

Анализ сыворотки гемодиализных пациентов спомощью комбинации SERS и машинного обучения

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

Д.Ю. Коновалова
Самарский государственный
медицинский университет
Самара, Россия
snowflake0605@mail.ru

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

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

П.А. Лебедев
Самарский государственный
медицинский университет
Самара, Россия
palebedev@yahoo.com

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

В.П. Захаров
Самарский национальный
исследовательский университет им. академика С.П. Королева
Самара, Россия
ssauvz@mail.ru

Аннотация—Целью данной работы является многомерный анализ характеристик SERS сыворотки крови у гемодиализных больных, включающий построение классификационных моделей (PLS-DA, CNN) по наличию/отсутствию терминальной стадии хронической болезни почек на диализе и определение наиболее информативных спектральных полос для идентификации пациентов на диализе по распределению важности переменных в построенной модели.

Ключевые слова— SERS, сыворотка, CNN, машинное обучение, PLS.

1. ВВЕДЕНИЕ

В настоящее время одним из наиболее доступных методов первичной оценки состояния организма человека является анализ крови [1]. Быстрый и точный анализ образца крови является основой клинической диагностики. Повышение прогностической значимости исследования крови при выявлении патологических процессов возможно при исследовании комплекса изменений компонентного состава крови. Рамановская спектроскопия является перспективным методом для решения этой задачи [2]. Основными преимуществами рамановской спектроскопии при анализе биологических тканей являются: высокая специфичность анализа, простота пробоподготовки, неразрушающий анализ, возможность анализа небольшого образца (около 1 мкм) [3]. Однако возможности рамановской спектроскопии при исследовании многокомпонентных биологических объектов ограничены низким уровнем полезного сигнала. Улучшение качества рамановского сигнала и повышение информативности анализа возможно при использовании поверхностно-усиленной рамановской спектроскопии (SERS).

Следует отметить, что при анализе многокомпонентных биологических тканей методами

рамановской спектроскопии и SERS полученные спектральные характеристики представляют собой суперпозицию спектрального вклада всех компонентов, входящих в исследуемый образец. Следовательно, анализ и интерпретация спектральных характеристик биологических тканей связаны с проблемой мультиколлинеарности и множественного перекрытия спектров. Выбор метода многомерного анализа, соответствующего особенностям решаемой задачи, является важным фактором получения статистически достоверного результата. Однако для оценки возможностей подходов, основанных на сочетании машинного обучения и рамановских методов в клинических задачах, следует внимательно рассмотреть проблему переобучения и анализа многомерных моделей. Целью работы на данном этапе является анализ *in vitro* спектральных характеристик сыворотки крови у гемодиализных больных с использованием комбинации SERS в ближнем инфракрасном диапазоне и методов машинного обучения. Используемый подход основан на построении классификационной модели выявления диализных больных и выявлении наиболее информативных полос SERS, связанных с различиями между группой диализных больных и группой больных с 1-3а стадиями ХБП.

2. МАТЕРИАЛЫ И МЕТОДЫ

Серебряные структуры на основе высушенного коллоида серебра использованы для усиления рамановского рассеяния в ближнем инфракрасном диапазоне. Коллоид серебра получали восстановлением из водного раствора нитрата серебра цитратом натрия при температуре 95 °С в течение 20 минут.

На данном этапе был проведен анализ сыворотки крови человека *in vitro*. Исследование проведено у 136 человек, из них 78 человек с хронической болезнью почек (ХБП) 1-3а стадии (в зависимости от скорости

клубочковой фильтрации) и 58 гемодиализных больных с терминальной стадией ХБП.

Каждому образцу сыворотки соответствует априорная информация о принадлежности к той или иной группе и биохимические характеристики. Таким образом, анализ данных реализуется посредством обучения с учителем. Анализ устойчивости построенных моделей и выбор оптимальных параметров во избежание переобучения осуществлялись с помощью многократной k-кратной кросс-валидации ($k = 10$). При построении моделей важность предикторов в решении задачи классификации оценивалась на основе переменной важности в проекционном распределении (VIP). Базовое решение неглубокого обучения реализовано на основе дискриминантного анализа с проекцией на латентные структуры (PLS-DA) для задачи классификации и на основе проекции на латентные структуры (PLS) для задач регрессии. Решение на основе глубокого обучения реализовано с использованием отдельной одномерной сверточной нейронной сети (CNN) для каждой задачи. В качестве сигнала обратной связи для обучения тензоров весов использована функция потерь *categorical_crossentropy* для решения задачи различения гемодиализных больных с терминальной стадией ХБП и больных с 1-3а стадиями ХБП. CNN обучается с использованием алгоритма adamax. Количество эпох определялось по локальному минимуму функции потерь при кросс-валидации.

