

между моделями статистически не значимо.

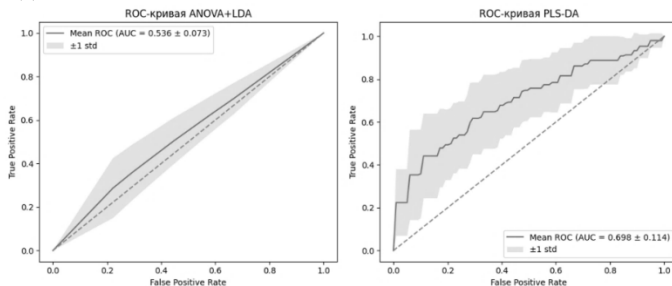


Рисунок 2 – ROC-кривые моделей на основе ANOVA с LDA и PLS-DA для спектров с без вычета базовой линии

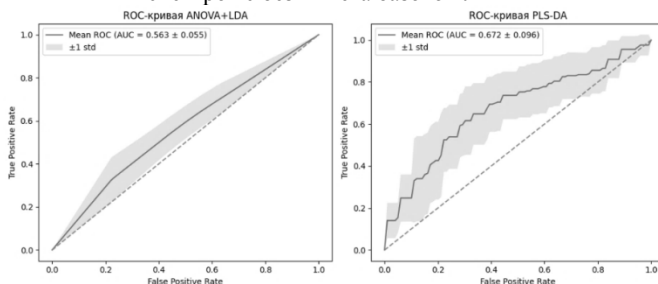


Рисунок 3 – ROC-кривые моделей на основе ANOVA с LDA и PLS-DA для спектров с вычетом базовой линии

Список использованных источников

1. Чупров И.Н., Сыдилов А.А., Заславский Д.В., Насыров Р.А. Дерматогонкопатология / Иллюстрированное руководство для врачей, 2021.
2. English D.R. Factors influencing the number needed to excise: excision rates of pigmented lesions by general practitioners, 2004.

Сальникова Екатерина Олеговна, студ. гр. 6402-120304D, fume890@outlook.com.
Христофорова Юлия Александровна, к.физ.-мат.н., доцент каф. ЛБС khristoforova.yua@ssau.ru.

УДК 543.424.2; 004.93

АНАЛИЗ ВЛИЯНИЯ АУГМЕНТАЦИИ ДАННЫХ НА СТРУКТУРУ РАМАНОВСКИХ СПЕКТРОВ СЫВОРОТКИ КРОВИ В ПРОСТРАНСТВЕ ГЛАВНЫХ КОМПОНЕНТ

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

Ключевые слова: PCA, аугментация, спектры рамановского рассеяния, дисперсия.

В исследовании использовались рамановские спектры сыворотки крови пациентов с заболеваниями различных нозологических групп: сердечно-сосудистыми (хроническая сердечная недостаточность и ишемическая болезнь сердца), аутоиммунными (системная красная волчанка и ревматоидное заболевание), а также респираторными заболеваниями. Исходная выборка включала 1245 спектров. Для анализа структуры данных и оценки влияния аугментации был применён метод главных компонент (РСА). Он позволяет снизить размерность спектральных данных и проанализировать распределение образцов в пространстве главных компонент, сохраняя максимальную долю дисперсии исходных данных [1].

На первом этапе РСА был применён к исходной выборке. Далее к тем же данным были применены три метода аугментации, каждый из которых увеличивал размер выборки в 5 раз. Аугментация позволяет искусственно расширить набор данных, сохраняя его статистические свойства, что особенно важно при ограниченном объёме экспериментальных выборок. Добавление случайного шума – к исходному спектру добавляется случайный сигнал с заданным распределением, что имитирует естественные флуктуации измерений и повышает устойчивость моделей к шуму. Сдвиг по оси волновых чисел заключается в небольшом смещении спектра вдоль горизонтальной оси, что позволяет моделировать инструментальные вариации. Масштабирование интенсивности изменяет амплитуду сигнала путём умножения на случайный коэффициент. Такой подход имитирует изменения условий регистрации [2]. Таким образом, для каждого метода аугментации формировалась выборка объёмом 6225 спектров, а суммарный объём проанализированных данных составил 16185 спектров. На рисунке 1 представлены спектры рамановского рассеяния сыворотки крови исходного образца и с добавлением случайного шума.

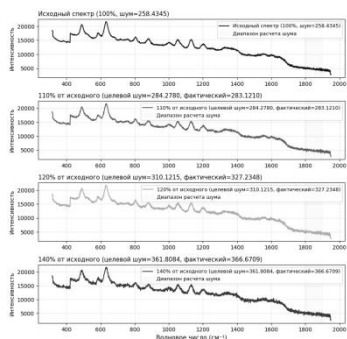


Рисунок 1 – Спектры рамановского рассеяния сыворотки крови исходного образца и с добавлением случайного шума с разными параметрами

На рисунке 2 представлено сравнение проекции аугментированной выборки на первые две главные компоненты на примере добавления шума

с проекцией исходных спектров. Центры распределений аугментированных данных в пространстве главных компонент практически совпадают с центром исходной выборки. Таким образом, синтетически сгенерированные спектры не смещают данные в пространстве признаков.

