышения точности и полноты резекции глиальных опухолей, поскольку радикальное удаление опухоли в настоящий момент является наиболее эффективным лечением онкологических заболеваний головного мозга [1]. Однако при этом возникает задача определения границ опухоли для обеспечения полноты резекции с целью предотвращения рецидива, с одной стороны, и для недопущения избыточной резекции и возникновения нейрпатологических последствий — с другой [2]. Основными универсальными интраоперационными методами контроля границ удаляемой опухоли до сих пор остаются позитронно-эмиссионная томография с применением компьютерной томографии (ПЭТ-КТ), магнитно-резонансная томография (МРТ) и гистохимический анализ, поскольку иные методы, например флуоресцентное окрашивание, могут оказаться неспецифичными для ряда диагнозов. Однако эти методы времязатратны, а томографические к тому же отличаются высокой стоимостью из-за необходимости оборудования специальных операционных блоков [3].

Масс-спектрометрия (МС) с прямой ионизацией позволяет за короткое время получить данные о молекулярном строении образца [4–6]. Однако на текущий момент подавляющее большинство вычислительных инструментов для работы с масс-спектрометрическими данными включают работу со спектрами, полученными с помощью МС в тандеме с газовой или жидкостной хроматографией. Эти данные отличаются тем, что количество пиков в каждом скане такого спектра намного меньше, чем в сканах, полученных МС с прямой ионизацией [7, 8]. Простота в подготовке образца и скорость анализа в случае МС с прямой ионизацией позволяют получить значительно более сложные масс-спектры, т. е. большее количество данных за срок, исчисляемый минутами. В то же время для анализа таких данных необходимо применять автоматизированные методы обработки и сложные алгоритмы анализа [9–11], поэтому большое внимание должно быть уделено контролю качества данных и их предварительной обработке [12].

Масс-спектрометрические данные представляют собой упорядоченные по времени наборы сканов. Каждый скан представляет собой упорядоченный по шкале отношения массы иона к его заряду ( $m/z$ ) профиль интенсивностей тока ионов, накопленных прибором за определенный интервал времени. На этапе предварительной обработки необходимо преобразовать этот скан в набор, состоящий из интенсивностей и значений  $m/z$  пиков, выделенных в скане. Обычно для этого выполняют такие действия, как нормализация значений интенсивностей, определение и удаление шума, определение и выравнивание положения пиков [13–15]. Большое разнообразие в подходах к предварительной обработке данных МС говорит о том, что в зависимости от природы образцов, использованных в исследовании, конструкции масс-спектрометра, режима сбора ионов и типа дальнейшего анализа перечисленные выше действия должны выполняться с различными параметрами.

В данной статье описана разработка методики определения параметров предварительной обработки масс-спектров с целью унификации масс-спектрометрических данных для дальнейшего автоматизированного анализа на примере экспериментальных данных масс-спектрометрии без пробоподготовки, полученных в ходе исследования образцов тканей опухолей головного мозга человека.

## МАТЕРИАЛЫ И МЕТОДЫ

В работе использовали масс-спектрометрические данные, полученные при обработке образцов тканей головного мозга человека с диагнозами глиобластома и астроцитомы IV степени злокачественности (согласно классификации ВОЗ 2021 г., [16]) и патологии неопухоловой природы, полученных в ходе хирургического лечения лекарственно-резистентной эпилепсии. Всего было исследовано 307 образцов тканей 74 пациентов. Данные были получены с помощью масс-спектрометра Thermo LTQ XL Orbitrap ETD (Thermo Fisher Scientific; США) с картриджной ионизацией [3, 17]. Каждый образец был разделен на две части, одну часть отправляли на стандартный гистохимический анализ для получения медицинского заключения по этому образцу, а из оставшейся части выделяли три фрагмента объемом примерно по 1 мм<sup>3</sup>, каждый из которых подвергали масс-спектрометрическому исследованию. Протокол масс-спектрометрического исследования включает анализ и детектирование ионов в восьми разных режимах, каждый из которых характеризуется полярностью ионов, разрешением детектора и шириной диапазона значений  $m/z$  регистрируемых ионов. Сбор ионов в каждом режиме выполняли по два раза.

Накопленные экспериментальные данные подвергали процедуре предварительной обработки с различными значениями параметров, описание которых дано в разделе «Результаты исследования». Процедура предварительной обработки заключалась в калибровке интенсивностей пиков, деформации пиков по отношению к скану с максимальным общим ионным током (TIC), взаимной деформации всех пиков среди сканов одного режима регистрации ионов и фильтрации редких и низкоинтенсивных пиков. Для каждого режима регистрации ионов были получены отдельные наборы сканов. Каждый набор сканов преобразовывался в матрицу интенсивностей пиков, которая использовалась для обучения классификационной модели. В качестве предикторов при обучении моделей выступали столбцы матрицы, содержащие распределения интенсивностей пиков среди всех сканов данного режима, а в качестве отклика были взяты гистологические диагнозы пациентов. Для обучения и проверки моделей были использованы масс-спектрометрические данные, полученные с образцов тканей головного мозга 33 пациентов с диагнозами глиобластома и семи пациентов с диагнозами патологий неопухоловой природы. Доступный для каждого режима набор данных подвергали разделению на тренировочную и проверочную группы в соотношении 3 : 1 соответственно, при этом деление выполняли таким образом, чтобы разные сканы, полученные от одного образца, присутствовали в обеих группах для снижения степени переобучения моделей.

Анализ данных выполняли на компьютере под управлением ОС Ubuntu 16.04 с установленным пакетом R версии 3.4.4 и R-пакетами MALDIquant [18], caret [19], glmnet [20], ggplot2 [21], и для этого полученные от масс-спектрометра данные были преобразованы из исходного формата Thermo Finnigan в открытый NetCDF [22] формат с помощью разработанного в лаборатории программного обеспечения [23].

