

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

На правах рукописи

Матвеева Ирина Александровна

**МЕТОД МУЛЬТИМОДАЛЬНОГО АНАЛИЗА РАМАНОВСКОГО РАССЕЯНИЯ И
ДЕРМАТОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ ДЛЯ ДИАГНОСТИКИ
НОВООБРАЗОВАНИЙ КОЖИ**

2.2.12. Приборы, системы и изделия медицинского назначения

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:

Захаров Валерий Павлович,

доктор физико-математических наук, профессор

Самара – 2024

Оглавление

ВВЕДЕНИЕ.....	4
1 ОПТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА И ДИАГНОСТИКИ НОВООБРАЗОВАНИЙ КОЖИ.....	14
1.1 Новообразования кожи и их диагностика.....	14
1.2 Рамановская спектроскопия для диагностики новообразований кожи	19
1.3 Цифровая дерматоскопия для диагностики новообразований кожи	30
1.4 Многомерный анализ данных для диагностики заболеваний	39
1.5 Постановка задачи диссертационного исследования.....	42
2 ДВУХКАНАЛЬНАЯ ЭКСПЕРИМЕНТАЛЬНАЯ СИСТЕМА И ОПТИЧЕСКИЕ ДААННЫЕ.....	44
2.1 Система регистрации рамановских спектров кожи и дерматоскопических изображений	44
2.2 Описание выборки оптических данных	47
3 РАЗРАБОТКА МЕТОДА АНАЛИЗА РАМАНОВСКИХ СПЕКТРОВ НОВООБРАЗОВАНИЙ КОЖИ С ВЫДЕЛЕНИЕМ КОМПОНЕНТНОГО СОСТАВА	53
3.1 Применение метода разрешения многомерных кривых для компонентного анализа рамановских спектров кожи.....	53
3.2 Компонентный анализ рамановских спектров.....	60
3.2.1 Разделение экспериментальных спектров.....	60
3.2.2 Разделение <i>in vivo</i> рамановских спектров	62
3.2.3 Изучение вклада отдельных компонентов в рамановские спектры различных новообразований	70
3.3 Идентификация новообразований кожи по спектральным данным	75
3.3.1 Классификация на основе компонентного профиля.....	75
3.3.2 Классификация новообразований на основе спектральных отсчетов и компонентного профиля	82

3.3.3	Классификация компонентных профилей новообразований методом градиентного бустинга.....	84
3.4	Выводы по главе	89
4	ИНТЕЛЛЕКТУАЛЬНЫЙ МЕТОД АНАЛИЗА ДЕРМАТОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ НОВООБРАЗОВАНИЙ КОЖИ.....	90
4.1	Архитектура сверточной нейронной сети для классификации дерматоскопических изображений	90
4.2	Идентификация новообразований кожи по дерматоскопическим изображениям	94
4.3	Выводы по главе	104
5	МУЛЬТИМОДАЛЬНЫЙ АНАЛИЗ СПЕКТРАЛЬНЫХ И ПРОСТРАНСТВЕННЫХ ПРИЗНАКОВ НОВООБРАЗОВАНИЙ.....	105
5.1	Идентификация новообразований кожи на основе мультимодального совместного анализа рамановских спектров и дерматоскопических изображений	105
5.2	Выводы по главе	113
	ЗАКЛЮЧЕНИЕ	115
	СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	117
	СПИСОК ЛИТЕРАТУРЫ	120
	Приложение А. Акт об использовании результатов диссертационного исследования.....	147

ВВЕДЕНИЕ

Диссертация посвящена разработке метода мультимодального многомерного анализа спектральных данных рамановского рассеяния и пространственных данных дерматоскопических изображений для диагностики новообразований кожи различных нозологий.

Актуальность темы.

Онкологические новообразования кожи являются одними из наиболее часто диагностируемых форм рака, что вызвано как генетическими факторами, образом жизни и прямым воздействием ультрафиолетового излучения солнца, так и возможностью их выявления при визуальном осмотре или с помощью цифровой дерматоскопии. Однако визуальная диагностика новообразований кожи существенно зависит от квалификации и профессионального опыта врача, врач ставит верный диагноз только в 40-80% случаев ввиду наличия множества различных нозологий новообразований кожи, имеющих схожие отличительные признаки [1]. Прогноз при онкологических заболеваниях кожи и метод лечения зависят от типа новообразования, его стадии и локализации. Золотым стандартом в онкологии является гистологическое исследование, но оно требует взятия пробы и не применимо при первичном осмотре или массовом мониторинге. Актуальной задачей является развитие инструментальных средств и методов обработки информации, позволяющих врачу произвести своевременную и точную диагностику новообразований кожи.

Степень разработанности темы.

Перспективными инструментами для неинвазивного исследования кожной ткани признаны такие оптические методы как ультрафиолетовая (УФ) флуоресцентная диагностика (Е.Ворисова [2]), оптическая когерентная томография (Г.В.Геликонов [3]), спектроскопия рамановского рассеяния (В.П.Захаров [4], И.А.Братченко [5], J.Popp [6], A. Mahadevan-Jansen [7], Z.Huang [8], J.Zhang [9]). Каждый из перечисленных методов обладает рядом преимуществ за счет признаков, которые невозможно или трудно выделить другими способами. Так,

оптическая когерентная томография (ОКТ) позволяет выделить градиенты неоднородности по глубине, но не включает спектральные признаки новообразований; УФ флуоресцентная диагностика и анализ обычных оптических изображений ввиду широкой полосы перекрывающихся спектров основных флуорофоров кожи имеет ограничения по точности даже при использовании нейронных сетей, обученных на большом количестве изображений (А. Esteva [10]); цифровая дерматоскопия, прежде всего, выделяет пространственные текстурные признаки (неоднородность плотности, изрезанность границы, градиенты пространственных частот), при этом теряется информация о характерных химических изменениях, происходящих при развитии новообразования. Для анализа дерматоскопических изображений успешно применяются алгоритмы машинного обучения (Т. Brinker [11], А. Mahbod [12]). Однако реализация методов глубокого обучения (А. Takiddin [13]) требует использования больших обучающих выборок. Такой объем данных получить затруднительно, поэтому необходимо применение методов аугментации и дообучения классификатора при добавлении новых данных. Существующие ограничения каждого из методов привели к созданию и развитию мультимодальных методов диагностики (J. Ropp, В. В. Тучин, В. П. Захаров [14]), эффективная реализация которых возможна путем объединения в единую систему всех признаков, выделяемых каждым из методов.

Наибольшей чувствительностью к изменению химического состава новообразований обладают методы рамановской спектроскопии (RS) за счет регистрации интенсивностей колебательно-вращательных колебаний функциональных групп молекул нуклеиновых кислот, белков, липидов и углеводов. В качестве признаков в алгоритмах RS-методов выбираются спектральные отсчеты и полосы рассеяния, зарегистрированные в условиях высокого уровня шумов (включая сопутствующую инфракрасную (ИК) флуоресценцию), что исключает определение пространственных градиентов распределения химических компонент. В работе И. А. Братченко [5] применялся дискриминантный анализ при помощи проекции на латентные структуры (PLS-DA), обеспечивая статистически надежный результат. Метод главных компонент

(РСА) успешно использовался для получения различий в спектральных данных в работах В.П.Захарова [4] и М.Wu [15]. Дальнейшее развитие оптическая диагностика рака кожи получила с развитием алгоритмов классификации на основе нейронных сетей, демонстрирующих высокую чувствительность и специфичность в реальном времени. Так, в работах R.Luo [16] и И.А.Братченко [17] показано, что применение сверточных нейронных сетей (CNN) для анализа RS спектров позволяет добиться точности определения типа новообразования кожи, сравнимой с диагностической точностью квалифицированных дерматологов. Однако в большинстве исследований этот подход представлен в виде «черного ящика», сосредотачиваясь исключительно на проблеме классификации, а критерии классификации очень часто остаются неизвестными, не позволяют явно провести их физическую интерпретацию и объективно обосновывать результаты исследования на основе изменения химического состава новообразований, которое фактически лежит в основе методов RS.

В последнее время широкое применение в биологии и медицине для определения концентрации компонентов смеси по спектрам нашел метод разрешения многомерных кривых (MCR). Для биохимической интерпретации RS спектров новообразований кожи L.Silveira [18] предложил использовать их разложение по RS спектрам чистых химических веществ, однако данный метод слабо сходилась ввиду сильного изменения спектров колебаний молекул в химически связанных компонентах. Существенно лучшей сходимости удалось добиться X.Feng и J.Tunnell [19] за счет использования разложения по RS спектрам отдельных клеточных структур, зарегистрированных при микроскопическом исследовании тонких слоев кожи. Тем не менее, в клинических условиях RS спектры новообразований снимаются со значительно большего объема кожи, поэтому алгоритмы и методы их биохимической интерпретации требуют развития и уточнения с учетом зависимости характеристик мод колебаний молекул от их связей и окружения в реальной биологической ткани.

В связи с этим возникает необходимость разработки нового мультимодального подхода, объединяющего алгоритмы и методы анализа спектральных (рамановская

спектроскопия) и пространственных (цифровая дерматоскопия) признаков с одновременной биохимической интерпретацией результата исследования, позволяющей врачу уверенно и объективно диагностировать новообразования кожи.

Цель диссертационной работы.

Целью диссертационной работы является разработка метода идентификации новообразований кожи и интерпретации их компонентного состава на основе мультимодального многомерного анализа спектральных данных рамановского рассеяния и дерматоскопических изображений новообразований кожи.

Задачи диссертационной работы.

1. Исследование способов восстановления и интерпретации компонентного состава образцов кожной ткани при многомерном разрешении спектральных данных рамановского рассеяния в ближнем инфракрасном диапазоне при возбуждении лазерным излучением на длине волны 785 нм;

2. Разработка метода и программно-алгоритмического обеспечения распознавания рамановских спектров кожи при решении классификационных задач определения типа новообразования с учетом выделения спектральных признаков с интерпретацией компонентного состава исследуемого образца;

3. Разработка метода интеллектуального распознавания дерматоскопических изображений при решении классификационных задач определения типа новообразований методами машинного обучения;

4. Разработка метода идентификации типа новообразования на основе мультимодального совместного анализа рамановских спектров и дерматоскопических изображений.

Научная новизна работы.

1. Предложен алгоритм выделения и интерпретации компонентного состава ткани кожи на основе разрешения многомерных кривых спектральных данных *in vivo* рамановского рассеяния в ближнем инфракрасном диапазоне при возбуждении на длине волны 785 нм с использованием метода чередующихся наименьших квадратов, учитывающий неотрицательность и замкнутость спектральных

профилей выделяемых компонентов, а также позволяющий автоматическое выделение вклада оптико-электронного тракта прибора регистрации спектров. Показана сходимость разложения по выделяемым компонентам, обеспечивающая ошибку интерпретации спектра кожи не более 5% при использовании не менее 8 групп химических компонентов кожи.

2. Предложен метод распознавания рамановских спектров новообразований, зарегистрированных в диапазоне $792-1874 \text{ см}^{-1}$ в условиях высоких шумов, и разработано программно-алгоритмическое обеспечение (свидетельство о государственной регистрации программы для ЭВМ № 2023685208 получено 23.11.2023) для идентификации типа новообразования на основе алгоритма градиентного бустинга с количеством деревьев не более 100 при разрешении спектральных признаков с выделением вклада меланина, белков, воды и липидов. Модели бинарной классификации доброкачественных и злокачественных новообразований, злокачественной меланомы и доброкачественных пигментных новообразований, злокачественной меланомы и себорейного кератоза характеризуются ROC AUC 0,82 (0,77-0,86, 0,95 CI), 0,82 (0,76-0,89, 0,95 CI) и 0,87 (0,79-0,94, 0,95 CI), соответственно, что превосходит значения ROC AUC моделей классификации перечисленных нозологий методом PLS-DA на 7%, 13% и 6%, соответственно.

3. Предложен метод идентификации новообразований кожи на основе мультимодального анализа спектральных признаков новообразования с интерпретацией компонентного состава методом градиентного бустинга и пространственных признаков дерматоскопического изображения, выделяемых нейронной сетью модифицированной архитектуры VGG16, содержащей 5 сверточных слоев. За счет комплексного взаимодополнения выделяемых пространственных и спектральных признаков метод характеризуется высокой устойчивостью и значением ROC AUC 0,94 (0,90-0,98, 0,95 CI) при бинарной классификации злокачественной меланомы и доброкачественных пигментных новообразований, что выше значений ROC AUC при использовании только

пространственных признаков или только спектральных признаков на 7% и 12%, соответственно.

Соответствие паспорту специальности 2.2.12 Приборы, системы и изделия медицинского назначения.

Область исследования соответствует пунктам направлений исследований:

14. Методы, модели и алгоритмы, включая распознавание образов, для медицинских информационных и интеллектуальных систем обеспечивающих повышение эффективности медико-биологических исследований и врачебных решений;

19. Методы и средства регистрации, анализа и интерпретации медицинских изображений;

21. Методы и средства искусственного интеллекта для медико-биологических исследований;

22. Специальное программное обеспечение, предназначенное для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека и проведения медицинских исследований.

Положения, выносимые на защиту.

1. Алгоритм выделения и интерпретации компонентного состава ткани кожи на основе разрешения многомерных кривых спектральных данных *in vivo* рамановского рассеяния в ближнем инфракрасном диапазоне при возбуждении на длине волны 785 нм с использованием метода чередующихся наименьших квадратов, учитывающий неотрицательность и замкнутость спектральных профилей выделяемых компонентов, обеспечивающий ошибку интерпретации спектра кожи не более 5% при выделении не менее 8 групп химических компонентов кожи. (Соответствие пункту 14 паспорта специальности).

2. Метод распознавания рамановских спектров новообразований, зарегистрированных в диапазоне 792-1874 см⁻¹ в условиях высоких шумов, и программно-алгоритмическое обеспечение для идентификации типа новообразования на основе алгоритма градиентного бустинга с количеством

деревьев не более 100 при разрешении спектральных признаков с выделением вклада меланина, белков, воды и липидов, реализующее бинарную классификацию для идентификации доброкачественных и злокачественных новообразований, злокачественной меланомы и доброкачественных пигментных новообразований, злокачественной меланомы и себорейного кератоза с ROC AUC, равными 0,82 (0,77-0,86, 0,95 CI), 0,82 (0,76-0,89, 0,95 CI) и 0,87 (0,79-0,94, 0,95 CI), соответственно. (Соответствие пунктам 14, 22 паспорта специальности).

3. Метод идентификации новообразований кожи на основе мультимодального анализа спектральных признаков новообразования с интерпретацией компонентного состава методом градиентного бустинга и пространственных признаков дерматоскопического изображения, выделяемых нейронной сетью модифицированной архитектуры VGG16, содержащей 5 сверточных слоев, обеспечивающий повышение ROC AUC модели классификации меланомы и доброкачественных пигментных новообразований до 0,94 (0,90-0,98, 0,95 CI) за счет комплексного взаимодополнения выделяемых пространственных и спектральных признаков. (Соответствие пунктам 19, 21 паспорта специальности).

Практическая значимость.

Практическая ценность результатов работы заключается в разработанных методах, моделях и алгоритмах мультимодального многомерного анализа, позволяющих обеспечить распознавание типа новообразования и повысить эффективность диагностики онкологических заболеваний кожи за счет объединения пространственных, спектральных и компонентных признаков, а также в разработке программно-алгоритмического обеспечения, реализующего указанные алгоритмы и методы. Алгоритмы и программные модули внедрены в научно-исследовательскую и клиническую деятельность ГБУЗ «Самарский областной клинический онкологический диспансер» (см. приложение А).

Практическая значимость результатов работы подтверждена их использованием в рамках гранта по программе развития Самарского университета на 2021-2030 годы в рамках программы стратегического академического лидерства «Приоритет 2030» по теме «Методы интеллектуальной обработки и анализа

мультимодальных изображений для построения и отслеживания состояния цифровых двойников антропогенных и биологических объектов» (соглашение № ПР-НУ/2.1-03-2022, шифр темы 21в-Р020-208), гранта Самарского университета на развитие стартап-проектов в 2022 году в целях реализации Программы развития Самарского университета на 2021-2030 годы в рамках программы стратегического академического лидерства «Приоритет 2030» по теме «Разработка программно-аппаратного дерматоскопического комплекса для автоматической неинвазивной диагностики новообразований кожи» (соглашение № ПР-НУ/2.5-02-2022/01 от 26.09.2022) и гранта по теме «Разработка неинвазивной технологии определения биохимического состава кожи», конкурс «УМНИК-Фотоника» в рамках программы УМНИК Фонда содействия инновациям (договор № 18063ГУ/2022 от 23.11.2022).

Достоверность полученных результатов.

Достоверность результатов работы обеспечивается за счёт использования сертифицированных оптических приборов и строгости примененных методов математического анализа и машинного обучения. Достоверность представленных научных результатов подтверждается воспроизводимостью экспериментальных и расчетных данных, а также их соответствием результатам работ других авторов. Основные результаты работы опубликованы в российских и международных рецензируемых научных журналах.

Методы исследования.

В качестве экспериментальных методов исследования в диссертационной работе использовались методы объемной рамановской спектроскопии и интеллектуального анализа дерматоскопических изображений, математическая обработка данных осуществлялась такими методами машинного обучения как логистическая регрессия, дискриминантный анализ при помощи проекции на латентные структуры, градиентный бустинг, метод разрешения многомерных кривых с использованием метода чередующихся наименьших квадратов и сверточные нейронные сети. Обработка экспериментальных данных и построение моделей интеллектуального анализа осуществлялись с использованием среды разработки Jupyter Notebook на языке программирования Python.

Личный вклад автора.

Личный вклад автора заключается в самостоятельном выполнении предварительной обработки данных, разработке методов анализа спектральных и пространственных данных, реализации алгоритмов глубокого обучения, выполнении интеллектуального анализа спектральных данных рамановского рассеяния и дерматоскопических изображений, построении моделей классификации, разработке метода мультимодального анализа спектральных и пространственных признаков, визуализации и интерпретации компонентного состава и полученных результатов. Изложенные в диссертации оригинальные результаты получены автором лично либо при его непосредственном участии. Постановка задач и обсуждение результатов проводились совместно с научным руководителем.

Публикации.

По теме диссертации опубликовано 20 научных работ, в том числе 7 публикаций в научных изданиях, рекомендованных Высшей аттестационной комиссией и/или рецензируемых базами данных Scopus и Web of Science, и получено 2 свидетельства о государственной регистрации программы для ЭВМ.

Апробация результатов диссертации.

Результаты диссертационной работы были представлены на XVIII, XIX, XX Всероссийских молодежных Самарских конкурсах-конференциях научных работ студентов и молодых исследователей по оптике и лазерной физике, г. Самара (2020, 2021, 2022 гг.); XXV, XXVI, XXVII Международных конференциях для молодых ученых и студентов Saratov Fall Meeting – International School for Junior Scientists and Students on Optics, Laser Physics & Biophotonics, г. Саратов (2021, 2022, 2023 гг.); Международном симпозиуме The 13th Winter Symposium on Chemometrics, г. Москва (2022 г.); Международном онлайн-семинаре Raman-based biosensors. European Biosensor Symposium – Digital Seminar Series (2021 г.), Международной конференции Photonics Europe, г. Страсбург, Франция (2020, 2022 гг.); Всероссийских научно-технических конференциях «Актуальные проблемы радиоэлектроники и телекоммуникаций», г. Самара (2022, 2023 гг.); VII, VIII, IX

Международных конференциях и молодежных школах «Информационные технологии и нанотехнологии», г. Самара (2021, 2022, 2023 гг.); Международной конференции «Terahertz and microwave radiation: generation, detection and applications TERA-2020», г. Томск (2020 г.); Международной конференции «Технические средства систем управления и связи = International Scientific Forum on Control and Engineering», г. Астрахань (2022 г.); XVII Международной научно-технической конференции «Оптические методы исследования потоков», г. Москва (2023 г.); Всероссийской диановской конференции по волоконной оптике, г. Пермь (2023 г.); Всероссийской научной конференции с международным участием «Невская фотоника – 2023», г. Санкт-Петербург (2023 г.).

Научно-квалификационная работа аспиранта по теме диссертации одержала победу во Всероссийском инженерном конкурсе в 2023 году.

Структура и объем диссертации.

Диссертация состоит из введения, пяти глав, заключения, списка сокращений и условных обозначений, списка литературы, включающего 251 наименование, и одного приложения. Работа изложена на 147 страницах машинописного текста, содержит 31 рисунок, 15 таблиц.

1 ОПТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА И ДИАГНОСТИКИ НОВООБРАЗОВАНИЙ КОЖИ

1.1 Новообразования кожи и их диагностика

По данным Всемирной организации здравоохранения (ВОЗ), на 2020 год приходится около 1,2 млн новых случаев немеланомных опухолей кожи [20], а всего в 2020 году было зарегистрировано более 1,5 миллиона новых случаев рака кожи. По оценкам Международного агентства по исследованию рака (IARC), в 2020 году во всем мире было диагностировано около 325 000 новых случаев меланомы и 57 000 человек умерли от этой болезни. Более того, исследователи прогнозируют рост заболеваемости раком кожи. Так, новое исследование ученых IARC показало, что количество новых случаев меланомы кожи в год увеличится более чем на 50% с 2020 по 2040 год [21].

По данным статистики онкологических заболеваний, в Российской Федерации по данным на 2022 год насчитывается 72,7 больных злокачественной меланомой кожи на 100 тыс. населения и 298,9 больных раком кожи (всех типов кроме меланомы) [22]. При этом летальность больных с меланомой в течение года с момента установления диагноза (из числа больных, впервые взятых на учет в предыдущем году) составляет 7,5 % [22].

Существует множество различных нозологий новообразований кожи. Чаще всего их разделяют на две большие группы по типу роста: злокачественные и доброкачественные. Среди злокачественных новообразований обычно выделяют меланому и немеланомные виды рака кожи. Ко второй группе относится одно из самых распространенных злокачественных новообразований кожи — базальноклеточная карцинома (БСС). Его доля в структуре зарегистрированных случаев злокачественных новообразований эпидермального происхождения в РФ составляет 60–80% [23].

БСС возникает из кератиноцитов вблизи базального слоя и наружной корневой оболочки волосяных фолликулов. Большинство базальноклеточных карцином

встречаются на участках кожи, подвергающихся воздействию ультрафиолетового излучения солнечного света, причем 75-85% поражений обнаруживаются в областях головы и шеи. ВСС чаще всего встречается у пожилого населения и чаще поражает мужчин, не дает метастазов и при лечении имеет благоприятный прогноз в практически 100% случаев [23].

Плоскоклеточный рак (SCC) является вторым по распространенности немеланомным раком кожи после ВСС. Наиболее значимые факторы риска SCC включают воздействие солнца, возраст и подавление иммунитета. SCC поражает в основном участки, подверженные воздействию солнца, такие как лицо, волосистая часть головы, шея, руки и кисти. Частота заболеваемости данным видом рака увеличивается с возрастом. SCC чаще встречается у мужчин, чем у женщин (соотношение 3:1) [24].

Одним из наиболее опасных типов рака является злокачественная меланома (ММ), развивающаяся из клеток-меланоцитов [25]. Хотя она встречается реже прочих злокачественных новообразований кожи (4% от общего числа выявленных случаев рака кожи), за счёт высокой вероятности метастазирования меланома является причиной 80% всех случаев смерти от рака кожи [26]. В 70 % случаев злокачественная меланома кожи развивается на месте доброкачественного пигментного образования – невуса. Это называется малигнизацией [27]. Пигментный невус (PN, NE) – это доброкачественная опухоль из вырабатывающих меланин клеток, встречающаяся примерно у 90% населения. Выделяют несколько клинико-морфологических типа меланом, отличающихся по этиологии и клиническому течению: поверхностно распространяющаяся, меланома типа злокачественного лентиго, акральная лентигозная меланома, узловатая меланома. Также встречается амеланоцитарная (беспигментная) меланома – опухоль в форме папулы, бляшки, бородавки, лишённая пигмента, часто с изъязвлением [28].

Помимо перечисленных злокачественных новообразований существует огромное количество разнообразных доброкачественных новообразований. К таким опухолям можно отнести уже упомянутые невусы, дерматофибром (DF), папилломы различных видов (PP), себорейный кератоз (SK) и др.

Для всех видов новообразований кожи первичным средством диагностики служит визуальный осмотр врача-дерматолога. Тщательный сбор анамнеза, знание признаков малигнизации, осмотр кожных покровов представляют большую ценность для постановки диагноза [26]. Важное значение имеет определение диагностических признаков малигнизации (трансформации доброкачественных клеток в злокачественные) новообразований на коже по ABCD-системе (см. таблицу 1.1) и Глазговской 7-точечной системе (см. таблицу 1.2).

Таблица 1.1 – ABCD-система признаков малигнизации образований на коже [29]

Обозначение	Признак	Характеристика признака
A	Asymmetry (асимметрия)	Форма асимметричная
B	Boundary (границы)	Контрастные и неровные границы
C	Color (цвет)	Неравномерный синеватый или коричнево-черный цвет
D	Dimension (величина)	Диаметр более 6 мм

Таблица 1.2 – Глазговская 7-точечная система признаков малигнизации образований на коже [30]

Точка	Характеристика признака
1	Изменение размеров, объема
2	Изменение формы, очертаний
3	Изменение цвета
4	Воспаление
5	Образование корки, кровоточивость
6	Изменение ощущений, чувствительности
7	Диаметр более 7 мм

При наличии любого из признаков, перечисленных в таблицах 1.1 и 1.2, новообразование следует рассматривать как потенциально злокачественное [23]. Однако, несмотря на наличие систем критериев злокачественности, качество диагностики рака кожи путем визуального осмотра во многом зависит от уровня квалификации и профессионального опыта врача. Например, врач верно определяет

меланому только в 40-80% случаев (в зависимости от квалификации) [1, 31]. Внешне меланома имеет характеристики, схожие с пигментным невусом, из которого (или на фоне которого) меланома может развиваться в результате ряда внешних воздействий. Невус, в свою очередь, представлен широким диапазоном форм, оттенков и размеров, существенно усложняющим диагностику «на глаз». Кроме того, поскольку перерождение диспластического невуса в меланому происходит на клеточном уровне, отличить их *in situ* считается невозможным.

Наиболее полную клиническую картину даёт взятие биопсии, предусматривающее забор образца новообразования и дальнейшее его гистологическое исследование специалистом-патоморфологом [32]. Из-за «агрессивного» поведения и высоких рисков метастазирования при внешнем воздействии подобная процедура, как правило, не применяется [33, 34]. Вместо этого, удалению подлежит весь поражённый участок и небольшая зона заведомо здоровой ткани вокруг, а лабораторные исследования проводятся постфактум [35, 36].

Перспективными инструментами для исследования кожной ткани признаны оптические методы. Возможность проведения тестов биоткани *in vivo*, их быстрого анализа, неинвазивность и точность исследований делают оптические методы наиболее подходящими для этой задачи. Методы оптической спектроскопии стали основой для развития направления неинвазивного контроля и диагностики биоткани, получившего название «оптическая биопсия», которое может применяться, в частности, для раннего обнаружения злокачественных новообразований кожи [37, 38]. Целью развития методов оптической биопсии является возможность неинвазивной регистрации оптических характеристик биотканей новообразований в естественных условиях и их интерпретации, т.е. обеспечение диагностика новообразований только на основе оптических измерений [39]. Это позволит устранить необходимость хирургического взятия проб образцов ткани для проведения гистологического исследования, либо, как минимум, снизить их необходимое количество за счет предварительного целеуказания локации опухоли при оптической биопсии. Кроме того, быстрота

получения информации и неинвазивность исследования позволяют снизить эмоциональную травму пациента в ожидании диагноза. Это, наряду с высокой скоростью оптического анализа биоткани, проводимого в реальном режиме времени, и определяет возможность обнаружения онкологических новообразований на ранних этапах ее развития.

В настоящее время оптическая диагностика рака кожи осуществляется с использованием следующих оптических методов: конфокальная микроскопия [40], оптическая когерентная томография [41, 42, 43], многофотонная томография [44], различные типы спектроскопии, мультиспектральная визуализация [45] и т.д. Большинство результатов исследований, проведенных с использованием этих технологий, показывают, что они могут быть сравнимы по точности с диагностикой врачом-дерматологом [45]. Так, в исследовании Н.А.Наенсле [1] средняя чувствительность и специфичность 58 дерматологов для классификации 100 поражений кожи составила 86,6% ($\pm 9,3\%$) и 71,3% ($\pm 11,2\%$), что соответствует точности диагностики (доле верно определенных диагнозов) 79% ($\pm 0,06\%$). Эксперты в области дерматоскопии показывают более высокие средние значения чувствительности, специфичности и ROC AUC, чем начинающие специалисты – 89% ($\pm 9,2\%$), 74,5% ($\pm 12,6\%$), 0,82 ($\pm 0,06$) против 82,9% ($\pm 7,1\%$), 67,6% ($\pm 6,3\%$), 0,75 ($\pm 0,04$) соответственно [1]. Точность современных методов оптической диагностики приближается к 80%, а в некоторых случаях и превосходит данное значение. Вместе с тем, следует иметь в виду, что кроме высокой точности немаловажным фактором является простота и быстрота получения результата исследования.

Однако в данный момент методы оптической биопсии не превосходят по точности «золотой стандарт» диагностики – гистологическое исследование – и поэтому рассматриваются не как самостоятельный способ диагностики, а как дополнительные инструменты, помогающие оперативно (фактически в реальном масштабе времени) предсказывать тип заболевания, а также использоваться в качестве оптического целеуказания, исключающего ошибку в месте забора пробы для гистологического исследования.

Авторы обзора [45] считают, что одним из способов преодоления ограниченных значений чувствительности и специфичности является комбинированное использование нескольких оптических методов. Более того, использование таких методов в сочетании с технологиями искусственного интеллекта, а именно машинным обучением и глубоким обучением, позволит улучшить эффективность диагностики без необходимости привлечения опытных дерматологов [46].

Известно, что эффективность лечения в значительной степени зависит от стадии заболевания, на которой оно обнаружено. Следовательно, крайне необходимым является развитие таких мультимодальных технических решений и методов, которые можно эффективно применять не только в специализированных онкологических учреждениях, но и при ежегодной диспансеризации населения или на стадии первичного осмотра врачом первичного звена.

На основании вышеизложенного в данной работе основное внимание уделяется развитию методов диагностики новообразований с использованием спектроскопии рамановского рассеяния и цифровой дерматоскопии.

1.2 Рамановская спектроскопия для диагностики новообразований кожи

В основе рамановской спектроскопии лежит явление рамановского рассеяния, или комбинационного рассеяния, излучения, при котором частоты квантов рассеянного излучения сдвигаются относительно кванта падающего света на частоты, соответствующие собственным частотам колебательно-вращательных состояний молекул исследуемого вещества (рисунок 1.1). Это явление было открыто в 1928 году одновременно советскими физиками Г. С. Ландсбергом и Л. И. Мандельштамом при исследовании рассеяния света в кристаллах [47] и индийскими физиками Ч. В. Раманом и К. С. Кришнаном при исследовании рассеяния света в жидкостях [48]. Следует отметить, что вероятность рамановского рассеяния по сравнению с рэлеевским рассеянием, сохраняющим частоту, крайне мала, фактически из 10^9 фотонов только один испытывает сдвиг

частоты при рассеянии. В связи с этим фундаментальной проблемой рамановской спектроскопии является выделение слабого сигнала на фоне интенсивного рэлеевского рассеяния, что требует предварительной фильтрации рассеянного излучения с использованием отрезающих фильтров с большой оптической плотностью (OD6 и выше).

Как видно из рисунка 1.1, при обмене энергией между фотоном и молекулой, последняя может как приобрести (стоксово рассеяние), так и потерять (антистоксово рассеяние) часть своей энергии в соответствии с правилами квантования. Соответственно, стоксовы переходы формируют в спектре рамановского рассеяния пики, сдвинутые в длинноволновую область спектра относительно длины волны возбуждающего излучения, а антистоксово рассеяние, наоборот, – в коротковолновую область спектра.

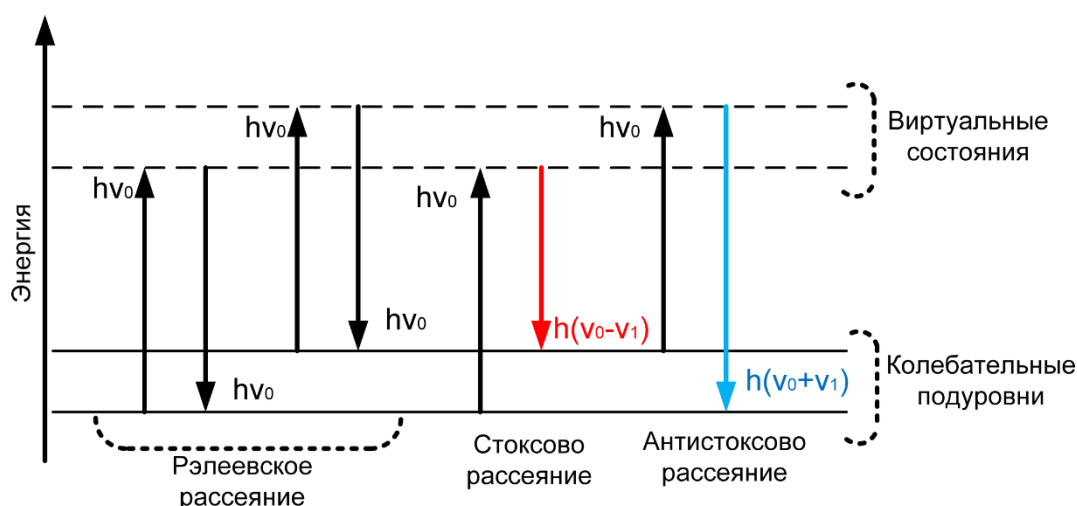


Рисунок 1.1 – Иллюстрация принципа рамановского рассеяния света (h – постоянная Планка, ν_0 – частота падающего кванта, ν_1 – сдвиг частоты)

Следует отметить, что антистоксовы линии в спектре формируются за счет переходов, стартующих из возбужденных состояний. Поскольку в равновесном состоянии распределение молекул по вращательно-колебательным состояниям определяется распределением Больцмана, интенсивность антистоксова рассеяния ниже, чем стоксова рассеяния [49]. С другой стороны, обычно наряду с рамановским рассеянием в исследуемом образце возбуждается флуоресценция, спектральный диапазон которой совпадает со стоксовыми пиками. В связи с этим

возникает дилемма, какую область рамановского спектра использовать для анализа: в случае использования пиков антистоксова рассеяния необходимо выделять сверхслабые сигналы, но в отсутствии флуоресцентного фона (когерентная антистоксова рамановская спектроскопия (CARS)); во втором случае необходимо выделять стоксовы пики на фоне мощного флуоресцентного сигнала (стоксова объемная рамановская спектроскопия). Однако и в том, и в другом случае спектры представляют собой спектральные пики специфических колебательно-вращательных мод молекул исследуемого образца, что позволяет идентифицировать образец и охарактеризовать его состав [50]. Спектроскопия рамановского рассеяния света является эффективным методом исследования строения молекул и процессов их взаимодействия с окружающей средой. Каждое химическое соединение имеет свой специфический рамановский спектр, на основании анализа которого можно идентифицировать это соединение и обнаружить его в смеси.

Известно, что в процессе развития заболевания в организме пациента происходят метаболические изменения, что приводит к изменению биохимического состава биологических тканей и жидкостей. Например, в случае диабета высокий уровень глюкозы в крови может заставить почки выводить его из организма с мочой [51]. В случае атеросклероза нездоровый липидный профиль крови является одним из факторов риска развития. Связь между биохимическими маркерами и так называемыми западными заболеваниями, такими как рак груди, рак простаты, ишемическая болезнь сердца и рак толстой кишки, описана в [52]. Многие исследователи отмечают изменения в профилях свободных аминокислот в плазме (PFAA) у больных раком [53, 54, 55]. Таким образом, определение состава биологических тканей и жидкостей может быть очень ценным для диагностики большого количества различных заболеваний. Сравнительный анализ рамановских спектров злокачественных и доброкачественных новообразований позволяют определять конкретные типы ткани кожи [6].

Классическими объектами для интерпретации рамановских спектров биомолекул являются белки и их компоненты. К настоящему времени накоплен

большой экспериментальный материал по спектроскопии рамановского рассеяния света биологическими молекулами. Анализируя положение максимума и форму контура полосы (например, разлагая экспериментальную полосу по спектрам отдельных структур), можно получать количественные оценки содержания отдельных типов структуры в исследуемом белке. Высокое разрешение и чувствительность метода спектроскопии рамановского рассеяния позволяют осуществить быструю идентификацию и анализ состава веществ [6].

В объемной рамановской спектроскопии, как правило, стоксов сигнал регистрируется при непрерывном освещении лазером с высокой спектральной стабильностью. Интенсивность сигнала увеличивается для коротких длин волн, но собственная флуоресценция ткани также увеличивается, в то время как глубина проникновения света в ткань уменьшается из-за поглощения хромофоров, в частности гемоглобина и меланина, так что соотношение полезного сигнала к общему фону уменьшается. Кроме того, коэффициент поглощения $\mu_a(\lambda)$ в видимом спектральном диапазоне сильно зависит от длины волны λ и типа исследуемой биоткани, что препятствует сопоставлению интенсивностей разных пиков рамановского рассеяния. Выбор длины волны обычно предполагает компромисс между интенсивностью рамановского сигнала и уровня автофлуоресцентного сигнала исследуемой биологической ткани, и зависит от конкретной цели исследования. Оптимальные оптические окна для исследования рамановского рассеяния кожи составляют от 700 до 900 нм и около 1064 нм [56].

Технический прогресс в лазерной технике и производстве высокочувствительных приемников оптического излучения позволил значительно повысить эффективность исследования биологических тканей и жидкостей с применением рамановской спектроскопии и/или автофлуоресцентного анализа [6, 56, 57, 58, 59, 60, 61] не только в лаборатории, но и непосредственно в клинической практике. Возможность проводить тесты *in vivo* и неинвазивность делают рамановскую спектроскопию потенциальным лидером среди новых методов анализа биологических тканей и жидкостей [62]. Поскольку рамановский спектр содержит характерные полосы, обусловленные присутствием определенных

химических связей в исследуемом объекте, это делает каждый спектр своеобразным «отпечатком пальца» исследуемой биоткани [56]. Многие исследователи отмечают, что развитие новообразований не приводит к появлению новых пиков в спектре по сравнению с рамановским спектром окружающей здоровой ткани, что говорит о перерождении ткани без появления принципиально новых химических веществ. Однако происходящие в области новообразования биохимические процессы влияют на концентрацию различных биохимических компонентов, что приводит к изменению измеренных рамановских спектров и позволяет дифференцировать различные патологии. Поскольку отличительной особенностью биологических молекул является их многозвенность, то каждое биохимическое вещество характеризуется множеством колебательных мод, которые формируют набор пиков в спектральной области «отпечатка пальца», как правило частично перекрывающийся с наборами пиков, характерными для других веществ. Причем интенсивность того или иного колебания в значительной мере определяется окружением и связями, что делает задачу определения компонентного состава биологической ткани нетривиальной задачей. Именно поэтому во многих исследованиях спектры рамановского рассеяния анализируются только по интенсивностям мод отдельных колебаний без восстановления компонентного состава исследуемого образца, то есть априорно предполагая, что изменения в зарегистрированном спектре произошли из-за изменения связей или концентраций химических веществ.

