

Применение доменной адаптации для диагностики природных вод по спектрам комбинационного рассеяния света

Л. С. Утегенова* К. А. Бузанов†

Московский государственный университет имени М. В. Ломоносова,
физический факультет, кафедра квантовой электроники
Россия, 119991, Москва, Ленинские горы, д. 1, стр. 2

(Поступила в редакцию 11.06.2025; подписана в печать 03.07.2025)

Определение ионного состава и концентрации веществ в речных и сточных водах представляет собой важную научно-практическую задачу экологического мониторинга. В настоящей работе решается задача по определению концентраций ионов Zn^{2+} , Cu^{2+} , Li^{+} , Fe^{3+} , Ni^{2+} , NH_4^{+} , SO_4^{2-} и NO_3^{-} в водных растворах с помощью спектроскопии комбинационного рассеяния (КР) и искусственных нейронных сетей (НС). Спектроскопия КР позволяет проводить быстрый и дистанционный анализ многокомпонентных жидких сред в режиме реального времени. Однако анализ спектров сталкивается с рядом проблем вследствие сложного химического состава природных вод и наличия флуоресцентного фона, обусловленного растворёнными органическими веществами. Поэтому необходимо использовать НС. В рамках исследования с целью доменной адаптации НС реализуется метод переноса обучения, основанный на предварительном обучении моделей на репрезентативной выборке спектральных данных, не содержащих флуоресцентной составляющей (исходный домен), и дальнейшем дообучении НС на спектрах с флуоресцентным фоном (целевой домен). Такой подход позволил повысить эффективность обучения модели и точность определения концентраций ионов в реальных речных водах (Москва-река, Яуза, Битца, Сетунь).

PACS: 33.20.-t.

УДК: 53.06

Ключевые слова: спектроскопия комбинационного рассеяния, экологический мониторинг, ионный состав, искусственные нейронные сети, доменная адаптация, перенос обучения.

ВВЕДЕНИЕ

Тяжелые металлы, обладая высокой токсичностью и способностью к биоаккумуляции, представляют серьезную угрозу для водных экосистем и здоровья человека [1–3]. В связи с этим разработка эффективных методов мониторинга концентраций тяжелых металлов в природных водах является одной из актуальных задач современной экологии. Существующие лабораторные методы анализа, такие как ионная хроматография, амперометрия, потенциометрия, проточный инжекционный анализ, вольтамперометрия, несмотря на высокую точность, требуют значительных временных затрат и сложной пробоподготовки, что делает их непригодными для оперативного мониторинга. Для дистанционного и экспрессного анализа многокомпонентных водных сред мощным инструментом является метод лазерной спектроскопии комбинационного рассеяния (КР) света, который обладает возможностью неинвазивных измерений и высокой скоростью анализа [4–8]. В настоящее время существует большое количество научных работ, демонстрирующих хорошую чувствительность спектроскопических методов и высокий потенциал их практического применения в различных областях — от пищевой промышленности до экологического мониторинга [9–13].

Как показывают исследования, традиционные калибровочные методы, основанные, например на концентрационной зависимости площади под валентной полосой воды [14] или положения и интенсивности характерных пиков [15], имеют существенные недостатки. В частности, они эффективны только для однокомпонентных систем, демонстрируют низкую точность при анализе природных вод и не учитывают комплексные межмолекулярные взаимодействия, что значительно ограничивает их практическое применение для мониторинга реальных водных сред. Однако подобные методы позволяют решить самые разные задачи, такие как определение хлоридности водных растворов [16] и количества микропластика в водных средах [17].

Современные перспективы развития методов анализа водных сред связаны с активным внедрением технологий машинного обучения. Особый научный и практический интерес представляет применение искусственных нейронных сетей, обладающих рядом уникальных преимуществ. Эти алгоритмы демонстрируют высокую эффективность при выявлении сложных нелинейных зависимостей в спектроскопических данных, что особенно важно при анализе многокомпонентных систем [18]. Важным свойством НС является их устойчивость к различным мешающим факторам, таким как флуоресцентный фон или шумы измерений. Однако применение адаптивных моделей сталкивается с проблемой дефицита обучающих данных, особенно при анализе реальных природных и технологических водных сред, где изменчивость состава и условий существенно усложняет формирование репрезентативных выборок. В данном исследовании предлагаются следующие под-

* utegenova.ls20@physics.msu.ru

† buzanov.ka21@physics.msu.ru

ходы, призванные обойти проблему нехватки данных. Во-первых, используется обучение НС как на большой базе спектров модельных растворов в дистиллированной воде (исходный домен), так и на малой базе спектров реальных речных вод Москвы-реки, Яузы, Битцы и Сетуни (целевой домен). Во-вторых, решением проблемы объединения разрозненных баз данных целевого и исходного доменов, необходимых для обучения НС, может стать доменная адаптация (ДА) — направление машинного обучения, которое исследует методы улучшения точности моделей в целевом домене за счет переноса знаний из исходного домена, где объем размеченной выборки достаточен для обучения [19]. Современные исследования подтверждают высокую эффективность методов доменной адаптации нейронных сетей в решении обратных задач лазерной спектроскопии [20–22]. В настоящей работе процедура ДА была проведена с использованием метода переноса обучения [23].

