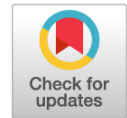


DOI: <https://doi.org/10.17816/fm16333>

EDN: DYJUPY



Цифровой сравнительный таксономический анализ диатомовых сообществ при установлении места утопления: технический отчёт

В.Г. Недугов¹, Г.В. Недугов²¹ Самарский национальный исследовательский университет имени академика С.П. Королёва, Самара, Россия;² Самарский государственный медицинский университет, Самара, Россия

АННОТАЦИЯ

Обоснование. Установление места утопления в судебной медицине основано на сравнительном анализе сообществ диатомовых водорослей. Однако существующая методология недостаточно формализована: стандартные экологические подходы неприменимы к парному сравнению образцов (из водоёма и трупа), а выводы в значительной степени основываются на субъективной визуальной оценке. Это накладывает ограничения на объективную и обоснованную интерпретацию результатов, снижая доказательную силу экспертизы.

Цель работы. Разработка формализованного метода цифрового парного сравнительного таксономического анализа диатомового планктона и его реализация в виде кроссплатформенного веб-приложения для судебно-медицинской практики.

Методы. Работа представляет разработку протокола сравнительного таксономического анализа и его реализацию в виде веб-приложения, совместимого с операционными системами Windows, Linux, Android и iOS. Протокол включает три блока взаимодополняющих метрик для парного сравнения: общее сходство (дистанции Брея–Кёртиса и Хелленджера с бутстрэп-тестом); структура доминирования (индексы Жаккара, Сёренсена–Дайса, Симпсона, ранговые корреляции); стабильность и индикаторность (адаптированные метрики Дюффрена–Лежандра). Для обоснования пороговых значений метрик и оценки их диагностической эффективности проведено имитационное моделирование методом Монте-Карло. Код приложения составляли на языках программирования JavaScript и Python.

Результаты. Разработан формализованный метод парного сравнительного таксономического анализа, заменяющий субъективную оценку расчётом объективных метрик. Создано бесплатное онлайн-приложение Qualitative Dual Sample Taxa Analyzer. На основе моделирования методом Монте-Карло для каждой метрики установлены статистически обоснованные пороговые значения, обеспечивающие чувствительность и специфичность не менее 95%. Ключевым элементом приложения является бутстрэп-процедура, гарантирующая статистическую проверку гипотезы о происхождении проб из одного водоёма. Комбинация метрик позволяет проводить многоуровневый анализ таксономического состава экстрактов планктона, а веб-интерфейс делает метод доступным для экспертов, автоматизируя расчёты, стандартизируя процедуру и формируя детализированный отчёт.

Заключение. Предложенный метод сравнительного таксономического анализа и программный инструмент представляют комплексное решение задачи объективной и статистически валидируемой оценки сходства диатомовых сообществ. Их внедрение в практику способно повысить доказательную силу, воспроизводимость и стандартизацию судебно-медицинских заключений по установлению места утопления.

Ключевые слова: сравнительный анализ; таксономический состав; диатомовый планктон; место утопления; бутстрэп-тест; веб-приложение; моделирование методом Монте-Карло; технический отчёт.

Как цитировать:

Недугов В.Г., Недугов Г.В. Цифровой сравнительный таксономический анализ диатомовых сообществ при установлении места утопления: технический отчёт // Судебная медицина. 2026. Т. 12, № 1. С. 48–62. DOI: 10.17816/fm16333 EDN: DYJUPY

DOI: <https://doi.org/10.17816/fm16333>

EDN: DYJUPY

Digital Comparative Taxonomic Analysis of Diatom Communities for Determining the Site of Drowning: A Technical Report

Vladimir G. Nedugov¹, German V. Nedugov²¹ Samara National Research University named after academician S.P. Korolev, Samara, Russia;² Samara State Medical University, Samara, Russia

ABSTRACT

BACKGROUND: Determination of the drowning site in forensic practice is based on comparative analysis of diatom communities. However, current methodology remains insufficiently formalized: standard ecological approaches are not applicable to paired sample comparisons (water body vs cadaver), and conclusions largely rely on subjective visual assessment. This limits objective and evidence-based interpretation and reduces the evidentiary value of forensic conclusions.

AIM: To develop a formalized method for digital paired comparative taxonomic analysis of diatom plankton and implement it as a cross-platform web application for forensic practice.

METHODS: This work describes the development of a protocol for comparative taxonomic analysis and its implementation as a web application compatible with Windows, Linux, Android, and iOS. The protocol includes three complementary metric blocks for paired comparison: overall similarity (Bray–Curtis and Hellinger distances with bootstrap testing); dominance structure (Jaccard, Sørensen–Dice, and Simpson indices, as well as rank correlations); and stability and indicator value (adapted Dufrière–Legendre metrics). To justify threshold values and evaluate diagnostic performance, Monte Carlo simulation was performed. The application was developed using JavaScript and Python.

RESULTS: A formalized method for paired comparative taxonomic analysis was developed, replacing subjective assessment with calculation of objective metrics. A free online application, Qualitative Dual Sample Taxa Analyzer, was created. Based on Monte Carlo simulation, statistically justified threshold values were established for each metric, ensuring sensitivity and specificity of at least 95%. A key feature of the application is a bootstrap procedure that enables statistical testing of the hypothesis that the samples originate from the same water body. The combination of metrics allows multilevel analysis of the taxonomic composition of plankton extracts, whereas the web interface facilitates accessibility for forensic experts by automating calculations, standardizing procedures, and generating detailed reports.

CONCLUSION: The proposed method and software tool provide a comprehensive solution for objective and statistically validated assessment of similarity between diatom communities. Their implementation may improve the evidentiary strength, reproducibility, and standardization of forensic conclusions regarding the determination of the site of drowning.

Keywords: comparative analysis; taxonomic composition; diatom plankton; drowning site; bootstrap test; web application; Monte Carlo simulation; technical report.

To cite this article:

Nedugov VG, Nedugov GV. Digital Comparative Taxonomic Analysis of Diatom Communities for Determining the Site of Drowning: A Technical Report. *Russian Journal of Forensic Medicine*. 2026;12(1):48–62. DOI: 10.17816/fm16333 EDN: DYJUPY

DOI: <https://doi.org/10.17816/fm16333>

EDN: DYJUPY

用于确定溺水地点的硅藻群落数字比较分类学分析：技术报告

Vladimir G. Nedugov¹, German V. Nedugov²¹ Samara National Research University named after academician S.P. Korolev, Samara, Russia;² Samara State Medical University, Samara, Russia

摘要

论证：法医学中溺水地点的判定基于硅藻群落的比较分析。然而现有方法论缺乏充分的形式化标准：传统生态学方法不适用于样本（水体和尸体）的配对比较，且结论很大程度上依赖主观视觉评估。这限制了结果的客观合理阐释，降低了鉴定的证据效力。

目的：开发形式化的数字配对比较分类学分析方法用于硅藻浮游生物分析，并实现为跨平台网络应用程序以供法医学实践应用。

方法：本研究设计了一套比较分类学分析方案，并开发了兼容Windows、Linux、Android和iOS操作系统的网页应用程序。该方案包含三个互补的指标模块用于成对比较：总体相似度（采用Bray-Curtis距离和Hellinger距离配合自举检验）；优势结构（Jaccard指数、Sørensen-Dice指数、Simpson指数及等级相关分析）；稳定性与指示性（改进的Dufrene-Legendre指标）。通过蒙特卡洛模拟法确定各指标的阈值并评估其诊断效能。应用程序采用JavaScript和Python编程语言编写代码。

结果：开发了一种形式化的成对比较分类学分析方法，该方法通过客观度量计算取代了主观评估。创建了免费的在线应用程序Qualitative Dual Sample Taxa Analyzer。通过蒙特卡洛模拟为每个度量指标确立了统计上合理的阈值，确保敏感性和特异性不低于95%。该应用的核心是自助法（bootstrap）程序，可统计验证样本是否来自同一水体的假设。度量指标的组合支持对浮游生物提取物进行多层次分类学组成分析，而网络界面通过自动化计算、标准化流程和生成详细报告，使该方法可供专家便捷使用。

结论：所提出的比较分类学分析方法和软件工具为硅藻群落相似性的客观且统计可验证的评估提供了完整解决方案。将其应用于实践可增强溺水地点鉴定中法医结论的证据效力、可重复性和标准化程度。

