

## ОЦЕНКА ЭФФЕКТИВНОСТИ КЛАССИФИКАЦИИ СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЯНИЯ КОЖИ ОТ СООТНОШЕНИЯ СИГНАЛ/ШУМ

А.А. Мельникова<sup>1</sup>, Ю.А. Христофорова<sup>1</sup>

<sup>1</sup>Самарский университет, г. Самара  
e-mail: angelina.melnikova.2021@internet.ru

### Аннотация

В работе исследовано влияние соотношения сигнал/шум на качество классификации спектров комбинационного рассеяния кожи *ex vivo*. Построены бинарные классификационные модели для 377 спектров. Изменение соотношения сигнал/шум моделировалось за счёт добавления шума и сглаживания спектров. Показано, что снижение соотношения сигнал/шум ухудшает качество классификации, а увеличение – улучшает. Статистически значимое улучшение площади под ROC-кривой получено при сравнении зашумлённых и сглаженных спектров.

**Ключевые слова:** *ex vivo* спектры комбинационного рассеяния кожи, соотношение сигнал/шум, PLS-DA, бинарная классификация, ALS, рак кожи.

### Введение

В последние десятилетия во всем мире отмечается рост заболеваемости злокачественными новообразованиями кожи (ЗНК). По оценкам Всемирной организации здравоохранения (ВОЗ), ежегодно регистрируется от 2 до 3 миллионов случаев немеланомного рака кожи (НРК) и около 132 000 случаев меланомы [1]. Спектроскопия комбинационного рассеяния (КР) показала большой потенциал во многих областях биомедицины [2], поскольку позволяет получать информацию о молекулярной структуре образца и обнаруживать патологические изменения. Однако эффективность анализа зависит от качества спектральных данных, в частности, от соотношения сигнал/шум (SNR). Низкая интенсивность сигналов биомолекул приводит к тому, что шум может маскировать информативные пики [3], а стандартные способы повышения SNR не всегда применимы. В связи с этим изучение влияния SNR на точность классификации необходимо для повышения надёжности диагностики.

Цель данной работы – изучить влияние SNR на качество бинарной классификации *ex vivo* спектров кожи при искусственно варьируемом уровне шума.

### 1. Материалы и методы

Для реализации поставленной задачи были использованы *ex vivo* спектры КР кожи пациентов с гистологически подтвержденными доброкачественными и злокачественными новообразованиями. Для каждого пациента спектры регистрировались портативным прибором при длине волны 785 нм как в области поражения, так и на непоражённых участках кожи. Время экспозиции составляло 3 секунды с пятикратным накоплением при температуре детектора –15 °С, что

обеспечивало приемлемый уровень SNR. Измерения проводились в области волновых чисел 100–2300  $\text{см}^{-1}$ , для последующего анализа выбирался диапазон 400–2100  $\text{см}^{-1}$ . Общее количество анализируемых спектров составило 377: 183 – контрольная группа и 194 – патологическая ткань. Спектры в указанных областях были подвергнуты процедуре удаления фона флуоресценции с использованием метода асимметричных наименьших квадратов (ALS) с тремя ключевыми параметрами:  $\lambda$ ,  $\rho$  и  $N$  [4].  $N$  было принято равным 10, что является стандартным значением, тогда как  $\lambda = 10^4$  и  $\rho = 0,001$  подобраны экспериментально. Полученные после данной обработки спектры рассматривались в качестве исходных данных для дальнейшего анализа.

Для изучения влияния SNR на качество классификации рассматривались три случая:

- 1) исходные спектры;
- 2) спектры, искусственно зашумленные гауссовским шумом с нулевым средним и стандартным отклонением (СКО) 7 % от максимальной интенсивности исходного спектра для снижения SNR;
- 3) спектры, сглаженные фильтром Савицкого-Голея с шириной окна равной 15 и полиномом 0-го порядка для повышения SNR.

На рисунке 1(а, б, в) показаны спектры для трёх рассматриваемых случаев: исходные спектры, зашумленные спектры и сглаженные спектры.

