Статья поступила в редакцию: 25.09.2026
Статья принята к публикации: 30.09.2026
Статья опубликована: 07.10.2026

УДК 543.42+004.9+631.147

DOI: 10.32743/UniChem.2026.148.10.23482

Хемометрический анализ образцов мёда: современные аналитические платформы, многомерная классификация и машинное обучение

Исматова Нозима Рахматиллоевна
соискатель, PhD
Самаркандский государственный университет им. Ш. Рашидова
Республика Узбекистан, г. Самарканд
E-mail: nozimarahmatovna63@gmail.ru

Аронбаев Дмитрий Маркиэлович
канд. хим. наук, доц.
Самаркандский государственный университет им. Ш. Рашидова
Республика Узбекистан, г. Самарканд
E-mail: diron51@mail.ru

Аронбаев Сергей Дмитриевич
д-р хим. наук, проф.
Самаркандский государственный университет им. Ш. Рашидова
Республика Узбекистан, г. Самарканд

Chemometric analysis of honey samples: modern analytical platforms, multivariate classification, and machine learning

Ismatova Nozima
PhD Candidate
Samarkand State University named after Sharof Rashidov
Republic of Uzbekistan, Samarkand

Aronbaev Dmitry
candidate of Chemical Sciences, Associate Professor
Samarkand State University named after Sharof Rashidov
Republic of Uzbekistan, Samarkand

Aronbaev Sergey
doctor of Chemical Sciences, Professor
Samarkand State University named after Sharof Rashidov
Republic of Uzbekistan, Samarkand

Библиографическое описание: Исматова Н.Р., Аронбаев Д.М., Аронбаев С.Д. Хемометрический анализ образцов мёда: современные аналитические платформы, многомерная классификация и машинное обучение // Universum: химия и биология : электрон. научн. журн. 2026. 10(148). URL: https://7universum.com/ru/nature/archive/item/23482

Аннотация

Достоверная оценка подлинности мёда осложняется высокой природной вариабельностью его состава, обусловленной ботаническим и географическим происхождением, климатическими и сезонными факторами, технологией получения и условиями хранения. Перекрывание диапазонов содержания отдельных компонентов ограничивает информативность традиционного однофакторного контроля и определяет необходимость перехода к анализу многомерных химических профилей.

Цель. Систематизировать современные хемометрические подходы к аутентификации мёда, сопоставить аналитические возможности инструментальных платформ и методов многомерной классификации и сформулировать практические принципы построения воспроизводимой системы установления ботанического и географического происхождения и выявления фальсификации.

Материалы и методы. Обобщены преимущественно наиболее актуальные научные публикации 2023–2026 гг., отражающие современное состояние хемометрического анализа мёда. Рассмотрены FTIR/NIR, Raman, UV–Vis и флуоресцентная спектроскопия, ЯМР, HPLC/LC–MS, GC–MS, ICP–MS, LIBS и вольтамперометрический электронный язык в сочетании с PCA, HCA, PLS/PLS-DA, SIMCA/DD-SIMCA, SVM/SVR, Random Forest, нейросетевыми и ансамблевыми методами. Особое внимание уделено предобработке данных, отбору информативных признаков, предупреждению переобучения и независимой валидации моделей.

Результаты и научная новизна. На основе сравнительного анализа показано, что наибольший потенциал имеет не отдельный инструментальный метод, а интеграция взаимодополняющих спектральных, хроматографических, элементных и электрохимических профилей. Научная новизна обзора состоит в системном рассмотрении аналитической платформы, хемометрической обработки и процедуры валидации как единой системы аутентификации, а также в разработке концепции её адаптации к региональным особенностям мёда Центральной Азии.

Практическое значение. Предложена многоуровневая схема исследования мёда Узбекистана, предусматривающая создание региональной референсной библиотеки и сочетание доступных скрининговых методов UV–Vis, ATR-FTIR и вольтамперометрического профилирования с выборочным элементным анализом ICP-OES/ICP-MS и последующей интеграцией данных. Такой подход может стать основой национальной системы контроля подлинности, выявления фальсификации и подтверждения происхождения мёда.

Abstract

Reliable assessment of honey authenticity is complicated by the high natural variability of its composition, which is influenced by botanical and geographical origin, climatic and seasonal factors, production practices, and storage conditions. Overlapping concentration ranges of individual constituents limit the discriminatory power of conventional univariate analysis and highlight the need for approaches based on multidimensional chemical profiles.

Purpose. This review aims to systematize current chemometric approaches to honey authentication, compare the analytical capabilities of instrumental platforms and multivariate classification methods, and formulate practical principles for developing reproducible systems for determining botanical and geographical origin and detecting adulteration.

Materials and methods. The review focuses primarily on the most recent scientific literature published in 2023–2026, reflecting the current state of chemometric analysis of honey. FTIR/NIR, Raman, UV–Vis and fluorescence spectroscopy, NMR, HPLC/LC–MS, GC–MS, ICP–MS, LIBS, and voltammetric electronic tongue systems are considered in combination with PCA, HCA, PLS/PLS-DA, SIMCA/DD-SIMCA, SVM/SVR, Random Forest, neural-network, and ensemble approaches. Particular attention is given to data preprocessing, selection of informative features, prevention of overfitting, and independent model validation.

Results and scientific novelty. Comparative analysis demonstrates that the greatest potential lies not in any single instrumental technique, but in the integration of complementary spectral, chromatographic, elemental, and electrochemical profiles. The novelty of this review lies in considering the analytical platform, chemometric data processing, and validation strategy as interconnected components of a unified authentication system, and in developing a framework for adapting this approach to the regional characteristics of Central Asian honey.

Practical significance. A multilevel strategy is proposed for studying honey produced in Uzbekistan. It involves establishing a regional reference library and combining accessible screening techniques, including UV–Vis, ATR-FTIR, and voltammetric profiling, with selective ICP-OES/ICP-MS elemental analysis and subsequent data integration. This approach may provide a practical basis for a national system for honey authenticity control, adulteration detection, and verification of botanical and geographical origin.

Ключевые слова: мёд, хемометрия, аутентификация, фальсификация, PCA, PLS-DA, машинное обучение, FTIR, NIR, Raman, электронный язык, спектральный отпечаток.

Keywords: honey, chemometrics, authentication, adulteration, PCA, PLS-DA, machine learning, FTIR, NIR, Raman spectroscopy, electronic tongue, spectral fingerprinting.

1. Введение

Мёд относится к пищевым продуктам с высокой природной вариабельностью состава. Помимо преобладающих фруктозы и глюкозы, в нём присутствуют вода, дисахариды и олигосахариды, органические кислоты, аминокислоты, белки и ферменты, фенольные соединения, пигменты, летучие вещества и минеральные элементы. Их относительное содержание зависит от нектароносных растений, почвенно-климатических условий, сезона, зрелости мёда и технологии обработки. Поэтому два подлинных образца одного заявленного типа могут заметно различаться, а фальсифицированный продукт после целенаправленной коррекции отдельных физико-химических показателей может формально соответствовать нормативным диапазонам.

