

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., Tytikova 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

спектров методом разрешения многомерных кривых (MCR). Такой метод сокращения объема данных позволяет физически интерпретировать получаемые в результате анализа компоненты и оценить вклад этих компонентов в исходный рамановский спектр.

Третий набор данных – анамнез пациентов. Анамнез играет важную роль в диагностике заболеваний кожи. Информация о ранее перенесенных болезнях, аллергиях, наследственности и образе жизни может помочь врачу получить полную картину состояния пациента [4]. В данной работе представлены 15 различных факторов риска.

Четвертый набор данных представляет собой комбинацию второго и третьего наборов данных и включает в себя относительные концентрации тридцати компонентов и все представленные факторы риска.

В результате работы были разработаны модели классификации для трех случаев: доброкачественные новообразования (350 образцов) против злокачественных новообразований (266 образцов) (случай 1), злокачественная меланома (70 образцов) против пигментного невуса (170 образцов) (случай 2), злокачественная меланома (70 образцов) против пигментного невуса и себорейного кератоза (283 образца) (случай 3).

Для каждого случая были разработаны модели на основе алгоритма многослойного перцептрона. В качестве позначков классификации использовались следующие комбинации: спектральные отсчеты рамановских спектров («отсчеты»), относительные концентрации, полученные в результате MCR анализа («MCR»), анамнез пациентов («анамнез»), относительные концентрации, полученные в результате MCR анализа, и анамнез пациентов («MCR+анамнез»).

Метрикой оценки качества работы моделей выбрана ROC-кривая и площадь под ней (ROC AUC). Указанные метрики для каждой модели классификации представлены в таблице 1.

Таблица 1 – ROC AUC разработанных моделей классификации

Случай классификации	Набор данных	Многослойный перцептрон
1	Отсчеты	0,65±0,11
	MCR	0,68±0,12
	Анамнез	0,67±0,10
	MCR+анамнез	0,72±0,08
2	Отсчеты	0,65±0,09
	MCR	0,72±0,12
	Анамнез	0,60±0,11
	MCR+анамнез	0,73±0,10
3	Отсчеты	0,62±0,08
	MCR	0,76±0,09
	Анамнез	0,66±0,09
	MCR+анамнез	0,84±0,10

Как видно из таблицы 1, все самые эффективные модели были разработаны с использованием набора данных, где в качестве признаков классификации используются относительные концентрации, полученные в результате MCR анализа, и анамнез пациентов. При использовании набора данных «MCR+анамнез» не только увеличивается эффективность моделей классификации, но и повышается скорость вычислений (по сравнению с использованием набора данных «Отсчеты»), а также появляется возможность физической интерпретации результатов.

Список использованных источников

1. Bratchenko I. A., Bratchenko L. A., Moryatov A. A., Khristoforova Y. A., Artemyev D. N., Myakinin O. O., Orlov A. E., Kozlov S. V., Zakharov V. P., In vivo diagnosis of skin cancer with a portable Raman spectroscopic device / *Experimental Dermatology*, 2021, т. 30, № 5, С. 652-663.
2. Santos I. P., van Doorn R., Caspers P. J., Bakker Schut T. C., Barroso E. M., Nijsten T. E. C., Noordhoek Hegt V., Koljenović S., Puppels G. J., Improving clinical diagnosis of early-stage cutaneous melanoma based on Raman spectroscopy / *British journal of cancer*, 2018, т. 119, № 11, С. 1339-1346.
3. Araújo D. C., Veloso A. A., de Oliveira Filho R. S., Giraud M. N., Raniero L. J., Ferreira L. M., Bitar R. A., Finding reduced Raman spectroscopy fingerprint of skin samples for melanoma diagnosis through machine learning / *Artificial Intelligence in Medicine*, 2021, т. 120, С. 102161.
4. Русак Ю. Э., Ефанова Е. Н., Васильева Е. А., Схема написания истории болезни дерматологического больного, 2018.

Томникова Ксения Евгеньевна, студент гр. 6231-120404D, ksetomnikova@yandex.ru.
Матвеева Ирина Александровна, к.т.н., доцент каф. лазерных и биотехнических систем, matveeva.ia@ssau.ru.

УДК 543.424.2; 004.93

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА СПЕКТРОВ РАМАНОВСКОГО РАССЕЯНИЯ КОЖИ ПАЦИЕНТОВ С ХРОНИЧЕСКОЙ СЕРДЕЧНОЙ НЕДОСТАТОЧНОСТЬЮ

Д.С. Афанасьев, Ю.А. Христофорова
«Самарский национальный исследовательский университет имени академика С.П. Королёва», г. Самара

Ключевые слова: Метод k ближайших соседей, Random Forest классификатор, спектры рамановского рассеяния, хроническая сердечная недостаточность

Хроническая сердечная недостаточность (ХСН) остается одной из ключевых проблем современной кардиологии из-за высокой