

ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ» РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи



Волков Егор Николаевич

**МЕТОДЫ АВТОМАТИЧЕСКОГО АНАЛИЗА
РЕТИНАЛЬНЫХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ
НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ**

Специальность 1.2.1 –
«Искусственный интеллект и машинное обучение»

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
к. ф.-м. н., доцент
Аверкин Алексей Николаевич

Москва – 2025

Оглавление

Стр.

Введение	8
Глава 1. Методика полуавтоматического формирования проблемно-ориентированных коллекций данных	14
1.1 Теоретические основы создания медицинских коллекций данных	14
1.1.1 Требования к формированию коллекций медицинских данных	14
1.1.2 Обзор литературы и источников	16
1.2 Описание предложенной методики	18
1.3 Существующие коллекции данных ОК-томограмм	20
1.4 Предподготовка изображений	21
1.5 Определение качества изображения	23
1.6 Объяснительный искусственный интеллект	24
1.6.1 Объяснительный искусственный интеллект: общая характеристика	24
1.6.2 Метод объяснительного ИИ САМ	26
1.7 Предварительная разметка с помощью базовой модели	28
1.8 Экспертная разметка	30
1.9 Характеристики полученной коллекции данных	31
1.10 Выводы к первой главе	32
Глава 2. Методы анализа ОК-томограмм	33
2.1 Теоретические основы анализа ОК-томограмм	33
2.1.1 Оптическая когерентная томография	33
2.1.2 Диабетический макулярный отёк	36
2.1.3 Обзор литературы и источников	38
2.2 Описание предлагаемого метода	39
2.3 Коллекции данных	41
2.4 Предподготовка изображений	42
2.5 Предварительная классификация	43
2.5.1 Модель EfficientNetB0	43
2.5.2 Обучение и валидация, результаты	46

	Стр.
2.6	Метод объяснительного ИИ 46
2.7	Сегментация экземпляров биомаркеров ДМО 47
2.7.1	Модель YOLOv8 47
2.7.2	Модель BiFPN 50
2.7.3	Механизм внимания Coordinate Attention 51
2.7.4	Модификация базовой модели 52
2.7.5	Метрики оценки качества 54
2.7.6	Обучение, валидация, результаты 56
2.8	Определение количественных характеристик экземпляров биомаркеров ДМО 57
2.9	Классификация на основе нечётких правил 58
2.9.1	Нечёткий классификатор и лингвистические правила . . . 58
2.9.2	Формирование корпуса лингвистических правил 60
2.10	Выводы ко второй главе 62
Глава 3.	Метод анализа фундус-снимков 68
3.1	Теоретические основы анализа фундус-снимков 68
3.1.1	Офтальмоскопия 68
3.1.2	Диабетическая ретинопатия 69
3.1.3	Обзор литературы и источников 71
3.2	Описание предлагаемого метода 74
3.3	Коллекции данных 76
3.4	Предподготовка изображений 79
3.5	Предварительная классификация 83
3.5.1	Модель VGG16 84
3.5.2	Метрики оценки качества 87
3.5.3	Обучение, валидация, результаты 87
3.6	Морфологические преобразования 89
3.6.1	Метод поиска анатомических зон глазного дна 89
3.6.2	Сегментация зоны фовеа 93
3.6.3	Разделение на квадранты 94
3.7	Семантическая сегментация биомаркеров ДР 95
3.7.1	Модель U-Net 96
3.7.2	Фильтры Габора 99

	Стр.
3.7.3 Модификация базовой модели	100
3.7.4 Метрики оценки качества	103
3.7.5 Обучение, валидация, результаты	103
3.8 Метод объяснительного ИИ	106
3.8.1 Объяснительный ИИ в анализе фундус-снимков	106
3.8.2 Метод ОИИ Grad-CAM	108
3.8.3 Особенности практического применения Grad-CAM	109
3.9 Метод водораздела	111
3.10 Определение количественных характеристик	114
3.11 Классификация на основе логических правил	115
3.12 Выводы к третьей главе	117
Заключение	119
Список рисунков	120
Список таблиц	121
Список литературы	122
Благодарности	146

Введение

Актуальность темы. Диабетическая ретинопатия (ДР) и диабетический макулярный отёк (ДМО) входят в число наиболее значимых осложнений сахарного диабета, оказывающих выраженное влияние на качество жизни пациентов и ведущих к прогрессирующему снижению зрения. Оба состояния требуют регулярного контроля и своевременного выявления структурных изменений сетчатки, особенно в ранних стадиях, когда возможна терапевтическая коррекция. При этом распространённость сахарного диабета и его офтальмологических осложнений продолжает увеличиваться: по данным регистра Минздрава РФ, в 2023 году общее число больных СД превысило 5,5 млн человек, из которых более 30 % имеют признаки диабетической ретинопатии, а около 10 % – клинически значимый макулярный отёк [1].

В практике скрининга и мониторинга состояния сетчатки применяются фундус-фотография и оптическая когерентная томография (ОКТ). Первая является удобным, неинвазивным методом массового обследования, вторая – стандартом верификации и оценки тяжести ДМО. ОКТ позволяет выявлять интравитреальные кисты, утолщение макулы, субретинальную жидкость и другие маркеры отёка, а также отслеживать динамику состояния при проведении терапии. Однако эффективность этих методов на практике ограничивается доступностью квалифицированных специалистов, неоднородностью протоколов и значительными затратами времени на интерпретацию каждого снимка.

Технологии искусственного интеллекта (ИИ), в том числе методы глубокого обучения и объяснительного ИИ, демонстрируют высокий потенциал в задачах анализа фундус-снимков и ОКТ-снимков. С их помощью можно автоматизировать как распознавание признаков ДР (микроаневризмы, кровоизлияния, неоваскуляризация), так и детектирование структур, характерных для ДМО (интравитреальные кисты, экссудаты, отёчные зоны). При этом применение методов объяснительного ИИ позволяет дополнительно визуализировать зоны, на основе которых модель приняла решение, что критически важно для увеличения доверия к технологии.

Актуальность разработки таких решений особенно высока в условиях территориальной протяжённости России, неравномерного распределения медицинской инфраструктуры, а также в свете тенденций цифровизации здра-

воохранения. Внедрение автоматизированных систем интерпретации фундус- и ОКТ-изображений способствует повышению доступности офтальмологической помощи, стандартизации заключений и снижению времени анализа. Кроме того, эти подходы позволяют расширить применение скрининговых программ за пределами специализированных центров.

Разработка и внедрение алгоритмов анализа медицинских изображений напрямую поддерживается стратегическими инициативами государства. В частности, Национальная стратегия развития искусственного интеллекта в Российской Федерации на период до 2030 года, утверждённая Указом Президента РФ от 10 октября 2019 года № 490, подчёркивает необходимость создания и внедрения интеллектуальных технологий, повышающих качество и доступность медицинской помощи, с обязательной ориентацией на прозрачность и воспроизводимость результатов [2].

Целью работы является разработка и исследование новых методов автоматического анализа ретинальных изображений (снимков фундус-камеры и оптических когерентных томограмм) на основе нейронных сетей глубокого обучения для повышения точности диагностики заболеваний сетчатки глаза на основе интеллектуальной идентификации и оценки биомаркеров. В качестве результата диссертационного исследования предлагаются разработанные методы автоматического анализа ретинальных изображений двух модальностей, реализованные в специализированном программном обеспечении.

Для достижения поставленной цели были решены задачи:

1. Разработать методику автоматизированного формирования специализированных коллекций данных для задач анализа ретинальных изображений, использующую нейросетевую оценку качества, визуальные промпты и перекрёстные экспертные аннотации.
2. Создать специализированную коллекцию данных для сегментации экземпляров биомаркеров диабетического макулярного отёка на оптических когерентных томограммах.
3. Разработать метод автоматического анализа ретинальных изображений, основанный на сегментации экземпляров и применении формализованных экспертных знаний и методов объяснительного искусственного интеллекта.

4. Создать метод поиска и сегментации анатомических зон сетчатки глаза на ретинальных изображениях, устойчивый к артефактам и не требующий предварительного обучения.
5. Формализовать и решить задачу анализа биомаркеров диабетического макулярного отёка на оптических когерентных томограммах как задачу сегментации экземпляров с целью последующего применения логических правил для прогнозирования состояния пациентов.

Научная новизна диссертационного исследования заключается в следующем:

1. Предложена новая методика формирования проблемно-ориентированных коллекций данных, отличающиеся от существующих нейросетевой оценкой качества изображений с применением метода объяснительного ИИ, интеграцией базовой модели на основе визуальных промптов, а также применением перекрестных экспертных аннотаций. Впервые получена специализированная коллекция данных для задачи сегментации экземпляров биомаркеров диабетического макулярного отёка на ОКТ-снимках, содержащая все основные классы биомаркеров.
2. Предложен новый метод анализа специальных видов изображений на основе сегментации экземпляров и применения формализованных в виде логических правил экспертных знаний, а также методов объяснительного ИИ. Преимущества нового метода показаны на основе решения задач анализа ретинальных изображений.
3. Предложен новый метод поиска и сегментации анатомических зон глазного дна на основе анализа бинаризованных гистограмм цветового канала, отличающийся нечувствительностью к артефактам и отсутствием необходимости обучения.
4. Впервые формализована и решена задача анализа биомаркеров диабетического макулярного отёка на ОКТ-снимках как задача сегментации экземпляров, что позволило применить логические правила для оценки прогноза пациентов.

Теоретическая и практическая значимость работы заключается в разработке эффективных методов автоматического анализа ретинальных изображений с применением нейронных сетей глубокого обучения, направленных на

решение актуальных задач ранней диагностики офтальмологических заболеваний. Результаты работы внедрены в реальную практику при создании специализированного программного обеспечения для системы поддержки принятия врачебных решений, используемой в бизнес-продукте компании-резидента фонда «Сколково» ООО «МАКАО ИТ».

Разработанные методы и программные решения обладают высокой степенью адаптивности и могут быть использованы в других направлениях медицинской визуализации. Помимо клинического применения, полученные результаты могут использоваться в образовательных программах, научных исследованиях и при формировании высококачественных специализированных коллекций данных.

Положения, выносимые на защиту:

1. Методика формирования специализированных коллекций данных для автоматического анализа ретинальных изображений, основанная на нейросетевой оценке качества изображений, применении метода объяснительного искусственного интеллекта, интеграции моделей на основе визуальных промптов и перекрёстных экспертных аннотациях.
2. Метод автоматического анализа ретинальных изображений, сочетающий сегментацию экземпляров, формализованные логические правила экспертных знаний и методы объяснительного искусственного интеллекта, повышающий точность диагностики заболеваний сетчатки глаза.
3. Метод поиска и сегментации анатомических зон сетчатки глаза на ретинальных изображениях, основанный на анализе бинаризованных гистограмм цветового канала, отличающийся нечувствительностью к артефактам и отсутствием необходимости обучения.

Апробация работы. Основные результаты работы докладывались на 19 международных и всероссийских конференциях: XXVI, XXVII и XXVIII International Conference on Soft Computing and Measurement "SCM (Санкт-Петербург, 2023, 2024, 2025); IV, V и VI International Conference on Neural Networks and Neurotechnologies "Neuro NT (Санкт-Петербург, 2023, 2024, 2025); XXI Национальная конференция по искусственному интеллекту с международным участием "КИИ (Смоленск, 2023); XVI Международная конференция "Безопасность АЭС и подготовка кадров (Обнинск, 2023); XXXI и XXXII

Международная конференция "Математика. Компьютер. Образование. (Дубна, 2024; Пущино, 2025); International Conference on Information Processes and System Development and Quality Assurance "IPSQDA (Санкт-Петербург, 2024); XIII Конгресс молодых учёных ИТМО, (Санкт-Петербург, 2024); International and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems "ITTMM (Москва, 2024); XXIX Всероссийская научно-практическая конференция студентов, аспирантов и молодых специалистов государственного университета "Дубна (Дубна, 2024); XII Международная научно-практическая конференция "Интегрированные модели и мягкие вычисления в искусственном интеллекте (Коломна, 2024); VIII International Conference on Deep Learning in Computational Physics "DLCP 2024 (Москва, 2024); XV Международная конференция "Интеллектуализация обработки информации (Гродно, 2024); XXVII Российская научная конференция "Инжиниринг предприятий и управление знаниями (Москва, 2024); XXVII International Conference «Digital Signal Processing and Its Applications – DSPA-2025», (Москва, 2025).

Достоверность научных достижений подтверждается корректным использованием методов, обоснованием постановки задач, а также экспериментальными исследованиями на открытых данных и данных промышленных партнёров, демонстрирующими эффективность предложенных методов и алгоритмов, корректность работы разработанных программных решений.

Публикации. Всего по результатам диссертационного исследования опубликовано 18 работ. В журналах из перечня ВАК РФ по специальности 1.2.1 – «Искусственный интеллект и машинное обучение» — 2, в журналах из перечня ВАК по иным специальностям — 3, в изданиях, индексируемых в международных базах Scopus и Web of Science — 13. Результаты отражены в 3 отчётах НИР.

Личный вклад. Все результаты, представленные в работах, включённых в диссертационное исследование, получены лично соискателем под непосредственным научным руководством кандидата физико-математических наук, доцента Аверкина Алексея Николаевича.

В рамках исследования были разработаны оригинальные методики анализа ОКТ-изображений с применением объяснительного ИИ [3–7], модифицированы архитектуры нейронных сетей [6; 8], предложены гибридные подходы

с использованием нечётких множеств [9—11], а также реализованы алгоритмы автоматизированной оценки качества изображений [12]. Отдельное внимание уделено задачам анализа фундус-снимков и классификации стадий диабетической ретинопатии [13; 14], включая формализацию правил и интеграцию объяснительного ИИ в системы поддержки принятия решений [15—22].

Структура и объем диссертации. Диссертация состоит из введения, 3 глав, заключения. Полный объём диссертации составляет 143 страницы, включая 18 рисунков и 18 таблиц. Список литературы содержит 228 наименований.

Глава 1. Методика полуавтоматического формирования проблемно-ориентированных коллекций данных

1.1 Теоретические основы создания медицинских коллекций данных

1.1.1 Требования к формированию коллекций медицинских данных

В задачах разработки алгоритмов глубокого обучения для автоматического анализа медицинских изображений ключевое значение имеет качество, воспроизводимость и корректность исходных данных. Для этого требуется формировать специальные коллекции медицинских данных, соответствующие установленным в Российской Федерации стандартам и регламентам, с учётом технических, юридических и этико-информационных ограничений.

Под коллекцией медицинских данных в данной работе понимается организованное множество взаимосвязанных изображений, метаданных и аннотаций, формируемое по единым протоколам сбора, предобработки, хранения и доступа. Такая коллекция отличается от произвольного набора тем, что включает формализованную структуру, описание источников, протоколы экспертной разметки, документированную историю обработки, фиксированные обучающие и тестовые подмножества.

Обработка медицинских изображений в целях обучения нейросетевых моделей осуществляется в рамках требований Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных» [23]. Закон устанавливает принципы законности, минимизации, ограниченности цели обработки и необходимости защиты данных. В медицинской сфере дополнительные регламенты определены приказами Минздрава России [24], а также методическими документами по реализации Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ).

Порядок обезличивания медицинских изображений, содержащих персональные данные, регламентирован ГОСТ Р 55036–2012 [25], устанавливающим

архитектуру процедур псевдонимизации. Дополнительно применяется Приказ Роскомнадзора № 996 [26], определяющий допустимые методы обезличивания и контроль остаточного риска реидентификации.

Однако при использовании заранее обезличенных открытых датасетов, изначально опубликованных без привязки к личности пациента, требование локальной реализации процедур обезличивания может быть существенно смягчено. Такие данные могут быть использованы в качестве визуальной базы для повторной экспертной аннотации и формирования новой, клинически значимой коллекции.

Сформированная коллекция медицинских изображений, предназначенная для обучения, валидации и тестирования нейросетевых моделей, должна удовлетворять следующим критериям:

- Репрезентативность. Коллекция должна охватывать реальные вариации клинических сценариев, источников, оборудования и патологий, обеспечивая устойчивость модели к смещению данных [27].
- Полнота и целостность. Необходимо сохранять метаданные: идентификатор исследования, параметры съёмки, модель прибора, номер серии. Целостность обеспечивается контрольными суммами, управление версиями – через ведение журналов изменений.
- Контроль качества изображений. Применяется отбор изображений по техническим критериям (резкость, контрастность, отсутствие артефактов), включая автоматизированную предфильтрацию [28].
- Аннотирование. Аннотации выполняются врачами-экспертами по согласованной методологии. Требуется стадия перекрёстной валидации с консенсусной корректировкой.
- Разделение на подмножества. Деление на обучающую, валидационную и тестовую выборки должно производиться по пациентам, не по изображениям. Структура фиксируется и документируется.
- Формат хранения. Используется формат DICOM согласно ГОСТ Р 12052–2009 [29], допускается параллельное хранение изображений в PNG/TIFF с метаданными в JSON.
- Справочники и онтологии. Используемая терминология должна быть согласована с НСИ в составе ЕГИСЗ [27].

Методически корректное формирование коллекций медицинских изображений требует соблюдения действующих нормативов РФ, ГОСТ по форматам,

а также практик обеспечения качества, безопасности и воспроизводимости, что создаёт надёжную основу для обучения, валидации и последующего внедрения моделей искусственного интеллекта в клиническую практику.

1.1.2 Обзор литературы и источников

Развитие компьютерного анализа медицинских изображений сопровождалось переходом от локальных, узко специализированных наборов к масштабным мультицентровым коллекциям и многоступенчатым конвейерам формирования данных. Ранние систематические обзоры фиксируют, что качество коллекций определяет верхнюю границу достижимой точности алгоритмов и устойчивость к доменным сдвигам [30].

Наиболее традиционный подход опирается на ручную экспертную аннотацию по заранее регламентированным протоколам с двойным чтением и консенсусом «золотого стандарта»; межэкспертное согласие оценивают статистическими мерами и (или) алгоритмами объединения разметок, что продемонстрировано в бенчмарке BRATS [31]. Этот метод обеспечивает наивысшее клиническое доверие к меткам, но характеризуется высокой трудоёмкостью, ограниченной масштабируемостью и чувствительностью к скрытым смещениям выборки.

Для снижения издержек сформировался класс полуавтоматизированных конвейеров: первичная разметка создаётся вспомогательными моделями, после чего уточняется экспертом; активное обучение приоритизирует наиболее информативные случаи для разметки, что даёт выигрыш по стоимости и времени [32]. Одновременно широкое распространение получили weakly-supervised подходы, где метки извлекаются из текстовых отчётов и кодов нозологий: крупные рентгенологические наборы MIMIC-CXR, CheXpert и ChestX-ray8 построены на правилах обработки радиологических описаний и допускают нестандартизированные ярлыки, позволяя обучать модели на сотнях тысяч снимков при минимальном участии экспертов [33–35].

Недостаток таких меток – их неточность; потому всё чаще применяются процедуры очистки и калибровки неопределённости, например на основе confident learning [36] или инструментов обработки отрицаний и модальностей в

медицинских текстах [37]. Параллельно активно развиваются подходы самообучения и обучения с частично размеченными данными, при которых большая часть обучающего корпуса остаётся неразмеченной, а разметка применяется лишь к ограниченному подмножеству. Такой подход позволяет существенно снизить трудозатраты при формировании датасетов и одновременно повысить переносимость моделей между различными учреждениями и типами оборудования [30].

Важным подходом стало использование синтетических данных: генеративные модели применяют для балансировки редких классов и более лёгкого расширения обучающих выборок, что в ряде задач улучшает распознавание и сегментацию [38; 39]. Вместе с тем синтетика несёт риск переноса артефактов генератора и деградации обобщающей способности, если её доля чрезмерна. На инфраструктурном уровне важную роль играют открытые репозитории и бенчмарки, такие как TCIA, BRATS, MSD, MIMIC-CXR и CheXpert. Они формируют стандарты документирования, аннотирования и оценки моделей, а также позволяют выявлять типовые проблемы и ограничения в дизайне соревнований по медицинскому анализу изображений [40—42].

Сопоставление методик позволяет выделить устойчивые закономерности. Ручная аннотация остаётся эталоном по клинической достоверности, но проигрывает по масштабируемости; полуавтоматизированные циклы с активным обучением дают компромисс между стоимостью и качеством; weakly-supervised подходы и самообучающиеся схемы обеспечивают наилучший масштаб ценой процедур контроля шума и неопределённости; синтетические данные полезны для балансировки, но требуют внешней проверки.

Сквозным ограничением большинства публикаций является то, что контур формирования датасетов практически не включает методы объяснительного искусственного интеллекта и базовые модели общего назначения. Хотя методы ОИИ (атрибуции, визуализации, локальные аппроксимации) активно исследуются в контексте интерпретации моделей [43], они редко применяются как инструмент валидации аннотаций, диагностики ошибок. Аналогично, несмотря на наличие крупных предобученных моделей, в большинстве публикаций отсутствует их систематическое использование для предварительной разметки и отбора.

Таким образом, надёжные датасеты формируются как гибридные, мультицентровые и документированные коллекции с внешней валидацией, тогда как

вопросы интеграция ОИИ и базовых моделей в контур качества остаётся открытой задачей [42; 44—46].

1.2 Описание предложенной методики

1. Загрузка изображения. На вход конвейера поступает исходный ОКТ В-срез. На этом шаге выполняется проверка формата и целостности файла, а также регистрация версии данных.

2. Нейросетевая оценка качества изображения. Классификатор на основе искусственной нейронной сети определяет пригодность снимка. В качестве сопроводительной информации формируются карты значимости, визуализирующие зоны, повлиявшие на решение. Изображения, признанные некачественными, автоматически исключаются или отправляются на повторную верификацию.

3. Предварительная обработка изображений. Выполняются масштабирование с сохранением пропорций, нормализация интенсивностей и фильтрация шума. Цель – приведение всех данных к единому стандарту для дальнейшей обработки и сегментации.

4. Предварительная аннотация на основе базовой модели с визуальными промптами. Базовая модель получает изображение и начальные подсказки (точки или рамки) и формирует черновые маски предполагаемых объектов. Это позволяет автоматически сгенерировать первичную структуру аннотации, минимизируя объём ручной работы.

5. Экспертная аннотация офтальмологом. Специалист вручную проверяет и корректирует автоматически сгенерированные маски: уточняет границы, удаляет ложные участки, разъединяет пересекающиеся объекты. Аннотация осуществляется в формате сегментации экземпляров.

6. Перекрёстная валидация аннотаций. Второй эксперт (офтальмолог, имеющий научную степень) независимо проверяет разметку. При наличии расхождений проводится обсуждение и достигается согласованное решение.

7. Сохранение в коллекцию данных. Для каждого изображения сохраняется полный набор: оригинал, финальные маски. Таким образом формируется

стандартизированная коллекция, ориентированная на задачу сегментации экземпляров биомаркеров диабетического макулярного отёка.

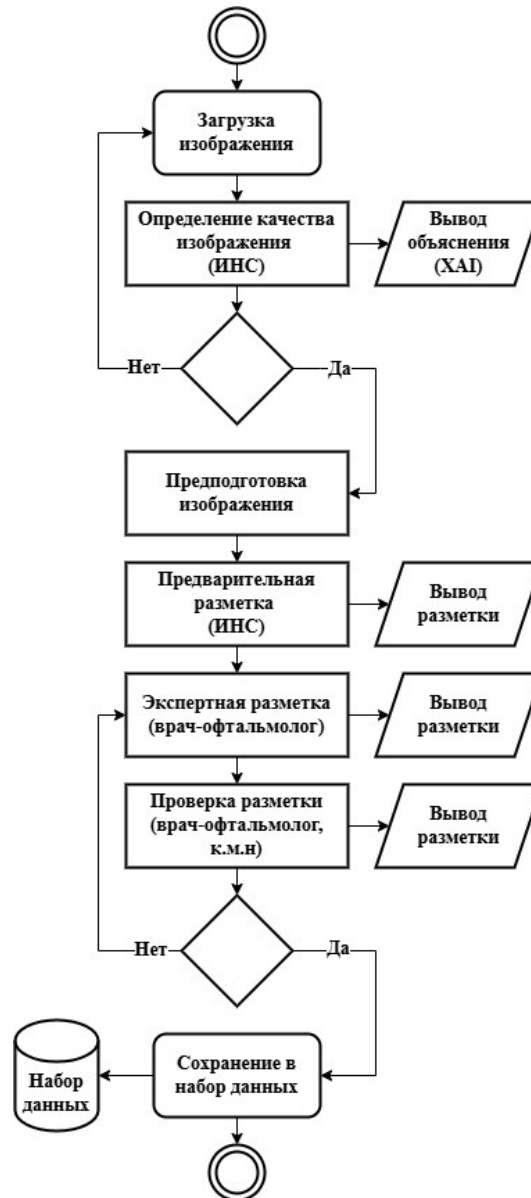


Рисунок 1.1 — Блок-схема предлагаемой методики полуавтоматического формирования коллекции данных

Предложенная методика отличается комплексностью и новизной: объединены автоматическая оценка качества с использованием объяснимой ИИ-модели, модуль предварительной аннотации на основе визуальных промптов и перекрёстная экспертная валидация. Методика легко масштабируется и может быть адаптирована к другим задачам анализа медицинских изображений.

1.3 Существующие коллекции данных ОК-томограмм

Одним из первых и принципиально важных этапов методики полуавтоматического формирования проблемно-ориентированных коллекций данных, изложенной в данной главе, является отбор исходных изображений, обладающих клинической значимостью и пригодных для последующей разметки. В контексте настоящей работы эта задача решается за счёт выборки изображений из уже существующих открытых коллекций, предназначенных для задачи классификации.

Несмотря на широкий охват патологий макулярной зоны и высокое техническое качество, большинство таких коллекций не содержит необходимых классов аннотаций, требуемых для сегментации биомаркеров диабетического макулярного отёка (ДМО). В частности, в них отсутствуют маски и локализованные метки для структур типа DRIL, EZD, IRC и других, что делает невозможным прямое использование этих данных в задачах пространственного анализа и количественной оценки патологий.

В связи с этим ключевым подходом стало использование классификационных коллекций в качестве источника изображений, удовлетворяющих требованиям по анатомической области (макула), техническому качеству и типу патологии, с целью их дальнейшей экспертной реаннотации в формате сегментации экземпляров. Для этого был проведён анализ доступных классификационных коллекций, на основе которого определены подходящие наборы для отбора изображений.

Коллекция Kermany (OCT2017) является одной из наиболее известных и широко используемых в задачах классификации. Данные собраны в клиниках США и Китая, в том числе в Shiley Eye Institute (UCSD) и Beijing Tongren Eye Center, с использованием устройства Heidelberg Spectralis OCT. Изображения сгруппированы по четырём категориям: CNV (хороидальная неоваскуляризация), DME (диабетический макулярный отёк), DRUSEN (возрастные изменения) и NORMAL. Объём данных составляет от 84 000 до более 100 000 В-срезов [47; 48]. Разметка представлена на уровне классов изображений, без локализованных масок.

Коллекция OCTID была сформирована в Sankara Nethralaya (Индия) с использованием устройства Cirrus HD-OCT (Zeiss). Она включает более 500

изображений, классифицированных по категориям: NORMAL, AMD, DR, CSR, МН. Все изображения имеют высокое аксиальное разрешение (до 5 мкм), ширину скан-линии 2 мм и размер 512×1024 пикселя. Некоторые изображения сопровождаются ручной аннотацией анатомических слоёв сетчатки. Несмотря на относительно небольшой объём, OCTID содержит высококачественные фовеа-центрированные В-срезы, представляющие интерес для последующей ручной сегментации биомаркеров [49].

Коллекция OCTDL включает более 2000 В-срезов, полученных с помощью устройства Optovue Avanti RTVue XR. Набор охватывает широкий спектр макулярных и сосудистых патологий, включая AMD, DME, ERM, RAO, VID и другие. Аннотация проводилась поэтапно: предварительная классификация, двойная разметка офтальмологами и финальная верификация экспертом. Разметка предоставлена только на уровне диагноза, без экзemplярных масок. Благодаря разнообразию состояний и стандартизированным протоколам съёмки, коллекция подходит в качестве источника изображений для последующей реаннотации [50].

Анализ открытых классификационных коллекций показал, что они обладают клинически разнообразным содержанием и высоким техническим качеством изображений, но не содержат аннотаций, достаточных для задач сегментации экземпляров биомаркеров ДМО. Это исключает их прямое использование в задачах количественного анализа. В связи с этим был реализован отбор изображений по критериям качества, анатомической релевантности и типу патологии, за которым следовала экспертная аннотация. Такой подход позволил сформировать специализированную коллекцию, пригодную для обучения и объективной валидации современных моделей сегментации в офтальмологии.

1.4 Предподготовка изображений

Современные методы автоматизированной и полуавтоматизированной обработки медицинских изображений требуют формально однородного входного датасета как на уровне геометрических характеристик, так и с точки зрения распределения интенсивностей. Это особенно критично при работе с изображениями оптической когерентной томографии (ОКТ), где структура дан-

ных может варьироваться в зависимости от условий сканирования, моделей устройств и индивидуальных анатомических особенностей пациента. Предподготовка изображений перед аннотированием направлена на устранение этой вариабельности, обеспечивая условия для точной и воспроизводимой разметки, которая в дальнейшем используется при обучении и валидации сегментационных моделей [51; 52].

В условиях различной яркости и контраста изображений, полученных с различных ОКТ-сканеров, нормализация позволяет привести значения интенсивностей к единому масштабу, что критически важно для обеспечения согласованности аннотаций и улучшения сходимости моделей. Наиболее простой и часто применяемой формой нормализации является min–max-нормировка, определяемая как:

$$I_{\text{norm}}(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}}, \quad (1.1)$$

где $I(x, y)$ – исходное значение интенсивности пикселя, I_{\min} и I_{\max} – минимальные и максимальные значения яркости в изображении. В результате интенсивности преобразуются в диапазон $[0, 1]$, что делает данные совместимыми с типичными требованиями большинства нейросетевых фреймворков и облегчает визуальное восприятие при ручной аннотации [52].

Изображения, получаемые с различных ОКТ-устройств, могут иметь произвольные размеры, что осложняет пакетную обработку и построение общих масок. Приведение изображений к фиксированному размеру (M, N) достигается посредством масштабирования с сохранением пропорций. Формально преобразование координат можно описать как:

$$I_{\text{resized}}(x', y') = I(\alpha x', \beta y'), \quad \alpha = \frac{m}{M}, \quad \beta = \frac{n}{N}, \quad (1.2)$$

где (m, n) – исходные размеры изображения, а (x', y') – координаты в результирующем пространстве. Данная процедура обеспечивает сопоставимость аннотаций по различным пациентам и устройствам без искажения анатомических структур [51].

ОКТ-снимки содержат выраженный спекл-шум, обусловленный интерференционной природой метода визуализации [53]. Для его подавления применяют медианный фильтр, который эффективно устраняет точечные шумы без размытия границ анатомических структур. Пусть $\mathcal{N}_{k \times k}(x, y)$ – окно размером $k \times k$,

центрированное в точке (x, y) . Тогда медианная фильтрация задаётся выражением:

$$I_{\text{med}}(x, y) = \text{median} \{I(u, v) \mid (u, v) \in \mathcal{N}_{k \times k}(x, y)\}. \quad (1.3)$$

В отличие от линейных фильтров, медианный сохраняет резкие границы, что критично при ручной разметке границ слоёв сетчатки, патологических включений или отслоений [54].

Процесс предподготовки ОКТ-снимков перед аннотированием должен быть максимально стандартизированным и в то же время щадящим для изображений. Применение min–max-нормировки, унификация размеров изображений и медианная фильтрация шума позволяют создать однородный и качественный набор данных, пригодный для точной ручной аннотации и последующего использования в обучении нейросетевых моделей сегментации. Выбор этих операций основан на балансе между сохранением анатомически значимой информации и устранением мешающих факторов, что подтверждается современными исследованиями в области медицинского компьютерного зрения [55–57].

1.5 Определение качества изображения

Задача фильтрации изображений по техническому качеству является обязательным этапом формирования обучающих выборок в медицинских задачах компьютерного зрения. Для автоматизации этого этапа классификации ОКТ-снимков на качественные и некачественные предлагается формализация в виде задачи бинарной классификации. Под "качественными" понимаются изображения, обладающие достаточной контрастностью, отсутствием артефактов и чёткой стратификацией слоёв сетчатки, а к "некачественным" отнесены срезы с расфокусировкой, шумами, артефактами движения и низким сигналом.

Пусть задана выборка изображений $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^N$, где $I_i \in \mathbb{R}^{H \times W}$ – входное изображение, $y_i \in \{0, 1\}$ – бинарная метка (1 – качественное изображение, 0 – некачественное). Модель $f_{\theta}(I) \in [0, 1]$ аппроксимирует вероятность принадлежности изображению к классу "качественные". Обучение модели осуществляется посредством минимизации функции бинарной кросс-энтропии:

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^N [y_i \log f_{\theta}(I_i) + (1 - y_i) \log(1 - f_{\theta}(I_i))]. \quad (1.4)$$

В качестве базовой архитектуры используется свёрточная нейронная сеть с интеграцией модуля пространственного внимания, предложенная в работе [5]. Данный модуль позволяет выделить релевантные фрагменты изображения за счёт пространственной фокусировки, усиливая сигналы от анатомически значимых областей и подавляя шумы. Архитектура включает 8 свёрточных слоёв, слой нормализации, пулинг и полносвязную классификационную голову. Модуль внимания встроен перед выходным слоем классификатора.

По результатам валидации на тестовом подмножестве получены следующие показатели точности: ассигасу составила 96,2 % на классе "качественные изображения" и 95,4 % на классе "некачественные изображения". Такие значения подтверждают надёжность модели в задаче предфильтрации изображений.

Предложенная модуль системы используется как входной фильтр пайплайна: изображения, классифицированные как некачественные, автоматически исключаются из процесса аннотирования и обучению, что позволяет существенно сократить долю шумных и малополезных примеров в обучающей выборке и повысить итоговую стабильность сегментационных моделей.

1.6 Объяснительный искусственный интеллект

1.6.1 Объяснительный искусственный интеллект: общая характеристика

Объяснительный искусственный интеллект (ОИИ, англ. Explainable Artificial Intelligence, XAI) – это направление в области ИИ, которое объединяет подходы, обеспечивающие интерпретируемость и прозрачность моделей, обычно рассматриваемых как «чёрные ящики». Его основная цель – сделать процессы принятия решений алгоритмами ИИ понятными для человека, формализуемыми и воспроизводимыми. В условиях, когда глубокие модели

достигают выдающихся результатов за счёт увеличения сложности, ОИИ становится необходимым связующим звеном между машинной эффективностью и человеческой интерпретацией [17].

В литературе принято различать несколько ключевых понятий. Термин «прозрачность» (англ. transparency) характеризует степень понимания структуры и параметров модели. «Интерпретируемость» (англ. interpretability) означает возможность объяснить внутренние процессы модели, тогда как «объяснимость» (англ. explainability) относится к результатам методов, обеспечивающих понимание того, почему была принята определённая моделью гипотеза [58; 59]. ОИИ-методы подразделяются на встроенные (model-specific) и внешние (model-agnostic), а также по времени применения — на предварительные, встроенные в обучение и постфактум-объяснения, применяемые после получения вывода [60].

Согласно обзору [61], на практике преобладают постфактум-объяснения, обеспечивающие визуальную интерпретацию решений готовых моделей, особенно в случае сверточных нейросетей. Это обусловлено тем, что архитектуры современных глубоких моделей часто не проектируются с учётом интерпретируемости, и внесение изменений в их структуру может негативно сказаться на точности. Вместо этого используется визуализация активаций (например, Grad-CAM), анализ значений градиентов, построение суррогатных моделей и другие методы, позволяющие получить представление о работе модели без необходимости вмешательства в её архитектуру.

Работа [17] подчёркивает, что ОИИ не является изолированной технологией, а интегрируется в состав гибридных интеллектуальных систем как модуль, обеспечивающий визуальную расшифровку предсказаний нейросетей. В частности, визуальное объяснение формируется в виде тепловой карты, где цветовая интенсивность отражает степень вклада отдельных областей изображения в итоговый результат модели. Такая форма объяснения используется как интерфейс между пространственными признаками и восприятием пользователя. Применение методов ОИИ в контексте анализа медицинских изображений рассматривается как ключевое направление развития клинических систем поддержки принятия решений.