Проблема особенно значима для дорогих монофлорных и региональных сортов. Наиболее распространёнными нарушениями являются добавление дешёвых сахарных сиропов, смешивание мёда различного происхождения, неверное указание ботанического или

географического происхождения, интенсивное нагревание и длительное хранение. Современные обзоры подчёркивают, что надёжная аутентификация всё чаще строится не на одном маркере, а на совокупности аналитических сигналов и многомерном статистическом моделировании [14, 19, 34, 36].

Хемометрический подход принципиально отличается от традиционной интерпретации результата «показатель — норматив». Каждый образец рассматривается как точка в многомерном пространстве признаков. Такими признаками могут быть интенсивности поглощения на сотнях длин волн, токи на множестве потенциалов, площади хроматографических пиков, концентрации элементов или метаболитов. Задача хемометрии состоит в извлечении устойчивой структуры из этого массива, выявлении скрытых групп, построении классификационных границ и количественных прогнозных моделей.

Цель настоящего обзора — систематизировать современные направления хемометрического анализа мёда, показать аналитическую ценность различных инструментальных платформ, сопоставить алгоритмы классификации и регрессии и сформулировать практические рекомендации для построения воспроизводимой системы аутентификации образцов мёда, включая перспективы её применения к продукции Узбекистана и Центральной Азии.

2. Материалы и методы

Настоящая работа выполнена в формате аналитического обзора современных методов хемометрического анализа, аутентификации и выявления фальсификации мёда. Поиск научных публикаций проводили с использованием международных библиографических и реферативных баз данных Scopus, Web of Science Core Collection и PubMed. Для расширенного поиска релевантных публикаций и отслеживания связанных работ дополнительно использовали Google Scholar, а корректность библиографических метаданных и цифровых идентификаторов публикаций (DOI) при необходимости уточняли по данным Crossref.

Основное внимание уделяли публикациям 2023–2026 гг., характеризующим современный этап развития спектроскопических, хроматографических, масс-спектрометрических, элементных и электрохимических методов анализа мёда в сочетании с хемометрией и алгоритмами машинного обучения. Более ранние исследования включали в обзор в тех случаях, когда они имели существенное методологическое значение для рассматриваемого направления.

Поиск проводили по комбинациям англоязычных ключевых слов: honey, authentication, adulteration, chemometrics, machine learning, botanical origin, geographical origin, spectroscopy, FTIR, NIR, Raman, UV–Vis, fluorescence, NMR, HPLC, LC–MS, GC–MS, ICP–MS, LIBS, voltammetric electronic tongue, PCA, PLS-DA, SIMCA, DD-SIMCA, SVM и Random Forest. Ключевые слова объединяли с использованием логических операторов AND и OR.

При формировании литературной выборки приоритет отдавали оригинальным рецензируемым исследованиям и современным обзорным публикациям, посвящённым: (1) установлению ботанического и географического происхождения мёда; (2) выявлению фальсификации сахарными сиропами и смесями различных типов мёда; (3) получению

спектральных, хроматографических, элементных, метаболомных и электрохимических fingerprints; (4) применению методов многомерного статистического анализа и машинного обучения для классификации, моделирования класса и количественного прогнозирования.

В анализ включали публикации, содержащие достаточную информацию об используемой аналитической платформе, характере получаемых данных, применённых хемометрических алгоритмах и результатах классификации или количественного моделирования. Особое внимание уделяли исследованиям, в которых представлены сведения о предобработке данных, отборе информативных признаков, формировании обучающей и тестовой выборок, перекрёстной и/или внешней валидации моделей.

3. Результаты и обсуждение

Проведённый анализ литературы показывает, что современная аутентификация мёда развивается в направлении перехода от определения отдельных химических показателей к анализу многомерных аналитических fingerprints. При этом диагностическая способность метода определяется не только химической информативностью используемой аналитической платформы, но и способом предобработки данных, выбором хемометрического алгоритма, структурой обучающей выборки и корректностью валидации модели. Поэтому результаты анализа целесообразно рассматривать последовательно на трёх уровнях: формирование многомерного аналитического сигнала, извлечение из него классификационной или прогностической информации и оценка устойчивости полученной модели.

3.1. Многомерная структура аналитических данных и необходимость хемометрического подхода

В классическом лабораторном контроле измеряют влажность, электропроводность, свободную кислотность, содержание гидроксиметилфурфурола, диастазное число, пролин, редуцирующие сахара и сахарозу. Эти параметры необходимы для оценки качества, но их дискриминирующая способность в отношении происхождения ограничена. Для географической классификации турецкого соснового мёда, например, использовалась совокупность изотопных, мелиссопалинологических и физико-химических характеристик; информативность возникала именно за счёт многомерного сочетания показателей [35].

Математически набор данных можно представить матрицей X размерности $n \times p$, где n — число образцов, а p — число аналитических переменных. При спектроскопии p легко достигает сотен или тысяч, тогда как число независимых образцов часто ограничено десятками. Возникает типичная проблема $p \gg n$: высокая коллинеарность, шум, риск случайных корреляций и переобучения. Поэтому корректная стратегия должна включать предварительную обработку, снижение размерности или отбор признаков, независимую валидацию и оценку неопределённости.

Не менее важна биологическая структура данных. Образцы одного производителя, одной пасеки или одного года могут быть более похожи друг на друга, чем действительно независимые образцы. Если технические повторности одного и того же мёда случайно распределить одновременно в обучающую и тестовую выборки, модель получит завышенную точность. Разделение train/test желательно выполнять на уровне исходных образцов, партий, пасек или сезонов, а не отдельных спектров.

Общая последовательность этапов хемометрической аутентификации мёда — от формирования аналитического сигнала и предобработки данных до построения модели, её валидации и принятия классификационного решения — представлена на рисунке 1.