3. РЕЗУЛЬТАТЫ

База данных спектров сыворотки SERS была подвергнута многофакторному анализу. В таблице 1 представлены характеристики построенных моделей классификации: для обучающего набора данных (обозначается «train»), для проверочного набора данных (обозначается «test»).

Таблица 1. Специфичность, чувствительность, точность различения группы терминальной стадии почечной недостаточности и группы 1-3а стадий ХБП

Модель	Чувствительность		Специфичность		Точность	
	train	test	train	test	train	test
PLS - DA	0,71	0,69	0,95	0,96	0,84	0,81
CNN	1	0,92	0,99	0,95	0,99	0,94

Результаты, представленные в таблице 1 демонстрируют, что при классификации испытуемых по наличию/отсутствию терминальной стадии почечной недостаточности при анализе спектральных характеристик SERS сыворотки базовое решение без привлечения глубокого обучения уступает решению на основе CNN. Следует отметить, что как базовая модель решения, так и модель CNN стабильны. Специфичность

0,95, чувствительность 0,92 и точность 0,94 для обнаружения целевых субъектов на основе решения для глубокого обучения достаточны для клинического использования. Анализ парной корреляции Пирсона между распределением VIP при построении модели PLS-DA и распределением VIP при построении модели CNN демонстрирует коэффициент корреляции 0,61 и отсутствие значимой корреляции. Модель на основе базового решения и модель на основе CNN определили как наиболее информативные следующие полосы спектра: 720 - 750 cm^{-1} (соответствует растяжению CС, пролину), 990 - 1030 cm^{-1} (соответствует фенилаланину, $\nu(\text{CO})$, $\nu(\text{CC})$, $\delta(\text{OCH})$), 1220 - 1255 cm^{-1} (соответствует липидам), 1380 - 1415 cm^{-1} (соответствует δCH_3) [4].

4. ЗАКЛЮЧЕНИЕ

В работе выполнено исследование SERS характеристик сыворотки крови у диализных больных с терминальной стадией ХБП. На основе многомерного анализа были выделены информативные спектральные полосы, связанные с терминальной стадией ХБП во время диализа. На следующем этапе работы описанный подход будет применен на расширенной выборке для выявления спектрального вклада биохимических компонентов сыворотки (альбумина, ферритина, холестерина, глюкозы и др.), что позволит улучшить описание патолого-ассоциированных спектральных характеристик сыворотки крови у диализных больных с терминальной стадией ХБП. В целом, описанный подход может стать основой для мониторинга состояния здоровья диализных больных и может быть применен для изучения других патологических состояний организма человека.

БЛАГОДАРНОСТИ

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

ЛИТЕРАТУРА

- [1] Watson, J. What do tests do for doctors? A qualitative study of blood testing in UK primary care / J. Watson, I. de Salis, J. Banks, C. Salisbury // Family Practice. – 2017. – Vol. 34(6). – P. 735-739.
- [2] Atkins, C.G. Raman Spectroscopy of Blood and Blood Components / C.G. Atkins, K. Buckley, M.W. Blades, R.F.B. Turner // Applied Spectroscopy. – 2017. – Vol. 71(5). – P. 767-793.
- [3] Bergholt, M.S. Raman Spectroscopy: Guiding Light for the Extracellular Matrix / M.S. Bergholt, A. Serio, M.B. Albro // Bioengineering and Biotechnology. – 2019. – Vol. 7. – P. 303. DOI: 10.3389/fbioe.2019.00303.
- [4] Cao, X. Label-Free Detection of Human Serum Using Surface-Enhanced Raman Spectroscopy Based on Highly Branched Gold Nanoparticle Substrates for Discrimination of Non-Small Cell Lung Cancer / X. Cao, Z. Wang, L. Bi, J. Zheng // Journal of Chemistry. – 2018. – Vol. 2018. – P. 1-13.