关键词：比较分析；分类学组成；硅藻浮游生物；溺水地点；自助法检验；网络应用程序；蒙特卡洛模拟；技术报告。

引用本文：

Nedugov VG, Nedugov GV. 用于确定溺水地点的硅藻群落数字比较分类学分析：技术报告. *Russian Journal of Forensic Medicine*. 2026;12(1):48–62. DOI: 10.17816/fm16333 EDN: DYJUPY

收到: 01.01.2026

接受: 12.03.2026

发布日期: 24.03.2026

ОБОСНОВАНИЕ

Установление места утонления является одной из важнейших и одновременно сложных задач судебно-медицинской диагностики данного вида смерти [1–3]. На практике решение указанной задачи основывается на сравнительном исследовании сообществ диатомовых водорослей (планктона), извлечённых из внутренних органов трупа и из предполагаемого водоёма [4–6]. Однако, несмотря на десятилетия применения, методология сравнительного таксономического анализа в судебно-медицинской альгологии остаётся недостаточно формализованной и в значительной степени субъективной [3, 7, 8]. Существующие экологические подходы к анализу биоразнообразия [например, кластерный анализ, многомерное шкалирование, анализ сходства (Analysis of Similarities, ANOSIM)] рассчитаны на сравнение множества проб и зачастую неприменимы ввиду требования к большому размеру выборки или избыточны для принципиальной задачи экспертизы — парного сравнения образцов: контрольного (из водоёма) и исследуемого (трупного) [9]. Рассчитываемые отдельные экологические индексы, несмотря на их значимость для интерпретации результатов диатомового анализа, при самостоятельном применении не позволяют сделать однозначный количественный вывод о происхождении экстрактов планктона и служат лишь вспомогательным инструментом, не давая интегральной оценки сходства, учитывающей всю структуру сообщества [10–12]. На практике в большинстве случаев экспертное заключение вообще традиционно опирается лишь на качественную визуальную оценку сходства сообществ диатомей, что неизбежно влечёт за собой риск интерпретационной вариабельности и затрудняет статистическое обоснование выводов.

Указанное методологическое ограничение обуславливает необходимость разработки специализированного аналитического протокола, отвечающего специфике судебно-медицинских исследований. Он должен решать три основные задачи.

- Во-первых, обеспечивать количественную оценку сходства обнаруженных диатомовых сообществ с помощью объективных метрик, таких как дистанция Брея–Кёртиса и Хелленджера, учитывающих не только наличие, но и обилие таксонов [13, 14].
- Во-вторых, включать анализ структуры сообщества: выявление и сравнение ядра доминирующих таксонов, а также оценку индикаторных таксонов (например, по адаптированной логике метода Дюффрена–Лежандра) и общей стабильности, где интерпретация результатов должна быть напрямую калибрована на задачу установления общего источника происхождения проб [15–20].
- В-третьих, методология должна предусматривать статистическую валидацию результатов парного сравнения, позволяя вычислять значимость наблюдаемых

различий при минимальном количестве сравниваемых образцов [11, 21].

Потребности практики также диктуют необходимость создания специализированного программного обеспечения, реализующего данный метод сравнительного таксономического анализа. Экспертам-альгологам необходим доступный и интуитивно понятный инструмент, который автоматизирует ввод данных и сложные расчёты, обеспечивает единообразие процедуры анализа и сформирует структурированный отчёт. Разработка такого программного обеспечения не только стандартизирует исследовательский процесс, минимизируя человеческую ошибку, но и сделает передовые методы многомерной статистики доступными для рутинной экспертной работы, что представляется важным шагом на пути цифровизации и объективизации судебно-медицинского экспертного заключения [22, 23].

Таким образом, разработка формализованного метода и программного инструмента для парного сравнительного таксономического анализа при утонлении является решением актуальной научно-практической задачи, направленным на повышение доказательной силы и воспроизводимости диатомологического метода в судебной медицине.

ЦЕЛЬ

Разработка формализованного метода цифрового парного сравнительного таксономического анализа диатомового планктона и его реализация в виде кроссплатформенного веб-приложения для судебно-медицинской практики.

МЕТОДЫ

Работа представляет разработку и реализацию в виде веб-приложения формализованного метода парного сравнительного таксономического анализа векторов данных о диатомовых сообществах экстрактов двух образцов: контрольного, извлечённого из предполагаемого водоёма, и исследуемого — из трупного биоматериала. Протокол метода, определяемого как цифровой сравнительный таксономический анализ (ЦСТА), представляет последовательность вычисления взаимодополняющих метрик, объединённых в три аналитических блока:

- общее сходство (несходство Брея–Кёртиса, дистанция Хелленджера, бутстрэп-тест значимости);
- структура доминирования (определение доминирующих таксонов по порогу или топ- k , индекс доминирования Симпсона, относительная разница в доминировании, ранговые корреляции);
- стабильность и индикаторность (адаптированные метрики Дюффрена–Лежандра для каждого таксона с последующей агрегацией в композитный индекс индикаторности и общий индекс стабильности).

Имитационное моделирование методом Монте-Карло

Для обоснования пороговых значений метрик протокола ЦСТА и оценки их диагностической эффективности проведено имитационное моделирование методом Монте-Карло. Моделирование выполняли на языке Python 3 с использованием библиотек `numpy`, `scipy.stats` и `matplotlib`. Основной целью являлось получение эмпирических распределений каждой метрики для двух сценариев:

- H_0 (нулевая гипотеза) — пробы происходят из одного и того же диатомового сообщества (одного водоёма);
- H_1 (альтернативная гипотеза) — пробы происходят из статистически разных сообществ (разных водоёмов).

Относительные обилия генерировали с использованием логнормального распределения, параметр формы которого определяли коэффициентом выравнивания:

$$\sigma = 1/\text{evenness}, \quad (1)$$

где σ — параметр формы логнормального распределения; *evenness* — параметр выравнивания.

Параметры моделирования:

- общее число потенциально возможных таксонов в водоёме принято равным 100;
- количество активных таксонов в каждом сообществе случайным образом выбирали из диапазона от 10 до 50, неактивные таксоны имели нулевую вероятность встречаемости;
- диапазон варьирования параметра выравнивания составлял от 0,3 до 3,0;
- для учёта эффекта редкости введён порог редкости 0,001: если расчётная вероятность таксона оказывалась ниже этого порога, таксон считали необнаруженным, а его вероятность обнулялась перед извлечением пробы;
- для моделирования случайной вариации при отборе проб использовали мультиномиальное распределение;
- для учёта возможных ошибок идентификации и подсчёта введён параметр ошибки пробы, равный 0,01, при срабатывании которого к счётчику таксона случайным образом добавлялось или вычиталось от 1 до 4 створок;
- в сценарии H_0 объёмы проб фиксированы и составляли 1000 особей в каждой пробе;
- в сценарии H_1 объёмы проб для каждого сообщества выбирали случайным образом из диапазона от 200 до 2000.

Генерация данных

Для сценария H_0 генерировали одно эталонное сообщество с параметром *evenness*=1,0 (умеренная выравниванность). Из этого сообщества с использованием вышеописанных параметров извлекали 10 000 пар проб, имитирующих случайную вариацию при повторном отборе из одной генеральной совокупности.

Для сценария H_1 на каждой итерации независимо генерировали два сообщества с параметрами *evenness*, равномерно распределёнными в диапазоне [0,3; 3,0]. Из каждого сообщества извлекали по одной пробе со случайным объёмом из диапазона [200; 2000] также с учётом редкости и ошибок. Всего генерировали 10 000 пар проб, отражающих различия между разными водоёмами.

Выделение доминирующих таксонов

Для всех метрик, использующих понятие «множество доминирующих таксонов» (индексы Жаккара, Сёренсена–Дайса, процент совпадения, стабильность присутствия и производные композитные индексы), применяли метод «топ-3» — три таксона с наибольшим относительным обилием в пробе.

Расчёт метрик и определение пороговых значений

Для каждой сгенерированной пары проб вычисляли все метрики, включённые в протокол ЦСТА. После завершения моделирования для каждой метрики построены эмпирические распределения при H_0 и H_1 .