1. ОБЪЕКТЫ ИССЛЕДОВАНИЯ

В данной работе исследовались растворы, содержащие 10 неорганических солей: $\text{Zn}(\text{NO}_3)_2$, $\text{Zn}(\text{SO}_4)$, $\text{Cu}(\text{NO}_3)_2$, $\text{Cu}(\text{SO}_4)$, $\text{Li}(\text{NO}_3)$, $\text{Fe}(\text{NO}_3)_3$, $\text{Ni}(\text{SO}_4)$, $\text{Ni}(\text{NO}_3)_2$, $(\text{NH}_4)_2\text{SO}_4$, $\text{NH}_4(\text{NO}_3)$ (производство «Бамлюкс»). Исследуемые растворы содержали 8 ионов: 6 катионов (Zn^{2+} , Cu^{2+} , Li^+ , Fe^{3+} , Ni^{2+} , NH_4^+) и 2 анионов (SO_4^{2-} , NO_3^-). Для приготовления водных растворов солей использовалась бидистиллированная деионизованная вода (Millipore Simplicity UV water purification system).

Были приготовлены следующие серии растворов неорганических солей, различающихся растворителем:

- Серия 1: 3744 водных раствора солей в дистиллированной воде и 18 образцов дистиллированной воды (модельные растворы)
- Серия 2: 400 растворов в речной воде (Москва-река)
- Серия 3: 200 растворов в речной воде (Яуза)
- Серия 4: 200 растворов в речной воде (Битца)
- Серия 5: 200 растворов в речной воде (Сетунь)

Концентрация ионов в приготовленных растворах изменялась от 0 до максимальной, которая указана для каждого иона в таблице.

2. СПЕКТРОСКОПИЯ КОМБИНАЦИОННОГО РАССЕЯНИЯ СВЕТА

Исследования по спектроскопии КР света и флуоресценции водных растворов солей проводились на лазерном спектрометре. Принципиальная схема лазерного спектрометра КР показана на рис. 1. Возбуждение

Таблица. Максимальные концентрации ионов в приготовленных растворах

Ион	Максимальная концентрация, М
Zn^{2+}	1.089
Cu^{2+}	0.955
Li^+	0.466
Fe^{3+}	0.862
Ni^{2+}	0.972
NH_4^+	0.801
SO_4^{2-}	1.373
NO_3^-	4.906

сигналов КР и флуоресценции образцов осуществлялось излучением второй гармоники ИАГ-лазера (длина волны 532 нм, мощность 500 мВт). Зондирование образцов проводилось через дно кюветы. Контроль мощности лазерного излучения осуществлялся с помощью пластинки, отводящей 5% возбуждающего излучения на головку измерителя мощности излучения (Ophir 3-AP). Интегральные спектры (поляризованная и деполяризованная части вместе) регистрировались в 90° схеме эксперимента.

Система регистрации состояла из монохроматора (Acton 2500i, фокусное расстояние 500 мм, решётка 900 или 1800 штр/мм) и CCD камеры (Horiba Jobin Yvon, Sincerity-1024x256-OE). Ширина входной щели монохроматора составляла 50 мкм, что обеспечивало практическое разрешение 1.5 см^{-1} . Для подавления рассеяния на несмещённой частоте использовались интерференционные фильтры Semrock или фильтр Omega Optical. Система термостабилизации на основе элемента Пельтье/жидкого термостата «Сгуо-VT-01» позволяла контролировать температуру образца с точностью $0.1^\circ\text{C}/0.2^\circ\text{C}$. Полученные спектры регистрировались на чувствительность регистрационной системы.

3. АНАЛИЗ ВЛИЯНИЯ ИОНОВ НА СПЕКТРЫ КР ВОДНЫХ РАСТВОРОВ

В ходе работы был проведён анализ влияния каждого из 8 ионов: 6 катионов (Zn^{2+} , Cu^{2+} , Li^+ , Fe^{3+} , Ni^{2+} , NH_4^+) и 2 анионов (SO_4^{2-} , NO_3^-) на спектры КР.