Следует отметить, что рамановскую спектроскопию кожных покровов можно использовать не только для диагностики рака кожи [5], но и для диагностики почечной недостаточности [63], сердечно-сосудистых заболеваний [64, 65], что связано с накоплением в коже конечных продуктов гликирования при развитии патологии. Известны применения рамановской спектроскопии для оценки косметологического действия кремов на кожу [66], анализа дерматита кожи [67], а также дифференцирования ряда других заболеваний [56].

Применение рамановской спектроскопии для исследования рака кожи представлено в многочисленных исследованиях [68, 69, 70]. Оптическая биопсия

на основе применения объемной рамановской спектроскопии и автофлуоресцентного анализа демонстрирует высокий уровень точности (более 80%) в диагностике рака кожи различных нозологий [56, 71, 72, 73, 74, 75]. Обычно для *in vivo* исследований применяются рамановские системы с глубоким захлаживанием детектора, чтобы обеспечить уверенное выделение рамановского сигнала на фоне флуоресценции. Это, естественно, усложняет установку и требования к ее эксплуатации в клинических условиях. В этой связи следует отметить работу [76], в которой продемонстрирована возможность классификации здоровой кожи и немеланомных злокачественных и предраковых новообразований с помощью портативной системы рамановского рассеяния.

Интересны исследования изменения состава и распределения компонент биологических тканей по глубине, которые реализуются при применении рамановской спектроскопии в конфокальной микроскопии, что позволяет восстанавливать и интерпретировать распределение химического состава исследуемой ткани с поперечным и осевым разрешением до 2 мкм и менее [77]. Конфокальная рамановская спектроскопия была реализована как для *ex vivo*, так и *in vivo* исследований [78]. Однако в последнем случае требуется довольно длительное время накопления сигнала, что затрудняет применение метода в клинической практике.

В работе [79] проведено исследование *ex vivo* и *in vivo* срезов здоровой кожи и меланомы с помощью CARS для оценки распределения феомеланина в клетках и тканях. Современные когерентные методы рамановской спектроскопии открывают уникальные возможности анализа кожных тканей, однако высокие требования к чувствительности используемых установок ограничивают применение CARS спектроскопии для массового скрининга, особенно при обследованиях опухолей кожи. Вероятно, данный метод найдет свое применение в клинической практике для экспресс-анализа гистологических срезов без их заморозки и окрашивания.

Другой возможностью повышения эффективности исследования является использование поверхностно усиленной рамановской спектроскопией (SERS), использующей эффект усиления стоксова рассеяния на несколько порядков за счет

формирования поверхностного плазмонного резонанса вблизи поверхности металлических (как правило, серебряных или золотых) наночастиц [80]. Такое усиление обеспечивает высокую чувствительность и специфичность SERS в обнаружении специфических молекул или биомаркеров, что может значительно повысить точность обнаружения рака и других заболеваний [81]. Однако реализация SERS требует дополнительной трудоемкой подготовки (специализированные подложки или введение наночастиц) [56, 82], что ограничивает ее использование для массового скрининга.

Таким образом, для массовых исследований новообразований кожи наиболее доступным из вышеупомянутых спектроскопических методов является объемная рамановская спектроскопия из-за компромисса простоты и времени накопления сигнала по сравнению с другими методами рамановской спектроскопии.

Использование рамановской спектроскопии в медицинской диагностике облегчается развитием технических средств для регистрации рамановского рассеяния света: стабилизированных полупроводниковых лазеров ближнего инфракрасного (ИК) диапазона, гибких волоконно-оптических рамановских зондов с встроенными узкополосными и широкополосными отрезающими фильтрами и портативных спектрометров [83], позволяющих использовать их в клинических условиях. Однако, несмотря на технический прогресс, анализ экспериментальных рамановских спектров все еще остается сложной задачей. Причина этого в том, что спектры содержат большой объем информации обо всех химических веществах кожи. Многие компоненты кожной ткани имеют сходные рамановские спектры с перекрывающимися полосами. Кроме того, помимо различий, связанных с развитием заболевания, существуют внутригрупповые различия, связанные с сопутствующими заболеваниями и фенотипом кожи. Также важно понимать, что портативные рамановские системы регистрируют рамановский спектр в присутствии шумов (с низким соотношением сигнал/шум на уровне 3) по сравнению с системами рамановской микроскопии [5, 83]. Помимо полезного RS сигнала, необработанные («сырые») спектры также содержат сигнал флуоресценции и паразитную составляющую, связанную с влиянием оптико-

электронного тракта измерительного прибора. Поэтому предварительная обработка записанных спектров также является нетривиальной задачей. В настоящее время используется комплекс процедур предварительной обработки спектров, включая сглаживание для удаления высокочастотных помех и удаление базовой линии для исключения фоновой флуоресценции и низкочастотных помех, связанных с вкладом оптической системы [5]. Однако, следует иметь в виду, что в условиях высокого шума небольшая часть полезного сигнала также удаляется вместе с паразитными сигналами и помехами.

Традиционным методом обработки данных рамановского рассеяния является сравнение интенсивности пиков [84], когда для идентификации различий используют интенсивность нескольких характерных пиков или отношение пары характерных пиков. Такой метод подходит для небольших выборок спектральных данных с очевидными характерными пиками, существенно отличающимися для разных классов.

Для больших объемов данных часто используются различные методы статистического анализа и машинного обучения. Оптическая диагностика рака кожи улучшается с развитием алгоритмов автоматической классификации [85]. Многочисленные обзоры научных исследований в области рамановской спектроскопии показывают, что применение машинного обучения позволяет анализировать спектральные данные рамановского рассеяния в реальном времени с высокой чувствительностью и специфичностью. В качестве признаков в большинстве исследований выбираются спектральные отсчеты и полосы, соответствующие конкретным веществам в исследуемом образце [86, 87].

Среди методов машинного обучения для анализа рамановских спектров чаще всего применяются дискриминантный анализ при помощи проекции на латентные структуры (PLS-DA), метод главных компонент (PCA), метод опорных векторов (SVM), метод k-ближайших соседей (k-NN) и некоторые другие [15, 88, 89]. Метод PLS-DA является одним из наиболее распространенных подходов к анализу мультиколлинеарных спектральных данных, поскольку может обеспечить статистически надежный результат [5, 56]. Среди методов анализа спектров

рамановского рассеяния можно особенно отметить метод PCA – ортогональное линейное преобразование, которое обычно используется для уменьшения размерности путем проецирования данных на новые главные компоненты. Анализ спектров рамановского рассеяния света методом PCA успешно применен в [4], а также для визуализации и сжатия данных в [15]. Однако, несмотря на положительный опыт, этот метод имеет существенное ограничение в применении к реальным данным – нефизичность, то есть невозможность физической интерпретации компонентов, поскольку они «поглощают» дисперсию нескольких исходных переменных.

Способность выбирать важную информацию и распознавать закономерности и сходства в наборах данных делает статистические методы идеально подходящими для анализа спектральных данных [56]. В частности, эти различные статистические методы уже были использованы с целью понимания пространственного распределения биохимических компонентов в биотканях [90, 91], выявления потенциальных биомаркеров заболеваний [92], дифференциации здоровых биологических тканей от больных [93] и для определения стадий заболевания [94]. Достоинства статистических методов анализа делают их идеальными кандидатами для использования в задачах выделения информации из многомерных спектральных кривых [56].

Другим подходом, применяемым к анализу RS спектров, являются искусственные нейронные сети [56]. Чаще всего они используются в качестве классификатора образцов по заболеваниям [87, 95]. Большая часть исследований сосредоточена на применении сверточных нейронных сетей (CNN) [16]. CNN может напрямую извлекать признаки из входных данных и классифицировать наблюдаемые объекты. Исследования показывают, что точность классификации опухолей кожи с помощью CNN превосходит или находится на одном уровне с диагностической точностью квалифицированных дерматологов [17]. Также описывается успешное применение рекуррентных нейронных сетей (RNN) – одного из типов нейронных сетей с циклическими связями [95]. Преимущество RNN заключается в возможности анализа временной и семантической информации

в данных. Учитывая, что RS спектр можно рассматривать как последовательность взаимосвязанных спектральных отсчетов, RNN имеет высокий потенциал для анализа данных рамановской спектроскопии [89].

Большинство перечисленных исследований ориентировано на решение проблемы классификации, однако при этом остаются неизвестны критерии классификации [16]. Такой подход делает модели менее интерпретируемыми и неприемлемыми для клинических приложений [89].

В работе [18] для биохимической интерпретации RS спектров новообразований кожи предложено использовать их разложение по RS спектрам чистых химических веществ, однако данный метод слабо сходил ввиду сильной зависимости спектров колебаний молекул от их связей и окружения в реальной биологической ткани. Существенно лучшей сходимости удалось добиться в работе [19] за счет использования разложения по RS спектрам отдельных компонентов, зарегистрированных при микроскопическом исследовании тонких слоев кожи с точечным выделением клеточных структур.

В настоящем исследовании для интерпретации рамановских спектров кожи с низким отношением сигнал/шум предлагается использовать метод разрешения многомерных кривых (MCR). Этот метод широко используется для восстановления концентрационных профилей в хемометрике [96, 97]. В последнее время метод MCR нашел широкое применение в биологии и медицине для анализа спектральных данных, когда необходимо определить концентрации компонентов смеси по спектрам [97, 98]. Например, в работе [99] использовано разделение методом MCR при флуоресцентной визуализации. Авторы работы [100] использовали рамановскую спектроскопию и анализ MCR для обнаружения кератина и автоматической диагностики рака полости рта, а в работе [101] исследовали возможность отделения клеток рака молочной железы от нормальных эпителиальных клеток молочной железы с помощью рамановской микроскопии и анализа MCR. В [102] MCR разложение применено к спектрам рамановского рассеяния кожи, записанным с помощью рамановской микроскопии. Авторы сообщают, что можно восстановить спектры рамановского рассеяния,

соответствующие вкладу урокановой кислоты, увлажняющего фактора, меланина и т. д. В [19] образцы биологических тканей были исследованы *in situ* с помощью сканирующего рамановского микроскопа. Используя метод MCR, авторам удалось получить спектры комбинационно активных компонентов кожи, участвующих в биохимических изменениях при развитии различных новообразований. В работах [103, 104] исследователи используют метод MCR для количественного анализа фармацевтических и сельскохозяйственных образцов, а также для анализа вибрационных спектральных изображений биологического материала.

Важно упомянуть, что следует различать химический компонент, присутствующий в изучаемой системе, и главную компоненту, определяемую в методе PCA. Первый – это реальное вещество, имеющее свой спектр, а вторая – это абстрактная величина, характеризуемая вектором нагрузок. Однако между разложением PCA и задачей MCR имеется сходство. PCA позволяет оценить важнейший параметр – число химических компонентов в MCR, которое естественно равно числу главных компонент в PCA. Кроме того, PCA дает базис для целого класса методов разрешения кривых, называемых факторным анализом. Термин «базис» здесь нужно понимать буквально – матрица счетов и матрица концентраций образуют одно и то же линейное подпространство. Аналогично обстоит дело и для матриц нагрузок и чистых спектров. Это означает, что оценка матрицы чистых спектров может быть найдена как линейная комбинация векторов из матрицы нагрузок, и, соответственно, оценка матрицы концентрационных профилей может быть составлена из векторов счетов матрицы. Другими словами, векторы PCA счетов и нагрузок являются абстрактными концентрационными профилями и спектрами.

Большинство известных работ по MCR разложению рамановских спектров описывают использование установок рамановской микроскопии при анализе тонких срезов тканей кожи. Однако в клинических условиях RS спектры новообразований снимаются со значительно большего объема кожи, поэтому применение и сходимость MCR разложения для спектров, полученных в клинических условиях, требует отдельного исследования при применении

портативных установок с относительно небольшим соотношением сигнал/шум и ограниченным временем накопления сигнала, учитывая зависимости характеристик мод колебаний молекул от их связей и окружения в реальной биологической ткани.

1.3 Цифровая дерматоскопия для диагностики новообразований кожи

Визуальный осмотр области новообразования может быть дополнен дерматоскопическим анализом [105]. При дерматоскопическом осмотре врач оценивает изменение формы и пигментацию опухоли, которые трудно обнаружить невооруженным глазом [106].

Дерматоскопия – это инструментальный неинвазивный метод исследования новообразований, состоящий в визуализации эпидермиса при многократном оптическом увеличении с последующим анализом набора признаков. Дерматоскопия проводится амбулаторно, не требует от пациента никакой подготовки и практически не имеет противопоказаний. В отличие от классического визуального осмотра цифровая дерматоскопия позволяет зарегистрировать цифровое изображение исследуемого участка кожи при стандартизированном освещении и применить к изображению различные способы интеллектуального анализа. Дерматоскопию можно классифицировать в зависимости от типа используемого режима освещения (поляризованная и неполяризованная дерматоскопия), от того, контактирует ли пластина дерматоскопа с поверхностью кожи (контактная и бесконтактная дерматоскопия), а также в зависимости от типа самого дерматоскопа (ручной или видеодерматоскоп) [107].

Свет, попадающий на поверхность кожи, испытывает диффузное френелевское отражение, которое значительно ввиду того, что роговой слой кожи имеет высокий коэффициент преломления (обычно около 1,55). Для визуализации более глубоких структур кожи необходимо уменьшить эффект отражения от границы кожи с воздухом. Для этого чаще всего используют иммерсионные методы (иммерсионные дерматоскопы), выравнивающие скачок коэффициента

преломления на границе. Это достигается путем наложения стеклянной пластины (коэффициент преломления 1,52), контактирующей с роговым слоем кожи, а между пластиной и кожей вводят иммерсионную жидкость или гель (см. рисунок 1.2, слева).

В идеале коэффициент преломления иммерсионной жидкости должен быть согласован с коэффициентом преломления кожи, чтобы минимизировать отражения от границы и пропускать больше света через роговой слой [108]. Для иммерсионной дерматоскопии может использоваться неполяризованный источник света. В качестве иммерсионной жидкости можно использовать ультразвуковой гель, жидкий парафин, минеральное масло или воду. Использование прозрачного ультразвукового геля имеет преимущество, так как он не растекается. Использование минерального масла не рекомендуется из-за его канцерогенного действия [107].



Рисунок 1.2 – Виды дерматоскопии:

слева – с масляной иммерсией (неполяризованный свет),

справа – кросс-поляризованная (поляризованный свет)

Уменьшение интенсивности отраженного от границы излучения (а точнее его компенсация) при регистрации изображения также может достигаться использованием поляризованного излучения (поляризационная дерматоскопия). Для этого используются два скрещенных поляризационных фильтра. Один устанавливается в источник излучения для формирования поляризационной подсветки места исследования. Второй кросс-поляризационный фильтр

устанавливается на входе цифровой камеры. Поскольку при «зеркальном» отражении от границы раздела поляризация излучения подсветки сохраняется, то установка кросс-поляризационного фильтра гарантирует отсутствие попадания отраженного от границы света в цифровую камеру. Поскольку кожа является мутной средой, то излучение, преломленное на границе и попавшее в кожу, испытывает многократное рассеяние, теряя таким образом свою поляризацию. В процессе рассеяния свет проникает на 60-100 мкм вглубь поверхности кожи (см. рисунок 1.2, справа) [107, 108]. Соответственно, диффузно рассеянное назад излучение из слоя кожи становится деполяризованным и проходит через кросс-поляризационный фильтр, установленный на входном окне камеры.

Следует отметить, что поляризационный дерматоскоп может работать и в режиме неполяризованного излучения, что достигается путем изменения положения поляризационного фильтра на камере. Основное различие между неполяризованной и поляризованной дерматоскопией заключается в глубине визуализируемых структур. Режим неполяризованного излучения лучше подходит для осмотра новообразований, находящихся на поверхности кожи, а поляризованная дерматоскопия – для оценки более глубоких слоев [108].

В настоящий момент разработано множество моделей дерматоскопов: от простого ручного дерматоскопа без встроенной камеры или специализированных насадок для цифровых камер [109] до высокотехнологичных кросс-поляризационных [110] или иммерсионных дерматоскопов [111]. Последние позволяют получать качественные изображения высокого разрешения с выводом на экран компьютера, автоматически вычислять критерии согласно стандартным правилам ABCD [29] или 7-балльного контрольного списка [30]. Кроме того, при обследовании кожи методом дерматоскопии можно исследовать выраженность сосудистых сетей и наличие серо-белых областей с синим тоном (так называемой, бело-голубой вуали) [56, 112].

В целом, цифровой дерматоскоп – это один из наиболее часто применяемых медицинских приборов для диагностики кожных заболеваний [113]. Однако необходимо понимать, что учет всех перечисленных критериев в процессе изучения

дерматоскопического снимка вызывает сложности даже у обученного и опытного медицинского специалиста, не говоря уже о врачах общего профиля. Качество диагностики рака кожи путем визуального осмотра во многом зависит от уровня квалификации и профессионального опыта врача. Например, доля верно определенных меланом врачом общего профиля составляет около 40% [40]. Требование высокой квалификации врача приводит к тому, что качественный и детальный анализ морфологии новообразования может проводить только квалифицированный врач-онколог с многолетним опытом [114].

Анализ дерматоскопических изображений является комплексной задачей [92], включающей различные методы выделения характерных для конкретной нозологии пространственных признаков, а также методы предобработки изображений, позволяющие стандартизовать их свойства. Среди методов предварительной обработки дерматоскопических изображений можно выделить такие методы, как сегментацию и удаление со снимков шумовых структур, например, волос. Так, авторы работы [115] в своем исследовании использовали алгоритмы вейвлет-преобразования для сегментации дерматоскопических изображений. Авторы работы [116] сосредоточились на применении вариационного автоэнкодера к задаче удаления волос на дерматоскопических снимках. В работе [90] для исключения структур волос из изображений был предложен алгоритм, основанный на традиционных методах обработки изображений. Авторы сообщают, что интеграция предложенного алгоритма привела к улучшению показателей производительности архитектуры AlexNet: доля правильных ответов увеличилась на 0,9%, чувствительность на 1,4%, специфичность на 0,6% [90].

При решении задач классификации и анализа дерматоскопических изображений в качестве признаков классификации часто используют параметры цвета и яркости новообразования, а также формы, симметрии и площади, по сути, следуя известному правилу ABCD. Например, такой подход применяется в работах [117, 118, 119]. В работе [120] авторами было использовано правило ABCD для анализа границ и цвета кожных образований, а классификация велась с помощью многослойного перцептрона на основе обучения с обратным распространением

ошибки. Текстурные параметры позволяют количественно определять характеристики изображения с помощью статистического подхода, учитывающего гистограмму изображения, фильтры Габора и т.д. В работе [121] для улучшения изображения и удаления волос используются фильтр Габора и геодезические активные контуры, а для извлечения признаков использовался метод оценки ABCD.

Наилучшее распознавание поражений кожи обеспечивают методы SVM, k-NN [122] и случайный лес [13]. В исследовании [123] предложен метод извлечения признаков с использованием вейвлет-преобразования и локального двоичного шаблона, а извлеченные функции были классифицированы с использованием SVM. Авторы работы [124] представили метод обнаружения и сегментации кожных заболеваний с использованием алгоритма динамического разрезания графа и классификации с помощью наивного байесовского классификатора. Оценка предложенного подхода с использованием набора данных ISIC 2017 показала точность на уровне 91,7%, чувствительность – 70,1%, специфичность – 72,7%.

Для задач классификации могут использоваться методы компьютеризированной диагностики, реализующей схему поиска изображений по содержанию (CBIR), в которой из базы данных выбираются несколько изображений, визуальнo наиболее похожих на входное изображение. Такой подход использовали авторы работы [125], проводя диагностику меланомы поиском похожих дерматоскопических изображений в базе данных. Такой же подход применен в ряде других исследований [93, 94, 126]. Другим популярным подходом к анализу дерматоскопических изображений является подход на основе составления карты активации классов (CAM) [127, 128, 129].

Алгоритмы машинного обучения показывают хорошую производительность при работе с несколькими сотнями признаков, но с увеличением числа признаков их точность снижается [46]. Для преодоления описанных недостатков применяются методы глубокого обучения [91]. Подавляющее их большинство основано на нейронных сетях [130, 131, 132, 133*, 134*, 135, 136*]. Иногда методы глубокого обучения применяются в сочетании с традиционными классификаторами [12, 137].

Благодаря возможности извлекать сложные признаки из больших объемов данных глубокое обучение применяется в задачах классификации изображений и разделения, например, злокачественной меланомы и других новообразований кожи. Авторы статьи [115] использовали в своей работе несколько архитектур нейронных сетей, таких как AlexNet, VGG16 и др. CNN для классификации дерматоскопических изображений в своих работах так же использовали авторы статей [116, 125, 135]. Точности классификации на этапе тестирования модели превышают 70-80%.

Авторы статьи [116] используют рассеянное вейвлет-преобразование. Когда необработанные дерматоскопические изображения подаются на вход нейронной сети, а значения признаков, полученных при сегментации, выступают в качестве дополнительных входных данных для полносвязного слоя, предлагаемый метод дает точность классификации более 95% при идентификации меланомы.

Авторы работы [138] предложили систему классификации новообразований кожи с использованием CNN, применив трансферное обучение к начальной версии архитектуры GoogLeNet V3.

В статье [139] представлен классификационный подход для обнаружения кожных заболеваний, включающий предварительную обработку изображений кожи с последующим извлечением признаков с помощью MobileNet V2 и классификацией с использованием сети долгой краткосрочной памяти (LSTM). Предложенный подход достиг точности 90,21%.

Авторы работы [140] предложили новую модель классификации кожных заболеваний, основанную на слиянии DenseNet и ConvNeXt. Использование сильных сторон DenseNet и ConvNeXt позволило достичь точности 95,29% на наборе данных HAM10K.

Другая гибридная модель, которая сочетает в себе различные предварительно обученные модели глубокого обучения (VGG19, InceptionV3, ResNet50, DenseNet201 и Xception) с традиционными классификаторами машинного обучения (логистическая регрессия, SVM, решающие деревья), предложена в работе [141].

Несмотря на то, что большинство исследований используют известные топологии нейронных сетей, по-прежнему имеется большой потенциал для их оптимизации. Например, авторы работы [142] применили вейвлет-преобразование для извлечения признаков и глубокое обучение для классификации признаков с целью повышения эффективности классификации поражений кожи. Используя гибридную модель, которая объединяет глубокое трансферное обучение, CNN и градиентный бустинг (GBM), авторы исследования [143] предложили новую ансамблевую стратегию классификации меланомы, которая позволила добиться точности 99,1%, чувствительности 99,4% и специфичности 98,8%. Авторы работы [144] использовали 11 различных моделей CNN в своем исследовании для выявления рака кожи. Среди этих конфигураций, рассмотренных в исследовании, лучшей оказалась DenseNet169, показавшая точность на уровне 92,25%.

Авторы исследования [145] усовершенствовали алгоритм бинарной дискриминации ММ и меланоцитарного невуса, в результате чего CNN превзошла 136 из 157 дерматологов с разным уровнем опыта с точки зрения средней специфичности и чувствительности. Впоследствии эти же авторы сравнили диагностическую эффективность CNN с участием 144 дерматологов. В этом исследовании принимались во внимание только изображения с подтвержденной гистологией достоверностью. Результаты показали, что классификация ММ на основе CNN значительно превосходила результаты младших и сертифицированных дерматологов (82,3% против 68,9%/63,2% чувствительности и 77,9% против 58,0%/65,2% специфичности) [11].

В работе [146] предложен алгоритм, основанный на бинарной классификации акральной ММ и меланоцитарного невуса. CNN достигла средних уровней чувствительности, специфичности и точности, которые были сопоставимы с уровнями опытных дерматологов (92,6%, 71,8% и 81,9% против 96,6%, 67,0% и 81,4%), что иллюстрирует потенциал автоматизированной диагностики акрального подтипа ММ на основе CNN.

В работе [1] авторы сравнили диагностическую точность CNN с результатами, полученными 58 дерматологами: CNN значительно превосходила средние

показатели участников (при чувствительности дерматологов 88,9%: специфичность 82,5% против 75,7%).

Наиболее полное на сегодняшний день многоклассовое исследование автоматических алгоритмов классификации и сравнение с результатами классификации медицинскими специалистами представлено в работе [147]. Авторы исследования сравнили результаты, полученные с помощью 139 алгоритмов классификации, с результатами, полученными 283 сертифицированными дерматологами, 118 ординаторами-дерматологами и 83 врачами общей практики. В задаче различения ММ и шести других кожных заболеваний алгоритмы поставили в среднем 19,9 правильных диагнозов из 30, при этом люди получили в среднем 17,9 правильных диагнозов.

Авторы обзорного исследования [13] сообщают, что в большинстве работ по анализу дерматоскопических изображений используются различные типы предварительно обученных сверточных нейронных сетей (92,3% от проанализированных в обзоре [13] работ) с различными архитектурами: ImageNet, ResNet, Inception, AlexNet, MobileNet, VGG, Xception, DenseNet и GoogleNet. В некоторых исследованиях использовались собственноручно обученные нейронные сети на основе CNN (28,2%), различные комбинации CNN вместе с другими моделями – гибридные модели (12,8%), а также ансамблевые модели (10,3%) [13].

В таблице 1.3 представлены основные исследования, посвященные интеллектуальному анализу наборов изображений новообразований кожи, свободно доступных в сети Интернет: ISIC [148], HAM100K [149], ISBI [148], MED-Node [150], PH2 Database [151]. В перечисленных исследованиях точность распознавания злокачественной меланомы (доля верно диагностированных новообразований) варьируется от 66,3 до 92,1% в зависимости от используемой архитектуры нейронной сети.

Таким образом, точность распознавания новообразований кожи методами машинного обучения варьируется от исследования к исследованию и составляет от 73 до 97% точности [13]. Сверточные нейронные сети обычно выделяют определенные признаки, переходя в процессе обучения от абстрактных деталей к

Таблица 1.3 – Обзор результатов исследований по анализу наборов изображений новообразований кожи, свободно доступных в сети Интернет, с помощью нейронных сетей

Автор, год	Данные	Архитектура	Случай классификации	ROC AUC	Доля верных ответов, %
H.Haenssle, 2018 [1]	ISIC [148]	Inception v4	MM vs. PL	0,80-0,82	-
S.Kassani, 2019 [135]	ISIC 2018 [148]	AlexNet ResNet50 VGG16 VGG19 Xception	MM vs. др.	-	78,5-80,5 86,4-92,1 84,4-88,4 84,6-88,7 66,3-90,3
D.Gavrilov, 2019 [138]	ISIC [148]	Inception v3	MM vs. NE	0,96	-
R.Maron, 2019 [152]	HAM100K [149]	ResNet50	MM vs. др.	0,902	
R.Seeja, 2019 [153]	ISBI 2016 [148]	U-Net + VGG16	MM vs. PL	-	83,1
B.Gazioglu, 2021 [154]	ISBI 2017, 2018, 2019 [148]	DenseNet121 ResNet50 AlexNet VGG16	MM vs. PL	-	89,22 87,96 84,5 86,2
R.Indraswari, 2022 [155]	ISIC [148], ISBI 2016 [148], MED-Node [150], PH2 Database [151]	MobileNetV2	MM vs. PL	0,97	85

MM – злокачественная меланома, PL – пигментные новообразования, NE – невус, ROC AUC – площадь под рабочей характеристикой приемника

более конкретным, самостоятельно выявляют закономерности и зависимости между различными параметрами. Во время проведения дерматоскопического исследования визуальные методы оценки играют важную роль, поэтому высокая точность классификации на данном этапе приобретает существенную актуальность. Таким образом, качественно обученные модели нейронных сетей, способные предсказывать и принимать решения на основе большего количества данных, позволяют повысить надежность такого метода в дерматологии как визуальный осмотр. Однако в большинстве работ по нейросетевой классификации изображений кожи используются большие обучающие выборки. Такой объем

данных формировать затруднительно, поэтому интересно применение методов аугментации и дообучения классификатора на дополнительных наборах данных.

1.4 Многомерный анализ данных для диагностики заболеваний

Ключевыми шагами на пути к более эффективной диагностике рака кожи являются повышение чувствительности и специфичности [156], одним из способов достижения этого является использование мультимодальных установок. Так, например, рамановская спектроскопия в сочетании с методами визуализации расширяет возможности диагностики рака кожи за счет получения изображений с высоким разрешением наряду со спектральными данными исследуемых тканей кожи [157, 158].

Один из методов, демонстрирующий достоинства использования сразу нескольких типов данных, – это гиперспектральная визуализация (HSI). Она объединяет пространственные пиксели со спектральными данными для создания 3D-куба данных. Такое объединение пространственной и спектральной информации позволяет количественно оценивать физические и химические свойства исследуемого образца [159].

Объединение рамановской спектроскопии с другими спектроскопическими методами, например, флуоресцентной или ИК спектроскопией, также расширяет объем получаемой информации и позволяет более тщательно исследовать ткани кожи. Например, ИК спектроскопия может выявить изменения во вторичных структурах белков в тканях ММ, дополняя тем самым знания, полученные с помощью рамановской спектроскопии [160]. В исследовании сообщается о совместном использовании многоволновой рамановской и фотолюминесцентной микроспектроскопии для изучения *in vitro* образцов кожи [161]. Высокую степень точности диагностики рака кожи демонстрирует оптическая биопсия, в которой используются как рамановская спектроскопия, так и аутофлуоресценция [70, 162]. Портативная недорогая спектроскопическая установка, объединяющая анализ спектров рамановского рассеяния света и аутофлуоресценции, была опробована на

когорте из 56 пациентов (19 ММ, 18 ВСС и 19 доброкачественных новообразований). Точность классификации новообразований кожи (дифференциация ММ от других новообразований, ВСС от других новообразований, ММ от ВСС, ММ от доброкачественных новообразований) достигла диапазона точности от 89 до 100% [56, 83]. Последующее исследование, охватывающее 617 случаев новообразований кожи (615 пациентов, 70 ММ, 122 ВСС, 12 SCC и 413 доброкачественных опухолей), продемонстрировало ROC AUC 0,75 (0,71-0,79, 95% CI) для случая классификации злокачественных и доброкачественных опухолей, 0,69 (0,63-0,76, 95% CI) для случая ММ и пигментных опухолей) и 0,81 (0,74-0,87, 95% CI) для случая классификации ММ и СК [5].

Интересна интеграция спектрального анализа с изучением уникальных характеристик пациентов. Авторы работы [163] сообщают, что благодаря объединению спектральных данных с тремя факторами риска удалось повысить эффективность распознавания 70 ММ и 283 доброкачественных пигментных новообразований.

В работе [164] представлен гибридный мультимодальный прибор, объединяющий спектроскопию рамановского рассеяния света и лазерно-искровую эмиссионную спектрометрию. В результате применения прибора была повышена точность распознавания и классификации ММ и здоровых тканей, фиксированных формалином и залитых в парафин [164].

Гибридные системы, объединяющие рамановскую спектроскопию с другими диагностическими методами, такими как ОКТ или высокочастотный ультразвук, обеспечивают получение как морфологической, так и биохимической информации, что позволяет проводить комплексную оценку поражений кожи [88]. Комбинированные модули, объединяющие ОКТ с клеточным разрешением и рамановскую спектроскопию в ближнем инфракрасном диапазоне, были разработаны для дифференциации различных раковых клеток кожи от здоровых клеток несколькими научными группами [165, 166, 167]. Причина такой популярности комбинации рамановской спектроскопии и ОКТ заключается в

следующем. ОКТ позволяет получать изображения с высоким разрешением, по которым можно выявить различия в форме, плотности клеток и строении ткани между раковыми и здоровыми клетками. Рамановская спектроскопия дополняет данные информацией о химическом составе клеток. Кроме того, рамановская спектроскопия идентифицирует специфические клеточные биомаркеры, выявляет метаболические изменения и изменения в биомолекулах, таких как ДНК, РНК и белки, что в совокупности способствует комплексной оценке клеточных характеристик. Другими словами, этот интегративный подход использует возможности обоих методов для повышения диагностической точности при различении клеток рака кожи и здоровых клеток [168, 169].

В работе [170] для быстрого скрининга рака кожи *in vivo* авторы применяют тримодальную систему, сочетающую в себе ОКТ, оптоакустические методы для точного определения глубины опухоли, и методы рамановской спектроскопии, которая позволяет идентифицировать тип поражения, еще больше расширяя диагностические возможности системы.

Таким образом, мультимодальные диагностические подходы позволяют получить более целостное представление о тканях кожи, помогая принимать более обоснованные клинические решения [171]. Ожидается, что в ближайшее десятилетие будут появляться подходы к анализу биологических тканей, основанные на методах визуализации в сочетании с рамановской спектроскопией, позволяющие получать изображения высокого разрешения наряду с химическими данными. Эта синергия позволит проводить глубокий анализ злокачественных и доброкачественных новообразований кожи, что значительно поможет раннему выявлению и диагностике рака кожи [171].

Мультимодальный диагностический подход на основе алгоритмов и методов анализа спектральных (рамановская спектроскопия) и пространственных (цифровая дерматоскопия) признаков с одновременной биохимической интерпретацией результата исследования, позволит расширить набор анализируемых признаков, что положительно отразится на эффективности распознавания новообразований кожи.

1.5 Постановка задачи диссертационного исследования

Представленный обзор литературы свидетельствует об актуальности задач, связанных с разработкой новых подходов к проведению неинвазивной оптической диагностики кожи с целью выявления социально значимых заболеваний человека. Результаты многочисленных исследований показывают, что наиболее широко применяемым методом диагностики кожных патологий является цифровая дерматоскопия, которая позволяет успешно детектировать рак кожи за счет выделения специфических пространственных признаков различных типов новообразований. Среди спектральных методов оптической биопсии наиболее мощным средством неинвазивного анализа изменения химического состава биотканей в процессе развития новообразований является рамановская спектроскопия.

Несмотря на успешное развитие перечисленных оптических методов, каждый из них, тем не менее, не превосходит результаты гистологического исследования, обеспечивая несколько более низкие значения чувствительности и специфичности [45]. Также следует отметить насущную потребность наряду с предсказанием иметь качественную и количественную интерпретацию регистрируемых данных, получение которой затруднительно без информации о компонентном составе новообразования.

Для улучшения качества диагностики целесообразно использовать комбинацию нескольких оптических методов. Возникает необходимость разработки нового мультимодального физически интерпретируемого подхода к анализу спектральных данных рамановского рассеяния и дерматоскопических изображений. Для достижения поставленной цели представляется необходимым решение следующих задач:

1. Исследовать способы восстановления и интерпретации компонентного состава образцов кожной ткани при многомерном разрешении спектральных данных рамановского рассеяния в ближнем ИК диапазоне при возбуждении лазерным излучением на длине волны 785 нм;

2. Разработать метод и программно-алгоритмическое обеспечение распознавания рамановских спектров кожи при решении классификационных задач определения типа новообразования с учетом выделения спектральных признаков с интерпретацией компонентного состава исследуемого образца;

3. Разработать метод интеллектуального распознавания дерматоскопических изображений при решении классификационных задач определения типа новообразований методами машинного обучения;

4. Разработать метод идентификации типа новообразования на основе мультимодального совместного анализа рамановских спектров и дерматоскопических изображений.

Перечисленные направления исследований определяют содержание и структуру данной диссертационной работы.

2 ДВУХКАНАЛЬНАЯ ЭКСПЕРИМЕНТАЛЬНАЯ СИСТЕМА И ОПТИЧЕСКИЕ ДАННЫЕ

2.1 Система регистрации рамановских спектров кожи и дерматоскопических изображений

Все рамановские спектры и дерматоскопические изображения новообразований кожи, используемые в данной диссертации, получены с помощью экспериментальной системы, схема которой показана на рисунке 2.1.

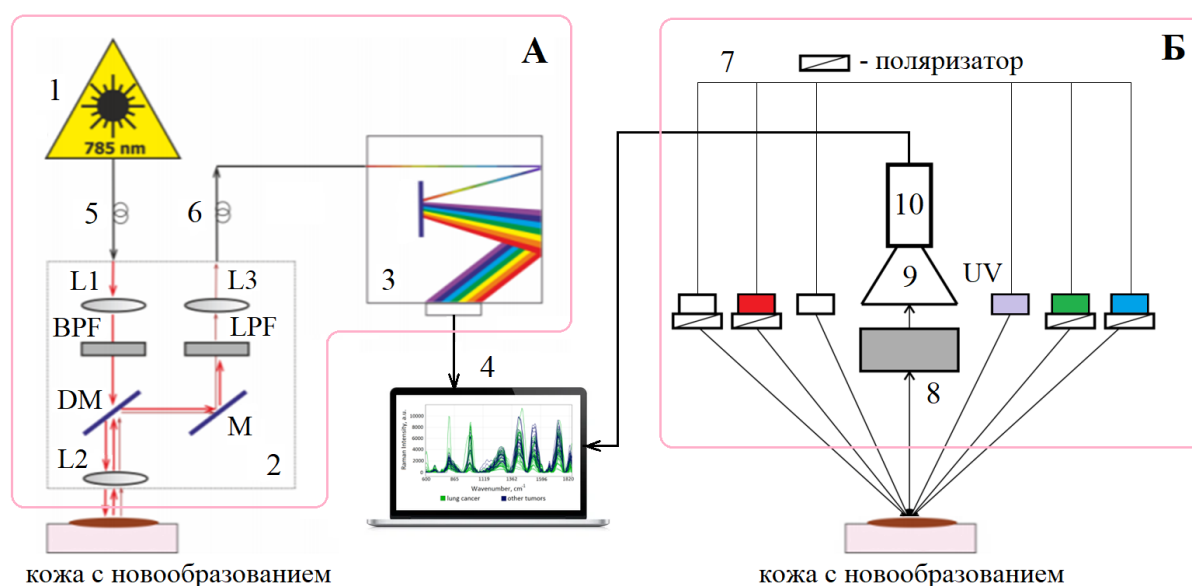


Рисунок 2.1 – Структурная схема двухканальной экспериментальной системы для *in vivo* исследований кожи человека. А – Портативная рамановская установка (1 – лазер, 2 – оптоволоконный зонд, 3 – спектрометр, 4 – ПК, 5 – волокно возбуждения, 6 – собирающее волокно, L1, L2, L3 – линзы, BPF – полосовой фильтр, DM – дихроичное зеркало, M – зеркало, LPF – широкополосный фильтр) .
Б – Мультиспектральный цифровой дерматоскоп (7 – цифровой блок светодиодных излучателей, 8 – объектив, 9 – камера, 10 – микроконтроллер, UV – ультрафиолетовый светодиод)

Канал регистрации рамановских спектров А представляет собой портативную оптоволоконную установку с интегрированными рамановским спектрометром, стабилизированным лазером и оптическим зондом, которая позволяет одновременно регистрировать спектры рамановского рассеяния и автофлуоресценцию в ближней ИК области [83]. Длина волны лазерного излучения в системах рамановской спектроскопии должна попадать в «терапевтическое окно прозрачности» кожи (диапазон от 700 до 1300 нм), в котором поглощение биологических тканей относительно невелико, что обеспечивает возможность регистрации спектров *in vivo* при относительно небольшой плотности мощности лазерного излучения на поверхности кожи с соблюдением мер лазерной безопасности. В качестве источника излучения в портативной рамановской установке используется термостабилизированный диодный лазерный модуль LML-785.0RB-04 с центральной длиной волны 785 нм (обеспечивает оптимальное соотношение между интенсивностью рамановского сигнала и флуоресценцией) с шириной спектра 0,1 нм и стабильностью центральной линии генерации ± 10 пм [56].