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В 2012 г. было показано, что различия в масс-спектрах опухолевых и неопухоловых тканей головного мозга могут

быть использованы для построения классификаторов, автоматизированно определяющих присутствие злокачественной ткани в биопсийном материале [24]. На рис. 1 показаны пики двух масс-сканов, полученных с образцов тканей пациентов с диагнозами глиобластома и патология неопухолевой природы.

Процедура предварительной обработки масс-спектрометрических данных состоит из нескольких этапов. На первом этапе производят оценку шумового сигнала и определение отношения «сигнал/шум» для всех сканов:

$$SNR = \frac{I_s}{I_n},$$

где  $I_s$  — интенсивность сигнала,  $I_n$  — интенсивность шума. Существует несколько способов определения интенсивности шумового сигнала в цифровых данных, например, с помощью среднего абсолютного отклонения (MAD) или с помощью регрессии с адаптивной шириной полосы (SuperSmoother) [25]. На последующих этапах низкоинтенсивные пики, для которых отношение «сигнал/шум» меньше, чем заданное значение SNR, будут исключены из спектра. Под действием переменных факторов окружающей среды и случайных флуктуаций положения максимумов в скане могут слегка изменяться, и для того, чтобы компенсировать такие изменения, следующим шагом производят выравнивание профилей в разных сканах. В качестве референсного используют скан с максимальным TIC, так как предполагается, что этот

скан имеет наибольшее зарегистрированное количество ионов, его профиль содержит наибольшее количество различных ионных пиков. Здесь каждый профиль подвергается деформации вдоль оси  $m/z$ , так чтобы быть максимально похожим на референсный профиль. Максимально допустимая величина такой деформации задается с использованием предельного допуска на отклонение (TA). Затем производится детектирование пиков — преобразование профиля скана в набор отдельных пиков. Для этого весь профиль разбивают на несколько частей, размер каждой части определяют параметром «полуширина окна» (HWS) — это диапазон точек  $m/z$ , внутри которого производится поиск точки с максимальным значением интенсивности. Эта точка назначается пиком в данной части профиля. Затем производится выравнивание положения идентичных пиков среди всего набора сканов. Здесь идентичными считают пики, положения  $m/z$  которых различаются не больше, чем заданный допуск на отклонение при детектировании пиков (TBP). На заключительном этапе производят удаление редких пиков и объединение пиков из всех сканов в общую матрицу интенсивностей.

Таким образом, в результате предварительной обработки масс-спектрометрических данных образуется матрица [26], число строк в которой определено количеством сканов, полученных в ходе эксперимента, а число колонок — это объединенное количество пиков из всех сканов. Очевидно, что описанные выше параметры (SNR, TA, HWS

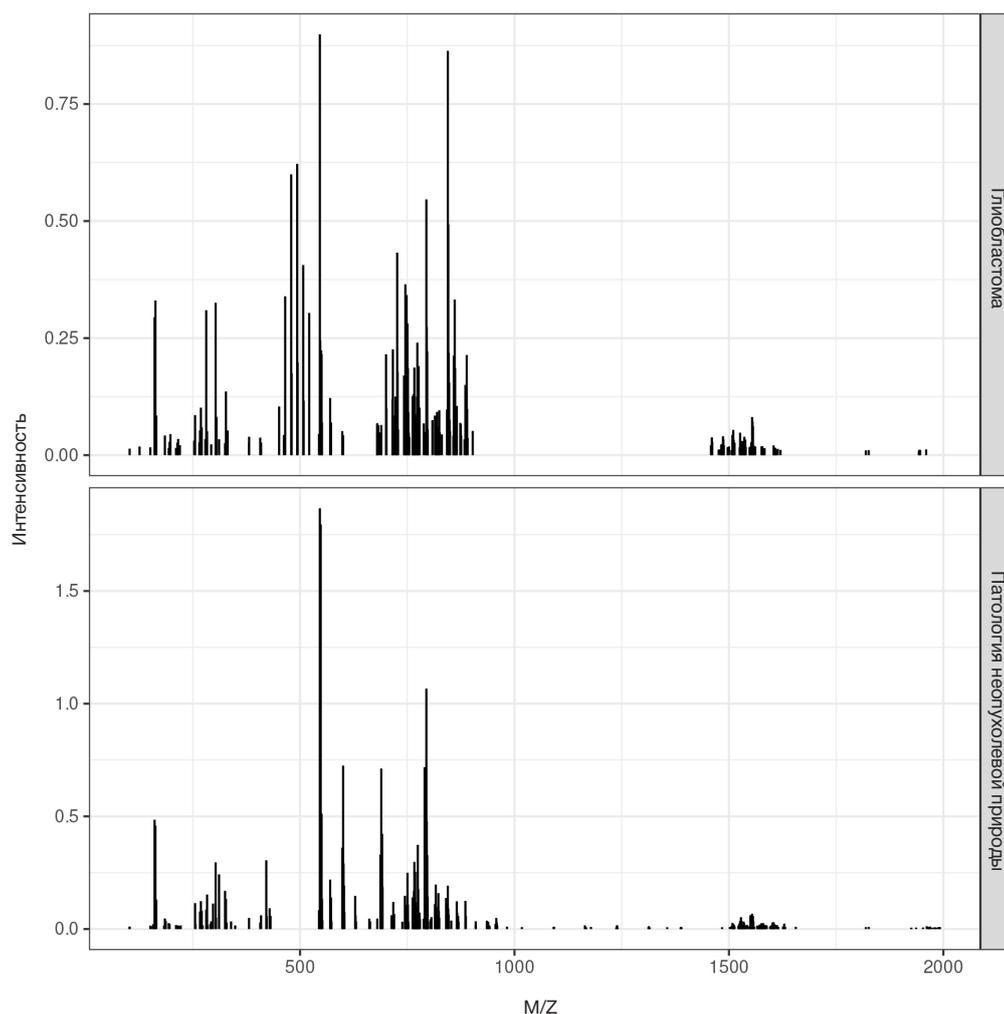


Рис. 1. Сравнение пиков масс-сканов опухолевых и неопухолевых образцов

Таблица 1. Оптимальные значения SNR, соответствующие LASSO-моделям с минимальным значением AIC