В приготовленных модельных растворах содержалось 6 однокомпонентных (содержащих 1 катион) растворов. Низкочастотная область спектров КР (с центром диапазона 560 нм) однокомпонентных растворов представлена на рис. 2. Поканальные статистические показатели в анализируемых спектрах позволяют количественно оценить вклад отдельных ионов в спектры КР. Величина стандартного отклонения в каждом

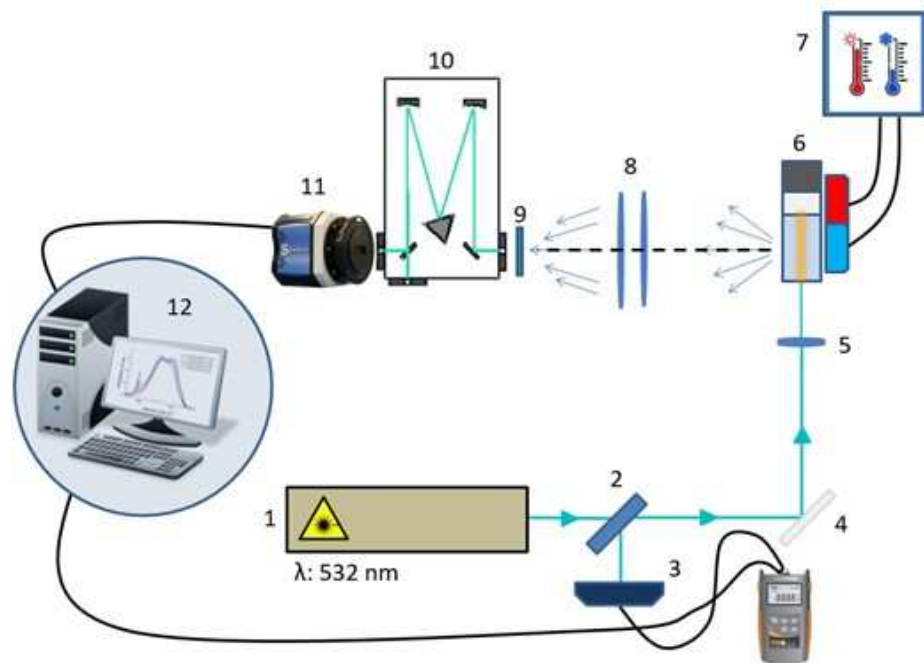


Рис. 1. Схема лазерного спектрометра: 1 — лазер накачки; 2 — отводная пластинка; 3 — измеритель мощности лазерного излучения; 4 — зеркало; 5 — фокусирующая линза; 6 — термостабилизированная кювета; 7 — термостат; 8 — система светосбора; 9 — фильтр; 10 — монохроматор; 11 — CCD-камера; 12 — компьютер