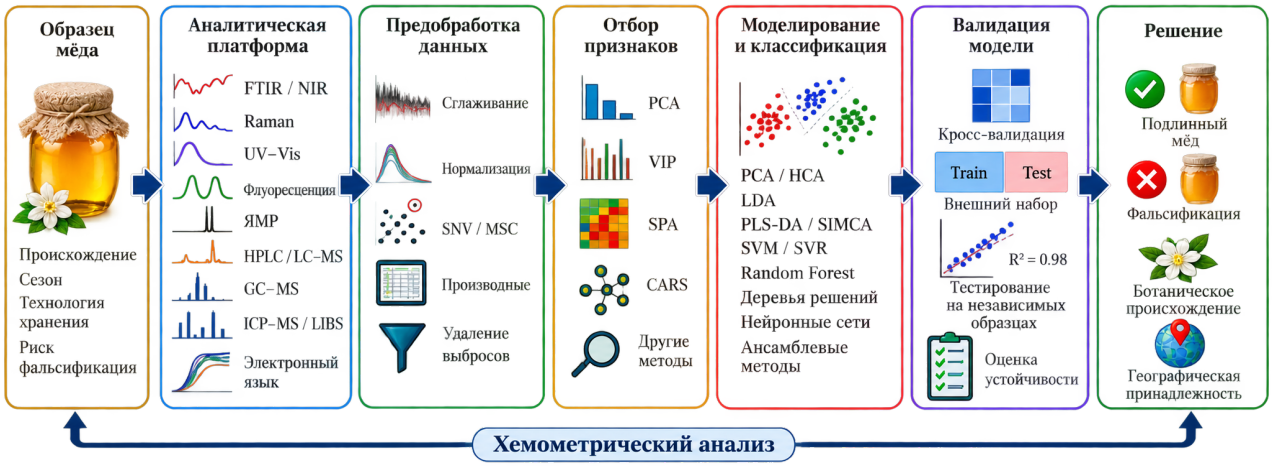


Рисунок 1. Основные этапы хемометрического анализа и аутентификации мёда

3.2. Хемометрические алгоритмы

Метод главных компонент (PCA) остаётся базовым инструментом первичной оценки структуры данных. Он преобразует исходные коррелированные переменные в ортогональные главные компоненты, последовательно объясняющие максимальную долю дисперсии. Для мёда PCA позволяет визуализировать кластеры ботанических типов, географических регионов, уровней фальсификации и технологических воздействий. Однако PCA является неконтролируемым методом: наличие разделения на графике scores не является доказательством высокой классификационной способности.

Иерархический кластерный анализ (HCA) полезен как независимое средство проверки естественного группирования. Результат зависит от масштабирования переменных, метрики расстояния и метода объединения кластеров. В задачах с переменными разных размерностей autoscaling обычно предпочтительнее, тогда как для спектральных данных часто используют mean-centering после специальной спектральной предобработки.

Partial Least Squares (PLS) широко используется для количественного прогнозирования доли примеси, содержания сахаров, влаги и других показателей. Для классификации применяется PLS-DA. В современных работах по мёду PLS-DA успешно сочеталась с HPLC-UV, Raman, UV–Vis, флуоресцентными и масс-спектрометрическими отпечатками [5, 9, 13, 25, 32].

SIMCA и DD-SIMCA относятся к class-modeling подходам. Это особенно важно для аутентификации: практический вопрос часто состоит не в том, к какому из нескольких известных классов отнести образец, а соответствует ли он заявленному классу вообще. Для средиземноморских медов DD-SIMCA показала высокую пригодность для ботанической и географической проверки по UV–Vis данным [11].

Support Vector Machine (SVM) эффективен при сложных нелинейных границах и сравнительно небольших выборках. Для количественных задач используется SVR. При UV–Vis определении сахарозной фальсификации SVR обеспечивал высокую точность прогноза и превосходил PLSR в конкретном наборе данных [31]. Random Forest устойчив

к нелинейности и взаимодействиям переменных, допускает оценку важности признаков и успешно применялся для FTIR, UV–Vis–NIR, ICP–MS и LIBS профилей [1, 6, 12, 18]. Нейронные сети и глубокое обучение становятся актуальными при больших массивах спектров или при объединении нескольких аналитических модальностей; ансамблирование и meta-learning показали высокую устойчивость при аутентификации мёда [24].

Сравнительная характеристика основных хемометрических методов, их аналитических задач, преимуществ и ограничений при исследовании мёда обобщена в таблице 1.

Таблица 1.

Основные хемометрические методы для анализа мёда

Метод	Задача	Преимущества	Ограничения
PCA	Разведочный анализ	Кластеры, выбросы, снижение размерности	Не подтверждает классификацию
PLS/PLS-DA	Регрессия/классификация	Коллинеарные данные	Риск переобучения
SIMCA/DD-SIMCA	Моделирование класса	Подлинность заявленного класса	Нужна репрезента-тивная библиотека
SVM/SVR	Классификация/регрессия	Нелинейные границы	Настройка гиперпараметров
Random Forest	Классификация/регрессия	Устойчивость, важность признаков	Дисбаланс классов
ANN/CNN	Нелинейное моделирование	Data fusion, сложные зависимости	Требует больше данных

3.3. FTIR и NIR спектроскопия

Инфракрасная спектроскопия является одним из наиболее активно развивающихся направлений быстрой аутентификации мёда. ATR-FTIR практически не требует пробоподготовки и формирует интегральный отпечаток, в котором доминируют колебания углеводов, воды, органических кислот и минорных органических компонентов. NIR менее специфична по отдельным полосам, но удобна для быстрого недеструктивного контроля и портативных измерений. Современные обзоры показывают, что именно сочетание IR-спектров с хемометрией превращает широкие перекрывающиеся полосы в информативный классификационный сигнал [34,4].

В работе по экономически мотивированной фальсификации клеверного мёда FTIR в сочетании с хемометрией позволила быстро выявлять добавки рисового и кукурузного сиропов; сообщалась специфичность около 96 % для исследованных уровней примеси [22]. ATR-FTIR совместно с алгоритмами машинного обучения применяли и для распознавания смесей мёда, когда монофлорный продукт фактически представлял смесь нескольких сортов [3].

Критически важна предварительная обработка спектров. Используются коррекция базовой линии, сглаживание Savitzky–Golay, первая и вторая производные, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), нормализация и mean-centering. Недавнее исследование показало, что выбор preprocessing существенно меняет результат ML-моделей для mid-FTIR обнаружения фальсификации [17]. Для региональной классификации FTIR особенно привлекателен благодаря низкой стоимости одного измерения: в 2025 г. для

средиземноморских медов показана классификация географического и ботанического происхождения с RF и DD-SIMCA, причём для большинства задач точность превышала 90 % [12].

3.4. Raman, UV–Vis и флуоресцентная спектроскопия

Raman-спектроскопия дополняет IR, поскольку чувствительна к иным типам молекулярных колебаний и допускает измерение через прозрачную упаковку. В 2024 г. FT-Raman в сочетании с PLS-DA использовали для различения ботанического происхождения и года сбора; после предобработки и отбора переменных правильность прогноза превышала 97 % [9]. Raman-подход применялся и для обнаружения сиропной фальсификации с использованием PCA и регрессионных моделей [20].

В 2025 г. объединение IR- и Raman-признаков с алгоритмами машинного обучения продемонстрировало дополнительную ценность data fusion [10].