Режим сканирования	SNR	TA = TBP, ppm
Negative, High, 120–2000	1,5	20
Negative, High, 500–1000	2	2000
Negative, Low, 120–2000	1,5	20
Negative, Low, 500–1000	2	2000
Positive, High, 120–2000	1,5	2000
Positive, High, 500–1000	2	2000
Positive, Low, 120–2000	1,5	20
Positive, Low, 500–1000	2	2000

и TBP) существенным образом влияют на количество пиков в матрице интенсивностей и вопрос, какие именно значения они должны принимать в каждом конкретном режиме сбора ионов, не является тривиальным.

В классических задачах определения модели, которая наилучшим образом описывает экспериментальные данные [27, 28], с использованием метода регуляризации и набора критериев построения модели определяются информационные критерии [29], экстремальные значения которых соответствуют оптимальным значениям этих критериев. В нашей работе минимальное значение классического информационного критерия Акаике (AIC) [30] было использовано для определения оптимального значения SNR. Оптимальность остальных параметров, а именно HWS, TA и TBP, была определена по экспертной оценке качества обработки спектров.

## Параметр SNR

Определение оптимального значения параметра SNR выполняли с использованием критерия Акаике классификационных LASSO-моделей. Для этого составляли комбинацию значений параметров SNR, TA и TBP, выполняли предварительную обработку масс-спектров, строили матрицу интенсивностей, а затем тренировали LASSO-модель, где в качестве тренировочных данных использовали эту матрицу и диагноз пациента. Тренировку моделей производили с кросс-валидацией 5/10, выбор наилучшей модели осуществляли по метрике «Точность».

Комбинации параметров составляли из множеств значений:

$$\text{SNR} = \{1.5, 2\}$$

$$\text{TA} = \text{TBP} = \{20, 200, 2000\}$$

Оптимальной комбинацией параметров назначали ту, при которой полученная модель обладала наименьшим значением AIC. Оптимальные значения параметров представлены в табл. 1.

Для предотвращения появления отрицательных интенсивностей шумового сигнала в скане к набору точек (M/Z, Интенсивность) слева и справа были добавлены по 100 нулевых точек, в результате чего оценку шумового сигнала производили на расширенном диапазоне значений M/Z при неизменном количестве значимых пиков в спектре.

## Параметры HWS, TA, TBP

Оптимальность параметров HWS, TA и TBP определяли путем проведения экспертной оценки качества обработки спектров. Для этой цели было разработано интерактивное Shiny-приложение Mass-spectrum observer, которое позволяет исследовать, как изменяются форма спектра, положения пиков и характеристики матрицы интенсивностей определенного масс-скана при варьировании значений этих параметров. Исходный код приложения доступен в GitHub-репозитории [31], демонстрационная версия приложения доступна в библиотеке Shiny-приложений

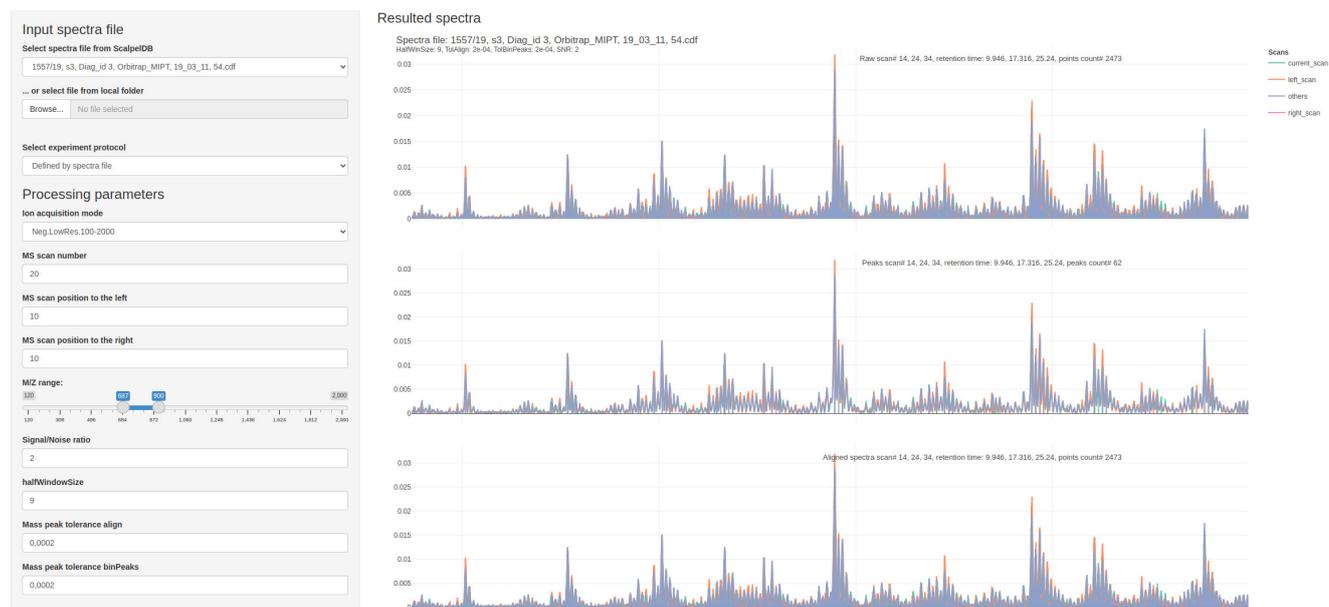


Рис. 2. Принт-скрин окна приложения Mass-spectrum observer с панелью управления параметрами предварительной обработки спектров