Портативная рамановская установка работает следующим образом. Лазерное излучение модуля 1 по оптоволокну 5 вводится в оптоволоконный зонд 2 с помощью коллимирующей линзы L1 и направляется на полосовой фильтр ВРФ, предназначенный для фильтрации слабоинтенсивного рамановского рассеяния, которое может возбуждаться в кварцевом волокне 5. Это гарантирует отсутствие паразитного сигнала при облучении кожи. После фильтра ВРФ излучение проходит через дихроичное зеркало DM и фокусируется линзой L2 на исследуемом образце. Эта же линза L2 собирает диффузно рассеянное назад излучение из зоны облучения, которое содержит рэлеевское рассеяние, автофлуоресценцию и рамановское рассеяние, возбужденные лазером в исследуемой кожной ткани. Собранное излучение через дихроичное зеркало DM и зеркало М перенаправляется в приемный канал, где проходит широкополосный фильтр LPF и с помощью согласующей линзы L3 вводится в волновод 6 и подается в спектрометр 3. Широкополосный фильтр LPF служит для отрезания диффузно-рассеянного

возбуждающего лазерного излучения, обеспечивая тем самым на входе спектрометра 3 только излучение в области спектральной полосы выше 790 нм [56]. Таким образом, оптоволоконный зонд 2 выполняет роль приемопередатчика, фокусирующего возбуждающее лазерное излучение на поверхности кожи или иного исследуемого объекта, а также собирающего и фильтрующего рассеянное излучение, обеспечивая тем самым регистрацию с помощью спектрометра 3 только автофлуоресценции и рамановского рассеяния. Спектрометр 3 раскладывает рассеянное излучение в спектр и передает данные на персональный компьютер 4 [56].

Аналогичные рамановские установки использовались, например, для анализа кожи человека при мониторинге диабета [172, 173]. В отличие от лабораторных систем [57, 69, 174] с высокой степенью охлаждения детектора (от -60 до -120 °С), в портативной рамановской установке используется спектрометр QE65 Pro от Ocean Optics Inc. с ПЗС-детектором, работающим при -15°С [83]. Это значительно снижает габариты, вес и стоимость системы. Однако для обеспечения приемлемого соотношения сигнал/шум необходимо накопление сигнала, и необходимое время регистрации спектров увеличивается с 0,1-1 сек до 30-60 сек [56].

Канал регистрации дерматоскопических изображений представляет собой мультиспектральный цифровой дерматоскоп Б, который состоит из цифровой блока светодиодных излучателей 7, объектива 8, камеры 9 и микроконтроллера 10 [175, 176]. В качестве цифровой камеры используется Basler acA1920-25uc (RGB, разрешение ~ 13 мкм/пиксель), позволяющая регистрировать изображения в поляризованной видимом свете (1920×1080, 48 бит, 96 точек на дюйм, формат .tiff). Характеристики объектива позволяют захватывать изображение от 10 мм по каждой из осей.

В качестве подсветки в блоке излучателей 7 используется несколько светодиодов:

- два ультрафиолетовых LEUVA77V20RV00 (365 нм, 1 Вт, полуширина 9 нм) для возбуждения автофлуоресценции кожи в видимом диапазоне спектра;

- четыре белых FM-5630WDS-460W-R80 (39 лм, 4000 К), причем два светодиода с поляризованной пленкой для визуализации более глубоких слоев кожи и два неполяризованных;

- два синих CREEXREBLU-L1-0000-00K01 (30 лм, 465-485 нм);
- два зеленых CREEXRCGRN-L1-0000-00N01 (52 лм, 520-535 нм);
- два красных CREEXPCRED-L1-0000-00301 (46 лм, 620-630 нм).

Благодаря такому блоку излучателей цифровой дерматоскоп способен работать в нескольких режимах: с поляризованной подсветкой, с неполяризованной подсветкой и в режиме съемки серии мультиспектральных изображений. В режиме неполяризованной подсветки 90% света отражается зеркально, что позволяет оценить поверхностный слой кожи, то есть рисунок поверхности эпидермиса. Это может быть полезно при диагностике злокачественной меланомы. В режиме поляризованной подсветки поляризационный фильтр отсекает зеркальную компоненту, что позволяет получить изображение более глубокого слоя кожи и анализировать особенности кожи в глубине эпидермиса (до 1 мм). Ультрафиолетовое излучение используется для оценки флуоресценции кожи (при развитии новообразований может меняться концентрация флуоресцирующих веществ). Красная, зеленая и синяя подсветки позволяют построить карты распределения концентраций основных хромофоров кожи в исследуемом участке кожи [175, 176].

2.2 Описание выборки оптических данных

Исследование новообразований кожи *in vivo* проводилось в Самарском областном клиническом онкологическом диспансере с мая 2017 года по декабрь 2019 года. В исследовании участвовали пациенты, обследованные врачами общей практики в поликлиниках Поволжья и направленные в Самарский областной клинический онкологический диспансер с подозрением на рак кожи [56].

Все пациенты были не моложе 18 лет. Перед *in vivo* исследованием у всех пациентов было получено информированное согласие. Протоколы диагностики

тканей кожи *in vivo* одобрены этическим комитетом Самарского государственного медицинского университета (Самарская область, Самара, Россия, протокол № 132, 29 мая 2013 г., клинические исследования подпадают под Кодекс этики врача России, утвержденного на 4-й конференции Российской медицинской ассоциации и в рамках Хельсинкской декларации Всемирной медицинской ассоциации).

Пациентам проводились спектральные измерения рамановского рассеяния и дерматоскопические снимки очагов поражения кожи. Поражения не рассматривались для включения в выборку, если они были менее 3 мм в поперечном направлении (меньше диаметра зондирующего луча канала регистрации рамановского рассеяния), располагались на участке тела, который был недоступен для рамановского зонда или дерматоскопа, инфицирован или ранее подвергнулся биопсии, удалению или был травмирован. Также для каждого пациента проводились измерения рамановского рассеяния с области здоровой кожи [56].

Регистрация рамановского рассеяния и автофлуоресценции проводилась со спектральным разрешением 0,2 нм в диапазоне от 800 до 1000 нм. Все измерения спектров *in vivo* проводились при времени накопления сигнала 60 с. Мощность лазера на выходе оптоволоконного зонда составляла примерно 30 мВт. Рассеянное излучение собиралось с верхнего слоя кожи толщиной от 1 до 2 мм, наибольший вклад в рамановский сигнал дает преимущественно дерма (приблизительно 70%) [176, 177]. Оптический детектор был расположен непосредственно над образцом ткани на расстоянии от 7 до 8 мм. Диаметр луча зондирующего излучения на исследуемой ткани составлял примерно 3 мм. Это соответствует плотности мощности лазерного излучения 1,56 Вт/см² на коже. Данная плотность мощности лазерного излучения не превышает максимально допустимый предел воздействия согласно межгосударственному стандарту [178] (аналог ANSI) 1,63 Вт/см² и не вызывает повреждений кожи или дискомфорта у пациентов [56].

По причине очень низкой эффективности используемого детектора в области 920-1000 нм анализ спектральных данных проводился в диапазоне от 803 до 920 нм, что соответствует рамановскому сдвигу от 285 до 1874 см⁻¹. Каждый из зарегистрированных спектров представляет собой дискретный набор из 529

параметров (предикторов) – спектральных отсчетов на заданной длине волны заданной амплитуды, измеряемой в относительных единицах интенсивности [56].

Для анализа использовались как необработанные («сырые») спектры в диапазоне 837-920 нм, что соответствует 792-1874 см⁻¹, так и предварительно обработанные спектры, в которых исключалась автофлуоресценция и выделялся вклад только рамановского рассеяния. Высокий уровень флуоресценции и низкое отношения сигнал/шум препятствовало уверенному выделению пиков рамановского рассеяния в области 300-1200 см⁻¹ [83]. Особенно сильно этот эффект проявляется при исследовании меланоцитарных новообразований, где автофлуоресценция меланина перекрывает слабый сигнал рамановского рассеяния [7]. Ввиду быстрого спада интенсивности флуоресценции с увеличением длины волны во всех экспериментальных случаях удавалось уверенно регистрировать и выделять пики рамановского рассеяния в диапазоне 1200-1800 см⁻¹. Поэтому для обеспечения общности результатов анализ новообразований проводился в более узком спектральном диапазоне от 860 до 920 нм, что соответствует 1114-1874 см⁻¹. Кроме выделения спектрального диапазона, предобработка спектров включала удаление базовой линии (полином 6-ой степени) и сглаживание фильтром Савицкого-Голея (производная 0-го порядка, ширина окна – 15). Пример экспериментальных рамановских спектров кожи до (содержат сигнал автофлуоресценции) и после предварительной обработки показан на рисунке 2.2.

Регистрация дерматоскопических изображений проводилась в режиме неполяризованной подсветки. Дерматоскоп располагался вплотную к образцу ткани перпендикулярно поверхности кожи. Примеры получаемых дерматоскопических изображений представлены на рисунке 2.3.

Каждого пациента осматривал онколог. Все новообразования, анализируемые в данном исследовании, оценивались врачами-терапевтами и онкологами как подозрительные на предмет наличия злокачественного новообразования, в связи с чем для каждого из исследованных новообразований была взята эксцизионная биопсия. Диагноз был установлен с помощью гистопатологического анализа [56].

В данное исследование включено 602 пациента. Подробное распределение данных по типам новообразований представлено в таблице 2.1.

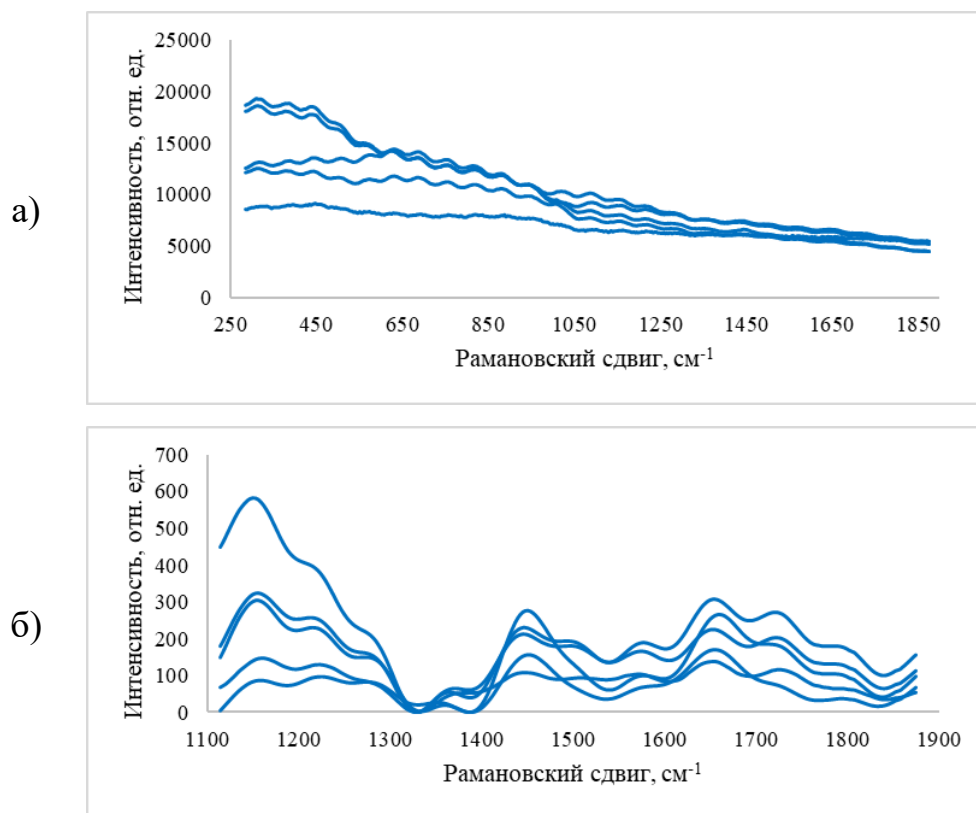


Рисунок 2.2 – Пример *in vivo* спектров рамановского рассеяния здоровой кожи до (а) и после (б) предварительной обработки

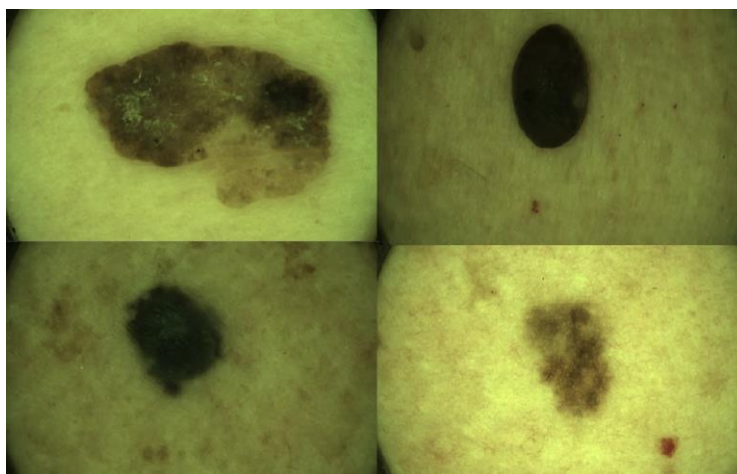


Рисунок 2.3 – Примеры дерматоскопических изображений злокачественной меланомы, полученных в режиме неполяризованной подсветки

Таблица 2.1 – Сводка новообразований, исследованных с помощью рамановской спектроскопии и цифровой дерматоскопии

Истинный диагноз		Количество рамановских спектров	Количество дерматоскопических изображений СОКОД	Количество дерматоскопических изображений НАМ10К [153]
Здоровая кожа		602	-	-
Злокачественные новообразования	BCC	120	-	-
	SCC	12	-	-
	MM	65	65	1113
Доброкачественные новообразования	SK	111	107	1099
	NE	167	166	6705
	DF	26	-	-
	PP	61	-	-
	HE	40	-	-

Набор спектральных данных рамановского рассеяния (RS спектров) содержал 602 спектра здоровой кожи, 197 злокачественных новообразований (65 злокачественных меланом (MM), 120 базальноклеточных карцином (BCC), 12 плоскоклеточных карцином (SCC)) и 405 доброкачественных новообразований (26 дерматофибром (DF), 61 папиллому (PP), 40 гемангиом (HE), 111 образцов себорейного кератоза (SK) и 167 невусов (NE) (всех типов)). Часть из перечисленных новообразований была одновременно исследована с помощью опытного образца цифрового дерматоскопа. Окончательный набор дерматоскопических изображений содержал 65 опухолей злокачественной меланомы (MM), 166 опухолей невусов (NE) (всех типов) и 107 образцов себорейного кератоза (SK). Данный набор изображений будем называть СОКОД. Размеры кожных новообразований варьировались от 0,3 до 5 см [56].

Наряду с набором СОКОД для анализа дерматоскопических изображений использовался свободно распространяемый набор НАМ10К, полученный в работах [179, 180]. Необходимость использования дополнительного набора изображений связана с применением для классификации изображений нейронных сетей (см. четвертую главу), качество обучения которых зависит от объема обучающей выборки. В связи с этим было принято решение о предварительной тренировки нейронной сети на дополнительном наборе данных НАМ10К.

Набор данных НАМ10К состоит из 10015 дерматоскопических изображений (224×224, 24 бита, 96 точек на дюйм, формат .jpeg), среди которых содержатся все важные диагностические категории пигментных поражений: актинический кератоз и внутриэпителиальная карцинома/болезнь Боуэна, базально-клеточная карцинома, доброкачественные кератозоподобные поражения (солнечные лентиго/себорейный кератоз и кератозы, подобные плоскому лишаяю), дерматофиброма, меланома, меланоцитарные невусы и сосудистые поражения (ангиомы, ангиокератомы, пиогенные гранулемы и кровоизлияния). Эти изображения были собраны в течение 20 лет на кафедре дерматологии Медицинского университета Вены, Австрия [179] и в клинике рака кожи Клиффа Розендаля в Квинсленде, Австралия [180]. Диагноз для 50% новообразований был подтвержден гистологическим исследованием. Для остальных случаев заключение было получено после проведения обследования экспертом либо с помощью конфокальной микроскопии. Из набора НАМ10К была выделена только подгруппа изображений новообразований, нозология которых совпадала с нозологиями новообразований, исследованных с помощью цифрового дерматоскопа. Для данной подгруппы в настоящей работе сохранено наименование набор НАМ10К. Он включает 1113 изображений злокачественной меланомы (MM), 6705 изображений невусов (NE) (всех типов) и 1099 изображений себорейного кератоза (SK).

3 РАЗРАБОТКА МЕТОДА АНАЛИЗА РАМАНОВСКИХ СПЕКТРОВ НОВООБРАЗОВАНИЙ КОЖИ С ВЫДЕЛЕНИЕМ КОМПОНЕНТНОГО СОСТАВА

В данной главе описана разработка физически интерпретируемого способа анализа данных оптической биопсии, полученных с использованием портативного спектрометра, для *in vivo* диагностики и определения типа новообразований кожи в клинической практике. Для анализа спектральных данных с выделением биофизических компонент тканей предложено использовать метод MCR, что позволит явным образом выделить закономерности в изменениях биохимического состава кожи в процессе развития заболевания только на основании измерений RS спектров. Выявление данных закономерностей и их использование для распознавание различных новообразований целесообразно вести методами машинного обучения, эффективность применения которых уже продемонстрирована при анализе полных рамановских спектров.

3.1 Применение метода разрешения многомерных кривых для компонентного анализа рамановских спектров кожи

Метод разрешения многомерных кривых (MCR) традиционно используется для решения проблемы разрешения кривых в спектроскопии [97, 98]. Следует также отметить, что метод MCR может быть чувствителен к шуму, который препятствуют корректному разделению компонентов с близко расположенными пиками. Поскольку RS спектры регистрировались портативной рамановской установкой с соотношением сигнал/шум, равным 3, то возможность применения метода MCR требует отдельного исследования. С этой целью автором была изучена возможность применения метода MCR для анализа RS спектров смесей протеиногенных аминокислот [181*, 182*] и влияние шума в спектрах на результат их анализа [183*, 184*], а также была опробована возможность применения метода MCR к рамановским спектрам кожи [185*, 186*, 187*]. Было установлено, что

уверенное разделение нескольких спектрально подобных (перекрывающихся) компонентов сложной смеси методом MCR, восстановление их спектров и оценка концентрации аминокислот в смеси достигается даже когда RS спектры имеют низкое отношение сигнал/шум (от 3 до 5) [183*]. Это дает основания полагать, что метод MCR принципиально применим и для анализа спектральных данных оптической биопсии.

Предположим, что каждая отдельная химическая компонента вносит аддитивный вклад в результирующий экспериментально регистрируемый спектр кожи. Возможность аддитивного разложения подтверждается исследованиями L.Silveira [18] и X.Feng [19], в которых они сравнивали экспериментальный спектр с синтезированным спектром химически чистых компонент. В результате данных исследований получено удовлетворительное согласие экспериментальных и синтезированных спектров, а разница в основном была обусловлена тем, что компоненты кожи имеют более сложные спектры по сравнению с чистыми химическими веществами за счет внутренних связей. Спектр каждого компонента кожи может быть представлен как вектор-столбец s_i размерностью J , которая определяется количеством спектральных отсчетов (волновых чисел) в зарегистрированных спектрах. Преобразование спектральной длины волны и волнового числа сдвига в рамановском спектре выглядит следующим образом:

$$\Delta\tilde{\nu} = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1} \right) \times 10^7, \quad (3.1)$$

где $\Delta\tilde{\nu}$ – рамановский сдвиг в см^{-1} , λ_0 – длина волны возбуждения в нм, λ_1 – длина волны рамановского спектра в нм.

Тогда в предположении применимости принципа аддитивности каждый экспериментально зарегистрированный спектр кожи будет просто линейной комбинацией спектров s_i :

$$d = c_1 s_1^T + c_2 s_2^T + \dots + c_A s_A^T, \quad (3.2)$$

где d – вектор спектральных значений кожи размерности J , A – количество компонент, дающих вклад в спектр, c_1, c_2, \dots, c_A – концентрации компонентов кожи, а s_1, s_2, \dots, s_A – их спектры.

Уравнение (3.2) можно записать в более компактной форме:

$$\mathbf{D}_{(i \times j)} = \mathbf{C}_{(i \times k)} \mathbf{S}_{(k \times j)}^T, \quad (3.3)$$

где $\mathbf{D}_{(i \times j)}$ – матрица экспериментально зарегистрированных спектров (в строках i -ые спектральные отсчеты для j -го эксперимента), $\mathbf{C}_{(i \times k)}$ – матрица относительных концентраций компонентов (строки содержат концентрационный профиль для каждого эксперимента), а $\mathbf{S}_{(k \times j)}$ – матрица спектров компонентов (восстанавливаемые спектры компонентов расположены в столбцах матрицы). Причем, в отличие от подхода авторов работ [18, 19], матрица спектров \mathbf{S} заранее не известна, и ее следует восстанавливать одновременно с матрицей концентраций по экспериментально зарегистрированному спектру кожи \mathbf{D} . Это нетривиальная задача, поскольку приведенное выше выражение (3.3) не имеет однозначного решения, и при восстановлении спектров и концентраций компонентов будет неизменно возникать ошибка. Поэтому уравнение (3.3) следует переписать в виде:

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E}, \quad (3.4)$$

где \mathbf{E} – матрица ошибок [97], которая в общем случае определяется экспериментальной ошибкой восстановления и наличием шумов.

Основное преимущество решения уравнения (3.4) методом MCR заключается в восстановлении компонентного состава исследуемого новообразования на основании экспериментально зарегистрированного спектра, что позволяет надеяться на возможность классификации и определения типа новообразования. Однако, сходимость решения (минимизация матрицы ошибок \mathbf{E}) уравнения (3.4) является необходимым, но далеко не достаточным условием возможности физической интерпретации компонентов. Получающиеся в результате MCR анализа спектры компонентов, по сути, напрямую связаны с реальными химическими веществами, входящими в исследуемое новообразование. Однако одним из ограничений метода MCR является сложность разделения веществ со схожими рамановскими спектрами, когда вещество не содержит уникальных рамановских пиков в спектре [97]. Следовательно, прежде чем проводить классификацию новообразований на основании восстановленных

концентрационных профилей необходимо идентифицировать каждую выделенную спектральную компоненту и соотнести ее с определенным веществом.

Для разделения рамановских спектров с помощью метода MCR использован графический интерфейс MCR-ALS GUI 2.0 [188], где в качестве оценки неизвестных концентраций и значений интенсивности спектров компонентов используется регрессионная модель с применением метода чередующихся наименьших квадратов (ALS) [189]. Это простой и одновременно очень мощный метод решения задачи MCR. В отличие от большинства других методов решения задачи MCR, ALS не опирается на разложение PCA. Вместо этого он использует принцип наименьших квадратов, последовательно на каждом шаге корректируя матрицы D и S для минимизации ошибки E . Алгоритм метода представлен на рисунке 3.1.

Для решения уравнения (3.4) необходимо задать количество членов разложения A , которых будет достаточно для корректного восстановления компонентного профиля новообразований. Определение величины A можно осуществить, оценивая ранг матрицы D , при которой достигается сходимость разложения.

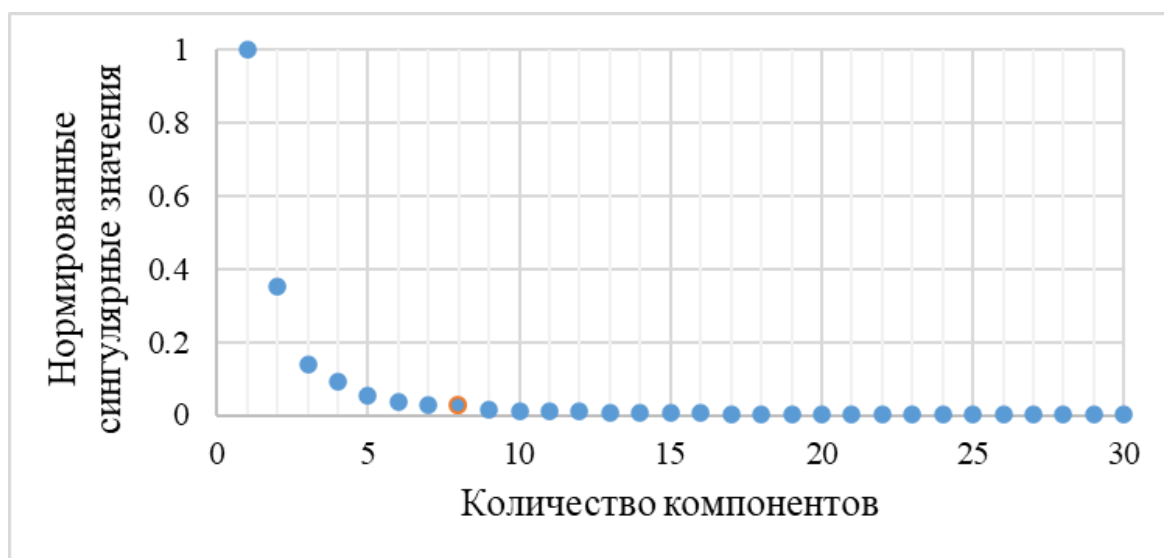
Для этого воспользуемся методом сингулярного разложения (SVD) [190], которое сводится к нахождению диагональной матрицы Σ , удовлетворяющей соотношению

$$D = U\Sigma V^T \quad (3.5)$$

где U и V – ортогональные унитарные матрицы ранга $m \times m$ и $n \times n$ соответственно, n – количество отсчетов в каждом экспериментальном спектре, m – число анализируемых экспериментов, а $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$, $k = \min\{m, n\}$ и $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq 0$. В матрице Σ можно выделить первые r строк и столбцов, которые фактически определяют количество компонент, вносящих наибольший вклад в разложение (3.5). Реконструируя исходную матрицу спектров с использованием меньшего объема компонент и сравнивая ее с исходной, можно оценить сходимость метода (см. рисунок 3.2).



Рисунок 3.1 – Схема алгоритма компонентного разложения RS спектров



№	1	2	3	4	5	6	7	8
Значение	1,000	0,353	0,140	0,090	0,053	0,035	0,029	0,026
№	9	10	11	12	13	14	15	Сумма 16-30
Значение	0,013	0,011	0,010	0,009	0,007	0,006	0,005	0,029

Рисунок 3.2 – График сингулярных значений набора рамановских спектров здоровой кожи

Видно, что первые восемь компонентов описывают не менее 95% информации о компонентном составе кожи на основании спектрального анализа (сумма первых восьми отсчетов равна более 95% суммы всех 30 отсчетов). Данный вывод согласуется с выводами авторов работы [19], которые показали, что при использовании детерминированного спектрального набора химически чистых веществ удастся уверенно использовать для анализа только восемь компонент. Следовательно, при применении метода MCR для восстановления концентрационного профиля требуется использовать не менее восьми компонент. Однако данный вывод нуждается в дополнительном исследовании, так как необходимо обеспечить не только сходимость метода разложения, но и соответствие восстановленных спектров компонент тому или иному химическому веществу или группе веществ.

Для оценки и задания начальных значений спектров и концентраций компонентов использовался алгоритм SIMPLISMA [191]. Далее в соответствии с алгоритмом, представленным на рисунке 3.1, на каждой итерации вычислялась новая оценка матриц C и S с использованием метода ALS:

$$C = DS(S^T S)^{-1}, \quad (3.6)$$

$$S^T = (C^T C)^{-1} C^T D. \quad (3.7)$$

Дополнительно на каждой итерации проводилась проверка ограничений, обеспечивающих физичность восстанавливаемых параметров: спектры компонентов и профили концентраций должны быть неотрицательны. В случае обнаружения отрицательных значений они заменялись нулями, а итерационная процедура (3.6) или (3.7) повторялась. Реконструированная матрица D^* , полученная на основании оценок концентраций (3.6) и спектров компонентов (3.7), сравнивалась с матрицей экспериментально зарегистрированных спектров D . Итерационная оптимизация продолжалась до достижения критерия сходимости:

$$\frac{\sqrt{\sum_{i,j=1}^{M,N} (D_{k+1} - D_k)^2}}{\sqrt{\sum_{i,j=1}^{M,N} D_k^2}} < 0,05, \quad (3.8)$$

где D_k – реконструированная матрица $M \times N$ на k -ой итерации, $k = 1, \dots, \infty$.

Другими словами, оптимизация останавливалась, когда разброс между двумя последовательными итерациями становился меньше порогового значения, равного 0,05%. При выбранном значении порога для всех исследованных случаев сходимость обеспечивалась при числе итераций, не превосходящем 70.

Помимо ограничения неотрицательности в каждой итерации используется условие замкнутости набора спектральных компонент, приводящее к необходимости нормировки спектров компонент на максимальное значение с сохранением суммарного спектрального профиля за счет балансировки профиля концентраций путем умножения на максимальное значение спектральной интенсивности соответствующей компоненты. Каждая составляющая восстанавливаемого таким образом профиля концентраций пропорциональна

объемной концентрации соответствующего компонента с точностью до постоянного коэффициента, не зависящего от состава.

3.2 Компонентный анализ рамановских спектров

Известно, что новообразования формируются из патологического разрастания клеток кожи, то есть характерный набор спектральных компонент новообразования не должен значительно отличаться от аналогичного набора для здоровой кожи. Разница должна проявляться преимущественно в компонентном профиле. В связи с этим для восстановления спектрального набора компонент в процессе итеративной оптимизации MCR использовались только спектры здоровой кожи.

3.2.1 Разделение экспериментальных спектров

Полный экспериментальный спектр (без предобработки) содержит как вклад флуоресценции, так и пики рамановского рассеяния. Излучение каждого вида несет качественно разную информацию о компонентном составе исследуемой биологической ткани. Более того, если для различных флуорофоров кожи характерны широкие перекрывающиеся полосы флуоресценции, то рамановское рассеяние имеет узкие пики, то есть имеет место разномасштабная зависимость от частоты для каждого из типов излучений.

Был проведен MCR анализ полных экспериментальных спектров здоровой кожи, в результате которого было выделено девять компонентов. Восстановленные нормированные спектры компонентов кожи (в порядке их извлечения в результате MCR) представлены на рисунке 3.3. Видно, что наибольший вклад в разложение вносят компоненты 2, 3, 4, 5, 6, для которых наблюдается характерный для автофлуоресценции монотонный спад интенсивности с ростом величины рамановского сдвига. Это ожидаемо ввиду значительно более высокой интенсивности автофлуоресценции по сравнению с интенсивностью рамановского рассеяния. Компонент 9 может быть отождествлен с шумами измерительной системы, а пики рамановского рассеяния наблюдаются только для компонент 7 и 8. Причем можно идентифицировать (см. таблицу 3.1) только самые сильные

колебательные моды белков (деформационные С-Н колебания 1318 см^{-1} и валентные С-О колебания) и кератина (изгибные С-Н колебания 1450 см^{-1}) [18, 192]. Белки являются основными составляющими верхних слоев кожной ткани, поэтому неудивительно, что они вносят такой большой вклад в спектр кожи и их можно различить с помощью MCR даже в присутствии флуоресценции.

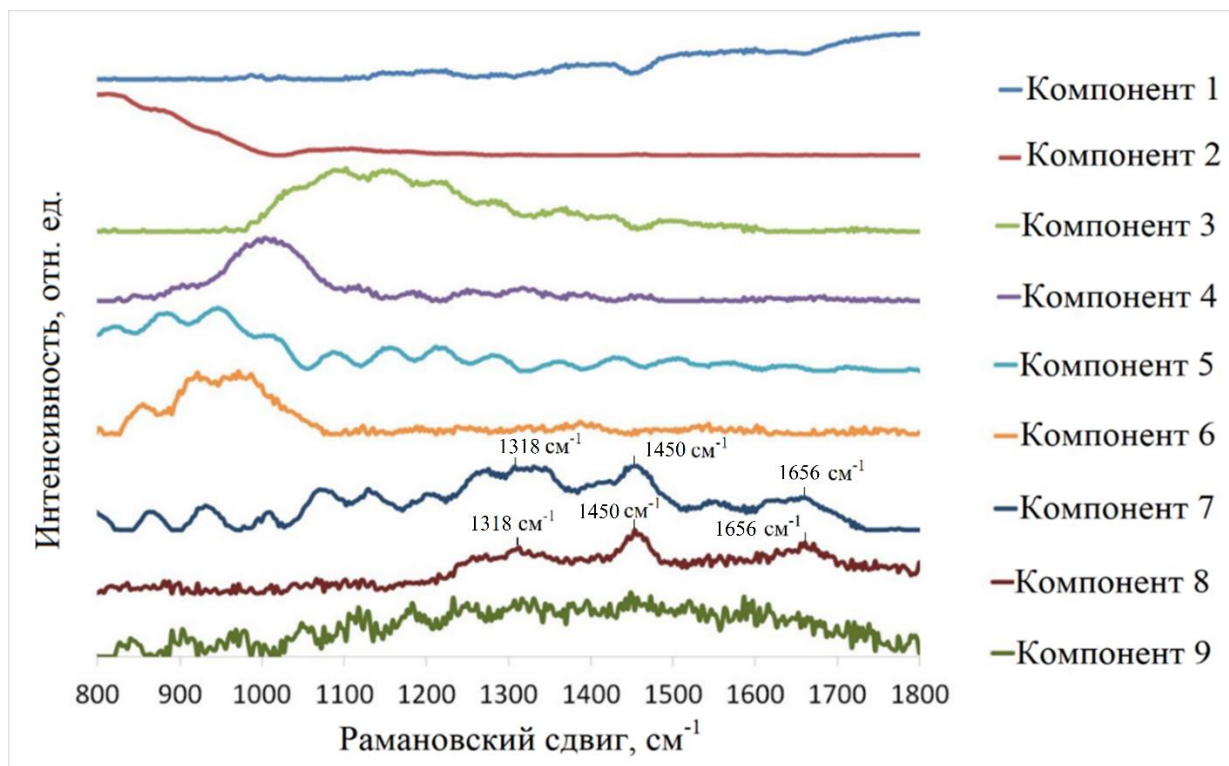


Рисунок 3.3 – Восстановленные спектры компоненты MCR-разложения полных экспериментальных спектров кожи

Таблица 3.1 – Рамановские пики, выделенные на спектрах компонентов кожи (см. рисунок 3.3), выделенных с помощью MCR-анализа

Рамановский сдвиг, см^{-1}	Колебание	Источник литературы
1318	деформационные С-Н колебания (белки)	[18]
1450	изгибные С-Н колебания (кератин)	[192]
1656	Амид I валентные С-О колебания (белки)	[8]

Интересно отметить, что в большинстве полученных спектров (см. компоненты 3, 4, 5, 6, 7, 9) видна синусоидальная составляющая. Наиболее вероятной причиной такой зависимости является вклад оптико-электронного

тракта системы регистрации. Аналогичная составляющая, представляющая вклад оптики, была получена в [102]. Применительно к используемой портативной установке такой эффект могут давать оптические фильтры оптоволоконного зонда – это хорошо видно на рисунке 3.4, где приведен спектр фильтра 785 nm RazorEdge® ultrasteep long-pass edge фирмы Semrock [193], аналог которого использовался в портативной установке.

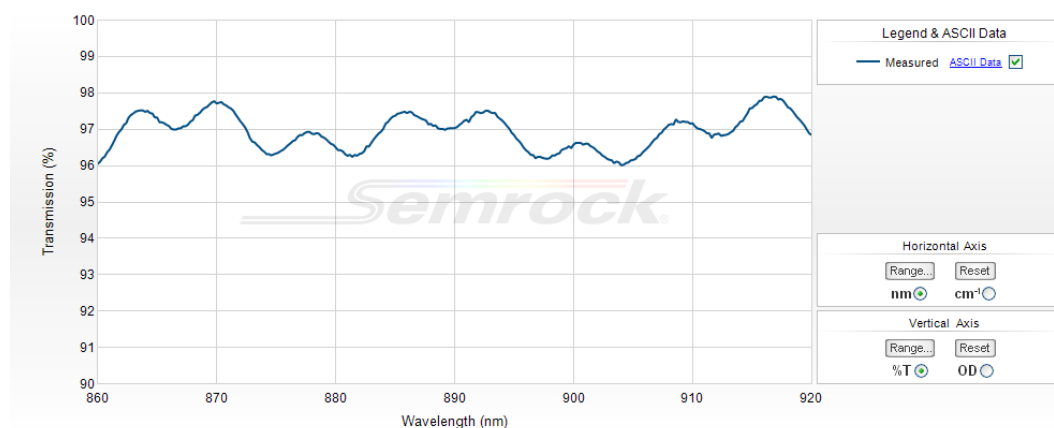


Рисунок 3.4 – Спектр фильтра 785 nm EdgeBasic™ best-value long-pass edge (Semrock) [193]

Отсутствие явно выраженных экстремумов в компонентах 2, 3, 4, 5, 6 не позволяет использовать их для идентификации компонентного профиля флуорофоров. Это указывает на необходимость удаления сигнала автофлуоресценции из полного спектра и перехода к MCR анализу только рамановского рассеяния.

