

solution / Analyst, 2017, т. 142, № 21, С. 4067-4074.

Новикова Виктория Васильевна, студент гр. 6302-010302D, victorianovikovaa@mail.ru.

Матвеева Ирина Александровна, к.т.н., доцент каф. лазерных и биотехнических систем, matveeva.ia@ssau.ru.

УДК 004.891.3:543.424.2:616-71

## **АВТОМАТИЗАЦИЯ ДИАГНОСТИКИ ХРОНИЧЕСКОЙ СЕРДЕЧНОЙ НЕДОСТАТОЧНОСТИ С ИСПОЛЬЗОВАНИЕМ РАМАНОВСКОЙ СПЕКТРОСКОПИИ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

И.А. Пименова, И.А. Матвеева

«Самарский национальный исследовательский университет имени  
академика С.П. Королёва», г. Самара

**Ключевые слова:** чередующиеся наименьшие квадраты, метод разрешения многомерных кривых, рамановская спектроскопия, сыворотка крови, хроническая сердечная недостаточность, машинное обучение

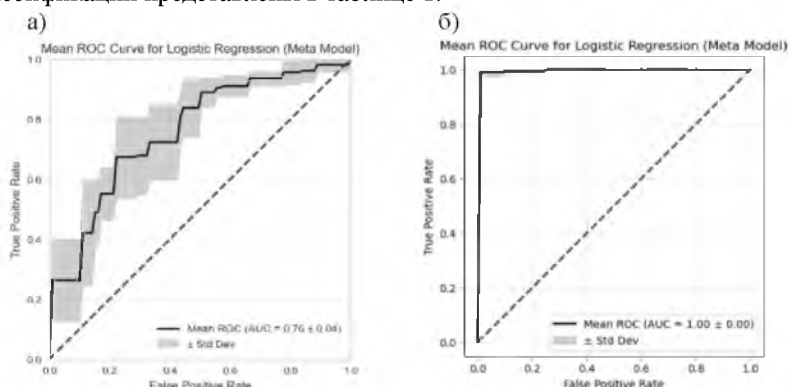
Хроническая сердечная недостаточность (ХСН) остаётся одним из самых распространенных сердечно-сосудистых заболеваний во всём мире. Своевременная диагностика ХСН имеет ключевое значение для назначения эффективной терапии и профилактики осложнений. Однако существующие методы требуют значительных затрат ресурсов и могут не выявлять заболевание на ранних стадиях. В связи с этим актуальным направлением является разработка новых, неинвазивных и высокочувствительных методов диагностики, в частности, на основе рамановской спектроскопии [1].

В исследовании использовались *in vitro* спектры сыворотки крови, полученные с применением технологии поверхностно-усиленной рамановской спектроскопии (SERS), подробно описанной в [2]. Всего проанализировано 229 образцов, из которых 180 принадлежали пациентам с подтверждённой ХСН, а 49 – контрольной группе. Для повышения информативности спектров и снижения размерности данных был применён метод разрешения многомерных кривых на основе алгоритма чередующихся наименьших квадратов (MCR-ALS). Его применение позволило эффективно выделить биохимически значимые компоненты спектра, связанные с молекулами белков, липидов и метаболитов.

Далее профили концентраций выделенных компонентов использовались в качестве признаков для построения моделей машинного обучения: логистической регрессии, метода опорных векторов (SVM), случайного леса и градиентного бустинга. Оценка качества моделей проводилась с использованием, 5-ти фолдовой кросс-валидации, ROC-кривых и метрики точности.

Наиболее высокие показатели классификации были достигнуты с помощью ансамблевого метода, объединяющего несколько моделей методом стекинга. Помимо профилей концентраций, в качестве дополнительных признаков для классификации использовались клинические данные пациентов: систолическое артериальное давление, диастолическое артериальное давление, конечный систолический размер, общий билирубин, гематокрит.

На рисунке 1 представлены ROC-кривые для моделей, использующих ансамблевый метод стекинга. В данном случае также применен метод MCR-ALS для сокращения числа признаков. Результаты оценки метрик классификации представлены в таблице 1.



а) в качестве признаков взяты только концентрации компонент; б) в качестве признаков взяты концентрации компонент и клинические данные

Рисунок 1 – ROC-кривые для случая классификации здоровых людей и пациентов с хронической сердечной недостаточностью

Таблица 1 – Метрики, полученные в результате классификации

| Модель                         | Accuracy (MCR)    | AUC (MCR)       |
|--------------------------------|-------------------|-----------------|
| Логистическая регрессия        | $0.83 \pm 0.05$   | $0.65 \pm 0.07$ |
| Метод опорных векторов         | $0.83 \pm 0.05$   | $0.64 \pm 0.13$ |
| Случайный лес                  | $0.83 \pm 0.06$   | $0.70 \pm 0.13$ |
| Градиентный бустинг            | $0.81 \pm 0.09$   | $0.64 \pm 0.19$ |
| Стекинг                        | $0.83 \pm 0.03$   | $0.72 \pm 0.04$ |
| Стекинг с клиническими данными | $0.85\% \pm 0.01$ | $1.00 \pm 0.00$ |

Применение современных методов машинного обучения открывает возможности для использования модели в скрининговых и амбулаторных условиях. Предложенный в работе метод интеллектуального анализа

спектров крови на основе SERS и MCR-ALS в сочетании с машинным обучением представляет собой перспективный инструмент для ранней диагностики ХСН.

#### Список использованных источников

1. Atkins C. G., Buckley K., Blades M. W., Turner R. F., Raman spectroscopy of blood and blood components / *Applied spectroscopy*, 2017, т. 71, № 5, С. 767-793.
2. Al-Sammarraie S. Z., Bratchenko L. A., Typikova E. N., Zakharov V. P., Bratchenko I. A., Lebedev P. A., Silver nanoparticles-based substrate for blood serum analysis under 785 nm laser excitation / *Journal of Biomedical Photonics & Engineering*, 2022, т. 8, № 1, С. 010301.

Пименова Ирина Александровна, студент гр. 6231-120404D, pimenova.0312@list.ru.

Матвеева Ирина Александровна, к.т.н., доцент каф. лазерных и биотехнических систем, matveeva.ia@ssau.ru.

УДК 004.891.3; 543.424.2; 616-71

### **ВЫБОР ОПТИМАЛЬНОГО НАБОРА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ИДЕНТИФИКАЦИИ РАМАНОВСКИХ СПЕКТРОВ КОЖИ**

К.Е. Томникова, И.А. Матвеева

«Самарский национальный исследовательский университет имени академика С.П. Королёва», г. Самара

**Ключевые слова:** рамановская спектроскопия, новообразования кожи, классификация, набор данных, признаки

Анализ рамановских спектров кожи является сложной задачей, поскольку спектральные данные несут в себе большое количество информации о химическом составе кожи, а также побочные сигналы [1]. Для идентификации рамановских спектров часто используют различные методы искусственного интеллекта [2, 3]. Для эффективной работы таких методов необходимо грамотно подготовить данные для обучения. Целью работы является выбор оптимального набора данных, благодаря которому будет достигнута наилучшая эффективность классификации рамановских спектров с помощью модели многослойного перцептрона.

Первый набор данных включает в себя только спектры рамановского рассеяния (спектральные отсчёты). Всего было зарегистрировано 615 спектров кожи с различными заболеваниями. Исследование проводилось в Самарском областном клиническом онкологическом диспансере.

Второй набор данных состоит из относительных концентраций тридцати компонентов, полученных благодаря анализу рамановских