Особое значение объяснительный ИИ приобретает в медицине, где последствия ошибок критичны, а доверие специалистов к результатам предсказаний модели напрямую зависит от наличия обоснованных и понятных объяснений.

Современные регуляторные требования (в том числе этические и юридические) требуют обоснованности и прозрачности алгоритмов, используемых в диагностике и оценке риска [59; 61]. Включение ОИИ в медицинские ИИ-системы позволяет не только повысить доверие со стороны врачей, но и выявлять случаи систематического смещения или переобучения модели, а также строить понятные пользователю отчёты о принятых решениях.

Таким образом, ОИИ выступает не только как инструмент валидации моделей, но и как методологическая основа для построения этически обоснованных, регламентируемых и доверенных систем искусственного интеллекта в медицине.

1.6.2 Метод объяснительного ИИ САМ

Метод САМ (англ. Class Activation Mapping) предназначен для построения пространственных карт, отражающих вклад признаков в решение модели по каждому классу, позволяющих выявлять области входного изображения, наиболее существенно повлиявшие на предсказание нейросети. Его основная идея заключается в том, что веса линейного классификатора, обученного на признаках последнего сверточного слоя, можно спроецировать обратно на пространственные карты активаций этого слоя, что делает возможным восстановление пространственного распределения вклада признаков в оценку выбранного класса [62].

САМ применяется в архитектурах, где после последнего сверточного слоя используется операция глобального усреднения по пространству (Global Average Pooling, GAP), за которой следует линейный классификатор. Пусть последний сверточный слой формирует набор карт признаков $\{A^k \in \mathbb{R}^{h \times w}\}_{k=1}^K$. Для каждого канала вычисляются усредненные по пространству значения:

$$F_k = \frac{1}{Z} \sum_{i=1}^h \sum_{j=1}^w A_{ij}^k, \quad Z = h \cdot w. \quad (1.5)$$

Значение выходного сигнала (логита) для класса c вычисляется как линейная комбинация усреднённых признаков:

$$y^c = \sum_{k=1}^K w_k^c F_k + b^c, \quad (1.6)$$

где w_k^c – вес связи канала k с классом c , а b^c – смещение. Подставив (1.1) в (1.2), получаем выражение:

$$y^c - b^c = \frac{1}{Z} \sum_{i=1}^h \sum_{j=1}^w \left(\sum_{k=1}^K w_k^c A_{ij}^k \right), \quad (1.7)$$

где выражение в скобках интерпретируется как карта активаций класса c :

$$L_{\text{SAM}}^c(i, j) = \sum_{k=1}^K w_k^c A_{ij}^k. \quad (1.8)$$

Таким образом, выход модели можно представить в виде суммы локальных вкладов по пространству карты признаков. На практике карту L_{SAM}^c билинейно масштабируют до размеров исходного изображения и накладывают в виде тепловой маски, что позволяет визуализировать вклад различных областей изображения в предсказание модели [62].

Тепловая карта САМ является классо-специфической и подчеркивает те участки изображения, признаки которых, взвешенные коэффициентами w_k^c , наиболее усиливают значение логита. Использование последних сверточных карт делает метод устойчивым к локальному шуму и позволяет акцентировать внимание на высокоуровневых семантических признаках. Простота вычислений и отсутствие необходимости в градиентной информации объясняют широкое распространение метода в задачах с частичной разметкой данных (weakly supervised localization) и в экспресс-аудите моделей [62; 63].

Однако метод имеет ряд ограничений. Во-первых, он зависит от архитектурных предпосылок: требуется наличие глобального усреднения и линейного классификатора непосредственно над сверточными картами. Во-вторых, разрешение карт ограничено размерностью пространственных признаков последнего сверточного слоя, что снижает точность локализации мелких объектов [64]. Наконец, САМ усиливает доминирующие каналы и может подавлять альтернативные признаки, что приводит к неполному объяснению. В отличие от градиентных подходов, например Grad-CAM, метод САМ не учитывает локальную производную целевой функции и потому менее универсален, но при

соблюдении архитектурных условий он дает объяснения, непосредственно согласованные с линейной моделью над признаками [65].

Значимость САМ для задач интерпретируемого машинного обучения заключается в том, что он демонстрирует прозрачную связь между признаками и решением модели: веса w_k^c отражают важность каналов, а карты активаций A^k задают их пространственное распределение. Такая факторизация позволяет выявлять неявные зависимости, например использование фоновых артефактов вместо целевых объектов, и способствует улучшению качества данных и архитектур. САМ является базовым методом, который часто применяется в комбинации с другими подходами (saliency-карты, LRP, интегрированные градиенты) для перекрестной валидации объяснений и повышения доверия к результатам глубоких моделей [66–69].

1.7 Предварительная разметка с помощью базовой модели

Предварительная разметка предназначена для ускорения формирования обучающих выборок за счёт автоматического получения черновых масок с последующей минимальной правкой экспертом. В контексте ОКТ-снимков эффективной основой для такого шага выступает базовая модель сегментации SAM (Segment Anything Model), способная по визуальному промπτу (точка или рамка) выделять объекты произвольной формы. На практике это уменьшает объём ручной обводки и повышает воспроизводимость аннотаций при подготовке клинических датасетов [70–73]. Полуавтоматические конвейеры, в которых модель выполняет предварительную сегментацию, а эксперт – верификацию и коррекцию, уже показали существенное сокращение трудозатрат при сохранении качества разметки [72; 73]. Результаты предварительной разметки ОКТ-снимка представлены на рис. 1.2.

Пусть $I \in \mathbb{R}^{H \times W}$ – ОКТ-изображение, $P = \{p_k\}_{k=1}^K$ – множество визуальных промπτов (точки $p_k = (x_k, y_k)$ или прямоугольные рамки), а $f_{\text{SAM}}(\cdot)$ – отображение модели SAM из пространства изображений и промπτов в множество бинарных масок $M = \{m_\ell\}_{\ell=1}^L$, $m_\ell \in \{0,1\}^{H \times W}$. Базовая формализация шага предварительной разметки:

$$M = f_{\text{SAM}}(I, P). \quad (1.9)$$

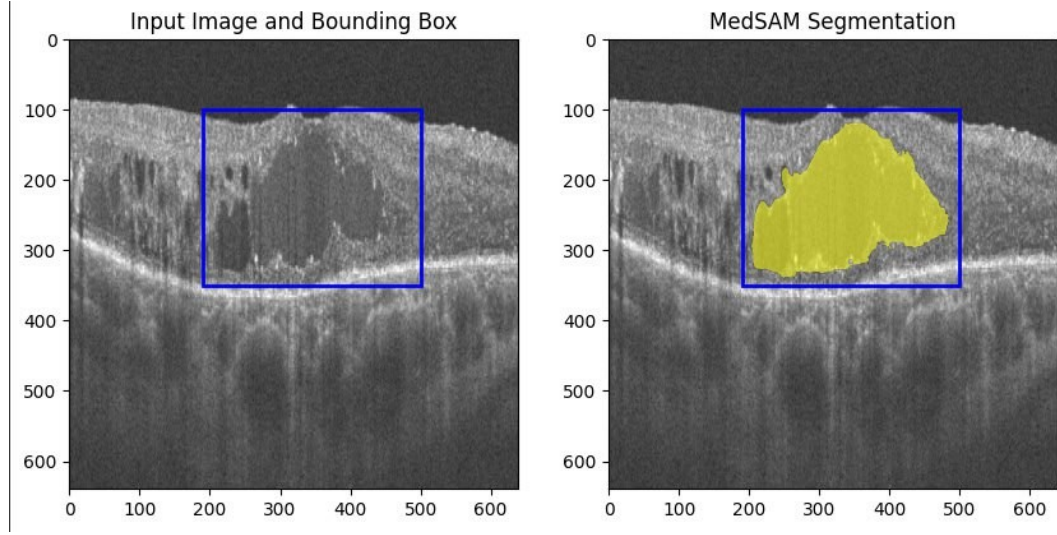


Рисунок 1.2 — Результат предварительной разметки с помощью базовой модели

При использовании SAM возвращает вероятностные карты $S = \{s_\ell\}_{\ell=1}^L$, $s_\ell \in [0,1]^{H \times W}$, которые сортируются по порогу:

$$m_\ell(x,y) = \{s_\ell(x,y) \geq \tau\}, \quad \tau \in (0,1). \quad (1.10)$$

В сценарии интерактивной предразметки промпт можно подбирать как решение задачи поиска, максимизирующей качество выходной маски по выбранному критерию Q :

$$P^* = \arg \max_{P \in \mathcal{A}} Q(f_{\text{SAM}}(I,P), g), \quad (1.11)$$

где \mathcal{A} – допустимое множество промптов, а g – эталонная маска.

Интеграция SAM как базовой модели в предразметку для ОКТ обеспечивает быстрый выпуск черновых масок слоёв и патологических структур, которые затем проходят верификацию и согласование специалистами. На собранном материале с восемью биомаркерами (DRIL, ERM, EZD, HE, HF, IRC, SRF, VMT) предварительная сегментация SAM показала практическую пользу: экспертная правка сводилась к локальной корректировке границ и устранению ложноположительных областей, после чего маски принимались в датасет. Общая трудоёмкость аннотации снизилась, а воспроизводимость повысилась.

1.8 Экспертная разметка

Экспертная разметка представляет собой завершающий этап формирования клинически валидной коллекции изображений. После предварительной автоматической сегментации черновые маски поступают на доработку двум врачам-офтальмологам экспертам в данной области, один из которых кандидат медицинских наук, обладающим опытом интерпретации анатомо-морфологических структур глазного дна. На этом этапе выполняется детализированная проверка, исправление ошибок предварительной разметки и формирование окончательного варианта аннотаций, пригодного для последующего использования в обучении и валидации моделей анализа медицинских изображений.

Экспертная разметка включает следующие операции:

- проверка технического качества изображений и соответствия аннотированных объектов целевым клиническим категориям;
- уточнение границ патологических структур, устранение ложноположительных и добавление пропущенных объектов;
- нормализация структуры масок, разделение слипшихся экземпляров и унификация классов;
- финальная верификация, включающая повторный просмотр и контроль согласованности принятых решений.

Ключевая задача данного этапа – обеспечение соответствия аннотаций клиническим критериям и морфологическим стандартам, принятым в офтальмологии. Разметка выполняется на уровне экземпляров: каждый отдельный очаг патологии выделяется как отдельный объект с присвоением класса и атрибутов. Это обеспечивает возможность количественной оценки признаков заболевания и формирует основу для обучения сегментационных моделей, ориентированных на интерпретируемые и стандартизированные результаты.

Экспертная разметка является связующим звеном между предварительными нейросетевыми предсказаниями базовой модели SAM и конечным формированием клинически достоверной коллекции данных. Именно она обеспечивает объективность, воспроизводимость и доверие к данным, что критически важно для задач построения и оценки интеллектуальных систем в офтальмологической практике.

1.9 Характеристики полученной коллекции данных

Разработанный набор данных предназначен для обучения и объективной валидации моделей сегментации в задаче автоматизированной диагностики диабетического макулярного отёка по изображениям оптической когерентной томографии. В коллекцию включено 900 полноразмерных В-срезов глазного дна с разрешением 512×1024 пикселя, полученных из открытых источников и прошедших полную ручную аннотацию в формате сегментации экземпляров. Разметка выполнена двумя независимыми врачами-офтальмологами с последующей верификацией по протоколу «золотого стандарта».

Всего в датасете содержится 6752 экземпляра биомаркеров, относящихся к восьми ключевым классам патологических изменений, связанных с ДМО. Распределение по классам: HE – 1140, HF – 882, IRC – 2310, EZD – 516, SRF – 502, VMT – 403, DRIL – 375, ERM – 624. Наличие широкого диапазона представленных структур от очагов отёка и экссудатов до дезорганизации внутренних слоёв, что обеспечивает как клиническую репрезентативность, так и полноту охвата возможных проявлений заболевания. На рис. 2.6 представлен пример изображений из описываемой коллекции данных.

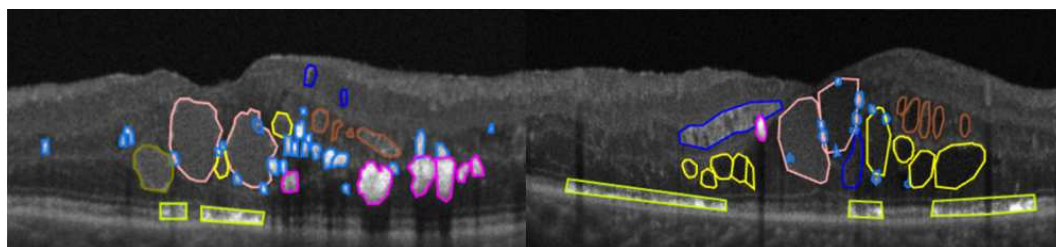


Рисунок 1.3 — Пример изображений коллекции данных биомаркеров ДМО

Каждая маска отражает точные границы отдельного очага, включая случаи перекрытия и сложной геометрии. Это позволяет использовать набор для обучения и тестирования современных архитектур сегментации, включая модели, ориентированные на мелкие и трудноразличимые структуры. Данная коллекция данных является первой, в котором представлены все восемь классов биомаркеров, связанных с диабетическим макулярным отёком, размеченных для решения задачи сегментации экземпляров. Создаёт основу для построения высокоточных систем поддержки принятия решений в офтальмологии.

1.10 Выводы к первой главе

1. Обоснована необходимость создания специализированных коллекций, ориентированных на сегментацию экземпляров биомаркеров, с учётом недостатков существующих наборов.
2. Разработана структура методики, включающая:
 - предварительную фильтрацию изображений на основе нейросетевой оценки качества;
 - автоматизированную предразметку с использованием визуальных промптов;
 - перекрёстную экспертную аннотацию.
3. Методика опирается на формализованный пайплайн, обеспечивающий воспроизводимость, снижение трудозатрат и повышение клинической достоверности данных.
4. Практическая реализация методики позволила сформировать первую в своём роде коллекцию ОКТ-снимков для сегментации экземпляров восьми биомаркеров ДМО.
5. Предложенная методика может быть адаптирована для других модальностей и задач анализа биомедицинских изображений, обеспечивая масштабируемый подход к формированию валидационных наборов.

Глава 2. Методы анализа ОК-томограмм

2.1 Теоретические основы анализа ОК-томограмм

2.1.1 Оптическая когерентная томография

Оптическая когерентная томография (англ. Optical Coherence Tomography, ОКТ) – неинвазивный метод послойной визуализации биологических тканей с микронным аксиальным разрешением, основанный на низкокогерентной интерферометрии ближнего инфракрасного излучения. Метод был впервые описан в 1991 году [74].

Благодаря прозрачности оптических сред глаза и выраженной слоистой структуре сетчатки оптическая когерентная томография (ОКТ) быстро стала стандартом визуализации в офтальмологии, поскольку сочетает микронное аксиальное разрешение (порядка 5–6 μm у клинических SD-ОСТ-систем) с полной неинвазивностью и возможностью многократного повторения исследования без лучевой нагрузки [75; 76]. Уникальное соотношение высокой детализации, безопасности и оперативности получения данных позволяют в одном скане визуализировать все десять гистологических слоёв сетчатки и количественно оценивать патологические изменения. Ряд клинических руководств, включая рекомендации NICE по активной форме возрастной макулярной дегенерации, прямо предписывают использовать ОКТ в качестве метода первичного выбора [77].

ОКТ-ангиография (ОСТ-А) дополнительно визуализирует микрососудистую сеть без контрастного агента за счёт анализа движущихся эритроцитов. В совокупности это делает ОКТ незаменимой при макулярной дегенерации, диабетической ретинопатии, глаукоме и других патологиях глаза.

Оптическая когерентная томография основана на низкокогерентной интерферометрии в интерферометре Майкельсона, где излучение делится на объектный и эталонный пучки. Интерференционный сигнал регистрируется, когда оптическая разность хода Δl не превышает длину когерентности источника. Интенсивность описывается:

$$I(\Delta l) = I_s + I_r + 2\sqrt{I_s I_r} \operatorname{Re}[\gamma(\Delta l)] \cos\left(\frac{2\pi\Delta l}{\lambda_0}\right), \quad (2.1)$$

где $\gamma(\Delta l)$ — степень когерентности, λ_0 — центральная длина волны.

Аксиальное разрешение для источника с гауссовым спектром [78]:

$$\Delta z \approx \frac{2 \ln 2}{\pi} \cdot \frac{\lambda_0^2}{\Delta \lambda}, \quad (2.2)$$

что при $\lambda_0 = 840$ нм и спектральной ширине источника $\Delta \lambda = 50$ нм даёт теоретически ≈ 6.2 μm (практически 6–7 μm с учётом дисперсии).

Латеральное разрешение ограничивается дифракцией фокусирующей оптики:

$$\Delta x \approx \frac{1.22 \lambda_0}{2NA}, \quad (2.3)$$

при числовой апертуре объектива $NA = 0.04$ и $\lambda_0 = 840$ нм, $\Delta x \approx 12.8$ μm .

Таким образом, аксиальное разрешение определяется спектром источника и отвечает за разделение слоёв сетчатки, в то время как латеральное зависит от оптики и задаёт поперечную детализацию структур внутри слоёв.

Выделяют следующие типы данных, основанные на способе визуального отображения результатов сканирования:

- А-скан — одномерный профиль отражения по глубине;
- В-скан — поперечное сечение (последовательность А-сканов при сканировании вдоль поверхности);
- en face — аксиальное сечение;
- 3D-кубы (объёмная реконструкция исследуемой области) формируются из серий В-сканов.

Согласно [79] современные ОКТ-системы подразделяются на 3 поколения:

- TD-ОКТ (Time-Domain). Светоделитель расщепляет падающий свет на эталонную длину волны, которая настраивается для различных глубин, в то время как свет образца отражается от наблюдаемой ткани и рекомбинируется на датчике для получения единого профиля ткани.
- SD-ОКТ (Spectral-Domain). Принцип действия аналогичен TD-ОКТ, но датчик преобразован в спектрометр для разделения отражённого света на длины волн (λ) с помощью дифрактора. Комбинация отдельных изображений, полученных с помощью λ , создаёт преобразование Фурье для получения информации о глубине (А-сканирование).

- SS-OCT (Swept-Source). Также основан на БПФ, как и SD-OCT. В качестве источника света используется широкоугольный перестраиваемый лазер. Фотоприемник преобразует передаваемые сигналы в шум (изображения).

Таблица 1 — Классификация ОКТ-систем

Тип	λ_0 (нм)	Скорость, тыс А-сканов/с	Аксиальное разрешение, мкм
TD-OCT	810	0.4 k	8–10
SD-OCT	840	20–52	5–7
SS-OCT	~1050	100–400	5.3

Оптическая когерентная томография остаётся единственным полностью неинвазивным методом, способным отображать все десять гистологических слоёв сетчатки с микронной точностью и при этом давать числовые метрики толщины, объёма и кровообращения. Именно поэтому ОКТ фигурирует как обследование «первой линии» в международных руководствах, например, в «Preferred Practice Pattern» Американской академии офтальмологии по глаукоме [80]. Тем не менее для полной клинической картины офтальмологи используют и другие технологии, каждая из которых восполняет ограничения ОКТ (табл. 2). Ниже кратко перечислены их сильные и слабые стороны чтобы показать, почему именно ОКТ остаётся «опорной» методикой, а не взаимозаменяемой с альтернативами.

Альтернативные технологии дополняют, но не вытесняют ОКТ:

- УБМ спасает, когда роговица или хрусталик непрозрачны, но её пространственное разрешение в 3–4 раза хуже.
- FA остаётся эталоном для оценки утечки, однако инвазивность и системные риски ограничивают частоту её применения.
- SLO и автофлуоресценция дают метаболическую информацию, но без третьего измерения.
- 7T MPT открывает окно к метаболизму сетчатки, однако по детализации проигрывает ОКТ на два порядка.

Таким образом, в большинстве клинических сценариев именно ОКТ обеспечивает критически важную комбинацию микронного разрешения, полной неинвазивности, высокой воспроизводимости и количественного анализа, что

делает её краеугольным камнем как диагностики, так и всех современных исследований, включая настоящий проект.

Несмотря на микронное разрешение и неинвазивность, ОКТ чувствительна к ряду физических явлений, способных исказить изображение или занижать достоверность количественных измерений. Знание этих ограничений критично: неправильная трактовка артефакта может привести к ложноположительной (или ложноотрицательной) диагностике, в то время как своевременное распознавание проблемы позволяет выбрать альтернативный метод (УБМ, FA) либо повторить сканирование после коррекции настроек. Примеры таких ограничений и артефактов:

- Оптические помутнения медиальных сред. Катаракта, корнеальные рубцы и гемофтальм рассеивают и поглощают ИК-излучение, снижая отношение сигнал/шум (SNR) и делая глубокие слои неразличимыми. Для корректировки переходят на более длинную волну (1050–1310 нм), расширяют зрачок, усредняют несколько В-сканов.
- Артефакты движения глаза. Мигание вызывают смещение строчек В-скана, ступенчатую геометрию 3D-кубов и ложные разрывы. Для устранения используют eye-tracking, пост-обработку с выравниванием слоёв и увеличение частоты сканирования.
- Спекл-шум. Когерентная интерференция создаёт зернистую структуру (speckle), маскирующую тонкие детали и ухудшающую работу ИИ-алгоритмов. Для подавления используют частотное усреднение и шумоподавители на основе глубокого обучения.

Комплексное понимание физики артефактов и протоколов их минимизации делает ОКТ не просто высокоразрешающей камерой, но надёжным количественным инструментом.

2.1.2 Диабетический макулярный отёк

Диабетический макулярный отёк (ДМО) определяется как осложнение сахарного диабета, проявляющееся накоплением жидкости в центральной зоне сетчатки, утолщением макулы и нарушением её послойной архитектуры, что приводит к снижению центрального зрения [85]. В контексте визуализации

основным инструментом оценки ДМО является оптическая когерентная томография (ОКТ), позволяющая послойно отображать морфологию макулярной области и количественно оценивать отёк (толщину, объём) и сопутствующие структуры.

К числу характерных ОКТ-биомаркеров ДМО относятся:

- DRIL (disorganization of retinal inner layers) – дезорганизация внутренних слоёв сетчатки; на ОКТ визуализируется как участок с отсутствием чёткой стратификации, без различимых границ между слоями [86].
- ERM (epiretinal membrane) – эпиретинальная мембрана; определяется как гиперрефлективная линия на внутренней поверхности сетчатки, сопровождается деформацией фовеального контура [87].
- EZD (ellipsoid zone disruption) – нарушение зоны эллипсоидов; проявляется как прерывание или исчезновение гиперрефлективной полосы зоны фоторецепторов [88].
- HE (hard exudates) – твёрдые экссудаты; визуализируются как локальные гиперрефлективные включения в наружных слоях сетчатки, часто с эффектом затенения [89].
- HF (hyperreflective foci) – гиперрефлективные фокусы; представлены множественными мелкими точечными гиперрефлективными элементами без заднего затенения, отличающиеся от HE меньшим размером [90].
- IRC (intraretinal cysts) – интраретинальные кисты; определяются как гипорефлективные полости с тонкой стенкой, могут сливаться и деформировать фовеальный контур [91].
- MH (macular hole) – макулярное отверстие; щелевидный или округлый дефект в центральной зоне сетчатки, может быть ламеллярным или полнослойным [87].
- SRF (subretinal fluid) – субретинальная жидкость; гипорефлективное пространство между нейросенсорной сетчаткой и пигментным эпителием, контуры как правило чёткие [92].
- VMT (vitreomacular traction) – витреомакулярная тракция; частичное отслоение стекловидного тела с точечной фиксацией в фовеа, тянущее внутреннюю поверхность сетчатки [87].

Совокупный анализ вышеуказанных ОКТ-биомаркеров позволяет количественно и качественно оценить выраженность ДМО, дифференцировать морфологические подтипы и проводить мониторинг эффективности терапии.

Их типизация лежит в основе алгоритмов компьютерного зрения и стандартизации офтальмологической оценки в клинической практике [85].

2.1.3 Обзор литературы и источников

Для формирования настоящего обзора литературы применены рекомендации PRISMA 2020: поиск проводился в профильных научных базах (PubMed/MEDLINE, Scopus, IEEE Xplore) с использованием ключевых слов (optical coherence tomography/OCT, retina, segmentation, instance segmentation, DRIL, ERM, EZD, HE, HF, IRC, IRF, SRF, VMT). Выполнялись автоматическая дедупликация записей, двухэтапный скрининг (по заголовку и аннотации, затем полнотекстовый анализ). Критерии включения: оригинальные исследования, использующие ОКТ-B-scan для задач сегментации (семантической или экземпляров), опубликованные в рецензируемых журналах и конференционных сборниках; языки публикаций – русский и английский. Исключались работы, посвящённые только классификации без сегментации, использующие иные модальности визуализации, а также статьи без доступа к полному тексту или без указания метрик качества.

Систематизированные публикации по анализу ОКТ-изображений демонстрируют преобладание семейства U-Net и его модификаций в задачах семантической сегментации: Residual U-Net для гиперрефлективных фокусов [93], базовый U-Net для биомаркеров ДМО и сопутствующих патологий [94–96], Attention U-Net для совместной сегментации слоёв и жидкостей [97], U-Net++ для повышения точности локализации [98], а также применение U-Net в специфических когортах [99]. К альтернативам относятся SegNet для эпиретинальной мембраны [100], Mask R-CNN для областей утраты эллипсоидной зоны [101], ансамбли U-Net для сегментации жидкостей [102], архитектуры семейства FCN [103], а также адаптация универсальных сегментаторов SAM [104; 105] и SAM2/MedSAM2 [50]. При этом подавляющее большинство работ ограничивается семантической постановкой без разделения пересекающихся объектов; полноценное применение сегментации экземпляров к ОКТ-биомаркерам ранее не описано, за редкими частными попытками [101]. Сводное сравнение представлено в таблице 3.

Корпус данных охватывает как открытые, так и приватные наборы: RETOUCH [106], AROI [107], OIMHS [108], широко используемые коллекции Kermany и Lu [47; 109], а также современные OCTDL и OCT5k [110; 111]. В ряде работ использованы закрытые клинические наборы различного объёма [94; 101]. Комбинирование крупных открытых баз с клиническими примерами повышает обобщаемость и устойчивость метрик.

По охвату биомаркеров чаще всего сегментируются интравитреальная жидкость и кисты (IRF/IRC) [95; 96; 112], нарушения эллипсоидной зоны (EZD) [100; 101], субретинальная жидкость (SRF) [97; 104], отслойка пигментного эпителия (PED) [102; 103], гиперрефлективные фокусы (HF) [93]. Реже отмечаются ERM и DRIL. Детальный охват классов приведён в таблице 4.

Компоненты объяснимого ИИ в анализируемых работах внедрены ограниченно: за исключением механизмов внимания в Attention U-Net [97], большинство моделей остаются "чёрными ящиками" что снижает клиническую интерпретируемость и подчёркивает перспективность интеграции ХАИ-модулей.

В итоге, рассмотрения данных таблиц 3 и 4 : доминирование U-Net-подобных семантических схем, ограниченная интеграция ОИИ и отсутствие публикаций, формулирующих сегментацию ОКТ-биомаркеров как задачу сегментации экземпляров.

2.2 Описание предлагаемого метода

Предложенный метод представляет собой интерпретируемый конвейер анализа ОКТ-снимков для диагностики диабетического макулярного отёка. Он основан на последовательной интеграции нейросетевых моделей классификации и сегментации с логическим выводом. Структура метода включает следующие этапы:

1. Предобработка изображения. Выполняется нормализация размеров и контрастности изображения, подавление артефактов и автоматическая фильтрация по техническому качеству. Это обеспечивает однородность входных данных и стабильность работы последующих модулей.
2. Бинарная классификация. С использованием нейросетевой модели определяется наличие или отсутствие признаков диабетического маку-

- лярного отёка. В случае отсутствия патологии обработка завершается с информированием пользователя.
3. Формирование визуального объяснения. При положительном результате классификации формируется тепловая карта внимания (Grad-CAM), отображающая значимые для модели участки изображения. Это позволяет проводить аудит решений и повышает доверие к системе.
 4. Сегментация экземпляров. Активируется сегментационный модуль, выполняющий разметку отдельных объектов по классам биомаркеров. Разметка осуществляется в формате экземпляров, с сохранением индивидуальных масок для каждого проявления патологии.
 5. Анализ сегментированных масок. Для каждого класса биомаркеров подсчитывается количество объектов и их суммарная площадь. Полученные численные характеристики формируют основу для последующего анализа.
 6. Сравнение с историей наблюдений. Извлечённые характеристики сопоставляются с ранее сохранёнными результатами обследований конкретного пациента, что позволяет оценить динамику состояния и выявить признаки прогрессирования или регресса.
 7. Логический вывод. На основании текущих и исторических данных применяется логический классификатор, использующий базу правил. Это обеспечивает интерпретируемое заключение о наличии, типе и выраженности патологии.
 8. Формирование результатов. Пользователю предоставляются диагностические метрики, визуализации (включая маски и тепловые карты) и текстовое заключение, пригодное для медицинской документации и последующего мониторинга.

Метод сочетает точность нейросетевых алгоритмов с интерпретируемостью логического анализа и может быть адаптирован для интеграции в клинические информационные системы.

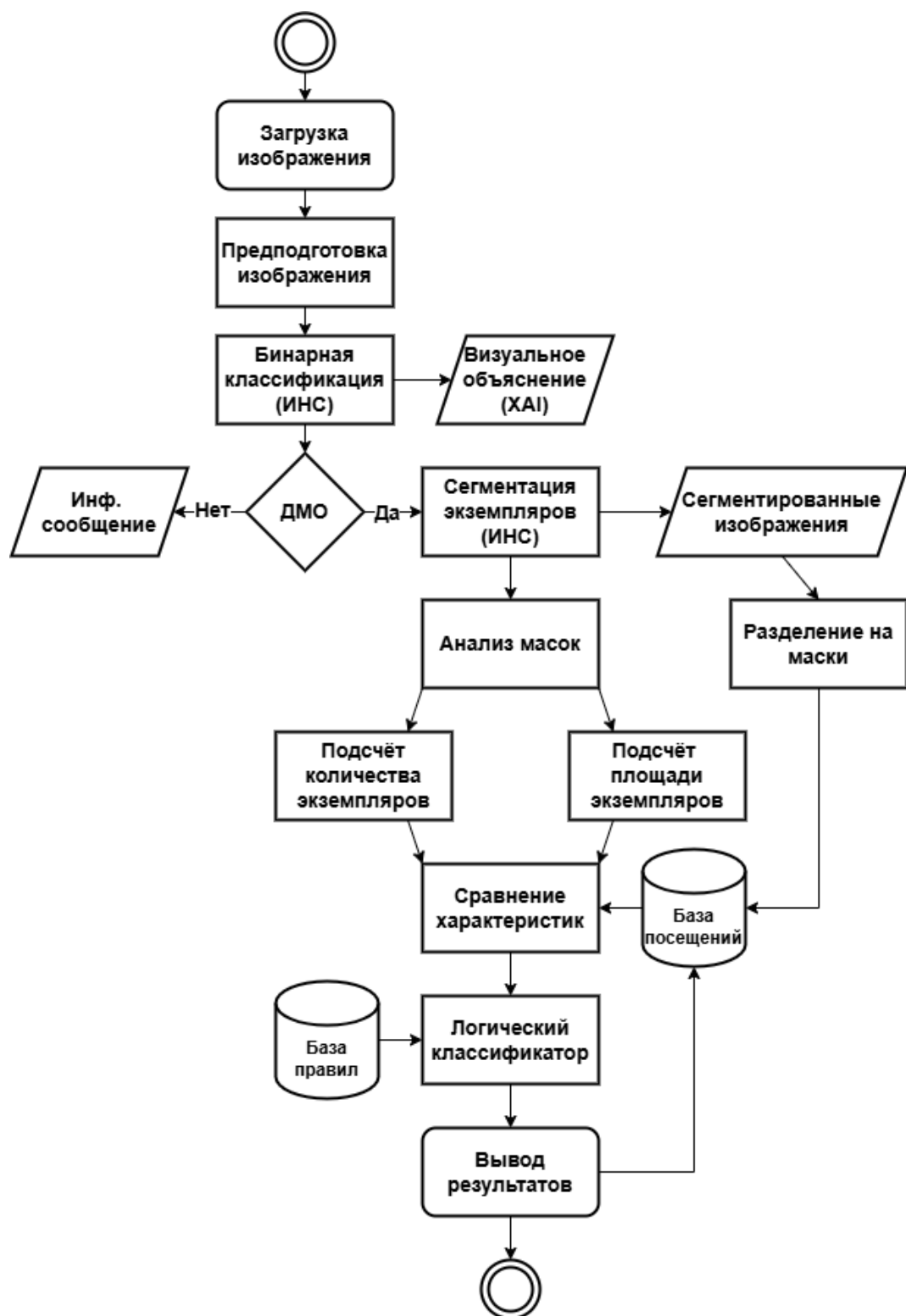


Рисунок 2.1 — Блок-схема предлагаемого метода анализа ОКТ-снимков

2.3 Коллекции данных

В работе использована коллекция OCTDL, содержащая более 2000 В-срезов, полученных на системе Optovue Avanti RTVue XR. Набор охватывает

широкий спектр макулярных и сосудистых патологий, включая AMD, DME, ERM, RAO, VID. Аннотация выполнена поэтапно: первичная классификация, двойная разметка офтальмологами и финальная верификация экспертом. Разметка предоставлена только на уровне диагноза, для задач классификации.

Также сформирован специализированный датасет для сегментации экземпляров биомаркеров ДМО, включающий 900 полноразмерных В-срезов с разрешением 512×1024 пикселя, отобранных из открытых источников и полностью размеченных в формате сегментации экземпляров. Разметку выполняли два независимых офтальмолога с последующей верификацией. Всего отмечено 6752 экземпляра по восьми классам: HE – 1140, HF – 882, IRC – 2310, EZD – 516, SRF – 502, VMT – 403, DRIL – 375, ERM – 624. Коллекция позиционируется как первая, где представлены все восемь указанных классов для решения задачи сегментации экземпляров; пример изображений показан на рис. 1.2. Для обучения моделей применялось разбиение 70 % / 15 % / 15 % (train/val/test).

Оба набора данных использованы в диссертации взаимодополняющим образом: OCTDL – для обучения и объективной валидации этапа классификации, собственная коллекция – для обучения и оценки моделей сегментации экземпляров биомаркеров ДМО.

2.4 Предподготовка изображений

В контексте методов анализа ОК-томограмм предподготовка изображений выполняет важную функцию стандартизации данных, обеспечивая их совместимость с алгоритмами машинного обучения и снижая влияние технических артефактов. Как подробно описано в пункте 1.4, данный этап включает масштабирование томограмм до фиксированного разрешения с сохранением пропорций, нормализацию яркости и сглаживание шума. Однако с учётом специфики входных требований большинства нейросетевых архитектур, в дальнейшем используется только операция нормализации, обеспечивающая приведение значений пикселей к допустимому диапазону без искажения структурной информации. Такой подход позволяет минимизировать вмешательство в исходные данные, сохраняя при этом необходимую совместимость с моделями глубокого обучения.

2.5 Предварительная классификация

2.5.1 Модель EfficientNetB0

EfficientNet — это семейство моделей СНС, предложенное Мингсином Таном и Куоком Ле в 2019 году для более эффективного использования параметров и вычислительных ресурсов при высокой точности классификации изображений [113]. Архитектура этих моделей была оптимизирована с использованием автоматического поиска архитектур (англ. Neural Architecture Search, NAS) в сочетании с принципом комбинированного масштабирования.

В отличие от большинства ранее предложенных архитектур СНС, EfficientNet объединяет в себе три ключевых конструктивных решения: во-первых, компактную и автоматически найденную базовую архитектуру (EfficientNet-B0); во-вторых, систематическое масштабирование всех трёх основных параметров модели — глубины, ширины и разрешения входящих изображений; в-третьих, использование высокоэффективных архитектурных блоков — MBConv-блоков, дополненных механизмом Squeeze-and-Excitation.

Архитектура EfficientNet организована как последовательность стадий, каждая из которых включает несколько однотипных сверточных блоков. При этом, пространственное разрешение входящих тензоров последовательно уменьшается от стадии к стадии, в то время как глубина увеличивается. Формально архитектуру модели можно описать как композицию функций, соответствующих каждой стадии:

$$\mathcal{N} = F_1 \circ F_2 \circ \dots \circ F_K(X), \quad (2.4)$$

где X — входной тензор, F_i — i -й сверточный блок или стадия, K — общее число стадий в архитектуре. Такое представление подчёркивает модульную природу архитектуры EfficientNet, позволяющую реализовать последовательную трансформацию признаков через каскад компактных и эффективно спроектированных сверточных блоков.