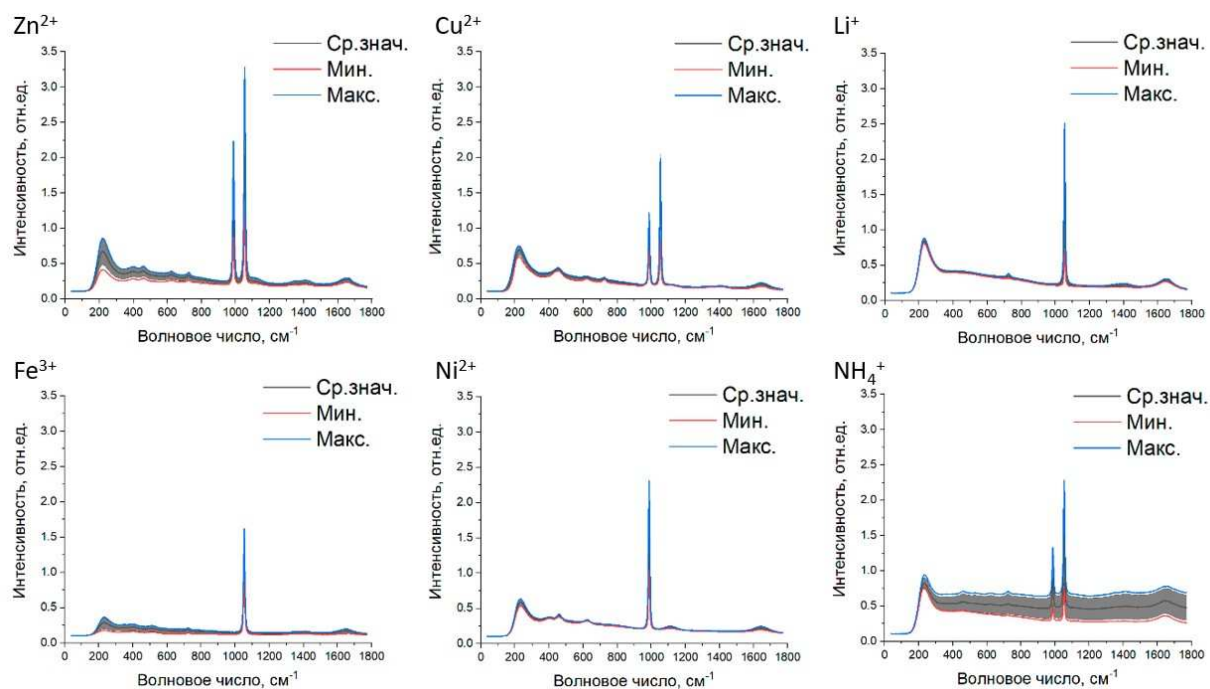


Рис. 2. Поканальные статистические показатели низкочастотных областей спектров КР однокаатионных растворов (центр диапазона 560 nm). Серым цветом обозначено значение стандартного отклонения в данном спектральном канале по набору данных

спектральном канале коррелирует со степенью влияния соответствующего иона на формирование спектра КР: более высокие значения стандартного отклонения

указывают на более значимый вклад данного иона. Число спектральных каналов в представленном диапазоне равнялось 1024.

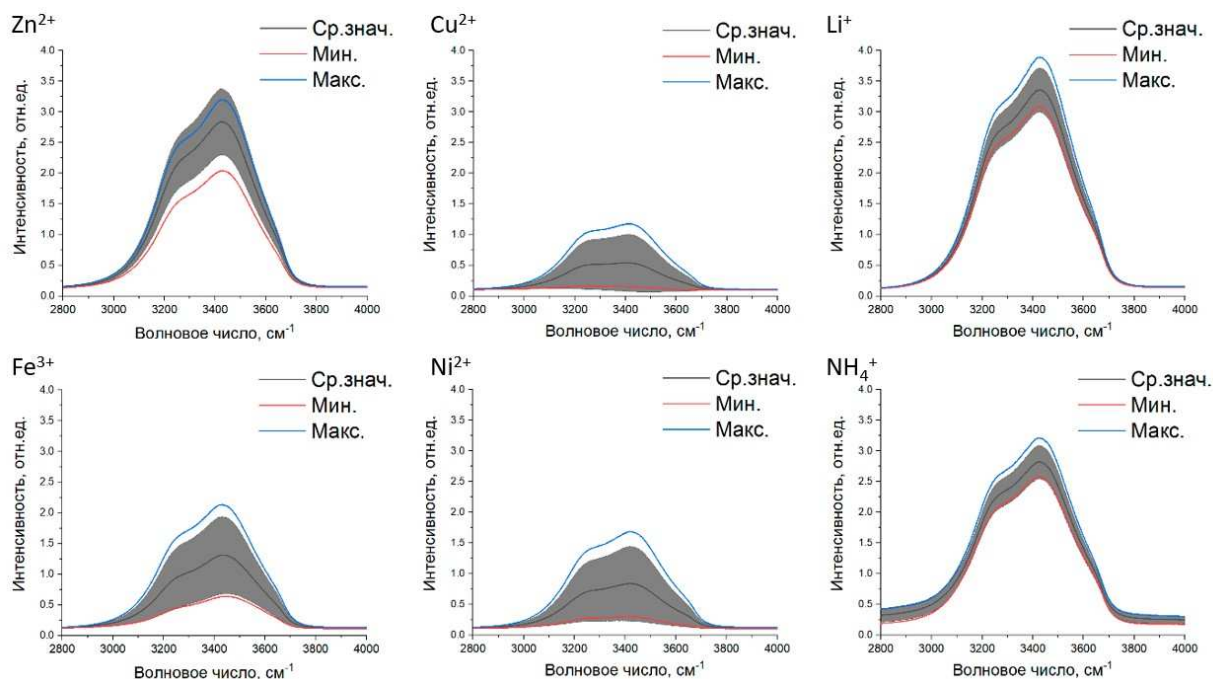


Рис. 3. Поканальные статистические показатели валентных полос КР (центр диапазона 650 нм) однокатионных растворов. Серым цветом обозначено значение стандартного отклонения в данном спектральном канале по набору данных

На спектрах видны характерные полосы SO_4^{2-} (1030–940 cm^{-1}) и NO_3^- (1100–1020 cm^{-1}). NH_4^+ имеет полосу с максимумом в районе 1600 cm^{-1} , однако она перекрывается с деформационной полосой воды. Другие ионы (Zn^{2+} , Cu^{2+} , Fe^{3+} , Li^+ , Ni^{2+}) не имеют характерных линий в спектрах КР.