3.2.2 Разделение *in vivo* рамановских спектров

Для выделения рамановских спектров все зарегистрированные спектры здоровой кожи и новообразований были подвергнуты предварительной обработке (см. раздел 2.2), которая обеспечивала исключение автофлуоресценции и фильтрацию шума. Характерный RS спектр здоровой кожи представлен на рисунок 3.5. Выделение спектральных компонент и определение их компонентного профиля проводилось методом MCR, описанным в разделе 3.1. На рисунке 3.6 приведены восстановленные спектральные компоненты при количестве компонентов разложения, равном 8, которые обеспечивают не менее 95% вклада в

экспериментально регистрируемый спектр в соответствии с SVD анализом. Восстановленный в результате MCR анализа спектральный профиль (MCR компоненты) идентифицировался на основании характерных пиков и известных спектров (пунктир) компонентов кожи (кератин, меланин, эластин, коллаген, вода, белки, липиды) (см. рисунок 3.6 и таблицу 3.2) [194].

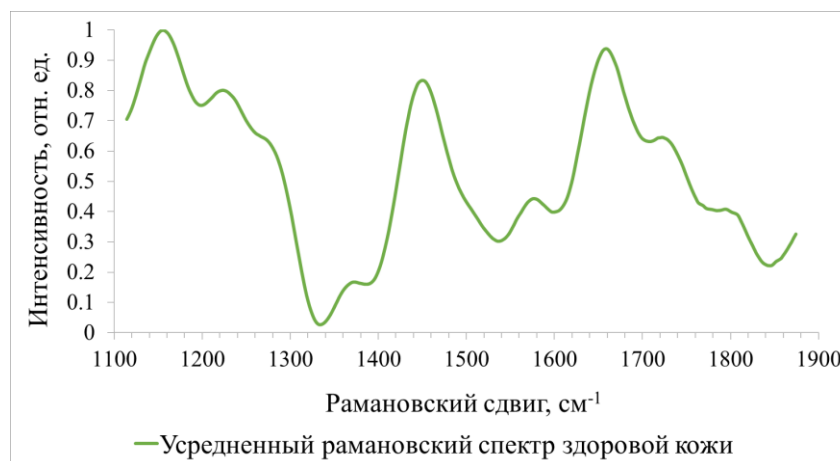


Рисунок 3.5 – Усредненный рамановский спектр здоровой кожи

Для подтверждения замкнутости восстановленных спектральных профилей компонентов, однозначно описывающих и идентифицирующих состав кожной ткани, было проведено отдельное исследование сходимости MCR метода с разложением на 5, 6, 12 и 30 компонентов, результаты которого представлены на рисунках 3.7 и 3.8 [195*]. Из анализа рисунка 3.7 следует, что при уменьшении количества компонент A до 5 или 6 восстанавливаемые спектральные профили сохраняют устойчивость, имеют те же характерные спектральные пики, что и первые MCR компоненты разложения с $A = 8$, то есть несут информацию о тех же компонентах, что подтверждает их преимущественный вклад в экспериментально регистрируемый RS спектр. При разложении на 12 компонентов все члены разложения выше 8 порядка были шумоподобными и не имели каких-либо отчетливых пиков. Использование еще большего количества членов разложения (30 и более) приводило к потере многопиковых спектральных профилей (рисунок 3.8), характерных для химических компонент кожи. На основании проведенного исследования далее использовалось MCR разложение с количеством компонентом, равным 8.

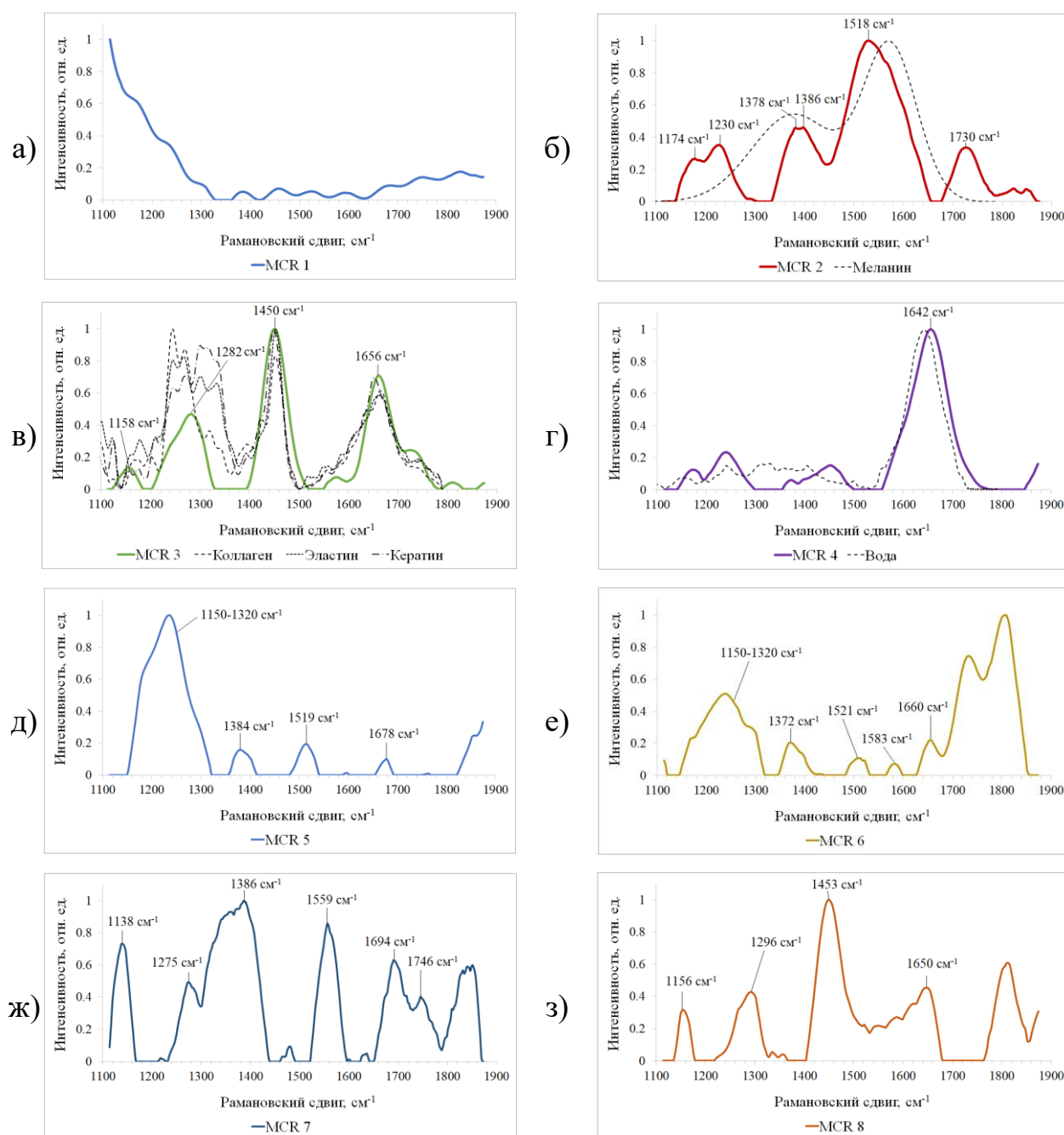
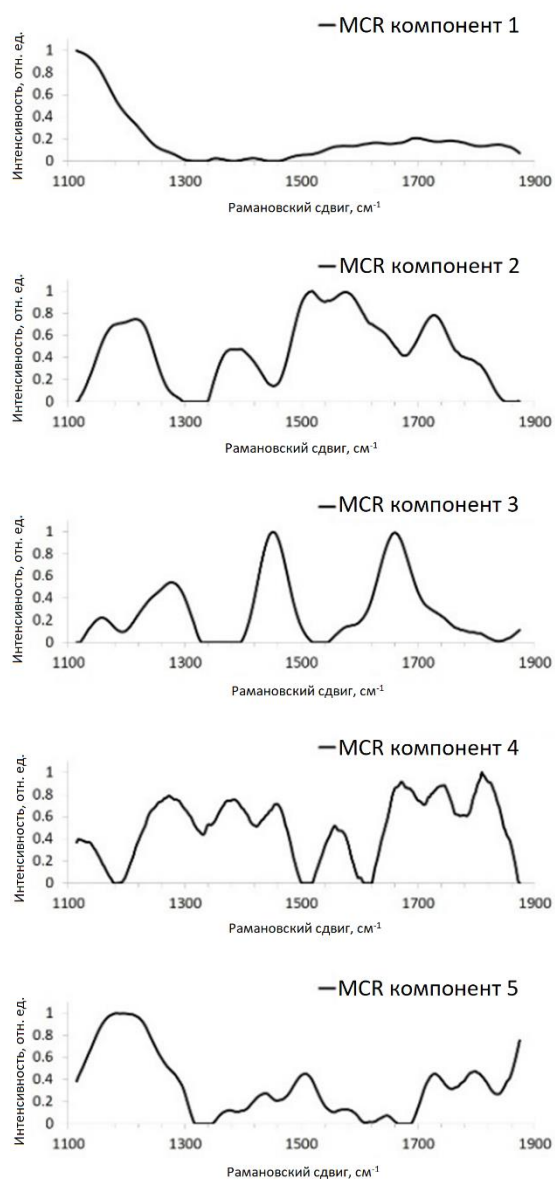


Рисунок 3.6 – RS спектры, полученные с помощью MCR анализа *in vivo* RS спектров кожи. Сплошные линии – спектры, полученные в данном исследовании, штриховые линии – спектры из работы [19]: а) MCR компонент 1 (вклад оптико-электронного тракта), б) MCR компонент 2 (меланин, естественный увлажняющий фактор), в) MCR компонент 3 (белки), г) MCR компонент 4 (вода), д) MCR компонент 5 (белки, липиды, нуклеиновые кислоты), е) MCR компонент 6 (белки, липиды), ж) MCR компонент 7 (белки, липиды, меланин, естественный увлажняющий фактор), з) MCR компонент 8 (белки, вода)

Таблица 3.2 – Рамановские пики, выделенные на спектрах компонентов кожи (см. рисунок 3.6), выделенных с помощью MCR анализа

Рамановский сдвиг, см ⁻¹	Колебание	Источник литературы
1138	пальмитиновая кислота, жирные кислоты	[196]
1150-1320	каротиноиды, тирозин, цитозин, гуанин, фенилаланин, аденин, Амид III	[197, 198, 199, 200, 201, 202]
1156, 1158	растяжение C–C и C–N	[197]
1174	деформационные колебания C–H	[102]
1230	Амид III	[102, 197, 203]
1248	растягивающие асимметричные PO ₂ колебания (нуклеиновые кислоты)	[204]
1275	Амид III растягивающие C–N и N–H колебания	[200]
1282	Амид III и CH ₂ основной цепи глицина и боковых цепей пролина	[203, 205, 206, 207]
1296	деформационные колебания CH ₂	[208]
1372	колебания оснований ДНК и РНК	[202, 209]
1378	линейное растяжение связей C–C с некоторым вкладом колебаний C–H в метильных и метиленовых группах	[8]
1384	CH ₃ колебания	[200]
1386	деформационное колебание C–H, валентное колебание C–C	[102]
1450	изгибные C–H колебания, ножничные колебания δ(CH ₂), колебания δ(CH ₂), δ(CH ₃)	[204, 207, 210, 211, 212]
1453	изгибные C–H колебания структурных белков	[207, 213, 214]
1518	цитозин, синфазные C=C и C–C валентные колебания полиеновой цепи	[198, 215]
1519, 1521	–C=C– (каротиноиды)	[200]
1559	триптофан	[209]
1583	изгибные C=C колебания фенилаланина	[209]
1642	изгибные колебания OH	[19, 216]
1650	Амид I, колебания C=C	[205]
1656	Амид I валентные C–O колебания	[211, 212]
1660	Амид I, C=C колебания в ненасыщенных жирных кислотах, липиды	[196, 197, 217]
1678	никотинамидадениндинуклеотид	[205]
1694	Амид I, валентные колебания C=O, C=C	[210, 211, 212]
1730	колебание C=O Амид I липидов	[196]
1746	растяжение C=O	[196]

Разложение на 5 компонентов



Разложение на 6 компонентов

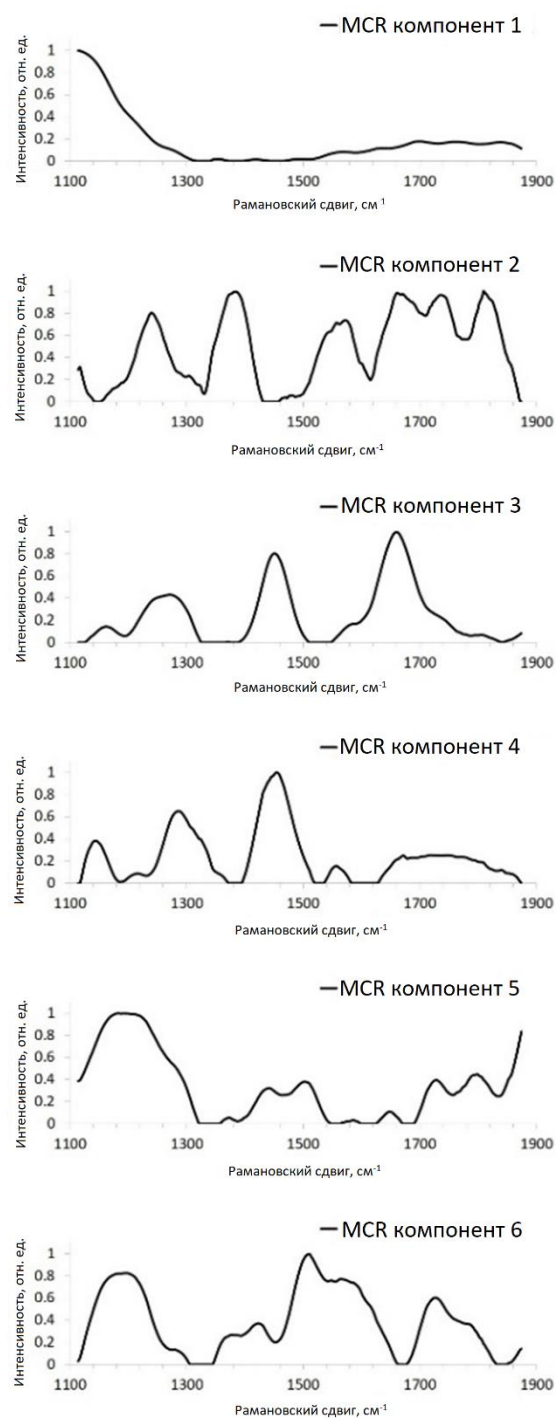


Рисунок 3.7 – RS спектры, полученные с помощью MCR анализа *in vivo* RS спектров кожи при разложении на 5 и 6 компонентов

Разложение на 12 компонентов

Разложение на 30 компонентов

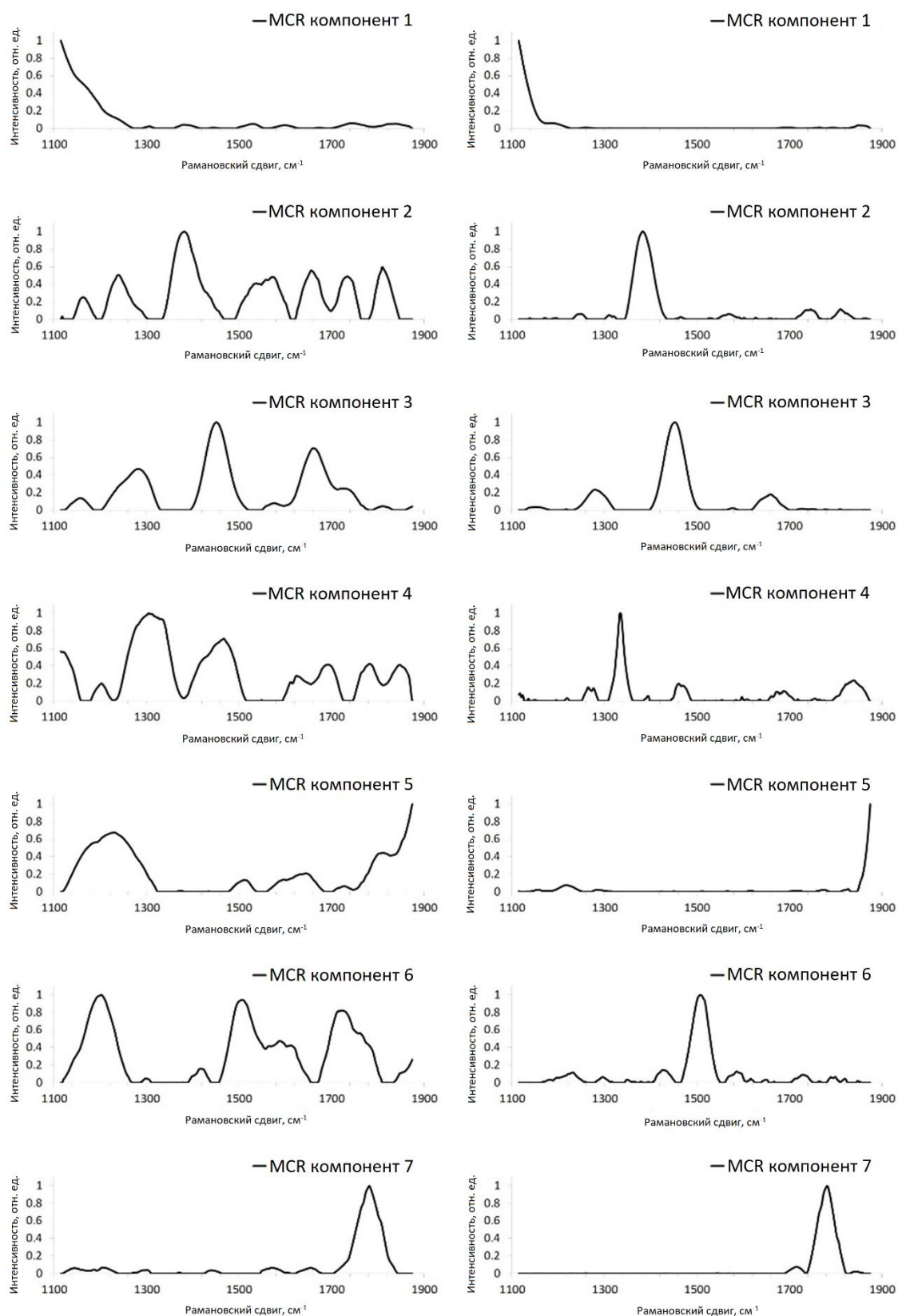


Рисунок 3.8 – RS спектры, полученные с помощью MCR анализа *in vivo* RS спектров кожи при разложении на 12 и 30 компонентов (представлены первые семь компонентов для каждого случая)

Первый компонент MCR разложения (см. рисунок 3.6а) имеет плавно спадающий вид с гармонической составляющей с характерным периодом, совпадающим с синусоидальными колебаниями, выявленными при исследовании полного спектра с автофлуоресценцией (рисунок 3.4). Естественным является отождествить данный компонент с вкладом в MCR разложение оптико-электронного тракта портативной установки. Интересно отметить, что сигнал аналогичной формы (гармонические колебания на фоне спадающей интенсивности) был получен в [102] в результате анализа спектров рамановского рассеяния кожи при микроскопическом исследовании. В диссертационной работе используется портативная установка с низким отношением сигнал/шум, поэтому вклад оптико-электронного тракта спектрометра более выражен. Способность идентифицировать паразитный сигнал, связанный с вкладом оптической системы, потенциально может быть использована при предварительной обработке RS сигналов.

Особенностью и преимуществом метода MCR является возможность физической интерпретации получаемых компонент. Так, компонента 2 (рисунок 3.6б) представляет преимущественно меланин, который имеет широкие пики с центрами 1378 см^{-1} и 1573 см^{-1} , что совпадает со спектром меланина (пунктир), зарегистрированного в работе [19]. Вместе с тем видно, что в спектр компоненты 2 также вносят вклад белки (1518 см^{-1}) [198, 215], валентные колебания С=О Амид I липидов (1730 см^{-1}) [196], а пики 1174 см^{-1} , 1230 см^{-1} и 1386 см^{-1} соответствуют вкладу естественного увлажняющего фактора (аминокислоты и их производные, такие как пирролидонкарбоновая и урокановая кислоты) [102].

Компонент 3 (рисунок 3.6в) представляет вклад различных белков (изгибные С–Н колебания кератина 1450 см^{-1} , Амид I валентные С–О колебания 1656 см^{-1} и Амид III колебания 1282 см^{-1} коллагена и эластина) [203, 205]. Сравнение с результатами, полученными в работе [19], показывает сходство этого спектра со спектрами коллагена (пунктир), эластина (точки) и кератина (штрих-пунктир) в диапазоне от 1400 до 1800 см^{-1} .

Компонента 4 (рисунок 3.6г) ассоциирована с вкладом воды, так как наиболее интенсивный пик 1642 см^{-1} соответствует изгибным колебаниям ОН. Сравнение с результатами микроскопического исследования [19] показывает почти полное совпадение спектров. Компоненты 5 и 6 имеют широкую интенсивную полосу от 1150 см^{-1} до 1320 см^{-1} , соответствующему вкладу Амид III растягивающим С–N и N–H колебаниям белков (1271 см^{-1}), скручивающим CH_2 колебаниям липидов и коллагена (1296 см^{-1}), асимметричным растягивающим С–N и деформационным С–H колебаниям белков (1318 см^{-1}), а также растягивающему асимметричному PO_2 колебанию нуклеиновых кислот (1248 см^{-1}). В спектре компонента 7 можно отметить пики 1138 см^{-1} (пальмитиновая кислота, жирные кислоты), 1275 и 1746 см^{-1} (липиды), 1386 см^{-1} (естественный увлажняющий фактор), 1559 см^{-1} (триптофан) и 1694 см^{-1} (коллаген). Спектр компонента 8 содержит интенсивные пики 1453 см^{-1} (белки) и 1650 см^{-1} (Амид I). Таким образом, каждая компонента соответствует в физическом смысле RS спектру вещества или группы веществ – компоненту кожи, что позволяет извлекать из спектральных данных важную диагностическую информацию, понятную для врача. В результате MCR анализа не удалось явно реконструировать спектры, связанные с вкладом ядер клеток, дезоксирибонуклеиновой кислоты (ДНК), тем не менее, компонента 6 содержит пики, соответствующие PO_2 колебанию нуклеиновых кислот (1248 см^{-1}), что может быть использовано при дискриминации ВСС.

Важно понимать, что причина наложения пиков и полос (именно это мешает корректной работе метода MCR) заключается не только в природе этих веществ, но и уровне шумов, при которых ведется регистрация рамановских спектров. Тем не менее, даже в случае использования портативной установки удалось явно выявить четыре компонента, которые преимущественно определяют вклад меланина, белков, воды и липидов. Остальные компоненты ассоциированы с вкладом групп нескольких веществ. Это позволяет использовать восстановленные с помощью метода MCR спектры компонентов для определения компонентных профилей новообразований, которые можно использовать в качестве информативных

признаков при классификации различных нозологий с использованием машинного обучения.

3.2.3 Изучение вклада отдельных компонентов в рамановские спектры различных новообразований

Ограничением MCR анализа является тот факт, что восстанавливаемый компонентный состав содержит информацию только об относительной объемной концентрации компонентов, что исключает возможность прямого биохимического анализа и сравнения значения концентраций компонентов кожи. Вместе с этим можно косвенно подтвердить правильность работы MCR метода, изучая общие закономерности в компонентном профиле распределения [218*, 219*]. На рисунке 3.9 представлены диаграммы размаха относительных объемных концентраций компонентов кожи, полученных методом MCR анализа с использованием базиса (см. рисунок 3.6). По рисунку 3.9 можно заметить, например, что для пигментных новообразований (SK, MM, NE) регистрируется повышенный средний уровень меланина, то есть повышение вклада компонента 2 в общий компонентный профиль. Более комплексную оценку эффективности анализа спектров методом MCR можно получить, используя классификационные модели.

В качестве иллюстрации возможности классификации рамановских спектров по нозологиям, исходя из полученных относительных концентраций компонентов кожи, выбраны несколько классификационных моделей [220*, 221*]: ВСС против здоровой кожи и остальных заболеваний, ВСС против MM, злокачественные новообразования (ВСС + MM) против доброкачественных новообразований (SK + NE), MM против NE. Данные модели применялись для исследования вклада выделенных компонентов в предсказания и классификацию указанных нозологий. Оценка качества классификации проводилась с использованием кривых рабочих характеристик приемника (ROC). Это графическая характеристика качества бинарного классификатора, которая представляет собой зависимость доли верных положительных классификаций (TPR) от доли ложных положительных классификаций (FPR) при варьировании порога решающего правила.

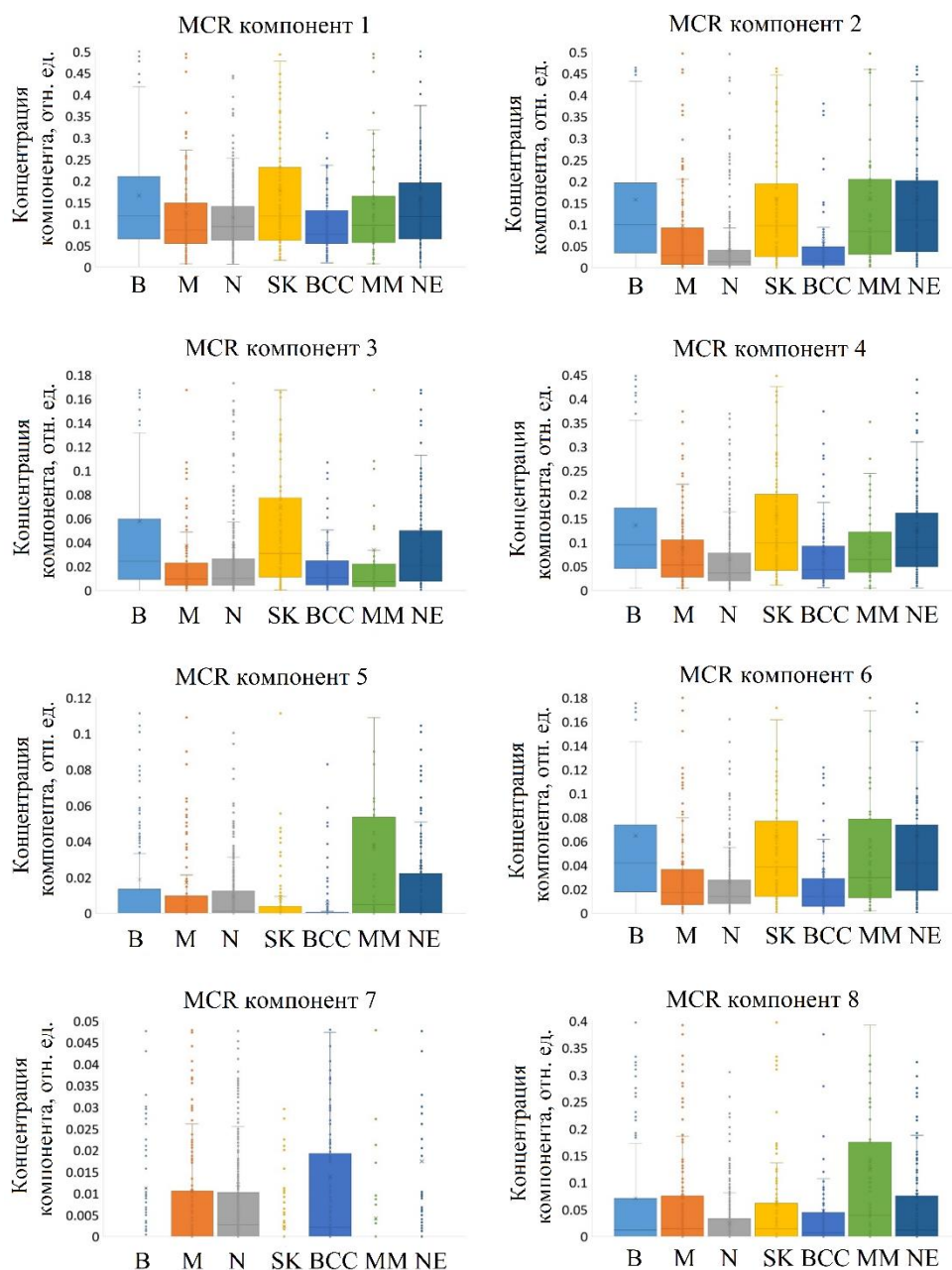


Рисунок 3.9 – Диаграммы размаха относительных объемных концентраций компонентов кожи, полученных методом MCR анализа с использованием базиса (см. рисунок 3.6): В – доброкачественные новообразования, М – злокачественные новообразования, N – здоровая кожа, SK – себорейный кератоз, BCC – базальноклеточная карцинома, MM – злокачественная меланома, NE – невус

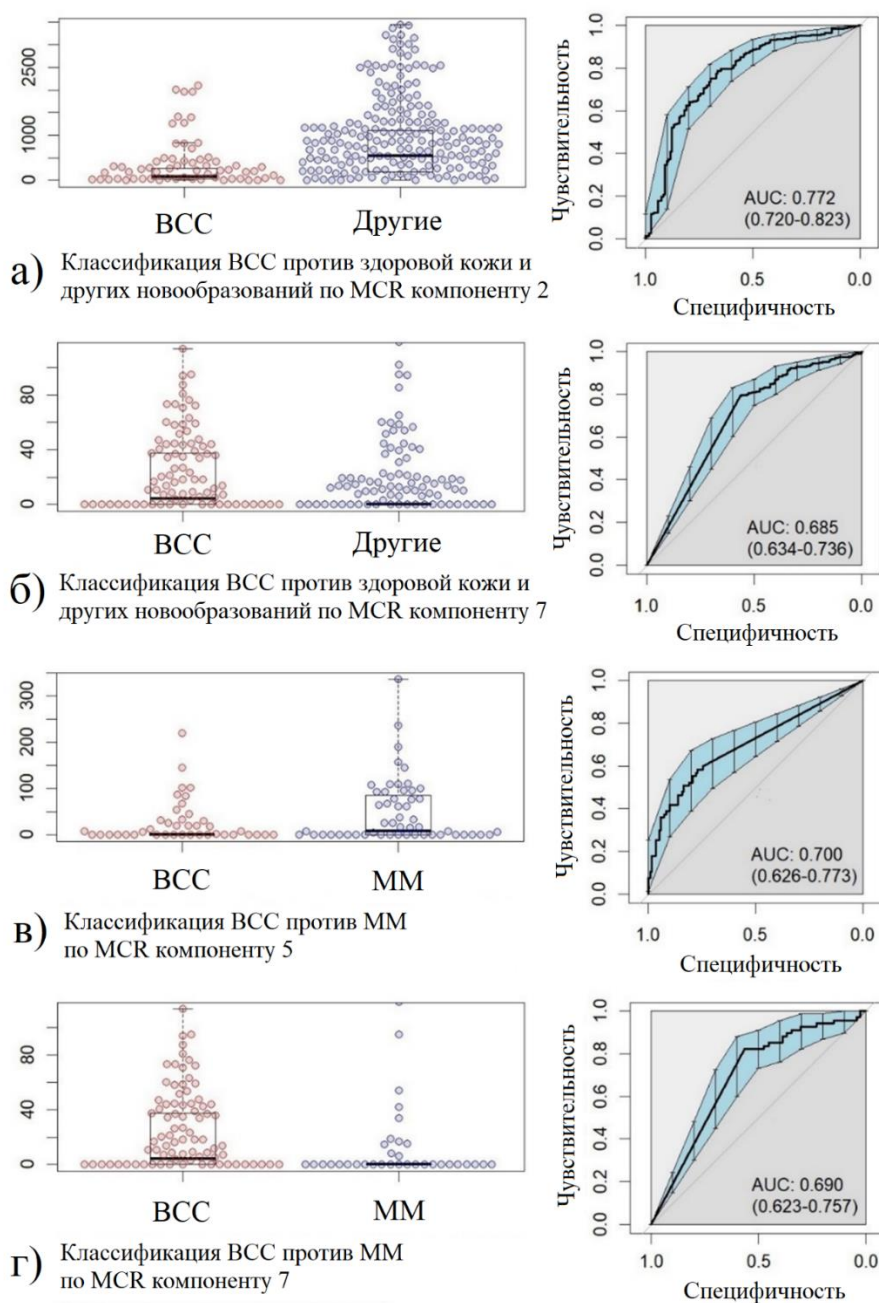


Рисунок 3.10 – Диаграммы размаха и ROC-кривые, соответствующие нескольким случаям классификации: а) BCC против здоровой кожи и остальных заболеваний («Другие») по компоненту 2, б) BCC против здоровой кожи и остальных заболеваний по компоненту 7, в) BCC против MM по компоненту 5, г) BCC против MM по компоненту 7

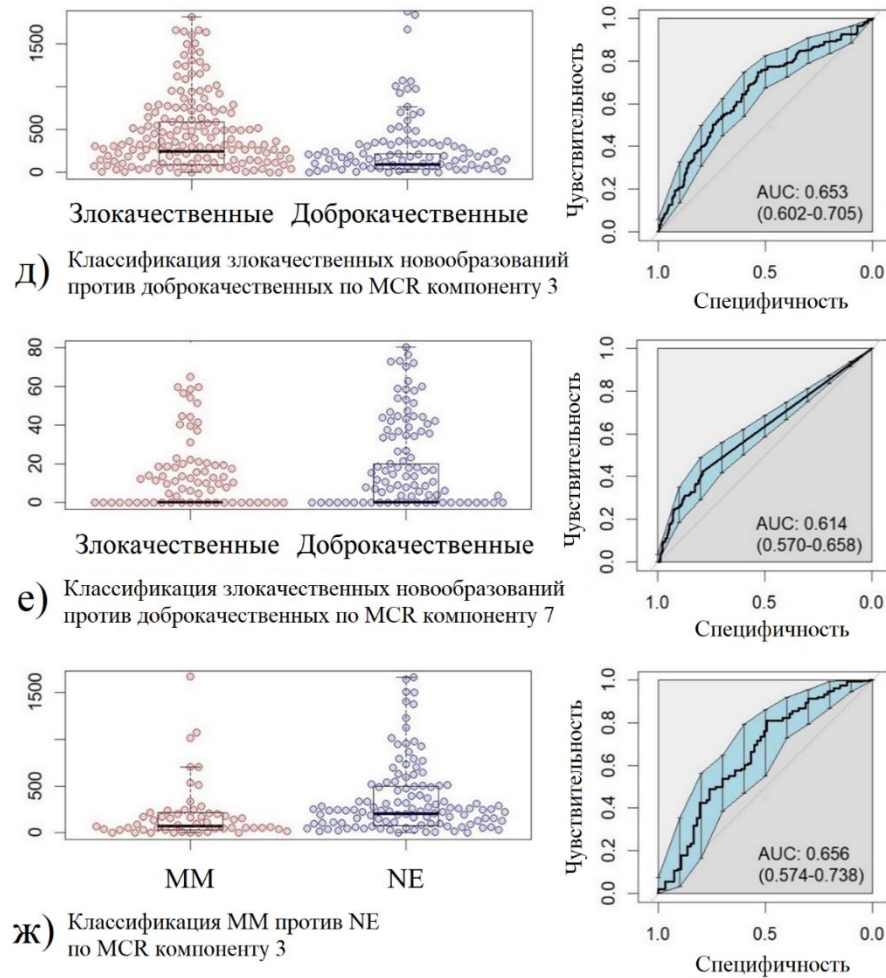


Рисунок 3.11 – Диаграммы размаха и ROC-кривые, соответствующие нескольким случаям классификации: д) злокачественные новообразования (BCC + MM) против доброкачественных новообразований (SK + NE) по компоненту 3, е) злокачественные новообразования (BCC + MM) против доброкачественных новообразований (SK + NE) по компоненту 7, ж) MM против NE по компоненту 3

Преимуществом ROC-кривой является её инвариантность относительно отношения цены ошибки I и II рода. ROC-кривая позволяет найти оптимальный порог классификации, который максимизирует баланс между чувствительностью (вероятностью правильного определения положительного класса) и специфичностью (вероятностью правильного определения отрицательного класса). Чем ближе ROC-кривая к верхнему левому углу графика, тем лучше работает классификатор.

В качестве скалярной характеристики сравнения различных классификационных моделей выбрана площадь под ROC кривой (AUC). Значение ROC AUC может быть интерпретировано как вероятность того, что классификатор правильно классифицирует случайно выбранный положительный пример выше, чем случайно выбранный отрицательный пример. Таким образом, чем выше значение площади под кривой ROC, тем более надежными будут предсказания классификатора [222].

На рисунках 3.10 и 3.11 показаны диаграммы размаха, ROC кривые и ROC AUC с 95% доверительным интервалом (CI), соответствующие этим случаям. Диаграммы размаха и ROC кривые были построены с использованием пакета pROC [223].

В случае ВСС против здоровой кожи и всех заболеваний, AUC составляет 0,772 (0,720-0,823, 95% CI) при разделении по компоненту 3 и 0,685 (0,634-0,736, 95% CI) при разделении по компоненту 7. Наилучшее разделение ВСС и ММ возможно по компонентам 5 и 7 – 0,700 (0,626-0,773, 95% CI) и 0,690 (0,623-0,757, 95% CI), соответственно.

В случае классификации злокачественных новообразований (ММ, ВСС) и доброкачественных новообразований (SK, NE) наивысшее значение AUC составляет 0,653 (0,602-0,705, 95% CI) при разделении по компоненту 3 и чуть ниже – 0,614 (0,570-0,658, 95% CI) – при разделении по компоненту 7. Интересно, что модели классификации ММ против NE демонстрируют AUC 0,656 (0,574-0,738, 95% CI). Это достаточно высокий результат с учетом того, что эти типы новообразований имеют схожие внешние проявления и сопоставимое количество меланина.

Полученные результаты подтверждают возможность применения метода MCR для анализа спектральных данных оптической биопсии. Метод позволяет выделить из исследуемых рамановских спектров информативные признаки, которые можно использовать для дальнейшего анализа. Другими словами, предлагаемый подход позволяет извлекать из спектральных данных важную диагностическую информацию, понятную для врача. Это приближает данный метод к широко

используемым методам биохимического анализа биологических жидкостей, таких как кровь, моча и т.д.

3.3 Идентификация новообразований кожи по спектральным данным

3.3.1 Классификация на основе компонентного профиля

Рассмотрим применение и использование найденных в предыдущем разделе концентрационных профилей спектральных компонент новообразований в качестве признаков для идентификации спектральных данных рамановского рассеяния новообразований.