Каждая стадия в архитектуре характеризуется следующими параметрами:

- H_i, W_i — высота и ширина входной карты признаков (т.е. пространственное разрешение);
- C_i — число выходных каналов после стадии;
- L_i — количество повторений блока (глубина стадии);
- $k \times k$ — размер ядра свёртки;
- t — коэффициент расширения (expansion factor), который определяет, во сколько раз увеличивается число каналов внутри блока;
- s — шаг (stride), определяющий, происходит ли понижение разрешения (обычно $s = 2$ означает downsampling, $s = 1$ — сохранение размера).

Стадии отличаются друг от друга по следующим признакам:

- Разрешение входа уменьшается от стадии к стадии;
- Число каналов растёт, обеспечивая более абстрактное представление признаков;
- Размер ядра может быть 3×3 или 5×5 ;
- На более поздних стадиях число повторов блоков увеличивается.

Таблица 5. Характеристики архитектуры EfficientNet-B0

Стадия	Операция	Разрешение	Каналы	Повторы	Stride	
1	Conv2D	224×224	32	1	2	
2	MBConv1, SE	112×112	16	1	1	
3	MBConv6, SE	112×112	24	2	2	
4	MBConv6, SE	56×56	40	2	2	
5	MBConv6, SE	28×28	80	3	2	
6	MBConv6, SE	14×14	112	3	1	
7	MBConv6, SE	14×14	192	4	2	
8	MBConv6, SE	7×7	320	1	1	
9	Conv1×1 + Pool + FC	7×7	1280	1	—	

Основной элемент архитектуры EfficientNet — это MBConv-блок, впервые предложенный в MobileNetV2 [114], а затем усовершенствованный в EfficientNet за счёт добавления механизма Squeeze-and-Excitation [115]. MBConv-блок реализует инвертированную бутылочную структуру, направленную на сохранение информативных признаков при снижении количества операций и параметров.

Структура блока включает последовательные компоненты:

Во-первых, выполняется расширение (англ. expansion) – операция 1×1 -свёртки, увеличивающая число каналов входного тензора C_{in} до $C_{\text{exp}} = t \cdot C_{\text{in}}$, где t — коэффициент расширения. Входной тензор $X \in \mathbb{R}^{H \times W \times C}$ преобразуется следующим образом:

$$X_{\text{exp}} = \text{Conv}_{1 \times 1}(X), \quad \dim(X_{\text{exp}}) = H \times W \times tC. \quad (2.5)$$

Затем применяется глубинная свёртка (англ. depthwise convolution), осуществляемая отдельно для каждого канала:

$$X_{\text{dw}} = \text{DWConv}_{k \times k}(X_{\text{exp}}). \quad (2.6)$$

Следующим этапом является механизм Squeeze-and-Excitation, реализующий адаптивную перекалибровку каналов. Сначала выполняется глобальное усреднение по пространству:

$$z_c = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W X_{\text{dw}}^{(i,j,c)}, \quad (2.7)$$

где z_c — усреднённая активация по c -му каналу. Далее через двухслойную полносвязную сеть формируется вектор:

$$s = \sigma(W_2 \cdot \text{ReLU}(W_1 z)), \quad (2.8)$$

и выполняется масштабирование каналов:

$$X_{\text{se}} = X_{\text{dw}} \cdot s. \quad (2.9)$$

Затем применяется проекция (англ. projection) с использованием 1×1 -свёртки, возвращающей тензор к исходному числу каналов C :

$$X_{\text{proj}} = \text{Conv}_{1 \times 1}(X_{\text{se}}), \quad \dim(X_{\text{proj}}) = H' \times W' \times C. \quad (2.10)$$

Наконец, если $s = 1$ и размерности совпадают, используется остаточная связь:

$$Y = X_{\text{proj}} + X. \quad (2.11)$$

Таким образом, MBConv-блок реализует эффективную глубинную свёртку, усиливающую выраженность признаков, с одновременным сохранением компактности архитектуры.

После всех MBConv-стадий выходной тензор имеет форму $7 \times 7 \times 320$. Далее выполняются следующие операции:

- 1×1 свёртка до 1280 каналов;
- Global average pooling: $7 \times 7 \rightarrow 1 \times 1$;
- Полносвязный слой (FC) для классификации.

Если входное изображение имеет $C_{\text{in}} = 3$ канала и размер 224×224 , то выход EfficientNet-B0 – это вектор длины 1000 (для задачи ImageNet), который вычисляется по формуле:

$$\hat{y} = \text{softmax}(W_{\text{FC}} \cdot \text{GAP}(X_{\text{final}}) + b) \quad (2.12)$$

где GAP – операция глобального усреднения, $W_{\text{FC}} \in \mathbb{R}^{1000 \times 1280}$ – матрица весов полносвязного слоя.

В качестве функции активации в EfficientNet используется SiLU – Sigmoid Linear Unit, определяемая следующим образом:

$$\text{SiLU}(x) = x \cdot \sigma(x) \quad (2.13)$$

где $\sigma(x)$ – сигмоида. Эта функция показала лучшие результаты по сравнению с ReLU в глубоких архитектурах [116]. Для нормализации после каждой свёртки в EfficientNet применяется Batch Normalization [117].

2.5.2 Обучение и валидация, результаты

Обучение модели EfficientNetB0 на датасете OCTDL выполнено с использованием отложенной выборки. Для оценки качества применены метрики F1-score и Recall. Модель достигла устойчиво высоких результатов (выше 95 %), что подтверждает её применимость для автоматического анализа ОКТ-снимков. Результаты классификации по классам представлены в табл. 6.

2.6 Метод объяснительного ИИ

В данной работе для получения визуального объяснения этапа предварительной классификации ОКТ-снимков на основе нейросетевой архитектуры EfficientNetB0 применяется метод Class Activation Mapping (CAM). Выбор CAM

обусловлен тем, что архитектура EfficientNet содержит слой глобального усреднения по каналам (Global Average Pooling, GAP), за которым следует линейный классификатор, что соответствует необходимым условиям применения данного метода. САМ позволяет восстановить пространственное распределение признаков, задействованных в процессе классификации, и выделить области изображения, наиболее существенно влияющие на итоговое решение модели.

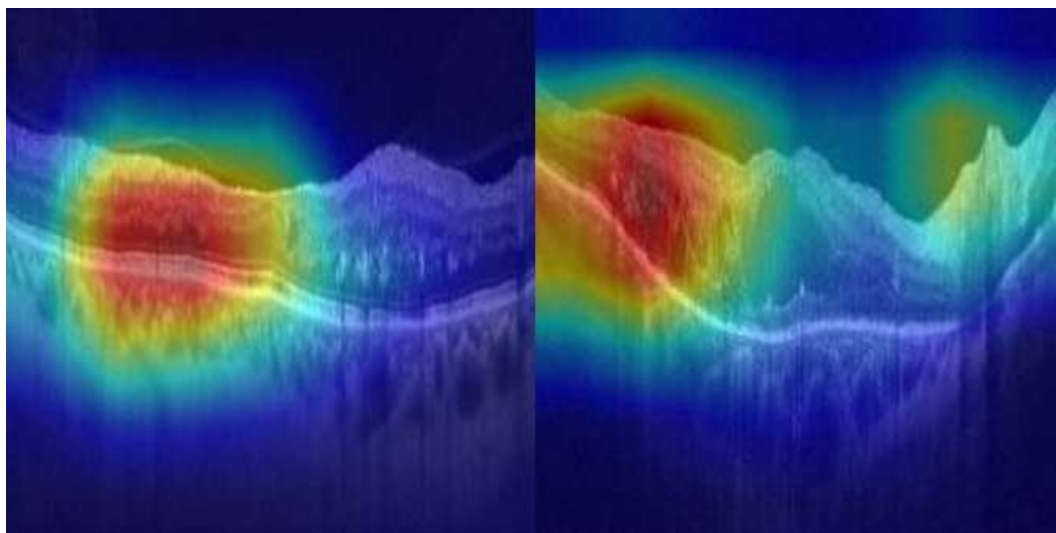


Рисунок 2.2 — Пример получения визуального объяснения методом САМ

2.7 Сегментация экземпляров биомаркеров ДМО

2.7.1 Модель YOLOv8

Модель YOLOv8 представляет собой современное развитие семейства одноэтапных нейросетевых архитектур для анализа изображений. Модель объединяет задачи детекции, сегментации и позовой оценки в единую модульную структуру с поддержкой масштабирования, anchor-free подхода и разделённого head-a, обеспечивая высокую точность при работе в реальном времени. Эволюция YOLO началась с YOLOv1 [118], продолжилась через YOLOv2/YOLO9000 [119], YOLOv3 [120], YOLOv4 [121] и YOLOv5 [122], и достигла современной реализации в YOLOv8 [123].

YOLOv8-seg состоит из трёх основных компонентов: блока извлечения признаков (backbone), блока агрегации признаков (neck) и блока выходных предсказаний (head). В качестве backbone используется архитектура C2f (Cross-Stage-Partial with fuse), пришедшая на смену CSP-блокам. Она обеспечивает улучшенный поток градиента и компактность модели за счёт конкатенации промежуточных выходов нескольких bottleneck-блоков. Верхняя часть backbone включает модуль быстрого пирамидального пула SPPF, позволяющий учитывать многошкальный контекст без потери производительности.

Блок агрегации признаков представлен PAN (Path Aggregation Network) [124], реализующим двунаправленную агрегацию: восходящий поток усиливает локальные признаки, а нисходящий — распространяет семантическую информацию, что особенно важно для точной сегментации мелких объектов.

Блок выходных предсказаний построен по принципу разделения задач (decoupled head): регрессия и классификация обрабатываются независимо, что повышает точность и устойчивость обучения. Для детекции используется anchor-free подход [125], избавляющий от необходимости подбора якорей и повышающий переносимость модели. Сегментация реализуется с помощью прототипной масочной ветви: на основе карты признаков генерируются прототипы, а маска восстанавливается линейной комбинацией с учётом коэффициентов для каждого сегмента [126]. Общая схема архитектуры модели YOLOv8 представлена на рис. 2.3.

Модель масштабируется в пяти вариантах (n, s, m, l, x), позволяя выбрать баланс между точностью и скоростью в зависимости от вычислительных ресурсов [123]. Характеристики модельного ряда приведены в таблице 7.

В процессе обучения используется оптимизатор AdamW, косинусное или одноцикловое планирование скорости обучения, ЕМА (экспоненциальное сглаживание весов) и набор аугментаций (Mosaic, MixUp и др.). Инференс оптимизирован за счёт поддержки FP16, ONNX/TensorRT и class-agnostic NMS.

Функция потерь представляет собой сумму трёх компонент:

$$\mathcal{L} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{mask}} L_{\text{mask}}. \quad (2.14)$$

где L_{cls} — кросс-энтропия по классам:

$$L_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C [y_{i,c} \log \hat{p}_{i,c} + (1 - y_{i,c}) \log(1 - \hat{p}_{i,c})]. \quad (2.15)$$

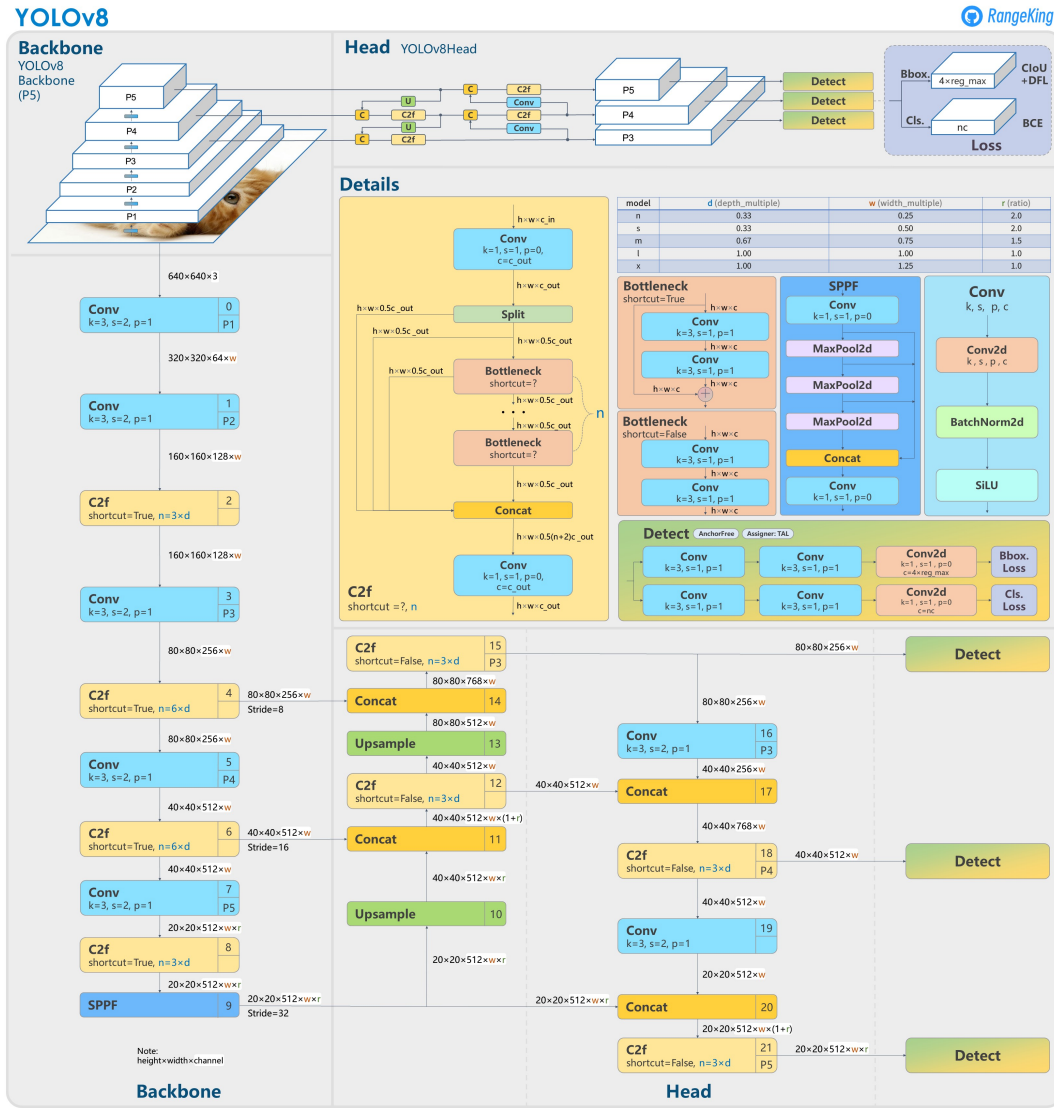


Рисунок 2.3 — Архитектура модели YOLOv8

Компонент локализации рассчитывается через IoU или его обобщения:

$$L_{\text{box}} = \frac{1}{N} \sum_{i=1}^N \left(1 - \text{IoU}(B_i, \hat{B}_i) \right). \quad (2.16)$$

Компонент маски формулируется как бинарная кросс-энтропия:

$$L_{\text{mask}} = -\frac{1}{M} \sum_{j=1}^M [m_j \log \hat{m}_j + (1 - m_j) \log(1 - \hat{m}_j)]. \quad (2.17)$$

Таким образом, YOLOv8-seg представляет собой компактную, точную и масштабируемую архитектуру, пригодную для задач анализа изображений в реальном времени.

2.7.2 Модель BiFPN

Модель BiFPN (англ. Bidirectional Feature Pyramid Network) предназначена для эффективной агрегации признаков из различных уровней иерархии пространства признаков в СНС. В отличие от классических FPN [127] и PAN [128], где направление передачи информации строго определено (вверх или вниз), архитектура BiFPN обеспечивает двунаправленную агрегацию признаков с использованием взвешенного объединения, в котором веса являются обучаемыми параметрами (см. рис. 2.4). Это позволяет адаптивно определять вклад каждого уровня в итоговое представление и повышает точность при ограниченных вычислительных ресурсах [129].

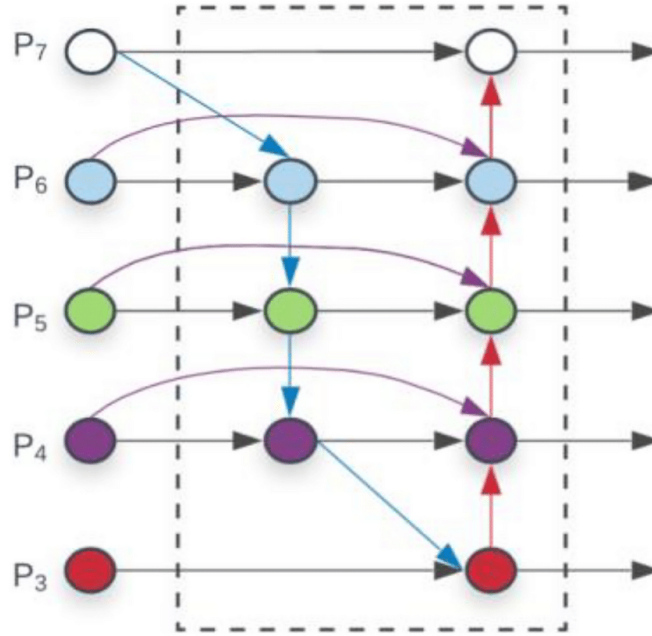


Рисунок 2.4 — Архитектура BiFPN

Пусть $\{X_i\}_{i=1}^m$ – входные карты признаков, приведённые к одному масштабу. BiFPN вычисляет нормализованные веса по формуле:

$$\tilde{w}_i = \frac{\max(0, w_i)}{\varepsilon + \sum_{j=1}^m \max(0, w_j)}, \quad (2.18)$$

и формирует результат объединения:

$$Y = \varphi \left(\sum_{i=1}^m \tilde{w}_i \cdot X_i \right), \quad (2.19)$$

где Φ – последовательность операций свёртки, нормализации и активации. Такая схема обеспечивает устойчивую комбинацию признаков, а веса автоматически подстраиваются под каждый уровень.

BiFPN упрощает структуру графа пирамиды, удаляя узлы с единственным входом и добавляя перекрёстные соединения между соседними уровнями. Каждый блок BiFPN повторяется несколько раз, реализуя итеративную обработку. Благодаря разделимым по каналам свёрткам, вычислительная нагрузка значительно снижается, что особенно важно для мобильных применений.

Архитектура легко масштабируется: количество каналов и повторов блоков увеличивается согласованно с масштабом модели [130], обеспечивая оптимальный баланс между точностью и вычислительными затратами.

BiFPN обладает рядом архитектурных преимуществ, обеспечивающих её эффективность и универсальность: обучаемое взвешенное объединение признаков позволяет заменить фиксированную сумму адаптивной, регулируемой на основе данных; двунаправленная агрегация информации между уровнями повышает согласованность семантических и пространственных признаков; повторяющиеся блоки агрегации упрощают реализацию и масштабирование; архитектура легко интегрируется в модели различной сложности без необходимости её модификации; ограничение весов обеспечивает устойчивое поведение во время обучения.

2.7.3 Механизм внимания Coordinate Attention

Механизм Coordinate Attention (CA) предназначен для повышения качества извлечения признаков в сверточных нейросетях путём одновременного моделирования важности каналов и пространственного положения объектов. В отличие от SE-блока, где глобальное усреднение по пространству полностью теряет координатную информацию [131], CA сохраняет её, разлагая агрегацию на две ортогональные оси, что особенно важно для обнаружения мелких или вытянутых объектов, где ключевую роль играют ориентация и локализация [132]. Графическое представление механизма CA представлена на рис. 2.5.

Пусть входная тензорная карта признаков обозначена как $X \in \mathbb{R}^{C \times H \times W}$, где C — число каналов, H и W — высота и ширина карты.

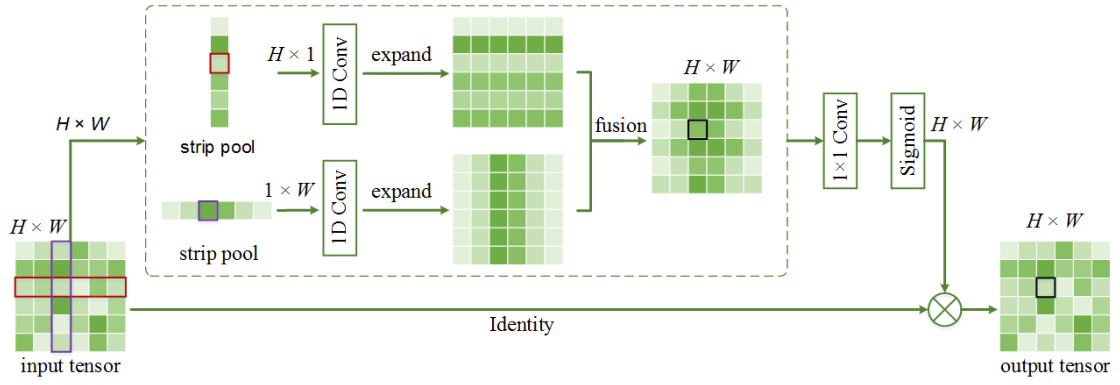


Рисунок 2.5 — Структура механизма внимания Coordinate Attention

Сначала выполняется раздельное усреднение по ширине и высоте:

$$p_h(c, y) = \frac{1}{W} \sum_{x=1}^W X(c, y, x), \quad p_w(c, x) = \frac{1}{H} \sum_{y=1}^H X(c, y, x), \quad (2.20)$$

где $p_h \in \mathbb{R}^{C \times H \times 1}$, а $p_w \in \mathbb{R}^{C \times 1 \times W}$ – эти признаки объединяются, проходят через еденичную свёртку с уменьшением размерности в $r = 16$ раз, затем через нормализацию и нелинейность. После этого применяются две независимые еденичные свёртки и сигмоида:

$$a_h = \sigma(\text{Conv}_{1 \times 1}^{(h)}(f_h)), \quad a_w = \sigma(\text{Conv}_{1 \times 1}^{(w)}(f_w)). \quad (2.21)$$

Затем исходные признаки масштабируются с помощью полученных весов:

$$Y(c, y, x) = X(c, y, x) \cdot a_h(c, y) \cdot a_w(c, x). \quad (2.22)$$

Такой подход позволяет учитывать пространственную ориентацию и глобальный контекст без существенного увеличения числа параметров. Coordinate Attention превосходит SE и CBAM [133] в задачах, где важно учитывать расположение и форму объектов. Он обеспечивает высокую точность при низкой вычислительной стоимости и легко интегрируется в существующие архитектуры, включая мобильные.

2.7.4 Модификация базовой модели

Модель-модификация YOLOv8-BiFPN-Coordinate Attention направлена на повышение точности сегментации мелких деталей за счёт усовершенст-

ния путей агрегации признаков и внедрения пространственно-чувствительного внимания в архитектуру YOLOv8-seg.

В рамках модификации стандартный блок агрегации признаков заменяется на BiFPN, позволяющую объединять признаки с разных уровней разрешения более гибко. В отличие от обычного суммирования, в BiFPN используется обучаемое взвешивание признаков карт, что позволяет модели самостоятельно определять, какие уровни более информативны для конкретного объекта. Благодаря двунаправленному распространению информации улучшается согласование признаков, особенно важных при обработке мелких или вытянутых объектов. Это положительно сказывается на точности локализации и сегментации тонких контуров, которые часто теряются при использовании стандартных иерархических структур.

Дополнительное улучшение достигается за счёт внедрения механизма Coordinate Attention, встроенного как в блоки извлечения признаков, так и между уровнями агрегации. В отличие от традиционных модулей внимания, Coordinate Attention сохраняет пространственную структуру признаков, позволяя модели акцентировать внимание не только на важнейших каналах, но и на их положении в изображении. Это особенно важно при работе с мелкими или вытянутыми структурами, такими как сосуды, трещины, контуры объектов и другие элементы, представленные малым числом пикселей.

Интеграция BiFPN и Coordinate Attention даёт комплексное улучшение: BiFPN усиливает многоуровневое представление признаков, а Coordinate Attention позволяет направленно фильтровать и усиливать пространственно важные регионы. Такая комбинация даёт прирост точности масочной сегментации без существенного увеличения вычислительной нагрузки. Особенно заметен эффект при работе с изображениями, насыщенными мелкими деталями – в задачах медицинской диагностики.

Таким образом, предложенная модификация архитектуры YOLOv8, дополненная BiFPN и Coordinate Attention, демонстрирует устойчивое улучшение сегментации за счёт усиления пространственного и многоуровневого внимания, что делает её особенно перспективной для задач, требующих точного выделения мелких объектов и структур.

2.7.5 Метрики оценки качества

Для оценки качества моделей в задачах сегментации применяются специальные метрики, позволяющие количественно определить степень совпадения предсказанной и истинной масок классов. В данной работе рассматриваются две метрики: Intersection over Union (IoU) и коэффициент сходства Dice (DSC). Обе метрики применимы как в бинарной, так и в многоклассовой сегментации.

Метрика Intersection over Union, также известная как индекс Жаккара, определяется как отношение площади пересечения предсказанной и истинной областей к площади их объединения:

$$IoU = \frac{area_{intersection}}{area_{union}}, \quad (2.23)$$

где $area_{intersection}$ – площадь пересечения областей, а $area_{union}$ – площадь их объединения.

В многоклассовой сегментации IoU рассчитывается отдельно для каждого класса k :

$$IoU_k = \frac{TP_k}{TP_k + FP_k + FN_k}, \quad (2.24)$$

где:

- TP_k – число пикселей, правильно отнесённых к классу k ;
- FP_k – число пикселей, ошибочно предсказанных как класс k ;
- FN_k – число пикселей класса k , не распознанных моделью.

Отмечается, что метрика IoU может быть чувствительна к ошибкам на границах объектов, особенно при сегментации крупных структур [134].

Коэффициент Dice используется для измерения степени схожести между двумя множествами пикселей и определяется следующим образом:

$$DSC = \frac{2TP}{2TP + FP + FN}. \quad (2.25)$$

В многоклассовой сегментации коэффициент Dice также вычисляется отдельно для каждого класса k :

$$DSC_k = \frac{2TP_k}{2TP_k + FP_k + FN_k}, \quad (2.26)$$

где TP_k , FP_k , FN_k имеют то же значение, что и в формуле для IoU_k .

Коэффициент Dice близок к единице при хорошем совпадении предсказанной и истинной масок. Однако при перекрытии экземпляров одного класса возможны переоценки точности сегментации [135].

Для более детальной оценки качества модели сегментации, особенно в задачах с несколькими классами, целесообразно рассчитывать метрики отдельно для каждого класса. Такой подход позволяет выявить, насколько эффективно модель обнаруживает объекты конкретного класса, а также отличает их от остальных.

В задачах сегментации оценка ведётся на уровне пикселей – каждый пиксель изображения рассматривается как отдельный пример, принадлежащий одному из C классов. Для класса c определяются следующие значения:

- TP_c – количество пикселей, правильно предсказанных как класс c ;
- FP_c – количество пикселей, ошибочно предсказанных как класс c ;
- FN_c – количество пикселей истинного класса c , нераспознанных моделью;
- TN_c – количество пикселей, правильно предсказанных как не принадлежащие классу c .

На основе этих величин определяются метрики чувствительности и специфичности для каждого класса:

$$Sensitivity_c = \frac{TP_c}{TP_c + FN_c}, \quad (2.27)$$

$$Specificity_c = \frac{TN_c}{TN_c + FP_c}. \quad (2.28)$$

Чувствительность (англ. sensitivity) отражает долю пикселей класса c , правильно распознанных моделью. Она показывает, насколько эффективно модель обнаруживает все области заданного класса. Высокое значение чувствительности особенно важно в медицинских задачах, где пропуск патологических областей может привести к клинически значимым ошибкам.

Специфичность (англ. specificity) характеризует способность модели правильно исключать пиксели, не принадлежащие классу c , и минимизировать количество ложноположительных предсказаний. Это важно для снижения числа ложных срабатываний и повышения доверия к модели при практическом применении.

Таким образом, использование чувствительности и специфичности, рассчитанных по каждому классу, позволяет получить более полную картину производительности модели сегментации, особенно в условиях несбалансированных классов и медицински значимых объектов.

2.7.6 Обучение, валидация, результаты

Для решения задачи сегментации экземпляров биомаркеров ДМО на ОКТ-снимках использовались базовая модель YOLOv8-seg и её модифицированная версия YOLOv8l-BiFPN-Coordinate Attention. Обе модели дообучались на специализированном датасете, включающем вручную размеченные экземпляры восьми классов биомаркеров: DRIL, ERM, EZD, HE, HF, IRC, IRF, SRF, VMT. Разбиение коллекции данных (900 изображений): 70 % – обучение, 15 % – валидация, 15 % – тест.

Модифицированная модель обеспечила устойчивое улучшение качества сегментации, особенно на мелких биомаркерах. В табл. 8 приведены значения метрик IoU, Sensitivity и Specificity для каждой категории. На рис. 2.6 представлен пример сегментации экземпляров биомаркеров ДМО на ОКТ-снимках.

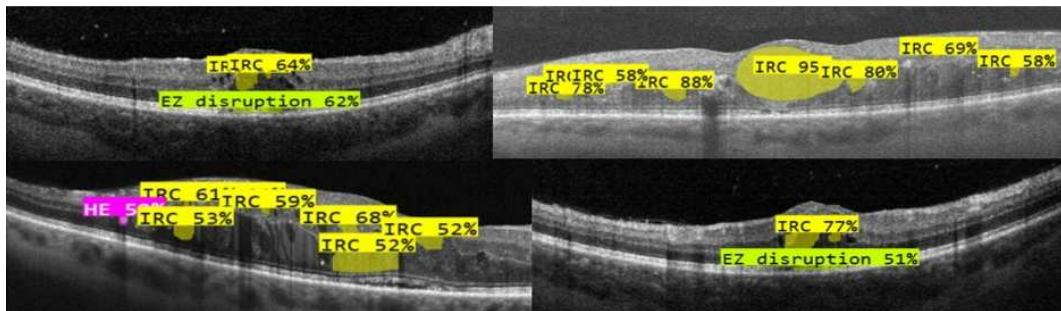


Рисунок 2.6 — Пример сегментации изображений с помощью YOLOv8-BiFPN-CA

Наибольший прирост наблюдается у классов с наименьшими объектами (HF, HE, DRIL, EZD), что подчёркивает эффективность модифицированной архитектуры при сегментации мелких деталей. Несмотря на небольшое увеличение времени инференса по сравнению с базовой моделью, вычислительные затраты остаются приемлемыми для большинства клинических сценариев.

Модель сохраняет высокую скорость и пригодна для интеграции в рабочие процессы анализа ОКТ-снимков.

2.8 Определение количественных характеристик экземпляров биомаркеров ДМО

Рассматривается задача количественной оценки площади экземпляров биомаркеров ДМО на основе бинарных масок, полученных после сегментации экземпляров нейросетевой моделью. Площадь каждого экземпляра может быть измерена либо в пикселях, либо в микрометрах в квадрате – в зависимости от того, известен ли масштаб изображения. При известном масштабе площадь рассчитывается как произведение количества пикселей в маске на площадь одного пикселя:

$$A = s_x s_y \sum M_{ij},$$

где s_x, s_y – физические размеры пикселя по горизонтали и вертикали в микрометрах, M_{ij} – бинарная маска экземпляра. Если изображение было масштабировано, масштаб пересчитывается с учётом коэффициентов изменения размеров. В случае, когда физический масштаб недоступен, площадь выражается в пикселях:

$$A_{\text{px}} = \sum M_{ij}.$$

При необходимости точности для мелких объектов возможна постобработка с использованием субпиксельной аппроксимации границ. Полученные значения могут агрегироваться по классам и использоваться для дальнейшего анализа в абсолютных или относительных единицах.

2.9 Классификация на основе нечётких правил

2.9.1 Нечёткий классификатор и лингвистические правила

Целью является разработка методов классификации патологий на основе диагностических критериев, формализованных в виде нечётких лингвистических правил. Эффективность нечёткой классификации в задачах медицинской диагностики определяется как качеством описания объектов (например, изображений или клинических данных), так и степенью формализации экспертных знаний. В настоящем разделе, опираясь на подход, представленный в [136], рассматривается постановка задачи нечёткой классификации, ориентированной на диагностику заболеваний с применением анализа визуальной информации, основанной на экспертных знаниях, представленных в виде лингвистических правил. Такой подход позволяет учитывать неопределённость и вариативность признаков, характерных для клинической практики.

Пусть объект, подлежащий классификации (изображение), обозначается как A и представляет собой конечный набор признаков $A = \{A_1, \dots, A_n\}$, извлечённых с его помощью с использованием искусственной нейронной сети. Каждый признак A_i ассоциирован с множеством U_i описательных характеристик (в данном случае – количественных значений), а также с множеством $\{a_{i1}, \dots, a_{in_i}\}$ ($1 \leq n_i$) лингвистических значений. Каждое лингвистическое значение a_{ij} связано с функцией принадлежности $\mu_{a_{ij}}(u_i)$, определённой на универсальном множестве U_i .

Нечёткий классификатор – это алгоритм, предназначенный для отнесения объекта O , описываемого совокупностью признаков $\mathbf{u} = \{u_1, \dots, u_n\}$, к одному или нескольким классам $C_k \in \{C_1, \dots, C_K\}$ на основе степени принадлежности, а не жёсткого распределения. Пусть каждому признаку u_i соответствует множество лингвистических значений $\{a_{i1}, \dots, a_{in_i}\}$, функция принадлежности $\mu_{a_{ij}}(u_i) : U_i \rightarrow [0,1]$ отображает степень принадлежности признака значению a_{ij} . Нечёткие правила r_j имеют вид:

$$\text{если } A_1 = a_{1j}, \dots, A_n = a_{nj}, \text{ то } \mu_{r_j}(O) = \pi_j,$$

где $\pi_j \in [0, 1]$ – степень выполнения правила r_j для объекта O . Классификация объекта O осуществляется путём агрегирования выводов всех нечётких правил $r = 1, \dots, R$, имеющих степени срабатывания $\pi_r \in [0, 1]$, с использованием весов уверенности принадлежности к каждому классу C_k из (2.29). Для каждого класса $k \in \{1, \dots, K\}$ вычисляется агрегированная степень принадлежности:

$$S_k(O) = \sum_{r=1}^R \pi_r \eta_{r,k}, \quad (2.29)$$

где $\eta_{r,k} \in [0, 1]$ – степень принадлежности, задаваемая выводом r -го правила для класса C_k . Нормированная степень принадлежности объекта O классу C_k определяется следующим образом:

$$M_k^*(O) = \frac{S_k(O)}{\sum_{q=1}^K S_q(O)}, \quad k = 1, \dots, K, \quad (2.30)$$

в результате чего формируется выходной вектор принадлежностей:

$$M^*(O) = (M_1^*(O), M_2^*(O), \dots, M_K^*(O)), \quad \sum_{k=1}^K M_k^*(O) = 1.$$

При необходимости получения жёсткой классификации выбирается класс с максимальной нормированной степенью принадлежности:

$$\hat{k} = \arg \max_{k \in \{1, \dots, K\}} M_k^*(O). \quad (2.31)$$

Если некоторое правило r не формирует вывод по классу C_k , соответствующее значение $\eta_{r,k}$ полагается равным нулю. [137; 138].

Кроме того, существует K классов C_k , $k \in \{1, \dots, K\}$, которые в рамках данного исследования соответствуют стадиям диабетической ретинопатии. Информация о классах рассматривается как множество правил r , полученных на основе формализованных критериев диагностики заболевания (клинические рекомендации, публикации и т. д.).

Согласно постановке задачи, множество r лингвистических правил представлено в следующем виде:

$$\left\{ \begin{array}{l} \text{если } A_1 = a_{i1}^1, A_2 = a_{i2}^1, \dots, A_n = a_{in}^1, \\ \text{то } \mu_{C_1}(O) = \eta_1^1, \dots, \mu_{C_k}(O) = \eta_k^1, \\ \vdots \\ \text{если } A_1 = a_{i1}^r, A_2 = a_{i2}^r, \dots, A_n = a_{in}^r, \\ \text{то } \mu_{C_1}(O) = \eta_1^r, \dots, \mu_{C_k}(O) = \eta_k^r, \end{array} \right. \quad (2.32)$$

где объект O имеет значения признаков, указанных в левой части правил, η – степень уверенности в том, что объект O принадлежит определённому классу в соответствии с r -м правилом, а μ – функция принадлежности, определяющая степень принадлежности объекта O к соответствующему классу.