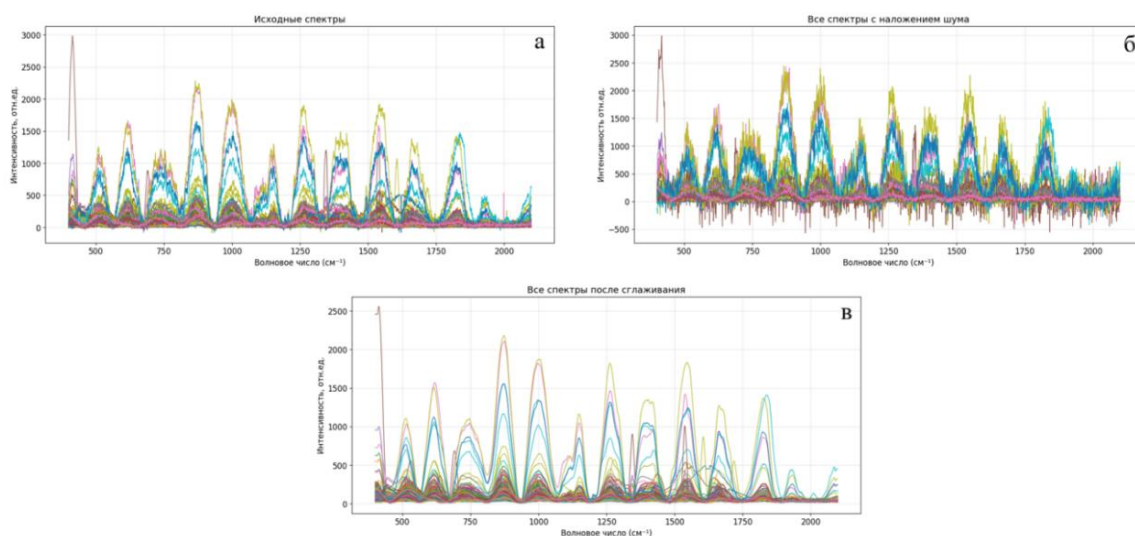


Рисунок 1 – *Ex vivo* спектры кожи для трех случаев: исходные спектры (а), зашумленные спектры (б), сглаженные спектры (в)

Для каждого случая SNR рассчитывалось как отношение максимальной интенсивности сигнала в области 1610–1700  $\text{см}^{-1}$  к СКО шума в области 1885–1985  $\text{см}^{-1}$  по формуле (1):

$$\text{SNR} = \frac{I_{\max}}{\sigma_{\text{noise}}} . \quad (1)$$

Для построения классификаторов применялся метод частичных наименьших квадратов с дискриминантным анализом (PLS-DA) с кросс-валидацией методом K-fold по 10 фолдам между классами «норма» и «патология». Непосредственно перед

классификацией спектры нормировались по СКО. Эффективность классификации оценивалась с использованием метрик ассигасу, площади под кривой ROC (ROC AUC).

## 2. Результаты

На рисунке 2 представлены диаграммы разброса, иллюстрирующие распределение значений SNR *ex vivo* спектров кожи для трёх рассматриваемых случаев.

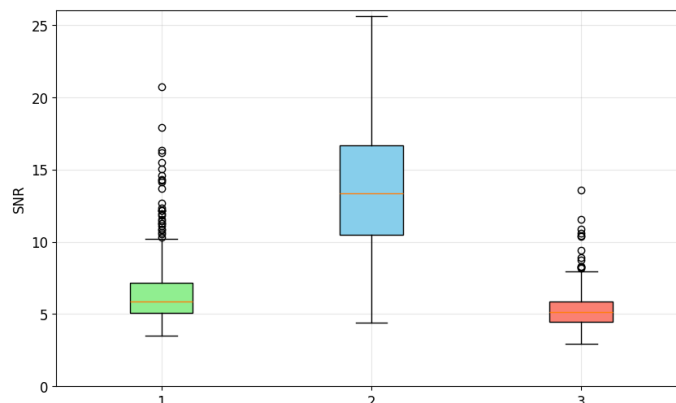


Рисунок 2 – Распределение значений SNR *ex vivo* спектров кожи для трёх случаев:

1 – исходные спектры, 2 – сглаженные спектры, 3 – зашумленные спектры

На рисунке 3 (а, б, в) представлены ROC-кривые для трёх рассматриваемых случаев, отражающие качество бинарной классификации при различных уровнях SNR.