Пороговые значения, разделяющие области подтверждения H_0 и H_1 , устанавливали таким образом, чтобы специфичность (доля проб H_0 , правильно классифицированных как «один водоём») и чувствительность (доля проб H_1 , правильно классифицированных как «разные водоёмы») составляли не менее 95% для каждой метрики. Для метрик, где большее значение соответствует большему сходству, за нижнюю границу «высокого сходства» принимался 5-й перцентиль распределения H_0 , а за верхнюю границу «низкого сходства» — 95-й перцентиль распределения H_1 . Для метрик, где большее значение соответствует большему различию, за верхнюю границу «высокого сходства» принимался 95-й перцентиль распределения H_0 , а за нижнюю границу «низкого сходства» — 5-й перцентиль распределения H_1 . Если распределения H_0 и H_1 перекрывались настолько, что указанные условия не могли быть выполнены одновременно, метрику признавали недостаточно информативной и исключали из итоговой шкалы интерпретации.

Языки программирования и технологии

Код онлайн-приложения написан на JavaScript с использованием библиотеки React, а также на Python 3 с использованием фреймворка Django и Django REST Framework для обеспечения API расчётных функций. В коде применяются библиотеки:

- `numpy` — для векторных вычислений;
- `scipy` — для статистических расчётов (ранговые корреляции);
- `plotly` — для визуализации графиков.

Приложение совместимо с операционными системами Windows, Linux, Android и iOS, запускается в современных веб-браузерах.

РЕЗУЛЬТАТЫ

При выполнении настоящей работы разработан протокол ЦСТА, адаптированный для задачи идентификации места утонения. Протокол включает многомерный статистический анализ на основе бутстрэп-процедуры с использованием дистанций Брея–Кёртиса и Хелленджера, а также комплексный анализ доминирующих таксонов и индикаторных видов. Подробная характеристика вычисляемых метрик приведена ниже.

1. Метрики дистанций (несходства)

Основными метриками ЦСТА являются оценки дистанций (несходства) составов таксонов диатомового планктона, обнаруженных при альгологическом исследовании образцов среды утонения и трупного материала, в частности несходство Брея–Кёртиса и дистанция Хелленджера [13, 14]. Обе метрики измеряют степень перекрытия видовых составов планктона между образцами и являются композиционными, поскольку основываются на процентных соотношениях таксонов, вследствие чего нечувствительны к пропорциональным изменениям их абсолютных величин. Оценки обеих метрик могут принимать значения от 0 до 1, при этом 0 означает идентичный состав, а 1 — отсутствие общих таксонов.

В рамках решаемой задачи формулы указанных метрик принимают следующий вид:

$$BC_{ij} = 1 - \sum_{k=1}^n \min(p_{ik}, p_{jk}), \quad (2)$$

где BC_{ij} — несходство Брея–Кёртиса между таксономическими составами образцов i и j ; p_{ik}, p_{jk} — доли таксона k в образцах i и j ; n — количество таксонов; \min — минимальное значение в паре;

$$H_{ij} = \sqrt{\sum_{k=1}^n (\sqrt{p_{ik}} - \sqrt{p_{jk}})^2}, \quad (3)$$

где H_{ij} — ненормированная дистанция Хелленджера между таксономическими составами образцов i и j ; p_{ik}, p_{jk} — доли таксона k в образцах i и j ; n — количество таксонов.

2. Анализ доминирующих таксонов

2.1. Определение доминирующих таксонов

Базовым компонентом анализа доминирующих таксонов, в значительной степени определяющим его результаты, является способ установления множества доминирующих таксонов. Стандартный подход к определению элементов данного множества сводится к произвольному выбору порога доминирования. Таксоны, относительные обилия которых достигают или превышают выбранный порог, идентифицируют как доминирующие.

Помимо указанного способа, в рамках ЦСТА также реализован альтернативный метод определения множества доминирующих таксонов, основанный на выделении топ- k таксонов.

2.2. Формулы сходства доминирующих таксонов

Стандартный подход к оценке сходства доминирующих таксонов заключается в вычислении метрик Жаккара и Сёренсена–Дайса [15–17]. Индекс Жаккара оценивают как:

$$J = \frac{|A \cap B|}{|A \cup B|}, \quad (4)$$

где J — индекс Жаккара; A и B — множества доминирующих таксонов в каждом из двух образцов; $|A \cap B|$ — количество общих доминирующих таксонов; $|A \cup B|$ — количество таксонов в объединении множеств A и B [15].

Коэффициент Сёренсена–Дайса вычисляют по формуле [16, 17]:

$$SD = \frac{2|A \cap B|}{|A| + |B|}, \quad (5)$$

где SD — метрика Сёренсена–Дайса; $|A \cap B|$ — количество общих доминирующих таксонов; $|A|$ и $|B|$ — количество элементов во множествах A и B .

Помимо указанных метрик в качестве естественного развития заложенных в них идей в вычислительный алгоритм ЦСТА включена интуитивно понятная и легко интерпретируемая метрика — процентное совпадение доминирующих таксонов:

$$O = \frac{|A \cap B|}{\min(|A|, |B|)} \times 100\%, \quad (6)$$

где O — показатель совпадения, %; $|A \cap B|$ — количество общих доминирующих таксонов; \min — минимальное значение в паре; $|A|$ и $|B|$ — количество элементов во множествах A и B .

2.3. Анализ ранговой корреляции доминирования

Помимо перечисленных метрик структуры доминирования ЦСТА также выполняют сравнение рангов обилий таксонов диатомового планктона в двух образцах путём вычисления коэффициентов ранговой корреляции Спирмена и Кендалла.

Коэффициент ранговой корреляции Спирмена оценивается по формуле:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (7)$$

где ρ — обозначение коэффициента Спирмена; d_i — разность рангов таксона i в образцах A и B ; n — количество таксонов.

Коэффициент ранговой корреляции Кендалла вычисляется иначе:

$$\tau = \frac{2(P - Q)}{n(n - 1)}, \quad (8)$$

где τ — обозначение коэффициента Кендалла; P — количество согласованных пар рангов; Q — количество несогласованных пар; n — общее количество пар рангов.

Диапазоны значений обоих коэффициентов могут изменяться от -1 до 1 , однако в контексте сравнения двух проб из одного водоёма отрицательные значения коэффициентов маловероятны, поскольку сравниваются ранжировки одних и тех же таксонов.

2.4. Индекс доминирования Симпсона

Ещё одной метрикой, измеряющей доминирование и реализованной в вычислительном алгоритме ЦСТА, является индекс Симпсона, определяемый по формуле:

$$D = \sum p_i^2, \quad (9)$$

где D — выборочная оценка параметра λ Симпсона; p_i — относительное обилие таксона i [18]. Параметр λ Симпсона представляет меру концентрации для бесконечной популяции, в которой особи распределены по группам, и обозначает вероятность, что две особи, случайно и независимо выбранные из популяции, принадлежат к одной и той же группе [18].

Индекс Симпсона измеряет доминирование: чем значение этой метрики больше, тем меньше разнообразие таксонов в экстракте планктона. Для количественного сравнения индексов Симпсона для двух образцов протокол ЦСТА с помощью масштабирующего множителя, представляющего максимально наблюдаемое доминирование в образцах, предполагает вычисление нормированного абсолютного различия в степени доминирования:

$$\Delta D = \frac{|D_1 - D_2|}{\max(D_1, D_2)}, \quad (10)$$

где ΔD — относительная разница в доминировании; D_1 и D_2 — индексы Симпсона для образцов 1 и 2; \max — максимальное значение в паре.

Диапазон значений параметра ΔD варьирует от 0 до 100%, при этом идентичной структуре доминирования соответствует значение 0%, а максимально возможному различию, когда в одном сообществе наблюдается полное доминирование, а в другом — равномерное распределение таксонов — 100%.

3. Анализ индикаторных видов

Поскольку для сравнения двух образцов классический анализ индикаторных видов Дюффрена–Лежандра неприменим [19, 20], в ЦСТА предусмотрены адаптированные подходы, заключающиеся в вычислении метрик относительной разницы обилий, стабильности присутствия, стабильности обилий и композитного индекса индикаторности.

3.1. Относительная разница обилий

Данная метрика показывает, насколько различается обилие конкретного таксона между двумя образцами:

$$RD_i = \frac{|p_{iA} - p_{iB}|}{\max(p_{iA}, p_{iB})} \times 100\%, \quad (11)$$

где RD_i — относительная разница обилий; p_{iA} и p_{iB} — относительные обилия таксона i в образцах A и B ; \max — максимальное значение в паре. Максимум данный показатель достигает при отсутствии таксона в одном из образцов.

