

Распознавание поверхностно-усиленных рамановских спектров органических сред на основе глубокого обучения

Л.А. Братченко
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
shamina94@inbox.ru

С. Аль-Саммаррайе
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
saharzead@rocketmail.com

Е.Н. Тупикова
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
nil-6ssau@mail.ru

Ю.А. Христофорова
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
khristoforovajulia@gmail.com

И.А. Братченко
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
iabratchenko@gmail.com

Аннотация— В работе представлен анализ рамановских характеристик поликомпонентных органических сред с привлечением глубокого обучения, позволяющий улучшить идентификацию поликомпонентных коллоидных сред за счет обобщенной интерпретации совокупности спектральных признаков. Продемонстрирована одномерная сверточная нейронная сеть для распознавания поверхностно-усиленных рамановских спектров поликомпонентных коллоидных сред. Предложен метод идентификации поверхностно-усиленных рамановских спектральных особенностей поликомпонентных коллоидных сред, ассоциированных с особенностями их компонентного состава, включающий в себя анализ распределения важности переменных в модели одномерной сверточной нейронной сети.

Ключевые слова— Рамановский спектр, сверточная нейронная сеть, нелинейная зависимость, глубокое обучение

1. ВВЕДЕНИЕ

Применение поверхностного усиления на основе агломератов сферических частиц серебра позволяет усилить рамановский сигнал поликомпонентной органической коллоидной среды и снизить относительное влияние таких ограничивающих факторов, как фотолюминисцентный и флуоресцентный фон от образца, примесей и вклада посторонних источников спектрального шума, что улучшает информативность анализа компонентных особенностей органических сред. Однако, это не исключает негативного влияния на интерпретацию поверхностно-усиленных рамановских спектров и идентификацию поликомпонентных коллоидных сред таких явлений как мультиколлинеарность и множественное перекрытие спектров различных аналитов, входящих в состав исследуемой среды, а также влияние возможных флуктуаций сигнала рассеяния и степени однородности и стабильности инструментальной базы.

В работе исследован подход, заключающийся в выявлении компонентно-ассоциированных особенностей поверхностно-усиленных рамановских характеристик коллоидных сред в диапазоне $500-1900\text{ см}^{-1}$ в условиях нелинейной зависимости спектральных характеристик среды от особенностей ее состава с помощью одномерной сверточной нейронной сети (CNN) на

основе комбинации сверточной базы с остаточными связями, дополненной рекуррентным слоем.

2. МАТЕРИАЛЫ И МЕТОДЫ

База экспериментальных данных сформирована на основе регистрации характеристик поверхностно-усиленного рамановского рассеяния сыворотки крови пациентов с хронической болезнью почек (группа 58 диализных больных с терминальной стадией ХБП, группа 78 больных с начальными стадиями ХБП) Предложенный метод основан на многофакторном анализе экспериментальных данных для дискриминации образцов по степени тяжести ХБП, что ассоциировано с множественными изменениями компонентного состава сыворотки, и для определения содержания и спектрального вклада отдельных аналитов на примере креатинина и карбамида.

Предобработка поверхностно-усиленных рамановских спектров сыворотки состояла из нескольких последовательных этапов: сглаживание шумов, удаление автофлуоресцентного фона и нормализация. Сглаживание необработанных спектров выполнялось с помощью фильтра Савицкого-Голея с шириной окна фильтра 15, первым порядком полинома, используемого для сглаживания, и производной нулевого порядка (без производной). Во избежание переобучения, анализ устойчивости реализации метода анализа экспериментальных данных и выбор оптимальных параметров были реализованы с помощью k-кратной кросс-валидации ($k = 10$) [1]. Метод на основе глубокого обучения реализован с использованием отдельной одномерной CNN для каждой задачи обработки и распознавания экспериментальных данных. Сравнительный анализ характеристик сетей различных архитектур выполнен на основе генетического алгоритма, позволяющего эмпирически определить оптимальную архитектуру и гиперпараметры. В качестве меры успешности и критериев оценки тестируемых конфигураций CNN были выбраны следующие факторы: устойчивость модели, величина ошибки, точность для задачи классификации и визуализация распределения важности переменных. Для моделей CNN использовался алгоритм вычисления важности переменных на основе перестановок [2].

Важность переменной в построенной модели на основе перестановки определяется как уменьшение характеристик модели при случайной перестановке переменной в наборе параметров.

3. РЕЗУЛЬТАТЫ

Без использования многомерного анализа в качестве отличительных особенностей спектральных характеристик сыворотки крови гемодиализных больных можно выделить увеличение интенсивности пиков 1004 см^{-1} (C-C, фенилаланин), 1240 см^{-1} (амид III), 1395 см^{-1} (CH, C=O) по сравнению с пациентами 1-3а групп ХБП. Следует отметить наличие внутригрупповой дисперсии и перекрытия спектральных признаков изучаемых групп, а также наличие мультиколлинеарности. По этой причине база данных спектров сыворотки была подвергнута многофакторному анализу. Для обнаружения целевых субъектов с помощью решения на основе глубокого обучения специфичность 0,95, чувствительность 0,92 и точность 0,94 достаточны для клинического использования.