Таким образом, результирующая система является нечётким классификатором, основанным на множестве r правил, сформированных на основе формализованных критериев диагностики заболевания.

2.9.2 Формирование корпуса лингвистических правил

В данном разделе представлены нечёткие лингвистические правила для оценки биомаркеров поражения сетчатки на основании оптической когерентной томографии (ОКТ), сформированные на основе диагностических и прогностических критериев, изложенных в Клинических рекомендациях Министерства здравоохранения РФ [139], Национальном руководстве по офтальмологии [140], а также иных руководствах по диагностике данного заболевания. Каждое правило связывает значения биомаркеров с клинически значимыми выводами, используя лингвистические переменные и термы.

1. Дезорганизация внутренних слоёв сетчатки (DRIL). Переменная: протяжённость DRIL; лингвистические термы: малая (trap), умеренная (tri), большая (trap); область определения: [0 мм, 2 мм]. Переменная: прогноз; лингвистические термы: хороший (trap), умеренный (tri), плохой (trap); область определения: [0, 1]. Правило: ЕСЛИ протяжённость DRIL большая, ТО прогноз остроты зрения плохой.

2. Нарушение целостности эллипсоидной зоны (EZD). Переменная: целостность EZ; лингвистические термы: разрушена (trap), частично (tri), интактна

(trap); область определения: $[0, 1]$. Переменная: прогноз; лингвистические термы: умеренное (tri), необратимое (trap), сохранённое (trap); область определения: $[0, 1]$. Правило: ЕСЛИ эллипсоидная зона разрушена, ТО снижение зрения необратимо.

3. Твёрдые эксудаты (HE). Переменная: диаметр включений; лингвистические термы: малый (trap), пограничный (tri), большой (trap); область определения: $[0, 60 \text{ мкм}]$. Переменная: тень; лингвистические термы: отсутствует (trap), умеренная (tri), выраженная (trap); область определения: $[0, 1]$. Переменная: прогноз; лингвистические термы: стабильный (trap), ухудшение (trap); область определения: $[0, 1]$. Правило: ЕСЛИ диаметр большой И тень выраженная, ТО прогноз – ухудшение.

4. Гиперрефлективные фокусы (HF). Переменная: количество ГРФ; лингвистические термы: мало (trap), умеренно (tri), много (trap); область определения: $[0, 50]$. Переменная: воспаление; лингвистические термы: слабое (trap), выраженное (trap); область определения: $[0, 1]$. Правило: ЕСЛИ количество ГРФ много, ТО воспаление выраженное, отклик на стероиды хороший.

5. Интравитреальные кисты / интравитреальная жидкость (IRC / IRF). Переменная: размер кисты; лингвистические термы: мелкая (trap), средняя (tri), гигантская (trap); область определения: $[0, 300 \text{ мкм}]$. Переменная: наличие перемычек; лингвистические термы: нет (trap), частично (tri), есть (trap); область определения: $[0, 1]$. Переменная: прогноз; лингвистические термы: хороший (trap), умеренный (tri), плохой (trap); область определения: $[0, 1]$. Правила: ЕСЛИ размер гигантская, ТО прогноз плохой. ЕСЛИ перемычки есть, ТО ответ на anti-VEGF хороший.

6. Субретинальная жидкость (SRF). Переменная: длительность SRF; лингвистические термы: острая (trap), подострая (tri), хроническая (trap); область определения: $[0, 12 \text{ мес}]$. Переменная: прогноз; лингвистические термы: положительный (trap), ухудшение (trap); область определения: $[0, 1]$. Правила: ЕСЛИ SRF острая, ТО отклик на терапию положительный. ЕСЛИ SRF хроническая, ТО прогноз – ухудшение.

7. Витреомакулярное натяжение (VMT). Переменная: сила тракции; лингвистические термы: нет (trap), умеренная (tri), выраженная (trap); область определения: $[0, 100]$. Переменная: прогноз; лингвистические термы: стабилен (trap), ДМО (trap), улучшение (trap); область определения: $[0, 1]$. Правила: ЕС-

ЛИ тракция выраженная, ТО формируется ДМО. ЕСЛИ тракция устранена, ТО прогноз – улучшение.

Таким образом, сформированы лингвистические переменные и правила, отражающие экспертные знания в области диагностики диабетической ретинопатии и макулярного отёка на основе данных ОКТ. Эти правила могут быть использованы в системах нечёткого вывода и принятия решений в клинической практике.

2.10 Выводы ко второй главе

1. Проведён систематический обзор литературы, который показал преобладание архитектур семейства U-Net в задачах семантической сегментации. Выявлено отсутствие работ, рассматривающих сегментацию биомаркеров диабетического макулярного отёка (ДМО) в формате instance segmentation. Ограниченное внедрение методов объяснимого искусственного интеллекта подчёркивает актуальность разработки интерпретируемых решений в медицинских нейросетевых системах.
2. Сформирована специализированная коллекция ОКТ-снимков с полной ручной аннотацией восьми биомаркеров (DRIL, ERM, EZD, HE, HF, IRC, SRF, VMT) в формате сегментации экземпляров, что обеспечило возможность обучения и объективной оценки моделей, ориентированных на точное выделение отдельных патологических структур.
3. Разработана архитектура интеллектуального анализа ОКТ-изображений, включающая:
 - классификационную модель EfficientNetB0 с визуализацией решений методом Class Activation Mapping;
 - модифицированную модель YOLOv8, дополненную модулями BiFPN и Coordinate Attention, что позволило повысить точность сегментации, особенно для мелких объектов, при сохранении вычислительной эффективности.
4. На основе результатов сегментации разработан интерпретируемый нечёткий классификатор, основанный на формализованных лингвистических правилах. Это обеспечило переход от количественного описания

экземпляров к диагностически значимой интерпретации, соответствующей экспертной клинической логике.

5. Впервые формализована и решена задача анализа биомаркеров диабетического макулярного отёка на ОКТ-снимках как задача сегментации экземпляров, что позволило применить логические правила для оценки прогноза пациентов и интегрировать результаты в систему поддержки принятия врачебных решений.

Таблица 2 — Ключевые методы визуализации глаза и их диагностические ниши

Метод	Разрешение (акси- ал./лат., мкм)	Глубина, мм	Расширенные особенности
ОКТ (SD/SS)	3–6 / 10–15	2–3 (840 нм) 4–5 (1310 нм)	Золотой стандарт послойной томографии; количественные карты толщины RNFL и макулы; объёмное 3D-сканирование до 400 тыс А-сканов/с (SS-OCT) [81]
ОКТ-ангиография	3–5 / 10–15	2–3	Капиллярная перфузия без красителя; выявляет неоваскуляризацию и зоны неперфузии; сопоставима по чувствительности с FA при AMD и DR [82]
Ультразвуковая биомикроскопия (25–50 МГц)	20 / 40–50	4–5	Проходит мутные среды; незаменима для цилиарного тела и угла передней камеры; контактная методика
Флуоресцеиновая ангиография (FA)	– / 10–20	сосудистое русло	«Золотой стандарт» оценки утечки и неперфузии, но требует внутривенного красителя; риск анафилактики
Конфокальная SLO / автофлуоресценция	– / 10–15	0.5–1	Метаболическое картирование пигментного эпителия; микропериметрия; нет глубины, только 2D [83]
MP-томография (7T)	– / 300–500	>10	Функциональные и метаболические карты; высокая стоимость и низкая детализация для микроструктур [84]

Таблица 3 — Обзор исследований по теме сегментации биомаркеров ДМО

Исследование	Год	Тип задачи	Набор данных	Модель
Schlegl [93]	2018	Сем. сег.	Приватный (AMD/DME/RVO)	Residual U-Net
Ganjee [112]	2020	Сем. сег.	OPTIMA, UMN, Kermany (откр.)	Markov Random Field
De Silva [101]	2021	Сем. сег.	Приватный (168 глаз)	Mask R-CNN
Hsu [94]	2022	Сем. сег.	Приватный (3084 В-срезы)	U-Net
Tang [100]	2023	Сем. сег.	Приватный (468 глаз)	SegNet
Rahil [102]	2023	Сем. сег.	RETOUCH (откр.)	U-Net Ensemble
Ganjee [95]	2023	Сем. сег.	OPTIMA, UMN, Kermany (откр.)	U-Net
Melinščak [97]	2023	Сем. сег.	AROI (откр.)	Attention U-Net
Wang [103]	2023	Сем. сег.	AROI (откр.)	D3T-FCN
Daanouni [98]	2024	Сем. сег.	AROI (откр.)	U-Net++
George [96]	2024	Сем. сег.	Kermany, Lu (откр.)	U-Net
Qiu [104]	2023	Сем. сег.	AROI (откр.)	SAM
Fazekas [105]	2023	Сем. сег.	RETOUCH (откр.)	SAM, SAMe
Kulyabin [50]	2024	Сем. сег.	OIMHS, AROI (откр.)	SAM2, MedSAM
Hensman [99]	2025	Сем. сег.	Приватный (XLRS, 37 сканов)	U-Net

Таблица 4 — Охват биомаркеров в исследованиях

Исследование	DRIL	ERM	EZD	HE	HF	IRC	IRF	MH	PED	SRF	Т
Schlegl [93]	—	—	—	—	✓	—	—	—	—	—	—
Ganjee [112]	—	—	—	—	—	✓	—	—	—	—	—
De Silva [101]	—	—	✓	—	—	—	—	—	—	—	—
Hsu [94]	—	—	✓	—	—	✓	—	—	—	✓	—
Tang [100]	—	✓	—	—	—	—	—	—	—	—	—
Rahil [102]	—	—	—	—	—	—	✓	—	✓	—	—
Ganjee [95]	—	—	—	—	—	✓	—	—	—	—	—
Melinščak [97]	—	—	—	—	—	—	✓	—	✓	—	—
Wang [103]	—	—	—	—	—	✓	—	✓	—	—	—
Daanouni [98]	—	—	—	—	—	—	✓	—	✓	—	—
George [96]	—	—	—	—	—	—	✓	—	—	—	—
Qiu [104]	—	—	—	—	—	—	✓	—	—	—	—
Fazekas [105]	—	—	—	—	—	—	✓	—	✓	—	—
Kulyabin [50]	—	—	—	—	—	✓	✓	✓	✓	—	—
Hensman [99]	—	—	—	—	—	✓	—	—	—	—	—
Наше исследование	✓	✓	✓	✓	✓	✓	✓	—	—	✓	—

Таблица 6 — Результаты классификации на датасете OCTDL

Класс	F1-score (%)	Recall (%)
AMD	97.2	98.1
DME	96.3	95.9
ERM	95.6	95.1
RAO	96.9	96.0
VID	95.4	95.2
Среднее	96.3	96.1

Таблица 7 — Сравнение версий YOLOv8-seg по количеству параметров и времени инференса

Модель	Параметры, М	GFLOPs	Инференс, ms	FPS
YOLOv8n-seg	3.2	8.7	6.2	~161
YOLOv8s-seg	11.2	28.8	8.9	~112
YOLOv8m-seg	25.9	78.9	12.7	~79
YOLOv8l-seg	43.7	165.2	19.5	~51
YOLOv8x-seg	68.2	257.8	28.2	~35

Таблица 8 — Метрики сегментации на YOLOv8 и YOLOv8-BiFPN-CA

Класс	IoU (base)	IoU (mod)	Sens. (base)	Sens. (mod)	Spec. (base)
DRIL	0.86 ± 0.021	0.91 ± 0.017	0.87 ± 0.020	0.92 ± 0.015	0.985 ± 0.014
ERM	0.88 ± 0.018	0.90 ± 0.015	0.89 ± 0.017	0.91 ± 0.014	0.987 ± 0.012
EZD	0.85 ± 0.024	0.91 ± 0.018	0.86 ± 0.023	0.92 ± 0.015	0.984 ± 0.015
HE	0.85 ± 0.026	0.90 ± 0.020	0.87 ± 0.024	0.91 ± 0.017	0.985 ± 0.014
HF	0.85 ± 0.025	0.91 ± 0.019	0.86 ± 0.023	0.92 ± 0.016	0.984 ± 0.016
IRC	0.89 ± 0.017	0.91 ± 0.014	0.90 ± 0.016	0.92 ± 0.013	0.988 ± 0.012
IRF	0.88 ± 0.018	0.90 ± 0.015	0.89 ± 0.017	0.91 ± 0.014	0.987 ± 0.012
SRF	0.90 ± 0.015	0.92 ± 0.013	0.91 ± 0.014	0.93 ± 0.011	0.989 ± 0.011
VMT	0.87 ± 0.019	0.89 ± 0.017	0.88 ± 0.018	0.90 ± 0.015	0.986 ± 0.013

Глава 3. Метод анализа фундус-снимков

3.1 Теоретические основы анализа фундус-снимков

3.1.1 Офтальмоскопия

Офтальмоскопия – это метод визуализации глазного дна, основанный на регистрации отражённого и рассеянного света от внутренних структур глаза: сетчатки, сосудистой оболочки и пигментного эпителия. В офтальмологической практике применяются прямая и непрямая офтальмоскопия, цифровые фундус-камеры и сканирующие лазерные офтальмоскопы. Независимо от метода, получение изображения включает: освещение сетчатки, оптическое формирование изображения в системе «глаз–камера», проекцию на сенсор и последующее преобразование сигнала в цифровую форму [141; 142].

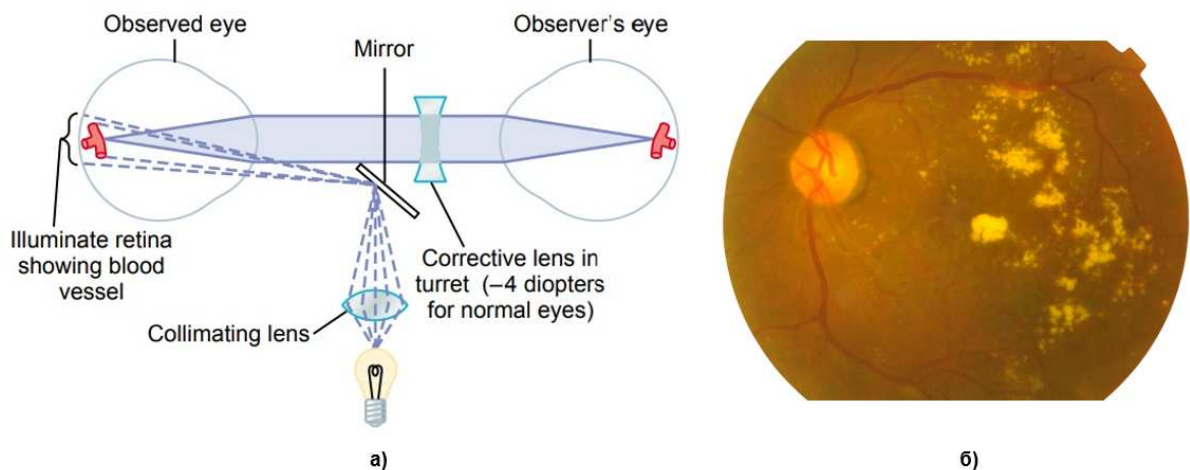


Рисунок 3.1 — а) Схема процесса офтальмоскопии; б) Фундус-снимок

Фундус-снимки используются для диагностики широкого круга офтальмологических заболеваний, включая диабетическую ретинопатию [143; 144], возрастную макулярную дегенерацию [145], глаукому [146], гипертоническую ретинопатию [147], сосудистые окклюзии, увеиты, новообразования хориоидеи [148] и другие патологии заднего отрезка глаза. Визуализируемыми на снимке структурами глаза являются: сосуды и капилляры, диск зрительного нерва, макула, пигментные и очаги дегенеративных процессов.

С физической точки зрения изображение на сенсоре формируется как результат отражательной способности тканей и оптической передачи системы. Геометрически фундус-снимок – центральная проекция сферической поверхности глазного дна, что вызывает дисторсии и виньетирование. Яркость пикселя аппроксимируется как

$$I(x, y) \approx L(x, y) \cdot R(x, y) + \eta(x, y), \quad (3.1)$$

где L — освещённость, R — отражательные свойства тканей, η — шумы и артефакты регистрации [142].

Фундус-снимок представляет собой цифровое изображение размерности $I \in \mathbb{R}^{H \times W \times 3}$, содержащее информацию о внутренней поверхности глаза. Перед его использованием в диагностических системах проводится коррекция освещённости, устранение дисторсий, цветовая нормализация и усиление контраста.

3.1.2 Диабетическая ретинопатия

Диабетическая ретинопатия (ДР) – это хроническое, прогрессирующее нейроваскулярное заболевание глазного дна, развивающееся как осложнение сахарного диабета (СД). Оно характеризуется нарушением проницаемости сосудистой стенки, окклюзией капилляров, ишемией и патологической неоваскуляризацией сетчатки. ДР является одной из ведущих причин слепоты в трудоспособном возрасте и требует системного наблюдения и диагностики [85].

По данным эпидемиологических исследований, частота ретинопатии у больных СД 1-го типа достигает 27,2 %, при СД 2-го типа – около 13 % [149]. Основные модифицируемые факторы риска – хроническая гипергликемия и артериальная гипертензия, способствующие микрососудистому повреждению [150]. В связи с ростом распространённости СД, потребность в скрининге ДР и автоматизированном анализе фундус-изображений возрастает. Современные алгоритмы анализа медицинских изображений требуют формализованного описания биомаркеров, отражающих стадию и активность заболевания [151].

Клиническая классификация ДР включает следующие стадии: непролиферативная (NPDR), пролиферативная (PDR), а также выделяется диабетический макулярный отёк (DME) как отдельная клинически значимая форма.

NPDR подразделяется на лёгкую, умеренную и тяжёлую. Переход от NPDR к PDR обусловлен прогрессирующей ишемией, появлением интратретиальных микрососудистых аномалий (IRMA), венозных изменений и неоваскуляризации [85; 152].

К числу основных биомаркеров ДР, визуализируемых при офтальмоскопии и на фундус-снимках, относятся:

- Микроаневризмы (МА) – первые морфологически наблюдаемые признаки ДР, представляют собой мелкие точечные расширения капилляров. Выглядят как округлые тёмно-красные точки, преимущественно в парамакулярной зоне [153].
- Внутрисетчаточные кровоизлияния (НЕМ) – различаются по форме: точечные/пятнистые (внутренние слои сетчатки) и пламенивидные (вдоль нервных волокон). Увеличение их числа отражает нарастание тяжести NPDR и риск перехода к более тяжёлой стадии [85].
- Твёрдые экссудаты (EX) – представляют собой липидные отложения жёлтого цвета с чёткими границами, формируются в результате хронической экссудации. Связаны с наличием макулярного отёка [153].
- Мягкие экссудаты (SE, «ватные очаги») – белёсые пятна в слое нервных волокон, соответствуют микроинфарктам сетчатки, свидетельствуют о выраженной ишемии [150].
- Интратретиальные микрососудистые аномалии (IRMA) и венозные изменения – извитость, «бисеринчатость» – указывают на тяжёлую непролиферативную стадию и риск развития PDR [152].
- Неоваскуляризация (NV) – рост новых сосудов на поверхности сетчатки и диске зрительного нерва, основной диагностический критерий перехода в PDR. Эти сосуды хрупкие и склонны к кровоизлияниям, приводящим к ухудшению зрения [85].
- Макулярный отёк (DME) – утолщение центральной зоны сетчатки с отложениями экссудатов. Является основной причиной снижения центрального зрения при любом уровне ДР [154].

На рисунке ниже представлена типовая визуализация ключевых биомаркеров диабетической ретинопатии на цветном фундус-снимке. Обозначены наиболее характерные морфологические элементы, используемые в офтальмоскопии и автоматизированной интерпретации изображений.

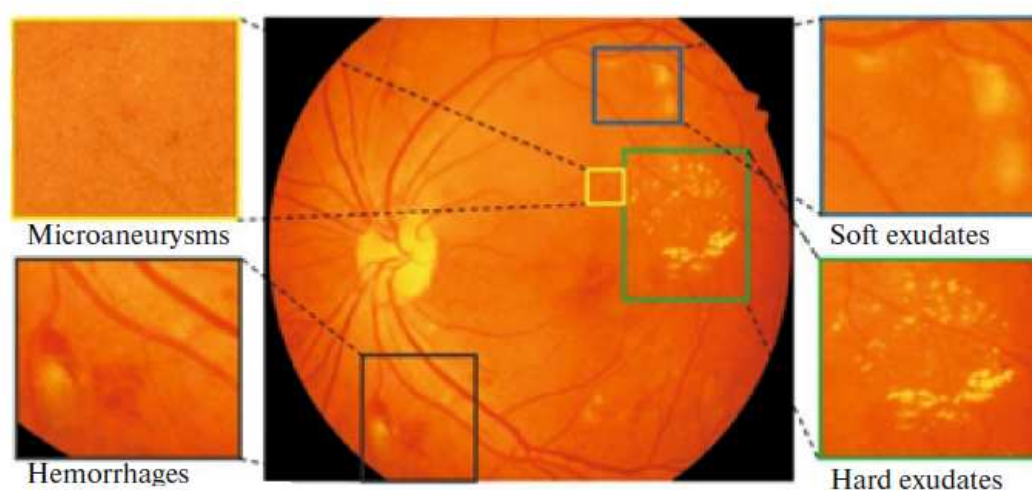


Рисунок 3.2 — Визуализация биомаркеров диабетической ретинопатии на фундус-снимке: микроаневризмы, кровоизлияния, эксудаты

Выявление и количественная оценка указанных биомаркеров лежит в основе классификации и выбора тактики ведения пациента. Например, наличие только микроаневризм свидетельствует о лёгкой NPDR и требует наблюдения, тогда как сочетание IRMA, венозной бисеринчатости и множественных кровоизлияний – о тяжёлой форме, требующей активного вмешательства. Неоваскуляризация – абсолютное показание для начала панретиальной лазерной коагуляции или терапии ингибиторами VEGF. DME, как самостоятельный фенотип, лечится преимущественно интравитреальными инъекциями анти-VEGF препаратов [85].

Таким образом, систематизированное описание биомаркеров диабетической ретинопатии на фундус-снимках является ключевым условием для разработки и внедрения автоматизированных методов скрининга. Учитывая клиническую значимость каждого из признаков и доступность цифровой визуализации, фундус-фотография остаётся стандартом скрининга, а алгоритмический анализ изображений на её основе – перспективным направлением для раннего выявления и профилактики слепоты при диабете.

3.1.3 Обзор литературы и источников

Сегментация биомаркеров ДР на фундус-снимках является ключевым компонентом автоматизированных СППВР. Основными объектами сегментации

служат микроаневризмы (МА), кровоизлияния (НЕ), твёрдые (ЕХ) и мягкие (СЕ) экссудаты, а также неоваскуляризация (NV) и преретинальные кровоизлияния (PRH). В данной области активно применяются нейросетевые методы, в первую очередь архитектуры на основе U-Net. Задачи классифицируются как семантическая сегментация или сегментация экземпляров в зависимости от способа представления итоговой маски.

Для формирования аналитической базы был проведён систематический поиск и отбор публикаций по протоколу PRISMA-2020. Поисковый запрос охватывал ключевые термины: diabetic retinopathy, segmentation, U-Net, instance segmentation, Mask R-CNN, IDRiD, FGADR. Источники: PubMed, Scopus, IEEE Xplore, ScienceDirect, SpringerLink (2019–2025). Включались только рецензируемые журнальные статьи и материалы конференций с описанием нейросетевых моделей, доступных датасетов и метрик. Исключались препринты, обзоры, работы рассматривающие задачи классификации или детекции. В результате было отобрано 31 исследование, удовлетворяющее критериям включения.

Ниже представлена итоговая таблица (табл. 9), содержащая систематизированные сведения о годе публикации, типе задачи, сегментируемых биомаркерах, используемых архитектурах, датасетах и применении методов объяснительного искусственного интеллекта.

Анализ отобранных 31 работы позволяет выделить ряд устойчивых технических и методологических тенденций. Во-первых, тип решаемой задачи в подавляющем большинстве случаев (28 из 31 работ) представлен в виде задачи семантической сегментации. Исключение составляет работа [155], где реализована сегментация экземпляров с использованием Mask R-CNN для подсчёта микроаневризм.

Во-вторых, среди сегментируемых биомаркеров наибольшее внимание уделяется стандартной четвёрке: МА, НЕ, ЕХ, СЕ. Полный набор этих маркеров используется более чем в 75% работ. Узкоспециализированные модели предложены для отдельных поражений – МА [156–158], НЕ [159; 160], ЕХ [161; 162], NV и PRH [163].

По частоте используемости лидирует колекция данных IDRiD, применённая в 26 работах. Она часто комбинируется с DDR и FGADR. Отдельные исследования используют собственные приватные наборы изображений [163].

В части архитектур нейросетей преобладают модели на базе U-Net и его модификаций, включая Attention U-Net, Dual-Branch U-Net, CBAM U-Net,

Таблица 9 — Обзор исследований по теме сегментации биомаркеров ДР

№	Автор(ы)	Год	Тип задачи	Биомаркеры	Датасет	Архитектура
1	Guo et al.	2019	Сем. сег.	MA, HE, EX, SE	IDRiD	L-Seg
2	Xu et al.	2021	Сем. сег.	MA, HE, EX, SE	IDRiD	FFU-Net
3	Wan et al.	2021	Сем. сег.	MA, HE, EX, SE	IDRiD	EAD-Net
4	Garifullin et al.	2021	Сем. сег.	MA, HE, EX, SE	IDRiD	Байесовская CNN
5	He et al.	2022	Сем. сег.	MA, HE, EX, SE	IDRiD, DDR	PMCNet
6	Huang et al.	2022	Сем. сег.	MA, HE, EX, SE	IDRiD, DDR	RTNet
7	Guo & Peng	2022	Сем. сег.	MA, HE, EX, SE	IDRiD, DDR	CARNet
8	Yan et al.	2022	Сем. сег.	MA, HE, EX, SE	IDRiD, FGADR	MSLF-Net
9	Skouta et al.	2022	Сем. сег.	HE	IDRiD, DIARETDB1	Модифицированный U-Net
10	Yin et al.	2023	Сем. сег.	MA, HE, EX, SE	IDRiD	Dual-Branch U-Net
11	Liu et al.	2023	Сем. сег.	MA, HE, EX, SE	IDRiD, FGADR	Many-to-many reassembly
12	Ullah et al.	2023	Сем. сег.	MA, HE, EX, SE	FGADR, IDRiD	SSMD-UNet
13	Zhao et al.	2023	Сем. сег.	MA, HE, EX, SE	Публичные	NAU-Net
14	Manan M. A. et al.	2023	Сем. сег.	EX	IDRiD	Residual encoder-decoder
15	Wang et al.	2023	Сем. сег.	MA, HE, EX, SE	IDRiD, DDR	ViT-Adapter + гиперб. эмбеддинги

UNet++. С 2022 года наметилась тенденция к интеграции модулей визуальных трансформеров (RTNet [164], ViT-Adapter [165], Prior-Guided Transformer [166]), а также к внедрению каскадных решений (CARNet). Активно развиваются облегчённые модели, ориентированные на мобильные приложения и массовый скрининг (BiSeNetV2 [167], MLNet [168], SegDRoWS [169]).

Важным аспектом анализа стало использование методов объяснительного искусственного интеллекта. Только в 4 работах (12%) указано явное применение методов ОИИ:

- [170] – байесовская неопределённость;
- [163; 164; 166] – карты внимания как механизм интерпретации.

Во всех остальных публикациях ОИИ не используется, что подчёркивает дефицит интерпретируемости в исследованиях по сегментации биомаркеров ДР.

Таким образом, на сегодняшний день основное внимание в литературе уделяется расширению архитектур, комбинированию датасетов, улучшению показателей метрик. Однако объяснимость, сегментация экземпляров остаются недостаточно охваченными направлениями. Отдельно следует отметить, что представленные исследования не реализуют полноценных диагностических пайплайнов и не позволяют проводить постановку диагноза в соответствии с клиническими рекомендациями и консенсусами. Это существенно ограничивает возможность их применения в клинической практике и подчёркивает необходимость разработки комплексных интеллектуальных систем, интегрирующих сегментацию с процессами грейдинга и принятия решений.

3.2 Описание предлагаемого метода

Предложен новый метод анализа специальных видов изображений, основанный на сочетании нейросетевой обработки и формализованных логических правил, отражающих экспертные знания. Метод направлен на извлечение интерпретируемых и воспроизводимых признаков из медицинских изображений и их последующую оценку с использованием объяснимых решений. Основной акцент сделан на переход от пиксельной семантической сегментации к анализу экземпляров объектов, что позволяет количественно оценивать выраженность

патологических признаков и привязывать их к анатомическим зонам. Метод продемонстрирован на задаче анализа фундус-снимков глаза.

Пошаговое описание метода:

1. Загрузка и предобработка изображений. На вход подаются цветные фундус-снимки. Предобработка включает выравнивание освещённости, нормализацию цветового пространства, масштабирование до стандартизированного разрешения. Эти операции создают унифицированное представление изображения, минимизируя влияние внешних факторов на качество последующей обработки.
2. Бинарная классификация. Нейросетевая модель определяет наличие или отсутствие признаков диабетической ретинопатии (ДР). При отрицательном результате пользователю выдаётся информационное сообщение. В случае положительного класса активируются шаги локализации и сегментации признаков.
3. Поиск зоны фовеа и сегментация макулы. Определяется координата фовеа и выделяется макулярная область как ключевая анатомическая зона. Это необходимо для локализации изменений, напрямую влияющих на центральное зрение, и позволяет пространственно нормировать признаки.
4. Разделение изображения на квадранты. Фундус-снимок разделяется на четыре анатомически значимых сектора (квадранта), что позволяет учитывать пространственное распределение биомаркеров, важное для клинической классификации стадий заболевания.
5. Семантическая сегментация и получение масок биомаркеров. С помощью нейросетевой модели выполняется многоклассовая сегментация фундус-снимка. В результате формируются пиксельные маски микроаневризм, кровоизлияний, твёрдых и мягких экссудатов, неоваскуляризации и других признаков. Эти маски служат основой для дальнейшего извлечения экземпляров объектов. Одновременно формируется визуальное объяснение работы модели (например, тепловые карты внимания), что обеспечивает проверяемость и интерпретируемость локализации выявленных признаков.
6. Сегментация экземпляров и подсчёт. Маски преобразуются в отдельные экземпляры патологических объектов с применением морфологических операций и метода водораздела. Для каждого экземпляра определяется

принадлежность конкретному квадранту. Выполняется подсчёт экземпляров каждого класса и агрегирование их характеристик.

7. Классификация стадии ДР по экспертным правилам. На основе количественных признаков и их пространственного распределения выполняется логическая классификация стадии заболевания. Классификатор использует набор формализованных экспертных правил: от жёстких условий (например, наличие неоваскуляризации) до пороговых сочетаний биомаркеров по числу и локализации.

На рис. 3.3 представлена схема предложенного метода анализа фундус-снимков.

Метод объединяет три ключевых компонента: сегментацию экземпляров для количественного анализа биомаркеров, логический вывод на основе формализованных экспертных правил и объяснимость, обеспечивающую интерпретируемость принятого решения. Такое сочетание делает предложенный подход не только технологически обоснованным, но и практически применимым в контексте клинической офтальмологии, где важны воспроизводимость, прозрачность и соответствие экспертной практике.

3.3 Коллекции данных

Разработка интеллектуальных систем анализа фундус-снимков требует опоры на верифицированные коллекции изображений, снабжённые экспертной аннотацией. Эти коллекции позволяют реализовывать обучение нейросетей, обеспечивать воспроизводимость экспериментов и объективную оценку результатов. В настоящем исследовании были использованы четыре наиболее репрезентативные публичные коллекции фундус-снимков: APTOS 2019 – для задач классификации, а MAPLES-DR, FGADR и IDRiD – для задач семантической сегментации.

Коллекция APTOS 2019 Blindness Detection была создана в рамках международного соревнования на платформе Kaggle и включает 3662 цветных изображения глазного дна [171]. Каждому изображению присвоен класс в соответствии с пятиуровневой шкалой тяжести диабетической ретинопатии (DR): от 0 (No DR) до 4 (Proliferative DR). Изображения были получены с различным

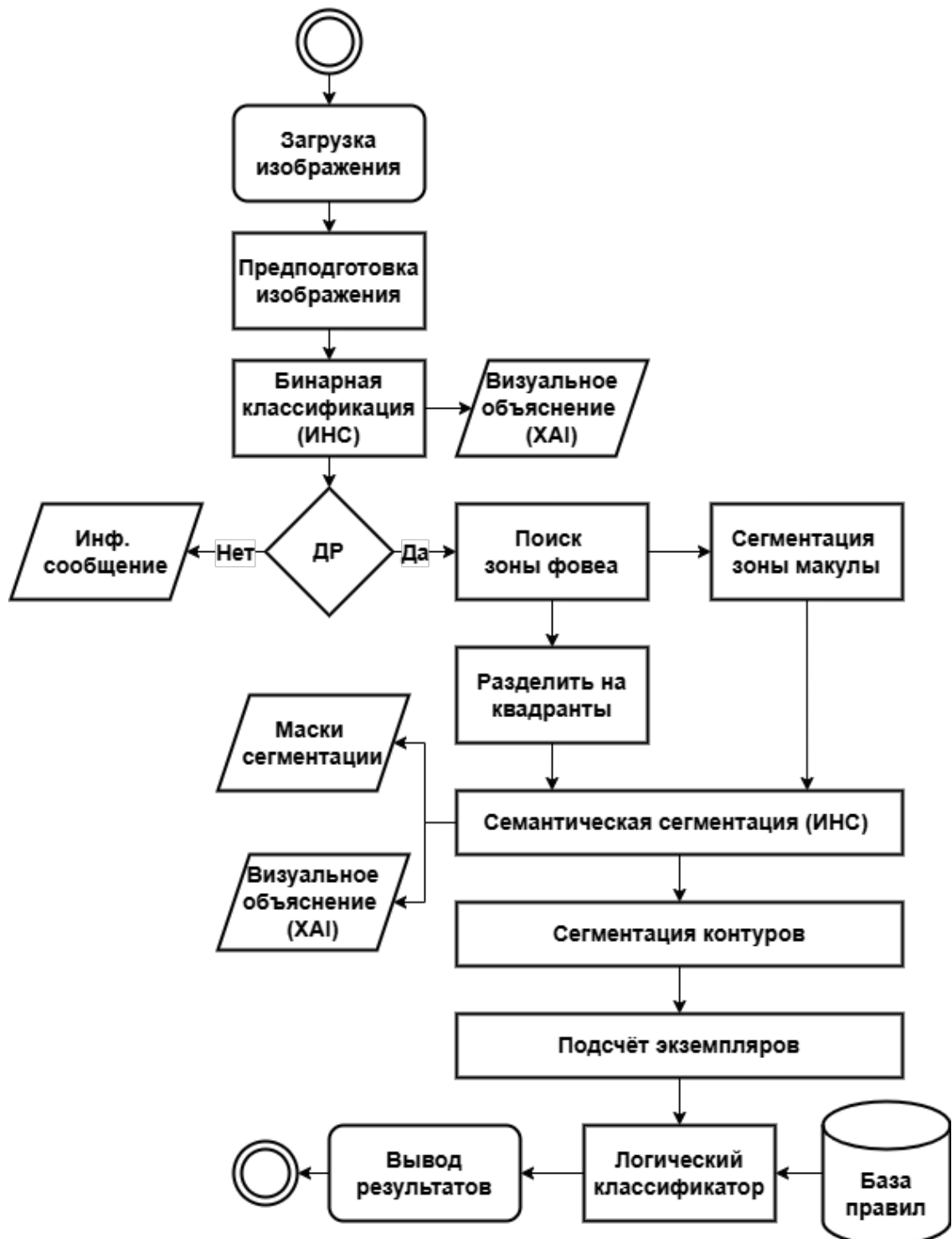


Рисунок 3.3 — Блок-схема предлагаемого метода анализа фундус-изображений

качеством, степенью освещённости и разрешением, что требует обязательной предобработки (усреднение, нормализация, выравнивание гистограммы и др.). Изначальное распределение классов демонстрирует выраженный дисбаланс: почти половина изображений не содержит признаков DR, в то время как тяжёлые формы представлены значительно реже. В рамках настоящей работы была выполнена бинаризация классов (No DR против DR) и балансировка через геометрические аугментации, включая повороты от 45° до 270° , что позволило добиться равномерного распределения классов для обучения моделей классификации.