3.2. Стабильность присутствия

В протоколе ЦСТА используется расширенная формула стабильности присутствия, дающая три возможных значения вместо двух стандартных и позволяющая перейти от бинарной логики (есть/нет) к более тонкой градации, что особенно полезно для таксонов с низкой или переменной встречаемостью. Кроме того, предусмотрена дополнительная возможность выбора одного из двух подходов к вычислению данной метрики.

Первый подход строится на анализе порога встречаемости. В соответствии с алгоритмом оценки данной метрики таксон считается «присутствующим», если его относительное обилие превышает установленный порог, например 1 или 5%. Тогда индикатору присваивается значение 1, если таксон присутствует в обоих образцах (его обилие в каждом из них превышает установленный порог). Если таксон присутствует только в одном из двух образцов (его обилие выше установленного порога в одном и ниже в другом), то индикатору присваивается значение 0,5. Индикатор равен 0, если таксон отсутствует в обоих образцах (его обилие в каждом из них ниже установленного порога).

Альтернативный подход основывается на выборе вместо фиксированного порога топ- k первых k таксонов. Это более адаптивный подход, подстраивающийся под уровень доминирования, характерный для каждого конкретного образца. Алгоритм метода включает выполнение следующих операций:

- определение значения k и топ- k самых обильных таксонов в каждом образце;
- оценка вхождения каждого таксона из имеющегося набора данных в топ- k каждого из двух образцов;
- присвоение индикатору значения 1, если таксон входит в топ- k обоих образцов, значения 0,5 — в топ- k только одного из двух образцов, и значения 0 — таксон не входит в топ- k ни одного из двух образцов.

Выбор значения k является ключевым в реализации второго подхода. Это объясняется тем, что слишком маленькое k (например, 3) может не учитывать важные таксоны, а слишком большое (например, 25) — нивелировать сам принцип «доминирования». Значение k целесообразно выбирать на основе кумулятивной кривой доминирования, принимая за него количество таксонов,

численность которых составляет 80% общего количества обнаруженных диатомей.

Высокая стабильность ($SP=1$) означает, что сообщество в обоих экстрактах планктона определяют одни и те же доминирующие таксоны. Умеренная стабильность ($SP=0,5$) свидетельствует о доминировании некоторых таксонов лишь в одном из образцов, что указывает на начальные различия в структуре обнаруженных в них диатомовых сообществ. Нулевая стабильность ($SP=0$) отдельных таксонов доказывает, что они не определяют структуру диатомовых сообществ ни в одном из образцов.

3.3. Коэффициент стабильности обилий

Метрика показывает степень сходства обилий таксона между образцами:

$$SC_i = \frac{\min(p_{iA}, p_{iB})}{\max(p_{iA}, p_{iB})}, \quad (12)$$

где SC_i — степень сходства обилий; p_{iA} и p_{iB} — относительные обилия таксона i в образцах A и B ; \max — максимальное значение в паре; \min — минимальное значение в паре. Это наиболее объективная метрика для таксонов со средним обилием, диапазон возможных значений которой варьирует от 0 до 100%.

3.4. Композитный индекс индикаторности для двух образцов

Метрика является интегральным индексом, объединяющим значения трёх индикаторов по отдельным таксону — относительной разницы обилий, стабильности присутствия и стабильности обилий:

$$Ind_i = \frac{(1 - RD_i) + SC_i + SP_i}{3} \times 100\%, \quad (13)$$

где Ind_i — композитный индекс индикаторности; RD_i — относительная разница обилий; SC_i — степень сходства обилий; SP_i — стабильность присутствия. Область значений композитного индикатора также варьирует от 0 до 100%.

3.5. Общий индекс стабильности сообщества

Поскольку композитный индекс индикаторности рассчитывается в отношении каждого отдельного таксона, в протокол ЦСТА включён расчёт общего индекса стабильности по всем таксонам:

$$IS = \frac{\sum (Ind_i \times p_i)}{\sum p_i}, \quad (14)$$

где IS — общий индекс стабильности; Ind_i — композитный индекс индикаторности; p_i — относительное обилие таксона i .

Общий индекс стабильности носит интегральный характер, обобщая данные композитного индекса индикаторности по всем таксонам. Он чувствителен к сдвигам в ядре сообщества, поскольку придаёт больший вес изменениям в обильных видах.

Имитационное моделирование методом Монте-Карло с контролируемыми уровнями специфичности и чувствительности ($\geq 95\%$) позволили разработать шкалу интерпретации значений метрик ЦСТА, представленную в табл. 1.

4. Бутстрэп-процедура

Комплекс вышеперечисленных метрик представляет мощный вспомогательный показатель, но не самостоятельный критерий, дающий количественную оценку вероятности происхождения экстрактов планктона из одного водоёма. Именно поэтому в протокол ЦСТА включён наиболее важный бутстрэп-компонент, позволяющий решить указанную задачу. Суть последней сводится к тому, что у эксперта есть два вектора (пробы), каждый из которых описывает количество особей для n таксонов. Необходимо проверить гипотезы:

- H_0 (нулевая гипотеза): пробы взяты из одного диатомового сообщества (одного водоёма с одинаковой структурой сообщества), различия между ними случайны и обусловлены лишь естественной изменчивостью и ошибками взятия пробы и пробоподготовки;

Таблица 1. Шкала интерпретации метрик происхождения экстрактов диатомового планктона

Метрика	Подтверждение H_0 (один водоём)	Неопределённая зона	Подтверждение H_1 (разные водоёмы)
Несходство Брея–Кёртиса	$\leq 0,131$	0,132–0,696	$\geq 0,697$
Дистанция Хелленджера	$\leq 0,141$	0,142–0,762	$\geq 0,763$
Индекс Жаккара ¹	$\geq 0,500$	0,201–0,499	$\leq 0,200$
Коэффициент Сёрнсена–Дайса ¹	$\geq 0,667$	0,334–0,666	$\leq 0,333$
Процентное совпадение доминирующих таксонов ¹ , %	$\geq 66,7$	33,4–66,6	$\leq 33,3$
Ранговая корреляция Спирмена	$\geq 0,958$	0,168–0,957	$\leq 0,167$
Ранговая корреляция Кендалла	$\geq 0,886$	0,147–0,885	$\leq 0,146$
Общий индекс стабильности сообщества	$\geq 0,632$	0,086–0,631	$\leq 0,085$

Примечание. ¹ — для метрик, зависящих от выделения доминирующих таксонов, применяли метод «топ-3». В силу дискретной природы этих метрик указанные диапазоны соответствуют границам, установленным по 5 и 95 перцентилем распределений, и позволяют однозначно интерпретировать любое возможное дискретное значение. Все пороги обеспечивают специфичность и чувствительность не менее 95%. Метрики, проявившие дискриминационную способность меньше заявленной, не включены в шкалу интерпретации.

- H_1 (альтернативная гипотеза): пробы взяты из статистически разных диатомовых сообществ (из разных водоёмов).

Наиболее эффективное решение указанной задачи достигается с помощью бутстрэппинга, в рассматриваемом контексте представляющего метод оценки значимости различий между двумя экологическими пробами путём многократной генерации псевдовыборок из объединённого пула данных, моделирующего гипотетический «общий водоём», и сравнения наблюдаемых дистанций Брея–Кёртиса и Хелленджера с распределением данных метрик, полученных на этих псевдовыборках. Таким образом, рассматриваемый метод является гибридным подходом, сочетающим идею нулевой гипотезы (один водоём) с бутстрэп-методологией генерации данных, что особенно эффективно при таком малом числе образцов.

Пошаговый алгоритм бутстрэп-процедуры включает следующие этапы.