На рисунке 1 представлено распределение важности переменных матрицы спектров сыворотки при построении моделей классификации гемодиализных больных с терминальной стадией почечной недостаточности и больных с ХБП 1-3а стадии.

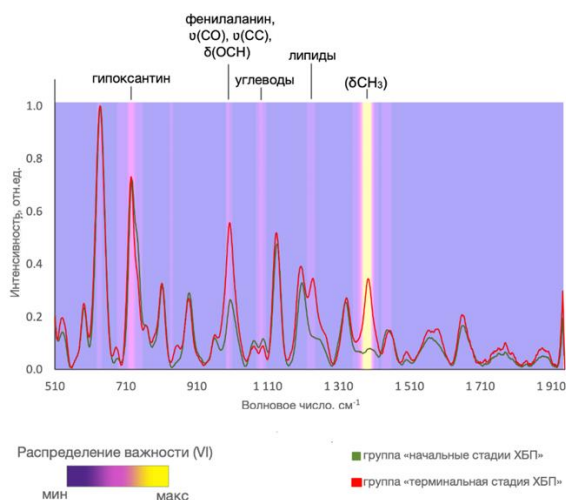


Рис. 1. Распределение важности переменных поверхностно-усиленных рамановских спектров сыворотки в классификации образцов по патолого-ассоциированному признаку

Реализация метода анализа экспериментальных данных на основе CNN выделила в качестве наиболее информативных следующие полосы спектра: полоса $720\text{--}750\text{ см}^{-1}$ с максимумом пика при 724 см^{-1} (соответствует гипоксантину), полоса $990\text{--}1030\text{ см}^{-1}$ с максимумом пика при 1001 см^{-1} (фенилаланин, $\nu(\text{CO})$, $\nu(\text{CC})$, $\delta(\text{OCH})$),

полоса $1089\text{--}1110\text{ см}^{-1}$ с максимумом пика при 1095 см^{-1} (углеводы), полоса $1220\text{--}1255\text{ см}^{-1}$ с максимумом пика при 1238 см^{-1} (липиды), полоса $1380\text{--}1415\text{ см}^{-1}$ с максимумом пика при 1393 см^{-1} (δCH_3) [3-6]. Были построены регрессионные модели корреляции характеристик поверхностно-усиленного рамановского рассеяния сыворотки и уровня креатинина и карбамида, определяемых стандартным биохимическим анализом. Коэффициент детерминации R^2 уровня креатинина составил 0,81 для модели CNN. Коэффициент детерминации уровня карбамида составил 0,82. В целом полученные характеристики методов обработки экспериментальных данных в задачах регрессии креатинина и карбамида сопоставимы с результатами, полученными другими авторами

4. ЗАКЛЮЧЕНИЕ

Применение интеллектуального метода идентификации сред с привлечением глубокого обучения позволяет снизить негативное влияние факторов, ассоциированных с мультиколлинеарностью, флуктуациями сигнала рассеяния и степенью стабильности инструментальной базы, на спектральный анализ компонентных особенностей коллоидных сред.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Российского научного фонда No. 21-75-10097, <https://rscf.ru/project/21-75-10097/>.

ЛИТЕРАТУРА

- [1] Guo, J. Diagnosis of chronic kidney diseases based on surface-enhanced Raman spectroscopy and multivariate analysis / J. Guo, Z. Rong, Y. Li, S. Wang, W. Zhang, R. Xiao // *Laser Physics*. – 2018. – Vol. 28(7). – P. 075603.
- [2] Date, Y. Application of a deep neural network to metabolomics studies and its performance in determining important variables / Y. Date, J. Kikuchi // *Analytical chemistry*. – 2018. – Vol. 90(3). – P. 1805-1810.
- [3] Wang, J. Raman spectroscopic study of cervical precancerous lesions and cervical cancer / J. Wang, CX. Zheng, CL. Ma, XX. Zheng, XY. Lv, GD. Lv, J. Tang, GH. Wu // *Lasers in Medical Science*. – 2021. – Vol. 36(9). – P. 1855-1864.
- [4] Feng, S. Gastric cancer detection based on blood plasma surface-enhanced Raman spectroscopy excited by polarized laser light / S. Feng, R. Chen, J. Lin, J. Pan, Y. Wu, Y. Li, J. Chen, H. Zeng // *Biosensors and Bioelectronics*. – 2011. – Vol. 26(7). – P. 3167-3174.
- [5] De Gelder, J. Reference database of Raman spectra of biological molecules / J. De Gelder, K. De Gussem, P. Vandenabeele, L. Moens // *Journal Raman spectroscopy*. – 2007. – Vol. 38. – P. 1133-1147.
- [6] Cheng, W. T. Micro - Raman spectroscopy used to identify and grade human skin pilomatrixoma / W. T. Cheng, M. T. Liu, H. N. Liu, S. Y. Lin // *Microscopy research and technique*. – 2005. – Vol. 68(2). – P. 75-79.