На рис. 3 представлены поканальные статистические показатели валентных полос ОН-групп (3800–2800 cm^{-1}) спектров КР однокомпонентных растворов (с центром диапазона 650 нм). Число спектральных каналов в данном диапазоне равнялось 881. Как следует из представленных валентных полос, все катионы вносят довольно существенный вклад в изменение формы полосы валентных колебаний молекул воды.

Были получены спектры КР каждого из приготовленных растворов солей в двух диапазонах с центрами 560 нм и 650 нм.

4. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ К ЗАДАЧЕ ОПРЕДЕЛЕНИЯ КОНЦЕНТРАЦИЙ ИОНОВ В ВОДЕ

На первом этапе решения поставленной задачи к пяти имеющимся базам данных были применены НС с различной архитектурой. В ходе поиска оптимального количества скрытых слоёв и количества нейронов был выбран многослойный перцептрон, имеющий 2 скрытых слоя по 32 и 8 нейронов, соответственно (рис. 4). На вход НС подавались объединенные спектры — 2 диапазона с центрами 560 нм и 650 нм. Размерность входных данных равнялась 1905, размер-

ность выходных данных равнялась 1 (концентрация конкретного иона). В качестве функции потерь, показывающей ошибку в предсказаниях сети во время обучения, была выбрана среднеквадратичная ошибка (СКО). Активационной функцией являлась сигмоида, а для оптимизации обучения был выбран алгоритм Adam.

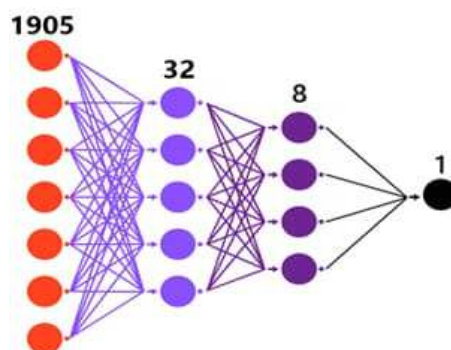
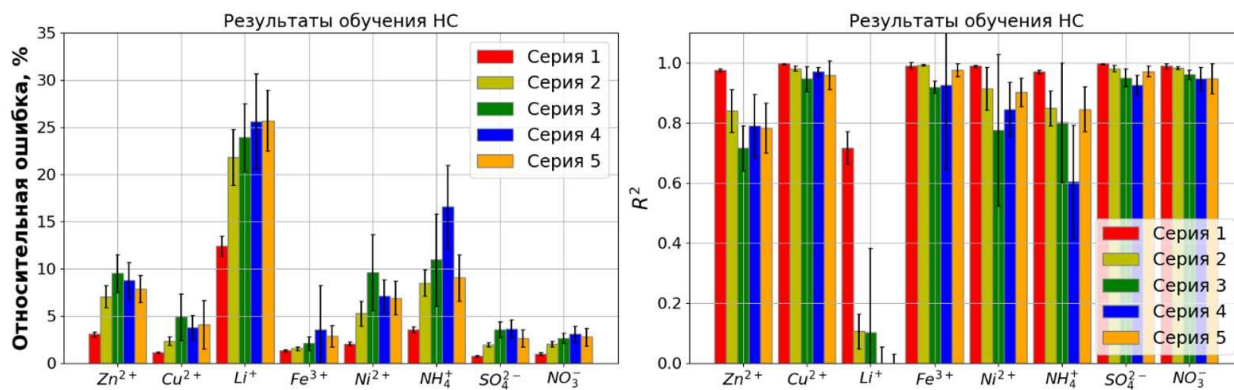


Рис. 4. Архитектура модели

При обучении НС очень важным является оценка точности, с которой НС решает поставленную задачу. Кроме того, одной из проблем при обучении НС является переобучение, связанное с тем, что НС перестает учиться на обобщенных признаках, а начинает воспринимать специфические признаки каждого примера. Поэтому вся база данных была разделена на три набора в соотношении 7:2:1: на обучающую выборку — данные, на которых непосредственно обучалась НС, вали-

Рис. 5. Относительная ошибка и R^2 определения концентраций ионов для пяти баз спектральных данных

дационную выборку, на которой в ходе обучения происходила проверка переобучения, и тестовую выборку — данные, на которых происходило тестирование НС после обучения.

