

МНОГОКЛАССОВОЕ РАЗДЕЛЕНИЕ РАМАНОВСКИХ СПЕКТРОВ КОЖИ ПРИ РАЗЛИЧНЫХ ЗАБОЛЕВАНИЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДА ОПОРНЫХ ВЕКТОРОВ

А.М. Сергеева¹, Ю.А. Христофорова¹

¹Самарский университет, г. Самара
e-mail: amsergeeva@yandex.ru

Аннотация

В работе рассматривается метод опорных векторов как многоклассовый классификатор для анализа спектров рамановского рассеяния кожи пациентов с хронической сердечной недостаточностью, хронической болезнью почек, неврологическими заболеваниями и здоровых лиц. Для балансировки классов случайным образом отбирались пациенты из групп хронической сердечной недостаточности и неврологических заболеваний, другие классы применялись полностью. Среднее значение точности классификации по 5 итерациям составило $0,85 \pm 0,02$ (среднее $\pm$ среднеквадратичное отклонение).

Ключевые слова: метод опорных векторов, спектры рамановского рассеяния, хроническая сердечная недостаточность, хроническая болезнь почек, неврологические заболевания.

Введение

Неинфекционные заболевания, такие как хроническая болезнь почек (ХБП), хроническая сердечная недостаточность (ХСН) и неврологические заболевания, несмотря на современные достижения медицины, остаются лидерами по уровням заболеваемости и летальности [1]. Ранняя диагностика этих состояний критически важна, однако существующие подходы – от биохимического скрининга до инструментальных и неврологических тестов – не всегда обеспечивают выявление патологии на ранних этапах, зачастую инвазивны и не гарантируют высокой точности дифференциации [2].

Рамановская спектроскопия кожи является перспективным неинвазивным методом, позволяющим получать молекулярно-биохимическую информацию о тканях в реальном времени. Благодаря высокой чувствительности к малым изменениям в структуре тканей, рамановская спектроскопия может служить эффективным инструментом для раннего выявления патологических процессов и мониторинга состояния пациентов. Интеграция спектроскопических данных с методами машинного обучения позволит повысить точность и скорость диагностики, снизить инвазивность процедур и расширить возможности скрининга заболеваний на ранней стадии [3].

Цель данной работы – разработка мультиклассового классификатора на основе метода опорных векторов (SVM), способного различать пациентов с ХСН, ХБП, неврологическими заболеваниями и здоровых.

1. Материалы и методы

В исследовании были использованы спектры рамановского рассеяния кожи, снятые с предплечья пациентов с различными диагнозами: хронической сердечной недостаточностью (ХСН), хронической болезнью почек (ХБП), неврологическими заболеваниями (НЗ) и у здоровых участников. Всего в выборку вошли 271 пациент с ХСН, 90 пациентов с ХБП, 340 пациентов с НЗ и 80 здоровых участников.

Спектры были зарегистрированы в диапазоне $400\text{--}1800\text{ см}^{-1}$ и прошли предобработку, включающую сглаживание с использованием фильтра Савицкого-Голея и нормализацию на стандартное отклонение, что позволило снизить шум и стандартизировать интенсивности для всех спектров.

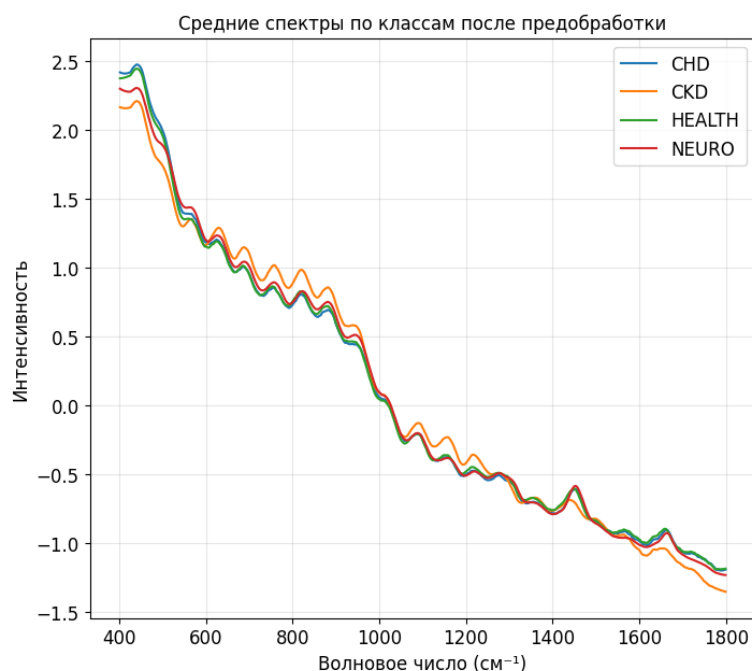


Рисунок 1 – График средних спектров по классам после предобработки (CHD – ХСН, CKD – ХБП, HEALTH – здоровые, NEURO – с неврологическими заболеваниями)

Для анализа использовался метод опорных векторов (SVM) с радиальным базисным ядром (RBF), что обеспечивало выявление нелинейных зависимостей в спектральных данных [4]. В конвейер обработки данных (pipeline) были включены стандартизация признаков и отбор наиболее информативных переменных с помощью F-критерия (SelectKBest, $k=50$), что повышало устойчивость модели и предотвращало переобучение.

Для оценки обобщающей способности модели применялась пятифолдовая групповая кросс-валидация (GroupKFold), учитывающая принадлежность спектров к конкретным пациентам, чтобы избежать утечек данных между обучающей и тестовой выборками. Для балансировки классов случайным образом отбирались 90 пациентов из групп ХСН и неврологических заболеваний, тогда как для ХБП и здоровых участников использовались все доступные данные. Для повышения надежности результаты эксперимента усреднялись по пяти итерациям с различными значениями генератора случайных чисел. Качество классификации оценивалось с использованием стандартных

мультиклассовых метрик: точности (accuracy), средней точности по классам (precision), средней полноты (recall) и F1-меры (F1-score) с макроусреднением.