Коллекция MAPLES-DR (MESSIDOR Anatomical and Pathological Labels for Explainable Screening of Diabetic Retinopathy) разработана исследовательской группой Cheriet et al. на базе изображений Messidor [172]. В отличие от оригинальной коллекции, MAPLES-DR включает аннотации на уровне пикселей для десяти структур: диск зрительного нерва (optic disc), сосудистая сеть, макула, мягкие и твёрдые экссудаты, кровоизлияния, микроаневризмы, неоваскуляризация, ватные пятна и друзы. Маски создавались вручную офтальмологами и валидировались независимыми экспертами. Помимо сегментации, каждому изображению сопоставлены уровни DR и макулярного отёка (DME), что позволяет использовать коллекцию в задачах совместной классификации и объяснимой сегментации. Разрешения варьируются от 1440×960 до 2304×1536 пикселей. Коллекция обладает высоким качеством аннотаций, однако классы представлены неравномерно: такие категории, как друзы, мягкие экссудаты и неоваскуляризация, встречаются реже других, что снижает сбалансированность для этих классов.

Коллекция FGADR (Fine-Grained Annotated Diabetic Retinopathy), опубликованная Zhou et al. [173], представляет собой наиболее масштабную на сегодняшний день подборку фундус-снимков с детальной аннотацией. Она включает 2842 изображения с разрешением 1280×1280 пикселей, полученные с различных камер (Canon CR-2, Topcon TRC-NW8, Kowa VX-10). Коллекция разделена на два поднабора: 1842 изображения снабжены пиксельными масками (seg-set), ещё 1000 размечены по стадии DR (grade-set). Сегментируемые классы включают микроаневризмы, кровоизлияния, твёрдые и мягкие экссудаты, интравитреальные микрососудистые аномалии (IRMA) и неоваскуляризацию. Аннотации выполнены двумя лицензированными офтальмологами с последующей верификацией. Авторы акцентируют внимание на том, что

FGADR ориентирован на тонкую диагностику, включая смешанные и пограничные проявления DR. Распределение по классам варьируется: наиболее полно представлены МА, НЕМ, EX и SE, в то время как IRMA и NV встречаются значительно реже.

Коллекция IDRiD (Indian Diabetic Retinopathy Image Dataset), представленная Porwal et al. [174], предназначена для оценки методов диагностики и локализации при DR. Она содержит 516 цветных изображений высокого разрешения (4288×2848 пикселей), снятых камерой Kowa VX-10 (FOV 50°). Из них 81 изображение размечено по пиксельным маскам для следующих классов: микроаневризмы, кровоизлияния, мягкие и твёрдые экссудаты, а также диск зрительного нерва. Кроме того, аннотации включают координаты центра фovea и диска, что позволяет использовать коллекцию для задач локализации и топографического анализа. Распределение масок между изображениями является умеренно сбалансированным: микроаневризмы и экссудаты представлены шире, чем кровоизлияния. Коллекция подходит для задач валидации на изображениях высокого разрешения и для сравнения с другими источниками.

На основе трёх указанных коллекций (MAPLES-DR, FGADR, IDRiD) была сформирована объединённая обучающая выборка. В неё включены только те изображения, для которых доступны аннотации по пяти ключевым патологическим признакам: микроаневризмы (МА), кровоизлияния (НЕМ), твёрдые экссудаты (EX), мягкие экссудаты (SE) и неоваскуляризация (NV). Все изображения были приведены к унифицированному пространственному размеру, маски – к согласованной структуре. Избыточные и неполные данные были удалены. Объединённая коллекция использовалась для обучения и оценки моделей семантической сегментации.

3.4 Предподготовка изображений

Шаг 1 – Исходные параметры изображений и приведение масштаба. Коллекции данных, рассматриваемые в данной работе, включают изображения с различными исходными разрешениями. В частности, в наборе IDRiD исполь-

Таблица 10 — Количественные характеристики и сбалансированность коллекций данных

Коллекция	Назначение	Кол-во изображений	Пиксельные маски	Разрешение	Сбалансированность
APTOS 2019	Классификация	3662	нет	варьируется	нет
MAPLES-DR	Сегментация	198	10 классов	1440–2304	нет
FGADR	Сегм., градация	2842 (1842+1000)	6 классов	1280×1280	да
IDRiD	Сегментация	516 (81 размечено)	5 классов	4288×2848	да

Таблица 11 — Источники классов в объединённой коллекции

Класс	патологий	MAPLES-DR	FGADR	IDRiD	Использованы
МА – микроаневризмы		да	да	да	да
НЕМ – кровоизлияния		да	да	да	да
ЕХ – твёрдые экссудаты		да	да	да	да
SE – мягкие экссудаты		да	да	да	да
NV – неоваскуляризация		да	да	нет	да

зуются макулоцентрированные фундус-снимки с полем зрения около 50° и разрешением 4288×2848 пикселей (формат JPEG, 8 бит на канал, RGB). В наборе FGADR представлены аннотированные изображения высокого качества с типичным разрешением 3072×2048 пикселей. Изображения из MAPLES-DR (на базе MESSIDOR) охватывают диапазон разрешений от 1440×960 до 2304×1536 пикселей и на этапе аннотирования были приведены к квадратному формату 1500×1500 пикселей.

Для унификации пространственного разрешения и обеспечения совместимости между наборами все изображения масштабируются до фиксированного квадратного размера:

$$S \times S = 1536 \times 1536 \text{ пикселей.}$$

Масштабирование изображений выполняется с коэффициентом $s = S/\min(H, W)$, при этом каждая точка x исходного изображения $I(x)$ сопоставляется точке $x' = sx$ в масштабированном изображении $I_S(x')$. Процедура сохраняет пропорции сторон, а при необходимости добавляется симметричный паддинг по меньшей стороне для получения квадратного изображения. Такой подход стандартизирует входные данные, упрощает выбор параметров CLANE и нейросетевых архитектур, а также исключает влияние различий в размере и геометрии снимков на последующую обработку.

Таблица 12 — Размеры изображений до и после приведения к единому масштабу

Характеристика	IDRiD	FGADR	MAPLES-DR
Исходное разрешение	4288×2848	3072×2048	1500×1500
Приведённый размер	1536×1536		

Шаг 2 – Нормализация яркости. Для приведения интенсивностей каждого цветового канала $c \in \{R, G, B\}$ к стандартному числовому диапазону используется линейная нормализация методом Min–Max. Преобразование основано на приведении значений пикселей к интервалу $[0, 1]$ по следующей формуле:

$$I_c^{(n)}(x) = \frac{I_c(x) - \min_{\Omega} I_c}{\max_{\Omega} I_c - \min_{\Omega} I_c}, \quad x \in \Omega, \quad (3.2)$$

где:

– $I_c(x)$ – значение пикселя в канале c исходного изображения;

- $\min_{\Omega} I_c$, $\max_{\Omega} I_c$ – минимальное и максимальное значение интенсивности в канале c по всем пикселям изображения;
- Ω – множество всех пикселей изображения.

Данная процедура обеспечивает равномерное растяжение динамического диапазона каждого канала в пределах $[0,1]$, что необходимо для стандартизации входных данных перед дальнейшими преобразованиями (в частности, перед применением CLANE).

Шаг 3 – Выделение зелёного канала. Поглощение гемоглобина максимально в зелёной области спектра ($\sim 520\text{--}560$ нм), поэтому сосудистое русло и ряд биомаркеров диабетической ретинопатии (микроаневризмы, геморрагии) контрастнее именно в зелёной проекции; красный канал склонен к пересвету, синий – к повышенному шуму [175–177]. Запишем Формально:

$$I_g(x) = I_G^{(n)}(x), \quad x \in \Omega. \quad (3.3)$$

Шаг 4 – билатеральный фильтр. Для подавления мелкого шума без размывания границ сосудов применяется билатеральная фильтрация к I_g :

$$J(x) = \frac{1}{W(x)} \sum_{\xi \in \mathcal{N}(x)} \exp \left(-\frac{\|x - \xi\|^2}{2\sigma_s^2} - \frac{(I_g(\xi) - I_g(x))^2}{2\sigma_r^2} \right) I_g(\xi), \quad (3.4)$$

где $W(x)$ – нормирующий множитель; $\sigma_s \approx 1.5\text{--}2.5$ пикс., $\sigma_r \approx 0.05\text{--}0.10$ в нормированном диапазоне [178].

Шаг 5 – Усиление контраста методом CLANE. CLANE (англ. Contrast Limited Adaptive Histogram Equalization) выполняет локальную гистограммную эквализацию с ограничением контраста. Изображение разбивается на равномерную сетку тайлов; в каждом тайле строится гистограмма, которая обрезается по заданному порогу, после чего её избыточные значения равномерно распределяются между всеми бинами. Затем значения интенсивностей преобразуются по нормированной интегральной гистограмме. На границах тайлов применяется билинейная интерполяция локальных отображений для обеспечения непрерывности результирующего изображения [179; 180].

Шаг 6 – Постнормировка. После применения CLANE результирующее изображение I_c может иметь значения интенсивностей, выходящие за пределы диапазона $[0,1]$. Для обеспечения корректной передачи контраста и согласованности с последующими этапами обработки (например, морфологической

Algorithm 1: CLAHE

Input: Original image I

Output: CLAHE-processed image I_c

- 1 1. Resize I to $M \times M$ (if needed); decompose into n tiles of size $m \times m$;
 - 2 2. For each tile t_n , compute histogram: $H_n \leftarrow \text{histogram}(t_n)$;
 - 3 3. Compute clip limit: $C_L \leftarrow N_{cl} \times N_{avg}$, where $N_{avg} \leftarrow \frac{N_x N_y}{N_{gray}}$;
 - 4 N_{gray} is the number of gray levels; N_x, N_y are tile dimensions;
 $N_{cl} \leftarrow 0.002$;
 - 5 4. Clip H_n using C_L : for all k such that $H_n(k) > C_L$, compute N_{ecl} and
distribute clipped pixels evenly: $N_{cp} \leftarrow N_{ecl}/N_{gray}$;
 - 6 5. Apply histogram equalization using the normalized cumulative
distribution function (CDF).;
 - 7 6. Perform bilinear interpolation of transformed values across neighboring
tiles to construct I_c ;
-

фильтрацией или пороговой сегментацией) проводится дополнительная линейная нормализация на весь диапазон значений. Итоговое изображение обозначим как I^* :

$$I^*(x) = \frac{I_c(x) - \min_{\Omega} I_c}{\max_{\Omega} I_c - \min_{\Omega} I_c}, \quad x \in \Omega, \quad (3.5)$$

где Ω – множество всех пикселей изображения.

Данная операция позволяет гарантировать, что все значения пикселей находятся в пределах $[0, 1]$, и нормализует локально усиленные области контраста, полученные в результате CLAHE.

3.5 Предварительная классификация

Цель данного этапа – отфильтровать изображения, не содержащие признаков искомого заболевания. На этом этапе решается задача бинарной классификации.

Пусть задана выборка

$$D = \{(\vec{x}_i, y_i)\}_{i=1}^N,$$

где $\vec{x}_i \in X = \mathbb{R}^m$ – множество объектов (изображений), представленных в векторной форме, а $y_i \in Y = \{-1, +1\}$ – метки классов, соответствующие двум состояниям: наличие заболевания и отсутствие заболевания. Таким образом, рассматривается случай бинарной классификации. Здесь $C = 2$ – количество классов в выборке, N – объём выборки.

Необходимо построить модель функции принятия решения

$$a : X \rightarrow Y,$$

которая аппроксимирует истинные метки y на всём множестве объектов X . В данном случае в роли функции принятия решений используется ИНС.

После получения классификационных метрик для изображения, поданного на вход сети (I_{prep}), принимается решение о дальнейшем его прохождении по конвейеру обработки. Если изображение классифицировано как диабетическая ретинопатия с вероятностью более 90 % или как отсутствие заболевания с вероятностью менее 90 %, оно передаётся на следующий этап. Если же изображение классифицировано как отсутствие заболевания с вероятностью более 90 %, дальнейший анализ прерывается, и пользователю отображается соответствующее информационное сообщение.

3.5.1 Модель VGG16

Модель VGG16 была предложена группой Visual Geometry Group (Оксфорд) в 2014 году в контексте соревнования ILSVRC и стала одной из ключевых вех развития глубоких сверточных сетей. Концептуальная новизна заключалась в использовании только малых свёрток размером 3×3 с единичными шагом и паддингом при последовательном увеличении глубины до 16 слоёв. Такая однородная архитектура с регулярным понижением разрешения посредством операции максимального объединения оказалась простой для реализации и масштабирования, дала существенный прирост качества и оказала влияние на конструкцию последующих моделей семейства глубоких CNN [181–183].

Архитектура VGG16 принимает на вход изображение $224 \times 224 \times 3$. Последовательность преобразований организована в пять сверточных блоков, каждый из которых заканчивается операцией максимального объединения 2×2

с шагом 2. Канальная размерность по блокам составляет от 64, до 512 каналов. В первых двух блоках по два сверточных слоя 3×3 , в последующих трёх – по три. Таким образом, пространственное разрешение уменьшается в 2 раза на каждом объединении: $224 \rightarrow 112 \rightarrow 56 \rightarrow 28 \rightarrow 14 \rightarrow 7$. После сверточной части следуют три полносвязных слоя с понижающейся размерностью за которыми расположен классификатор на 1000 классов, реализованный в виде финального слоя с функцией softmax. Во всех скрытых слоях применяется нелинейная функция активации ReLU. В ряде реализаций также используются dropout и L2-регуляризация [181]. Размерности слоёв и количество параметров VGG16 представлены в таблице 13.

С точки зрения обработки изображений, каскад малых свёрток реализует поэтапное пространственное агрегирование локальных признаков при сбалансированном соотношении между числом параметров и извлекаемой пространственной информацией. Две последовательные свёртки 3×3 обладают эффективным приёмным полем, эквивалентным свёртке 5×5 , однако требуют меньше параметров и позволяют ввести дополнительную нелинейность между слоями. Формально, выход k -го канала сверточного слоя в позиции (i, j) задаётся следующим образом:

$$y_{i,j,k} = \sum_{m=0}^2 \sum_{n=0}^2 \sum_{c=1}^{C_{in}} x_{i+m-1, j+n-1, c} w_{m,n,c,k} + b_k, \quad (3.6)$$

где x – входной тензор, w – веса свёртки, b_k – смещение. Функция активации имеет вид:

$$f(t) = \max(0, t), \quad (3.7)$$

а операция максимального объединения 2×2 со шагом 2 записывается:

$$z_{i,j,k} = \max_{(m,n) \in \{0,1\}^2} y_{i+m, j+n, k}. \quad (3.8)$$

Однородная структура и регулярное понижение размерностей упростили перенос архитектурных принципов VGG в более сложные модели. Последующие архитектуры, такие как ResNet [184] и DenseNet [185], развивали эту идею за счёт введения остаточных и плотных связей. Обзор применения и развития свёрточных сетей представлен в работах [182; 183].

Модель VGG16 демонстрирует согласованный набор проектных принципов: локальность обработки за счёт малых фильтров, устойчивость к смещению

Таблица 13 — Размерности слоёв и количество параметров VGG16

Слой	Размер выходного тензора	Ядро / шаг	Кол-во параметров
Conv1_1	$224 \times 224 \times 64$	$3 \times 3 \times 3 \times 64$	1 792
Conv1_2	$224 \times 224 \times 64$	$3 \times 3 \times 64 \times 64$	36 928
MaxPool1	$112 \times 112 \times 64$	$2 \times 2/2$	0
Conv2_1	$112 \times 112 \times 128$	$3 \times 3 \times 64 \times 128$	73 856
Conv2_2	$112 \times 112 \times 128$	$3 \times 3 \times 128 \times 128$	147 584
MaxPool2	$56 \times 56 \times 128$	$2 \times 2/2$	0
Conv3_1	$56 \times 56 \times 256$	$3 \times 3 \times 128 \times 256$	295 168
Conv3_2	$56 \times 56 \times 256$	$3 \times 3 \times 256 \times 256$	590 080
Conv3_3	$56 \times 56 \times 256$	$3 \times 3 \times 256 \times 256$	590 080
MaxPool3	$28 \times 28 \times 256$	$2 \times 2/2$	0
Conv4_1	$28 \times 28 \times 512$	$3 \times 3 \times 256 \times 512$	1 180 160
Conv4_2	$28 \times 28 \times 512$	$3 \times 3 \times 512 \times 512$	2 359 808
Conv4_3	$28 \times 28 \times 512$	$3 \times 3 \times 512 \times 512$	2 359 808
MaxPool4	$14 \times 14 \times 512$	$2 \times 2/2$	0
Conv5_1	$14 \times 14 \times 512$	$3 \times 3 \times 512 \times 512$	2 359 808
Conv5_2	$14 \times 14 \times 512$	$3 \times 3 \times 512 \times 512$	2 359 808
Conv5_3	$14 \times 14 \times 512$	$3 \times 3 \times 512 \times 512$	2 359 808
MaxPool5	$7 \times 7 \times 512$	$2 \times 2/2$	0
FC6	4096	$7 \times 7 \times 512 \rightarrow 4096$	102 764 544
FC7	4096	$4096 \rightarrow 4096$	16 781 312
FC8	1000	$4096 \rightarrow 1000$	4 097 000
Итого			138 357 544

объектов на изображении благодаря свёрточной структуре, а также последовательное объединение локальных признаков в более глобальные структуры. Такая архитектура обеспечивает устойчивое извлечение признаков при контролируемой сложности модели и остаётся эталонной во многих задачах компьютерного зрения.

3.5.2 Метрики оценки качества

Выбор метрик качества для оценки бинарной классификации изображений глазного дна обусловлен спецификой задачи предварительного скрининга, где приоритет отдается высокой чувствительности модели к патологическим случаям. Метрика *recall* позволяет контролировать долю правильно выявленных положительных классов (диабетическая ретинопатия) и минимизировать риск ложноотрицательных решений, критичных для последующего этапа диагностики. *F1-score*, как гармоническое среднее между *recall* и *precision*, отражает сбалансированность модели при ошибках разного рода и обеспечивает надёжную оценку в условиях возможного дисбаланса классов. Все указанные метрики формализованы ранее.

3.5.3 Обучение, валидация, результаты

Для предварительной бинарной классификации фундус-снимков использован модифицированная коллекция данных APTOS 2019, в которой изображения всех стадий диабетической ретинопатии (Mild, Moderate, Severe, Proliferative) были объединены в один класс – «диабетическая ретинопатия», а класс Normal сохранён как «норма». После процедур аугментации и балансировки итоговый корпус составил 4800 изображений, по 2400 в каждом классе. Разбиение выполнено в соотношении: 70% (3360 изображений) использованы для обучающей выборки, 15% (720 изображений) – для валидационной, оставшиеся 15% (720 изображений) – для тестовой. Во всех выборках сохранён баланс между классами [186; 187].

На этапе предварительной обработки изображения были масштабированы до 224×224 пикселей и нормированы по требованиям ImageNet. Для повышения обобщающей способности модели применялись стандартные методы аугментации не нарушающие пространственную структуру изображения: случайные повороты до $\pm 15^\circ$, горизонтальные и вертикальные отражения. Все аугментации применялись исключительно к обучающей выборке [188]. Для учё-

та возможного дисбаланса классов использовались стратифицированные батчи и веса классов в функции потерь [187].

В качестве базовой модели использована сверточная нейросеть VGG16 с предобученными на ImageNet весами. Исходный классификатор модели был заменён на последовательность: Global Average Pooling \rightarrow Dense(256, ReLU) \rightarrow Dropout(0.5) \rightarrow Dense(2, softmax). Обучение проводилось в два этапа: на первом этапе замороженные сверточные блоки не обновлялись, обучение велось только для новой головы в течение 5 эпох при скорости обучения $lr = 10^{-3}$ с оптимизатором Adam. На втором этапе размораживались блоки `conv4_` и `conv5_`, которые дообучались вместе с головой до 40 эпох при $lr = 10^{-4}$ и weight decay 10^{-4} [189; 190]. Использовались следующие приёмы: снижение скорости обучения при отсутствии улучшения (ReduceLROnPlateau), ранняя остановка по метрике macro-F1 с patience = 8 и сохранение лучших весов модели.

В качестве Функции потерь была выбрана взвешенная перекрёстная энтропия:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c \in \{\text{norm}, \text{dr}\}} w_c y_{i,c} \log p_{i,c}, \quad (3.9)$$

где w_c – вес класса, $y_{i,c}$ – истинная метка, $p_{i,c}$ – вероятность, предсказанная моделью, N – общее число примеров. При балансировке $w_c = 1$.

Подбор гиперпараметров производился с использованием стратифицированной 5-кратной кросс-валидации на обучающей выборке (70

На тестовой выборке из 720 изображений (360 «норма», 360 «диабетическая ретинопатия») модель достигла следующих результатов при пороговом значении 0.8:

Таблица 14 — Результаты VGG16 на тестовой выборке

Класс	Recall (%)	F1-score (%)
Норма	98.1	97.1
Диабетическая ретинопатия	97.8	97.9

Графическая интерпретация результатов представлена на рисунке 3.4, где показана матрица ошибок.

Таким образом, правильно классифицировано 705 из 720 изображений. Высокая чувствительность по патологическому классу подтверждает надёжность предложенного подхода как инструмента для предварительного фильтра

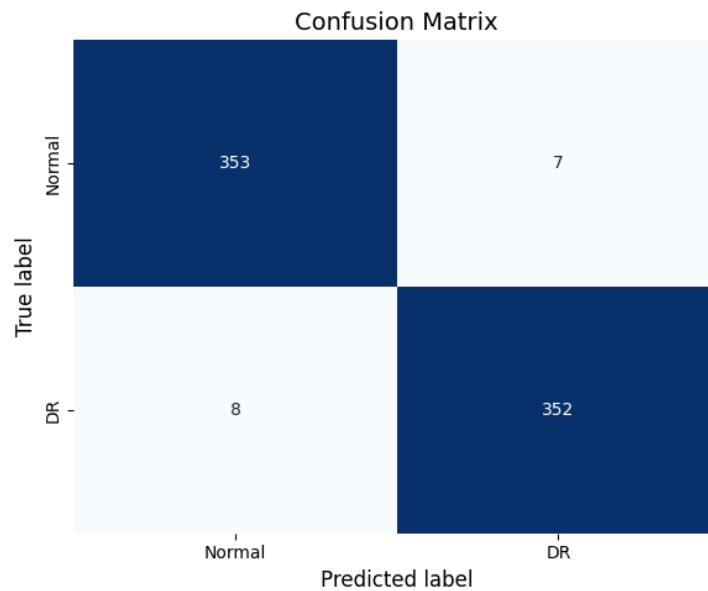


Рисунок 3.4 — Матрица ошибок VGG16

изображений. Данный этап позволяет отсеять очевидно нормальные случаи и направить подозрительные изображения на последующую детальную обработку. Высокая F1-оценка указывает на сбалансированную производительность модели, достигнутую за счёт использования трансферного обучения, а также применения современных приёмов регуляризации.

3.6 Морфологические преобразования

3.6.1 Метод поиска анатомических зон глазного дна

Автоматизированное определение анатомических ориентиров глазного дна, таких как центр оптического диска и зона фовеа, является ключевым этапом для последующей локализации патологий и топографического анализа распределения биомаркеров. Существующие подходы к решению этой задачи можно условно разделить на три группы. Первая включает морфологические методы, основанные на морфологических преобразованиях, геометрических характеристиках (площадь, круглость, компактность) и простых эвристиках. Вторая группа опирается на анатомо-геометрические соотношения между структурами глазного дна, включая относительное расположение сосудистого

русла, центра диска и фовеа. Третья представлена обучаемыми моделями глубокого обучения, использующими предварительную аннотацию и обучающиеся по изображению целиком. Основным недостатком обучаемых методов является зависимость от размеченных выборок и чувствительность к смещению распределения, тогда как классические подходы требуют ручной настройки и слабо устойчивы к шуму и артефактам. В данной работе предлагается новый метод локализации анатомических зон на основе анализа распределения яркости и морфологических операций, не требующий обучения и демонстрирующий устойчивость к условиям реальной съёмки.

Для анализа изображения каждый цветовой канал $c \in \{R, G, B\}$ представляется в виде гистограммы яркости $H_c(k)$, отражающей количество пикселей с интенсивностью k , где $k = 0, \dots, 255$ [191]. Для компенсации неравномерного освещения и фона изображения предварительно выполняется морфологическая коррекция, а именно: из исходного изображения вычитается его сглаженная версия, полученная с помощью операции открытия, что позволяет усилить локальные яркие структуры на фоне медленно меняющегося фона:

$$I_c^* = I_c - (I_c \circ b_r), \quad (3.10)$$

где \circ обозначает морфологическое открытие с круглым структурирующим элементом радиуса r [192].

Для дальнейшего поиска оптического диска и зоны фовеа из трёх каналов выбираются наиболее информативные: канал с наибольшим содержанием ярких пикселей и канал с наибольшей концентрацией тёмных значений. В первом случае анализируется правый хвост гистограммы:

$$c_{OD} = \arg \max_c \sum_{k=\lceil \mu_c + \alpha \sigma_c \rceil}^{255} H_c(k), \quad (3.11)$$

а во втором – левый:

$$c_F = \arg \max_c \sum_{k=0}^{\lfloor \mu_c - \alpha \sigma_c \rfloor} H_c(k), \quad (3.12)$$

где μ_c и σ_c – среднее и стандартное отклонение яркости канала c , $\alpha \in [1.0, 2.5]$ – масштабный коэффициент. Под хвостом гистограммы в данном контексте понимается часть распределения, выходящая за границы интервала

$[\mu_c - \alpha\sigma_c, \mu_c + \alpha\sigma_c]$; левый хвост соответствует тёмным значениям, правый – ярким значениям пикселей.

После выбора каналов выполняется пороговая бинаризация с использованием адаптивных порогов T_{OD} и T_F , полученных, например, методом Отсу:

$$B_{OD}(x,y) = \begin{cases} 1, & I_{c_{OD}}^*(x,y) \geq T_{OD}, \\ 0, & \text{иначе,} \end{cases} \quad B_F(x,y) = \begin{cases} 1, & I_{c_F}^*(x,y) \leq T_F, \\ 0, & \text{иначе.} \end{cases} \quad (3.13)$$

Пороговые значения уточняются после отсечения выбросов по крайним квантилям от общей гистограммы интенсивности.

После бинаризации и морфологической фильтрации в маске B_{OD} остаются связные компоненты $\{\Omega_i\}$, каждая из которых может потенциально соответствовать оптическому диску. Для выбора наилучшей области применяется интегральная оценка, учитывающая как геометрические, так и позиционные признаки компонент. Для каждой области Ω_i вычисляется агрегированная оценка, отражающая её соответствие предполагаемой геометрии и положению оптического диска:

$$S_i = \lambda_1 \cdot \frac{|\Omega_i|}{\max_j |\Omega_j|} + \lambda_2 \cdot \frac{4\pi|\Omega_i|}{P_i^2} - \lambda_3 \cdot e_i - \lambda_4 \cdot \rho_i, \quad (3.14)$$

где:

- $|\Omega_i|$ – площадь области (в пикселях);
- P_i – периметр области;
- $\frac{4\pi|\Omega_i|}{P_i^2}$ – показатель округлости, достигающий максимума при идеально круглой форме;
- e_i – эксцентриситет, измеряющий вытянутость формы ($e_i = 0$ соответствует кругу);
- ρ_i – нормализованное евклидово расстояние от центра масс области до центра изображения;
- $\lambda_k \geq 0$, $\sum_{k=1}^4 \lambda_k = 1$ – веса, определяющие значимость каждого критерия.

Компонента с максимальным значением оценки S_i выбирается как наиболее вероятный кандидат на оптический диск:

$$\Omega^* = \arg \max_i S_i. \quad (3.15)$$

Её центр вычисляется как центр масс:

$$\mathbf{c}_{OD} = \frac{1}{|\Omega^*|} \sum_{(x,y) \in \Omega^*} (x,y)^\top, \quad (3.16)$$

а диаметр — как диаметр окружности той же площади:

$$D_{OD} = 2\sqrt{\frac{|\Omega^*|}{\pi}}. \quad (3.17)$$

Для последующего поиска центра фовеа формируется область интереса с радиусами $[\kappa_1, \kappa_2]D_{OD}$, центрированная в точке \mathbf{c}_{OD} и смещённая в височную сторону (анатомически противоположную расположению диска) [193]. На практике параметры выбираются из диапазонов $\kappa_1 \in [2.0, 2.5]$, $\kappa_2 \in [3.0, 3.5]$.

Для определения центра фовеа $I_{c_F}^*$ используется свёртка с гауссовым ядром:

$$J = G_\sigma * I_{c_F}^*, \quad (3.18)$$

где G_σ — гауссов фильтр с параметром сглаживания σ , а $*$ — операция двумерной свёртки. Центр фовеа \mathbf{c}_F определяется как точка минимума полученного поля яркости внутри заданной области поиска:

$$\mathbf{c}_F = \arg \min_{\mathbf{p} \in \text{ROI}} J(\mathbf{p}). \quad (3.19)$$

Ниже приведён пример результата автоматического поиска центра зоны фовеа и оптического диска на фундус-снимке (рис. ??).

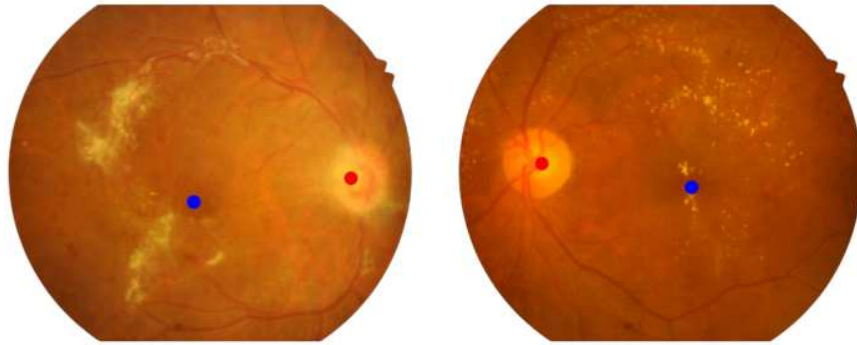


Рисунок 3.5 — Пример результата автоматического поиска центра зоны фовеа и оптического диска: центр ОД – красный, зоны фовеа – синий

Предложенный метод не содержит обучаемых параметров, базируется на морфологических преобразованиях и статистических характеристиках гистограммцветовых каналов, что обеспечивает её устойчивость к артефактам

съёмки (таким как блики, шум, неравномерная освещённость) и вариативности сосудистого рисунка. С точки зрения вычислительной сложности все этапы метода имеют линейную асимптотику по числу пикселей. Параметры r , α , κ_1 , κ_2 , σ выбираются из фиксированных диапазонов и не требуют настройки на размеченных данных.

3.6.2 Сегментация зоны фовеа

Зона фовеа играет ключевую роль в диагностике и количественном анализе патологических изменений при диабетической ретинопатии. Её корректная локализация необходима для оценки симметрии изменений и выделения центральной макулярной области. Учитывая, что ранее были определены координаты центра фовеа (c_x, c_y) и центра диска зрительного нерва (x_d, y_d) , возможно построить простую модель сегментации на основе круговой аппроксимации.

В рамках упрощённого подхода предполагается, что граница зоны фовеа может быть аппроксимирована окружностью фиксированного радиуса r_f , определяемого относительно размера диска или расстояния между центрами. Радиус зоны фовеа задаётся на основе анатомо-геометрического соотношения между фовеа и диском.

Если известна маска диска, радиус фовеа может быть выражен как доля от его эффективного радиуса:

$$r_f = \beta \cdot r_{\text{disc}}, \quad \beta \in [0,45; 0,60], \quad (3.20)$$

где $r_{\text{disc}} = \sqrt{A_{\text{disc}}/\pi}$, а A_{disc} — площадь маски диска (в пикселях) [194; 195].

Если маска диска отсутствует, используется расстояние между центром фовеа и центром диска:

$$d = \sqrt{(c_x - x_d)^2 + (c_y - y_d)^2}, \quad r_f = \alpha \cdot d, \quad \alpha \in [0,10; 0,18], \quad (3.21)$$

что также согласуется с анатомо-геометрическими соотношениями между положением фовеа и диска на фундус-снимках [195].

Полученная маска фовеа $F \subset D$ формируется по определению:

$$F = \left\{ (x, y) \in D \mid \sqrt{(x - c_x)^2 + (y - c_y)^2} \leq r_f \right\}. \quad (3.22)$$

На практике результат может быть дополнительно сглажен или отфильтрован. При наличии калибровки по пикселям на миллиметр, площадь и радиус зоны могут быть приведены к физическим единицам.

Данный метод отличается высокой воспроизводимостью, вычислительной простотой и не требует обучения. Его точность зависит от качества предварительного определения центра фовеа и может быть подтверждена метриками Dice и IoU при наличии эталонной разметки.

Такой подход может быть использован в практических системах анализа глазного дна для построения анатомических зон интереса, пространственной агрегации признаков, а также для извлечения количественных биомаркеров центральной макулярной области, в частности — плотности микрогеморрагий, твёрдых экссудатов и микрососудистых аномалий в пределах фовеального кольца [196].

3.6.3 Разделение на квадранты

Пусть $(c_x, c_y) \in \mathbb{Z}^2$ – координаты центра фовеа, найденные на предыдущем этапе обработки. Тогда изображение делится на четыре квадранта, определяемых по следующей схеме:

$$\begin{aligned} Q_1 &= \{(x, y) \in D \mid c_x \leq x < w, 0 \leq y < c_y\}, \\ Q_2 &= \{(x, y) \in D \mid c_x \leq x < w, c_y \leq y < h\}, \\ Q_3 &= \{(x, y) \in D \mid 0 \leq x < c_x, c_y \leq y < h\}, \\ Q_4 &= \{(x, y) \in D \mid 0 \leq x < c_x, 0 \leq y < c_y\}. \end{aligned} \tag{3.23}$$

Каждый квадрант представляет собой прямоугольную область, соответствующую одной из четырёх сторон изображения относительно центра фовеа. Объединение всех квадрантов даёт полную область изображения:

$$D = Q_1 \cup Q_2 \cup Q_3 \cup Q_4, \quad Q_i \cap Q_j = \emptyset, \quad \text{при } i \neq j. \tag{3.24}$$

Ниже представлен пример разбиения изображения на квадранты (рис. 3.6).

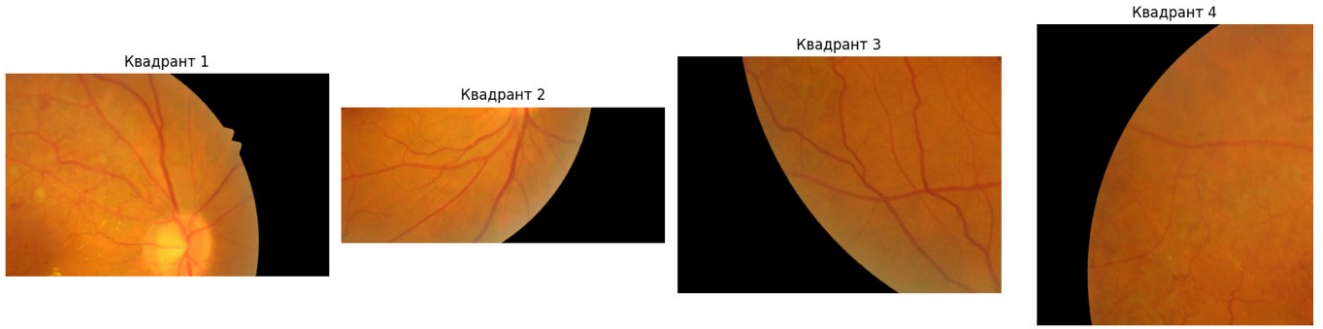


Рисунок 3.6 — Пример разделения изображения на квадранты относительно центра фовеа

Такое разбиение используется при локализованном анализе фундус-снимка, например, при вычислении распределения биомаркеров по анатомическим регионам сетчатки. Для каждого квадранта можно определить локальные маски классов биомаркеров и подсчитать как их суммарную площадь (в пикселях), так и количество отдельных экземпляров с помощью алгоритма связных компонент.

Разбиение на квадранты позволяет формализовать топографическую оценку распределения патологических изменений и использовать эту информацию в правилах логического вывода или принятия решений.