В работе исследуются три классификационные задачи:

1. Дискриминация злокачественных (MM, BCC, SCC) и доброкачественных (DF, PP, HE, SK, NE) новообразований;
2. Дискриминация MM и доброкачественных пигментных новообразований кожи (SK + NE);
3. Дискриминация MM и SK.

Важность построения именно таких классификационных моделей обусловлена тем, что первая модель важна с точки зрения доказательства принципиальной возможности применения MCR разложения для дискриминации злокачественных новообразований, вторая модель позволяет обосновать точность метода MCR по предсказанию наиболее опасного онкологического заболевания кожи – злокачественной меланомы, а последняя задача интересна с точки зрения исследования возможности метода для разделения меланоцитарных злокачественных (MM) и доброкачественных (NE) опухолей, для которых характерна одинаково высокая концентрация меланина.

В качестве способа классификации применяются два метода: логистическая регрессия и PLS-DA. Логистическая регрессия выбрана как самый простой способ анализа взаимосвязи между несколькими независимыми переменными (предикторами) и зависимой переменной [224]. Предикторами являются относительные концентрации компонентов, описывающие их вклад в RS спектр.

По сути, логистическая регрессия вычисляет вероятность того, что данное наблюдение принадлежит к определенному классу, и строит гиперплоскость, наилучшим образом разделяющую данные. Метод PLS-DA широко используется в биомедицинских исследованиях и позволяет построить регрессионную модель на основе значительных различий в исследуемой выборке, проецируя прогнозируемые переменные и наблюдаемые переменные в новое пространство [225]. В работе [5] метод PLS-DA применяется для анализа описанных во второй главе спектральных данных, поэтому его применение анализу компонентных профилей новообразований позволит судить о том, насколько полно MCR компоненты отражают информацию, содержащуюся в спектральных отсчетах рамановских спектров.

Стабильность полученных классификационных моделей проверялась k -кратной перекрестной валидацией (кросс-валидацией). Перекрестная проверка проводилась для $k = 10$ по следующему алгоритму: 10% объектов в выборке исключаются из нее и становятся тестовой выборкой. Модель строится на основе оставшейся части данных. Затем возвращается исключенная часть объектов, в качестве тестовой выборки снова выбираются случайные 10% объектов, еще не побывавшие в тестовой выборке, и модель строится снова. Цикл повторяется 10 раз, и итоговая характеристика модели рассчитывается как средняя по десяти получившимся тестовым результатам. Перекрестная проверка позволяет определить оптимальные параметры модели разделения классов, чтобы избежать переобучения.

Важно упомянуть, что используемые в работе спектральные данные рамановского рассеяния, описанные во второй главе, содержат несбалансированное количество разных видов новообразований, поэтому при формировании обучающей и тестовой выборки классы балансировались следующим образом. В выборку меньшего класса попадали все имеющиеся образцы, а выборка большего класса формировалась случайным образом из имеющихся образцов этого класса. Кроме того, внутри каждого класса каждой выборки сохранялось примерное соотношение диагнозов, представленное в исходном наборе данных. Учитывая

случайный характер формирования выборки, построение моделей классификации было проведено не менее 10 раз, и далее представлены результаты, полученные в результате усреднения метрик моделей по количеству проведенных экспериментов.

Таким образом, на каждом этапе построения модели дискриминации злокачественных и доброкачественных новообразований использовалось 65 MM, 120 BCC, 12 SCC, 14 DF, 31 PP, 20 HE, 56 SK и 83 NE, взятых случайно из общей выборки. На каждом этапе построения модели дискриминации MM и доброкачественных пигментных новообразований использовалось 65 MM, 26 SK и 39 NE, а модели дискриминации MM и SK – 65 MM и 65 SK.

Наряду с ROC AUC, для оценки качества классификатора используется матрица ошибок (см. таблицу 3.3). Если предсказанный класс соответствует фактическому, то исход классификации считается истинным, а в противном случае — ложным. Примеры положительного и отрицательного классов, для которых исход предсказания истинный, называются истинноположительными (TP) и истинноотрицательными (TN), соответственно. Примеры положительного и отрицательного классов, для которых исход предсказания является ложным, называются ложноположительными (FP) и ложноотрицательными (FN), соответственно.

Таблица 3.3 – Таблица, иллюстрирующая матрицу ошибок

	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	количество истинноотрицательных исходов (TN)	количество ложноположительных исходов (FP)
$y = 1$	количество ложноотрицательных исходов (FN)	количество истинноположительных исходов (TP)

По данным, представленным в матрице ошибок, вычислялись следующие дополнительные метрики качества модели бинарной классификации.

1. Доля правильных ответов (англ. accuracy – ACC) — доля правильно классифицированных примеров:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}. \quad (3.9)$$

Несмотря на то, что доля правильных ответов хорошо интерпретируется, на практике она используется достаточно редко, поскольку не информативна в случае дисбаланса классов.

2. Точность (англ. precision) – это отношение числа истинноположительных классификаций к общему числу положительных классификаций. Данная величина также известна как положительное прогностическое значение (англ. positive predictive value – PPV):

$$PPV = \frac{TP}{TP+FP}. \quad (3.10)$$

3. Полнота (англ. recall) — число истинноположительных классификаций относительно общего числа положительных примеров.

$$TPR = \frac{TP}{TP+FN}. \quad (3.11)$$

Таким образом, полноту можно рассматривать как чувствительность (англ. sensitivity), или способность классификатора обнаруживать определенный тип новообразования. Точность и полнота, в отличие от доли правильных ответов, не зависят от соотношения классов и могут применяться в условиях несбалансированных выборок.

4. Специфичность (англ. specificity) – доля истинноотрицательных (TNR) классификаций в общем числе отрицательных классификаций:

$$TNR = \frac{TN}{TN+FP}. \quad (3.12)$$

Данная величина показывает, насколько хорошо модель классифицирует отрицательные примеры.

5. Точность отрицательного прогноза (NPV) – доля верно классифицированных отрицательных примеров от общего числа примеров, классифицированных как отрицательные:

$$NPV = \frac{TN}{TN+FN}. \quad (3.13)$$

6. F1-мера объединяет в себе информацию о точности и полноте, поэтому позволяет находить баланс между ними:

$$F1 = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR}. \quad (3.14)$$

Результаты классификации новообразований методами логистической регрессии и PLS-DA на основе относительного компонентного состава, а именно ROC кривые предсказанных значений для каждого случая дискриминации представлены на рисунке 3.12 (кривые с обозначением «MCR»). Информация о значениях ROC AUC представлена в таблице 3.4.

В случае применения в качестве способа классификации логистической регрессии самое высокое значение ROC AUC составляет 0,83 (0,74-0,92, 95% CI) для случая дискриминации MM и SK. Разделение злокачественных и доброкачественных опухолей возможно с ROC AUC 0,78 (0,72-0,84, 95% CI), тогда как разделение MM и SK + NE демонстрирует всего 0,74 (0,67-0,82, 95% CI) ROC AUC. Применение PLS-DA демонстрирует значения ROC AUC несколько ниже, чем логистическая регрессия. Самая высокая ROC AUC так же получена для случая дискриминации MM и SK и составляет 0,80 (0,72-0,88, 95% CI). Худшие результаты на уровне 0,65 (0,59-0,71, 95% CI) дает разделение злокачественных и доброкачественных новообразований, а в случае разделения MM и SK + NE ROC AUC принимает значение на уровне 0,71 (0,63-0,80, 95% CI).

Полученные результаты анализа новообразований кожи на основе выделенных компонентных профилей целесообразно сравнить с данными PLS-DA анализа рамановских спектров кожных новообразований, полученных ранее в работе [5] на той же портативной установке. В работе [5] исследовались такие же три задачи бинарной классификации новообразований кожи, но без выделения компонентного профиля новообразований. Стоит отметить, что в исследовании [5] анализировались информативные полосы спектров новообразований в диапазоне от 800 см⁻¹ до 1800 см⁻¹, в то время как в данном диссертационном исследовании MCR признаки и компонентный профиль восстанавливались в спектральном диапазоне от 1100 см⁻¹ до 1800 см⁻¹. Однако и в том, и в другом случае построение моделей классификации проводилось с помощью PLS-DA, поэтому это позволяет

более предметно сравнивать результаты классификации на основе относительного компонентного профиля и результаты исследования [5].

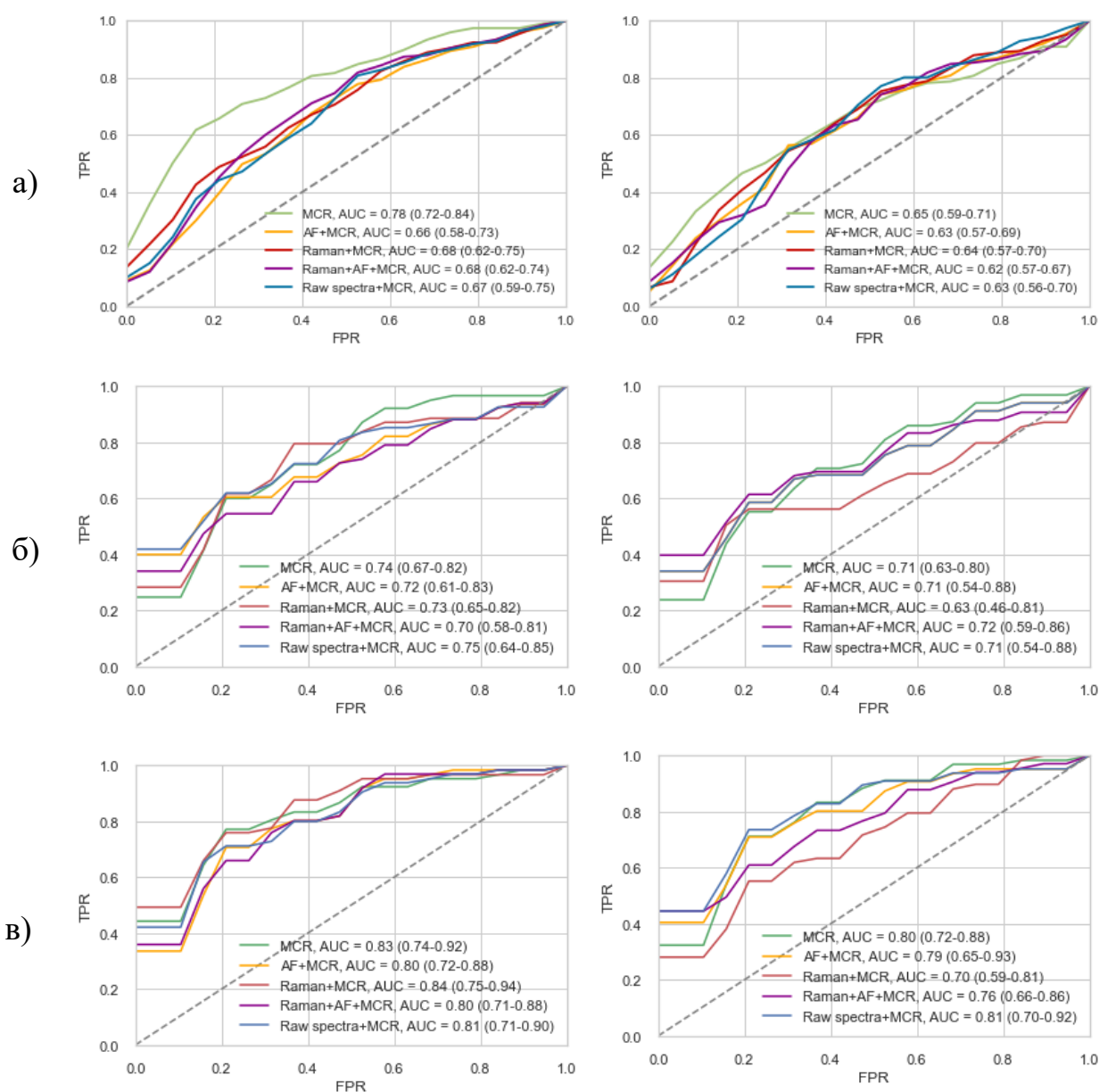


Рисунок 3.12 – Классификация на основе анализа логистической регрессии (слева) и PLS-DA (справа): а) злокачественные и доброкачественные новообразования; б) ММ и SK + NE; в) ММ и SK. Доверительные интервалы ROC AUC рассчитаны для уровня вероятности не менее 95%

Таблица 3.4 – Значения ROC AUC для моделей классификации, построенных в данной работе, и значения ROC AUC, полученные в исследовании [5]. Доверительные интервалы ROC AUC рассчитаны для уровня вероятности не менее 95%

Признаки \ Методы	PLS-DA [5]	PLS-DA + MCR	Логистическая регрессия + MCR	LightGBM + MCR
Дискриминация злокачественных и доброкачественных новообразований				
MCR	-	0,65 (0,59-0,71)	0,78 (0,72-0,84)	0,82 (0,77-0,86)
AF (300-1200 cm^{-1})	0,62 (0,57-0,67)	0,63 (0,57-0,69)	0,66 (0,58-0,73)	0,82 (0,79-0,86)
Raman (1200-1800 cm^{-1})	0,75 (0,71-0,79)	0,64 (0,57-0,70)	0,68 (0,62-0,75)	0,80 (0,76-0,85)
Raman (1200-1800 cm^{-1}) + AF (300-1200 cm^{-1})	0,75 (0,71-0,79)	0,62 (0,57-0,67)	0,68 (0,62-0,74)	0,81 (0,77-0,85)
Raw spectra (300-1800 cm^{-1})	0,60 (0,56-0,65)	0,63 (0,56-0,70)	0,67 (0,59-0,75)	0,82 (0,79-0,86)
Дискриминация ММ и доброкачественных пигментных новообразований кожи (SK + NE)				
MCR	-	0,71 (0,63-0,80)	0,74 (0,67-0,82)	0,82 (0,76-0,89)
AF (300-1200 cm^{-1})	0,66 (0,59-0,74)	0,71 (0,54-0,88)	0,72 (0,61-0,83)	0,81 (0,74-0,89)
Raman (1200-1800 cm^{-1})	0,61 (0,53-0,69)	0,63 (0,46-0,81)	0,73 (0,65-0,82)	0,83 (0,77-0,90)
Raman (1200-1800 cm^{-1}) + AF (300-1200 cm^{-1})	0,66 (0,59-0,74)	0,72 (0,59-0,86)	0,70 (0,58-0,81)	0,81 (0,73-0,89)
Raw spectra (300-1800 cm^{-1})	0,69 (0,63-0,76)	0,71 (0,54-0,88)	0,75 (0,64-0,85)	0,81 (0,73-0,90)
Дискриминация ММ и SK				
MCR	-	0,80 (0,72-0,88)	0,83 (0,74-0,92)	0,87 (0,79-0,94)
AF (300-1200 cm^{-1})	0,80 (0,73-0,87)	0,79 (0,65-0,93)	0,80 (0,72-0,88)	0,86 (0,81-0,92)
Raman (1200-1800 cm^{-1})	0,75 (0,67-0,83)	0,70 (0,59-0,81)	0,84 (0,75-0,94)	0,86 (0,77-0,94)
Raman (1200-1800 cm^{-1}) + AF (300-1200 cm^{-1})	0,81 (0,74-0,87)	0,76 (0,66-0,86)	0,80 (0,71-0,88)	0,85 (0,75-0,94)
Raw spectra (300-1800 cm^{-1})	0,80 (0,73-0,86)	0,81 (0,70-0,92)	0,81 (0,71-0,90)	0,85 (0,78-0,92)

Raw spectra – «сырые» рамановские спектры без предварительной обработки, AF – спектры автофлуоресценции в диапазоне 300-1200 cm^{-1} , Raman – предварительно обработанные рамановские спектры (убран фон флуоресценции)

ROC AUC, полученные в [5] для разных задач классификации и разных способов анализа спектров, варьировались в диапазоне от 0,60 (0,56-0,65, 95% CI) до 0,81 (0,74-0,87, 95% CI), в то время как в настоящем исследовании AUC ROC составляет от 0,65 (0,59-0,71, 95% CI) до 0,83 (0,74-0,92, 95% CI). Другими словами, полученные в этом исследовании результаты сравнимы с результатами исследования, представленного в [5], по эффективности классификации нозологий новообразований кожи. Однако несомненным преимуществом использования MCR разложения с выделением концентрационных профилей компонентов кожи в модели классификации является возможность ее физической интерпретации. Если авторы исследования [5] оперируют абстрактными математическими компонентами, то в данном исследовании в качестве признаков классификации выступают физически интерпретируемые и понятные медицинскому специалисту характеристики.

3.3.2 Классификация новообразований на основе спектральных отсчетов и компонентного профиля

Возникает вопрос, насколько полно спектральные компонентные профили, выделенные в результате MCR анализа, описывают анализируемые рамановские спектры новообразований. Для проверки гипотезы о том, что неизбежно потерянная в результате MCR анализа спектральная информация не вносит значительных изменений в эффективность работы моделей классификации, необходимо добавить в модель не только выделенные заранее методом MCR признаки, но и сами спектральные отсчеты рамановских спектров. Обработка экспериментальных данных проводилась аналогично с предыдущим пунктом с помощью логистической регрессии и PLS-DA. На рисунке 3.12 представлены ROC кривые для каждого случая дискриминации с использованием логистической регрессии (слева) и PLS-DA (справа). Информация о значениях ROC AUC представлена в таблице 3.4.

Всего было рассмотрено четыре спектроскопических задачи по аналогии с задачами в исследовании [5]:

1. Совместный анализ концентрационных профилей и спектров автофлуоресценции в диапазоне 300-1200 см^{-1} (кривая с обозначением «AF + MCR»);
2. Совместный анализ концентрационных профилей и спектров рамановского рассеяния в диапазоне 1200-1800 см^{-1} (кривая с обозначением «Raman + MCR»);
3. Совместный анализ концентрационных профилей, рамановского сигнала и автофлуоресценции (кривая с обозначением «Raman + AF + MCR»);
4. Совместный анализ концентрационных профилей и полных экспериментальных спектров (без предобработки) в диапазоне 300-1800 см^{-1} (кривая с обозначением «Raw spectra + MCR»).

Стабильность полученных классификационных моделей проверялась k-кратной перекрестной валидацией (кросс-валидацией).

Анализируя полученные ROC кривые, можно отметить, что в случае применения метода логистической регрессии использование объединенных спектральных признаков и концентрационных профилей не приводит к повышению эффективности классификации, значения ROC AUC в некоторых случаях становятся даже ниже, чем в моделях классификации, где в качестве признаков использовались только концентрационные профили (см. рисунок 3.12, а), слева). Так, значения ROC AUC объединенных моделей находятся на уровне от 0,66 (0,58-0,73, 95% CI) до 0,68 (0,62-0,75, 95% CI) для случая дискриминации злокачественных и доброкачественных новообразований, от 0,70 (0,58-0,81, 95% CI) до 0,75 (0,64-0,85, 95% CI) для случая дискриминации MM и SK + NE и от 0,80 (0,71-0,88, 95% CI) до 0,84 (0,75-0,94, 95% CI) для случая дискриминации MM и SK. Аналогичная ситуация наблюдается и при дискриминации новообразований методом PLS-DA. Так, значения ROC AUC объединенных моделей PLS-DA находятся на уровне от 0,62 (0,57-0,67, 95% CI) до 0,64 (0,57-0,70, 95% CI) для случая дискриминации злокачественных и доброкачественных новообразований, от 0,63 (0,46-0,81, 95% CI) до 0,72 (0,59-0,86, 95% CI) для случая дискриминации MM и SK + NE и от 0,70 (0,59-0,81, 95% CI) до 0,81 (0,70-0,92, 95% CI) для случая дискриминации MM и SK.

Также стоит сказать, что сложно выделить, в какой из четырех спектроскопических задач классификация производится более эффективно, так как ROC AUC имеют практически сравнимые значения в каждом из рассмотренных спектральных диапазонов: предобработанные спектры 1200-1800 cm^{-1} («Raman»), или полные экспериментальные спектры в диапазоне 300-1800 cm^{-1} («Raw spectra»). Это позволяет сделать вывод, что при использовании портативной установки основной вклад в модель классификации дает диапазон спектра 1200-1800 cm^{-1} , по которому восстанавливались концентрационные профили при MCR анализе.

Это позволяет сделать вывод, что MCR компоненты содержат достаточно полную информацию о RS спектре для дискриминации новообразований и нет необходимости вносить в модели классификации дополнительные признаки помимо относительных концентраций, выделенных в результате MCR анализа. Однако стоит отметить, что описанный выше метод логистической регрессии является достаточно примитивным для задач бинарной классификации, и его эффективность невысока. Поэтому логичным развитием представленного исследования является применение к новым признакам современных методов машинного обучения.

3.3.3 Классификация компонентных профилей новообразований методом градиентного бустинга

Особенностью данных, анализируемых в данном диссертационном исследовании, является их малый объем вследствие трудности формирования выборки, редкости некоторых заболеваний и т.д. Одним из способов работы с малыми данными разного типа является алгоритм градиентного бустинга (GBM) – ансамблевый алгоритм на основе моделей дерева решений, последовательно добавляемых в ансамбль для исправления ошибок, сделанных предыдущими моделями. Другими словами, алгоритм строит предсказание в виде ансамбля слабых предсказывающих моделей, из которых в итоге формируется одна, но уже более эффективная [226].

Было предложено использовать для реализации алгоритма градиентного бустинга его реализацию LightGBM [227]. В данной реализации для ускорения обучения и уменьшения вычислительной сложности метода применяется градиентная односторонняя (GOSS) – модификация градиентного бустинга, фокусирующая внимание на образцах, приводящих к большему градиенту. Это позволяет исключить большую долю образцов данных с небольшими градиентами. Помимо GOSS, в реализации LightGBM применяется один из подходов автоматического подбора признаков – объединение взаимоисключающих признаков (EFB). Описанные подходы помогают ускорить обучение алгоритма до 20 раз.

Известно успешное применение алгоритма LightGBM для диагностики диабета [228, 229], рака легких [230], заболеваний кожи [231]. В данном диссертационном исследовании метод LightGBM был применен как к модели «MCR», использующей только концентрационные профили, так и к четырем объединенным спектроскопическим моделям («AF + MCR», «Raman + MCR», «Raman + AF + MCR», «Raw spectra + MCR»), рассмотренным выше. Были исследованы все уже упомянутые случаи классификации: а) злокачественные и доброкачественные новообразования; б) MM и SK + NE; в) MM и SK. Стабильность полученных классификационных моделей проверялась k-кратной перекрестной валидацией (кросс-валидацией).

Метод LightGBM позволяет настроить ряд гиперпараметров, основными из которых являются количество деревьев и глубина дерева. Количество деревьев – это количество предсказывающих моделей, используемых в ансамбле. Глубина дерева определяет, насколько каждое дерево специализируется на обучающем наборе данных. Градиентный бустинг обычно хорошо работает с деревьями, имеющими умеренную глубину, находящую баланс между обученностью и обобщенностью. Существует два основных способа управления сложностью дерева: через максимальную глубину дерева и максимальное количество терминальных узлов (листьев) дерева. В результате подбора гиперпараметров для решения задачи

дискриминации новообразований максимальное количество листьев было установлено равным 30, а максимальное количество деревьев – 100.

На рисунке 3.13 представлены ROC-кривые для каждого случая дискриминации с использованием алгоритма LightGBM.

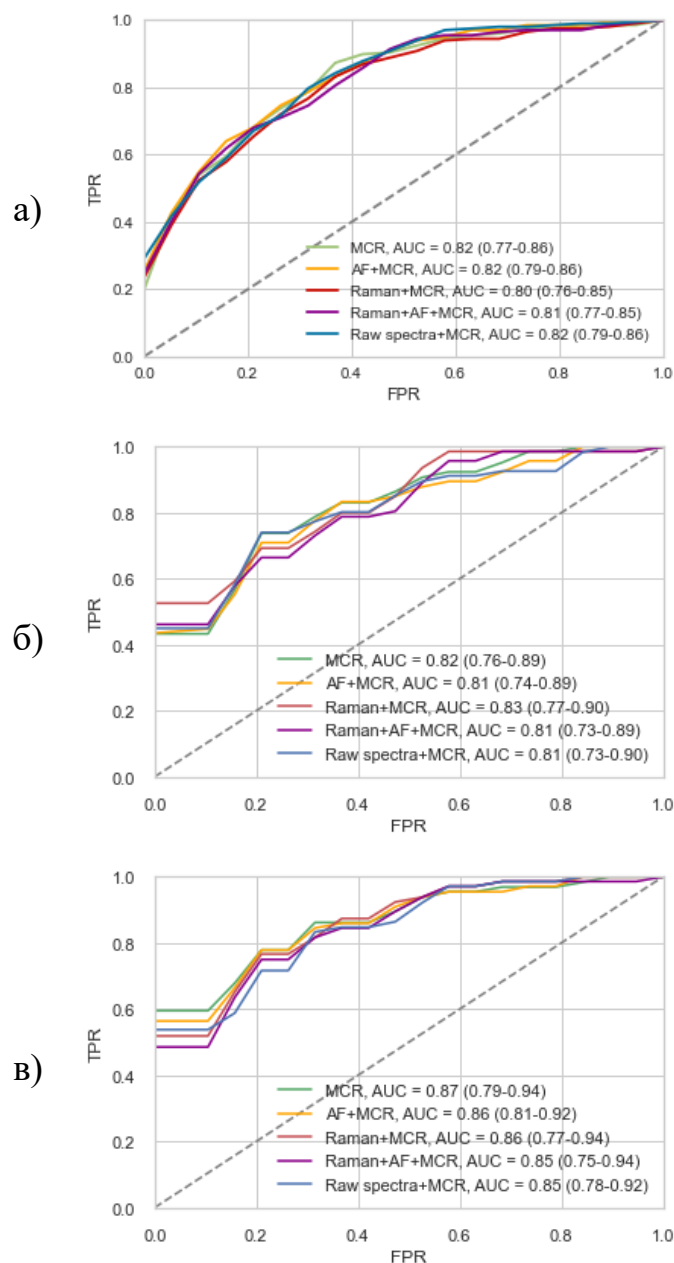


Рисунок 3.13 – Классификация на основе анализа LightGBM: а) злокачественные и доброкачественные новообразования; б) ММ и SK + NE; в) ММ и SK. Доверительные интервалы ROC AUC рассчитаны для уровня вероятности не менее 95%

В сводной таблице 3.4 представлены данные о площади под ROC кривой для перечисленных выше случаев классификации, и значения ROC AUC, полученные в исследовании [5]. Матрица ошибок и метрики качества модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) методом LightGBM на основе информации об относительном компонентном составе представлены в таблицах 3.5 и 3.6.

Таблица 3.5 – Матрица ошибок модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) методом LightGBM на основе информации об относительном компонентном составе

	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	52	13
$y = 1$	15	50

Таблица 3.6 – Метрики качества модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) методом LightGBM на основе информации об относительном компонентном составе

Метрика качества	Обозначение	Значение
Доля правильных ответов (accuracy)	ACC	0,78
Точность (precision)	PPV	0,79
Полнота (recall), чувствительность	TPR	0,77
Специфичность	TNR	0,80
Точность отрицательного прогноза	NPV	0,78
F1-мера	F1	0,78

Возвращаясь к рисунку 3.13, совершенно очевидно, что применение ансамблевого метода LightGBM позволяет увеличить эффективность дискриминации новообразований кожи. Так, для случая классификации злокачественных и доброкачественных опухолей ROC AUC составляет от 0,80 (0,76-0,85, 95% CI) до 0,82 (0,79-0,86, 95% CI) в зависимости от используемых признаков, что превышает значение ROC AUC, полученное в [5], на 5-7%. В случае

классификации MM и SK + NE площадь под кривой также возрастает по сравнению со значением 0,69 (0,63-0,76, 95% CI) [5] на 12-14% и составляет от 0,81 (0,73-0,89, 95% CI) до 0,83 (0,77-0,90, 95% CI). В случае классификации MM и SK площадь под кривой составляет от 0,85 (0,75-0,94, 95% CI) до 0,87 (0,79-0,94, 95% CI), что выше, чем результаты статьи [5], на 4-6%.

Сравнивая ROC кривые для разных спектроскопических задач в пределах одного случая классификации, можно заметить, что ROC AUC отличаются незначительно, что еще раз подтверждает выводы о том, что MCR компоненты содержат достаточно полную информацию о RS спектре. Добавление в модель дополнительных признаков в виде спектральных отсчетов не влияет на ее эффективность. Полученные модели классификации демонстрируют высокую устойчивость, о чем свидетельствуют малые доверительные интервалы ROC AUC (по сравнению с доверительными интервалами для моделей, получаемых с помощью метода PLS-DA и логистической регрессии).

На основе описанных выше алгоритмов был реализован программный модуль на языке Python для классификации рамановских спектров патологий кожи [232*]. Программа предназначена для классификации рамановских спектров кожных патологий, полученных с помощью портативной рамановской установки, описанной во второй главе. Задача классификации решается с помощью извлечения признаков алгоритмом градиентного бустинга. Для обучения классификатора на вход принимаются размеченные данные об относительных концентрациях компонентов в образцах, полученных на основании MCR анализа рамановских спектров. Программа формирует обучающую и тестовую выборки в соотношении 3:1. Также в программе реализована возможность предсказания класса новообразования на основе анализа его данных об относительном компонентом составе предварительно обученными моделями классификации здоровой кожи против кожи с заболеванием, злокачественных новообразований против доброкачественных, злокачественной меланомы против пигментных новообразований [233*, 234*]. Программа написана на языке Python версии 3.8.5, совместима с ЭВМ типа IBM-PC и операционной системой Windows.

3.4 Выводы по главе

1. Проведенное исследование и сравнение различных подходов к анализу спектральных данных оптической биопсии показало, что спектральные компоненты кожи, полученные в результате MCR разложения рамановских спектров, являются физически интерпретируемыми, а построенные на основании них концентрационные профили новообразований позволяют использовать их в качестве признаков классификации.

2. Предложенный подход к анализу спектральных данных рамановского рассеяния на основании компонентных профилей обеспечивает эффективность предсказания и дискриминации различных типов новообразований, сравнимую с эффективностью квалифицированных дерматологов [1].

3. Автоматизация процесса распознавания злокачественной меланомы кожи при сохранении возможности физической интерпретации получаемых данных об составе исследуемого участка кожи делает предложенный в этой главе алгоритм анализа спектральных данных оптической биопсии претендентом на широкое использование в системах контроля принятия решения, особенно для врачей общей практики.

4. Не смотря на превосходство модели над результатами ранее проведенных исследований, значения ROC AUC не превышают 0,90, что ниже, чем результаты гистологического исследования. Для повышения эффективности распознавания новообразований кожи необходимо учитывать дополнительные качественно иные признаки, например, пространственные и текстурные признаки дерматоскопических изображений.

4 ИНТЕЛЛЕКТУАЛЬНЫЙ МЕТОД АНАЛИЗА ДЕРМАТОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ НОВООБРАЗОВАНИЙ КОЖИ

В данной главе предложен способ анализа данных оптической биопсии для *in vivo* диагностики и определения типа опухолей кожи в клинической практике на основе применения искусственного интеллекта к дерматоскопическим изображениям, зарегистрированным с использованием цифрового дерматоскопа [175]. Для обучения нейронной сети необходима достаточно большая обучающая выборка изображений, и изображений набора СОКОД, зарегистрированных мультиспектральным цифровым дерматоскопом, для этой задачи недостаточно. В большинстве работ по анализу дерматоскопических изображений, описанных в обзорном исследовании [13], проблема недостатка экспериментальных данных решается использованием различных типов предварительно обученных сверточных нейронных сетей. В этой работе для повышения эффективности работы нейронной сети применяется трансферное обучение и аугментация данных.

4.1 Архитектура сверточной нейронной сети для классификации дерматоскопических изображений

Искусственные нейронные сети – это один из методов искусственного интеллекта, обучающий компьютеры проводить обработку данных способом, напоминающим работу человеческого мозга. Основными объектами для глубокого обучения являются слои, которые затем объединяются в сеть, исходные данные и соответствующие им истинные значения, функция потерь, определяющая сигнал обратной связи в ходе обучения, и оптимизатор, проводящий корректировку весов модели [235]. На первом этапе извлекается пакет обучающих примеров с их истинными значениями, после чего сеть обрабатывает пакет данных и получает пакет с предсказаниями. На следующем этапе происходит вычисление потерь. Другими словами, оцениваются несовпадения между истинными значениями и

предсказанными, и в зависимости от полученных результатов веса сети корректируются таким образом, чтобы уменьшить потери при обучении [236].

В полносвязных нейронных сетях прямого распространения каждый нейрон передает свою информацию всем остальным нейронам следующего слоя. Входной слой принимает данные в виде вектора. Каждый элемент вектора соответствует одному из признаков входных данных. Значения нейронов каждого слоя вычисляются путем умножения значений предыдущего слоя на веса, добавления смещения, применения функции активации и передачи значений следующему слою. После прохождения через все скрытые слои, данные попадают на выходной слой, где они используются для решения конкретной задачи [236]. На практике этот тип сети имеет некоторые проблемы с точки зрения распознавания и классификации изображений, так как модель не учитывает пространственную структуру поступающих на вход данных, что приводит к потере части информации. Также из-за большого числа нейронов в слоях возрастает количество параметров, используемых в процессе обучения, что может привести к переобучению. Модель будет достигать большой точности на тренировочном наборе данных, но плохо распознавать тестовые примеры [237].

Для обработки изображений разработаны специальные сверточные нейронные сети (CNN), базирующиеся на использовании операции свертки в процессе обучения. Данная математическая операция помогает извлекать определенные признаки [238]. В отличие от полносвязных слоев, сверточные слои распознают локальный контекст на изображении, как это показано на рисунке 4.1 [239].

Свертка применяется к трехмерным тензорам. Для изображений в формате RGB значение третьей оси глубины равно трем. Каждый фильтр находит определенный признак, после чего создает карту признака, демонстрирующую, на каком участке изображения он был обнаружен. Затем собранная информация передается на следующий слой для дальнейшей обработки.

CNN успешно применяются для распознавания и классификации изображений [240, 241, 242*]. CNN является методом обучения с учителем, то есть обучение производится с использованием заранее размеченных данных. По сути, CNN

изучают взаимосвязь между входными объектами и метками классов и состоят из двух компонентов: скрытых слоев, в которых извлекаются функции, и полносвязных слоев, которые используются для фактической задачи классификации. Для успешного обучения CNN необходим большой объем размеченных классами данных. В случае обучения CNN на выборках небольшого объема возникает риск переобучения и, как следствие, неадекватное обобщение сети для неизвестных входных данных [243].

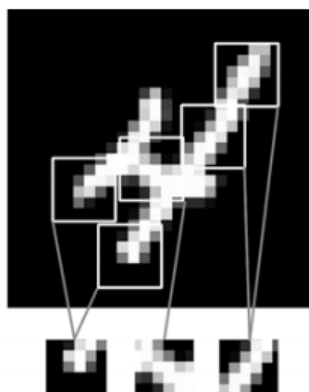


Рисунок 4.1 – Разбиение изображения на локальные шаблоны, иллюстрирующее выделение признаков сверточной нейронной сетью [239]

В представленной работе используется сверточная нейронная сеть на основе модифицированной архитектуры VGG16 (см. рисунок 4.2) [244*, 245*]. Выбор архитектуры основывался на анализе результатов работы [135] и обзорного исследования [13]. Согласно авторам, в настоящее время VGG является одной из наиболее часто применяемых архитектур для анализа дерматоскопических изображений. Однако выборка дерматоскопических изображений, зарегистрированных с помощью цифрового дерматоскопа, имеет небольшой объем (см. раздел 2.2), поэтому в целях избежания переобучения и нестабильности работы нейронной сети в архитектуре VGG16 было уменьшено количество сверточных слоев.



Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 32)	896
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0
conv2d_1 (Conv2D)	(None, 112, 112, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0
conv2d_2 (Conv2D)	(None, 56, 56, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 64)	0
conv2d_3 (Conv2D)	(None, 28, 28, 128)	73856
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 128)	0
conv2d_4 (Conv2D)	(None, 14, 14, 128)	147584
max_pooling2d_4 (MaxPooling2D)	(None, 7, 7, 128)	0
flatten (Flatten)	(None, 6272)	0
dropout (Dropout)	(None, 6272)	0
dense (Dense)	(None, 128)	802944
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 1)	129
=====		
Total params: 1,097,345		
Trainable params: 1,097,345		
Non-trainable params: 0		

Рисунок 4.2 – Архитектура сверточной нейронной сети

На вход нейронной сети поступают дерматоскопические изображения размерностью 224 x 224 пикселя. Как показано на рисунке 4.2, классификатор состоит из пяти чередующихся блоков слоев Conv2D и MaxPooling2D (с функцией активации ReLU) и полносвязных слоев Dense на выходе. В начале нейронная сеть работает с трехмерным тензором, а после чего происходит его преобразование в

одномерный. Слой сверточной нейронной сети Conv2D проводит операцию свертки, которая работает методом скользящего окна.

В данной работе используются окна размером 3×3 . Перемещаясь по карте признаков, окно извлекает трехмерный шаблон признаков с определенной формой (высота окна, ширина окна, глубина входа). Каждый, из полученных шаблонов при помощи операции умножения тензора на матрицу весов, получаемую в ходе обучения, преобразуется в одномерный вектор. Все полученные векторы собираются в трехмерную выходную карту.

Первый слой Conv2D на рисунке 4.2 имеет 32 фильтра, второй и третий – 64 фильтра, а четвертый и пятый 128 – фильтров. Второй слой MaxPooling2D, находящийся в чередующихся стеках архитектуры сети, необходим для уменьшения разрешения карты признаков, что позволяет сократить число обучаемых параметров. Принцип работы MaxPooling2D основан на операции выбора максимального значения из извлечённого окна входной карты признаков. В данной работе размер окон равен 2×2 , что позволяет уменьшить разрешение в два раза. На выходе общее число параметров обучения составляет 1 097 345. В качестве функции активации на выходном слое и оптимизатора эмпирическим путем выбраны сигмоида и Adam, соответственно. Для реализации нейронной сети на языке Python использовалась интерактивная оболочка Jupyter Notebook версии 6.4.12, широко применяющаяся для машинного обучения и работы, связанной с анализом данных, и открытые программные библиотеки для машинного обучения TensorFlow версии 2.3.0 и Keras версии 2.4.3.

4.2 Идентификация новообразований кожи по дерматоскопическим изображениям

Анализ пространственных характеристик дерматоскопических изображений будем проводить на примере одной классификационной задачи по выявлению наиболее опасной формы новообразований – дискриминации злокачественной

меланомы (ММ) и доброкачественных пигментных новообразований кожи (SK + NE).