UV–Vis спектроскопия привлекательна доступностью оборудования и высокой скоростью. Спектральные различия связаны с фенольными соединениями, продуктами реакции Майяра, пигментами и другими хромофорами. Для средиземноморских медов диапазон 220–550 нм в сочетании с PCA, RF, PLS-DA и DD-SIMCA позволял решать задачи происхождения [11]. Для бразильского полифлорного мёда UV–Vis и хемометрия применялись к выявлению добавки глюкозного сиропа [27]. В исследовании 2023 г. UV–Vis в сочетании с SVR и PLSR использовали для количественного прогнозирования сахарозы; SVR достиг R^2 около 0,98 [31]. В работе 2024 г. нейросетевая обработка UV–Vis данных позволяла обнаруживать фальсификацию различными сиропами при уровнях выше 10 % [16].

Флуоресцентные excitation–emission matrices (EEM) дают двумерный спектральный массив, чувствительный к ароматическим аминокислотам, фенольным веществам и продуктам преобразования сахаров. EEM в сочетании с SVM и PLS-DA успешно разделяли подлинные и фальсифицированные образцы [32]. Портативная LED-флуоресцентная система в 2024 г. была предложена для быстрой аутентификации stingless bee honey, демонстрируя тенденцию переноса хемометрии из стационарной лаборатории к полевому скринингу [33].

3.5. Хроматографические и масс-спектрометрические отпечатки

Нецелевой хроматографический fingerprinting позволяет использовать не только заранее известные маркеры, но и весь профиль образца. HPLC-UV в сочетании с PLS-DA обеспечивал 100 % классификацию исследованного набора мёда относительно различных сахарных сиропов и позволял количественно оценивать примесь, начиная примерно с 15 % [13]. Такой подход не требует полной идентификации каждого пика: для модели существенна воспроизводимая структура всего хроматографического отпечатка.

LC–MS ещё сильнее увеличивает размерность и химическую информативность. В 2025 г. была предложена robust multiclass feature-selection стратегия для нецелевого LC–MS анализа ботанического происхождения, специально учитывающая риск переобучения при тысячах признаков и ограниченном числе образцов [7]. В 2026 г. LC-LRMS fingerprints

и иерархическая последовательность парных PLS-DA моделей применялись для классификации мёда из десяти стран; для большинства сравнений получены высокие чувствительность и специфичность [25].

Летучий профиль, полученный HS-SPME-GC–MS, содержит важные ботанические маркеры. Для монофлорных медов Казахстана GC–MS совместно с хемометрией позволила выделить общие и дифференцирующие ароматические соединения, что особенно интересно для Центральной Азии [28]. Быстрый газохроматографический fingerprinting с ML также использовался для обнаружения фальсификации [30].

7. Ядерный магнитный резонанс (ЯМР), масс-спектрометрия с индуктивно связанной плазмой (ICP–MS) и лазерно-индуцированная эмиссионная спектроскопия (LIBS)

ЯМР сочетает воспроизводимость, количественность и возможность нецелевого профилирования. Анализ 424 образцов канадского мёда методом количественного ЯМР с PCA, PLS-DA и SIMCA показал возможность отделения канадских образцов от зарубежных и выявил потенциальные дискриминирующие метаболиты, включая тригонеллин и пролин [5]. Появление benchtop NMR снижает инфраструктурный барьер; компактные приборы уже комбинируются с хемометрией для скрининга сложных продуктов на основе мёда [29].

Минеральный профиль тесно связан с почвой, геологией, пылевым аэрозолем и растительным источником нектара. Для мёда из Мендосы в 2025 г. определяли 26 элементов методом ICP–MS и применяли деревья решений, RF и C5.0; происхождение удавалось прогнозировать с использованием ограниченного набора наиболее информативных элементов [6].

LIBS предлагает быстрый элементный fingerprint без сложной пробоподготовки. В 2025 г. LIBS и машинное обучение применены для классификации восьми ботанических типов греческого мёда, при этом RF использовали и для определения наиболее значимых спектральных признаков [18]. В более широком контексте LIBS рассматривается как перспективная технология аутентификации пищевых продуктов, особенно при сочетании с ML [26].

3.6. Электрохимические отпечатки и электронный язык

Электрохимический анализ интересен тем, что мёд содержит множество электроактивных соединений — фенольные кислоты, флавоноиды, продукты окисления и другие восстановители. Вольтамперометрический электронный язык не стремится селективно определить каждое вещество. Вместо этого массив электродов формирует многомерный электрохимический fingerprint, который затем анализируется PCA, LDA, SVM, PLS или другими алгоритмами.

Вольтамперометрический электронный язык был сопоставлен с 1H NMR при оценке фальсификации монофлорных медов сиропами; PLS-модели обеих платформ демонстрировали среднюю ошибку менее 5 % в условиях данного эксперимента [23]. Исследования многоэлектродных вольтамперометрических данных также показали высокую классификационную способность SVM [8]. Для региональных лабораторий электронный язык представляет особый интерес: потенциостат и набор серийных или screen-printed электродов существенно дешевле LC–HRMS и ЯМР.

Однако для реальной аутентификации необходимо обеспечить межэлектродную воспроизводимость, контролировать fouling, стандартизовать время контакта, температуру, разбавление, pH и процедуру регенерации. Перспективна комбинация электрохимического fingerprint с UV–Vis или FTIR данными и последующий low-level либо mid-level data fusion.

Таким образом, современные методы аутентификации мёда охватывают широкий спектр аналитических платформ, формирующих взаимодополняющие спектральные, хроматографические, элементные, метаболомные и электрохимические профили; их взаимосвязь с хемометрической обработкой схематически представлена на рис. 2.

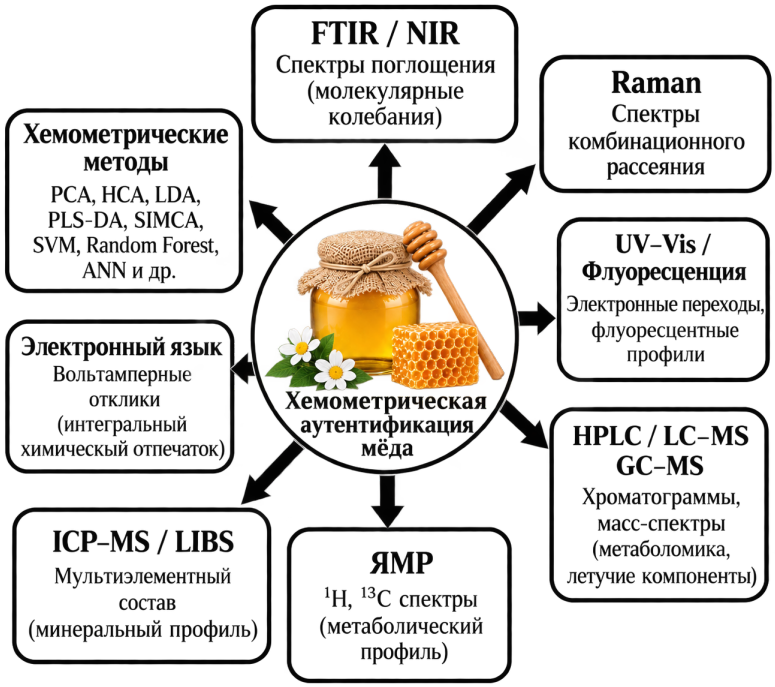


Рисунок 2. Современные аналитические платформы и типы данных, используемые для хемометрической аутентификации мёда

Основные типы данных, области применения и практические особенности рассмотренных аналитических платформ систематизированы в табл. 2.