3.7 Семантическая сегментация биомаркеров ДР

Целью является выделение контуров и классификация областей, содержащих патологии (микроаневризмы, кровоизлияния, твёрдые и мягкие экссудаты). На данном этапе рассматривается задача семантической сегментации.

Пусть каждый пиксель изображения I , представленного в виде матрицы размерности $W \times H$, соответствует некоторой дискретной функции:

$$u : \{1, \dots, M\} \times \{1, \dots, N\} \rightarrow \{0, \dots, 255\}^C,$$

отражающей информацию об интенсивности пикселей с координатами i, j , где количество каналов $C = 3$ для цветного изображения. Каждый пиксель идентифицируется соответствующим вектором x_j . Тогда сегментация изображения

представляет собой разбиение множества

$$X = \{x_j\}_{j=1}^d, \quad \text{где } d = WH,$$

на кластеры X_j , $j = \overline{1, l}$, при этом каждому кластеру X_j сопоставляется определённое цветовое значение, соответствующее анатомической области изображения или фону, который он выделяет.

Задача требует выполнения семантической сегментации, то есть всё множество пикселей изображения P_t должно быть разделено на n кластеров пикселей, каждый из которых отнесён к определённому классу, в соответствии со следующим выражением:

$$\bigcup_{i=1}^C X_j = P_t, \quad X_i \cap X_j \neq \emptyset, \quad i \neq j, \quad i, j = \overline{1, C}. \quad (1.1)$$

В данном случае ИНС выступает в роли функции принятия решений. После получения результатов сегментации (обнаруженные классы и соответствующие им вероятности при превышении порога в 80 %) они передаются на следующий этап.

3.7.1 Модель U-Net

Модель U-Net была предложена в 2015 году в Университете Фрайбурга для задач биомедицинской семантической сегментации и быстро стала стандартом в данной области благодаря сочетанию точности локализации и устойчивости на малых выборках [197]. Архитектура имеет U-образную форму и состоит из нисходящего пути кодировщика, уровня бутылочного горлышка и восходящего пути декодировщика. Кодировщик понижает пространственное разрешение и расширяет число каналов, извлекая контекст, а декодировщик восстанавливает разрешение и уточняет границы объектов [198; 199].

В классической реализации на вход подаётся изображение размером $572 \times 572 \times C$. Каждый блок кодировщика содержит две свёртки 3×3 с активацией ReLU и операцию максимального объединения 2×2 со шагом 2; число каналов удваивается при каждом понижении (64, 128, 256, 512), на уровне бутылочного горлышка – 1024 [197]. Восходящий путь симметричен: сначала транспонированная свёртка 2×2 (апсемплинг), затем конкатенация с признаками кодировщика

(англ. skip connection), после чего выполняются две свёртки 3×3 с ReLU. Финальный слой представляет собой свёртку 1×1 , проецирующую признаки в пространство классов, и поканальную softmax-нормализацию. В современных реализациях часто применяется same-padding, упрощающий выравнивание тензоров [200].

Размерности слоёв и количество параметров модели U-Net представлены в таблице 15. Схематичное изображение архитектуры представлено на рисунке 3.7.

Таблица 15 — Размерности слоёв и количество параметров модели U-Net

Слой	Размер выходного тензора	Описание операций	Параметры
Input	$572 \times 572 \times 1$	Входное изображение	—
Conv1_1	$570 \times 570 \times 64$	Conv 3×3 , ReLU	0.064 млн
Conv1_2	$568 \times 568 \times 64$	Conv 3×3 , ReLU	0.037 млн
MaxPool1	$284 \times 284 \times 64$	MaxPool 2×2	—
Conv2_1	$282 \times 282 \times 128$	Conv 3×3 , ReLU	0.074 млн
Conv2_2	$280 \times 280 \times 128$	Conv 3×3 , ReLU	0.148 млн
MaxPool2	$140 \times 140 \times 128$	MaxPool 2×2	—
Conv3_1	$138 \times 138 \times 256$	Conv 3×3 , ReLU	0.295 млн
Conv3_2	$136 \times 136 \times 256$	Conv 3×3 , ReLU	0.590 млн
MaxPool3	$68 \times 68 \times 256$	MaxPool 2×2	—
Conv4_1	$66 \times 66 \times 512$	Conv 3×3 , ReLU	1.180 млн
Conv4_2	$64 \times 64 \times 512$	Conv 3×3 , ReLU	2.359 млн
MaxPool4	$32 \times 32 \times 512$	MaxPool 2×2	—
Conv5_1	$30 \times 30 \times 1024$	Conv 3×3 , ReLU	4.719 млн
Conv5_2	$28 \times 28 \times 1024$	Conv 3×3 , ReLU	9.438 млн
UpConv1	$56 \times 56 \times 512$	Transposed Conv 2×2	2.097 млн
Concat1 + Conv6_1	$54 \times 54 \times 1024$	Conv 3×3 , ReLU	2.359 млн
Conv6_2	$52 \times 52 \times 512$	Conv 3×3 , ReLU	1.180 млн
...
Conv10	$388 \times 388 \times 2$	Conv 1×1 (2 класса)	0.002 млн
Итого (приблизительно)			≈ 31 млн

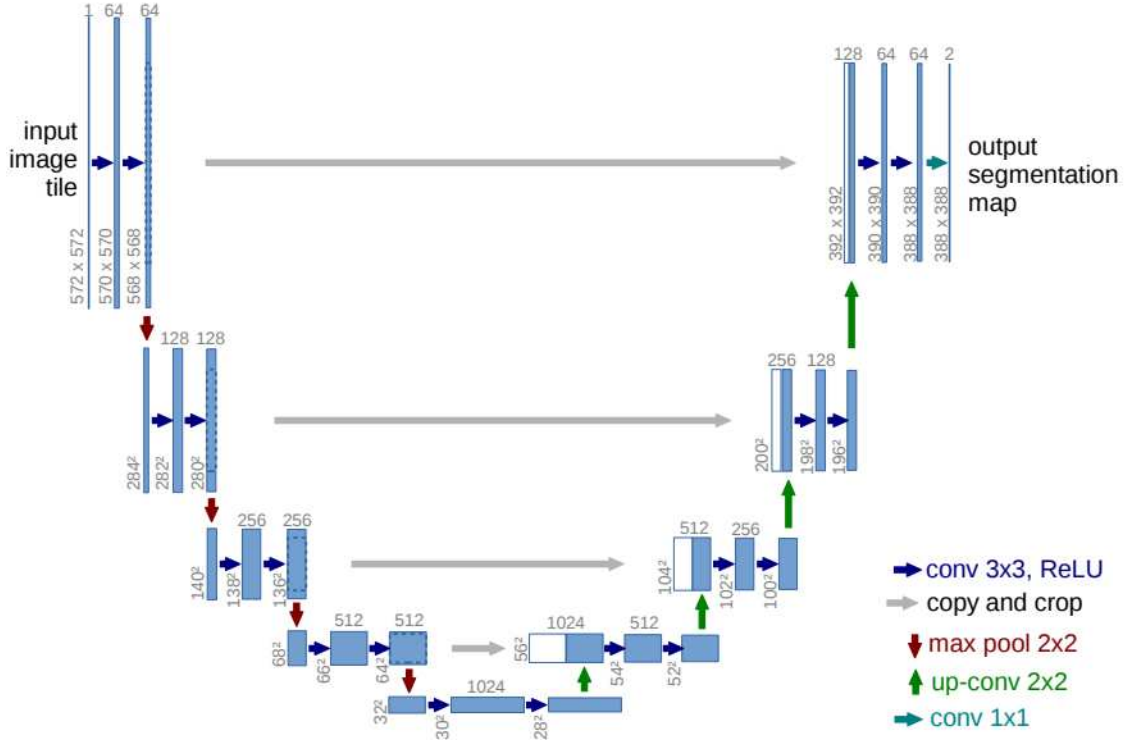


Рисунок 3.7 — Схема архитектуры модели U-Net [197]

Математическая формализация. Свёртка с ядром $K \times K$ и паддингом p :

$$y_{i,j,k} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \sum_{c=1}^{C_{in}} x_{i+m-p, j+n-p, c} \cdot w_{m,n,c,k} + b_k, \quad (3.25)$$

Функция активации ReLU:

$$\text{ReLU}(t) = \max(0, t), \quad (3.26)$$

Максимальное объединение 2×2 со шагом 2:

$$z_{i,j,k} = \max_{(m,n) \in \{0,1\}^2} y_{i+m, j+n, k}. \quad (3.27)$$

Транспонированная свёртка:

$$\text{TConv}_{w,s}(x) = \text{Conv}_w(\text{ups}_s(x)), \quad (3.28)$$

Прямые связи (skip connections):

$$G_l = \text{Concat}_c \left(\tilde{F}_l^{\text{dec}}, \text{Align}(F_l^{\text{enc}}) \right), \quad (3.29)$$

Финальная классификация пикселя с помощью softmax:

$$\hat{y}_{i,j} = \arg \max_{k \in \{1, \dots, K\}} \frac{e^{z_{i,j,k}}}{\sum_{m=1}^K e^{z_{i,j,m}}}, \quad (3.30)$$

U-Net обеспечивает компромисс между извлечением контекста и точной локализацией благодаря симметричной архитектуре, прямым связям и простоте операций. Модель стала основой для множества модификаций: UNet++ [199], Attention U-Net [201], 3D U-Net [202], nnU-Net [200].

3.7.2 Фильтры Габора

Фильтры Габора представляют собой линейные пространственно-частотные операторы, обеспечивающие одновременную локализацию в пространственной и частотной областях. Фильтр Габора представляет собой синусоиду, умноженную на двумерную гауссовую функцию. Такое сочетание позволяет минимизировать соотношение неопределённости между пространством и частотой, что делает габоровские фильтры эффективным инструментом в задачах обработки изображений [203; 204]. Они находят широкое применение при анализе ориентированных и полосовых структур, например, в задачах выделения текстур, линий и сосудистых образований. Особенно актуально использование таких фильтров в медицинских изображениях, где требуется надёжное распознавание вытянутых и направленных объектов, таких как кровеносные сосуды и биомаркеры диабетической ретинопатии (например, микроаневризмы и кровоизлияния) на фундус-снимках.

Математически двумерный фильтр Габора описывается функцией, представляющей собой произведение гауссовой и гармонической составляющих. В наиболее распространённой форме его импульсная характеристика выражается как:

$$G_R(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \varphi\right), \quad (3.31)$$

где x' и y' — координаты в системе, повернутой на угол θ , определяются следующим образом:

$$\begin{aligned} x' &= x \cos \theta + y \sin \theta, \\ y' &= -x \sin \theta + y \cos \theta. \end{aligned} \quad (3.32)$$

Здесь λ — длина волны синусоидальной составляющей (обратнопропорциональна частоте), θ — ориентация фильтра, φ — фазовый сдвиг, σ — ширина

гауссовой огибающей, γ – коэффициент формы, определяющий степень вытянутости фильтра. Такая реализация соответствует действительной части комплексной габоровской функции и широко применяется в задачах анализа изображений.

В задаче извлечения признаков из изображения с использованием банка габоровских фильтров, отличающихся по ориентации, удобно определить значение признака как индекс фильтра, дающего наименьший отклик в данной точке. Пусть $I(x, y)$ – изображение, а G_R – фильтр с ориентацией $\theta_j = \frac{j\pi}{J}$, где $j = 0, \dots, J - 1$. Тогда значение признака можно определить как:

$$F(x, y) = \arg \min_j \{(I * G_R)(x, y; \lambda, \theta_j, \varphi, \sigma, \gamma)\}, \quad (3.33)$$

где $*$ обозначает операцию свёртки.

Если применить банк фильтров с различными ориентациями, то в каждой точке изображения можно выбрать ту ориентацию, при которой отклик фильтра минимален, что позволяет оценить локальное направление структуры.

Для устойчивого покрытия пространственно-частотной плоскости применяется банк фильтров с параметрами, равномерно распределёнными по ориентации и логарифмически по частоте. Частота ω_n и угол θ_m могут быть заданы формулами:

$$\omega_n = \frac{\pi}{2} \cdot 2^{-(n-1)/2}, \quad \theta_m = \frac{\pi}{8}(m - 1), \quad (3.34)$$

где $n = 1, \dots, 5$, $m = 1, \dots, 8$. Ширина гауссовой огибающей обычно выбирается как $\sigma \approx \pi/\omega$, а фазовый сдвиг φ – из равномерного распределения $\mathcal{U}(0, \pi)$ [205].

Таким образом, фильтры Габора являются универсальным и строго формализованным инструментом выделения ориентированных признаков, применимым в различных областях технического и медицинского анализа изображений.

3.7.3 Модификация базовой модели

В настоящей работе реализована модификация архитектуры U-Net, в которой первый свёрточный блок энкодера заменён на обучаемый слой с ядрами, параметризуемыми габоровской функцией. Цель такого подхода – повысить

способность модели извлекать ориентированные полосовые признаки на ранних этапах обработки, что особенно важно для сегментации тонких сосудов, направленных контуров и биомаркеров диабетической ретинопатии на фундус-снимках и ОСТ-изображениях [206].

Каждое ядро обучаемого слоя представляется в виде габоровской функции, которая включает гармоническую составляющую с заданной частотой и ориентацией, с амплитудой, уменьшающейся по пространству в соответствии с двумерной гауссовой функцией. Параметры фильтра управляют длиной волны, направлением, фазовым сдвигом, масштабом и формой фильтра. Чтобы гарантировать положительные значения масштаба, длины волны и формы, параметры репараметризуются с помощью экспоненциальных преобразований. Для устойчивости в обучении ядро предварительно масштабируется до фиксированной нормы и вычитается среднее значение, чтобы сделать его отклик сбалансированным по амплитуде и не смещённым по яркости.

На этапе инициализации параметры фильтров выбираются из регулярного банка, охватывающего различные частоты и ориентации. Частоты дискретизируются по логарифмической сетке, а ориентации — равномерно:

$$\omega_n = \frac{\pi}{2} \cdot 2^{-\frac{(n-1)}{2}}, \quad \theta_m = \frac{\pi}{8}(m-1), \quad n = 1..5, m = 1..8. \quad (3.35)$$

Соответствующий масштаб определяется как

$$\sigma \approx \frac{\pi}{\omega_n}, \quad \varphi \sim \mathcal{U}(0, \pi). \quad (3.36)$$

Модифицированный блок встраивается в U-Net на месте стандартного первого свёрточного блока энкодера. Его структура имеет вид:

$$\text{GaborConv}(k \times k) \rightarrow \text{BatchNorm} \rightarrow \text{ReLU} \rightarrow \text{Conv}(3 \times 3). \quad (3.37)$$

Такой модуль сохраняет размерность признаков карт и совместим со структурой skip-соединений. Вместо двух обычных свёрток 3×3 , применяется одна обучаемая параметрическая свёртка с габоровскими ядрами и одна обычная. Размер ядра выбирается в диапазоне $k \in [7, 11]$, число выходных каналов подбирается так, чтобы не увеличивать размерность тензора признаков.

Несмотря на некоторое увеличение вычислительных затрат из-за наличия экспоненциальных и тригонометрических операций, такая модификация слоя обеспечивает более быструю и устойчивую сходимость модели. Обучаемые параметры ядра обновляются градиентным методом наравне с остальными

параметрами модели, с возможным уменьшением шага обучения для стабильности на ранних этапах.

Для повышения эффективности обучения в данной работе были использованы аналитические выражения градиентов по основным параметрам фильтра, что позволяет избежать неточностей численного дифференцирования. Ниже приведены выражения для производных по длине волны λ , фазе φ , масштабу σ , коэффициенту формы γ и ориентации θ :

$$\frac{\partial G}{\partial \lambda} = A \cdot \frac{2\pi x'}{\lambda^2} \sin \eta, \quad \frac{\partial G}{\partial \varphi} = -A \sin \eta, \quad (3.38)$$

$$\frac{\partial G}{\partial \sigma} = \frac{x'^2 + \gamma^2 y'^2}{\sigma^3} \cdot G, \quad \frac{\partial G}{\partial \gamma} = -\frac{\gamma y'^2}{\sigma^2} \cdot G, \quad (3.39)$$

$$\frac{\partial G}{\partial \theta} = -A \cos \eta \cdot \frac{x' \cdot \partial_\theta x' + \gamma^2 y' \cdot \partial_\theta y'}{\sigma^2} - A \sin \eta \cdot \frac{2\pi}{\lambda} \cdot \partial_\theta x', \quad (3.40)$$

где

$$A = \exp \left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right), \quad \eta = \frac{2\pi x'}{\lambda} + \varphi, \quad (3.41)$$

$$\partial_\theta x' = -x \sin \theta + y \cos \theta, \quad \partial_\theta y' = -x \cos \theta - y \sin \theta. \quad (3.42)$$

Градиенты по логарифмическим параметрам ℓ, s, g (введённым для обеспечения положительности λ, σ, γ) вычисляются по правилу цепочки. Например:

$$\frac{\partial G}{\partial s} = \frac{\partial G}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial s} = \sigma \cdot \frac{\partial G}{\partial \sigma}. \quad (3.43)$$

Это обеспечивает точную и стабильную передачу градиентов к параметрам формы ядра и позволяет эффективно обучать слой методом обратного распространения ошибки.

Полученные результаты показали, что внедрение GaborConv на первом уровне U-Net позволяет существенно повысить чувствительность модели к тонким и направленным структурам. Это особенно важно при работе с медицинскими изображениями, где ключевые признаки могут быть выражены через локальные, слабо контрастные и вытянутые формы. В рамках данной работы отмечено улучшение полноты сегментации сосудов и снижение количества пропусков микроаневризм и кровоизлияний на фундус-снимках.

3.7.4 Метрики оценки качества

В данной задаче в качестве метрик оценки качества сегментации используются коэффициент пересечения с объединением (Intersection over Union, IoU) и коэффициент сходства Dice (Dice Similarity Coefficient, DSC), математическая формализация которых приведена ранее. Эти метрики являются наиболее широко применяемыми в задачах семантической медицинской сегментации, поскольку позволяют количественно оценить степень совпадения предсказанных и истинных масок на уровне пикселей. Метрика IoU демонстрирует устойчивость к ложноположительным и ложноотрицательным предсказаниям за счёт учёта области объединения, в то время как коэффициент Dice обладает повышенной чувствительностью к совпадению форм и границ объектов, что особенно важно при анализе мелких или протяжённых структур. Совместное использование обеих метрик позволяет объективно оценить как точность локализации, так и полноту выявления сегментируемых объектов.

3.7.5 Обучение, валидация, результаты

Проведено сравнение двух архитектур: базовой U-Net и модифицированной Gabor-U-Net, в которой первый свёрточный блок энкодера заменён на обучаемый слой с ядрами, параметризуемыми функцией Габора. Такая модификация позволяет ввести в модель направленную чувствительность к структурам определённой ориентации и пространственной частоты, что особенно актуально при анализе изображений глазного дна с вытянутыми, линейными и текстурными образованиями.

Для обучения использовался объединённый корпус изображений, составленный из трёх открытых медицинских датасетов: IDRiD, FGADR и MAPLES-DR. Эти наборы содержат фундус-снимки с пиксельной разметкой биомаркеров диабетической ретинопатии. Объединённый датасет приведён к единому формату с выделением пяти целевых классов: микроаневризмы (MA), кровоизлияния (HEM), твёрдые экссудаты (EX), мягкие экссудаты (SE) и неоваскуляризация (NV). Изображения были нормализованы по размеру и освещению.

щённости, маски унифицированы по таксономии. При обучении использовалось взвешивание потерь с учётом дисбаланса классов.

В качестве исходной модели использовалась предобученная U-Net. Такая инициализация позволила повысить стабильность обучения и ускорить сходимость.

Функция потерь, использованная при дообучении обеих моделей, представляет собой сумму функции потерь Дайса и кросс-энтропии:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{CE}} + \beta \mathcal{L}_{\text{Dice}}, \quad (3.44)$$

где \mathcal{L}_{CE} – многоклассовая кросс-энтропия, $\mathcal{L}_{\text{Dice}}$ – усреднённая функции потерь Дайса по всем классам. В данной работе использовались равные веса: $\alpha = \beta = 1$.

Кросс-энтропия вычисляется как:

$$\mathcal{L}_{\text{CE}} = - \sum_{c=1}^C \sum_{i=1}^N y_{ic} \log p_{ic}, \quad (3.45)$$

где C – количество классов, N – число пикселей, y_{ic} – истинная метка принадлежности пикселя i классу c , p_{ic} – вероятность, предсказанная моделью для этого класса.

Функция потерь Дайса для отдельного класса имеет вид:

$$\mathcal{L}_{\text{Dice}}^{(c)} = 1 - \frac{2 \sum_{i=1}^N p_{ic} y_{ic} + \varepsilon}{\sum_{i=1}^N p_{ic}^2 + \sum_{i=1}^N y_{ic}^2 + \varepsilon}, \quad (3.46)$$

а итоговая рассчитывается как:

$$\mathcal{L}_{\text{Dice}} = \frac{1}{C} \sum_{c=1}^C \mathcal{L}_{\text{Dice}}^{(c)}, \quad (3.47)$$

где ε – малое положительное число для обеспечения численной устойчивости.

Модели дообучались с использованием оптимизатора Adam, пакетной нормализации (BatchNorm), размера входных патчей 512 пикселей и контроля сходимости по валидационным метрикам. Для оценки качества применялась 5-кратная стратифицированная кросс-валидация.

В таблицах 16 и 17 приведены результаты оценки качества сегментации.

Результаты показывают, что модифицированная архитектура превосходит базовую U-Net по всем целевым метрикам на всех классах биомаркеров. Прирост составляет от 1.3 до 3.0 процентных пунктов, что указывает на повышение

Таблица 16 — Результаты по метрике IoU (%), \pm стандартное отклонение

Класс	U-Net	Gabor-U-Net	Прирост
MA	85.6 \pm 1.6	88.1\pm1.5	+2.5
HEM	87.0 \pm 1.4	90.0\pm1.3	+3.0
EX	89.1 \pm 1.2	91.6\pm1.1	+2.5
SE	86.2 \pm 1.8	88.8\pm1.6	+2.6
NV	85.9 \pm 1.9	88.3\pm1.8	+2.4

Таблица 17 — Результаты по метрике Dice (%), \pm стандартное отклонение

Класс	U-Net	Gabor-U-Net	Прирост
MA	89.0 \pm 1.3	91.4\pm1.0	+2.4
HEM	90.2 \pm 1.1	92.0\pm0.9	+1.8
EX	90.7 \pm 1.0	92.0\pm0.8	+1.3
SE	89.3 \pm 1.4	91.0\pm1.1	+1.7
NV	89.0 \pm 1.6	90.9\pm1.3	+1.9

как точности модели. На рис. 3.8 представлен пример сегментации биомаркеров на фундус-снимке с помощью модифицированной архитектуры.

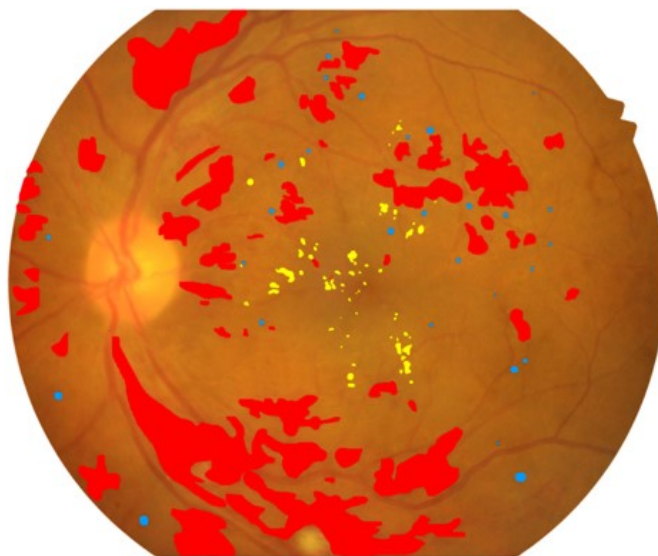


Рисунок 3.8 — Пример сегментации фундус-снимка

Наибольшее улучшение достигнуто для классов кровоизлияний и неоваскуляризации. Эти структуры обладают направленной или ветвящейся морфологией, которая хорошо соответствует чувствительности габоровских фильтров к ориентации и ширине текстур. Для классов микроаневризм и экссудатов улучшения оказались умеренными, что объясняется их изотропной, пятнистой формой. Тем не менее, даже в этих случаях направленная фильтрация улучшает контрастирование границ и снижает влияние фоновых искажений. Таким образом, внедрение обучаемого габоровского блока в структуру U-Net даёт количественные и качественные преимущества при сегментации биомаркеров диабетической ретинопатии на фундус-снимках.

3.8 Метод объяснительного ИИ

3.8.1 Объяснительный ИИ в анализе фундус-снимков

На фоне активного внедрения методов ОИИ в задачах анализа изображений лучевой диагностики (КТ, МРТ, рентген), применение подобных подходов в офтальмологических задачах на основе фундус-снимков остаётся фрагментарным. Несмотря на широкое использование СНС для автоматической классификации ДР, интерпретируемость таких моделей редко рассматривается как обязательное требование. Это затрудняет клиническое применение, особенно в условиях, где необходимо обоснование решений, принимаемых ИИ-моделью. В систематическом обзоре [151], охватывающем 46 оригинальных публикаций, лишь 18 посвящены задачам, связанным с ДР, и все они ограничиваются исключительно задачами классификации изображений, без перехода к детекции или сегментации микропатологических признаков.

Наиболее часто используемыми инструментами ОИИ в работах по классификации ДР являются методы визуализации важности признаков на основе активационных карт: CAM (Class Activation Mapping) [207], Grad-CAM [208], а также attention-механизмы и модели с множественным представлением экземпляров (multiple-instance learning, MIL). CAM-подобные методы отображают дискриминативные области изображения, ассоциированные с конкретным

классом, через активации свёрточных слоёв. Их популярность объясняется простотой реализации и совместимостью с типичными СНС-архитектурами. В ряде работ, таких как [209; 210], Grad-CAM использовался для верификации фокусировки модели на характерных для ДР поражениях. Более сложные подходы, использующие MIL и attention-механизмы, позволяют повысить разрешающую способность карт важности и дают более интерпретируемую локализацию [211; 212]. Модель-агностичные методы, такие как LIME [213], SHAP [214] и Integrated Gradients [215], применяются реже, но демонстрируют потенциал для локального анализа вклада отдельных сегментов изображения [216].

Как показывают опубликованные исследования, ОИИ в задачах классификации ДР служит, главным образом, для качественной верификации решений модели. Основной упор делается на визуальную интерпретацию тепловых карт и проверку их клинической обоснованности. При этом количественные метрики объяснимости (fidelity, robustness, completeness, sensitivity) в большинстве работ либо отсутствуют, либо оцениваются эпизодически. Так, в работе [211] приведена метрика AUPRC для соответствия карт внимания аннотированным поражениям, а в [217] проанализирована вариабельность объяснений при изменении условий визуализации. Однако во многих публикациях карты важности демонстрируют ограниченное пространственное разрешение, тенденцию к активации неспецифичных областей (например, диск зрительного нерва) и чувствительность к параметрам съёмки. В случае использования LIME и SHAP возникают дополнительные сложности, связанные с генерацией суперпикселей и стабильностью локальных объяснений [216].

Таким образом, современные публикации по применению ОИИ в диагностике ДР на фундус-снимках демонстрируют значительный прогресс в визуализации значимых признаков, но при этом ограничиваются только задачами классификации. Перспективные направления дальнейших исследований включают: разработку формализованных метрик объяснимости; интеграцию объяснимости в архитектуру модели на этапе обучения; перенос фокуса с post-hoc интерпретаций на объяснимые архитектуры; переход к задачам сегментации и количественной локализации очагов поражения. В условиях клинического применения требуется формализация протоколов валидации объяснений, верификация с участием офтальмологов и оценка влияния объяснений на принятие врачебных решений.

3.8.2 Метод ОИИ Grad-CAM

Принцип работы метода CAM (Class Activation Mapping), основанного на линейной комбинации карт признаков последнего сверточного слоя с фиксированными весами, уже был изложен ранее (см. п. X.X). В настоящем разделе рассматривается его развитие – метод Grad-CAM (Gradient-weighted Class Activation Mapping), обеспечивающий более универсальный и гибкий способ объяснения решений сверточных нейросетей.

Grad-CAM предназначен для генерации визуально интерпретируемых тепловых карт, указывающих участки входного изображения, внёсшие наибольший вклад в предсказание модели. Метод основан на использовании градиентов логитов по активациям сверточного слоя, что позволяет количественно оценить значимость пространственных признаков, формируемых на промежуточных уровнях нейросетевой архитектуры. В отличие от CAM, Grad-CAM не требует наличия архитектурных ограничений (например, глобального усреднения по каналам) и может быть применён к широкому классу моделей СНС [208].

Математически метод формализуется следующим образом. Пусть $A^k \in \mathbb{R}^{h \times w}$ – карта активаций k -го канала выбранного сверточного слоя, а y^c – логит для целевого класса c . Тогда веса каналов определяются как средние значения градиентов логита по пространственным координатам:

$$\alpha_k^c = \frac{1}{Z} \sum_{i=1}^h \sum_{j=1}^w \frac{\partial y^c}{\partial A_{ij}^k}, \quad Z = h \cdot w. \quad (3.48)$$

На их основе строится тепловая карта:

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right), \quad (3.49)$$

которая интерпретируется как карта важности пространственных признаков, масштабируется до размеров исходного изображения и визуализируется как наложение на вход. Применение функции ReLU позволяет исключить признаки, оказывающие отрицательное влияние на логит, и сфокусироваться на положительных вкладчиках в решение модели.

Так как метод работает с уже обученной моделью и не требует дополнительного обучения, он является подходом пост-фактум объяснения (англ.

post-hoc explainability) и может быть использован для анализа решений моделей в реальных приложениях. Его архитектурная независимость, вычислительная эффективность и возможность выбора уровня представления (через выбор сверточного слоя) обеспечивают гибкость интерпретации. Тепловые карты Grad-CAM позволяют визуально локализовать наиболее значимые для предсказания признаки, проводить аудит ошибок модели, интерпретировать некорректные или нестабильные решения, а также выявлять потенциальные источники артефактов и смещений [208; 218].

3.8.3 Особенности практического применения Grad-CAM

Практическое применение метода Grad-CAM для интерпретации результатов работы СНС в данной работе рассматривается в контексте объяснения результатов семантической сегментации, выполняемой модифицированной моделью U-Net. При сегментации реализуется отображение входного изображения $I \in \mathbb{R}^{H \times W \times 3}$ в тензор логитов $Y \in \mathbb{R}^{H \times W \times C}$, где каждый пиксель $p = (x, y)$ получает вектор значений $\{y^c(p)\}$ по классам $c = 1, \dots, C$. В рамках настоящей работы используется попиксельное объяснение, то есть объяснение принадлежности конкретного пикселя p^* заданному классу c . В этом случае целевая функция формулируется как:

$$T = y^c(p^*), \quad (3.50)$$

а градиенты вычисляются по активациям выбранного сверточного слоя. Такой подход позволяет локализовать признаки, оказывающие наибольшее влияние на классификацию данного пикселя, и сопоставлять их с клинически значимыми областями. Попиксельная интерпретация особенно актуальна в медицинских задачах, где необходимо анализировать поведение модели на границах классов и в областях с потенциальными артефактами, а также обеспечивать локальную интерпретируемость и выявление ошибок.

Качество тепловых карт Grad-CAM определяется выбором слоя, по активациям которого формируются объяснения. В данной работе в качестве целевого слоя выбран последний блок энкодера U-Net. Такой выбор обусловлен тем, что он обеспечивает широкое приёмное поле и учитывает глобальный

контекст изображения, формируемые признаки обладают высоким уровнем семантической абстракции, позволяя выявлять закономерности, действительно важные для классификации, а получаемые карты отличаются устойчивостью и интерпретируемостью, что критически важно для анализа медицинских изображений. Таким образом, использование последнего блока энкодера обеспечивает баланс между локализацией областей интереса и стабильностью объяснений, что делает его оптимальным для анализа решений модели U-Net в задачах диагностики [219].

Для практической реализации метода Grad-CAM применяются специализированные фреймворки. В данной работе используется библиотека `pytorch-grad-cam`, предоставляющая полнофункциональную реализацию базового метода и его модификаций (Grad-CAM++, Score-CAM и др.), включая поддержку сегментационных моделей. Она позволяет задавать целевые слои и функции, использовать режимы сглаживания (например, `aug_smooth`, `eigen_smooth`) и получать наглядные визуализации результатов [220].

Пример получения визуального объяснения для данной задачи представлен на рис. 3.9 показаны тепловые карты, наложенные на изображения глазного дна при семантической сегментации стадий диабетической ретинопатии.

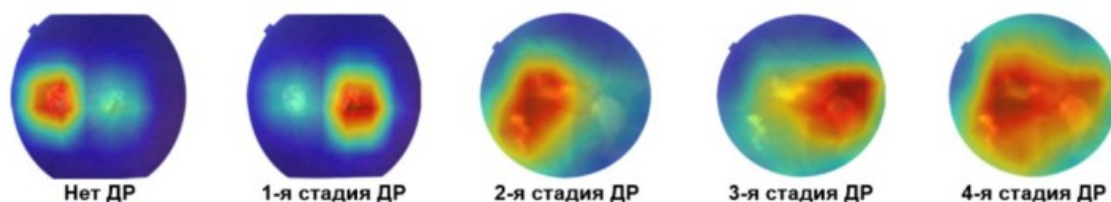


Рисунок 3.9 — Тепловые карты Grad-CAM для семантической сегментации стадий диабетической ретинопатии на фундус-снимке

Карты визуализируют области, которые вносят наибольший вклад в предсказание модели, позволяя отследить динамику изменения релевантных признаков по мере прогрессирования заболевания: от отсутствия патологий до четвёртой стадии диабетической ретинопатии. Можно отметить, что с увеличением тяжести заболевания тепловые карты охватывают всё более обширные участки сетчатки, что отражает участие в сегментации не только локальных изменений, но и диффузных патологических структур. Такой результат демонстрирует практическую значимость применения Grad-CAM для анализа и интерпретации решений нейросетей в медицинских задачах диагностики.

3.9 Метод водораздела

В задачах анализа фундус-снимков при диабетической ретинопатии часто требуется не только определить область присутствия определённого класса биомаркеров, но и выделить отдельные их экземпляры для дальнейшего подсчёта. Типичный пример – определение количества микрососудистых аномалий, геморагий, экссудатов и других структур в пределах одной анатомической области. Для этого необходимо выполнить переход от семантической сегментации, когда каждый пиксель отнесён к определённому классу, к сегментации экземпляров, при которой соприкасающиеся объекты внутри одного класса отдельно идентифицируются.

Метод водораздела (англ. watershed) представляет собой классический морфологический метод сегментации, интерпретирующий изображение как топографическую поверхность, где интенсивность или градиент соответствует высоте. Процесс затопления поверхности из локальных минимумов формирует бассейны притяжения, а линии водораздела представляют собой границы между ними [192; 221; 222].

Пусть $I : \Omega \subset \mathbb{Z}^2 \rightarrow \mathbb{R}$ – изображение, где $I(x)$ – интенсивность в точке $x \in \Omega$. Для подавления шумов используется сглаживание:

$$I_\sigma = G_\sigma * I, \quad (3.51)$$

где G_σ – гауссова свёртка. В качестве рельефа T применяют либо модуль градиента

$$T(x) = \|\nabla I_\sigma(x)\|_2, \quad (3.52)$$

либо морфологический градиент:

$$T(x) = (I_\sigma \oplus B)(x) - (I_\sigma \ominus B)(x), \quad (3.53)$$

где B – структурный элемент, \oplus и \ominus – операции дилатации и эрозии соответственно [191; 192].

Обозначим множество локальных минимумов рельефа T как $\mathcal{M}(T) = \{m_k\}$. Используя топографическую метрику, основанную на методе минимальной крутизны подъёма [223], расстояние от точки x до минимума m определяется как:

$$d_T(x, m) = \min_{\pi \in \mathcal{P}(x \rightarrow m)} \max_{y \in \pi} T(y), \quad (3.54)$$

где $\mathcal{P}(x \rightarrow m)$ – множество всех дискретных путей от x до m .

Бассейн притяжения минимума m_k определяется следующим образом:

$$CB(m_k) = \{x \in \Omega \mid d_T(x, m_k) < d_T(x, m_j), \forall j \neq k\}. \quad (3.55)$$

Линии водораздела определяются как:

$$W(T) = \Omega \setminus \bigcup_k CB(m_k), \quad (3.56)$$

что соответствует точкам, не принадлежащим ни одному бассейну [224].