Для использования описанной в предыдущем пункте нейронной сети анализируемые данные необходимо разделить на три выборки: обучающую, валидационную и тестовую. Обучающая выборка будет использоваться непосредственно для обучения классификатора, валидационная – используется в качестве сигнала обратной связи, чтобы проводить настройку параметров модели в процессе обучения, тестовая – необходима для проверки качества классификатора. Данные делились на обучающую, валидационную и тестовую выборки в соотношении 8:1:1. Стабильность полученных классификационных моделей проверялась k-кратной перекрестной валидацией (кросс-валидацией).

Поскольку, согласно данным второй главы, экспериментально зарегистрированные дерматоскопические изображения содержат несбалансированное количество разных типов новообразований, то при формировании обучающей, валидационной и тестовой выборки классы балансировались следующим образом. В выборку меньшего класса попадали все имеющиеся образцы, а выборка большего класса формировалась случайным образом из имеющихся образцов этого класса. Кроме того, внутри каждого класса каждой выборки сохранялось примерное соотношение диагнозов, представленное в исходном наборе данных. Учитывая случайный характер формирования выборки, построение моделей классификации было проведено не менее 10 раз, и далее представлены результаты, полученные в результате усреднения метрик моделей по количеству проведенных экспериментов. Таким образом, на каждом этапе построения модели дискриминации ММ и доброкачественных пигментных новообразований использовалось 65 ММ, 26 SK и 39 NE, случайно взятые из набора данных СОКОД, описанном во второй главе.

Как было отмечено ранее, по причине малого количества исходных данных в исследовании используется дополнительный набор дерматоскопических изображений – свободно распространяемый набор изображений HAM10K [149], который включает 1113 изображений злокачественной меланомы (ММ), 6705

изображений невусов (NE) (всех типов) и 1099 изображений себорейного кератоза (SK). Поскольку в данном наборе классы также являются несбалансированными, количество изображений меньшего класса (MM) было увеличено при помощи различных геометрических преобразований, например, вращения на случайный угол, сдвигов вправо или влево, приближения и т. д. В результате аугментации общее число изображений MM стало равным 7804. Примеры аугментированных изображений приведены на рисунке 4.3.

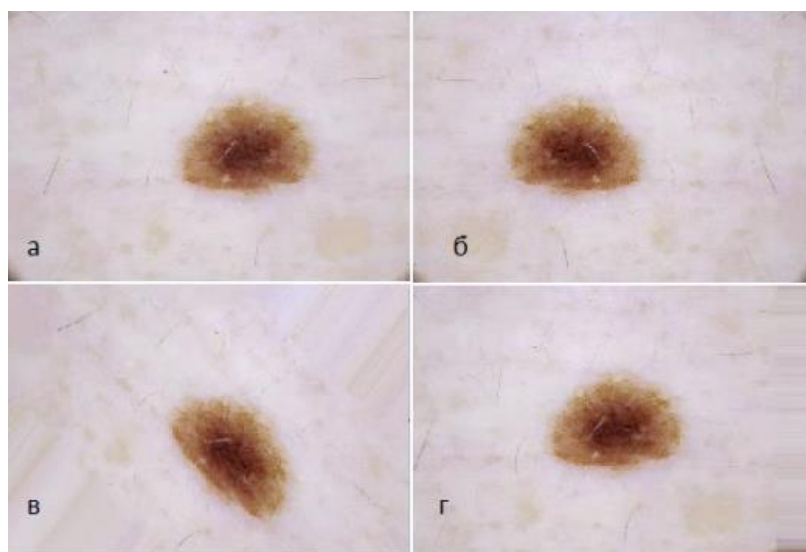


Рисунок 4.3 – Иллюстрация аугментации изображений из набора HAM10K (а – исходное изображение; б – изображение, отраженное по горизонтали; в – изображение, повернутое на угол 45 градусов; г – изображение, смещенное влево)

В случае набора данных СОКОД также осуществляется аугментация обоих классов. Она проходит в момент дообучения модели: на вход нейронной сети поступают изображения из обучающей выборки, и производится их аугментация (сдвиг, поворот, масштабирование, отражение). Таким образом, модель не сможет увидеть одни и те же новообразования дважды. Описанный метод аугментации данных осуществляется при помощи класса DataGenerator из библиотеки Keras [246]. Он позволяет загружать данные пакетами (батчами) и передавать их на вход сверточной нейронной сети, что экономит объем используемой памяти и ускоряет процесс обучения. В работе не применялась аугментация цветом, так как это может

исказать цветовые признаки на изображении: наличие сосудистых сетей и бело-голубой вуали. По этой же причине, а также на основании результатов ранее проведенных исследований [136*, 245*] было решено отказаться от предварительной цветокоррекции изображений набора СОКОД, так как она не привела к существенному улучшению распознавания изображений. Описание данных, используемых при обучении и тестировании нейронной сети, представлено в таблице 4.1.

Таблица 4.1 – Описание данных, используемых при обучении и тестировании нейронной сети

Изображения	НАМ10К	СОКОД
ММ (до аугментации)	1 133	65
ММ (после аугментации)	7 804	65
SK + NE (до аугментации)	7 804	65
SK + NE (после аугментации)	7 804	65
Всего	15 608	130
Обучающая (тренировочная) выборка	12 486	104 → 1 040*
Валидационная выборка	1 561	13
Тестовая выборка	1 561	13

*После аугментации обучающей выборки (сдвиг, поворот, масштабирование, отражение)

Успешность обучения глубоких моделей CNN определяется наличием достаточного количества размеченных обучающих данных. Так, например, авторы работы [10] использовали в своем исследовании 129 450 изображений, из которых 3374 были получены с дерматоскопических устройств. В данном диссертационном исследовании выборка изображений новообразований кожи, полученных с помощью опытного образца цифрового дерматоскопа [175], гораздо меньше. В случае классификации ММ против других пигментных новообразований существует возможность использовать сторонний набор данных [149], поэтому в качестве предмета исследований в данной главе выступает именно этот случай.

В качестве архитектуры нейронной сети используется архитектура, описанная в предыдущем подразделе. Процесс обучения построен на концепции трансферного

обучения. Трансферное обучение – это метод машинного обучения, при котором модель, разработанная для какой-либо задачи, повторно используется в качестве отправной точки для модели для второй задачи. Это позволяет снизить вычислительные и временные ресурсы, необходимые для обучения моделей нейронных сетей, и использовать уже накопленные нейросетью знания [247].

Успешное применение трансферного обучения было продемонстрировано в исследованиях, связанных с анализом медицинских изображений [248, 249], а также при анализе рака кожи [250]. Трансферное обучение особенно полезно, когда исследователям доступен небольшой набор обучающих данных. В этом случае для инициализации весов новой модели можно использовать веса модели, предварительно обученной на других, похожих данных [247].

В данной работе концепция трансферного обучения реализуется следующим образом. Вначале обучение нейронной сети проводится на данных HAM10K, далее нейронная сеть дообучается на втором наборе данных – СОКОД. Причем дообучение проводится в условиях, когда четыре верхних слоя нейронной сети заморожены [251*]. Замораживание слоев необходимо для того, чтобы сохранить общие признаки, выделенные классификатором ранее, и полученные веса. Такой подход позволяет учесть технические особенности используемого цифрового дерматоскопа, исключив возможные артефакты, возникающие в процессе его работы.

Оценка качества классификации проводилась с использованием ROC кривой, построенной для тестовой выборки, и ROC AUC. Для оценки процесса обучения используется график доли правильных ответов, которая рассчитывается в конце каждой эпохи обучения для тренировочной и валидационной выборки. Контролировать долю правильных ответов для валидационной выборки необходимо для избежания переобучения модели.

Обучение проводилось на компьютере с процессором Intel(R) Core(TM) i7-10700KF CPU @ 3.80GHz, оперативной памятью 32 ГБ и видеокартой GeForce RTX 3080.

На первом этапе нейронная сеть обучалась 20 эпох с использованием набора данных NAM10K. Оптимальное количество эпох было выявлено экспериментально на основании анализа графиков доли правильных ответов для тренировочной и валидационной выборок. Графики доли правильных ответов для случая 20 эпох обучения представлены на рисунке 4.4.

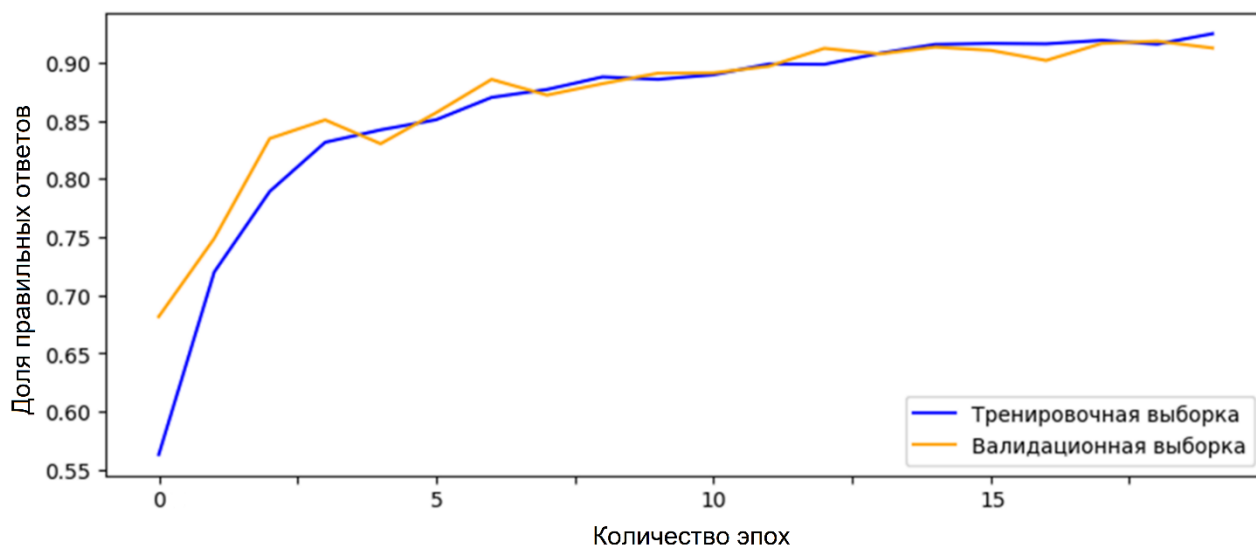


Рисунок 4.4 – Графики доли правильных ответов модели, обученной на наборе данных NAM10K, рассчитанные для тренировочной и валидационной выборки

Из рисунка 4.4 видно, что модель обучается плавно, число верно классифицированных изображений увеличивается. Графики не имеют выбросов, так как классы сбалансированы и имеют большой объем. ROC кривая модели классификации, рассчитанная на тестовой выборке, представлена на рисунке 4.5. ROC AUC составляет 0,97.

Для того, чтобы понять, на что обученная сверточная нейронная сеть обращает внимание при формировании прогноза, можно обратиться к карте внимания, представленной на рисунке 4.6.

На рисунке 4.6 заметно, что обученная нейронная сеть обращает свое внимание на границу новообразования, что свидетельствует о том, что она проводит анализ размера, неоднородности и формы объекта, на которые

впоследствии ориентируется в процессе принятия решения. По сути, нейронная сеть проводит анализ новообразования по уже упомянутому ранее критерию ABCD.

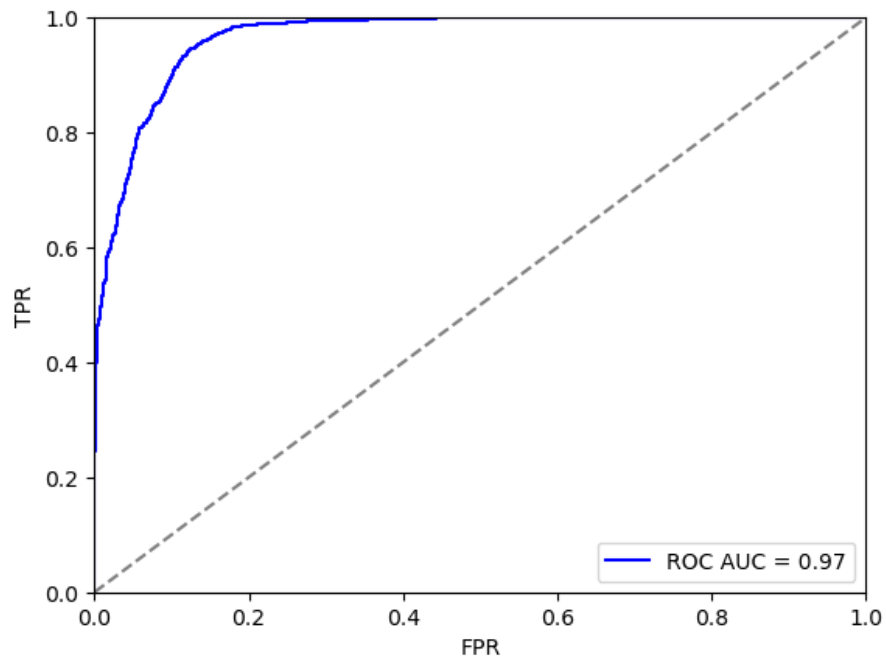


Рисунок 4.5 – ROC кривая модели, обученной на наборе данных HAM10K, рассчитанная для тестовой выборки

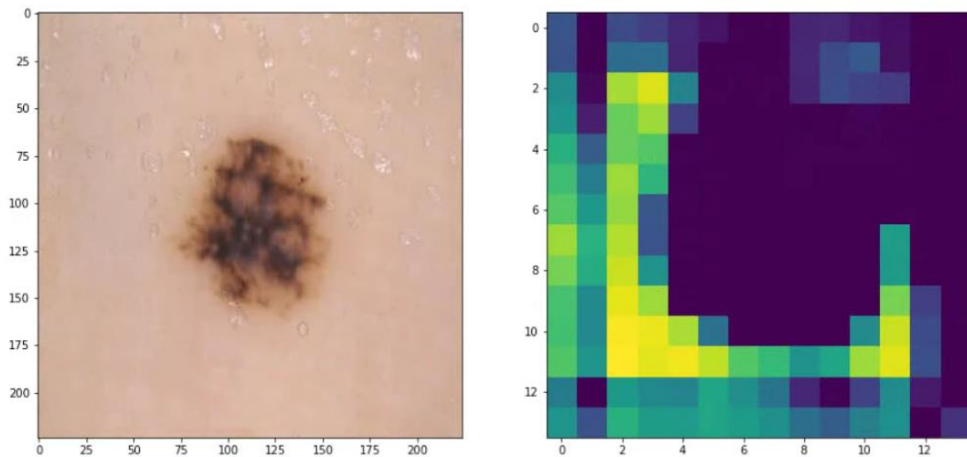


Рисунок 4.6 – Карта внимания для изображения из тестовой выборки набора данных HAM10K

На втором этапе нейронная сеть с замороженными весами дообучалась 300 эпох с использованием набора данных СОКОД. Оптимальное количество эпох было

выявлено экспериментально на основании анализа графиков доли правильных ответов для тренировочной и валидационной выборок. Графики доли правильных ответов для случая 300 эпох обучения представлены на рисунке 4.7.

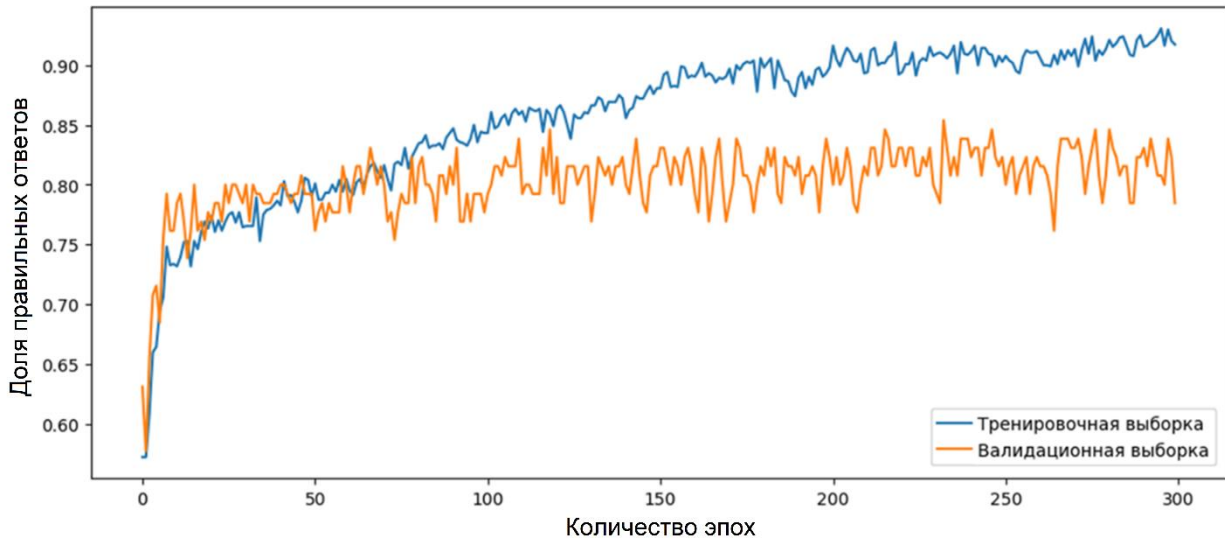


Рисунок 4.7 – Графики доли правильных ответов модели, обученной на наборе данных СОКОД, рассчитанные для тренировочной и валидационной выборок

Из рисунка 4.7 прослеживается общий тренд к увеличению доли правильных ответов для обеих выборок, однако модель ведет себя несколько нестабильно, что может быть обусловлено малым объемом используемых для дообучения данных. ROC кривая модели классификации, рассчитанная на тестовой выборке, представлена на рисунке 4.8. ROC AUC составляет 0,87 (0,82-0,93, 95% CI). Карта внимания для изображения из тестовой выборки набора данных СОКОД представлена на рисунке 4.9.

Анализируя рисунок 4.9, можно обратить внимание, что на этапе дообучения модель начинает обращать внимание на само новообразование, а именно оценивать не только его форму и размер, но и рельеф. Учитывая текстурные признаки, нейронная сеть анализирует неоднородность роста новообразования.

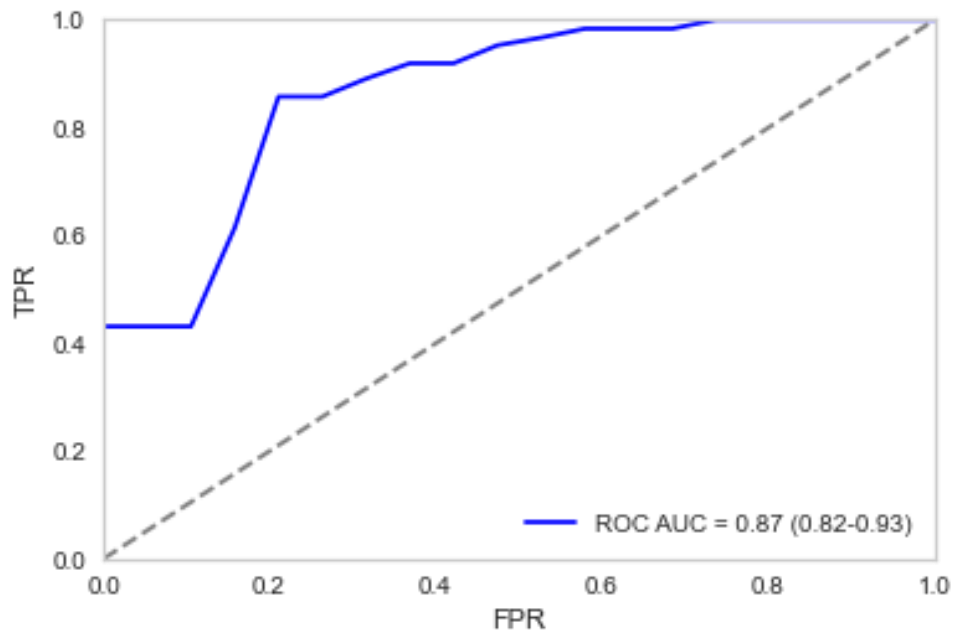


Рисунок 4.8 – ROC кривая модели, обученной на наборе данных СОКОД, рассчитанная для тестовой выборки. Доверительные интервалы ROC AUC рассчитаны для уровня вероятности не менее 95%

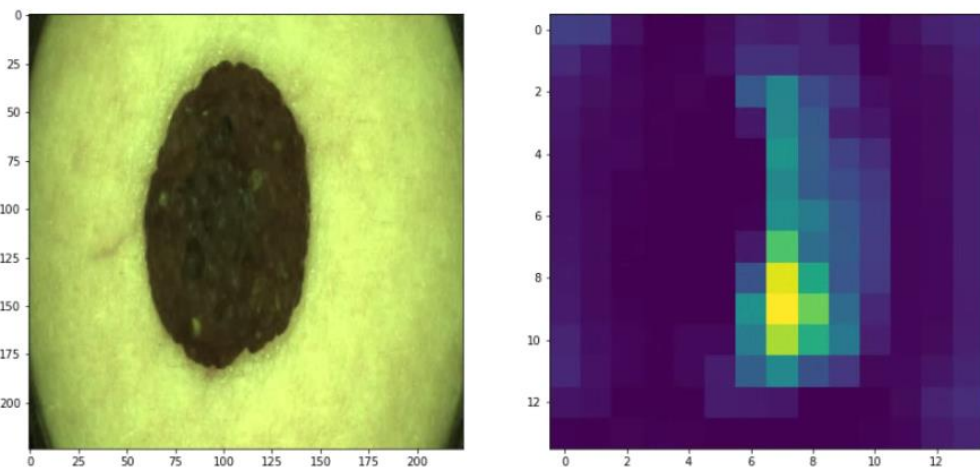


Рисунок 4.9 – Карта внимания для изображения из тестовой выборки набора данных СОКОД

Произведем полноценную оценку качества этого классификатора. Матрица ошибок и метрики качества модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) с помощью нейронной сети представлены в таблицах 4.2 и 4.3.

Таблица 4.2 – Матрица ошибок модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) с помощью нейронной сети

	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	55	10
$y = 1$	10	55

Таблица 4.3 – Метрики качества модели бинарной классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) с помощью нейронной сети

Метрика качества	Обозначение	Значение
Доля правильных ответов (accuracy)	ACC	0,85
Точность (precision)	PPV	0,85
Полнота (recall), чувствительность	TPR	0,85
Специфичность	TNR	0,85
Точность отрицательного прогноза	NPV	0,85
F1-мера	F1	0,85

Анализируя ROC кривую и представленные выше метрики, можно отметить хорошую способность модели классифицировать ММ и SK + NE. Так, для случая классификации ММ и SK + NE с помощью нейронной сети ROC AUC составляет 0,87 (0,82-0,93, 95% CI), что несколько выше ROC AUC аналогичной модели классификации на основе метода LightGBM с использованием концентрационного профиля, описанной в третьей главе, где ROC AUC составила 0,82 (0,76-0,89, 95% CI). Применение сверточной нейронной сети в комбинации с трансферным обучением и использованием стороннего набора данных для предварительного обучения сети позволяет получить устойчивую модель классификации, демонстрирующую точность на уровне 85%, что соизмеримо с результатами других исследований, посвященных распознаванию злокачественной меланомы с помощью анализа дерматоскопических изображений сверточными нейронными сетями [135].

Для повышения эффективности распознавания злокачественной меланомы может быть полезно дообучать текущую сохраненную реализацию нейронной сети на новых зарегистрированных цифровым дерматоскопом примерах.

4.3 Выводы по главе

1. Реализована и обучена сверточная нейронная сеть для анализа дерматоскопических изображений и дискриминации злокачественной меланомы и доброкачественных пигментных новообразований. Применен прием трансферного обучения, что позволило обучить нейронную сеть в условиях малого набора данных. Показано, что обученная нейронная сеть обращает свое внимание на пространственные характеристики новообразования, а не на его контуры, делая прогноз на основании текстурных признаков изображения.

3. Сравнение модели с результатами других исследований показало, что представленная модель не уступает по метрикам качества другими интеллектуальным моделям классификации с использованием нейронных сетей, обученным на наборе данных аналогичного размера.

3. Предлагаемый в этой главе подход к анализу дерматоскопических изображений позволяет выделять информативные пространственные признаки новообразований. Возможность автоматизация процедуры интеллектуального анализа дерматоскопического изображения практически в реальном режиме времени делает обученный нейросетевой классификатор претендентом на широкое использование в системах контроля принятия решения.

5 МУЛЬТИМОДАЛЬНЫЙ АНАЛИЗ СПЕКТРАЛЬНЫХ И ПРОСТРАНСТВЕННЫХ ПРИЗНАКОВ НОВООБРАЗОВАНИЙ

В предыдущих главах были разработаны два способа классификации новообразований: на основе анализа концентрационных профилей спектральных компонентов, выделенных в результате MCR разложения рамановских спектров, и интеллектуального анализа дерматоскопических изображений с помощью сверточной нейронной сети. Поскольку рамановские спектры кожи и дерматоскопические изображения содержат совершенно разные по своей этиологии признаки (изменение биохимического состава, плотности и пространственной структуры), естественно ожидать повышения метрик при объединении моделей классификации благодаря расширению набора анализируемых признаков. Исследование такого мультимодального подхода и является целью данной главы.

5.1 Идентификация новообразований кожи на основе мультимодального совместного анализа рамановских спектров и дерматоскопических изображений

В данной работе мультимодальный совместный анализ спектральных данных рамановского рассеяния и пространственных характеристик дерматоскопических изображений исследуется на примере классификационной задачи дискриминации MM и доброкачественных пигментных новообразований кожи (SK + NE). Отметим, что регистрация RS спектров и дерматоскопических изображений новообразований производилась одномоментно, что является основанием для построения многомерной модели идентификации новообразований.

Для реализации мультимодального метода необходимо понять, как именно можно объединить качественно различающиеся мультимодальные данные, имеющий разный порядок (двумерные цветные изображения и одномерные спектральные признаки). Самым простым и наиболее эффективным способом является объединение моделей классификации на самом последнем этапе, когда

модель получает предсказанные значения для каждого наблюдения. Для этой цели в работе применяется простейший способ анализа взаимосвязи между несколькими независимыми переменными и зависимой переменной – логистическая регрессия. Схема построения модели классификации, реализующей метод мультимодального анализа рамановского рассеяния и дерматоскопических изображений, представлена на рисунке 5.1.

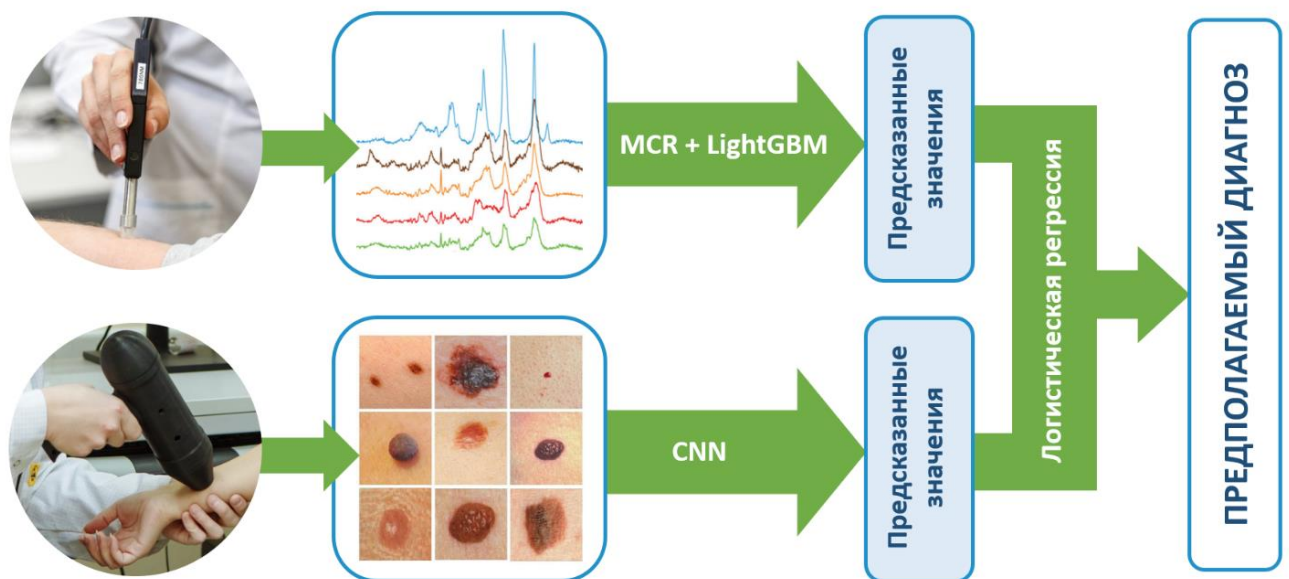


Рисунок 5.1 – Схема построения модели классификации, реализующей метод мультимодального анализа рамановского рассеяния и дерматоскопических изображений

Стабильность полученной объединенной классификационной модели проверялась k -кратной перекрестной валидацией (кросс-валидацией). На рисунке 5.2 представлена ROC-кривая полученной модели классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE). Матрица ошибок и метрики качества модели представлены в таблицах 5.1 и 5.2.

Анализируя представленные метрики, можно отметить хорошую способность модели классификации методом многомерного анализа различать ММ и SK + NE. Однако стоит отметить, что метрики качества, рассчитанные для порогового значения логистической регрессии 0,5, несколько уступают результатам,

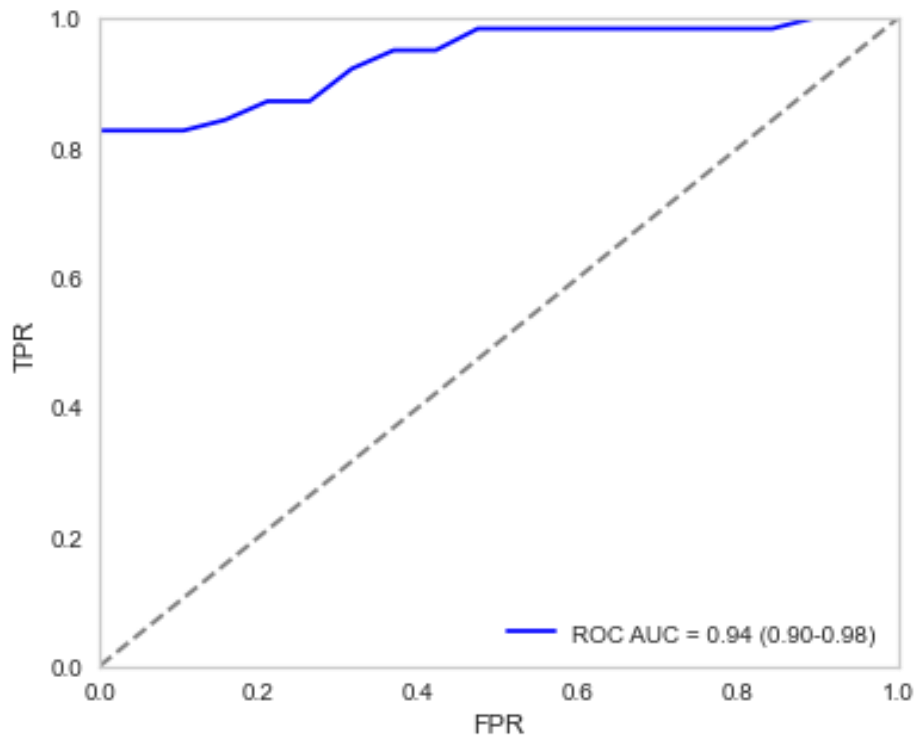


Рисунок 5.2 – ROC кривая модели классификации на основе многомерного анализа RS спектров и дерматоскопических изображений, рассчитанная для тестовой выборки. Доверительные интервалы ROC AUC рассчитаны для уровня вероятности не менее 95%

представленным в четвертой главе, где классификация производилась путем анализа дерматоскопических изображений с помощью CNN. Тем не менее, ROC AUC модели на основе многомерного анализа составляет 0,94 (0,90-0,98, 95% CI), что значительно выше по сравнению со значением 0,69 (0,63-0,76, 95% CI) из работы [5], а также превосходит ROC AUC моделей на основе анализа RS спектров методом LightGBM и на основе анализа изображений CNN, описанных в третьей и четвертой главах, где ROC AUC составили 0,82 (0,76-0,89, 95% CI) и 0,87 (0,78-0,95, 95% CI), соответственно.

Установив порог классификации на уровне 0,67 (две трети от максимально возможного предсказанного значения логистической регрессии), можно добиться специфичности на уровне 1,00. Для рассмотренного случая классификации это будет означать, что модель безошибочно идентифицирует в качестве ММ все образцы, которые действительно являются ММ. Чувствительность при этом равна

0,74, то есть примерно 25% образцов определены моделью как ММ, хотя на самом деле являются доброкачественными пигментными новообразованиями. Полученный результат позволяет говорить об эффективном использовании описанного метода многомерного анализа спектральных и пространственных данных оптической биопсии в клинической практике в программах скрининга населения.

Таблица 5.1 – Матрица ошибок модели классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) на основе многомерного анализа RS спектров и дерматоскопических изображений

	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	55	10
$y = 1$	11	54

Таблица 5.2 – Метрики качества модели классификации ММ и доброкачественных пигментных новообразований кожи (SK + NE) на основе многомерного анализа RS спектров и дерматоскопических изображений

Метрика качества	Обозначение	Значение (порог 0,5)	Значение (порог 0,67)
Доля правильных ответов (accuracy)	ACC	0,84	0,87
Точность (precision)	PPV	0,84	1,00
Полнота (recall), чувствительность	TPR	0,83	0,74
Специфичность	TNR	0,85	1,00
Точность отрицательного прогноза	NPV	0,83	0,79
F1-мера	F1	0,83	0,85

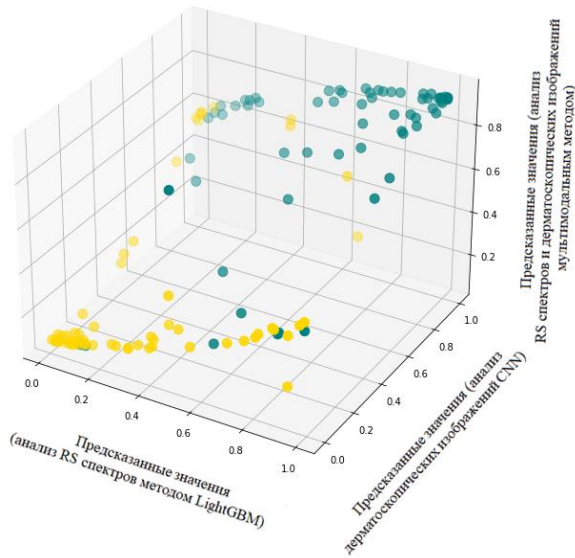
Увеличение ROC AUC модели на основе мультимодального анализа составляет по сравнению с ROC AUC моделей на основе анализа RS спектров методом LightGBM и на основе анализа изображений CNN объясняется возможностью модели использовать пространственные признаки для классификации, когда предсказание по спектральным (компонентным) признакам

не удастся, и наоборот. Это можно продемонстрировать с помощью анализа рисунка 5.3.

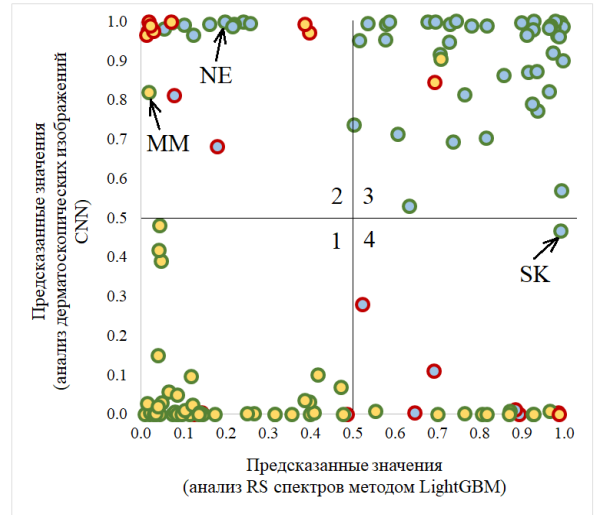
На рисунке 5.3 исследуемые новообразования показаны в трехмерном пространстве предсказанных значений трех моделей: на основе мультимодального анализа, на основе анализа RS спектров методом LightGBM и на основе анализа изображений CNN. Каждый маркер на рисунке 5.3 представляет собой одно из новообразований, а координаты маркера – это предсказанные значения моделей. Цвет маркера указывает на истинный диагноз образца, полученного в результате гистологического исследования: желтый – MM, синий – SK или NE. Предсказанные значения моделей находятся в диапазоне от 0 до 1, где 0 означает класс MM, а 1 означает класс доброкачественных пигментных новообразований (SK + NE). Если предсказанное значение меньше порогового значения (в данном случае 0,5), это означает, что модель классифицирует образец как MM; если предсказанное значение превышает пороговое значение, образец классифицируется как SK + NE. Цвет контура маркера указывает на правильность определения диагноза с помощью модели мультимодального анализа: зеленым обведены образцы, правильно распознанные мультимодальным методом; красным – распознанные неправильно.

Как видно по рисунку 5.3, ошибочной классификации подвергаются только часть пограничных случаев (когда модели на основе анализа RS спектров методом LightGBM и на основе анализа изображений CNN принимают разное решение, то есть, когда одна модель показывает предсказанное значение выше порога классификации, а другая модель – ниже) и случаи, когда обе модели приняли неверное решение. Эти случаи расположены на рисунке 5.3б на графиках во 2 и 4 квадрантах. Диаграммы разброса позволяют наглядно проследить, что происходит с образцами в результате применения метода мультимодального анализа.