Таблица 2.

Аналитические платформы и их роль в хемометрическом анализе мёда

Платформа	Тип данных	Основное применение	Практическая особенность
UV–Vis	Спектральный fingerprint	Сиропы, происхождение	Дёшево и быстро
FTIR/NIR	Вибрационный fingerprint	Фальсификация, сорт, происхождение	Минимальная пробоподготовка
Raman	Вибрационный fingerprint	Сорт, год, сиропы	Комплементарен IR
EEM fluorescence	2D спектры	Аутентификация	Высокая размерность
HPLC/LC–MS	Хроматографический/molecular fingerprint	Фальсификация, маркеры	Высокая химическая информативность
GC–MS	Летучие соединения	Ботаническое происхождение	Ароматические маркеры
NMR	Метаболомный fingerprint	Происхождение, подлинность	Референсный метод

Платформа	Тип данных	Основное применение	Практическая особенность
ICP–MS/LIBS	Элементный профиль	Географическое происхождение	Связь с геохимией
Электронный язык	Вольтамперометрический fingerprint	Скрининг, data fusion	Портативность

Для сопоставления современного состояния данного направления в табл. 3 обобщены наиболее показательные исследования 2023–2026 гг., различающиеся аналитической платформой, используемыми хемометрическими алгоритмами и решаемой задачей аутентификации

Таблица 3.

Современные исследования по хемометрической аутентификации мёда (2023–2026 гг.)

Год	Объект / задача	Аналитическая платформа	Хемометрия / ML	Основной результат	Ссылка
2023	Аутентификация и выявление фальсификации	EEM fluorescence	PLS-DA, SVM	Разделение подлинных и фальсифицированных образцов	[8]
2023	Прогнозирование фальсификации	UV–Vis	SVR, PLSR	SVR: $R^2 \approx 0{,}98$	[12]
2023	Сиропная фальсификация	FTIR	Chemometrics	Быстрый скрининг рисового/кукурузного сиропа	[19]
2023	Смеси различных типов мёда	ATR-FTIR	ML	Дискриминация смесей по спектральным fingerprints	[20]
2024	Фальсификация сахарными сиропами	HPLC-UV	PLS-DA	100% классификация исследованного набора; количественная оценка примеси	[6]
2024	Ботаническое происхождение и год сбора	FT-Raman	PLS-DA	Правильность прогноза >97%	[7]
2024	Географическое/ботаническое происхождение	UV–Vis	PCA, RF, PLS-DA, DD-SIMCA	Успешное моделирование происхождения средиземноморских медов	[11]
2024	Глюкозный сироп	UV–Vis	Chemometrics	Выявление фальсификации полифлорного мёда	[24]
2024	Различные сиропы	UV–Vis	Neural-network/chemometric analysis	Обнаружение фальсификации при уровнях >10%	[25]
2025	Географическое/ботаническое происхождение	FTIR	RF, DD-SIMCA	Для большинства задач точность >90%	[13]

Год	Объект / задача	Аналитическая платформа	Хемометрия / ML	Основной результат	Ссылка
2025	Региональная классификация	ICP-MS	Decision trees, RF, C5.0	Выделены информативные элементы для происхождения	[14]
2025	Ботаническое происхождение	LIBS	RF	Классификация восьми ботанических типов	[16]
2025	Ботаническое происхождение	LC-MS	Multiclass feature selection	Устойчивый отбор признаков при высокой размерности	[27]
2026	Географическое происхождение, 10 стран	LC-LRMS	Hierarchical PLS-DA	Высокие sensitivity и specificity для большинства сравнений	[9]

3.7. Предобработка, отбор признаков и валидация

Предобработка является частью модели и должна фиксироваться до внешней валидации. Для спектров типичны коррекция базовой линии, сглаживание, производные, SNV/MSC и центрирование. Для хроматографических данных добавляются выравнивание времени удерживания, нормализация интенсивности и обработка пропусков. Для ICP–MS необходимо учитывать значения ниже предела количественного определения и возможные batch-эффекты.

Отбор признаков уменьшает шум и повышает интерпретируемость, но при неправильной организации создаёт утечку информации. Любой supervised feature selection должен выполняться только внутри обучающей выборки или внутри каждого fold кросс-валидации. Использование всего набора, включая будущий test set, для выбора длин волн или метаболитов приводит к оптимистичной оценке точности. Современные методы включают VIP scores PLS, recursive feature elimination, genetic algorithms, permutation importance RF, SHAP-оценки и sparse-модели. Для IR-анализа фальсификации предложен генетический алгоритм отбора спектральных признаков [21], а для LC–MS — специализированная процедура устойчивого отбора [7].

Высокая accuracy сама по себе не доказывает пригодность метода. Для бинарной классификации следует приводить sensitivity, specificity, precision, recall, F1-score, ROC-AUC и confusion matrix. Для несбалансированных классов accuracy особенно опасна. Для количественных моделей необходимы R^2, RMSEC/RMSECV/RMSEP, MAE, bias и, при возможности, RPD.

Минимально приемлемая схема включает независимый test set, который не используется ни для выбора preprocessing, ни для отбора признаков, ни для настройки гиперпараметров. При небольшом n предпочтительна nested cross-validation. Если образцы собраны в разные годы, наиболее убедительной проверкой является temporal external validation: модель строится на одном или нескольких сезонах и проверяется на следующем.

С учётом рассмотренных источников основные методологические требования к формированию выборки, предобработке данных, построению и объективной валидации хемометрических моделей суммированы в табл. 4

Таблица 4.