Для минимизации влияния специфики разбиения данных на оценку эффективности модели проводилось пятикратное усреднение результатов по различным вариантам случайного разделения данных. Дополнительно для каждого варианта разбиения выполнялось пятикратное усреднение по независимым запускам обучения с различной случайной инициализацией весов нейронной сети, что позволило исключить зависимость результатов от начальных условий оптимизации.

Для определения качества обученной модели были выбраны следующие метрики: средняя абсолютная ошибка (САО), относительная ошибка (ОО) и скорректированный коэффициент детерминации (R^2):

$$\text{САО} = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|, \quad (1)$$

$$\text{ОО} = \frac{\text{САО}}{\Delta C} \cdot 100\%, \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (\bar{Y} - Y_i)^2}, \quad (3)$$

где X_i — значение, определённое моделью; Y_i — истинное значение; ΔC — диапазон изменения концентрации; $\bar{Y}$ — выборочное среднее.

5. ПЕРЕНОС ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ

В настоящей работе для повышения качества решения поставленной задачи в речной воде использовался метод переноса обучения (Transfer Learning), при котором НС, предварительно обученная на базе данных модельных растворов, адаптировалась для анализа природной речной воды.

Основная идея метода заключается в следующем: если одна задача решается относительно просто (например, в случае базовой серии данных, где высокая репрезентативность и отсутствие флуоресцентного фона упрощают обучение), то натренированную на ней модель можно использовать в качестве стартовой точки для решения более сложной задачи. Вместо случайной инициализации весов НС применяются веса модели, уже обученной на более простых данных.

Согласно теории глубокого обучения, начальные слои НС отвечают за выделение общих закономерностей, тогда как последние слои специализируются на более тонкой настройке зависимости, улавливаемой НС. В рассматриваемом методе происходит обучение на новой базе данных предобученной НС с «замороженными» первыми слоями (веса первых слоёв остаются неизменными в ходе обучения).

На начальном этапе исследования для определения оптимального подхода к переносу обучения были протестированы три варианта настройки нейронной сети:

- *Полное дообучение* — все слои НС обучались без заморозки весов;
- *Частичная заморозка* — фиксация весов первого скрытого слоя с дообучением остальных;
- *Двойная заморозка* — фиксация весов первых двух скрытых слоёв с дообучением последнего.

Эксперименты показали, что наилучшие результаты достигаются при полном дообучении сети без заморозки слоёв. Это можно объяснить относительно небольшой глубиной архитектуры (всего 3 слоя), из-за чего разделение на «общие» и «специфические» признаки становится менее выраженным, и все слои вносят сопоставимый вклад в решение задачи. Далее результаты по переносу обучения НС приведены для случая полного дообучения.