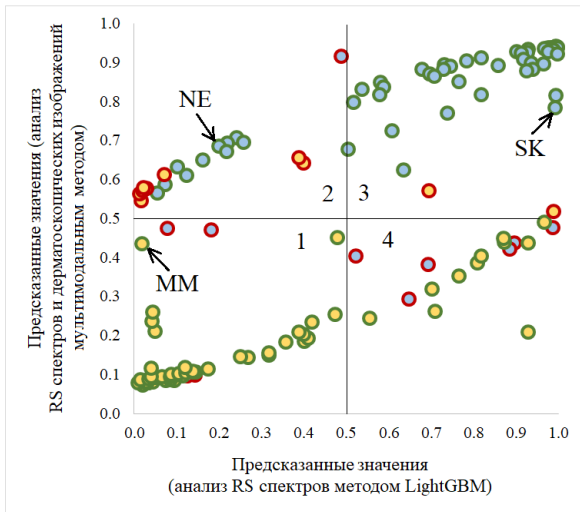
На рисунке 5.3б видно, что в проекции предсказанных значений CNN и LightGBM рассмотренные случаи (обозначены стрелками, подписи соответствуют истинному диагнозу) расположены во 2 и 4 квадрантах, что говорит о том, что модели на основе анализа RS спектров и дерматоскопических изображений относят эти новообразования к разным классам. Однако если посмотреть диаграмму в



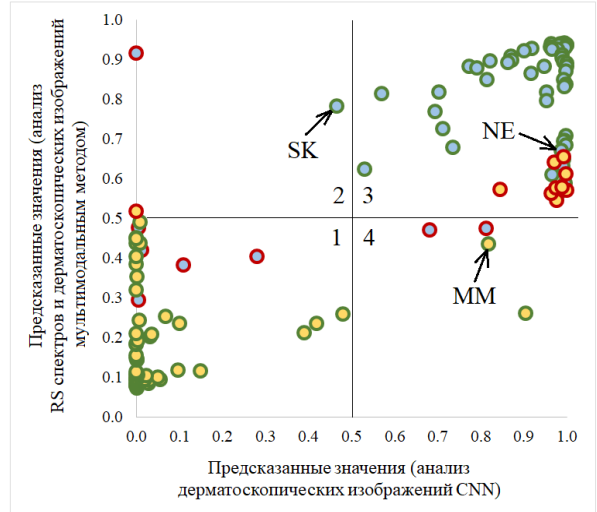
а)



б)



в)



г)

- гистология: SK+NE, мультимодальный метод: SK+NE
- гистология: SK+NE, мультимодальный метод: MM
- гистология: MM, мультимодальный метод: SK+NE

- гистология: MM, мультимодальный метод: MM
- гистология: SK+NE
- гистология: MM

Рисунок 5.3 – Диаграмма разброса предсказанных значений моделей на основе анализа RS спектров методом LightGBM, на основе анализа изображений CNN и модели на основе мультимодального анализа RS спектров и изображений: а) в трехмерном пространстве предсказанных значений всех моделей; б) проекция CNN и LightGBM; в) проекция мультимодального метода и LightGBM, г) проекция мультимодального метода и CNN. Стрелками отмечены образцы, представленные на рисунке 5.4 (3, 4 и 5 образцы)

проекции предсказанных значений метода мультимодального анализа (рисунки 5.3в, 5.3г), то можно заметить, что точки, обозначающие рассматриваемые новообразования, переходят в квадранты 1 и 3, соответствующие верным случаям распознавания. Так, образцы MM и SK расположены в 1 и 3 квадрантах рисунка 5.3в, а образец NE – в 3 квадранте рисунка 5.3г. Таким образом, метод мультимодального анализа новообразований позволяет решать спорные случаи, когда спектральные признаки указывают на один диагноз, а пространственные – на другой, и делать вывод о верном диагнозе.

На рисунке 5.4 представлены RS спектры некоторых конкретных новообразований, их относительный компонентный состав, определенный с использованием MCR, и дерматоскопические изображения. Справа представлены истинный диагноз (подтвержденный гистологией) и диагнозы, предсказанные моделями на основе анализа RS спектров методом LightGBM, на основе анализа изображений CNN и моделью на основе мультимодального анализа RS спектров и изображений. Третий, четвертый и пятый образцы, представленные на рисунке 5.4, соответствуют образцам, показанным стрелками на рисунке 5.3.

Рисунок 5.4 демонстрирует возможности мультимодального метода по сравнению с описанными ранее методами отдельного анализа признаков. В первых двух случаях модели на основе анализа RS спектров методом LightGBM и на основе анализа изображений CNN показывают верный результат, и применение мультимодального метода подтверждает эти результаты. В третьем случае модель на основе анализа изображений CNN показывает MM, и по дерматоскопическому изображению нетрудно заметить характерные признаки MM – изрезанность границ, градиент цвета и т.д. Однако истинный диагноз данного новообразования – SK, что подтверждает модель на основе анализа RS спектров методом LightGBM. Метод мультимодального анализа позволяет учесть решения обеих моделей и отнести образец к классу доброкачественных пигментных новообразований (SK + NE). В четвертом случае, наоборот, права оказывается модель на основе анализа изображений CNN, в то время как модель на основе анализа RS спектров методом LightGBM указывает на неверный диагноз. В пятом случае модель на основе

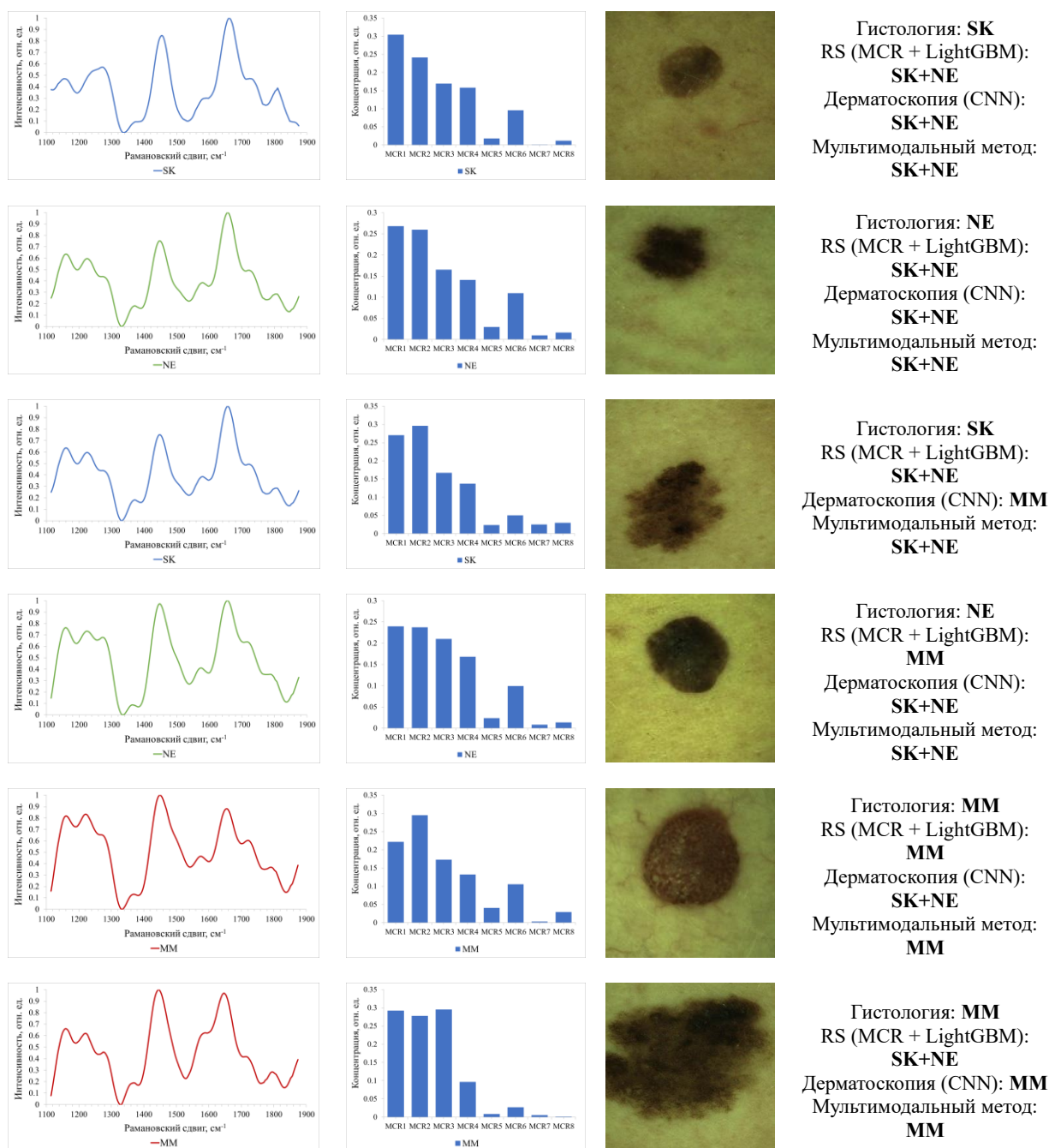


Рисунок 5.4 – Примеры новообразований каждого из исследуемых в работе видов: слева – RS спектры, посередине – профили концентраций, полученные в результате MCR, и дерматоскопические изображения, справа – истинные диагнозы (гистология) и предсказания моделей на основе анализа RS спектров методом LightGBM, на основе анализа изображений CNN и модели на основе мультимодального анализа RS спектров и изображений (сверху вниз). Решение классификаторов отнести образец к классу MM принималось в случае, если предсказанное значение $\leq 0,5$, и к доброкачественным пигментным новообразованиям (SK + NE), если $> 0,5$

анализа изображений снова ошибается, а модель на основе мультимодального анализа признаков обеспечивает верное распознавание. В шестом случае ошибочным является диагноз модели на основе анализа RS спектров. Таким образом, метод мультимодального анализа спектральных признаков, выделенных из RS спектров, и пространственных признаков позволяет более эффективно распознавать пограничные случаи. Применение мультимодального метода с совместным анализом спектральных признаков новообразования с интерпретацией компонентного состава и пространственных признаков позволяет дать прогноз, соответствующий гистологическому исследованию.

5.2 Выводы по главе

1. Подтверждена возможность мультимодального совместного анализа данных оптической биопсии кожи, а именно RS спектров и дерматоскопических изображений.

2. Объединение выделенных ранее методом MCR спектральных признаков новообразования с интерпретацией компонентного состава и пространственных признаков дерматоскопических изображений позволяет сохранить достоинства физической интерпретируемости результатов MCR анализа и одновременно с этим позволяет включить в анализ весь объем спектральной информации и текстурных особенностей новообразований, видимых на изображениях.

3. Эффективность бинарной классификации злокачественной меланомы и доброкачественных пигментных новообразований предложенным методом мультимодального анализа за счет комплексного взаимодополнения выделяемых пространственных и спектральных признаков выше, чем эффективность моделей при использовании только пространственных признаков или только спектральных признаков, что подтверждается увеличением ROC AUC на 7% и 12%, соответственно, а также увеличением доли правильных ответов на 2% и 9%, соответственно.

4. Автоматизация процесса распознавания злокачественной меланомы кожи по дерматоскопическим изображениям и рамановским спектрам кожи при сохранении возможности врачом самостоятельно рассмотреть и проанализировать изображение, а также получить информацию об компонентном составе исследуемой области кожи делает мультимодальный диагностический подход на основе регистрации спектральных и пространственных данных оптической биопсии и их последующего совместного анализа претендентом на широкое использование в системах контроля принятия решения.

ЗАКЛЮЧЕНИЕ

В представленной диссертационной работе разработан метод идентификации новообразований кожи и интерпретации их компонентного состава на основе мультимодального многомерного анализа спектральных данных рамановского рассеяния и дерматоскопических изображений новообразований кожи.

1. Предложен алгоритм выделения и интерпретации компонентного состава ткани кожи на основе разрешения многомерных кривых спектральных данных рамановского рассеяния в ближнем инфракрасном диапазоне при возбуждении на длине волны 785 нм, учитывающий неотрицательность и замкнутость спектральных профилей выделяемых компонентов, а также позволяющий автоматически выделять вклад оптико-электронного тракта спектрометра.

Алгоритм использует метод чередующихся наименьших квадратов и обеспечивает сходимость разложения на не менее 8 групп химических компонентов кожи с ошибкой интерпретации спектра не более 5%. Идентификация восстановленного спектрального профиля на основании характерных пиков показала возможность выделения вклада меланина, белков, воды и липидов. Данный подход позволяет извлекать из спектральных данных информативные признаки, которые можно использовать для дальнейшего анализа.

2. Предложен метод распознавания рамановских спектров новообразований, зарегистрированных в диапазоне $792-1874 \text{ см}^{-1}$ в условиях высоких шумов. Разработано программно-алгоритмическое обеспечение для идентификации типа новообразования на основании анализа спектральных признаков с выделением вклада меланина, белков, воды и липидов. В качестве метода анализа выбран градиентный бустинг реализации LightGBM с количеством деревьев не более 100.

ROC AUC построенных моделей бинарной классификации доброкачественных и злокачественных новообразований, злокачественной меланомы и доброкачественных пигментных новообразований, злокачественной меланомы и себорейного кератоза равны 0,82 (0,77-0,86, 0,95 CI), 0,82 (0,76-0,89, 0,95 CI) и 0,87 (0,79-0,94, 0,95 CI), соответственно, что превосходит эффективность

классификации перечисленных нозологий методом PLS-DA на 7%, 13% и 6%, соответственно. Предложенный подход к анализу спектральных данных рамановского рассеяния обеспечивает точность диагностики различных типов новообразований, сравнимую с точностью квалифицированных дерматологов.

3. Предложен метод идентификации новообразований кожи на основе мультимодального анализа спектральных признаков с интерпретацией компонентного состава и пространственных признаков, выделенных на основе анализа дерматоскопического изображения сверточной нейронной сетью. ROC AUC модели бинарной классификации злокачественной меланомы и доброкачественных пигментных новообразований равна 0,94 (0,90-0,98, 0,95 CI), что выше значений ROC AUC при использовании только пространственных признаков или только спектральных признаков на 7% и 12%, соответственно.

Объединение спектральных признаков новообразования с интерпретацией компонентного состава и пространственных признаков дерматоскопических изображений позволяет сохранить достоинства физической интерпретируемости результатов анализа методом MCR и одновременно с этим включить в анализ весь объем спектральной информации и текстурных особенностей новообразований.

Достигнутые уровни точности, чувствительности и специфичности классификации новообразований кожи свидетельствует о высоком клиническом потенциале предлагаемого подхода. Результаты показывают, что предлагаемые методы позволяют идентифицировать новообразования кожи в среднем более эффективно, чем медицинские специалисты. Разработанные алгоритмы и метод мультимодального анализа внедрены в научно-исследовательскую и клиническую деятельность ГБУЗ «Самарский областной клинический онкологический диспансер».

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ALS – метод чередующихся наименьших квадратов (англ. alternating least squares)

ANSI – Американский национальный институт стандартов (англ. American national standards institute)

AUC – площадь под кривой (англ. area under curve)

BCC – базальноклеточная карцинома (англ. basal cell carcinoma)

BPF – полосовой фильтр (англ. bandpass filter)

CAM – карта активации классов (англ. class activation map)

CARS – когерентная антистоксовая рамановская спектроскопия (англ. coherent anti-Stokes Raman spectroscopy)

CBIR – схема поиска изображений по содержанию (англ. content-based image retrieval)

CI – доверительный интервал (англ. confidence interval)

CNN – сверточные нейронные сети (англ. convolutional neural networks)

DF – дерматофиброма (англ. dermatofibroma)

DM – дихроичное зеркало (англ. dichroic mirror)

EFB – объединение взаимоисключающих признаков (англ. exclusive feature bundling)

FN – количество ложноотрицательных исходов (англ. false negative)

FP – количество ложноположительных исходов (англ. false positive)

FPR – доля ложных положительных классификаций (англ. false positive rate)

GBM – градиентный бустинг (англ. gradient boosted machine)

GOSS – градиентная односторонняя выборка (англ. gradient based one side sampling)

HE – гемангиома (англ. hemangioma)

IARC – Международное агентство по исследованию рака (англ. International Agency for Research on Cancer)

k-NN – метод k-ближайших соседей (англ. k-nearest neighbors)

LGBM, LightGBM – алгоритм градиентного бустинга (англ. light gradient boosted machine)

LPF – широкополосный фильтр (англ. long-pass filter)

LSTM – долгая краткосрочная память (англ. long short-term memory)

MCR – метод разрешения многомерных кривых (англ. multivariate curve resolution)

MCR-ALS – метод разрешения многомерных кривых с применением метода чередующихся наименьших квадратов (англ. multivariate curve resolution alternating least squares)

MM – злокачественная меланома (англ. malignant melanoma)

NE – невус (англ. nevus)

NPV – точность отрицательного прогноза (англ. negative predictive value)

PCA – анализ главных компонент (англ. principal component analysis)

PLS-DA – дискриминантный анализ методом частичных наименьших квадратов (англ. partial least squares discriminant analysis)

PP – папиллома (англ. papilloma)

PPV – точность, положительное прогностическое значение (англ. positive predictive value)

RNN – рекуррентные нейронные сети (англ. recurrent neural networks)

ROC – рабочая характеристика приёмника (англ. receiver operating characteristic)

RS – рамановская спектроскопия (англ. Raman spectroscopy)

SCC – плоскоклеточная карцинома (англ. squamous cell carcinoma)

SERS – поверхностно усиленная рамановская спектроскопия (англ. surface-enhanced Raman spectroscopy)

SK – себорейный кератоз (англ. seborrheic keratoses)

SNR – соотношение сигнал/шум (англ. signal to noise ratio)

SVD – метод сингулярного разложения (англ. singular value decomposition)

SVM – метод опорных векторов (англ. support vector machine)

TN – количество истинноотрицательных исходов (англ. true negative)

TNR – специфичность, доля истинноотрицательных классификаций (англ. true negative rate)

TP – количество истинноположительных исходов (англ. true positive)

TPR – чувствительность, доля верных положительных классификаций (англ. true positive rate)

ВОЗ – Всемирная организация здравоохранения

ДНК – дезоксирибонуклеиновая кислота

ИК – инфракрасный

ОКТ – оптическая когерентная томография

ПЗС – прибор с зарядовой связью

ПК – персональный компьютер

РНК – рибонуклеиновая кислота

СОКОД – Самарский областной клинический онкологический диспансер

ЭВМ – электронно-вычислительная машина

УФ – ультрафиолетовый

СПИСОК ЛИТЕРАТУРЫ

1. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists / H. A. Haenssle, C. Fink, R. Schneiderbauer [и др.] // *Annals of oncology*. – 2018. – Т. 29. – № 8. – С. 1836-1842.
2. Autofluorescence spectroscopy of cutaneous neoplasia under ultraviolet, visible and near infrared excitation / E. Borisova, D. Ivanov, B. Kolev [и др.] // *Tissue Optics and Photonics*. – SPIE, 2020. – Т. 11363. – С. 100-104.
3. Гладкова, Н. Мультимодальная оптическая когерентная томография в клинической медицине / Н. Гладкова, Г. Геликонов, Е. Киселева. – М.: ФИЗМАЛИТ, 2022. – 336 с.
4. Multiparametric spectral diagnosis of skin cancer / V. P. Zakharov, L. A. Bratchenko, Y. A. Khristoforova [и др.] // *Tissue Optics and Photonics*. – SPIE, 2020. – Т. 11363. – С. 105-110.
5. In vivo diagnosis of skin cancer with a portable Raman spectroscopic device / I. A. Bratchenko, L. A. Bratchenko, A. A. Moryatov [и др.] // *Experimental Dermatology*. – 2021. – Т. 30. – № 5. – С. 652-663.
6. Popp, J. Modern Raman spectroscopy for biomedical applications: A variety of Raman spectroscopical techniques on the threshold of biomedical applications / J. Popp, C. Krafft, T. Mayerhöfer // *Optik & Photonik*. – 2011. – Т. 6. – № 4. – С. 24-28.
7. Mahadevan-Jansen, A. Raman spectroscopy for the detection of cancers and precancers / A. Mahadevan-Jansen, R. R. Richards-Kortum // *Journal of biomedical optics*. – 1996. – Т. 1. – № 1. – С. 31-70.
8. Raman spectroscopy of in vivo cutaneous melanin / Z. Huang, H. Lui, X. Chen [и др.] // *Journal of biomedical optics*. – 2004. – Т. 9. – № 6. – С. 1198-1205.
9. Accuracy of Raman spectroscopy for differentiating skin cancer from normal tissue / J. Zhang, Y. Fan, Y. Song, J. Xu // *Medicine*. – 2018. – Т. 97. – № 34. – С. e12022.

10. Dermatologist-level classification of skin cancer with deep neural networks / A. Esteva, B. Kuprel, R. A. Novoa [и др.] // *nature*. – 2017. – Т. 542. – № 7639. – С. 115-118.
11. Deep neural networks are superior to dermatologists in melanoma image classification / T. J. Brinker, A. Hekler, A. H. Enk [и др.] // *European Journal of Cancer*. – 2019. – Т. 119. – С. 11-17.
12. Fusing fine-tuned deep features for skin lesion classification / A. Mahbod, G. Schaefer, I. Ellinger [и др.] // *Computerized Medical Imaging and Graphics*. – 2019. – Т. 71. – С. 19-29.
13. Artificial intelligence for skin cancer detection: scoping review / A. Takiddin, J. Schneider, Y. Yang [и др.] // *Journal of medical Internet research*. – 2021. – Т. 23. – № 11. – С. e22934.
14. Tuchin, V. V. Multimodal optical diagnostics of cancer / V. V. Tuchin, J. Popp, V. Zakharov. – Springer, 2020. – 605 с.
15. Wu, M. Recent advances in Raman spectroscopy for skin diagnosis / M. Wu, B. Gao, X. Wei // *Journal of Innovative Optical Health Sciences*. – 2023. – Т. 16. – № 03. – С. 2330003.
16. Luo, R. Deep learning for Raman spectroscopy: A review / R. Luo, J. Popp, T. Bocklitz // *Analytica*. – 2022. – Т. 3. – № 3. – С. 287-301.
17. Classification of skin cancer using convolutional neural networks analysis of Raman spectra / I. A. Bratchenko, L. A. Bratchenko, Y. A. Khristoforova [и др.] // *Computer Methods and Programs in Biomedicine*. – 2022. – Т. 219. – С. 106755.
18. Discriminating model for diagnosis of basal cell carcinoma and melanoma in vitro based on the Raman spectra of selected biochemicals / L. Silveira Jr, F. L. Silveira, B. Bodanese [и др.] // *Journal of biomedical optics*. – 2012. – Т. 17. – № 7. – С. 077003-077003.
19. Raman active components of skin cancer / X. Feng, A. J. Moy, H. T. Nguyen [и др.] // *Biomedical optics express*. – 2017. – Т. 8. – № 6. – С. 2835-2850.

20. Cancer statistics for the year 2020: An overview / J. Ferlay, M. Colombet, I. Soerjomataram [и др.] // *International journal of cancer*. – 2021. – Т. 149. – № 4. – С. 778-789.
21. Global burden of cutaneous melanoma in 2020 and projections to 2040 / M. Arnold, D. Singh, M. Laversanne [и др.] // *JAMA dermatology*. – 2022. – Т. 158. – № 5. – С. 495-503.
22. Шахзадова, А. Состояние онкологической помощи населению России в 2022 году / А. Шахзадова, В. Старинский, И. Лисичникова // *Сибирский онкологический журнал*. – 2023. – Т. 22. – № 5. – С. 5-13.
23. Чеботарев, В. Некоторые аспекты эпидемиологии и диагностики злокачественных новообразований кожи / В. Чеботарев, Ю. Закирова, others // *Креативная хирургия и онкология*. – 2020. – № 1. – С. 65-73.
24. Malignant skin neoplasms / C. Ricotti, N. Bouzari, A. Agadi, C. J. Cockerell // *Medical Clinics*. – 2009. – Т. 93. – № 6. – С. 1241-1264.
25. Imaging findings of malignant skin tumors: Radiological–pathological correlation / M. Kawaguchi, H. Kato, Y. Noda [и др.] // *Insights into Imaging*. – 2022. – Т. 13. – № 1. – С. 52.
26. Ascierto Paolo, A. Melanoma & Other Skin Cancers: Essentials for Clinicians / A. Ascierto Paolo, I. Lugowska, R. Plummer. – ESMO, 2021.
27. Романова, О. Клинический случай прогрессирующего диспластического невуса с переходом в меланому / О. Романова, Н. Артемьева, М. Солохина // *Клиническая дерматология и венерология*. – 2018. – Т. 17. – № 2. – С. 34-37.
28. Optical biopsy of amelanotic melanoma with Raman and autofluorescence spectra stimulated by 785 nm laser excitation / I. A. Bratchenko, Y. A. Khristoforova, L. A. Bratchenko [и др.] // *Journal of Biomedical Photonics & Engineering*. – 2021. – Т. 7. – № 2. – С. 020308.
29. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria / N. R. Abbasi, H. M. Shaw, D. S. Rigel [и др.] // *Jama*. – 2004. – Т. 292. – № 22. – С. 2771-2776.

30. Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: a diagnostic validation study / F. M. Walter, A. T. Prevost, J. Vasconcelos [и др.] // *British Journal of General Practice*. – 2013. – Т. 63. – № 610. – С. e345-e353.
31. Eikje, N. S. Vibrational spectroscopy for molecular characterisation and diagnosis of benign, premalignant and malignant skin tumours / N. S. Eikje, K. Aizawa, Y. Ozaki // *Biotechnology annual review*. – 2005. – Т. 11. – С. 191-225.
32. Cutaneous melanoma / G. V. Long, S. M. Swetter, A. M. Menzies [и др.] // *The Lancet*. – 2023. – Т. 402. – № 10400. – С. 485-502.
33. Friedman, R. J. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. / R. J. Friedman, D. S. Rigel, A. W. Kopf // *CA: a cancer journal for clinicians*. – 1985. – Т. 35. – № 3. – С. 130-151.
34. Challenges to effective cancer control in China, India, and Russia / P. E. Goss, K. Strasser-Weippl, B. L. Lee-Bychkovsky [и др.] // *The lancet oncology*. – 2014. – Т. 15. – № 5. – С. 489-538.
35. Rigel, D. S. The evolution of melanoma diagnosis: 25 years beyond the ABCDs / D. S. Rigel, J. Russak, R. Friedman // *CA: a cancer journal for clinicians*. – 2010. – Т. 60. – № 5. – С. 301-316.
36. Davis, L. E. Current state of melanoma diagnosis and treatment / L. E. Davis, S. C. Shalin, A. J. Tackett // *Cancer biology & therapy*. – 2019. – Т. 20. – № 11. – С. 1366-1379.
37. Optical biopsy of human skin-A tool for cutaneous tumours' diagnosis / E. Borisova, P. Pavlova, E. Pavlova [и др.] // *International Journal Bioautomation*. – 2012. – Т. 16. – № 1. – С. 53.
38. Research techniques made simple: emerging imaging technologies for noninvasive optical biopsy of human skin / G. Lentsch, E. G. Baugh, B. Lee [и др.] // *Journal of Investigative Dermatology*. – 2022. – Т. 142. – № 5. – С. 1243-1252.
39. New optical examination procedures for the diagnosis of skin diseases / K. Sies, J. Winkler, M. Zieger [и др.] // *Der Hautarzt*. – 2020. – Т. 71. – С. 101-108.

40. Ulrich, M. In vivo reflectance confocal microscopy for early diagnosis of nonmelanoma skin cancer / M. Ulrich, S. Lange-Asschenfeldt, S. González // *Actas Dermo-Sifiliográficas (English Edition)*. – 2012. – Т. 103. – № 9. – С. 784-789.
41. OCT imaging of skin cancer and other dermatological diseases / M. Mogensen, L. Thrane, T. M. Jørgensen [и др.] // *Journal of biophotonics*. – 2009. – Т. 2. – № 6-7. – С. 442-451.
42. Medical images classification for skin cancer using quantitative image features with optical coherence tomography / W. Gao, V. P. Zakharov, O. O. Myakinin [и др.] // *Journal of Innovative Optical Health Sciences*. – 2016. – Т. 9. – № 02. – С. 1650003.
43. Skin cancer texture analysis of OCT images based on Haralick, fractal dimension, Markov random field features, and the complex directional field features / D. S. Raupov, O. O. Myakinin, I. A. Bratchenko [и др.] // *Optics in Health Care and Biomedical Optics VII*. – SPIE, 2016. – Т. 10024. – С. 666-676.
44. Multiphoton laser tomography and fluorescence lifetime imaging of melanoma: morphologic features and quantitative data for sensitive and specific non-invasive diagnostics / S. Seidenari, F. Arginelli, C. Dunsby [и др.] // *PLoS One*. – 2013. – Т. 8. – № 7. – С. e70682.
45. Optical technologies for the improvement of skin cancer diagnosis: a review / L. Rey-Barroso, S. Peña-Gutiérrez, C. Yáñez [и др.] // *Sensors*. – 2021. – Т. 21. – № 1. – С. 252.
46. Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review / M. A. Kassem, K. M. Hosny, R. Damaševičius, M. M. Eltoukhy // *Diagnostics*. – 2021. – Т. 11. – № 8. – С. 1390.
47. Ландсберг, Г. Новое явление при рассеянии света / Г. Ландсберг, Л. Мандельштам // *Журнал Русского физико-химического общества*. – Т. 60. – № 4. – С. 335-338.
48. Raman, C. V. A new type of secondary radiation / C. V. Raman, K. S. Krishnan // *Nature*. – 1928. – Т. 121. – № 3048. – С. 501-502.

49. Сущинский, М. М. Комбинационное рассеяние света и строение вещества / М. М. Сущинский. – Наука, 1981.
50. Long, D. A. The raman effect / D. A. Long // *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. – 2002. – ISBN: 978-0-471-49028-9.
51. Barr, A. J. The biochemical basis of disease / A. J. Barr // *Essays in biochemistry*. – 2018. – Т. 62. – № 5. – С. 619-642.
52. Adlercreutz, H. Western diet and Western diseases: some hormonal and biochemical mechanisms and associations / H. Adlercreutz // *Scandinavian journal of clinical and laboratory investigation*. – 1990. – Т. 50. – № sup201. – С. 3-23.
53. Plasma free amino acid profiling of five types of cancer patients and its application for early detection / Y. Miyagi, M. Higashiyama, A. Gochi [и др.] // *PloS one*. – 2011. – Т. 6. – № 9. – С. e24143.
54. Plasma and tissue free amino acid profiles and their concentration correlation in patients with lung cancer / Q. Zhao, Y. Cao, C. Hu [и др.] // *Asia Pacific journal of clinical nutrition*. – 2014. – Т. 23. – № 3. – С. 429-436.
55. Breast and lung cancer are associated with a decrease in blood cell amino acid content / A. M. Proenza, J. Oliver, A. Palou, P. Roca // *The Journal of nutritional biochemistry*. – 2003. – Т. 14. – № 3. – С. 133-138.
56. Братченко, И. А. Мультимодальный флуоресцентный и Рамановский спектральный анализ тканей кожи человека и методы машинного обучения для диагностики новообразований кожи и патологий почек : специальность 1.5.2 – «Биофизика» : диссертация на соискание ученой степени доктора физико-математических наук / Братченко Иван Алексеевич. – Саратов, 2022. – 224 с.
57. Real-time Raman spectroscopy for in vivo skin cancer diagnosis / H. Lui, J. Zhao, D. McLean, H. Zeng // *Cancer research*. – 2012. – Т. 72. – № 10. – С. 2491-2500.
58. Integrated real-time Raman system for clinical in vivo skin analysis / J. Zhao, H. Lui, D. I. McLean, H. Zeng // *Skin Research and Technology*. – 2008. – Т. 14. – № 4. – С. 484-492.

59. Combined fluorescence-Raman spectroscopic setup for the diagnosis of melanocytic lesions / R. Cicchi, A. Cosci, S. Rossari [и др.] // *Journal of Biophotonics*. – 2014. – Т. 7. – № 1-2. – С. 86-95.
60. Optical techniques for the noninvasive diagnosis of skin cancer / M. A. Calin, S. V. Parasca, R. Savastru [и др.] // *Journal of cancer research and clinical oncology*. – 2013. – Т. 139. – С. 1083-1104.
61. Alfano, R. Optical biopsy for cancer detection / R. Alfano, Y. Pu // *Lasers for Medical Applications*. – Elsevier, 2013. – С. 325-367.
62. Khristoforova, Y. Raman-Based Techniques in Medical Applications for Diagnostic Tasks: A Review / Y. Khristoforova, L. Bratchenko, I. Bratchenko // *International Journal of Molecular Sciences*. – 2023. – Т. 24. – № 21. – С. 15605.
63. Raman spectroscopy of human skin for kidney failure detection / L. A. Bratchenko, I. A. Bratchenko, Y. A. Khristoforova [и др.] // *Journal of Biophotonics*. – 2021. – Т. 14. – № 2. – С. e202000360.
64. Raman spectroscopy in chronic heart failure diagnosis based on human skin analysis / Y. A. Khristoforova, L. A. Bratchenko, M. A. Skuratova [и др.] // *Journal of Biophotonics*. – 2023. – Т. 16. – № 7. – С. e202300016.
65. Comparative study of multivariate analysis methods of blood Raman spectra classification / L. A. Bratchenko, I. A. Bratchenko, A. A. Lykina [и др.] // *Journal of Raman Spectroscopy*. – 2020. – Т. 51. – № 2. – С. 279-292.
66. In-vivo study using Raman Spectroscopy to estimate the effect of fairness creams on skin / A. Daniel, W. Savarimuthu, Y. Gowripal, others // *J. Dermat. Cosmetol.* – 2018. – Т. 2. – № 5. – С. 57-59.
67. Use of Raman spectroscopy for the early detection of filaggrin-related atopic dermatitis / F. González, J. Alda, B. Moreno-Cruz [и др.] // *Skin Research and Technology*. – 2011. – Т. 17. – № 1. – С. 45-50.
68. Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through Raman spectroscopy and principal component analysis / B. Bodanese, F. L. Silveira, R. A. Zângaro [и др.] // *Photomedicine and laser surgery*. – 2012. – Т. 30. – № 7. – С. 381-387.

69. In vivo study for the discrimination of cancerous and normal skin using fibre probe-based Raman spectroscopy / J. Schleusener, P. Gluszczyńska, C. Reble [и др.] // *Experimental dermatology*. – 2015. – Т. 24. – № 10. – С. 767-772.
70. Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis / L. Lim, B. Nichols, M. R. Migden [и др.] // *Journal of biomedical optics*. – 2014. – Т. 19. – № 11. – С. 117003-117003.
71. In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy / C. A. Lieber, S. K. Majumder, D. L. Ellis [и др.] // *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery*. – 2008. – Т. 40. – № 7. – С. 461-467.
72. Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy / T. C. B. Schut, P. J. Caspers, G. J. Puppels [и др.] // *Journal of investigative dermatology*. – 2002. – Т. 119. – № 1. – С. 64-69.
73. In vivo near-infrared autofluorescence imaging of pigmented skin lesions: methods, technical improvements and preliminary clinical results / S. Wang, J. Zhao, H. Lui [и др.] // *Skin Research and Technology*. – 2013. – Т. 19. – № 1. – С. 20-26.
74. Advances in the in vivo Raman spectroscopy of malignant skin tumors using portable instrumentation / N. Kourkoumelis, I. Balatsoukas, V. Moulia [и др.] // *International journal of molecular sciences*. – 2015. – Т. 16. – № 7. – С. 14554-14570.
75. Biophysical basis of skin cancer margin assessment using Raman spectroscopy / X. Feng, M. C. Fox, J. S. Reichenberg [и др.] // *Biomedical optics express*. – 2019. – Т. 10. – № 1. – С. 104-118.
76. Discrimination of non-melanoma skin lesions from non-tumor human skin tissues in vivo using Raman spectroscopy and multivariate statistics / F. L. Silveira, M. T. Pacheco, B. Bodanese [и др.] // *Lasers in surgery and medicine*. – 2015. – Т. 47. – № 1. – С. 6-16.
77. Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo / K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice [и др.] // *Journal of Raman Spectroscopy*. – 2002. – Т. 33. – № 7. – С. 552-563.

78. Briançon, S. Confocal Raman microspectroscopy of the skin / S. Briançon // *European Journal of Dermatology*. – 2011. – Т. 21. – № 6. – С. 851-63.
79. In vivo coherent Raman imaging of the melanomagenesis-associated pigment pheomelanin / H. Wang, S. Osseiran, V. Igras [и др.] // *Scientific reports*. – 2016. – Т. 6. – № 1. – С. 37986.
80. Schlücker, S. Surface-Enhanced raman spectroscopy: Concepts and chemical applications / S. Schlücker // *Angewandte Chemie International Edition*. – 2014. – Т. 53. – № 19. – С. 4756-4795.
81. Analyzing the serum of hemodialysis patients with end-stage chronic kidney disease by means of the combination of SERS and machine learning / L. A. Bratchenko, S. Z. Al-Sammarraie, E. N. Tupikova [и др.] // *Biomedical Optics Express*. – 2022. – Т. 13. – № 9. – С. 4926-4938.
82. Silver nanoparticles-based substrate for blood serum analysis under 785 nm laser excitation / S. Z. Al-Sammarraie, L. A. Bratchenko, E. N. Tupikova [и др.] // *Journal of Biomedical Photonics & Engineering*. – 2022. – Т. 8. – № 1. – С. 010301.
83. Portable spectroscopic system for in vivo skin neoplasms diagnostics by Raman and autofluorescence analysis / Y. A. Khristoforova, I. A. Bratchenko, O. O. Myakinin [и др.] // *Journal of biophotonics*. – 2019. – Т. 12. – № 4. – С. e201800400.
84. Raman spectroscopic study of benign and malignant ovarian tissues / L. Fan, H. Wang, X. Wu [и др.] // *Laser Physics*. – 2022. – Т. 32. – № 3. – С. 035601.
85. Magalhaes, C. The role of AI classifiers in skin cancer images / C. Magalhaes, J. Mendes, R. Vardasca // *Skin Research and Technology*. – 2019. – Т. 25. – № 5. – С. 750-757.
86. Machine learning of Raman spectroscopy data for classifying cancers: a review of the recent literature / N. Blake, R. Gaifulina, L. D. Griffin [и др.] // *Diagnostics*. – 2022. – Т. 12. – № 6. – С. 1491.
87. Detection of skin cancer by classification of Raman spectra / S. Sigurdsson, P. A. Philipsen, L. K. Hansen [и др.] // *IEEE transactions on biomedical engineering*. – 2004. – Т. 51. – № 10. – С. 1784-1793.