Рекомендуемые требования к построению и валидации хемометрических моделей аутентификации мёда

Этап	Рекомендуемый подход	Что следует исключить	Основной критерий
Формирование выборки	Независимые партии, пасеки, регионы и сезоны	Использование технических повторностей как независимых образцов	Репрезентативность
Train/test split	Разделение на уровне исходных образцов/партий	Попадание одного образца одновременно в train и test	Независимость test set
Предобработка	Выбирать preprocessing только внутри CV	Оптимизация по всему dataset	Отсутствие data leakage
Отбор признаков	VIP, RFE, GA, permutation importance, sparse methods	Feature selection до разделения train/test	Стабильность признаков
Cross-validation	Stratified/grouped CV	Случайное смешивание зависимых образцов	Средняя метрика ± SD
Настройка модели	Nested cross-validation	Подбор гиперпараметров по test set	Unbiased performance
Классификация	Independent external test	Только accuracy	Sensitivity, specificity, precision, F1, ROC-AUC
Регрессия	Независимый prediction set	Только R^2 калибровки	R^2, RMSEP, MAE, bias, RPD
Проверка устойчивости	Permutation testing	Интерпретация случайного разделения как закономерности	p-value / распределение случайных моделей
Внешняя валидация	Другие партии, пасеки, сезоны/регионы	Только внутренняя CV	External predictive performance
Temporal validation	Обучение на предыдущих сезонах → следующий сезон	Смешивание всех лет	Устойчивость во времени
Переносимость	Межприборная/межлабораторная проверка	Один прибор и одна лаборатория	Robustness / calibration transfer

3.8. Практическая схема исследования мёда Узбекистана

Для мёда Узбекистана целесообразно создавать не единичную классификационную модель, а многоуровневую референсную библиотеку. Образцы должны собираться непосредственно у пасечников с документированием района, высоты, даты отбора, доминирующей растительности и технологической обработки. Для каждого заявленного ботанического типа желательно подтверждение мелиссопалинологией. Особую ценность могут представлять хлопковый, горный разнотравный, акациевый, люцерновый, подсолнечниковый и другие регионально характерные типы.

Оптимальный минимальный аналитический набор может включать стандартные физико-химические показатели, UV–Vis, ATR-FTIR, вольтамперометрический fingerprint на screen-printed электродах и элементный профиль ICP-OES/ICP-MS для поднабора образцов. Такая архитектура позволяет построить дешёвый скрининговый уровень и одновременно получить высокоинформативный референсный уровень.

На основе рассмотренных подходов для условий Узбекистана предлагается многоуровневая система, объединяющая доступные скрининговые методы с более информативным референсным уровнем и последующей хемометрической интеграцией данных (рис. 3)

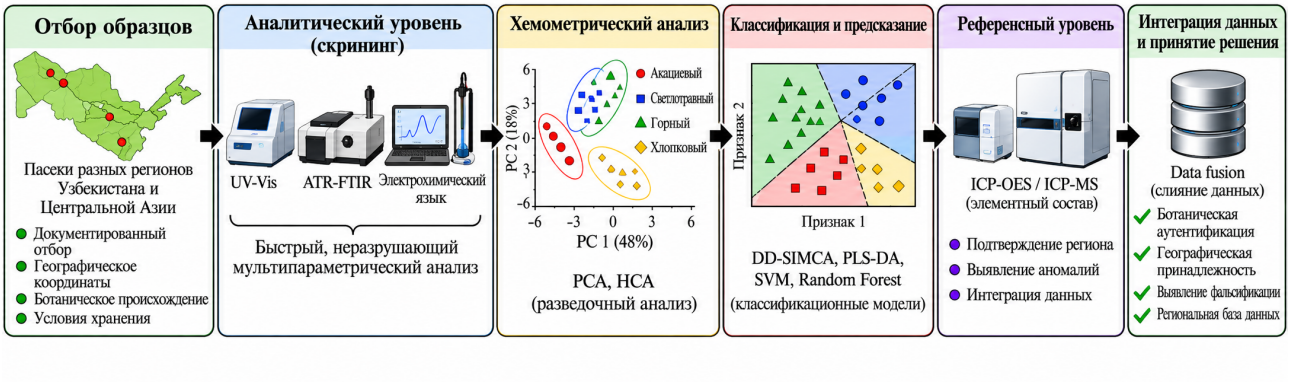


Рисунок 3. Предлагаемая многоуровневая система исследования и аутентификации мёда Узбекистана и Центральной Азии

Хемометрический конвейер обработки данных (chemometric pipeline) рекомендуется строить следующим образом: контроль качества данных, выявление и обоснованная оценка аномальных образцов (outliers); отдельный анализ методом главных компонент (PCA, Principal Component Analysis) для каждого блока данных; сравнение процедур предварительной обработки данных (preprocessing) только в рамках перекрёстной валидации (cross-validation); построение моделей методом дискриминантного анализа на основе частичных наименьших квадратов (PLS-DA, Partial Least Squares Discriminant Analysis), моделирования аналогий классов на основе расстояния до модели (DD-SIMCA, Data-Driven Soft Independent Modeling of Class Analogy), методом опорных векторов (SVM, Support Vector Machine) и случайного леса (RF, Random Forest); оценка сбалансированной точности (balanced accuracy), чувствительности (sensitivity), специфичности (specificity) и F1-меры (F1-score); проведение перестановочного теста (permutation test); проверка модели на внешней тестовой выборке (external test set), сформированной из образцов независимых пасек; объединение данных (data fusion) спектральных и электрохимических блоков; химическая интерпретация наиболее информативных признаков.

Для количественной модели фальсификации следует готовить смеси подлинного мёда с несколькими типами сиропов, например 1, 2.5, 5, 10, 20, 30 и 40 % (m/m), причём тестовые серии желательно готовить из иных партий мёда и сиропа, чем обучающие. Это позволит определить не только статистическую точность, но и реальный practical detection level.

Важным развитием будет моделирование неизвестного adulterant: модель должна уметь отклонить образец как «не соответствующий подлинному классу», даже если конкретный сироп отсутствовал в обучении.

Отдельная перспективная задача — географическая классификация долины Зарафшана и соседних регионов. Здесь можно ожидать вклад минерального профиля, особенностей растительности и фенольного состава. Однако доказательство географической специфичности потребует многолетнего отбора и независимой проверки, иначе сезонный эффект может быть ошибочно интерпретирован как региональный.

3.9. Тенденции 2025–2026 гг

Современная хемометрия мёда смещается от модели «один прибор — один классификатор» к объединению комплементарных данных. Low-level fusion объединяет исходные признаки нескольких платформ; mid-level fusion — отобранные признаки или scores; high-level fusion — решения отдельных моделей. Для небольших выборок mid-level подход часто наиболее рационален, поскольку ограничивает размерность.

Вторая тенденция — переход от традиционных линейных методов к нелинейным и ансамблевым алгоритмам. RF, SVM, XGBoost, нейронные сети и meta-learning способны давать очень высокие показатели, но без независимых партий такая точность может отражать особенности набора данных, а не универсальную закономерность [1,24,17,15].

Третья тенденция — портативность. NIR, UV–Vis, fluorescence, Raman, LIBS и электрохимические сенсоры постепенно переходят в компактный формат. В перспективе лабораторная референсная модель может использоваться для калибровки дешёвых полевых устройств. Ключевой проблемой становится calibration transfer между приборами, партиями сенсоров и лабораториями.

Наконец, формируется концепция молекулярной идентичности мёда, объединяющая метаболомику, спектроскопию, элементный анализ, ДНК-подходы и AI. Для реального внедрения необходимы гармонизированные протоколы отбора, открытые референсные библиотеки и межлабораторная проверка [19,2].