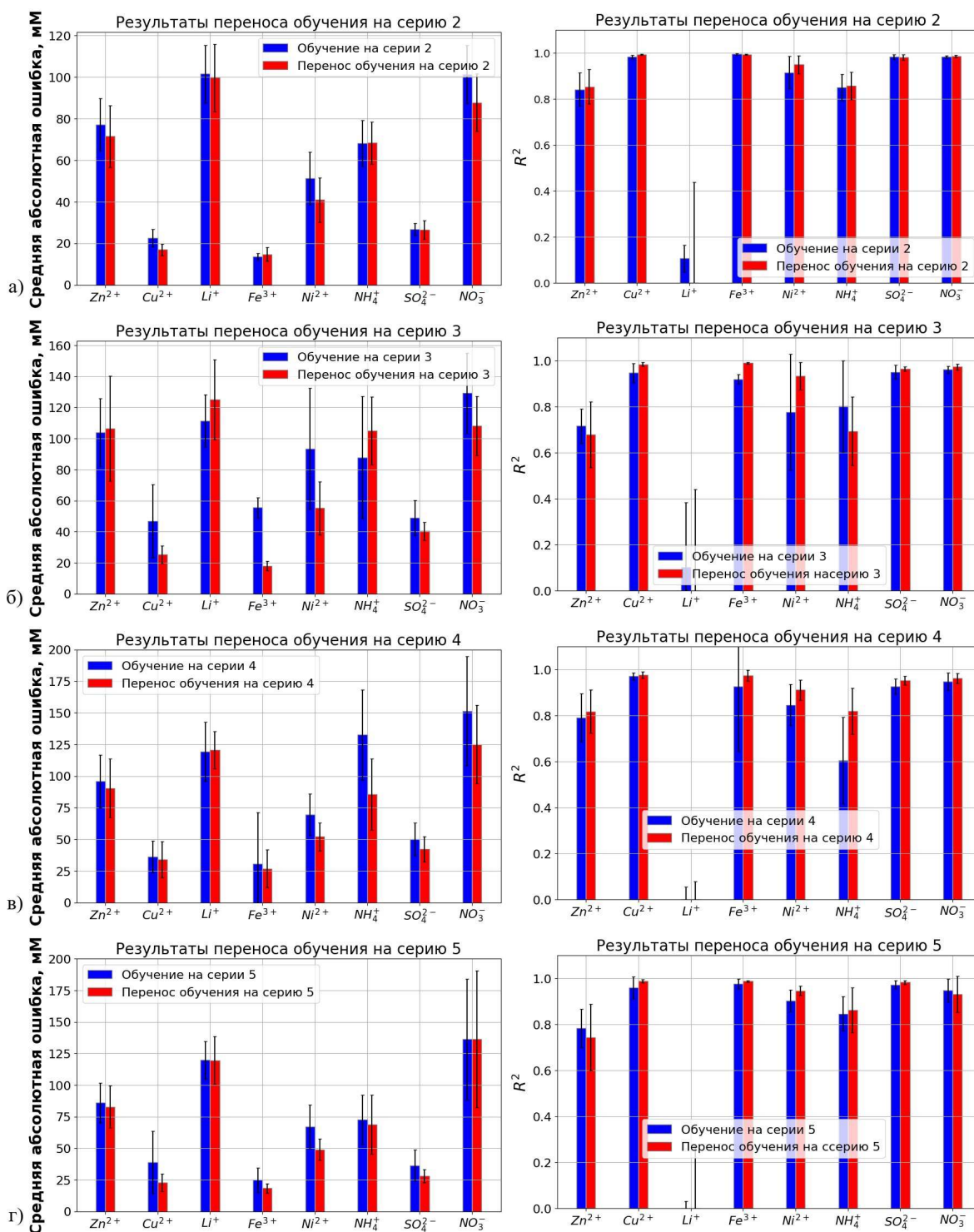


Рис. 6. САО и R^2 определения концентраций ионов с помощью метода переноса обучения для баз данных речных растворов: а — Москвы-реки, б — Яузы, в — Битцы, г — Сетуни

6. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На рис. 5 представлены относительная ошибка и R^2 , полученные при обучении на пяти сериях для каждого

иона. Как можно видеть на данных гистограммах, точность определения концентраций ионов в речных водах ниже, чем в дистиллированной воде, что обуславливается наличием флуоресцентного фона в спектрах КР речной воды.

На рис. 6 представлены CAO и R^2 определения концентраций ионов НС, обученными только на базах данных целевого домена и с помощью метода переноса обучения. Метод переноса обучения оказался эффективен для ряда ионов в сравнении с обучением НС без переноса обучения: удалось уменьшить CAO определения концентраций ионов Cu^{2+} , Fe^{3+} и Ni^{2+} во всех речных растворах, в среднем, на 31%, 35% и 31% соответственно. Для оставшихся ионов Zn^{2+} , NH_4^+ , SO_4^{2-} и NO_3^- удалось понизить CAO только на 2%, 7%, 19% и 11% соответственно. Результаты для катиона Li^+ в данном случае не рассчитываются, поскольку обратную задачу по определению концентрации этого иона, в принципе, не удалось решить с применением НС.

ЗАКЛЮЧЕНИЕ

Проведенное исследование продемонстрировало эффективность применения метода переноса обучения НС для решения задачи определения ионного состава речных вод по спектрам КР света.

Разработанная методика показала заметное улучшение точности определения концентраций ионов по сравнению с обучением только на малой базе целевого домена. Применение переноса обучения позволило повысить точность решения поставленной задачи в среднем для всех ионов:

- на 7% для серии 2 (растворы Москвы-реки);
- на 11% для серии 3 (растворы Яузы);
- на 16% для серий 4 и 5 (растворы Битцы и Сетуни).

Полученные результаты позволяют сделать вывод о перспективности дальнейшего развития методов доменной адаптации для задач спектроскопического анализа природных и сточных вод.

Л.С. Утегенова выражает благодарность Фонду развития теоретической физики и математики «БАЗИС» за финансовую поддержку научного проекта № 24-2-1-72-1.