88. Novel aspects of Raman spectroscopy in skin research / D. Lunter, V. Klang, D. Kocsis [и др.] // *Experimental Dermatology*. – 2022. – Т. 31. – № 9. – С. 1311-1329.
89. Qi, Y. Recent application of Raman spectroscopy in tumor diagnosis: from conventional methods to artificial intelligence fusion / Y. Qi, Y. Liu, J. Luo // *PhotoniX*. – 2023. – Т. 4. – № 1. – С. 22.
90. Barın, S. An improved hair removal algorithm for dermoscopy images / S. Barın, G. E. Güraksın // *Multimedia Tools and Applications*. – 2024. – Т. 83. – № 3. – С. 8931-8953.
91. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts / S. Haggemüller, R. C. Maron, A. Hekler [и др.] // *European Journal of Cancer*. – 2021. – Т. 156. – С. 202-216.
92. Explainable artificial intelligence in skin cancer recognition: A systematic review / K. Hauser, A. Kurz, S. Haggemüller [и др.] // *European Journal of Cancer*. – 2022. – Т. 167. – С. 54-69.
93. Barata, C. Improving the explainability of skin cancer diagnosis using CBIR / C. Barata, C. Santiago // *International Conference on Medical Image Computing and Computer-Assisted Intervention*. – Springer, 2021. – С. 550-559.
94. Using content-based image retrieval of dermoscopic images for interpretation and education: A pilot study / M. Sadeghi, P. Chilana, J. Yap [и др.] // *Skin Research and Technology*. – 2020. – Т. 26. – № 4. – С. 503-512.
95. Rapid diagnosis of lung cancer and glioma based on serum Raman spectroscopy combined with deep learning / C. Chen, W. Wu, C. Chen [и др.] // *Journal of Raman Spectroscopy*. – 2021. – Т. 52. – № 11. – С. 1798-1809.
96. Garrido, M. Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes / M. Garrido, F. Rius, M. Larrechi // *Analytical and bioanalytical chemistry*. – 2008. – Т. 390. – С. 2059-2066.
97. Juan, A. de. Multivariate Curve Resolution: 50 years addressing the mixture analysis problem–A review / A. de Juan, R. Tauler // *Analytica Chimica Acta*. – 2021. – Т. 1145. – С. 59-78.

98. Noothalapati, H. Biological and medical applications of multivariate curve resolution assisted Raman spectroscopy / H. Noothalapati, K. Iwasaki, T. Yamamoto // *Analytical Sciences*. – 2017. – Т. 33. – № 1. – С. 15-22.
99. Xu, H. In-vivo fluorescence imaging with a multivariate curve resolution spectral unmixing technique / H. Xu, B. W. Rice // *Journal of biomedical optics*. – 2009. – Т. 14. – № 6. – С. 064011-064011.
100. Automatic and objective oral cancer diagnosis by Raman spectroscopic detection of keratin with multivariate curve resolution analysis / P.-H. Chen, R. Shimada, S. Yabumoto [и др.] // *Scientific reports*. – 2016. – Т. 6. – № 1. – С. 20097.
101. Identification of molecular basis for objective discrimination of breast cancer cells (MCF-7) from normal human mammary epithelial cells by Raman microspectroscopy and multivariate curve resolution analysis / K. Iwasaki, A. Araki, C. M. Krishna [и др.] // *International Journal of Molecular Sciences*. – 2021. – Т. 22. – № 2. – С. 800.
102. Blind source separation of molecular components of the human skin in vivo: non-negative matrix factorization of Raman microspectroscopy data / B. Yakimov, A. Venets, J. Schleusener [и др.] // *Analyst*. – 2021. – Т. 146. – № 10. – С. 3185-3196.
103. Azzouz, T. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples / T. Azzouz, R. Tauler // *Talanta*. – 2008. – Т. 74. – № 5. – С. 1201-1210.
104. Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS) / J. Felten, H. Hall, J. Jaumot [и др.] // *Nature protocols*. – 2015. – Т. 10. – № 2. – С. 217-240.
105. Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study / P. Carli, V. de Giorgi, A. Chiarugi [и др.] // *Journal of the American Academy of Dermatology*. – 2004. – Т. 50. – № 5. – С. 683-689.
106. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting / M. Vestergaard, P. Macaskill, P. Holt, S. Menzies // *British Journal of Dermatology*. – 2008. – Т. 159. – № 3. – С. 669-676.

107. Adya, K. A. Dermoscopy: an overview of the principles, procedure and practice / K. A. Adya // *Dermoscopy-Histopathology Correlation: A Conspectus in the Skin of colour*. – 2021. – С. 1-13.
108. Nirmal, B. Dermoscopy: Physics and principles / B. Nirmal // *Indian Journal of Dermatopathology and Diagnostic Dermatology*. – 2017. – Т. 4. – № 2. – С. 27-30.
109. Sonthalia, S. Dermoscopy overview and extradiagnostic applications / S. Sonthalia, S. Yumeen, F. Kaliyadan. – StatPearls Publishing, 2019. – URL: <https://www.ncbi.nlm.nih.gov/books/NBK537131/> (дата обращения: 29.01.2024).
110. Патент US 6,587,711 B1. Spectral polarizing tomographic Dermatoscope: опубл. 01.07.2003 / R. R. Alfano, Y. Budansky, J. Luo. – Google Patents, 2003.
111. Differences between polarized light dermoscopy and immersion contact dermoscopy for the evaluation of skin lesions / C. Benvenuto-Andrade, S. W. Dusza, A. L. C. Agero [и др.] // *Archives of dermatology*. – 2007. – Т. 143. – № 3. – С. 329-338.
112. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis / G. Argenziano, G. Fabbrocini, P. Carli [и др.] // *Archives of dermatology*. – 1998. – Т. 134. – № 12. – С. 1563-1570.
113. Michalak-Stoma, A. Usefulness of dermoscopy to provide accurate assessment of skin cancers / A. Michalak-Stoma, K. Małkińska, D. Krasowska // *Clinical, Cosmetic and Investigational Dermatology*. – 2021. – С. 733-746.
114. A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: a systematic review / S. C. Chen, D. M. Bravata, E. Weil, I. Olkin // *Archives of dermatology*. – 2001. – Т. 137. – № 12. – С. 1627-1634.
115. Wavelet transform fuzzy algorithms for dermoscopic image segmentation / H. Castillejos, V. Ponomaryov, L. Nino-de-Rivera [и др.] // *Computational and mathematical methods in medicine*. – 2012. – Т. 2012. – С. 578721.
116. Hair removal in dermoscopy images using variational autoencoders / D. Bardou, H. Bouaziz, L. Lv, T. Zhang // *Skin Research and Technology*. – 2022. – Т. 28. – № 3. – С. 445-454.

117. A comparison of machine learning methods for the diagnosis of pigmented skin lesions / S. Dreiseitl, L. Ohno-Machado, H. Kittler [и др.] // *Journal of biomedical informatics*. – 2001. – Т. 34. – № 1. – С. 28-36.
118. Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions / G. Capdehourat, A. Corez, A. Bazzano [и др.] // *Pattern Recognition Letters*. – 2011. – Т. 32. – № 16. – С. 2187-2196.
119. An efficient machine learning approach for the detection of melanoma using dermoscopic images / Z. Waheed, A. Waheed, M. Zafar, F. Riaz // *2017 International conference on communication, computing and digital systems (C-CODE)*. – IEEE, 2017. – С. 316-319.
120. Alencar, F. E. S. Development of a system classification of images dermoscopic for mobile devices / F. E. S. Alencar, D. C. Lopes, F. M. M. Neto // *IEEE Latin America Transactions*. – 2016. – Т. 14. – № 1. – С. 325-330.
121. Kasmi, R. Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule / R. Kasmi, K. Mokrani // *IET Image Processing*. – 2016. – Т. 10. – № 6. – С. 448-455.
122. Collaborative human-AI (CHAI): Evidence-based interpretable melanoma classification in dermoscopic images / N. C. Codella, C.-C. Lin, A. Halpern [и др.] // *Understanding and Interpreting Machine Learning in Medical Image Computing Applications: First International Workshops, MLCN 2018, DLF 2018, and iMIMIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings 1*. – Springer, 2018. – С. 97-105.
123. Fusion of structural and textural features for melanoma recognition / F. Adjed, S. J. Safdar Gardezi, F. Ababsa [и др.] // *IET Computer Vision*. – 2018. – Т. 12. – № 2. – С. 185-195.
124. Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier / V. Balaji, S. Suganthi, R. Rajadevi [и др.] // *Measurement*. – 2020. – Т. 163. – С. 107922.
125. Компьютеризированная диагностика меланомы на базе поиска похожих дерматоскопических изображений в базе данных / В. Левчук, В. Ковалев, В.

Баркалин, В. Лозовский // Известия Национальной академии наук Беларуси. Серия физико-технических наук. – 2016. – № 2. – С. 86-91.

126. Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features / P. Tschandl, G. Argenziano, M. Razmara, J. Yap // *British Journal of Dermatology*. – 2019. – Т. 181. – № 1. – С. 155-165.

127. Exploring the correlation between deep learned and clinical features in melanoma detection / T. Chowdhury, A. R. Bajwa, T. Chakraborti [и др.] // *Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25*. – Springer, 2021. – С. 3-17.

128. Fusing metadata and dermoscopy images for skin disease diagnosis / W. Li, J. Zhuang, R. Wang [и др.] // *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*. – IEEE, 2020. – С. 1996-2000.

129. A mutual bootstrapping model for automated skin lesion segmentation and classification / Y. Xie, J. Zhang, Y. Xia, C. Shen // *IEEE transactions on medical imaging*. – 2020. – Т. 39. – № 7. – С. 2482-2493.

130. Li, Y. Skin lesion analysis towards melanoma detection using deep learning network / Y. Li, L. Shen // *Sensors*. – 2018. – Т. 18. – № 2. – С. 556.

131. Refianti, R. Classification of melanoma skin cancer using convolutional neural network / R. Refianti, A. B. Mutiara, R. P. Priyandini // *International Journal of Advanced Computer Science and Applications*. – 2019. – Т. 10. – № 3. – С. 409-417.

132. Maiti, A. Improving detection of Melanoma and Naevus with deep neural networks / A. Maiti, B. Chatterjee // *Multimedia Tools and Applications*. – 2020. – Т. 79. – № 21. – С. 15635-15654.

133*. VGG convolutional neural network classification of hyperspectral images of skin neoplasms / B. V. Grechkin, V. O. Vinokurov, Y. A. Khristoforova, **I. A. Matveeva** // *Journal of Biomedical Photonics & Engineering*. – 2023. – Т. 9. – № 4. – С. 040304.

134*. Нейросетевой классификатор гиперспектральных снимков кожных патологий / В. О. Винокуров, **И. А. Матвеева**, Ю. А. Христофорова [и др.] // *Компьютерная оптика*. – 2021. – Т. 45. – № 6. – С. 879-886.

135. Kassani, S. H. A comparative study of deep learning architectures on melanoma detection / S. H. Kassani, P. H. Kassani // *Tissue and Cell*. – 2019. – Т. 58. – С. 76-83.
- 136*. Дерюгина, В. А. Нейросетевая классификация дерматоскопических изображений / В. А. Дерюгина, **И. А. Матвеева** // XX Всероссийский молодежный Самарский конкурс-конференция научных работ по оптике и лазерной физике, посвященный 100-летию со дня рождения Н.Г. Басова. – М.: Тривант, 2022. – С. 315-320.
137. A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system / D. de A. Rodrigues, R. F. Ivo, S. C. Satapathy [и др.] // *Pattern Recognition Letters*. – 2020. – Т. 136. – С. 8-15.
138. Use of neural network-based deep learning techniques for the diagnostics of skin diseases / D. Gavrilov, A. V. Melerzanov, N. Shchelkunov, E. Zakirov // *Biomedical Engineering*. – 2019. – Т. 52. – С. 348-352.
139. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM / P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz [и др.] // *Sensors*. – 2021. – Т. 21. – № 8. – С. 2852.
140. A skin disease classification model based on densenet and convnext fusion / M. Wei, Q. Wu, H. Ji [и др.] // *Electronics*. – 2023. – Т. 12. – № 2. – С. 438.
141. Almuayqil, S. N. Computer-Aided diagnosis for early signs of skin diseases using multi types feature fusion based on a hybrid deep learning model / S. N. Almuayqil, S. Abd El-Ghany, M. Elmogy // *Electronics*. – 2022. – Т. 11. – № 23. – С. 4009.
142. Serte, S. Wavelet-based deep learning for skin lesion classification / S. Serte, H. Demirel // *IET Image Processing*. – 2020. – Т. 14. – № 4. – С. 720-726.
143. A hybrid approach for melanoma classification using ensemble machine learning techniques with deep transfer learning / M. R. Thanka, E. B. Edwin, V. Ebenezer [и др.] // *Computer Methods and Programs in Biomedicine Update*. – 2023. – Т. 3. – С. 100103.

144. Deep learning methods for accurate skin cancer recognition and mobile application / I. Kousis, I. Perikos, I. Hatzilygeroudis, M. Virvou // *Electronics*. – 2022. – Т. 11. – № 9. – С. 1294.
145. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task / T. J. Brinker, A. Hekler, A. H. Enk [и др.] // *European Journal of Cancer*. – 2019. – Т. 113. – С. 47-54.
146. Acral melanoma detection using a convolutional neural network for dermoscopy images / C. Yu, S. Yang, W. Kim [и др.] // *PloS one*. – 2018. – Т. 13. – № 3. – С. e0193321.
147. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study / P. Tschandl, N. Codella, B. N. Akay [и др.] // *The lancet oncology*. – 2019. – Т. 20. – № 7. – С. 938-947.
148. The International Skin Imaging Collaboration. – URL: <https://www.isic-archive.com> (дата обращения: 29.01.2024).
149. Skin Cancer MNIST: HAM10000. – URL: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000> (дата обращения: 29.01.2024).
150. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images / I. Giotis, N. Molders, S. Land [и др.] // *Expert systems with applications*. – 2015. – Т. 42. – № 19. – С. 6578-6585.
151. PH² Database. – URL: <https://www.fc.up.pt/addi/ph2%20database.html> (дата обращения: 29.01.2024).
152. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks / R. C. Maron, M. Weichenthal, J. S. Utikal [и др.] // *European Journal of Cancer*. – 2019. – Т. 119. – С. 57-65.
153. Seeja, R. Melanoma Segmentation and Classification using Deep Learning / R. Seeja, A. Suresh // *International Journal of Innovative Technology and Exploring Engineering*. – 2019. – Т. 8. – № 12. – С. 2667-2672.

154. Gazioğlu, B. S. A. Effects of objects and image quality on melanoma classification using deep neural networks / B. S. A. Gazioğlu, M. E. Kamaşak // *Biomedical Signal Processing and Control*. – 2021. – Т. 67. – С. 102530.
155. Indraswari, R. Melanoma image classification based on MobileNetV2 network / R. Indraswari, R. Rokhana, W. Herulambang // *Procedia computer science*. – 2022. – Т. 197. – С. 198-207.
156. Canetta, E. Current and future advancements of raman spectroscopy techniques in cancer nanomedicine / E. Canetta // *International journal of molecular sciences*. – 2021. – Т. 22. – № 23. – С. 13141.
157. From Vibrations to Visions: Raman Spectroscopy's Impact on Skin Cancer Diagnostics / C. Delrue, R. Speeckaert, M. Oyaert [и др.] // *Journal of Clinical Medicine*. – 2023. – Т. 12. – № 23. – С. 7428.
158. Cui, S. Raman spectroscopy and imaging for cancer diagnosis / S. Cui, S. Zhang, S. Yue // *Journal of healthcare engineering*. – 2018. – Т. 2018. – С. 8619342.
159. Hyperspectral imaging in the medical field: Present and future / M. A. Calin, S. V. Parasca, D. Savastru, D. Manea // *Applied Spectroscopy Reviews*. – 2014. – Т. 49. – № 6. – С. 435-447.
160. Infrared and Raman spectroscopic studies of molecular disorders in skin cancer / J. Anastassopoulou, M. Kyriakidou, E. Malesiou [и др.] // *in vivo*. – 2019. – Т. 33. – № 2. – С. 567-572.
161. Multispectral Raman Differentiation of Malignant Skin Neoplasms In Vitro: Search for Specific Biomarkers and Optimal Wavelengths / E. Rimskaya, S. Shelygina, A. Timurzieva [и др.] // *International Journal of Molecular Sciences*. – 2023. – Т. 24. – № 19. – С. 14748.
162. Tissue diagnosis using power-sharing multifocal Raman micro-spectroscopy and auto-fluorescence imaging / F. Sinjab, K. Kong, G. Gibson [и др.] // *Biomedical optics express*. – 2016. – Т. 7. – № 8. – С. 2993-3006.
163. Combination of Optical Biopsy with Patient Data for Improvement of Skin Tumor Identification / Y. Khristoforova, I. Bratchenko, L. Bratchenko [и др.] // *Diagnostics*. – 2022. – Т. 12. – № 10. – С. 2503.

164. Evaluation of human melanoma and normal formalin paraffin-fixed samples using Raman and LIBS fused data / M. N. Khan, Q. Wang, B. S. Idrees [и др.] // *Lasers in Medical Science*. – 2022. – Т. 37. – № 5. – С. 2489-2499.
165. Development of a combined OCT-Raman probe for the prospective in vivo clinical melanoma skin cancer screening / M. Mazurenka, L. Behrendt, M. Meinhardt-Wollweber [и др.] // *Review of scientific instruments*. – 2017. – Т. 88. – № 10. – С. 105103.
166. Integration of cellular-resolution optical coherence tomography and Raman spectroscopy for discrimination of skin cancer cells with machine learning / C. You, J.-Y. Yi, T.-W. Hsu, S.-L. Huang // *Journal of Biomedical Optics*. – 2023. – Т. 28. – № 9. – С. 096005-096005.
167. A clinical instrument for combined raman spectroscopy-optical coherence tomography of skin cancers / C. A. Patil, H. Kirshnamoorthi, D. L. Ellis [и др.] // *Lasers in surgery and medicine*. – 2011. – Т. 43. – № 2. – С. 143-151.
168. Integrated system for combined Raman spectroscopy–spectral domain optical coherence tomography / C. A. Patil, J. Kalkman, D. J. Faber [и др.] // *Journal of biomedical optics*. – 2011. – Т. 16. – № 1. – С. 011007-011007.
169. Multimodal Raman spectroscopy and optical coherence tomography for biomedical analysis / S. Fitzgerald, J. Akhtar, E. Schartner [и др.] // *Journal of Biophotonics*. – 2023. – Т. 16. – № 3. – С. e202200231.
170. Trimodal system for in vivo skin cancer screening with combined optical coherence tomography-Raman and colocalized optoacoustic measurements / A. Varkentin, M. Mazurenka, E. Blumenröther [и др.] // *Journal of biophotonics*. – 2018. – Т. 11. – № 6. – С. e201700288.
171. Advances in the clinical application of Raman spectroscopy for cancer diagnostics / C. Kallaway, L. M. Almond, H. Barr [и др.] // *Photodiagnosis and photodynamic therapy*. – 2013. – Т. 10. – № 3. – С. 207-219.
172. Noninvasive assessments of skin glycated proteins by fluorescence and Raman techniques in diabetics and nondiabetics / F. R. Paolillo, V. S. Mattos, A. O. de Oliveira [и др.] // *Journal of biophotonics*. – 2019. – Т. 12. – № 1. – С. e201800162.

173. Wavenumber selection based analysis in Raman spectroscopy improves skin cancer diagnostic specificity / J. Zhao, H. Zeng, S. Kalia, H. Lui // *Analyst*. – 2016. – Т. 141. – № 3. – С. 1034-1043.
174. Коновалов, С. Г. Программно-аппаратный дерматоскопический комплекс для *in vivo* диагностики рака кожи / С. Г. Коновалов, О. А. Мельситов, О. О. Мякинин, А. А. Морятов, С. В. Козлов, И. А. Братченко, В. П. Захаров // *Оптические методы исследования потоков*. – М.: Издательство «Перо», 2019. – С. 142-148.
175. Dermatoscopy software tool for *in vivo* automatic malignant lesions detection / S. G. Konovalov, O. A. Melsitov, O. O. Myakinin [и др.] // *Journal of Biomedical Photonics & Engineering*. – 2018. – Т. 4. – № 4. – С. 040302.
176. Combined Raman and autofluorescence *ex vivo* diagnostics of skin cancer in near-infrared and visible regions / I. A. Bratchenko, D. N. Artemyev, O. O. Myakinin [и др.] // *Journal of Biomedical Optics*. – 2017. – Т. 22. – № 2. – С. 027005-027005.
177. Monte Carlo simulation of *in vivo* Raman spectral Measurements of human skin with a multi-layered tissue optical model / S. Wang, J. Zhao, H. Lui [и др.] // *Journal of biophotonics*. – 2014. – Т. 7. – № 9. – С. 703-712.
178. ГОСТ ИЕС 60825-1-2013. Межгосударственный стандарт. Безопасность лазерной аппаратуры. – Москва : Стандартинформ, 2014. – 83 с.
179. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC) / N. Codella, V. Rotemberg, P. Tschandl [и др.] // *arXiv preprint arXiv:1902.03368*. – 2019.
180. Tschandl, P. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions / P. Tschandl, C. Rosendahl, H. Kittler // *Scientific data*. – 2018. – Т. 5. – № 1. – С. 1-9.
- 181*. Possibilities for decomposing Raman spectra of amino acids mixture by Multivariate Curve Resolution (MCR) analysis / **I. A. Matveeva**, O. O. Myakinin, Y. A. Khristoforova [и др.] // *Fourth International Conference on Terahertz and Microwave Radiation: Generation, Detection, and Applications*. – SPIE, 2020. – Т. 11582. – С. 299-306.

182*. Анализ рамановских спектров методом разрешения многомерных кривых (MCR) / **И. А. Матвеева**, Ю. А. Христофорова, Л. А. Братченко [и др.] // XIX Всероссийский молодежный Самарский конкурс-конференция научных работ по оптике и лазерной физике. – М.: Тривант, 2021. – С. 239-241.

183*. The Effect of Noise in Raman Spectra on the Reconstruction of the Concentration of Amino Acids in the Mixture by Multivariate Curve Resolution (MCR) Analysis / **I. A. Matveeva**, L. A. Bratchenko, O. O. Myakinin [и др.] // Journal of Biomedical Photonics & Engineering. – 2021. – Т. 7. – № 2. – С. 020309.

184*. **Матвеева, И. А.** Влияние шума рамановских базисных спектров аминокислот на их восстановление из смеси методом разрешения многомерных кривых (MCR) / **И. А. Матвеева**, Л. А. Братченко, О. О. Мякинин // Информационные технологии и нанотехнологии (ИТНТ-2021): сб. тр. по материалам VII Междунар. конф. и молодеж. шк. (г. Самара, 20-24 сент.): в 3 т. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2021. – Т. 3. – С. 32112.

185*. Monte Carlo simulation of Raman light scattering and Multivariate Curve Resolution–Alternating Least Squares for determination of changes in skin tissue during the development of malignant neoplasms / **I. Matveeva**, O. Myakinin, Y. Khristoforova, V. Vinokurov // 2020 International Conference on Information Technology and Nanotechnology (ITNT). – IEEE, 2020. – С. 1-4.

186*. **Матвеева, И. А.** Разложение in vivo рамановских спектров кожи методом разрешения многомерных кривых / **И. А. Матвеева**, О. О. Мякинин, И. А. Братченко // Информационные технологии и нанотехнологии (ИТНТ-2022): сб. тр. по материалам VIII Междунар. конф. и молодеж. шк. (г. Самара, 23-27 мая): в 5 т. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2022. – Т. 5. – С. 51252.

187*. **Матвеева, И. А.** Анализ рамановских спектров кожи методом разрешения многомерных кривых / **И. А. Матвеева** // Всероссийская научно-техническая конференция «Актуальные проблемы радиоэлектроники и телекоммуникаций». – Самара: ООО «Артель», 2022. – С. 19-22.

188. Jaumot, J. MCR-ALS GUI 2.0: New features and applications / J. Jaumot, A. de Juan, R. Tauler // *Chemometrics and Intelligent Laboratory Systems*. – 2015. – Т. 140. – С. 1-12.
189. Assessment of new constraints applied to the alternating least squares method / A. De Juan, Y. Vander Heyden, R. Tauler, D. Massart // *Analytica Chimica Acta*. – 1997. – Т. 346. – № 3. – С. 307-318.
190. Chengwang, L. Singular Value Decomposition in Active Monitoring Data Analysis / L. Chengwang // *Handbook of Geophysical Exploration: Seismic Exploration*. – Elsevier, 2010. – Т. 40. – С. 421-430.
191. Windig, W. Interactive self-modeling mixture analysis / W. Windig, J. Guilment // *Analytical chemistry*. – 1991. – Т. 63. – № 14. – С. 1425-1432.
192. Carey, P. Biochemical applications of Raman and resonance Raman spectroscopes / P. Carey. – Elsevier, 2012. – 262 с.
193. 785 nm EdgeBasic™ best-value long-pass edge filter. – URL: <https://www.semrock.com/FilterDetails.aspx?id=BLP01-785R-25> (дата обращения: 29.01.2024).
194. Movasaghi, Z. Raman spectroscopy of biological tissues / Z. Movasaghi, S. Rehman, I. U. Rehman // *Applied Spectroscopy Reviews*. – 2007. – Т. 42. – № 5. – С. 493-541.
- 195*. Multivariate curve resolution alternating least squares analysis of in vivo skin Raman spectra / **I. Matveeva**, I. Bratchenko, Y. Khristoforova [и др.] // *Sensors*. – 2022. – Т. 22. – № 24. – С. 9588.
196. Near infrared Raman spectra of human brain lipids / C. Krafft, L. Neudert, T. Simat, R. Salzer // *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. – 2005. – Т. 61. – № 7. – С. 1529-1535.
197. Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy / I. Notingher, C. Green, C. Dyer [и др.] // *Journal of the Royal Society Interface*. – 2004. – Т. 1. – № 1. – С. 79-90.
198. Characterization by Raman spectroscopy of conformational changes on guanine–cytosine and adenine–thymine oligonucleotides induced by aminoxy analogues

of spermidine / A. Ruiz-Chica, M. Medina, F. Sánchez-Jiménez, F. Ramirez // *Journal of Raman spectroscopy*. – 2004. – Т. 35. – № 2. – С. 93-100.

199. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer / Z. Huang, A. McWilliams, H. Lui [и др.] // *International journal of cancer*. – 2003. – Т. 107. – № 6. – С. 1047-1052.

200. Raman spectroscopy for identification of epithelial cancers / N. Stone, C. Kendall, J. Smith [и др.] // *Faraday discussions*. – 2004. – Т. 126. – С. 141-157.

201. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma / W.-T. Cheng, M.-T. Liu, H.-N. Liu, S.-Y. Lin // *Microscopy research and technique*. – 2005. – Т. 68. – № 2. – С. 75-79.

202. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells / J. W. Chan, D. S. Taylor, T. Zwerdling [и др.] // *Biophysical journal*. – 2006. – Т. 90. – № 2. – С. 648-656.

203. Andrus, P. G. Cancer grading by Fourier transform infrared spectroscopy / P. G. Andrus, R. D. Strickland // *Biospectroscopy*. – 1998. – Т. 4. – № 1. – С. 37-46.

204. Distinctive molecular abnormalities in benign and malignant skin lesions: studies by Raman spectroscopy / M. Gniadecka, H. C. Wulf, O. F. Nielsen [и др.] // *Photochemistry and photobiology*. – 1997. – Т. 66. – № 4. – С. 418-423.

205. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study / R. Malini, K. Venkatakrishna, J. Kurien [и др.] // *Biopolymers: Original Research on Biomolecules*. – 2006. – Т. 81. – № 3. – С. 179-193.

206. In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy / Z. Huang, S. K. Teh, W. Zheng [и др.] // *Biosensors and Bioelectronics*. – 2010. – Т. 26. – № 2. – С. 383-389.

207. Dukor, R. K. Vibrational spectroscopy in the detection of cancer / R. K. Dukor // *Handbook of vibrational spectroscopy*. – 2006.

208. A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy / E. O. Faolain, M. B. Hunter, J. M. Byrne [и др.] // *Vibrational Spectroscopy*. – 2005. – Т. 38. – № 1-2. – С. 121-127.

209. Raman spectroscopy for optical diagnosis in normal and cancerous tissue of the nasopharynx—preliminary findings / D. P. Lau, Z. Huang, H. Lui [и др.] // *Lasers in surgery and medicine*. – 2003. – Т. 32. – № 3. – С. 210-214.
210. Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice / R. J. Lakshmi, V. Kartha, C. Murali Krishna [и др.] // *Radiation research*. – 2002. – Т. 157. – № 2. – С. 175-182.
211. Near-infrared Raman spectroscopy for assessing biochemical changes of cervical tissue associated with precarcinogenic transformation / S. Duraipandian, J. Mo, W. Zheng, Z. Huang // *Analyst*. – 2014. – Т. 139. – № 21. – С. 5379-5386.
212. Raman tweezers spectroscopy of live, single red and white blood cells / A. Bankapur, E. Zachariah, S. Chidangil [и др.] // *PLoS one*. – 2010. – Т. 5. – № 4. – С. e10427.
213. Prospects for in vivo Raman spectroscopy / E. Hanlon, R. Manoharan, T. Koo [и др.] // *Physics in Medicine & Biology*. – 2000. – Т. 45. – № 2. – С. R1.
214. Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries / L. Silveira Jr, S. Sathaiah, R. A. Zângaro [и др.] // *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery*. – 2002. – Т. 30. – № 4. – С. 290-297.
215. Marshall, C. P. The potential of Raman spectroscopy for the analysis of diagenetically transformed carotenoids / C. P. Marshall, A. Olcott Marshall // *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. – 2010. – Т. 368. – № 1922. – С. 3137-3144.
216. In vivo measurement of human dermis by 1064 nm-excited fiber Raman spectroscopy / S. Naito, Y.-K. Min, K. Sugata [и др.] // *Skin Research and Technology*. – 2008. – Т. 14. – № 1. – С. 18-25.
217. Naumann, D. Infrared and NIR Raman spectroscopy in medical microbiology / D. Naumann // *Infrared spectroscopy: new tool in medicine*. – SPIE, 1998. – Т. 3257. – С. 245-257.

218*. **Матвеева, И. А.** Метод разрешения многомерных кривых как инструмент анализа рамановских спектров / **И. А. Матвеева** // Всероссийская научная конференция с международным участием «Невская фотоника-2023». – 2023. – С. 159.

219*. **Матвеева, И. А.** Неинвазивный анализ компонентного состава кожи методом разрешения многомерных кривых / **И. А. Матвеева, В. А. Дерюгина** // Фотон-экспресс. – 2023. – Т. 6. – № 190. – С. 320-321.

220*. Analysis of Raman spectra using the multivariate curve resolution-alternating least squares (MCR-ALS) algorithm / **I. A. Matveeva, Y. A. Khristoforova, L. A. Bratchenko, V. P. Zakharov** // Biomedical Spectroscopy, Microscopy, and Imaging II. – SPIE, 2022. – Т. 12144. – С. 122-127.

221*. Classification of human skin Raman spectra using multivariate curve resolution (MCR) and partial least squares discriminant analysis (PLS-DA) / **I. Matveeva, Y. Khristoforova, A. Moryatov** [и др.] // Journal of Physics: Conference Series. – IOP Publishing, 2021. – Т. 2127. – С. 012065.

222. Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation / **К. Hajian-Tilaki** // Caspian journal of internal medicine. – 2013. – Т. 4. – № 2. – С. 627.

223. Package ‘pROC’. – URL: <https://cran.r-project.org/web/packages/pROC/pROC.pdf> (дата обращения: 29.01.2024).

224. Schober, P. Logistic regression in medical research / **P. Schober, T. R. Vetter** // Anesthesia & Analgesia. – 2021. – Т. 132. – № 2. – С. 365-366.

225. Brereton, R. G. Applied chemometrics for scientists / **R. G. Brereton**. – John Wiley & Sons, 2007. – 379 с.

226. Brownlee, J. A gentle introduction to the gradient boosting algorithm for machine learning / **J. Brownlee** // Machine Learning Mastery. – URL: <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/> (дата обращения: 29.01.2024).

227. LightGBM: A highly efficient gradient boosting decision tree / G. Ke, Q. Meng, T. Finley [и др.] // *Advances in neural information processing systems*. – 2017. – Т. 30. – С. 3149–3157.
228. An ensemble of Light Gradient Boosting Machine and adaptive boosting for prediction of type-2 diabetes / M. J. Sai, P. Chettri, R. Panigrahi [и др.] // *International Journal of Computational Intelligence Systems*. – 2023. – Т. 16. – № 1. – С. 14.
229. Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM) / D. D. Rufo, T. G. Debelee, A. Ibenthal, W. G. Negera // *Diagnostics*. – 2021. – Т. 11. – № 9. – С. 1714.
230. Wang, D. LightGBM: an effective miRNA classification method in breast cancer patients / D. Wang, Y. Zhang, Y. Zhao // *Proceedings of the 2017 international conference on computational biology and bioinformatics*. – 2017. – С. 7-11.
231. Vidhyalakshmi, A. AMLGB-: Efficient Model for Skin Disease Detection and Classification using Adaptive Machine for Light Gradient Boosting / A. Vidhyalakshmi, M. Kanchana // *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*. – IEEE, 2023. – С. 985-992.
- 232*. Свидетельство о государственной регистрации программы для ЭВМ № 2023685208. Программный модуль для классификации рамановских спектров патологий кожи: получено 23.11.2023 / К. Е. Томникова, **И. А. Матвеева**. – 2023.
- 233*. Томникова, К. Е. Анализ рамановских спектров кожи методами машинного обучения / К. Е. Томникова, **И. А. Матвеева** // *Всероссийская научно-техническая конференция «Актуальные проблемы радиоэлектроники и телекоммуникаций»*. – Самара: ООО «Артель», 2023. – С. 153-155.
- 234*. Томникова, К. Е. Применение методов машинного обучения для анализа рамановских спектров новообразований кожи / К. Е. Томникова, **И. А. Матвеева** // *XXI Всероссийская молодежная Самарская конкурс-конференция по оптике, лазерной физике и физике плазмы, посвященная 300-летию РАН*. – М.: Тривант, 2023. – С. 261-262.
235. Müller, B. *Neural networks: an introduction* / B. Müller, J. Reinhardt, M. T. Strickland. – Springer Science & Business Media, 1995. — 331 с.

236. Anderson, J. A. An introduction to neural networks / J. A. Anderson. – MIT press, 1995. — 672 с.
237. Petersen, P. Equivalence of approximation by convolutional neural networks and fully-connected networks / P. Petersen, F. Voigtlaender // Proceedings of the American Mathematical Society. – 2020. – Т. 148. – № 4. – С. 1567-1581.
238. Albawi, S. Understanding of a convolutional neural network / S. Albawi, T. A. Mohammed, S. Al-Zawi // 2017 International conference on engineering and technology (ICET). – IEEE, 2017. – С. 1-6.
239. Chollet, F. Deep learning with Python / F. Chollet. – Simon and Schuster, 2021. — 504 с.
240. Recent advances in convolutional neural networks / J. Gu, Z. Wang, J. Kuen [и др.] // Pattern recognition. – 2018. – Т. 77. – С. 354-377.
241. A survey of convolutional neural networks: analysis, applications, and prospects / Z. Li, F. Liu, W. Yang [и др.] // IEEE transactions on neural networks and learning systems. – 2021. – Т. 33. – № 12. – С. 6999-7019.
- 242*. Свидетельство о государственной регистрации программы для ЭВМ № 2022685245. Программный модуль для нейросетевой классификации гиперспектральных изображений патологий кожи: получено 22.12.2022 / В.О. Винокуров, Л.А. Братченко, В.П. Захаров, **И.А. Матвеева**, Ю.А. Христофорова, Д.Н. Артемьев. – 2023.
243. Raschka, S. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 / S. Raschka, V. Mirjalili. – Packt publishing ltd, 2017. – 595 с.
- 244*. Дерюгина, В. А. Сверточные нейронные сети для задачи классификации дерматоскопических изображений / В. А. Дерюгина, **И. А. Матвеева** // Всероссийская научно-техническая конференция «Актуальные проблемы радиоэлектроники и телекоммуникаций». – Самара: ООО «Артель», 2023. – С. 142-144.
- 245*. Дерюгина, В. А. Классификация дерматоскопических изображений с помощью сверточной нейронной сети / В. А. Дерюгина, **И. А. Матвеева** //

Информационные технологии и нанотехнологии (ИТНТ-2023): сб. тр. по материалам IX Междунар. конф. и молодеж. шк. (г. Самара, 17-23 апр. 2023 г.): в 6 т. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2023. – Т. 6. – С. 62612.

246. `tf.keras.preprocessing.image.ImageDataGenerator`. – URL: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (дата обращения: 29.01.2024).

247. Weiss, K. A survey of transfer learning / K. Weiss, T. M. Khoshgoftaar, D. Wang // *Journal of Big data*. – 2016. – Т. 3. – С. 1-40.

248. Transfer learning techniques for medical image analysis: A review / P. Kora, C. P. Ooi, O. Faust [и др.] // *Biocybernetics and Biomedical Engineering*. – 2022. – Т. 42. – № 1. – С. 79-107.

249. Transfer learning for medical image classification: a literature review / H. E. Kim, A. Cosa-Linan, N. Santhanam [и др.] // *BMC medical imaging*. – 2022. – Т. 22. – № 1. – С. 69.

250. Federated and Transfer Learning Methods for the Classification of Melanoma and Nonmelanoma Skin Cancers: A Prospective Study / S. Riaz, A. Naeem, H. Malik [и др.] // *Sensors*. – 2023. – Т. 23. – № 20. – С. 8457.

251*. Multidimensional Analysis of Dermoscopic Images and Spectral Information for the Diagnosis of Skin Tumors / **I. A. Matveeva**, A. I. Komlev, O. I. Kaganov [и др.] // *Journal of Biomedical Photonics & Engineering*. – 2024. – Т. 10. – № 1. – С. 010307.

Приложение А

Акт об использовании результатов диссертационного исследования

Министерство здравоохранения
Самарской области
**Государственное бюджетное
учреждение здравоохранения
«Самарский областной клинический
онкологический диспансер»
(ГБУЗ СОКОД)**



443031 г. Самара, ул. Солнечная 50
Тел.: (846) 994-06-99 приемная, факс 994-03-99
E-mail: info@samaraonko.ru
ИНН 6319077552; КПП 631901001
ОКВЭД 86.10, ОКПО 55919368

№ _____
на № _____ от _____

УТВЕРЖДАЮ

Заместитель главного врача
по научной работе д.м.н.

профессор  О.И. Каганов

_____ 2024 г.



АКТ

об использовании результатов диссертационного исследования

Результаты диссертационного исследования ассистента кафедры лазерных и биотехнических систем Самарского университета Матвеевой Ирины Александровны по теме «Метод мультимодального анализа рамановского рассеяния и дерматоскопических изображений для диагностики новообразований кожи», включая алгоритмы анализа данных, на которые получены свидетельства о государственной регистрации программы для ЭВМ № 2022685245 «Программный модуль для нейросетевой классификации гиперспектральных изображений патологий кожи» (дата выдачи 22.12.2022) и № 2023685208 «Программный модуль для классификации рамановских спектров патологий кожи» (дата выдачи 23.11.2023), внедрены в научно-исследовательскую и клиническую деятельность ГБУЗ «Самарский областной клинический онкологический диспансер».

Заведующий консультативно-диагностической
поликлиникой

А.В. Копылов

Заведующий консультативным
отделением № 1

М.Г. Байчорова