Для подавления пересегментации используют маркерно-управляемый водораздел [221; 222]. Пусть $F = \bigcup_i F_i$ – множество маркеров объектов, B – маркеры фона. Для модификации рельефа T с целью ограничения сегментации только заданными маркерами выполняется фиксация минимумов в областях $F \cup B$. Обновлённая функция рельефа записывается как:

$$T' = I_{F \cup B}(T) = R_T(T - \lambda), \quad (3.57)$$

где $I_{F \cup B}(T)$ – оператор фиксации минимумов в маркерах $F \cup B$, R_T – морфологическая реконструкция по маске T , $\lambda > 0$ – параметр, подавляющий неглубокие минимумы [192].

С теоретической точки зрения, алгоритм водораздела может быть интерпретирован как построение минимального остовного леса на графе смежности пикселей, где корнями служат маркеры [225]. Эффективные реализации используют очереди приоритетов и обеспечивают квазилинейную временную сложность [222; 224].

Алгоритм водораздела реализован в библиотеке OpenCV через функцию `cv::watershed`, предназначенную для выполнения маркерно-управляемой сегментации [226]. На вход подаются: исходное изображение в формате BGR и карта маркеров в виде 32-битной целочисленной матрицы. Положительные значения маркируют отдельные объекты, ноль обозначает неизвестную зону, а результат работы включает границы, помеченные значением -1 .

В отличие от классического варианта, маркерно-управляемая реализация ограничивает рост сегментированных областей строго от заданных маркеров, что снижает пере-сегментацию и увеличивает устойчивость к шуму [221; 222]. Такой подход позволяет интегрировать априорную информацию, например, выходы модели сегментации или морфологически полученные маски.

Типовой конвейер применения алгоритма водораздела в OpenCV включает следующие этапы:

1. Предобработка и вычисление топографии:
 - сглаживание изображения (например, гауссово фильтрование);
 - вычисление градиента: операторы Sobel, Laplacian или морфологический градиент.
2. Получение бинарной маски объектов (если отсутствует готовая семантическая маска):
 - глобальное пороговое преобразование (например, метод Оtsu) или выходы сегментационной нейросети.
3. Формирование маркеров:
 - верный фон: дилатация бинарной маски;
 - верный объект: дистанционное преобразование $D = \text{dist}(M)$ и порог по доле от $\max D$; локальные максимумы D формируют маркеры объектов;
 - неизвестная зона: вычисляется как `unknown = sure_bg - sure_fg`.
4. Маркировка компонент: с использованием `connectedComponents`.
5. Вызов функции водораздела: `cv.watershed(image_bgr, markers)`; результат — карта меток с экземплярами и границами.

Такой конвейер обеспечивает воспроизводимую и модульную обработку, пригодную для задач биомедицинской визуализации, промышленного анализа и количественной морфометрии. Эффективность и качество сегментации во многом зависят от параметров порогов, глубины подавления локальных минимумов и формы структурного элемента [192; 224; 225].

Применительно к данной работе алгоритм водораздела использовался для разделения масок, полученных на этапе семантической сегментации биомаркеров диабетической ретинопатии, таких как микрогеморрагии, твёрдые экссудаты, микроаневризмы и другие (рис. ??).

Таким образом, использование метода водораздела позволило получить отдельную маску для каждого отдельного экземпляра биомаркера, относящегося к заданному классу. Это существенно расширяет возможности анализа: становится возможным выполнять не только классификацию наличия патологии, но и точный подсчёт её проявлений, а также оценку морфологических характеристик (например, площади, формы, плотности распределения). Такой

подход критически важен при построении интерпретируемых моделей диагностики и прогнозирования степени поражения сетчатки при диабетической ретинопатии.

3.10 Определение количественных характеристик

Цель данного этапа – получение количественных характеристик распределения биомаркеров диабетической ретинопатии в пределах зоны фовеа, с учётом её пространственного деления. На предыдущих шагах фовеа была сегментирована в виде окружности, а затем разделена на четыре квадранта, обозначенные как Q_1, Q_2, Q_3, Q_4 . Каждая маска семантической сегментации, соответствующая определённому классу биомаркеров (например, микроаневризмы, геморрагии, твёрдые экссудаты), также была ограничена соответствующим квадрантом. Для устранения наложения и точного подсчёта экземпляров внутри каждого квадранта применялся метод водораздела.

Для формального описания введём обозначения. Пусть $M_c(x, y)$ – бинарная маска класса c , а $Q_i \subset D$ – область i -го квадранта зоны фовеа, где $i = 1, 2, 3, 4$. Тогда маска класса в данном квадранте определяется как:

$$M_{c,i}(x, y) = M_c(x, y) \cap Q_i. \quad (3.58)$$

После применения алгоритма водораздела к $M_{c,i}$ получаем множество непересекающихся сегментов (экземпляров):

$$S_{c,i} = \{S_{c,i}^1, S_{c,i}^2, \dots, S_{c,i}^{N_{c,i}}\}, \quad (3.59)$$

где $N_{c,i}$ – число найденных экземпляров класса c в квадранте Q_i . Суммарное количество элементов класса c по всей фовеальной зоне определяется как:

$$N_c = \sum_{i=1}^4 N_{c,i}. \quad (3.60)$$

Для представления результатов удобно использовать матрицу подсчётов:

$$\mathbf{N} = (N_{c,i}), \quad c = 1, \dots, C; \quad i = 1, \dots, 4, \quad (3.61)$$

где C – количество классов биомаркеров. Каждая строка такой матрицы описывает пространственное распределение экземпляров конкретного класса по квадрантам.

Полученные величины позволяют перейти от пиксельного уровня (семантическая сегментация) к количественному анализу экземпляров. Это критически важно при построении интерпретируемых признаков, оценке локального распределения биомаркеров и сравнении пространственных паттернов патологий у разных пациентов.

3.11 Классификация на основе логических правил

Одним из ключевых этапов анализа биомаркеров ДР после выполнения сегментации является переход от количественного описания признаков к диагностической интерпретации. В данной работе этот переход реализуется через логический классификатор, основанный на формализованных клинических правилах.

Суть подхода заключается в том, чтобы преобразовать матрицу подсчитанных экземпляров биомаркеров (по классам и квадрантам) в диагностическое решение о текущей стадии ДР: непролиферативная (NPDR), препролиферативная (PPDR) или пролиферативная (PDR). Для этого используются пороговые и логические условия, полученные из официальных клинических рекомендаций Российской Федерации [139] и международных стандартов [143; 227; 228].

Для наглядности в таблице 18 приведены клинические характеристики стадий ДР согласно национальным рекомендациям.

Пусть для каждого класса биомаркеров $c \in \{MA, HEM, EX, SE, NV\}$ и квадранта Q_i , $i = 1, \dots, 4$ получены количества экземпляров $N_{c,i} \in \mathbb{N}_0$. Обозначим матрицу подсчётов

$$\mathbf{N} = (N_{c,i}) \in \mathbb{N}_0^{C \times 4}, \quad C = 5.$$

Введём пороговые параметры $\Theta = \{\theta_{MA}, \theta_{HEM}, \theta_{4Q}, \theta_{NV}\}$. Логический классификатор задаётся как функция:

$$f : \mathbb{N}_0^{C \times 4} \times \Theta \longrightarrow \{NPDR, PPDR, PDR\}.$$

Таблица 18 — Характеристика изменений на глазном дне при различных стадиях ДР [139]

Стадия ДР	Характеристика изменений на глазном дне
Непролиферативная	Микроаневризмы, мелкие интравитреальные кровоизлияния
Препролиферативная	Диагностируется при наличии хотя бы одного из признаков: – венозные деформации в 2-х и более квадрантах; – множественные ИРМА в одном и более квадрантах; – множественные геморрагии в виде пятен в 4-х квадрантах
Прролиферативная	Неоваскуляризация диска зрительного нерва и/или сетчатки, преретинальные и интравитреальные кровоизлияния (гемофтальм), фиброзная ткань в области кровоизлияний и вдоль новообразованных сосудов

Логические правила классификации стадий ДР:

- Прролиферативная ДР (PDR): наличие неоваскуляризации хотя бы в одном квадранте:

$$\exists i : N_{NV,i} \geq \theta_{NV}.$$

- Препролиферативная ДР (PPDR): множественные геморрагии и/или микроаневризмы во всех квадрантах:

$$Q_{\text{count}}(N_{HEM,i} + N_{MA,i} \geq \theta_{4Q}) = 4.$$

- Непролиферативная ДР (NPDR): наличие микроаневризм или кровоизлияний:

$$N_{MA} \geq \theta_{MA} \quad \vee \quad N_{HEM} \geq \theta_{HEM},$$

при условии, что не выполняются условия для PDR и PPDR.

Итоговое правило классификации:

$$f(\mathbf{N}; \Theta) = \begin{cases} \text{PDR}, & \text{если } \exists i : N_{NV,i} \geq \theta_{NV}, \\ \text{PPDR}, & \text{иначе если } Q_{\text{count}}(N_{HEM,i} + N_{MA,i} \geq \theta_{4Q}) = 4, \\ \text{NPDR}, & \text{иначе если } N_{MA} \geq \theta_{MA} \vee N_{HEM} \geq \theta_{HEM}. \end{cases}$$

Пороговые значения подбираются в соответствии с диагностическим критериям, представленным в клинических рекомендациях и руководствах [139; 143; 228].

3.12 Выводы к третьей главе

1. Предложен и реализован метод анализа фундус-снимков, включающий последовательность этапов: бинарная классификация изображений по признаку наличия патологий, локализация ключевых анатомических зон, сегментация экземпляров биомаркеров и финальная логическая интерпретация результатов. Метод разработан как компонент методики формирования проблемно-ориентированных коллекций данных и способен обеспечивать интеграцию в прикладные конвейеры.
2. Обоснована и практически реализована интеграция визуально-объяснимых механизмов интерпретации решений нейросетей на основе тепловых карт (Grad-CAM). Полученные визуальные объяснения позволяют проследивать причинно-следственные связи между входным изображением и результатом модели, что повышает прозрачность и воспроизводимость в условиях клинического применения.
3. Разработана модификация архитектуры сегментационной модели семейства U-Net, основанная на внедрении параметрически обучаемых фильтров Габора, ориентированных на извлечение локальных текстурных признаков и направленной градиентной структуры. Доказано, что модификация позволяет улучшить качество сегментации слабовыраженных биомаркеров на фундус-снимках.
4. Предложен подход к интерпретируемой классификации стадий ретинопатии на основе формализованных экспертных правил. Использование логического классификатора на этапе финального принятия решения

позволяет совместить преимущества глубокой нейросетевой сегментации и экспертного клинического опыта.

Заключение

Разработана методика формирования специализированных коллекций данных для автоматического анализа ретинальных изображений. Методика опирается на нейросетевую оценку качества, использование объяснимого ИИ для верификации релевантности признаков, интеграцию моделей с визуальными промптами для ускоренной предварительной разметки и перекрёстные экспертные аннотации. Такая организация конвейера повышает воспроизводимость экспериментов, снижает трудоёмкость ручной разметки и улучшает согласованность данных.

Предложен метод автоматического анализа ретинальных изображений, сочетающий сегментацию экземпляров, формализованные логические правила экспертных знаний и средства объяснимого ИИ. Переход к экземплярной постановке позволил выделять мелкие и пересекающиеся структуры, а последующая формализация знаний – агрегировать количественные показатели в диагностически значимые заключения. Это привело к повышению точности решения клинических задач распознавания патологий сетчатки.

Разработан метод поиска и сегментации анатомических зон сетчатки на фундус-снимках на основе анализа бинаризованных гистограмм цветового канала. Подход характеризуется нечувствительностью к артефактам и отсутствием необходимости в предварительном обучении, что расширяет его применимость при вариативности данных и уменьшает требования к объёму размеченных выборок.

Достижение поставленной цели – разработка интерпретируемых и устойчивых методов анализа ретинальных изображений на основе сегментации биомаркеров, экспертных правил и средств глубокого обучения. Обеспечено за счёт комплексного подхода, включающего формирование качественных датасетов, реализацию нейросетевых архитектур с объяснимым поведением и построение логического вывода на основе клинических знаний. Предложенные решения ориентированы на практическое применение в задачах поддержки принятия врачебных решений и соответствуют актуальным требованиям к точности, интерпретируемости и воспроизводимости медицинских ИИ-систем.

Список рисунков

1.1	Блок-схема предлагаемой методики полуавтоматического формирования коллекции данных	19
1.2	Результат предварительной разметки с помощью базовой модели . . .	29
1.3	Пример изображений коллекции данных биомаркеров ДМО	31
2.1	Блок-схема предлагаемого метода анализа ОКТ-снимков	41
2.2	Пример получения визуального объяснения методом CAM	47
2.3	Архитектура модели YOLOv8	49
2.4	Архитектура BiFPN	50
2.5	Структура механизма внимания Coordinate Attention	52
2.6	Пример сегментации изображений с помощью YOLOv8-BiFPN-CA . .	56
3.1	а) Схема процесса офтальмоскопии; б) Фундус-снимок	68
3.2	Визуализация биомаркеров диабетической ретинопатии на фундус-снимке: микроаневризмы, кровоизлияния, экссудаты	71
3.3	Блок-схема предлагаемого метода анализа фундус-изображений . . .	77
3.4	Матрица ошибок VGG16	89
3.5	Пример результата автоматического поиска центра зоны фове и оптического диска: центр ОД – красный, зоны фовеа – синий	92
3.6	Пример разделения изображения на квадранты относительно центра фовеа	95
3.7	Схема архитектуры модели U-Net [197]	98
3.8	Пример сегментации фундус-снимка	105
3.9	Тепловые карты Grad-CAM для семантической сегментации стадий диабетической ретинопатии на фундус-снимке	110

Список таблиц

1	Классификация ОКТ—систем	35
2	Ключевые методы визуализации глаза и их диагностические ниши .	64
3	Обзор исследований по теме сегментации биомаркеров ДМО	65
4	Охват биомаркеров в исследованиях	66
6	Результаты классификации на датасете OCTDL	66
7	Сравнение версий YOLOv8-seg по количеству параметров и времени инференса	67
8	Метрики сегментации на YOLOv8 и YOLOv8-BiFPN-CA	67
9	Обзор исследований по теме сегментации биомаркеров ДР	73
10	Количественные характеристики и сбалансированность коллекций данных	80
11	Источники классов в объединённой коллекции	80
12	Размеры изображений до и после приведения к единому масштабу .	81
13	Размерности слоёв и количество параметров VGG16	86
14	Результаты VGG16 на тестовой выборке	88
15	Размерности слоёв и количество параметров модели U-Net	97
16	Результаты по метрике IoU (%), \pm стандартное отклонение	105
17	Результаты по метрике Dice (%), \pm стандартное отклонение	105
18	Характеристика изменений на глазном дне при различных стадиях ДР [139]	116

Список литературы

1. *Шестакова М. В., Аметов А. С., Дедов И. И.* Состояние проблемы сахарного диабета в Российской Федерации: эпидемиология и стратегия действий // Сахарный диабет. — 2023. — Т. 26, № 1. — С. 4—12. — DOI: 10.14341/DM202312.
2. *Президент Российской Федерации.* Национальная стратегия развития искусственного интеллекта на период до 2030 года. — 2019. — Указ Президента РФ от 10.10.2019 № 490. <http://www.kremlin.ru/acts/bank/44731>.
3. *Volkov E. N., Averkin A. N.* Gradient-based explainable artificial intelligence methods for eye disease classification // 2023 IV International Conference on Neural Networks and Neurotechnologies (NeuroNT). — IEEE. 2023. — С. 6—9. — DOI: [10.1109/NeuroNT58640.2023.10175855](https://doi.org/10.1109/NeuroNT58640.2023.10175855).
4. *Volkov E. N., Averkin A. N.* Explainable artificial intelligence in clinical decision support systems // 2023 IV international conference on neural networks and neurotechnologies (NeuroNT). — IEEE. 2023. — С. 3—6. — DOI: [10.1109/NeuroNT58640.2023.10175852](https://doi.org/10.1109/NeuroNT58640.2023.10175852).
5. *Volkov E. N.* Application of Spatial Attention Module in Convolutional Neural Network for OCT Images Analysis // 2024 V International Conference on Neural Networks and Neurotechnologies (NeuroNT). — IEEE. 2024. — С. 1—4. — DOI: [10.1109/NeuroNT62606.2024.10585353](https://doi.org/10.1109/NeuroNT62606.2024.10585353).
6. *Volkov E., Averkin A.* Modification of U-Net Architecture Based on Gabor Filters with Trainable Parameters to Improve Biomarker Segmentation on Oct Images // 2025 VI International Conference on Neural Networks and Neurotechnologies (NeuroNT). — IEEE. 2025. — С. 66—68. — DOI: [10.1109/NeuroNT66873.2025.11049967](https://doi.org/10.1109/NeuroNT66873.2025.11049967).
7. *Volkov E., Linich A., Averkin A.* Active Contour Method Modification for Diabetic Macular Edema Area Quantification in OCT Images // 2025 27th International Conference on Digital Signal Processing and its Applications (DSPA). — IEEE. 2025. — С. 1—4. — DOI: [10.1109/DSPA64310.2025.10977883](https://doi.org/10.1109/DSPA64310.2025.10977883).

8. Волков Е. Н., Аверкин А. Н., Ярушев С. А. Интеграция механизма внимания в свёрточную нейронную сеть для улучшения результатов классификации офтальмологических заболеваний // Мягкие измерения и вычисления. — 2024. — Т. 80, № 7. — С. 105—114. — DOI: [10.36871/2618-9976.2024.07.009](https://doi.org/10.36871/2618-9976.2024.07.009).
9. Аверкин А. Н., Волков Е. Н., Ярушев С. А. Гибридный метод анализа изображений на основе технологий искусственного интеллекта и нечётких множеств // Известия Российской академии наук. Теория и системы управления. — 2025. — № 3. — С. 99—112. — DOI: [10.31857/S0002338825030103](https://doi.org/10.31857/S0002338825030103).
10. Averkin A., Volkov E., Yarushev S. Possibilities of Application of Neuro-Fuzzy Networks for Ophthalmologic Image Classification // Pattern Recognition and Image Analysis. — 2024. — Т. 34, № 3. — С. 610—616. — DOI: [10.1134/S1054661824700421](https://doi.org/10.1134/S1054661824700421).
11. Averkin A. N., Volkov E. N., Yarushev S. A. Hybrid Method of Image Analysis Based on Artificial Intelligence Technologies and Fuzzy Sets // Journal of Computer and Systems Sciences International. — 2025. — Т. 64, № 3. — С. 460—473. — DOI: [10.1134/S1064230725700431](https://doi.org/10.1134/S1064230725700431).
12. Volkov E., Averkin A. Fundus image quality assessment: a brief review of techniques // 2024 International Conference on Information Processes and Systems Development and Quality Assurance (IPS). — IEEE. 2024. — С. 50—54. — DOI: [10.1109/IPS62349.2024.10499480](https://doi.org/10.1109/IPS62349.2024.10499480).
13. Волков Е. Н., Аверкин А. Н. Оценка возможностей применения метода объяснительного ИИ Grad-CAM в задаче нейросетевой классификации стадий диабетической ретинопатии на снимках фундус-камеры // Мягкие измерения и вычисления. — 2024. — Т. 85, № 12. — С. 72—81. — DOI: [10.36871/2618-9976.2024.12.007](https://doi.org/10.36871/2618-9976.2024.12.007).
14. Volkov E. N., Averkin A. N. Hybrid Explainable Framework for Diabetic Retinopathy Classification from Fundus Images // 2024 XXVII International Conference on Soft Computing and Measurements (SCM). — IEEE. 2024. — С. 428—431. — DOI: [10.1109/SCM62608.2024.10554254](https://doi.org/10.1109/SCM62608.2024.10554254).

15. *Volkov E. N., Averkin A. N.* Possibilities of explainable artificial intelligence for glaucoma detection using the lime method as an example // 2023 XXVI International Conference on Soft Computing and Measurements (SCM). — IEEE. 2023. — С. 130—133. — DOI: [10.1109/SCM58628.2023.10159038](https://doi.org/10.1109/SCM58628.2023.10159038).
16. *Volkov E. N., Averkin A. N.* Explainable artificial intelligence in medical image analysis: State of the art and prospects // 2023 XXVI International Conference on Soft Computing and Measurements (SCM). — IEEE. 2023. — С. 134—137. — DOI: [10.1109/SCM58628.2023.10159033](https://doi.org/10.1109/SCM58628.2023.10159033).
17. *Аверкин А. Н., Волков Е. Н., Ярушев С. А.* Объяснительный искусственный интеллект в анализе цифровых изображений на основе нейронных сетей глубокого обучения // Известия Российской академии наук. Теория и системы управления. — 2024. — № 1. — С. 150—178. — DOI: [10.31857/S0002338824010122](https://doi.org/10.31857/S0002338824010122).
18. *Averkin A. N., Volkov E. N., Yarushev S. A.* Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis // Journal of Computer and Systems Sciences International. — 2024. — Т. 63, № 1. — С. 175—203. — DOI: [10.1134/S1064230724700138](https://doi.org/10.1134/S1064230724700138).
19. *Volkov E. N., Averkin A. N.* Local Explanations for Large Language Models: a Brief Review of Methods // 2024 XXVII International Conference on Soft Computing and Measurements (SCM). — IEEE. 2024. — С. 189—192. — DOI: [10.1109/SCM62608.2024.10554222](https://doi.org/10.1109/SCM62608.2024.10554222).
20. *Averkin A., Volkov E.* Explainable AI 2.0: Conceptual Shifts and New Requirements // 2025 XXVIII International Conference on Soft Computing and Measurements (SCM). — IEEE. 2025. — С. 171—173. — DOI: [10.1109/SCM66446.2025.11060242](https://doi.org/10.1109/SCM66446.2025.11060242).
21. *Volkov E., Sechin V., Averkin A.* Visual-Language Model Fine-Tuning via LoRA for Structured Medical Reports Generating for Lung X-Ray Skans // 2025 XXVIII International Conference on Soft Computing and Measurements (SCM). — IEEE. 2025. — С. 438—442. — DOI: [10.1109/SCM66446.2025.11060140](https://doi.org/10.1109/SCM66446.2025.11060140).
22. Перспективы применения объяснительного искусственного интеллекта в системах поддержки принятия врачебных решений на основе искусственных нейросетей / С. А. Ярушев [и др.] // Мягкие измерения и

- вычисления. — 2023. — Т. 70, № 9—1. — С. 56—67. — DOI: [10.36871/2618-9976.2023.09.006](https://doi.org/10.36871/2618-9976.2023.09.006).
23. *Российская Федерация*. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных». — 2006. — Дата обращения: 29.01.2025.
 24. *Минздрав России*. Приказ от 28.04.2011 № 364 «О создании единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ)». — 2011. — Дата обращения: 29.01.2025.
 25. *Росстандарт*. ГОСТ Р 55036–2012. Информатизация здоровья. Псевдонимизация. — 2012. — Дата обращения: 29.01.2025.
 26. *Роскомнадзор*. Приказ от 05.09.2013 № 996 «Об утверждении требований и методов по обезличиванию персональных данных». — 2013. — Дата обращения: 29.01.2025.
 27. *Стародубов В. И., Сидоров К. В., Зарубина Т. В.* Нормативно-справочная информация: принципы построения и перспективы развития на этапе создания единого цифрового контура в здравоохранении // Сибирский журнал клинической и экспериментальной медицины. — 2020. — № 4. — С. 14—21. — DOI: [10.29001/2073-8552-2020-15-4-14-21](https://doi.org/10.29001/2073-8552-2020-15-4-14-21).
 28. *Столбов А. П.* О стандартизации методов псевдонимизации персональных данных в здравоохранении // Проблемы стандартизации в здравоохранении. — 2017. — № 9/10. — С. 25—36. — DOI: [10.24075/brsj.2017.09-10.004](https://doi.org/10.24075/brsj.2017.09-10.004).
 29. *Росстандарт*. ГОСТ Р ИСО 12052–2009. Цифровые изображения и связь в медицине (DICOM). — 2009. — Дата обращения: 29.01.2025.
 30. A survey of deep learning in medical image analysis / G. Litjens [и др.] // Medical Image Analysis. — 2017. — Т. 42. — С. 60—88. — DOI: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).
 31. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) / B. H. Menze [и др.] // IEEE Transactions on Medical Imaging. — 2015. — Т. 34, № 10. — С. 1993—2024. — DOI: [10.1109/TMI.2014.2377694](https://doi.org/10.1109/TMI.2014.2377694).
 32. *Settles B.* Active learning literature survey : тех. отч. / University of Wisconsin–Madison. — 2009. — № 1648. — Дата обращения: 29.01.2025.

33. MIMIC-CXR: A large publicly available database of labeled chest radiographs / A. E. W. Johnson [и др.] // arXiv preprint arXiv:1901.07042. — 2019. — Дата обращения: 29.01.2025.
34. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison / J. Irvin [и др.] // Proceedings of the AAAI Conference on Artificial Intelligence. — 2019. — Дата обращения: 29.01.2025.
35. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases / X. Wang [и др.] // Proceedings of the IEEE CVPR. — 2017. — С. 3462–3471. — DOI: [10.1109/CVPR.2017.369](https://doi.org/10.1109/CVPR.2017.369).
36. *Northcutt C. G., Jiang L., Chuang I. L.* Confident learning: Estimating uncertainty and identifying mislabeled examples // Journal of Machine Learning Research. — 2021. — Т. 22. — С. 1–52. — Дата обращения: 29.01.2025.
37. NegBio: a high-performance tool for negation and uncertainty detection in radiology reports / Y. Peng [и др.] // Journal of the American Medical Informatics Association. — 2018. — Т. 25, № 10. — С. 1214–1223. — DOI: [10.1093/jamia/ocy114](https://doi.org/10.1093/jamia/ocy114).
38. Synthetic data augmentation using GAN for improved liver lesion classification on CT / M. Frid-Adar [и др.] // PLOS ONE. — 2018. — Т. 13, № 9. — e0202844. — DOI: [10.1371/journal.pone.0202844](https://doi.org/10.1371/journal.pone.0202844).
39. GANs in medical image analysis: From noise to application / S. Kazeminiya [и др.] // Medical Image Analysis. — 2021. — Т. 70. — С. 101993. — DOI: [10.1016/j.media.2020.101993](https://doi.org/10.1016/j.media.2020.101993). — Дата обращения: 29.01.2025.
40. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository / K. Clark [и др.] // Journal of Digital Imaging. — 2013. — Т. 26, № 6. — С. 1045–1057. — DOI: [10.1007/s10278-013-9622-7](https://doi.org/10.1007/s10278-013-9622-7).
41. A large annotated medical image dataset for the development and evaluation of segmentation algorithms (Medical Segmentation Decathlon) / A. L. Simpson [и др.] // arXiv:1902.09063. — 2019. — Дата обращения: 29.01.2025.
42. Why rankings of biomedical image analysis competitions are misleading—and what to do about it / L. Maier-Hein [и др.] // Nature Communications. — 2018. — Т. 9. — С. 5217. — DOI: [10.1038/s41467-018-05265-2](https://doi.org/10.1038/s41467-018-05265-2).

43. *Tjoa E., Guan C.* A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI // *IEEE Transactions on Neural Networks and Learning Systems*. — 2021. — Т. 32, № 11. — С. 4793–4813. — DOI: [10.1109/TNNLS.2020.3027314](https://doi.org/10.1109/TNNLS.2020.3027314).
44. Harmonization of cortical thickness measurements across scanners and sites / J.-P. Fortin [и др.] // *NeuroImage*. — 2018. — Т. 167. — С. 104–120. — DOI: [10.1016/j.neuroimage.2017.11.024](https://doi.org/10.1016/j.neuroimage.2017.11.024).
45. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study / J. R. Zech [и др.] // *PLOS Medicine*. — 2018. — Т. 15, № 11. — e1002683. — DOI: [10.1371/journal.pmed.1002683](https://doi.org/10.1371/journal.pmed.1002683).
46. Datasheets for datasets / T. Gebru [и др.] // *Communications of the ACM*. — 2021. — Т. 64, № 12. — С. 86–92. — DOI: [10.1145/3458723](https://doi.org/10.1145/3458723).
47. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning / D. S. Kermany [и др.] // *Cell*. — 2018. — Т. 172, № 5. — 1122–1131.e9. — DOI: [10.1016/j.cell.2018.02.010](https://doi.org/10.1016/j.cell.2018.02.010).
48. GABNet: global attention block for retinal OCT disease classification / X. Huang [и др.] // *Frontiers in Neuroscience*. — 2023. — Т. 17. — С. 1143422. — DOI: [10.3389/fnins.2023.1143422](https://doi.org/10.3389/fnins.2023.1143422).
49. OCTID: Optical Coherence Tomography Image Database / P. Gholami [и др.] // *arXiv preprint arXiv:1812.07056*. — 2018.
50. OCTDL: Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods / M. Kulyabin [и др.] // *Scientific Data*. — 2024. — Т. 11. — С. 365. — DOI: [10.1038/s41597-024-03182-7](https://doi.org/10.1038/s41597-024-03182-7).
51. *Гонсалес Рафаэль и Вудс Р.* Цифровая обработка изображений. — 4-е изд. — Москва : Техносфера, 2019.
52. *Russ J. C.* The Image Processing Handbook. — 7-е изд. — CRC Press, 2021. — DOI: [10.1201/9781003044543](https://doi.org/10.1201/9781003044543).
53. *Schmitt J. M., Xiang S.-H., Yung K. M.* Speckle in optical coherence tomography // *Journal of Biomedical Optics*. — 1999. — Т. 4, № 1. — С. 95–105. — DOI: [10.1117/1.429925](https://doi.org/10.1117/1.429925).

54. Optical Coherence Tomography Technology and Quality Improvement Methods for OCT Images of Skin: A Short Review / S. Adabi [и др.] // Biomedical Engineering and Computational Biology. — 2017. — Т. 8. — С. 1—5. — DOI: [10.1177/1179597217713475](https://doi.org/10.1177/1179597217713475).
55. Automatic segmentation of retinal layers in OCT images with intermediate AMD using U-Net and DexiNed / J. Sousa [и др.] // PLoS One. — 2021. — Т. 16, № 5. — e0251591. — DOI: [10.1371/journal.pone.0251591](https://doi.org/10.1371/journal.pone.0251591).
56. HDB-Net: hierarchical dual-branch network for retinal layer segmentation in diseased OCT images / Y. Chen [и др.] // Frontiers in Neuroscience. — 2024. — Т. 18. — С. 1410208. — DOI: [10.3389/fnins.2024.1410208](https://doi.org/10.3389/fnins.2024.1410208).
57. A comparison of deep learning U-Net architectures for retinal layer segmentation in OCT / J. Kugelman [и др.] // Scientific Reports. — 2022. — Т. 12. — С. 14340. — DOI: [10.1038/s41598-022-18155-4](https://doi.org/10.1038/s41598-022-18155-4).
58. Explaining Explanations: An Overview of Interpretability of Machine Learning / L. H. Gilpin [и др.] // Proceedings of the IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). — 2018. — С. 80—89. — DOI: [10.1109/DSAA.2018.00018](https://doi.org/10.1109/DSAA.2018.00018).
59. *Adadi A., Berrada M.* Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) // IEEE Access. — 2018. — Т. 6. — С. 52138—52160. — DOI: [10.1109/ACCESS.2018.2870052](https://doi.org/10.1109/ACCESS.2018.2870052).
60. *Saeed W., Omlin C.* Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities // Knowledge-Based Systems. — 2023. — Т. 263. — С. 110273. — DOI: [10.1016/j.knosys.2023.110273](https://doi.org/10.1016/j.knosys.2023.110273).
61. Explainable Artificial Intelligence (XAI) in Deep Learning-Based Medical Image Analysis / B. H. M. Van der Velden [и др.] // Medical Image Analysis. — 2022. — Т. 79. — С. 102470. — DOI: [10.1016/j.media.2022.102470](https://doi.org/10.1016/j.media.2022.102470).
62. Learning Deep Features for Discriminative Localization / B. Zhou [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — С. 2921—2929. — DOI: [10.1109/CVPR.2016.319](https://doi.org/10.1109/CVPR.2016.319).

63. Is Object Localization for Free? Weakly-Supervised Learning with Convolutional Neural Networks / M. Oquab [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2015. — С. 685—694. — DOI: [10.1109/CVPR.2015.7298630](https://doi.org/10.1109/CVPR.2015.7298630).
64. Zeiler M. D., Fergus R. Visualizing and Understanding Convolutional Networks // Computer Vision – ECCV 2014. — Springer, 2014. — С. 818—833. — DOI: [10.1007/978-3-319-10590-1_53](https://doi.org/10.1007/978-3-319-10590-1_53).
65. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization / R. R. Selvaraju [и др.] // Proceedings of the IEEE International Conference on Computer Vision (ICCV). — 2017. — С. 618—626. — DOI: [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74).
66. Simonyan K., Vedaldi A., Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps // arXiv preprint arXiv:1312.6034. — 2013.
67. On Pixel-Wise Explanations by Layer-Wise Relevance Propagation / S. Bach [и др.] // PLoS ONE. — 2015. — Т. 10, № 7. — e0130140. — DOI: [10.1371/journal.pone.0130140](https://doi.org/10.1371/journal.pone.0130140).
68. Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition / G. Montavon [и др.] // Pattern Recognition. — 2017. — Т. 65. — С. 211—222. — DOI: [10.1016/j.patcog.2016.11.008](https://doi.org/10.1016/j.patcog.2016.11.008).
69. Doshi-Velez F., Kim B. Towards a Rigorous Science of Interpretable Machine Learning // arXiv preprint arXiv:1702.08608. — 2017.
70. Segment anything in medical images / J. Ma, Y. He, F. Li [и др.] // Nature Communications. — 2024. — Т. 15. — С. 654. — DOI: [10.1038/s41467-024-44824-z](https://doi.org/10.1038/s41467-024-44824-z).
71. Segment Anything / A. Kirillov, E. Mintun, N. Ravi [и др.] // arXiv preprint arXiv:2304.02643. — 2023.
72. Improving the annotation process in computational pathology: a pilot study with manual and semi-automated approaches / G. Cazzaniga, F. Del Carro, A. Eccher [и др.] // Journal of Imaging Informatics in Medicine. — 2025. — Т. 38, № 2. — С. 1112—1119. — DOI: [10.1007/s10278-024-01248-x](https://doi.org/10.1007/s10278-024-01248-x).

73. *Galbusera F., Cina A.* Image annotation and curation in radiology: an overview for machine learning practitioners // *European Radiology Experimental*. — 2024. — Т. 8, № 1. — С. 11. — DOI: [10.1186/s41747-023-00408-y](https://doi.org/10.1186/s41747-023-00408-y).
74. *Optical coherence tomography* / D. Huang [и др.] // *Science*. — 1991. — Т. 254, № 5035. — С. 1178—1181. — DOI: [10.1126/science.1957169](https://doi.org/10.1126/science.1957169).
75. *Chopra R., Wagner S. K., Keane P. A.* Optical coherence tomography in the 2020s—outside the eye clinic // *Nature*. — 2021. — Т. 35. — С. 236—243. — DOI: [10.1038/s41433-020-01263-6](https://doi.org/10.1038/s41433-020-01263-6).
76. *Naholo E. S., Musa M. J., Zeppieri M.* Optical Coherence Tomography. — 2024. — StatPearls [Internet]. Treasure Island (FL). <https://www.ncbi.nlm.nih.gov/books/NBK606122/>.
77. *National Institute for Health and Care Excellence (NICE)*. Age-related macular degeneration: diagnosis and management. — London : National Institute for Health, Care Excellence (NICE), 2018. — Гл. 4. — (NICE Guideline, No. 82). — Summary of recommendations.
78. *Optical Coherence Tomography (OCT): Principle and Technical Realization* / S. Aumann [и др.] // *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics* / под ред. J. F. Bille. — Cham : Springer International Publishing, 2019. — С. 59—85. — DOI: [10.1007/978-3-030-16638-0_3](https://doi.org/10.1007/978-3-030-16638-0_3).
79. *Zeppieri M., Naholo E. S., Musa M. J.* Optical Coherence Tomography (OCT): A Brief Look at the Uses and Technological Evolution of Ophthalmology // *Medicina (Kaunas)*. — 2023. — Т. 59. — DOI: [10.3390/medicina59122114](https://doi.org/10.3390/medicina59122114).
80. *American Academy of Ophthalmology*. Primary Open-Angle Glaucoma Preferred Practice Pattern® Guidelines. — San Francisco, CA, 2020. — Ophthalmology Primary Open-Angle Glaucoma PPP Panel; Quality Eye Care High-Quality Care for All. <https://www.aao.org/guidelines>.
81. *Xia F., Hua R.* The Latest Updates in Swept-Source Optical Coherence Tomography Angiography // *Diagnostics (Basel, Switzerland)*. — 2023. — Т. 14. — DOI: [10.3390/diagnostics14010047](https://doi.org/10.3390/diagnostics14010047).