2. Результаты

Модель SVM продемонстрировала высокую способность классифицировать спектры рамановского рассеяния по четырём группам: ХСН, ХБП, здоровые и неврологические заболевания.

Результаты пяти итераций представлены в таблице 1, где значения указаны как среднее $\pm$ стандартное отклонение.

Таблица 1 – Результаты классификации SVM на пяти итерациях

Итерация	Точность общая (не требует macro)	Средняя точность по классам	Средняя полнота по классам	Средняя F1-мера по классам
1	0,85 $\pm$ 0,05	0,86 $\pm$ 0,04	0,85 $\pm$ 0,05	0,85 $\pm$ 0,05
2	0,82 $\pm$ 0,03	0,83 $\pm$ 0,03	0,82 $\pm$ 0,03	0,82 $\pm$ 0,03
3	0,84 $\pm$ 0,05	0,85 $\pm$ 0,05	0,84 $\pm$ 0,05	0,84 $\pm$ 0,05
4	0,86 $\pm$ 0,05	0,87 $\pm$ 0,05	0,86 $\pm$ 0,04	0,86 $\pm$ 0,05
5	0,87 $\pm$ 0,05	0,88 $\pm$ 0,05	0,87 $\pm$ 0,05	0,87 $\pm$ 0,05

Анализ результатов показал, что наилучшая производительность модели была достигнута в пятой итерации (seed=46), где общая точность составила 0,87 $\pm$ 0,05, средняя точность по классам – 0,88 $\pm$ 0,05, средняя полнота – 0,87 $\pm$ 0,05 и средняя F1-мера – 0,87 $\pm$ 0,05. Минимальные значения метрик наблюдались во второй итерации (seed=43): общая точность 0,82 $\pm$ 0,03, средняя точность по классам 0,83 $\pm$ 0,03, средняя полнота 0,82 $\pm$ 0,03 и средняя F1-мера 0,82 $\pm$ 0,03. В целом, несмотря на вариацию случайного отбора пациентов для балансировки классов, модель демонстрирует стабильные результаты с общей точностью на уровне примерно 82–87% и макро-усреднёнными метриками выше 0,82. Это подтверждает надежность метода опорных векторов для дифференциации пациентов с ХСН, ХБП, НЗ и здоровых участников по спектрам рамановского рассеяния кожи.

Заключение

В настоящем исследовании была разработана мультиклассовая модель на основе метода опорных векторов (SVM) для анализа рамановских спектров кожи с целью дифференциации пациентов с хронической сердечной недостаточностью (ХСН), хронической болезнью почек (ХБП), неврологическими заболеваниями и здоровых участников. Средняя общая точность по итерациям составила 0,85 $\pm$ 0,05, средняя точность по классам – 0,86 $\pm$ 0,04, средняя полнота – 0,85 $\pm$ 0,05, а средняя F1-мера – 0,85 $\pm$ 0,05. Наилучшие показатели достигнуты в пятой итерации (accuracy = 0,87 $\pm$ 0,05; precision_macro = 0,88 $\pm$ 0,05; recall_macro = 0,87 $\pm$ 0,05; F1_macro = 0,87 $\pm$ 0,05), что демонстрирует устойчивость модели при случайном отборе пациентов из группы ХСН и неврологических заболеваний.

Благодарности

Работа выполнена при поддержке гранта РФФ № 25-75-00146 (<https://rscf.ru/project/25-75-00146/>).

Список использованных источников

1. Noncommunicable diseases. – URL: <https://www.who.int/ru/news-room/factsheets/detail/noncommunicable-diseases> (дата обращения: 15.10.2025).
2. Dadang Muhammad Hasyim. The Importance of Early Detection in Disease Management / Dadang Muhammad Hasyim // ResearchGate. – 2024. – URL: https://www.researchgate.net/profile/Dadang-Muhammad-Hasyim/publication/377529921_The_Importance_of_Early_Detection_in_Disease_Management/links/65aab76abf5b00662e1e7175/The-Importance-of-Early-Detection-in-Disease-Management.pdf (дата обращения: 03.12.2025).
3. Lednev, I. Raman spectroscopy and machine learning for medical diagnostics and forensic purposes / I. Lednev // Advanced laser technologies ALT'21: Book of abstracts the 28th International Conference. – Moscow, 2021. – P. 8. – URL: <https://cyberleninka.ru/article/n/raman-spectroscopy-and-machine-learning-for-medical-diagnostics-and-forensic-purposes> (дата обращения: 05.12.2025).
4. Zhang, L. Raman spectroscopy and machine learning for the classification of breast cancers / L. Zhang // Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. – 2021. – V. 264. – P. 120300.

MULTICLASS SEPARATION OF SKIN RAMAN SPECTRA FOR DIFFERENT DISEASES USING SUPPORT VECTOR MACHINES

A.M. Sergeeva¹, Y.A. Khristoforova¹

¹*Samara National Research University, Samara
e-mail: amsergeeva@yandex.ru*

Abstract

This study investigates the use of support vector machines as a multiclass classifier for analyzing skin Raman spectra of patients with chronic heart failure, chronic kidney disease, neurological disorders, and healthy individuals. To balance the classes, patients were randomly selected from the chronic heart failure and neurological groups, while the remaining classes were used in full. The mean classification accuracy over five iterations was 0.85 ± 0.02 (mean $\pm$ standard deviation).

Keywords: support vector machine, Raman spectra, chronic heart failure, chronic kidney disease, neurological disorders.