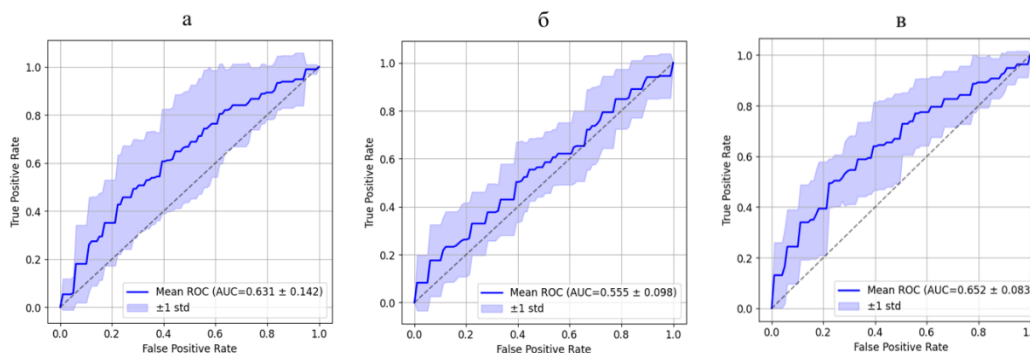


Рисунок 3 – ROC-кривые для трёх случаев: исходные спектры (а), зашумленные спектры (б), сглаженные спектры (в)

Из рисунка 2 видно, что медианное значение SNR для исходных спектров составляет около 6, незначительно превышая значение для зашумлённых спектров (около 5). Это сходство, вероятно, обусловлено невысокой амплитудой добавленного шума на фоне интенсивных пиков. При сглаживании медианное SNR увеличивается до 13 за счёт подавления высокочастотного шума при сохранении формы пиков.

На рисунке 3 визуально наблюдается, что для зашумлённых спектров ROC-кривая смещена вниз по сравнению с исходными данными, что указывает на снижение качества классификации. Исходные спектры демонстрируют несколько лучшее разделение классов, тогда как сглаженные спектры обеспечивают наиболее выраженное отделение классов. Количественные показатели подтверждают эти наблюдения: для исходных спектров ассигасу составляет 0,603, ROC AUC – 0,631, добавление шума снижает их до

0,538 и 0,555, а сглаживание повышает ROC AUC до 0,652 при сопоставимом значении ассигуры 0,601. Статистически значимое различие ROC AUC наблюдается только между зашумлёнными и сглаженными спектрами ( $p = 0,017$ ), что указывает на улучшение классификации при сглаживании, тогда как для исходных и зашумлённых спектров различия незначимы.

### **Заключение**

По результатам анализа *ex vivo* спектров кожи добавление шума не привело к статистически значимому ухудшению классификации относительно исходных данных, тогда как сглаживание спектров обеспечило статистически значимое улучшение качества разделения нормальной и патологической ткани по сравнению с зашумлёнными спектрами ( $p = 0,017$ ).

### **Список использованных источников**

1. The needs and experiences of people with a diagnosis of skin cancer / J. Barker, A. Kumar, W. Stanton [et al.] // JBI Libr Syst Rev. – 2009. – V. 7, № 16 (Suppl). – P. 1–9.
2. Recovery of Raman spectra with low signal-to-noise ratio using Wiener estimation / S. Chen, X. Lin, C. Yuen [et al.] // Optics Express. – 2014. – V. 22, № 10. – P. 12102–12114.
3. Surface-enhanced Raman scattering spectroscopy for potential noninvasive nasopharyngeal cancer detection / J. Lin, R. Chen, S. Feng [et al.] // Journal of Raman Spectroscopy. – 2012. – V. 43, № 4. – P. 497–502.
4. Eilers, P. H. C. Baseline Correction with Asymmetric Least Squares Smoothing / P. H. C. Eilers, H. F. M. Boelens // Leiden, 2005. – 12 p.

## **ESTIMATION OF THE IMPACT OF SIGNAL-TO-NOISE RATIO ON RAMAN SPECTRA CLASSIFICATION PERFORMANCE**

**A.A. Melnikova<sup>1</sup>, Y.A. Khristoforova<sup>1</sup>**

<sup>1</sup>*Samara University, Samara*  
*e-mail: angelina.melnikova.2021@internet.ru*

### **Abstract**

The study investigates the influence of the signal-to-noise ratio on the classification quality of *ex vivo* skin Raman spectra. Binary classification models were constructed for 377 spectra. Changes in the signal-to-noise ratio were simulated by adding noise and smoothing the spectra. It was shown that a decrease in the signal-to-noise ratio degrades classification performance, while an increase improves it. A statistically significant improvement in the area under the ROC curve was observed when comparing noisy and smoothed spectra.

**Keywords:** *ex vivo* skin Raman spectra, signal-to-noise ratio, PLS-DA, binary classification, ALS, skin cancer.