- Шаг 1. Расчёт наблюдаемых дистанций Брея–Кёртиса и Хелленджера между двумя исходными пробами.
- Шаг 2. Создание общего пула данных как взвешенного среднего нормализованных векторов обилий обеих проб с учётом их общего количества особей:

$$\mathbf{p}_0 = \frac{n_A \times \mathbf{p}_A + n_B \times \mathbf{p}_B}{n_A + n_B}, \quad (15)$$

где \mathbf{p}_0 — вектор оценок вероятностей встречи каждого таксона в гипотетической общей популяции, из которой, в соответствии с нулевой гипотезой, происходят обе пробы; n_A — общее количество особей в пробе A ; n_B — общее количество особей в пробе B ;

$$\mathbf{p}_A = \left(\frac{a_1}{n_A}, \frac{a_2}{n_A}, \dots, \frac{a_m}{n_A} \right), \quad (16)$$

где \mathbf{p}_A — вектор относительных обилий таксонов в пробе A , получаемый нормированием вектора количеств a на общее число особей n_A в этой пробе; $\mathbf{a}=(a_1, a_2, \dots, a_m)$ — вектор количеств особей по таксонам в пробе A ;

$$\mathbf{p}_B = \left(\frac{b_1}{n_B}, \frac{b_2}{n_B}, \dots, \frac{b_m}{n_B} \right), \quad (17)$$

где \mathbf{p}_B — вектор относительных обилий таксонов в пробе B , получаемый нормированием вектора количеств b на общее число особей n_B в этой пробе; $\mathbf{b}=(b_1, b_2, \dots, b_m)$ — вектор количеств особей по таксонам в пробе B .

- Шаг 3. Генерация путём случайного извлечения из общего пула двух псевдовыборок, по размеру равных исходным пробам. При этом для каждой бутстрэп-итерации t ($t=1, 2, \dots, 1000$) генерируются два случайных вектора $\mathbf{c}_A^{(t)}$ и $\mathbf{c}_B^{(t)}$ с помощью мультиномиального

распределения таким образом, что каждый вектор $\mathbf{c}_A^{(t)}$ и $\mathbf{c}_B^{(t)}$ представляет результат «случайного отбора» n_A и n_B особей соответственно из общей популяции с вероятностями таксонов, заданными вектором \mathbf{p}_0 . Для последующего расчёта дистанций векторы количеств нормализуют, чтобы получить векторы относительных обилий для псевдовыборок:

$$\mathbf{p}_A^{(t)} = \frac{\mathbf{c}_A^{(t)}}{n_A}, \quad (18)$$

где $\mathbf{p}_A^{(t)}$ — псевдовыборка для пробы A ; $\mathbf{c}_A^{(t)}$ — случайный вектор для пробы A ; n_A — общее количество особей в пробе A ; и

$$\mathbf{p}_B^{(t)} = \frac{\mathbf{c}_B^{(t)}}{n_B}, \quad (19)$$

где $\mathbf{p}_B^{(t)}$ — псевдовыборка для пробы B ; $\mathbf{c}_B^{(t)}$ — случайный вектор для пробы B ; n_B — общее количество особей в пробе B .

- Шаг 4. Оценка бутстрэп-дистанций Брея–Кёртиса и Хелленджера для полученной пары $(\mathbf{p}_A^{(t)}, \mathbf{p}_B^{(t)})$ псевдовыборок по тем же формулам, что и для исходных данных.
- Шаг 5. Повторение процесса генерации бутстрэп-выборки, описанного в шагах 3 и 4, 1000 раз с получением распределения дистанций, которые можно ожидать, если нулевая гипотеза верна (т. е. пробы действительно из одного водоёма).
- Шаг 6. Оценка сглаженного уровня значимости:

$$p = \frac{1+m}{1000+1}, \quad (20)$$

где p — уровень значимости; m — количество бутстрэп-итераций, в которых смоделированная дистанция больше или равна аналогичной наблюдаемой.

Малое значение уровня значимости предполагает, что столь большая наблюдаемая дистанция очень маловероятна, если бы пробы были из одного водоёма. В этом случае нулевую гипотезу отвергают и делают вывод, что различия по таксономическому составу экстрактов планктона статистически значимы. Если уровень значимости велик, нулевую гипотезу не отвергают, следовательно, наблюдаемую дистанцию легко можно получить случайно из одного сообщества (водоёма).

На основе изложенного протокола ЦСТА создано бесплатное кроссплатформенное онлайн-приложение Qualitative Dual Sample Taxa Analyzer¹, доступное по URL-адресу². Программа представляет научно-исследовательский прототип, реализующий предложенный метод ЦСТА, и предназначена для использования экспертами в качестве вспомогательного аналитического инструмента (калькулятора). Её применение в практической

¹ Свидетельство о государственной регистрации программы для ЭВМ № 2025694798/ 12.12.2025. Бюл. № 12. Недугов В.Г., Недугов Г.В. Qualitative Dual Sample Taxa Analyzer. Режим доступа: https://elibrary.ru/download/elibrary_87704714_29371316.PDF Дата обращения: 17.03.2026.

² Qualitative Dual Sample Taxa Analyzer [Internet]. 2025–2026. Режим доступа: <https://forensiscalc.ru/plankton-compare> Дата обращения: 17.03.2026.

деятельности осуществляют по усмотрению эксперта для объективизации и углублённого анализа результатов диатомологического исследования.

Приложение предоставляет гибкий ввод данных, который может быть выполнен как автоматически после загрузки XML-файлов, полученных при разметке SVS-сканов экстрактов планктона, так и путём ручного выбора числа актуальных таксонов и самостоятельного заполнения пользователем соответствующих текстовых полей. Разметку SVS-сканов экстрактов планктона также можно выполнить вручную или с помощью искусственных нейросетей. Приложение рассчитывает все метрики ЦСТА с формированием структурированного отчёта, а также визуализирует результаты бутстрэппинга.

Пример

При исследовании цифровых сканов минерализатов, полученных методом экстракции перекисью водорода из 40 мл воды предполагаемого водоёма утопления и 2 мл жидкости

клиновидной пазухи трупа, обнаружены соответственно 11 722 и 446 особей диатомового планктона, родовой состав которого приведён в табл. 2. Необходимо доказать или опровергнуть происхождение обоих экстрактов планктона из одного сообщества (водоёма), приняв порог доминирования равным 0,02, а топ доминирующих таксонов — 3.

Приложение Qualitative Dual Sample Taxa Analyzer по приведённым данным позволяет вычислить все метрики ЦСТА как в отношении каждого таксона (см. табл. 2), так и диатомовых сообществ обоих экстрактов (табл. 3). Применение разработанной шкалы интерпретации (см. табл. 1) показало, что большинство метрик уверенно подтверждают гипотезу о происхождении экстрактов из одного водоёма, поскольку их значения находятся в области H_0 с гарантированной специфичностью не менее 95%. Значения дистанции Хелленджера и ранговых корреляций Спирмена и Кендалла оказались в неопределённой зоне, что не противоречит нулевой гипотезе, но указывает на необходимость комплексной оценки. Решающим аргументом в пользу H_0 стал бутстрэп-тест,

Таблица 2. Таксономический состав экстрактов планктона из предполагаемого места утопления и трупного материала, а также результаты его сравнительного анализа

Род планктона	N_1	N_2	p_1	p_2	RD_i	SP_i	SC_i	Ind_i
<i>Stephanodiscus, Cyclotella, Discostella</i>	9966	380	0,850	0,852	0,002	1	0,998	0,999
<i>Pseudostaurosira</i>	28	0	0,002	0,000	1	0	0	0
<i>Caloneis</i>	4	0	0,000	0,000	1	0	0	0
<i>Fragilaria, Staurosira</i>	206	13	0,018	0,029	0,397	0,5* 1	0,603	0,569* 0,735
<i>Navicula</i>	171	7	0,015	0,016	0,071	0	0,929	0,62
<i>Staurosirella</i>	56	1	0,005	0,002	0,531	0	0,469	0,313
<i>Nitzschia</i>	52	3	0,004	0,007	0,34	0	0,66	0,44
<i>Brachysira</i>	7	0	0,001	0,000	1	0	0	0
<i>Gomphonema</i>	7	0	0,001	0,000	1	0	0	0
<i>Diatoma</i>	132	5	0,011	0,011	0,004	0	0,996	0,664
<i>Sellaphora</i>	51	2	0,004	0,004	0,03	0	0,97	0,647
<i>Asterionella</i>	376	15	0,032	0,034	0,046	1	0,954	0,969
<i>Gyrosigma</i>	28	0	0,002	0,000	1	0	0	0
<i>Ulnaria</i>	119	6	0,010	0,013	0,245	0	0,755	0,503
<i>Tabellaria</i>	93	4	0,008	0,009	0,115	0	0,885	0,59
<i>Pinnularia</i>	87	3	0,007	0,007	0,094	0	0,906	0,604
<i>Neidium</i>	80	2	0,007	0,004	0,343	0	0,657	0,438
<i>Surirella</i>	26	0	0,002	0,000	1	0	0	0
<i>Diploneis</i>	61	2	0,005	0,004	0,138	0	0,862	0,574
<i>Cymatopleura</i>	5	0	0,000	0,000	1	0	0	0
<i>Meridion</i>	64	3	0,005	0,007	0,188	0	0,812	0,541
<i>Cymbella</i>	26	0	0,002	0,000	1	0	0	0
<i>Rhoicosphenia</i>	37	0	0,003	0,000	1	0	0	0
<i>Eunotia, Cocconeis</i>	40	0	0,003	0,000	1	0	0	0