Заключение

Хемометрия превратилась из вспомогательной статистической процедуры в центральный компонент современной аутентификации мёда. Наиболее убедительные результаты достигаются тогда, когда аналитическая платформа генерирует богатый и воспроизводимый fingerprint, а математическая обработка строится с учётом структуры выборки, риска переобучения и необходимости внешней валидации.

FTIR/NIR, Raman, UV–Vis и флуоресцентная спектроскопия наиболее перспективны для быстрого рутинного скрининга; HPLC/LC–MS, GC–MS, ЯМР и ICP–MS — для построения референсных библиотек и поиска химических маркеров; LIBS и электронные языки — для развития портативных систем. PCA полезна для разведочного анализа, но окончательная аутентификация должна опираться на валидированные supervised или class-modeling алгоритмы.

Для Центральной Азии наиболее актуально создание собственных многолетних баз данных, поскольку модели, построенные на европейских или китайских образцах, нельзя автоматически переносить на узбекский мёд. Комбинация ATR-FTIR/UV–Vis, электрохимического fingerprint и выборочного элементного анализа с PCA, DD-SIMCA, PLS-DA, SVM и RF может стать реалистичной основой национальной системы скрининга подлинности и происхождения мёда.

Список литературы

1. Al Noman M.A., Barua Nijhum A., Hossain I., Islam M.S., Sifat I.M., Aziz M.G., Rahman A. Non-destructive adulterants detection in various honey types in Bangladesh using UV–VIS–NIR spectroscopy coupled with machine learning algorithms // LWT. – 2025. – T. 228. – Art. 118125. – DOI: 10.1016/j.lwt.2025.118125.
2. Álvarez-Suárez Дж. M., Majtan Дж., Tejera Э., Santos-Buelga K., González-Paramás A.M. Álvarez-Suárez J.M., Majtan J., Tejera E., Santos-Buelga C., González-Paramás A.M. The molecular identity of honey: toward reliable biochemical authentication // Trends in Food Science & Technology. – 2025. – T. 165. – Art. 105331. – DOI: 10.1016/j.tifs.2025.105331.
3. Berghian-Grosan C., Hategan A.R., David M., Magdas D.A. Untargeted metabolomic analysis of honey mixtures: Discrimination opportunities based on ATR-FTIR data and machine learning algorithms // Microchemical Journal. – 2023. – T. 188. – Art. 108458. – DOI: 10.1016/j.microc. 2023.108458.
4. Biswas A., Chaudhari S.R. Exploring the role of NIR spectroscopy in quantifying and verifying honey authenticity: A review // Food Chemistry. – 2024. – T. 445. – Art. 138712. – DOI: 10.1016/j.foodchem.2024.138712.
5. Burton I.W., Kompany-Zareh M., Haverstock S., Haché J., Martinez-Farina C.F., Wentzell P.D., Berrué F. Analysis and discrimination of Canadian honey using quantitative NMR and multivariate statistical methods // Molecules. – 2023. – T. 28, № 4. – Art. 1656. – DOI: 10.3390/ molecules28041656.
6. Canizo B.V., Diedrichs A.L., Fiorentini E.F., Brusa L., Sigrist M., Juricich J.M., Pellerano R.G., Wuilloud R.G. Intra-regional classification and quality evaluation of honey from Mendoza (Argentina) based on multi-elemental analysis and chemometrics // Journal of Food Composition and Analysis. – 2025. – T. 137. – Art. 106958. – DOI: 10.1016/j.jfca.2024.106958.
7. Chahal S., Tian L., Bilamjian S., Balogh F., De Leoz L., Anumol T., Cuthbertson D., Bayen S. Robust multiclass feature selection for the authentication of honey botanical origin via nontargeted LC-MS analysis // Analytical Chemistry. – 2025. – T. 97, № 24. – P. 12521–12530. – DOI: 10.1021/acs.analchem.4c06723.
8. Ciursa P., Oroian M. Voltammetric E-Tongue for honey adulteration detection // Sensors. – 2021. – T. 21, № 15. – Art. 5059. – DOI: 10.3390/s21155059.
9. David M., Magdas D.A. Authentication of honey origin and harvesting year based on Raman spectroscopy and chemometrics // Talanta Open. – 2024. – T. 10. – Art. 100342. – DOI: 10.1016/j.talo.2024.100342.

10. David M., Berghian-Grosan C., Magdas D.A. Honey differentiation using infrared and Raman spectroscopy analysis and machine-learning-based authentication models // Foods. – 2025. – T. 14, № 6. – Art. 1032. – DOI: 10.3390/foods14061032.
11. Dimakopoulou-Papazoglou D., Ploskas N., Koutsoumanis K., Katsanidis E. Identification of geographical and botanical origin of Mediterranean honeys using UV-vis spectroscopy and multivariate statistical analysis // Journal of Food Measurement and Characterization. – 2024. – T. 18. – P. 3923–3934. – DOI: 10.1007/s11694-024-02465-2.
12. Dimakopoulou-Papazoglou D., Serrano S., Rodríguez I., Ploskas N., Koutsoumanis K., Katsanidis E. FTIR spectroscopy combined with machine learning for the classification of Mediterranean honey based on origin // Journal of Food Composition and Analysis. – 2025. – T. 144. – Art. 107778. – DOI: 10.1016/j.jfca.2025.107778.
13. Egido C., Saurina J., Sentellas S., Núñez O. Honey fraud detection based on sugar syrup adulterations by HPLC-UV fingerprinting and chemometrics // Food Chemistry. – 2024. – T. 436. – Art. 137758. – DOI: 10.1016/j.foodchem.2023.137758.
14. En-Najy A., Kharbach M., Yu H., Vander Heyden Y., Bouklouze A. Honey authentication and adulteration assessments using spectroscopic techniques in tandem with chemometric tools // Journal of Food Composition and Analysis. – 2026. – T. 155. – Art. 109225. – DOI: 10.1016/j.jfca.2026.109225.
15. Ennahli S., Ajal E.A., Bajoub A., Hssaini L. Rapid prediction of honey adulteration using machine learning-assisted FTIR spectroscopy and chemometrics // Food Control. – 2026. – T. 187. – Art. 112146. – DOI: 10.1016/j.foodcont.2026.112146.
16. Geană E.I., Isopescu R., Ciucure C.T., Gîjiu C.L., Joșceanu A.M. Geană E.I., Isopescu R., Ciucure C.T., Gîjiu C.L., Joșceanu A.M. Honey adulteration detection via ultraviolet–visible spectral investigation coupled with chemometric analysis // Foods. – 2024. – T. 13, № 22. – Art. 3630. – DOI: 10.3390/foods13223630.
17. Hssaini L., Ennahli S. Preprocessing-dependent machine learning models enhance mid-FTIR detection of honey adulteration // Journal of Food Composition and Analysis. – 2025. – T. 148, Ч. 5. – Art. 108666. – DOI: 10.1016/j.jfca.2025.108666.
18. Kantemiris G., Kourelias P., Panagiotopoulou C., Stefas D., Couris S. Authentication of the botanical origin of honey via Laser Induced Breakdown Spectroscopy // Microchemical Journal. – 2025. – T. 212. – Art. 113537. – DOI: 10.1016/j.microc.2025.113537.
19. Khare T., Mosa K.A., Hamdy R., Elnaggar A., Malik S., Khan S.K., El-Keblawy A., Lamghari F., Alhmoudi A.M.S., Alyammahi K.M. Cutting-edge approaches for honey authentication: Chemical, molecular, and AI-driven strategies for botanical origin verification // Journal of Food Composition and Analysis. – 2025. – T. 146. – Art. 107974. – DOI: 10.1016/j.jfca.2025.107974.
20. Kou Z., Chen G., Li S., Yang Z., Ouyang L., Gong Y. Identification of honey adulterated with syrup by Raman spectroscopy and chemometrics // Food Science. – 2024. – T. 45, № 1. – P. 254–260. – DOI: 10.7506/spkx1002-6630-20230323-230.
21. Kumar N., Gupta K.K., Panchariya P.C. Infrared spectroscopy and genetic algorithm based spectral feature selection for rapid honey adulteration detection // Discover Food. – 2025. – T. 5, № 1. – Art. 380. – DOI: 10.1007/s44187-025-00619-9.