Таблица 2. Списки возможных значений параметров HWS, TA, TBP

Параметр	Значения для высокого разрешения	Значения для низкого разрешения
HWS	{3, 5, 7}	{7, 9, 11, 13, 15, 17, 19}
TA, ppm	{1, 20.8, 40.6, 60.4, 80.2, 100, 208, 406, 604, 802, 1·10 <sup>3</sup> }	{100, 325, 550, 775, 1·10 <sup>3</sup> }
TBP = m·TA	m := {0.1, 1, 10}	

со свободным доступом [32]. На рис. 2 и 3 представлены принт-скрины приложения.

Для параметров HWS, TA и TBP были определены списки возможных значений и для каждой комбинации этих значений проведены процедуры предварительной обработки масс-спектрометрических данных до получения матриц интенсивностей для каждого режима сбора ионов отдельно. Параметр TBP изменялся пропорционально параметру TA с тремя возможными значениями коэффициента пропорциональности. В табл. 2 приведены списки значений параметров.

Для каждой полученной матрицы интенсивностей было определено количество колонок, которое соответствует совокупному количеству пиков, полученных из профилей масс-сканов. Кроме того, в процессе построения матрицы интенсивностей производили определение количества близко расположенных друг к другу пиков в результирующих спектрах. Если расстояние между пиками оказывалось меньше, чем два разрешения прибора при детектировании ионов в данном режиме, то пики рассматривают как возможно дублирующиеся. Такие пики могут возникать в процессе преобразования профилей сканов в наборы отдельных пиков, например, в одном скане при слишком низких значениях параметра HWS, в результате чего относительно широкий по шкале m/z всплеск интенсивности будет представлен несколькими пиками спектра, или в сканах одного файла при низких значениях параметра TBP, из-за чего алгоритм не может составить список идентичных пиков в разных сканах. Дублирующиеся пики определяли внутри одного скана, во всех сканах одного фрагмента ткани, использованного в масс-спектрометрическом исследовании, и среди всех пиков матрицы интенсивностей. Дублирование пиков определяли в зависимости от разрешения масс-

спектрометра при данном режиме сбора ионов, для режима низкого разрешения было взято значение 800 при m/z = 400, для высокого разрешения — значение 30 000 при m/z = 400.

По изменениям этих четырех показателей в зависимости от параметров обработки были определены опорные значения параметров HWS, TA и TBP, которые затем проходили экспертную оценку с помощью Mass-spectrum observer. Результаты экспертной оценки представлены в табл. 3.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Результаты показывают тесную взаимосвязь между параметрами обработки масс-спектрометрических данных прямой ионизации с качеством получаемых спектров. Параметр SNR позволяет сократить число пиков в результирующем спектре, однако следует обращать внимание на наличие отрицательной оценки значений шумового сигнала, которая может возникать как артефакт в граничных областях спектров. В процессе детектирования пиков в профиле оценку шума используют для определения интенсивности пика на данном участке профиля, поэтому отрицательный шум может привести к появлению избыточного количества пиков в спектре. Это может не иметь большого значения в случае детектирования ионов в широком диапазоне M/Z, например, 120–2000, но для узкого диапазона 500–1000 это обстоятельство может быть существенным. В некоторых случаях эти артефакты удается устранить с помощью тонкой настройки метода SuperSmoother, например, путем изменения степени сглаженности при аппроксимации или путем сужения участка профиля, для которого производится оценка шума. Но для каждого отдельного масс-скана эти методы