Примечание. N_1 и N_2 — количество особей в экстрактах из водоёма и клиновидной пазухи соответственно; Ind_i — композитный индекс индикаторности; RD_i — относительная разница обилий; SC_i — степень сходства обилий; SP_i — стабильность присутствия. * — данные приведены для порога доминирования=0,02 (верхняя строка) и топ- $k=3$ (нижняя строка).

Таблица 3. Результаты сравнительного анализа родового состава экстрактов планктона и их интерпретация в контексте гипотезы происхождения из одного водоёма

Метрика	Значение	Порог для H_0^1	Интерпретация
Несходство Брея–Кёртиса	0,024	$\leq 0,131$	Подтверждение H_0
Дистанция Хелленджера	0,143	$\leq 0,141$	Неопределённая зона ²
Индекс Жаккара ³ (порог 0,02)	0,667	$\geq 0,500$	Подтверждение H_0^4
Индекс Жаккара ³ (топ-3)	1,000	$\geq 0,500$	Подтверждение H_0
Коэффициент Сёрнсена–Дайса ³ (порог 0,02)	0,800	$\geq 0,667$	Подтверждение H_0^4
Коэффициент Сёрнсена–Дайса ³ (топ-3)	1,000	$\geq 0,667$	Подтверждение H_0
Процент совпадения доминирующих таксонов ³ , % (порог 0,02)	100	$\geq 66,7$	Подтверждение H_0^4
Процент совпадения доминирующих таксонов ³ , % (топ-3)	100	$\geq 66,7$	Подтверждение H_0
Ранговая корреляция Спирмена	0,947	$\geq 0,958$	Неопределённая зона ²
Ранговая корреляция Кендалла	0,859	$\geq 0,886$	Неопределённая зона ²
Общий индекс стабильности сообщества	0,942	$\geq 0,632$	Подтверждение H_0

Примечание. ¹ — пороги для H_0 соответствуют значениям из табл. 1 и гарантируют специфичность не менее 95%; ² — нахождение метрики в неопределённой зоне не опровергает гипотезу H_0 , но требует совместного анализа с другими показателями; ³ — для метрик, зависящих от способа выделения доминирующих таксонов, приведены значения как для порога 0,02, так и для метода «топ-3»; ⁴ — шкала табл. 1 разработана для метода «топ-3», поэтому интерпретация значений, полученных с использованием порога 0,02, дана ориентировочно, в случае применения метода «топ-3» все значения однозначно подтверждают H_0 .

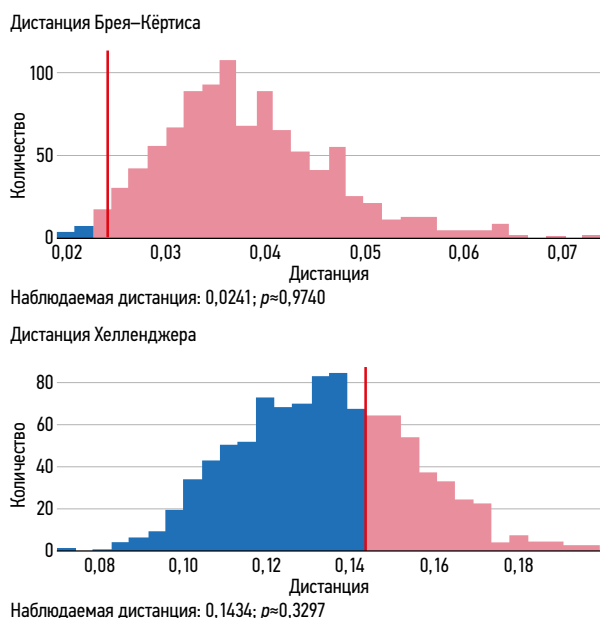


Рис. 1. Гистограммы дистанций Брея–Кёртиса и Хелленджера для 1000 бутстрэп-псевдовыборок, сгенерированных для данных из примера.

показавший наблюдаемые различия, не превышающие ожидаемых при случайной вариации внутри одного сообщества (рис. 1). Таким образом, несмотря на отдельные отклонения, совокупность данных однозначно свидетельствует в пользу происхождения экстрактов из одного водоёма.

ОБСУЖДЕНИЕ

Резюме основных результатов

Разработан формализованный метод ЦСТА, направленный на устранение ключевых методологических ограничений в судебно-медицинской альгологии. Они связаны с отсутствием алгоритмов объективной интерпретации диатомологических тестов и статистического обоснования экспертных выводов относительно сходства диатомовых сообществ в экстрактах планктона из трупного биоматериала и предполагаемой среды утонения. Основным результатом стала не только теоретическая проработка протокола ЦСТА, но и его реализация в виде доступного веб-приложения. Для объективизации шкалы интерпретации метрик впервые применено имитационное моделирование методом Монте-Карло, позволившее установить статистически обоснованные пороговые значения с контролируруемыми уровнями чувствительности и специфичности. Принципиальным преимуществом разработанной модели стал учёт широкого спектра факторов, приближающих симуляцию к реальным условиям судебно-медицинского анализа: переменное число активных таксонов, эффект редкости, ошибки идентификации и вариабельность объёмов проб. Это обеспечило высокую надёжность и практическую значимость полученных порогов.

Интерпретация результатов

Главной особенностью ЦСТА является создание методологии, адаптированной под специфическую задачу парного сравнения экстрактов планктона. Стандартные многомерные методы экологии для построения значимых моделей требуют множества проб из разных биотопов, что делает их неприменимыми в рутинной судебно-медицинской практике, где сравнивается всего два образца [9]. Предложенный протокол ЦСТА решает эту проблему путём комбинации взаимодополняющих метрик, каждая из которых применима для парного сравнения, и их статистической валидации через бутстрэп-тест.

Алгоритм ЦСТА заменяет субъективную визуальную оценку расчётом объективных, воспроизводимых величин. Ключевую роль играет использование дистанций Брея–Кёртиса и Хелленджера, которые, в отличие от простых индексов присутствия–отсутствия, интегрируют информацию об обилии таксонов, обеспечивая более тонкую и экологически обоснованную оценку общего сходства сообществ [13, 14].

Основопологающим элементом ЦСТА является внедрение процедуры бутстрэппинга для проверки значимости наблюдаемых дистанций. Этот подход позволяет ответить на вопрос, могли ли обнаруженные различия возникнуть случайно в условиях одного генерального сообщества, что предоставляет эксперту-альгологу ранее отсутствовавший в его арсенале статистически обоснованный критерий для принятия решения относительно происхождения диатомового планктона в трупном биоматериале.

Использование не какой-то одной, а системной комбинации метрик позволяет не только констатировать факт сходства или различия сравниваемых диатомовых сообществ, но и интерпретировать его природу [9, 24]. В частности, блок общего сходства даёт интегральную оценку и определяет степень сходства образцов в целом. Блок структуры доминирования смещает фокус на ядро сообщества, выявляя, совпадают ли доминирующие таксоны и их иерархия, что часто имеет решающее диагностическое значение. Блок стабильности и индикаторности позволяет оценить вклад каждого отдельного таксона в сходство или различие, выявляя потенциальные виды-индикаторы конкретного местообитания.