22. Limm W., Karunathilaka S.R., Mossoba M.M. Fourier Transform Infrared Spectroscopy and chemometrics for the rapid screening of economically motivated adulteration of honey spiked with corn or rice syrup // Journal of Food Protection. – 2023. – T. 86, № 4. – Art. 100054. – DOI: 10.1016/j.jfp.2023.100054.
23. Lozano-Torres B., Martínez-Bisbal M.C., Soto J., Borrás M.J., Martínez-Máñez R., Escriche I. Monofloral honey authentication by voltammetric electronic tongue: A comparison with ^{1}H NMR spectroscopy // Food Chemistry. – 2022. – T. 383. – Art. 132460. – DOI: 10.1016/j.foodchem.2022.132460.
24. Minho L.A.C., Conceição J.L., Barboza O.M., Santos Junior A.F., Santos W.N.L. Robust deep heterogeneous ensemble and meta-learning for honey authentication // Food Chemistry. – 2025. – T. 482. – Art. 144001. – DOI: 10.1016/j.foodchem.2025.144001.
25. Mostoles D., Mara A., Sanna G., Saurina J., Sentellas S., Núñez O. Authentication of honey geographical origin using LC-LRMS fingerprints // Food Analytical Methods. – 2026. – T. 19. – DOI: 10.1007/s12161-025-02953-1.
26. Nanou E., Pliatsika N., Stefas D., Polygenis D., Couris S. Laser induced breakdown spectroscopy as an emerging technique for olive oil, milk and honey authentication and traceability: A review // Journal of Food Composition and Analysis. – 2025. – T. 144. – Art. 107650. – DOI: 10.1016/j.jfca.2025.107650.
27. Nunes A., Azevedo G.Z. dos Santos B.R., Lima G.P.P., Moura S., Maraschin M. Application of UV–vis spectrophotometry and chemometrics to investigate adulteration by glucose syrup in Brazilian polyfloral honey // Food and Humanity. – 2024. – T. 2. – Art. 100194. – DOI: 10.1016/j.foohum.2023.12.002.
28. Ongalbek D. Şahin B., Berdesh T., Taş-Küçükaydın M., Tokul-Ölmez Ö., Yeskaliyeva B., Öztürk M. Biomarker aroma compounds of monofloral honey from Kazakhstan by GC–MS and chemometric analysis // Analytical Letters. – 2073. – T. 58, № 12. – P. 2056–2073. – DOI: 10.1080/00032719.2024.2345751.
29. Pujol C., Danoun S., Biasini G., Retailleau E., Masson J., Balayssac S., Gilard V. Benchtop NMR coupled with chemometrics: A workflow for unveiling hidden drug ingredients in honey-based supplements // Molecules. – 2086. – T. 29, № 9. – Art. 2086. – DOI: 10.3390/molecules29092086.
30. Punta-Sánchez I., Dymerski T., Pérez Calle J.L., Ruiz-Rodríguez A., Ferreiro-González M., Palma M. Detecting honey adulteration: Advanced approach using UF-GC coupled with machine learning // Sensors. – 2024. – T. 24, № 23. – Art. 7481. – DOI: 10.3390/s24237481.
31. Razavi R., Esmaeilzadeh Kenari R. Ultraviolet-visible spectroscopy combined with machine learning as a rapid detection method to predict adulteration of honey // Heliyon. – 2023. – T. 9, № 10. – DOI: 10.1016/j.heliyon.2023.e20973.
32. Ropciuc S., Dranca F., Pauliuc D., Oroian M. Honey authentication and adulteration detection using emission-excitation spectra combined with chemometrics // Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. – 2023. – T. 293. – Art. 122459. – DOI: 10.1016/j.saa.2023.122459.

33. Suhandy D., Al Riza D.F., Yulia M., Kusumiyati K., Telaumbanua M., Naito H. Rapid authentication of intact stingless bee honey by portable LED-based fluorescence spectroscopy and chemometrics // Foods. – 2024. – T. 13, № 22. – Art. 3648. – DOI: 10.3390/foods13223648.
34. Tarapoulouzi M., Mironescu M., Drouza C., Mironescu I.D., Agriopoulou S. Insight into the recent application of chemometrics in quality analysis and characterization of bee honey during processing and storage // Foods. – 2023. – T. 12, № 3. – Art. 473. – DOI: 10.3390/foods12030473.
35. Uçurum H. Ö., Tepe Ş., Yeşil E., Güney F., Karakuş S., Kolaylı S., Takma Ç., Duru M.E., Özkök A., Yücel B., Karaca Ü., Sorkun K., Baran A., Kılınç A., Köseoğlu M., Özsoy N., Kunduracı B.S., Türkaslan N., Atmaca H., Karaağaç H.E., Topal E., Dayıoğlu M., Kalkan M., Ekerbiçer Ü., Çakici N. Characterization of Turkish pine honey according to their geographical origin based on physicochemical parameters and chemometrics // European Food Research and Technology. – 2023. – T. 249, № 5. – P. 1317–1327. – DOI: 10.1007/s00217-023-04215-y.
36. Zhang X.H., Gu H.W., Liu R.J., Qing X.D., Nie J.F. A comprehensive review of the current trends and recent advancements on the authenticity of honey // Food Chemistry: X. – 2023. – T. 19. – Art. 100850. – DOI: 10.1016/j.fochx.2023.100850.