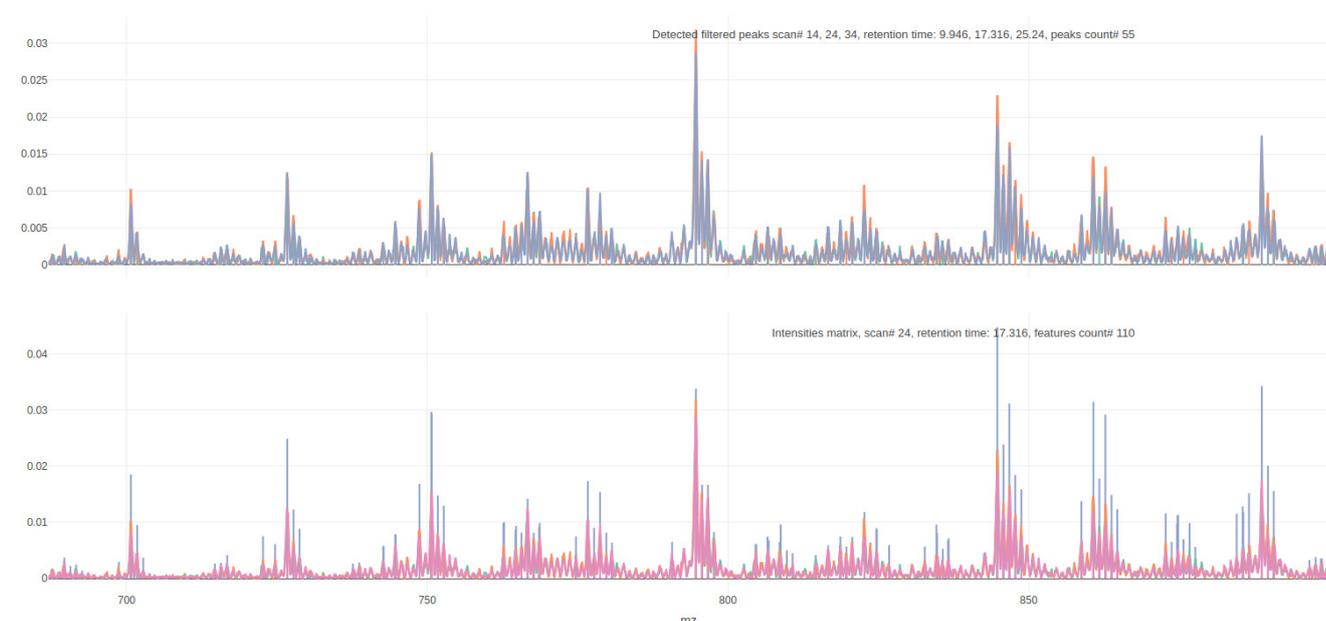


Рис. 3. Принт-скрин окна приложения Mass-spectrum observer с графиками, соответствующими спектрам после процедуры предварительной обработки

Таблица 3. Оптимальные параметры HWS, TA, TBP, полученные с помощью экспертной оценки

Режим накопления ионов	TA, ppm	TBP, ppm	HWS
Negative, High, 120–2000	40,6	40,6	3
Negative, High, 500–1000	60,4	60,4	3
Negative, Low, 120–2000	775	$7,75 \times 10^3$	13
Negative, Low, 500–1000	$1 \times 10^3$	$1 \times 10^3$	13
Positive, High, 120–2000	60,4	60,4	3
Positive, High, 500–1000	60,4	60,4	3
Positive, Low, 120–2000	$1 \times 10^3$	$1 \times 10^4$	13
Positive, Low, 500–1000	$1 \times 10^3$	$1 \times 10^4$	13

могут давать разные результаты, поэтому в качестве более устойчивого метода устранения отрицательных значений был выбран метод фиктивного расширения массива данных.

Выбор значений HWS, TA и TBP следует осуществлять, в первую очередь, с учетом разрешения прибора. Увеличение полуширины окна в процессе преобразования профиля в матрицу интенсивностей с одной стороны позволяет отсеивать артефактные и дублирующие пики (рис. 4), но с другой стороны слишком большие значения этого параметра приводят к исключению значимых пиков из последующего анализа (рис. 5). Значения допусков на отклонения в положениях пиков при выравнивании и детектировании также находятся в тесной связи с полушириной окна и, соответственно, с разрешением, а также с другими особенностями масс-спектрометра, которые обусловлены дрейфом масс и методами оцифровки сигнала. При этом значение TBP не должно быть меньше значения TA, поскольку такая конфигурация значений всегда приводит к увеличению среднего числа возможных дублирующих пиков. Это происходит из-за того, что алгоритму выравнивания пиков не хватает

допуска на сдвиг идентичных пиков в разных сканах даже после выравнивания всех сканов на скан с наибольшим ионным током, чтобы убрать дублирующиеся пики. Следует также отметить, что изменение ширины диапазона без изменения разрешения и полярности регистрируемых ионов не оказывает существенного влияния на значения параметров, что является ожидаемым результатом.

## ВЫВОДЫ

Разработан универсальный подход определения оптимальных значений параметров предварительной обработки данных, полученных с помощью МС с прямой ионизацией. Применение подхода продемонстрировано на данных, полученных с образцов тканей головного мозга человека с использованием масс-спектрометра Thermo LTQ XL Orbitrap ETD. Разработанный подход может быть использован для определения оптимальных значений параметров предварительной обработки данных, полученных при исследованиях других типов образцов и с применением другого масс-спектрометрического

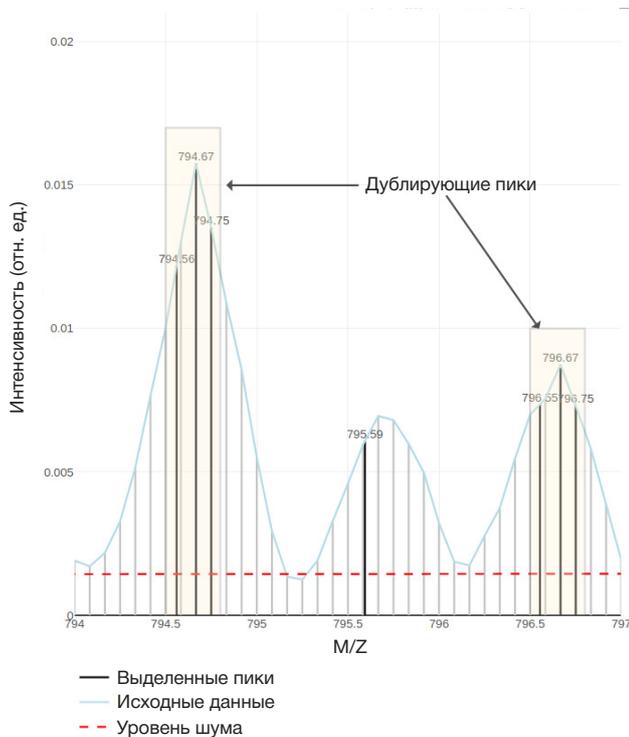


Рис. 4. Определение положения пиков. Появление пиков, расстояние между которыми меньше, чем два разрешения прибора при данном режиме регистрации ионов (дублирующие пики), в масс-скане отрицательных ионов широкого диапазона низкого разрешения при неоптимальных значениях параметров обработки