-
- [1] Бактыбаева З.Б., Сулейманов Р.А., Валеев Т.К., Рахматуллин Н.Р. // Здоровье населения и среда обитания. **7**, № 292. 14 (2017).
- [2] Costa M. // Toxicology and Applied Pharmacology. **375**. 1 (2019).
- [3] Azevedo E., Dias C., Dantas R. et al. // 2012 IEEE Latin America Conference on Cloud Computing and Communications (LatinCloud). Porto Alegre, Brazil, 43–48 (2012).
- [4] Li Z., Deen M., Kumar S., Selvaganapathy P. // Sensors. **14**, N 9. 17275 (2014).
- [5] Li Z., Wang J., Li D. // Applied Spectroscopy Reviews. **51**, N 4. 333 (2015).
- [6] Шаманаев В.С. // Оптика атмосферы и океана. **33**, № 7. 516 (2020).
- [7] Климкин А.В., Карапузиков А.А., Коханенко Г.П. и др. // Оптика атмосферы и океана. **33**, № 3. 205 (2020).
- [8] Patsaeva S.V., Dolenko T.A., Burikov S.A., Yuzhakova V.I. // Atmospheric and Oceanic Optics. **27**, N 4. 291 (2014).
- [9] Guo Z., Dong R., Ye W. // DOI: 10.2139/ssrn.4950034.
- [10] Post C., Heyden N., Reinartz A. et al. // Sensors. **22**, N 13. 4668 (2022).
- [11] Li Z., Wang J., Li D. // Appl. Spectroscopy Rev. **51**, N 4. 313 (2016).
- [12] Kirsanov D., Babain V., Agafonova-Moroz M. et al. // Radiochimica Acta. **100**, N 3. 185 (2012).
- [13] Sadecka J., Jakubiková M. // Journal of Food Science and Technology. **57**, N 7. 2545 (2020).
- [14] Tian Z. et al. // Analytical Letters. **53**, N 13. 2034 (2020).
- [15] Durickovic I., Marchetti M. // IET Science, Measurement & Technology. **8**, N 3. 122 (2014).
- [16] Dubessy J., et al. // Applied Spectroscopy. **56**, N 1. 99 (2002).
- [17] Luo Y., et al. // Science of The Total Environment. **926**. 171925 (2024).
- [18] Dolenko S., Isaev I., Burikov S., Dolenko T. et al. // Pattern Recognition and Information Processing (PRIP'2023). Minsk: BSU, 2023.
- [19] Farahani A., Voghoei S., Rasheed K., Arabnia H.R. // Advances in Data Science and Information Engineering. (2021).
- [20] Bowler A.L., Ozturk S., Rady A., Watson N. // Sensors. **22**, N 19. 7239 (2022).
- [21] Zhang Z., Zhong H., Li Y. et al. // Cellulose. **31**. 551 (2024).
- [22] Utegenova L.S., Sarmanova O.E., Burikov S.A. et al. // Bull. Russ. Acad. Sci. Phys. **88** (Suppl 2). S160 (2024).
- [23] Pan S., Yang Q. // IEEE TKDE. (2010).

Application of Domain Adaptation for Diagnostics of Natural Waters by Raman Spectroscopy

L.S. Utegenova^a, K.A. Buzanov^b

¹Faculty of Physics, Lomonosov Moscow State University. Moscow 119991, Russia
E-mail: ^autegenova.ls20@physics.msu.ru, ^bbuzanov.ka21@physics.msu.ru

Determination of ionic composition and concentration of substances in river and waste water is an important scientific and practical task of environmental monitoring. In this paper, the problem of determining the concentrations of Zn^{2+} , Cu^{2+} , Li^{+} , Fe^{3+} , Ni^{2+} , NH_4^{+} , SO_4^{2-} and NO_3^{-} ions in aqueous solutions is solved using Raman spectroscopy and artificial neural networks (NN). Raman spectroscopy allows to perform rapid and remote analysis of multicomponent liquid media in real time. However, spectral analysis faces a number of problems due to the complex chemical composition of natural waters and the presence of fluorescent background caused by dissolved organic substances. Therefore, NN must be used. In the study, the transfer learning method is implemented for the purpose of domain adaptation of the NN, based on preliminary training of models on a representative sample of spectral data that do not contain a fluorescent component (source domain), and further additional training of the NN on spectra with a fluorescent background (target domain). This approach made it possible to increase the efficiency of model training and the accuracy of determining ion concentrations in real river waters (Moscow River, Yauza, Bitsa, Setun).

PACS: 42.62.Fi, 87.64.kp, 07.05.Mh.

Keywords: Raman spectroscopy, environmental monitoring, ionic composition, artificial neural networks, domain adaptation, transfer learning.

Received 11 June 2025.

Сведения об авторах

1. Утегенова Луиза Салаватовна — студентка; e-mail: utegenova.ls20@physics.msu.ru.
2. Бузанов Кирилл Александрович — студент; e-mail: buzanov.ka21@physics.msu.ru.