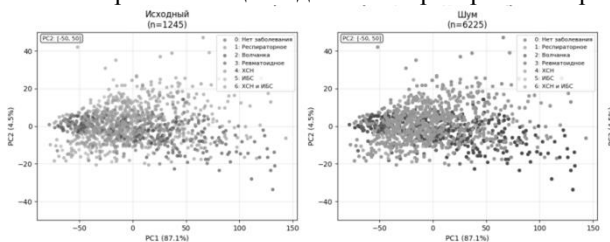


Рисунок 2 – Сравнение проекции аугментированной выборки на первые две главные компоненты на примере добавления шума с проекцией исходных спектров

Первая главная компонента (PC1) объясняет 87,1% общей дисперсии данных, вторая главная компонента (PC2) – 4,5%, что в сумме составляет более 91% дисперсии. Это указывает на то, что основная вариативность спектров эффективно описывается в низкоразмерном пространстве. После применения методов аугментации к данным повторно был выполнен PCA-анализ. Для всех вариантов аугментации значения объяснённой дисперсии PC1 оставались в диапазоне 87,02–87,18%, PC2 – 4,48–4,51%, что свидетельствует о сохранении направления максимальной вариативности и стабильности структуры пространства признаков.

Для количественной оценки структуры данных была рассчитана величина отношения межклассовой дисперсии к внутриклассовой. Для исходной выборки данная метрика составила 0,043 по PC1 и 0,852 по PC2. Низкие значения указывают на значительное перекрытие классов, что характерно для биологических спектров и отражает сложность задачи их разделения. После аугментации значения по PC1 остаются неизменными для всех методов и составляют = 0,043. Это указывает на сохранение структуры данных вдоль главного направления дисперсии. Для второй главной компоненты наблюдается небольшой разброс значений: 0,852 при добавлении шума, 0,859 при масштабировании, 0,857 при спектральном сдвиге.

Совместный анализ PCA-графиков и численных метрик показывает, что применённые методы аугментации не нарушают статистическую структуру исходной выборки. Сохранение положения центров распределений и близкие значения дисперсий по главным компонентам свидетельствуют о том, что синтетически сгенерированные спектры остаются согласованными с реальными экспериментальными данными.

Метрика отношения межклассовой к внутриклассовой дисперсии подтверждает, что аугментация не приводит к искусственному улучшению разделимости классов, а значит не вносит смещения, способного исказить последующее обучение моделей машинного обучения.

Список использованных источников

1. Nodehi, A. Recent advances in principal component analysis for directional data / A. Nodehi, M. Moghimbeygi // Journal of Multivariate Analysis. – 2025. – Vol. 206.

2. Zhao J. et al. Improving skin cancer detection by Raman spectroscopy using convolutional neural networks and data augmentation // Frontiers in Oncology. – 2024. – Т. 14. – С. 1320220.

Дьяченко Анжелика Игоревна, студ. гр. 6402-120304D, diychenkosmav@gmail.com.

Христофорова Юлия Александровна, к.физ.-мат.н., доцент каф. ЛБС khristoforova.yua@ssau.ru.

УДК 543.424.2; 004.93

ПРИМЕНЕНИЕ МЕТОДА ОПОРНЫХ ВЕКТОРОВ ДЛЯ АНАЛИЗА СПЕКТРОВ РАМАНОВСКОГО РАССЕЯНИЯ КОЖИ ПРИ РАЗЛИЧНЫХ ЗАБОЛЕВАНИЯХ

А.М. Сергеева, Ю.А. Христофорова

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

Ключевые слова: Метод опорных векторов, спектры рамановского рассеяния, хроническая сердечная недостаточность, хроническая болезнь почек.

Неинфекционные заболевания, такие как хроническая болезнь почек (ХБП, или СКД) и хроническая сердечная недостаточность (ХСН, или СНД), представляют собой серьезную проблему для здравоохранения из-за высокого уровня заболеваемости и смертности населения. Ранняя диагностика данных заболеваний крайне важна для своевременного начала лечения [1]. Традиционные методы, такие как анализ уровня креатинина и скорости клубочковой фильтрации для ХБП или электрокардиография и анализ на тропонины для ХСН, не всегда обладают достаточной чувствительностью на ранних стадиях или могут быть инвазивными. Кроме того, важной клинической задачей является дифференциальная диагностика этих заболеваний между собой. В связи с этим существует потребность в разработке неинвазивных и высокоточных методов скрининга и диагностики.

Целью настоящей работы является разработка мультиклассового классификатора на основе метода опорных векторов (SVM) для анализа рамановских спектров кожи с целью дифференциации пациентов с хронической болезнью почек, хронической сердечной недостаточностью и здоровых пациентов.

В исследовании были проанализированы спектры рамановского рассеяния, снятые с кожи предплечья пациентов. Исходная выборка включала