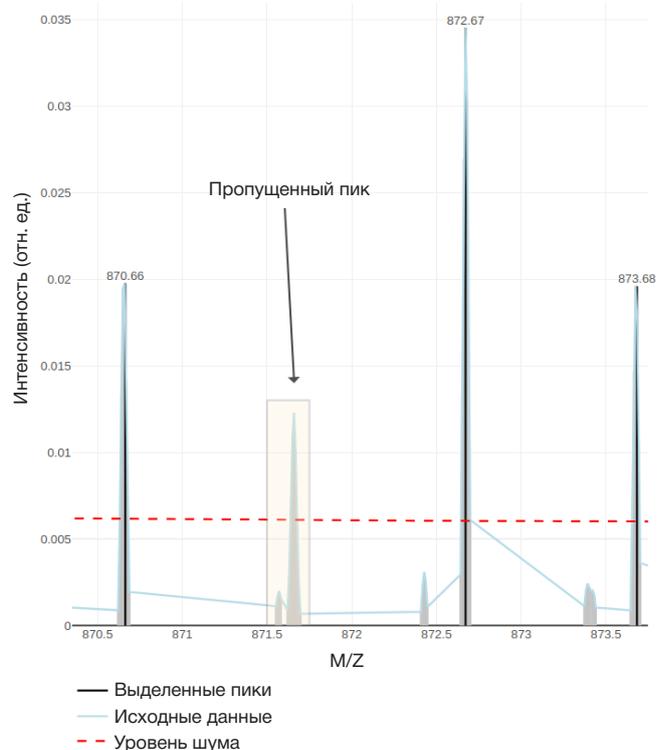


Рис. 5. Определение положения пиков. Пропуск значимого пика в масс-скане отрицательных ионов узкого диапазона высокого разрешения при неоптимальных значениях параметров обработки

оборудования. Результаты работы показывают, что при использовании МС с прямой ионизацией в клинике как быстрой и более доступной альтернативы традиционным методам интраоперационного мониторинга необходимо тщательно выработать настройку параметров обработки масс-спектрометрических данных. Параметры необходимо определять с учетом масс-спектрометра и условий проведения исследования. В частности, параметр SNR, определяющий число пиков в результирующих спектрах, следует выбирать, исходя из типа исследуемой ткани и способа ионизации, и значение в пределах 1,5–2 можно

принимать за нижнюю границу. При выравнивании профилей сканов и детектировании пиков значения полуширины окна (HWS) и допуска на модификацию скана (TA) следует выбирать в соответствии с разрешением используемого масс-спектрометра, а допуск на отклонение при выравнивании пиков спектров (TBP) не следует выбирать меньше, чем значение параметра TA. Для выбора оптимальных значений этих параметров среди нескольких возможных можно использовать как методы машинного обучения, так и экспертную оценку качества получаемых спектров.