Для устранения субъективности при интерпретации численных значений метрик проведено имитационное моделирование методом Монте-Карло с учётом максимального числа реалистичных факторов. В модели варьировало число активных таксонов (от 10 до 50), что отражает естественную неполноту таксономических списков. Логнормальное распределение с регулируемой выравненностью (*evenness* от 0,3 до 3,0) позволило воспроизвести широкий спектр структур доминирования — от сообществ с резким преобладанием немногих видов до высокоравномерных. Порог редкости (0,001) имитировал ситуацию, когда редкие, но присутствующие в водоёме таксоны

могут не попасть в пробу. Для моделирования случайной вариации при отборе проб использовали мультиномиальное распределение, являющееся стандартным инструментом для описания многомерных дискретных данных. Введение случайных ошибок идентификации и подсчёта (с вероятностью 0,01) позволило оценить устойчивость метрик к типичным погрешностям лабораторного этапа. Наконец, в сценарии H_1 варьировали объёмы проб (200–2000 особей), что соответствует реальному разбросу числа обнаруженных створок. Совокупность перечисленных усовершенствований делает разработанную модель наиболее адекватной реальным условиям судебно-медицинского диатомового анализа среди известных подходов.

Разработка веб-приложения, воспроизводящего протокол ЦСТА, решает несколько прикладных задач:

- стандартизации;
- доступности;
- документирования.

Использование программы гарантирует, что разными экспертами расчёт однотипных метрик будет проводиться по единому алгоритму, минимизируя технические ошибки и вариабельность. Веб-интерфейс, не требующий установки сложного программного обеспечения и глубоких знаний в программировании, делает передовые методы статистики доступными для любого эксперта-альголога. Автоматическая генерация отчёта, включающего все рассчитанные показатели и графики, создаёт прозрачную и проверяемую основу для экспертного заключения, соответствующего принципам доказательной медицины и требованиям судопроизводства.

Таким образом, представленный метод ЦСТА и его программная реализация являются шагом к созданию цифровой экосистемы для эксперта-альголога, которая будет включать инструменты для цифровой микроскопии, автоматической таксономии и статистического сравнения проб [10, 22, 23]. Интеграция таких инструментов в единый рабочий процесс — перспективное направление для полной цифровой трансформации судебно-медицинской диатомологии [8, 10, 25].

Ограничения

Ограничения ЦСТА определяются его зависимостью от качества входных данных. Прежде всего ЦСТА мало пригоден для сравнения экстрактов планктона из водоёма и извлечённых из так называемых закрытых органов — органов (и/или их содержимого) из бассейна большого круга кровообращения (почек, печени, сердца, костного мозга и т. д.), поскольку форма и размеры створок диатомовых водорослей влияют на прохождение последних через аэрогематический барьер [6, 24, 26]. Данное обстоятельство закономерно приводит к искажениям таксономического состава диатомовых сообществ указанных органов. Не исключено, что избирательная проницаемость аэрогематического барьера может приводить к увеличению обилия крупноразмерных видов планктона

и в лёгких. В связи с этим наиболее предпочтительным объектом ЦСТА со стороны трупа следует считать жидкость из клиновидной пазухи и лёгкие при её отсутствии [27].

Кроме того, чтобы оба сравниваемых экстракта диатомового планктона были приготовлены с использованием одного метода химического разложения. В противном случае разная степень агрессивности, например кислотного и перекисного разложения, приведёт к разрушению панцирей наименее устойчивых видов диатомей, что обязательно негативно отразится на результатах ЦСТА [5, 28, 29].

Точность ЦСТА зависит от корректности идентификации и подсчёта диатомей на этапе исследования изготовленных минерализатов, поскольку метод не заменяет квалификацию альголога, а усиливает её на этапе интерпретации [3, 7].

Часть метрик — индексы Жаккара, Сёренсена–Дайса, процент совпадения доминирующих таксонов, стабильность присутствия и производные от неё композитные индексы — непосредственно зависят от способа определения множества доминирующих таксонов. Представленные в работе пороговые значения указанных метрик получены для метода выделения доминирующих таксонов «топ-3». При использовании иных способов определения доминирования (например, фиксированного порога обилия) необходима дополнительная калибровка порогов на основе имитационного моделирования по аналогии с выполненной в настоящей работе процедурой.

Важным методологическим допущением является фиксация числа доминирующих таксонов на уровне $k=3$. Хотя теоретически обоснован адаптивный выбор k (например, на основе кумулятивной кривой доминирования), в целях унификации и воспроизводимости результатов в протоколе ЦСТА используется фиксированное значение. Это может снижать чувствительность метода при анализе сообществ с очень низким или, напротив, высоким видовым богатством. Для водоёмов с обеднённой флорой целесообразно рассмотреть возможность увеличения k , что потребует дополнительной калибровки пороговых значений.

При интерпретации результатов ЦСТА следует учитывать, что приведённые в табл. 1 показатели чувствительности и специфичности относятся к каждой метрике в отдельности. Совокупную вероятность ошибочного вывода при одновременном использовании нескольких метрик в условиях комплексного анализа в настоящей работе не оценивали, однако она может отличаться от индивидуальных значений.

Ещё одним ограничением является способ формирования общего пула вероятностей p_0 в бутстрэп-процедуре. Используемое взвешенное среднее нормализованных векторов обилий может приводить к сглаживанию редких таксонов и завышению вклада доминирующих при существенных различиях между пробами. Разработка альтернативных подходов к построению p_0 , лишённых этих недостатков, представляет отдельную методологическую задачу. Вопрос о чувствительности результатов бутстрэп-теста к способу построения p_0 также требует

дополнительного исследования и выходит за рамки настоящей работы.

Предлагаемый метод ЦСТА представляет пилотную разработку, направленную на формализацию и объективизацию этапа интерпретации результатов диатомового анализа. На данный момент методика не прошла внешнюю валидацию на независимых данных, что является задачей дальнейших исследований. Представленные в работе пороговые значения и операционные характеристики получены исключительно на основе имитационного моделирования и требуют дополнительной проверки в реальных экспертных условиях.

ЗАКЛЮЧЕНИЕ

Разработанный метод ЦСТА и его программная реализация представляют комплексное решение задачи объективной и статистически валидируемой оценки сходства диатомовых сообществ из предполагаемого водоёма утопления и трупного биоматериала. Предложенный протокол преодолевает субъективность традиционного подхода за счёт использования системы взаимодополняющих метрик, охватывающих интегральное сходство, структуру доминирования и индикаторную значимость отдельных таксонов.

Ключевым элементом работы стало обоснование пороговых значений метрик методом имитационного моделирования Монте-Карло, которое учитывает широкий спектр факторов, приближающих модель к реальным условиям судебно-медицинского анализа: переменное число активных таксонов, эффект редкости, ошибки идентификации и вариабельность объёмов проб. Это позволило для каждой метрики установить количественные критерии «высокого сходства» с контролируруемыми уровнями специфичности и чувствительности не менее 95%. Тем самым интерпретация результатов диатомового анализа переводится из области субъективных экспертных оценок в строгую количественную плоскость, обеспеченную статистическими характеристиками.

Внедрение разработанного инструмента в судебно-медицинскую экспертную практику способно повысить доказательную силу, воспроизводимость и стандартизацию заключений по установлению места утопления, что соответствует общей тенденции к цифровизации и внедрению объективных методов в судебно-медицинскую экспертизу.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. В.Г. Недугов — участие в разработке протокола сравнительного анализа, написание кода онлайн-приложения, имитационное моделирование методом Монте-Карло, редактирование текста рукописи, рассмотрение и одобрение окончательного варианта рукописи; Г.В. Недугов — концепция и дизайн работы, разработка протокола сравнительного анализа, тестирование онлайн-приложения, сбор и анализ литературных данных, создание графических материалов, написание и редактирование текста рукописи. Все авторы одобрили рукопись (версию для публикации), а также согласились нести ответственность за все аспекты работы, гарантируя надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой её части.

Этическая экспертиза. Неприменимо.

Источники финансирования. Отсутствуют.

Раскрытие интересов. Авторы заявляют об отсутствии отношений, деятельности и интересов за последние три года, связанных с третьими лицами (коммерческими и некоммерческими), интересы которых могут быть затронуты содержанием статьи.

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения (текст, иллюстрации, данные).

Доступ к данным. Авторы сообщают, что все данные представлены в статье.

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовали.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовал один член редакционной коллегии журнала.