82. *Haner N. U., Dysli C., Munk M. R.* Imaging in retinal vascular disease: A review // *Clinical & Experimental Ophthalmology*. — 2023. — Т. 51, № 3. — С. 217–228. — DOI: [10.1111/ceo.14203](https://doi.org/10.1111/ceo.14203).
83. *Mohankumar A., Gurnani B.* Scanning Laser Ophthalmoscope. — 2023. — StatPearls [Internet]. <https://www.ncbi.nlm.nih.gov/sites/books/NBK587354/>.
84. MR-EYE: High-Resolution MRI of the Human Eye and Orbit at Ultrahigh Field (7T) / R. K. Glarin [и др.] // *Magnetic Resonance Imaging Clinics of North America*. — 2021. — Т. 29, № 1. — С. 103–116. — DOI: [10.1016/j.mric.2020.09.004](https://doi.org/10.1016/j.mric.2020.09.004).
85. *Демидова Т. Ю., Кожевников А. А.* Диабетическая ретинопатия: история, современные подходы к ведению, перспективы профилактики и лечения // *Сахарный диабет*. — 2020. — Т. 23, № 1. — С. 95–105. — DOI: [10.14341/DM10273](https://doi.org/10.14341/DM10273).
86. Disorganization of the retinal inner layers as a predictor of visual acuity in eyes with center-involved diabetic macular edema / J. K. Sun, S. H. Radwan, M. K. Durbin [и др.] // *Ophthalmology*. — 2014. — Т. 121, № 11. — С. 2153–2161. — DOI: [10.1016/j.ophtha.2014.04.034](https://doi.org/10.1016/j.ophtha.2014.04.034).
87. The International Vitreomacular Traction Study Group classification of vitreomacular adhesion, traction, and macular hole / J. S. Duker, P. K. Kaiser, S. Binder [и др.] // *Ophthalmology*. — 2013. — Т. 120, № 12. — С. 2611–2619. — DOI: [10.1016/j.ophtha.2013.07.042](https://doi.org/10.1016/j.ophtha.2013.07.042).
88. Association of pathomorphology in diabetic macular edema with visual acuity following intravitreal ranibizumab / T. Murakami, K. Nishijima, T. Akagi [и др.] // *Retina*. — 2012. — Т. 32, № 7. — С. 1396–1407. — DOI: [10.1097/IAE.0b013e31824d2d35](https://doi.org/10.1097/IAE.0b013e31824d2d35).
89. Hyperreflective dots: A new spectral-domain OCT biomarker in retinal diseases / G. Coscas, U. De Benedetto, F. Coscas [и др.] // *Ophthalmologica*. — 2014. — Т. 232, № 4. — С. 206–212. — DOI: [10.1159/000365705](https://doi.org/10.1159/000365705).

90. Optical coherence tomographic hyperreflective foci: A morphologic sign of lipid extravasation in diabetic macular edema / M. Bolz, U. Schmidt-Erfurth, G. Deak [и др.] // Investigative Ophthalmology and Visual Science. — 2009. — T. 50, № 5. — С. 2309—2314. — DOI: [10.1167/iovs.08-2682](https://doi.org/10.1167/iovs.08-2682).
91. *Otani T., Kishi S., Maruyama Y.* Patterns of diabetic macular edema with optical coherence tomography // American Journal of Ophthalmology. — 1999. — T. 127, № 6. — С. 688—693. — DOI: [10.1016/S0002-9394\(99\)00033-1](https://doi.org/10.1016/S0002-9394(99)00033-1).
92. Optical coherence tomography patterns of diabetic macular edema: Prediction of visual outcome / J. Lee, B. G. Moon, A. R. Cho [и др.] // Retina. — 2016. — T. 36, № 9. — С. 1637—1645. — DOI: [10.1097/IAE.0000000000000953](https://doi.org/10.1097/IAE.0000000000000953).
93. Fully automated segmentation of hyperreflective foci in optical coherence tomography images / T. Schlegl, H. Bogunovic, S. Klimescha [и др.] // arXiv. — 2018. — eprint: [1805.03278](https://arxiv.org/abs/1805.03278).
94. Automatic Segmentation of Diabetic Macular Edema Biomarkers in Optical Coherence Tomography with Deep Learning / Y.-T. Hsu, C.-M. Yang, T.-T. Wu [и др.] // Biomedicines. — 2022. — T. 10, № 11. — С. 2767. — DOI: [10.3390/biomedicines10112767](https://doi.org/10.3390/biomedicines10112767).
95. *Ganjee R., Moghaddam M. E., Nourinia R.* A generalizable approach based on the U-Net model for automatic intraretinal cyst segmentation in SD-OCT images // International Journal of Imaging Systems and Technology. — 2023. — T. 33, № 4. — С. 1647—1660. — DOI: [10.1002/ima.22893](https://doi.org/10.1002/ima.22893).
96. A two-stage CNN model for the classification and severity analysis of retinal and choroidal diseases in OCT images / N. George, L. Shine, N. Ambily [и др.] // International Journal of Intelligent Networks. — 2024. — T. 5. — С. 10—18. — DOI: [10.1016/j.ijin.2024.01.002](https://doi.org/10.1016/j.ijin.2024.01.002).
97. *Melinščak M.* Attention-based U-net: Joint Segmentation of Layers and Fluids from Retinal OCT Images // Proc. 46th MIPRO. — 2023. — С. 391—396.
98. *Daanouni O., Cherradi B., Tmiri A.* Automated end-to-end architecture for retinal layers and fluids segmentation on OCT B-scans // Multimedia Tools and Applications. — 2024. — DOI: [10.1007/s11042-024-19514-z](https://doi.org/10.1007/s11042-024-19514-z).

99. Deep learning model for detecting cystoid fluid collections on OCT in X-linked retinoschisis patients / P. Hensman, Y. El Allali, H. Almushattat [и др.] // *Acta Ophthalmologica*. — 2025. — DOI: [10.1111/aos.17495](https://doi.org/10.1111/aos.17495).
100. *Tang L., Zhang Z., Huang W.* Deep learning model to detect epiretinal membrane in optical coherence tomography images // *Ophthalmic Research*. — 2023. — Т. 66, № 2. — С. 144—150. — DOI: [10.1159/000529299](https://doi.org/10.1159/000529299).
101. Deep learning-based automatic detection of ellipsoid zone loss in spectral-domain OCT for hydroxychloroquine retinal toxicity screening / T. De Silva [и др.] // *Ophthalmology Science*. — 2021. — Т. 1, № 4. — С. 100060.
102. A deep ensemble learning-based CNN architecture for multiclass retinal fluid segmentation in OCT images / M. Rahil, B. N. Anoop, G. N. Girish [и др.] // *IEEE Access*. — 2023. — Т. 11. — С. 17241—17251. — DOI: [10.1109/ACCESS.2023.3244922](https://doi.org/10.1109/ACCESS.2023.3244922).
103. Self-guided optimization semi-supervised method for joint segmentation of macular hole and cystoid macular edema in retinal OCT images / M. Wang, T. Lin, Y. Peng [и др.] // *IEEE Transactions on Biomedical Engineering*. — 2023. — Т. 70, № 7. — С. 2013—2024. — DOI: [10.1109/TBME.2023.3234031](https://doi.org/10.1109/TBME.2023.3234031).
104. Learnable ophthalmology SAM / Z. Qiu, Y. Hu, H. Li [и др.] // *arXiv*. — 2023. — eprint: [2304.13425](https://arxiv.org/abs/2304.13425).
105. Adapting Segment Anything Model (SAM) for Retinal OCT / B. Fazekas, J. Morano, D. Lachinov [и др.] // *Ophthalmic Medical Image Analysis (OMIA), MICCAI Workshop*. — 2023. — С. 92—101.
106. RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge / H. Bogunović, F. Venhuizen, S. Klimescha [и др.] // *IEEE Transactions on Medical Imaging*. — 2019. — Т. 38, № 8. — С. 1858—1874. — DOI: [10.1109/TMI.2019.2901398](https://doi.org/10.1109/TMI.2019.2901398).
107. AROI: Annotated retinal OCT images database / M. Melinščak, M. Radmilović, Z. Vatavuk [и др.] // *44th Int. Convention on Information, Communication and Electronic Technology (MIPRO)*. — 2021. — С. 371—376. — DOI: [10.23919/MIPRO52101.2021.9596934](https://doi.org/10.23919/MIPRO52101.2021.9596934).

108. OIMHS: An optical coherence tomography image dataset based on macular hole manual segmentation / X. Ye, S. He, X. Zhong [и др.] // Scientific Data. — 2023. — Т. 10, № 1. — С. 769. — DOI: [10.1038/s41597-023-02675-1](https://doi.org/10.1038/s41597-023-02675-1).
109. Deep-learning based multiclass retinal fluid segmentation and detection in OCT images using a fully convolutional neural network / D. Lu, M. Heisler, S. Lee [и др.] // Medical Image Analysis. — 2019. — Т. 54. — С. 100—110. — DOI: [10.1016/j.media.2019.02.011](https://doi.org/10.1016/j.media.2019.02.011).
110. OCTDL: Optical coherence tomography dataset for image-based deep learning methods / M. Kulyabin, A. Zhdanov, A. Nikiforova [и др.] // Scientific Data. — 2024. — Т. 11, № 1. — С. 365. — DOI: [10.1038/s41597-024-03182-7](https://doi.org/10.1038/s41597-024-03182-7).
111. OCT5k: A dataset of multi-disease and multi-graded annotations for retinal layers / M. Arikan, M. Willoughby, J. Ongun [и др.] // Scientific Data. — 2025. — Т. 12, № 1. — С. 267. — DOI: [10.1038/s41597-024-04259-z](https://doi.org/10.1038/s41597-024-04259-z).
112. *Ganjee R., Moghaddam M. E., Nourinia R.* An unsupervised hierarchical approach for automatic intra-retinal cyst segmentation in SD-OCT images // Medical Physics. — 2020. — Т. 47, № 11. — С. 4872—4884. — DOI: [10.1002/mp.14361](https://doi.org/10.1002/mp.14361).
113. *Tan M., Le Q. V.* EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning. Т. 97 / под ред. K. Chaudhuri, R. Salakhutdinov. — PMLR, 06.2019. — С. 6105—6114. — (Proceedings of Machine Learning Research).
114. MobileNetV2: Inverted Residuals and Linear Bottlenecks / M. Sandler [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — С. 4510—4520.
115. *Hu J., Shen L., Sun G.* Squeeze-and-Excitation Networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — С. 7132—7141.
116. *Ramachandran P., Zoph B., Le Q. V.* Searching for Activation Functions // arXiv preprint arXiv:1710.05941. — 2018.

117. *Ioffe S., Szegedy C.* Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv preprint arXiv:1502.03167. — 2015.
118. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon [и др.] // CVPR. — 2016.
119. *Redmon J., Farhadi A.* YOLO9000: Better, Faster, Stronger // CVPR. — 2017.
120. *Redmon J., Farhadi A.* YOLOv3: An Incremental Improvement // arXiv preprint arXiv:1804.02767. — 2018.
121. *Bochkovskiy A., Wang C.-Y., Liao H.-Y. M.* YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv preprint arXiv:2004.10934. — 2020.
122. *al. G. J. et.* YOLOv5 — Ultralytics. — 2020. — <https://github.com/ultralytics/yolov5>.
123. *al. G. J. et.* YOLOv8 Documentation — Ultralytics. — 2023. — <https://docs.ultralytics.com>.
124. Path Aggregation Network for Instance Segmentation / S. Liu [и др.] // CVPR. — 2018.
125. FCOS: Fully Convolutional One-Stage Object Detection / Z. Tian [и др.] // ICCV. — 2019.
126. YOLACT: Real-Time Instance Segmentation / D. Bolya [и др.] // ICCV. — 2019.
127. Feature Pyramid Networks for Object Detection / T.-Y. Lin [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — DOI: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
128. Path Aggregation Network for Instance Segmentation / S. Liu [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — DOI: [10.1109/CVPR.2018.00913](https://doi.org/10.1109/CVPR.2018.00913).
129. *Tan M., Pang R., Le Q. V.* EfficientDet: Scalable and Efficient Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — DOI: [10.48550/arXiv.1911.09070](https://doi.org/10.48550/arXiv.1911.09070).

130. *Tan M., Le Q. V.* EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning (ICML). — 2019.
131. *Hu J., Shen L., Sun G.* Squeeze-and-Excitation Networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — С. 7132—7141. — DOI: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
132. *Hou Q., Zhou D., Feng J.* Coordinate Attention for Efficient Mobile Network Design // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2021. — С. 13713—13722. — DOI: [10.1109/CVPR46437.2021.01350](https://doi.org/10.1109/CVPR46437.2021.01350).
133. CBAM: Convolutional Block Attention Module / S. Woo [и др.] // Proceedings of the European Conference on Computer Vision (ECCV). — 2018. — С. 3—19. — DOI: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1).
134. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation / B. Cheng [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2021. — С. 15334—15342. — DOI: [10.1109/CVPR46437.2021.01508](https://doi.org/10.1109/CVPR46437.2021.01508).
135. *Hehn T., Kooij J. F. P., Gavrilă D. M.* Fast and Compact Image Segmentation using Instance Stixels // IEEE Transactions on Intelligent Vehicles. — 2022. — Т. 7, № 1. — С. 45—56. — DOI: [10.1109/TIV.2021.3067223](https://doi.org/10.1109/TIV.2021.3067223).
136. *Ryzhov A. P.* On the quality of classification of objects based on fuzzy rules // Intelligent Systems. — 2005. — Т. 9. — С. 253—264.
137. *Zadeh L. A.* The concept of a linguistic variable and its application to approximate reasoning—I // Information Sciences. — 1975. — Т. 8, № 3. — С. 199—249. — DOI: [10.1016/0020-0255\(75\)90036-5](https://doi.org/10.1016/0020-0255(75)90036-5).
138. An Introduction to Fuzzy Logic Applications in Intelligent Systems / под ред. R. R. Yager, L. A. Zadeh. — Boston, MA : Springer, 1992. — (The Springer International Series in Engineering and Computer Science).
139. *Министерство здравоохранения Российской Федерации, Ассоциация врачей-офтальмологов, Российская ассоциация эндокринологов.* Сахарный диабет: ретинопатия диабетическая, макулярный отек диабетический:

- клинические рекомендации. — Москва, 2023. — 76 с. <https://cr.minzdrav.gov.ru/recommend/17561>.
140. Офтальмология: национальное руководство / под ред. С. Э. Аветисов [и др.]. — 3-е, перераб. и доп. — Москва : ГЭОТАР-Медиа, 2024. — 952. — Серия «Национальные руководства».
 141. *Hecht E.* Optics. — 4-е изд. — San Francisco : Addison-Wesley, 2002. — С. 698.
 142. *Goodman J. W.* Introduction to Fourier Optics. — 3-е изд. — Englewood, CO : Roberts, Company, 2005. — С. 523.
 143. *Group E. T. D. R. S. R.* Grading diabetic retinopathy from stereoscopic color fundus photographs — ETDRS report number 10 // Ophthalmology. — 1991. — Т. 98, 5 Suppl. — С. 786—806.
 144. *Abramoff M. D., Garvin M. K., Sonka M.* Retinal imaging and image analysis // IEEE Reviews in Biomedical Engineering. — 2010. — Т. 3. — С. 169—208.
 145. *Ferris F. L. I., Wilkinson C. P., Bird A. e. a.* Clinical classification of age-related macular degeneration // Ophthalmology. — 2013. — Т. 120, № 4. — С. 844—851.
 146. *Kanski J. J., Bowling B.* Clinical Ophthalmology: A Systematic Approach. — 8-е изд. — London : Elsevier, 2016. — С. 912.
 147. *Wong T. Y., Mitchell P.* Hypertensive retinopathy // New England Journal of Medicine. — 2004. — Т. 351, № 22. — С. 2310—2317.
 148. *Shields J. A., Shields C. L.* Intraocular Tumors: An Atlas and Textbook. — 3-е изд. — Philadelphia : Wolters Kluwer, 2015. — С. 608.
 149. *Дедов И. И., Шестакова М. В., Викулова О. К.* Эпидемиология сахарного диабета в Российской Федерации: клинико-статистический отчёт по данным федерального регистра сахарного диабета // Сахарный диабет. — 2017. — Т. 20, № 1. — С. 13—41. — DOI: [10.14341/DM8664](https://doi.org/10.14341/DM8664).
 150. *Klein R., Klein B. E. K.* Are individuals with diabetes seeing better? A long-term epidemiological perspective // Diabetes. — 2010. — Т. 59, № 8. — С. 1853—1860. — DOI: [10.2337/db09-1904](https://doi.org/10.2337/db09-1904).

151. *Averkina A. N., Volkova E. N., Yarusheva S. A.* Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis // Journal of Computer and Systems Sciences International. — 2024. — Т. 63, № 1. — С. 175–203. — DOI: [10.1134/S1064230724700138](https://doi.org/10.1134/S1064230724700138).
152. *Porta M., Kohner E.* Screening for diabetic retinopathy in Europe // Diabetic Medicine. — 1991. — Т. 8, № 3. — С. 197–198. — DOI: [10.1111/j.1464-5491.1991.tb01571.x](https://doi.org/10.1111/j.1464-5491.1991.tb01571.x).
153. Severity grading and early retinopathy lesion detection through hybrid Inception-ResNet architecture / S. Yasin, N. Iqbal, T. Ali [и др.] // Sensors. — 2021. — Т. 21, № 20. — С. 6933. — DOI: [10.3390/s21206933](https://doi.org/10.3390/s21206933).
154. A systematic literature review on diabetic retinopathy using an artificial intelligence approach / P. Bidwai, S. Gite, K. Pahuja [и др.] // Big Data and Cognitive Computing. — 2022. — Т. 6, № 4. — С. 152. — DOI: [10.3390/bdcc6040152](https://doi.org/10.3390/bdcc6040152).
155. A dual-learning ADRPPA algorithm for microaneurysm segmentation / V. Y. Wang, B. Zhao, H. Wang [и др.] // Scientific Reports. — 2024. — Т. 14. — С. 82884. — DOI: [10.1038/s41598-024-82884-9](https://doi.org/10.1038/s41598-024-82884-9).
156. Automatic detection of microaneurysms in fundus images / V. Raudonis [и др.] // Sensors. — 2023. — Т. 23, № 7. — С. 3431. — DOI: [10.3390/s23073431](https://doi.org/10.3390/s23073431).
157. *Vanaja C. B., Prakasam P.* Convolutional block attention gate-based Unet framework for microaneurysm segmentation using retinal fundus images // BMC Medical Imaging. — 2025. — Т. 25. — С. 83. — DOI: [10.1186/s12880-025-01625-0](https://doi.org/10.1186/s12880-025-01625-0).
158. Efficient microaneurysm segmentation in retinal images via a lightweight Attention U-Net for early DR diagnosis / M. Z. Tahir [и др.] // SLAS Technology. — 2025. — Т. 34. — С. 100323. — DOI: [10.1016/j.slast.2025.100323](https://doi.org/10.1016/j.slast.2025.100323).
159. *Skouta A., Elhassouny A., Kadoch B.* Hemorrhage semantic segmentation in fundus images for the diagnosis of diabetic retinopathy using modified U-Net // Journal of Big Data. — 2022. — Т. 9. — С. 126. — DOI: [10.1186/s40537-022-00632-0](https://doi.org/10.1186/s40537-022-00632-0).

160. *Tang S., Wu Q.* Research on recognition of diabetic retinopathy hemorrhage lesions based on fine tuning of segment anything model // Scientific Reports. — 2025. — T. 15. — C. 10292. — DOI: [10.1038/s41598-025-92665-7](https://doi.org/10.1038/s41598-025-92665-7).
161. Semantic segmentation of retinal exudates using a residual encoder–decoder architecture in diabetic retinopathy / M. A. Manan [и др.] // Microscopy Research and Technique. — 2023. — T. 86, № 11. — C. 1443–1460. — DOI: [10.1002/jemt.24345](https://doi.org/10.1002/jemt.24345).
162. Segmentation of hard exudate lesions in color fundus images using a two-stage CNN / Q. Van Do, M. C. Nguyen, T. T. Nguyen [и др.] // Knowledge-Based Systems. — 2024. — DOI: [10.1016/j.knosys.2023.110015](https://doi.org/10.1016/j.knosys.2023.110015).
163. A Deep Learning Segmentation Model for Detection of Active Proliferative Diabetic Retinopathy / S. Dinesen [и др.] // Ophthalmology and Therapy. — 2025. — DOI: [10.1007/s40123-025-01127-w](https://doi.org/10.1007/s40123-025-01127-w).
164. RTNet: Relation Transformer Network for Diabetic Retinopathy Multi-Lesion Segmentation / S. Huang [и др.] // IEEE Transactions on Medical Imaging. — 2022. — T. 41, № 6. — C. 1596–1607. — DOI: [10.1109/TMI.2022.3143833](https://doi.org/10.1109/TMI.2022.3143833).
165. Vision transformer adapter-based hyperbolic embeddings for multi-lesion segmentation in diabetic retinopathy / Z. Wang [и др.] // Scientific Reports. — 2023. — T. 13. — C. 11178. — DOI: [10.1038/s41598-023-38320-5](https://doi.org/10.1038/s41598-023-38320-5).
166. Prior-guided attention fusion transformer for multi-lesion segmentation of diabetic retinopathy / C. Xu, S. Wang, L. Zhang [и др.] // Scientific Reports. — 2024. — T. 14. — C. 20892. — DOI: [10.1038/s41598-024-71650-6](https://doi.org/10.1038/s41598-024-71650-6).
167. Lightweight and multi-lesion segmentation model for diabetic retinopathy based on BiSeNet V2 / W. Gao, J. Chen, Y. Li [и др.] // Computers in Biology and Medicine. — 2024. — T. 169. — C. 107854. — DOI: [10.1016/j.compbiomed.2023.107854](https://doi.org/10.1016/j.compbiomed.2023.107854).
168. Diabetic retinopathy lesion segmentation method based on multi-scale attention and lesion perception (MLNet) / Y. Bian [и др.] // Algorithms. — 2024. — T. 17, № 4. — C. 164. — DOI: [10.3390/algorithms17040164](https://doi.org/10.3390/algorithms17040164).

169. SegDRoWS: segmentation of diabetic retinopathy lesions by a whole-stage multi-scale feature fusion network / J. Liu [и др.] // Biomedical Signal Processing and Control. — 2025. — Т. 99. — С. 107581. — DOI: [10.1016/j.bspc.2025.107581](https://doi.org/10.1016/j.bspc.2025.107581).
170. *Garifullin A., Lensu L., Uusitalo H.* Deep Bayesian baseline for segmenting diabetic retinopathy lesions // Computers in Biology and Medicine. — 2021. — Т. 136. — С. 104725. — DOI: [10.1016/j.combiomed.2021.104725](https://doi.org/10.1016/j.combiomed.2021.104725).
171. APTOS 2019 Blindness Detection. — 2019. — дата обращения: 12.07.2025. <https://www.kaggle.com/competitions/aptos2019-blindness-detection>.
172. MAPLES-DR: MESSIDOR Anatomical and Pathological Labels for Explainable Screening of Diabetic Retinopathy / G. Lepetit-Aimon [и др.] // Scientific Data. — 2024. — Т. 11. — С. 914. — DOI: [10.1038/s41597-024-03739-6](https://doi.org/10.1038/s41597-024-03739-6).
173. A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability / Y. Zhou [и др.] // IEEE Transactions on Medical Imaging. — 2021. — Т. 40, № 3. — С. 818—828. — DOI: [10.1109/TMI.2020.3037771](https://doi.org/10.1109/TMI.2020.3037771).
174. Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research / P. Porwal [и др.] // Data. — 2018. — Т. 3, № 3. — С. 25. — DOI: [10.3390/data3030025](https://doi.org/10.3390/data3030025).
175. *Abràmoff M. D., Garvin M. K., Sonka M.* Retinal Imaging and Image Analysis // IEEE Reviews in Biomedical Engineering. — 2010. — Т. 3. — С. 169—208. — DOI: [10.1109/RBME.2010.2084567](https://doi.org/10.1109/RBME.2010.2084567).
176. Which Color Channel Is Better for Diagnosing Retinal Diseases Automatically in Color Fundus Photographs? / S. S. Biswas [и др.] // Life. — 2022. — Т. 12, № 7. — С. 973. — DOI: [10.3390/life12070973](https://doi.org/10.3390/life12070973).
177. Comparison of Multicolor Scanning Laser Imaging and Color Fundus Photography in Evaluating Vessel Whitening in BRVO / N. Unno [и др.] // Ophthalmic Research. — 2023. — Т. 66, № 1. — С. 413—420. — DOI: [10.1159/000528251](https://doi.org/10.1159/000528251).

178. *Tomasi C., Manduchi R.* Bilateral Filtering for Gray and Color Images // Proceedings of the 6th IEEE International Conference on Computer Vision. — 1998. — С. 839—846. — DOI: [10.1109/ICCV.1998.710815](https://doi.org/10.1109/ICCV.1998.710815).
179. *Zuiderveld K.* Contrast Limited Adaptive Histogram Equalization // Graphics Gems IV / под ред. P. S. Heckbert. — Academic Press, 1994. — С. 474—485. — DOI: [10.1016/B978-0-12-336156-1.50061-6](https://doi.org/10.1016/B978-0-12-336156-1.50061-6).
180. Adaptive Histogram Equalization and Its Variations / S. M. Pizer [и др.] // Computer Vision, Graphics, and Image Processing. — 1987. — Т. 39, № 3. — С. 355—368. — DOI: [10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X).
181. *Simonyan K., Zisserman A.* Very Deep Convolutional Networks for Large-Scale Image Recognition // arXiv preprint arXiv:1409.1556. — 2014. — С. 1—14.
182. *Rawat W., Wang Z.* Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review // Neural Computation. — 2017. — Т. 29, № 9. — С. 2352—2449. — DOI: [10.1162/NECO_a_00990](https://doi.org/10.1162/NECO_a_00990).
183. Convolutional neural networks: an overview for radiologists / R. Yamashita [и др.] // Insights into Imaging. — 2018. — Т. 9, № 4. — С. 611—629. — DOI: [10.1007/s13244-018-0639-9](https://doi.org/10.1007/s13244-018-0639-9).
184. Deep Residual Learning for Image Recognition / K. He [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — С. 770—778. — DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
185. Densely Connected Convolutional Networks / G. Huang [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — С. 4700—4708. — DOI: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
186. *Kohavi R.* A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection // Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI). — 1995. — С. 1137—1145.
187. *Buda M., Maki A., Mazurowski M. A.* A systematic study of the class imbalance problem in convolutional neural networks // Neural Networks. — 2018. — Т. 106. — С. 249—259. — DOI: [10.1016/j.neunet.2018.07.011](https://doi.org/10.1016/j.neunet.2018.07.011).

188. *Shorten C., Khoshgoftaar T. M.* A survey on Image Data Augmentation for Deep Learning // Journal of Big Data. — 2019. — Т. 6, № 60. — С. 1—48. — DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
189. *Kingma D. P., Ba J.* Adam: A Method for Stochastic Optimization // arXiv preprint arXiv:1412.6980. — 2015. — С. 1—15.
190. Dropout: A Simple Way to Prevent Neural Networks from Overfitting / N. Srivastava [и др.] // Journal of Machine Learning Research. — 2014. — Т. 15. — С. 1929—1958.
191. *Gonzalez R. C., Woods R. E.* Digital Image Processing. — 3rd. — Upper Saddle River, NJ : Prentice Hall, 2008.
192. *Soille P.* Morphological Image Analysis: Principles and Applications. — 2nd. — Berlin; Heidelberg : Springer, 2003.
193. *Pratt W. K.* Digital Image Processing. — 4th. — Hoboken : John Wiley Sons, 2016.
194. Human photoreceptor topography / C. A. Curcio [и др.] // Journal of Comparative Neurology. — 1990. — Т. 292, № 4. — С. 497—523. — DOI: [10.1002/cne.902920402](https://doi.org/10.1002/cne.902920402).
195. *Jonas J. B., Budde W. M.* Optic disc-fovea distance, rim width, and optic disc shape // British Journal of Ophthalmology. — 1998. — Т. 82, № 9. — С. 1029—1032. — DOI: [10.1136/bjo.82.9.1029](https://doi.org/10.1136/bjo.82.9.1029).
196. Visual acuity is correlated with the area of the foveal avascular zone in diabetic retinopathy / C. Balaratnasingam [и др.] // Investigative Ophthalmology & Visual Science. — 2015. — Т. 56, № 9. — С. 5200—5206. — DOI: [10.1167/iovs.15-16785](https://doi.org/10.1167/iovs.15-16785).
197. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation // Lecture Notes in Computer Science (MICCAI). Т. 9351. — 2015. — С. 234—241. — DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
198. U-Net: Deep learning for cell counting, detection, and morphometry / T. Falk [и др.] // Nature Methods. — 2019. — Т. 16. — С. 67—70. — DOI: [10.1038/s41592-018-0261-2](https://doi.org/10.1038/s41592-018-0261-2).

199. UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Z. Zhou [и др.] // IEEE Transactions on Medical Imaging. — 2020. — Т. 39, № 6. — С. 1856—1876. — DOI: [10.1109/TMI.2019.2959609](https://doi.org/10.1109/TMI.2019.2959609).
200. nnU-Net: a self-adapting framework for U-Net-based medical image segmentation / F. Isensee [и др.] // Nature Methods. — 2021. — Т. 18. — С. 203—211. — DOI: [10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z).
201. Attention U-Net: Learning Where to Look for the Pancreas / O. Oktay [и др.] // arXiv preprint arXiv:1804.03999. — 2018.
202. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation / Ö. Çiçek [и др.] // Lecture Notes in Computer Science (MICCAI). Т. 9901. — 2016. — С. 424—432. — DOI: [10.1007/978-3-319-46723-8_49](https://doi.org/10.1007/978-3-319-46723-8_49).
203. *Gabor D.* Theory of communication // Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering. — 1946. — Т. 93, № 26. — С. 429—457. — DOI: [10.1049/ji-3-2.1946.0074](https://doi.org/10.1049/ji-3-2.1946.0074).
204. *Daugman J. G.* Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters // Journal of the Optical Society of America A. — 1985. — Т. 2, № 7. — С. 1160—1169. — DOI: [10.1364/JOSAA.2.001160](https://doi.org/10.1364/JOSAA.2.001160).
205. A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA / H.-B. Deng [и др.] // International Journal of Information Technology. — 2005. — Т. 11, № 11. — С. 86—96.
206. *Volkov E., Averkin A.* Modification of U-Net Architecture Based on Gabor Filters with Trainable Parameters to Improve Biomarker Segmentation on OCT Images // 2025 VI International Conference on Neural Networks and Neurotechnologies (NeuroNT). — IEEE, 2025. — С. 66—68. — DOI: [10.1109/NeuroNT66873.2025.11049967](https://doi.org/10.1109/NeuroNT66873.2025.11049967).
207. Learning Deep Features for Discriminative Localization / B. Zhou [и др.] // Proc. CVPR. — 2016. — С. 2921—2929. — DOI: [10.1109/CVPR.2016.319](https://doi.org/10.1109/CVPR.2016.319).
208. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization / R. R. Selvaraju [и др.] // Proc. ICCV. — 2017. — С. 618—626. — DOI: [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74).

209. *Chetoui M., Akhloufi M. A.* Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets // *Journal of Medical Imaging*. — 2020. — Т. 7, № 4. — С. 044503. — DOI: [10.1117/1.JMI.7.4.044503](https://doi.org/10.1117/1.JMI.7.4.044503).
210. A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification / H. Jiang [и др.] // *Proc. EMBC*. — 2020. — С. 1560—1563. — DOI: [10.1109/EMBC44109.2020.9175884](https://doi.org/10.1109/EMBC44109.2020.9175884).
211. *Papadopoulos A., Topouzis F., Delopoulos A.* An interpretable multiple-instance approach for the detection of referable diabetic retinopathy in fundus images // *Scientific Reports*. — 2021. — Т. 11. — С. 14326. — DOI: [10.1038/s41598-021-93632-8](https://doi.org/10.1038/s41598-021-93632-8).
212. ExplAIIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis / G. Quellec [и др.] // *Medical Image Analysis*. — 2021. — Т. 72. — С. 102118. — DOI: [10.1016/j.media.2021.102118](https://doi.org/10.1016/j.media.2021.102118).
213. *Ribeiro M. T., Singh S., Guestrin C.* “Why Should I Trust You?”: Explaining the Predictions of Any Classifier // *Proc. KDD*. — 2016. — С. 1135—1144. — DOI: [10.1145/2939672.2939778](https://doi.org/10.1145/2939672.2939778).
214. *Lundberg S. M., Lee S.-I.* A Unified Approach to Interpreting Model Predictions // *Advances in Neural Information Processing Systems*. — 2017. — Т. 30. — С. 4765—4774.
215. *Sundararajan M., Taly A., Yan Q.* Axiomatic Attribution for Deep Networks // *Proc. ICML*. — 2017. — С. 3319—3328.
216. *Atci Ş. Y., Zengin M. A., Susam E.* Identifying Diabetic Retinopathy in the Human Eye: A Hybrid Approach Based on a Computer-Aided Diagnosis System Combined with Deep Learning // *Diagnostics*. — 2024.
217. *Malerbi F. K., al. et.* Heatmap analysis for artificial intelligence explainability in diabetic retinopathy detection: illuminating the rationale of deep learning decisions // *Annals of Translational Medicine*. — 2024. — in press.
218. Sanity Checks for Saliency Maps / J. Adebayo [и др.] // *Advances in Neural Information Processing Systems (NeurIPS)*. — 2018.
219. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation / L.-C. Chen [и др.] // *Computer Vision – ECCV 2018*. — 2018. — С. 833—851.

220. *Gildenblat J.* pytorch-grad-cam: PyTorch library for CAM methods. — 2021. — Дата обращения: 14.06.2025. <https://github.com/jacobgil/pytorch-grad-cam>.
221. *Beucher S., Meyer F.* The Morphological Approach to Segmentation: The Watershed Transformation // *Mathematical Morphology in Image Processing* / под ред. E. R. Dougherty. — New York : Marcel Dekker, 1993. — С. 433—481.
222. *Vincent L., Soille P.* Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 1991. — Т. 13, № 6. — С. 583—598. — DOI: [10.1109/34.87344](https://doi.org/10.1109/34.87344).
223. *Meyer F.* Topographic Distance and Watershed Lines // *Signal Processing*. — 1994. — Т. 38, № 1. — С. 113—125. — DOI: [10.1016/0165-1684\(94\)90060-4](https://doi.org/10.1016/0165-1684(94)90060-4).
224. *Roerdink J. B. T. M., Meijster A.* The Watershed Transform: Definitions, Algorithms and Parallelization Strategies // *Fundamenta Informaticae*. — 2000. — Т. 41, № 1/2. — С. 187—228.
225. Watershed Cuts: Minimum Spanning Forests and the Drop of Water Principle / J. Cousty [и др.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2009. — Т. 31, № 8. — С. 1362—1374. — DOI: [10.1109/TPAMI.2008.165](https://doi.org/10.1109/TPAMI.2008.165).
226. *Developers O.* Image Segmentation with Watershed Algorithm. — 2024. — Дата обращения: 12.07.2025. https://docs.opencv.org/4.x/d3/db4/tutorial_py_watershed.html.
227. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales / C. P. Wilkinson [и др.] // *Ophthalmology*. — 2003. — Т. 110, № 9. — С. 1677—1682.
228. *Cheung N., Mitchell P., Wong T. Y.* Diabetic retinopathy // *The Lancet*. — 2010. — Т. 376, № 9735. — С. 124—136.

Благодарности

Работа выполнена в рамках проектов государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124092700007-4, "Применение объяснительного искусственного интеллекта для интерпретации алгоритмов машинного обучения"; тема № 124040200033-9, "Математические методы анализа данных и прогнозирования").