## Литература

- Young RM, Jamshidi A, Davis G, Sherman JH. Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med* 2015; 1–15. <https://doi.org/10.3978/j.issn.2305-5839.2015.05.10>.
- Chanbour H, Chotai S. Review of intraoperative adjuncts for maximal safe resection of gliomas and its impact on outcomes. *Cancers*. 2022; 14: 5705. Available from: <https://doi.org/10.3390/cancers14225705>.
- Pekov SI, Bormotov DS, Nikitin PV, Sorokin AA, Shurkhay VA, Eliferov VA, et al. Rapid estimation of tumor cell percentage in brain tissue biopsy samples using inline cartridge extraction mass spectrometry. *Anal Bioanal Chem*. 2021; 413: 2913–22. Available from: <https://doi.org/10.1007/s00216-021-03220-y>.
- Eberlin LS, Norton I, Orringer D, Dunn IF, Liu X, Ide JL, et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc Natl Acad Sci*. 2013; 110: 1611–6. Available from: <https://doi.org/10.1073/pnas.1215687110>.
- Hänel L, Kwiatkowski M, Heikaus L, Schlüter H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Sci OA*. 2019; 5: FSO373. Available from: <https://doi.org/10.4155/fsoa-2018-0087>.
- Li L-H, Hsieh H-Y, Hsu C-C. Clinical application of ambient ionization mass spectrometry. *Mass Spectrom*. 2017; 6: S0060–S0060. Available from: <https://doi.org/10.5702/massspectrometry.S0060>.
- Huang M-Z, Yuan C-H, Cheng S-C, Cho Y-T, Shiea J. Ambient ionization mass spectrometry. *Annu Rev Anal Chem*. 2010; 3: 43–65. Available from: <https://doi.org/10.1146/annurev.anchem.111808.073702>.
- Shi L, Habib A, Bi L, Hong H, Begum R, Wen L. Ambient Ionization Mass Spectrometry: Application and Prospective. *Crit Rev Anal Chem*. 2022; 1–50. Available from: <https://doi.org/10.1080/10408347.2022.2124840>.
- Boiko DA, Kozlov KS, Burykina JV, Ilyushenkova VV, Ananikov VP. Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning. *J Am Chem Soc*. 2022; 144: 14590–606. Available from: <https://doi.org/10.1021/jacs.2c03631>.
- Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites*. 2020; 10: 1–23. Available from: <https://doi.org/10.3390/metabo10060243>.
- Piras C, Hale OJ, Reynolds CK, Jones AK (Barney), Taylor N, Morris M, et al. LAP-MALDI MS coupled with machine learning: an ambient mass spectrometry approach for high-throughput diagnostics. *Chem Sci*. 2022; 13: 1746–58. Available from: <https://doi.org/10.1039/D1SC05171G>.
- Seddiki K, Saudemont P, Precioso F, Ogrinc N, Wisztorski M, Salzet M, et al. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun*. 2020; 11. Available from: <https://doi.org/10.1038/s41467-020-19354-z>.
- Huang YC, Chung HH, Dutkiewicz EP, Chen CL, Hsieh HY, Chen BR, et al. Predicting breast cancer by paper spray ion mobility spectrometry mass spectrometry and machine learning. *Anal Chem*. 2020; 92: 1653–7. Available from: <https://doi.org/10.1021/acs.analchem.9b03966>.
- Iwano T, Yoshimura K, Inoue S, Odate T, Ogata K, Funatsu S, et al. Breast cancer diagnosis based on lipid profiling by probe electrospray ionization mass spectrometry. *Br J Surg*. 2020; 107: 632–5. Available from: <https://doi.org/10.1002/bjs.11613>.
- Zhou M, Guan W, Walker LDE, Mezencev R, Benigno BB, Gray A, et al. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol Biomarkers Prev*. 2010; 19: 2262–71. Available from: <https://doi.org/10.1158/1055-9965.EPI-10-0126>.
- Torp SH, Solheim O, Skjulsvik AJ. The WHO 2021 Classification of central nervous system tumours: a practical update on what neurosurgeons need to know — a minireview. *Acta Neurochir (Wien)*. 2022; 164: 2453–64. Available from: <https://doi.org/10.1007/s00701-022-05301-y>.
- Bormotov DS, Eliferov VA, Peregudova OV, Zavorotnyuk DS, Bocharov KV, Pekov SI, et al. Incorporation of a disposable ESI emitter into inline cartridge extraction mass spectrometry improves throughput and spectra stability. *J Am Soc Mass Spectrom*. 2023; 34: 119–22. Available from: <https://doi.org/10.1021/jasms.2c00207>.
- Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*. 2012; 28. Available from: <https://doi.org/10.1093/bioinformatics/bts447>.
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008; 28: 1–26. Available from: <https://doi.org/10.18637/jss.v028.i05>.
- Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010; 33: 1–22. Available from: <https://doi.org/10.18637/jss.v033.i01>.
- Wickham H. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York; 2016.
- Rew R, Davis G, Emmerson S, Cormack C, Caron J, Pincus R, et al. *Unidata NetCDF* 1989. Available from: <https://doi.org/10.5065/D6H70CW6>.
- Zavorotnyuk DS, Pekov SI, Sorokin AA, Bormotov DS, Levin N, Zhvansky E, et al. Lipid profiles of human brain tumors obtained by high-resolution negative mode ambient mass spectrometry. *Data*. 2021; 6: 1–7. Available from: <https://doi.org/10.3390/data6120132>.
- Eberlin LS, Norton I, Dill AL, Golby AJ, Ligon KL, Santagata S, et al. Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Research*. 2012; 72 (3): 645–54. Available from: <https://doi.org/10.1158/0008-5472.can-11-2465>.
- Friedman JH. *Smart user's guide*. Stanford Univ CA, Laboratory for Computational Statistics; 1984.
- Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*. 2005; 21: 1764–75. Available from: <https://doi.org/10.1093/bioinformatics/bti254>.
- Burnham KP, Anderson DR, editors. *Model Selection and Multimodel Inference*. New York, NY: Springer New York, 2004. Available from: <https://doi.org/10.1007/b97636>.
- Gustafsson F, Hjalmarsen H. Twenty-one ML estimators for model selection. *Automatica*. 1995; 31: 1377–92. Available from: [https://doi.org/10.1016/0005-1098\(95\)00058-5](https://doi.org/10.1016/0005-1098(95)00058-5).
- Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 2017. Доступна по ссылке: <https://github.com/ranalytics/data-mining>.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; 19: 716–23. Available from: <https://doi.org/10.1109/TAC.1974.1100705>.
- Zavorotnyuk DS. *MS Spectrum observer repository*. Available from: <https://github.com/zdens/MS-Spectrum-Observer/releases/tag/1.0>

(дата обращения: 29 февраля 2024 г.).

32. Zavorotnyuk DS. MS spectrum observer Demo. Available

from: <https://zdens.shinyapps.io/ms-spectrum-observer> (дата обращения: 29 февраля 2024 г.).