ADDITIONAL INFORMATION

Author contributions: V.G. Nedugov: methodology, software, formal analysis, writing—review & editing; G.V. Nedugov: conceptualization, methodology,

investigation, data curation, visualization, writing—original draft, writing—review & editing. All the authors approved the version of the manuscript to be published and agreed to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval: Not applicable.

Funding sources: No funding.

Disclosure of interests: The authors have no relationships, activities, or interests for the last three years related to for-profit or not-for-profit third parties whose interests may be affected by the content of the article.

Statement of originality: No previously published materials (text, images, or data) were used in this article.

Data availability statement: All data obtained in this study are available in this article.

Generative AI: No generative artificial intelligence technologies were used to prepare this article.

Provenance and peer-review: This article was submitted unsolicited and reviewed following the standard procedure. The peer-review process involved a member of the Editorial Board.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

- Tournois L, Hatsch D, Ludes B, Delabarde T. Automatic detection and identification of diatoms in complex background for suspected drowning cases through object detection models. *International Journal of Legal Medicine*. 2023;138(2):659–670. doi: 10.1007/s00414-023-03096-w EDN: YNYGHS
- Ludes BP, Chambre A, Delabarde T. The diatom test in the field of forensic medicine: a review of a long-standing question. *International Journal of Legal Medicine*. 2024;139(2):597–605. doi: 10.1007/s00414-024-03370-5 EDN: COETLL
- Tyr A, Lunetta P, Zilg B, et al. The medico-legal interpretation of diatom findings for the diagnosis of fatal drowning: a systematic review. *International Journal of Legal Medicine*. 2025;139(2):729–746. doi: 10.1007/s00414-024-03397-8 EDN: WYQIGL
- Carballeira R, Vieira DN, Febrero-Bande M, Muñoz Barús JJ. A valid method to determine the site of drowning. *International Journal of Legal Medicine*. 2017;132(2):487–497. doi: 10.1007/s00414-017-1708-1 EDN: LYFYBR
- Bhardwaj N, Ahluwalia AS, Pal SK, Mandotra SK. Potential of photoautotrophic microbial organisms in deciphering forensic issues. *Legal Medicine*. 2023;62:102223. doi: 10.1016/j.legalmed.2023.102223 EDN: XUIFGQ
- Bogusz M, Bogusz I, Żelazna-Wieczorek J. The possibilities and limitations of comparative diatomaceous analysis for confirming or excluding the site of an incident – Case studies. *Forensic Science International*. 2023;346:111644. doi: 10.1016/j.forsciint.2023.111644 EDN: OZIOTS
- Wu Y, Luo L, Li Y, et al. A comparative assessment of time-consuming and laborious diatom analysis—Brief experimentation with suggestion of automatic identification. *Forensic Science International*. 2024;355:111939. doi: 10.1016/j.forsciint.2024.111939 EDN: LHWLDI
- Tambuzzi S, Gentile G, Zoia R. Forensic Diatom Analysis: Where Do We Stand and What Are the Latest Diagnostic Advances? *Diagnostics*. 2024;14(20):2302. doi: 10.3390/diagnostics14202302 EDN: HDYOJX
- Legendre P, Legendre L. *Numerical ecology*. 3rd English edition. Amsterdam: Elsevier; 2012. ISBN: 9780444538697 Available from: <https://shop.elsevier.com/books/numerical-ecology/legendre/978-0-444-53868-0>
- Zhang J, Zhou Y, Vieira DN, et al. An efficient method for building a database of diatom populations for drowning site inference using a deep learning algorithm. *International Journal of Legal Medicine*. 2021;135(3):817–827. doi: 10.1007/s00414-020-02497-5 EDN: TIUNQQ
- Ludes B, Coste M, North N, et al. Diatom analysis in victim's tissues as an indicator of the site of drowning. *International Journal of Legal Medicine*. 1999;112(3):163–166. doi: 10.1007/s004140050224 EDN: AWDVLH
- Hürlimann J, Feer P, Elber F, et al. Diatom detection in the diagnosis of death by drowning. *International Journal of Legal Medicine*. 2000;114(1-2):6–14. doi: 10.1007/s004149900122 EDN: AVOFFT
- Bray JR, Curtis JT. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*. 1957;27(4):325–349. doi: 10.2307/1942268
- Hellinger E. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*. 1909;1909(136):210–271. doi: 10.1515/crll.1909.136.210
- Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist*. 1912;11(2):37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x
- Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302. doi: 10.2307/1932409
- Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter/Kongelige Danske Videnskaberne Selskab*. 1948;5(4):1–34.
- Simpson EH. Measurement of Diversity. *Nature*. 1949;163(4148):688–688. doi: 10.1038/163688a0
- Dufrêne M, Legendre P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*. 1997;67(3):345–366. doi: 10.2307/2963459
- Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. *Oecologia*. 2001;129(2):271–280. doi: 10.1007/s004420100716 EDN: ARUQVZ
- Horton BP, Boreham S, Hillier C. The development and application of a diatom-based quantitative reconstruction technique in forensic science. *Journal of Forensic Sciences*. 2006;51(3):643–650. doi: 10.1111/j.1556-4029.2006.00120.x
- Zhang J, Vieira DN, Cheng Q, et al. DiatomNet v1.0: A novel approach for automatic diatom testing for drowning diagnosis in forensically biomedical application. *Computer Methods and Programs in Biomedicine*. 2023;232:107434. doi: 10.1016/j.cmpb.2023.107434 EDN: SWTBQI
- Venkataramanan A, Kloster M, Burfeid-Castellanos A, et al. "UDE DIATOMS in the Wild 2024": a new image dataset of freshwater diatoms for training deep learning models. *GigaScience*. 2024;13:giae087. doi: 10.1093/gigascience/giae087 EDN: VGDUUH
- Du Y, Liu J, Li Q, et al. Concordance analysis of diatom types and patterns in lung tissue and drowning medium in laboratory animal model. *International Journal of Legal Medicine*. 2022;136(3):911–917. doi: 10.1007/s00414-021-02768-9 EDN: FIQOZZ
- Ma A, Chen M, Yu Q, et al. A simple optical microscopy-based diatom classification approach for forensic drowning site inference: a pilot study. *International Journal of Legal Medicine*. 2025;139(5):2207–2218. doi: 10.1007/s00414-025-03520-3 EDN: DFPZKZ

26. Lunetta P, Penttilä A, Hällfors G. Scanning and transmission electron microscopical evidence of the capacity of diatoms to penetrate the alveolo-capillary barrier in drowning. *International Journal of Legal Medicine*. 1998;111(5):229–237. doi: 10.1007/s004140050159 EDN: AWUXQN

27. Lin CY, Yen WC, Hsieh HM, et al. Diatomological investigation in sphenoid sinus fluid and lung tissue from cases of suspected drowning. *Forensic Science International*. 2014;244:111–115. doi: 10.1016/j.forsciint.2014.08.023

28. Ming M, Meng X, Wang E. Evaluation of four digestive methods for extracting diatoms. *Forensic Science International*. 2007;170(1):29–34. doi: 10.1016/j.forsciint.2006.08.022

29. Kakizaki E, Sonoda A, Shinkawa N, Yukawa N. A new enzymatic method for extracting diatoms from organs of suspected drowning cases using papain: Optimal digestion and first practical application. *Forensic Science International*. 2019;297:204–216. doi: 10.1016/j.forsciint.2019.02.008

ОБ АВТОРАХ

* **Недугов Герман Владимирович**, д-р мед. наук, доцент;
адрес: Россия, 443099, Самара, ул. Чапаевская, д. 89;
ORCID: 0000-0002-7380-3766;
eLibrary SPIN: 3828-8091;
e-mail: nedugovh@mail.ru

Недугов Владимир Германович;
ORCID: 0009-0007-7542-7235;
eLibrary SPIN: 2407-7937;
e-mail: nedugovvg@gmail.com

AUTHORS' INFO

* **German V. Nedugov**, MD, Dr. Sci. (Medicine), Assistant Professor;
address: 89 Chapaevskaya st, Samara, Russia, 443099;
ORCID: 0000-0002-7380-3766;
eLibrary SPIN: 3828-8091;
e-mail: nedugovh@mail.ru

Vladimir G. Nedugov;
ORCID: 0009-0007-7542-7235;
eLibrary SPIN: 2407-7937;
e-mail: nedugovvg@gmail.com

* Автор, ответственный за переписку / Corresponding author