## References

- Young RM, Jamshidi A, Davis G, Sherman JH. Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med* 2015; 1–15. <https://doi.org/10.3978/j.issn.2305-5839.2015.05.10>.
- Chanbour H, Chotai S. Review of intraoperative adjuncts for maximal safe resection of gliomas and its impact on outcomes. *Cancers*. 2022; 14: 5705. Available from: <https://doi.org/10.3390/cancers14225705>.
- Pekov SI, Bormotov DS, Nikitin PV, Sorokin AA, Shurkhay VA, Eliferov VA, et al. Rapid estimation of tumor cell percentage in brain tissue biopsy samples using inline cartridge extraction mass spectrometry. *Anal Bioanal Chem*. 2021; 413: 2913–22. Available from: <https://doi.org/10.1007/s00216-021-03220-y>.
- Eberlin LS, Norton I, Orringer D, Dunn IF, Liu X, Ide JL, et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc Natl Acad Sci*. 2013; 110: 1611–6. Available from: <https://doi.org/10.1073/pnas.1215687110>.
- Hänel L, Kwiatkowski M, Heikaus L, Schlüter H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Sci OA*. 2019; 5: FSO373. Available from: <https://doi.org/10.4155/fsoa-2018-0087>.
- Li L-H, Hsieh H-Y, Hsu C-C. Clinical application of ambient ionization mass spectrometry. *Mass Spectrom*. 2017; 6: S0060–S0060. Available from: <https://doi.org/10.5702/massspectrometry.S0060>.
- Huang M-Z, Yuan C-H, Cheng S-C, Cho Y-T, Shiea J. Ambient ionization mass spectrometry. *Annu Rev Anal Chem*. 2010; 3: 43–65. Available from: <https://doi.org/10.1146/annurev.anchem.111808.073702>.
- Shi L, Habib A, Bi L, Hong H, Begum R, Wen L. Ambient Ionization Mass Spectrometry: Application and Prospective. *Crit Rev Anal Chem*. 2022; 1–50. Available from: <https://doi.org/10.1080/10408347.2022.2124840>.
- Boiko DA, Kozlov KS, Burykina JV, Ilyushenkova VV, Ananikov VP. Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning. *J Am Chem Soc*. 2022; 144: 14590–606. Available from: <https://doi.org/10.1021/jacs.2c03631>.
- Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites*. 2020; 10: 1–23. Available from: <https://doi.org/10.3390/metabo10060243>.
- Piras C, Hale OJ, Reynolds CK, Jones AK (Barney), Taylor N, Morris M, et al. LAP-MALDI MS coupled with machine learning: an ambient mass spectrometry approach for high-throughput diagnostics. *Chem Sci*. 2022; 13: 1746–58. Available from: <https://doi.org/10.1039/D1SC05171G>.
- Seddiki K, Saudemont P, Precioso F, Ogrinc N, Wisztorski M, Salzet M, et al. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun*. 2020; 11. Available from: <https://doi.org/10.1038/s41467-020-19354-z>.
- Huang YC, Chung HH, Dutkiewicz EP, Chen CL, Hsieh HY, Chen BR, et al. Predicting breast cancer by paper spray ion mobility spectrometry mass spectrometry and machine learning. *Anal Chem*. 2020; 92: 1653–7. Available from: <https://doi.org/10.1021/acs.analchem.9b03966>.
- Iwano T, Yoshimura K, Inoue S, Odate T, Ogata K, Funatsu S, et al. Breast cancer diagnosis based on lipid profiling by probe electrospray ionization mass spectrometry. *Br J Surg*. 2020; 107: 632–5. Available from: <https://doi.org/10.1002/bjs.11613>.
- Zhou M, Guan W, Walker LDE, Mezencev R, Benigno BB, Gray A, et al. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol Biomarkers Prev*. 2010; 19: 2262–71. Available from: <https://doi.org/10.1158/1055-9965.EPI-10-0126>.
- Torp SH, Solheim O, Skjulsvik AJ. The WHO 2021 Classification of central nervous system tumours: a practical update on what neurosurgeons need to know — a minireview. *Acta Neurochir (Wien)*. 2022; 164: 2453–64. Available from: <https://doi.org/10.1007/s00701-022-05301-y>.
- Bormotov DS, Eliferov VA, Peregudova OV, Zavorotnyuk DS, Bocharov KV, Pekov SI, et al. Incorporation of a disposable ESI emitter into inline cartridge extraction mass spectrometry improves throughput and spectra stability. *J Am Soc Mass Spectrom*. 2023; 34: 119–22. Available from: <https://doi.org/10.1021/jasms.2c00207>.
- Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*. 2012; 28. Available from: <https://doi.org/10.1093/bioinformatics/bts447>.
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008; 28: 1–26. Available from: <https://doi.org/10.18637/jss.v028.i05>.
- Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010; 33: 1–22. Available from: <https://doi.org/10.18637/jss.v033.i01>.
- Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York; 2016.
- Rew R, Davis G, Emmerson S, Cormack C, Caron J, Pincus R, et al. Unidata NetCDF 1989. Available from: <https://doi.org/10.5065/D6H70CW6>.
- Zavorotnyuk DS, Pekov SI, Sorokin AA, Bormotov DS, Levin N, Zhvansky E, et al. Lipid profiles of human brain tumors obtained by high-resolution negative mode ambient mass spectrometry. *Data*. 2021; 6: 1–7. Available from: <https://doi.org/10.3390/data6120132>.
- Eberlin LS, Norton I, Dill AL, Golby AJ, Ligon KL, Santagata S, et al. Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Research*. 2012; 72 (3): 645–54. Available from: <https://doi.org/10.1158/0008-5472.can-11-2465>.
- Friedman JH. Smart user's guide. Stanford Univ CA, Laboratory for Computational Statistics; 1984.
- Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*. 2005; 21: 1764–75. Available from: <https://doi.org/10.1093/bioinformatics/bti254>.
- Burnham KP, Anderson DR, editors. Model Selection and Multimodel Inference. New York, NY: Springer New York, 2004. Available from: <https://doi.org/10.1007/b97636>.
- Gustafsson F, Hjalmarsen H. Twenty-one ML estimators for model selection. *Automatica*. 1995; 31: 1377–92. Available from: [https://doi.org/10.1016/0005-1098\(95\)00058-5](https://doi.org/10.1016/0005-1098(95)00058-5).
- Shitikov VK, Mastitsky SE. Klassifikacija, regressija i drugie algoritmy Data Mining s ispol'zovaniem R. 2017. Dostupna po ssylke: <https://github.com/ranalytics/data-mining>. Russian.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; 19: 716–23. Available from: <https://doi.org/10.1109/TAC.1974.1100705>.
- Zavorotnyuk DS. MS Spectrum observer repository. Available from: <https://github.com/zdens/MS-Spectrum-Observer/releases/tag/1.0> (data obrashhenija: 29 fevralja 2024 g.).
- Zavorotnyuk DS. MS spectrum observer Demo. Available from: <https://zdens.shinyapps.io/ms-spectrum-observer> (data obrashhenija: 29 fevralja 2024 g.).