

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»

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Москва – 2025

Работа выполнена в Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) в отделе № 14.

**Научный
руководитель:**

Аверкин Алексей Николаевич
кандидат физико-математических наук,
доцент, ведущий научный сотрудник
отдела № 14, ФИЦ ИУ РАН

**Официальные
оппоненты:**

Еремеев Александр Павлович
доктор технических наук, профессор,
профессор кафедры прикладной математики
и искусственного интеллекта,
ФГБОУ ВО «Национальный
исследовательский университет «МЭИ»

Ямаев Андрей Викторович
кандидат технических наук,
старший инженер,
ООО «КОУЛМЭН ТЕХ»

**Ведущая
организация:**

ФГБОУ ВО «Тверской государственный
университет» (ТвГУ)

Защита состоится «25» декабря 2025 г. в ____ : ____ часов на заседании
диссертационного совета 24.1.224.03 на базе ФИЦ ИУ РАН по
адресу: 119333, г. Москва, ул. Вавилова, д. 42.

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН и
на официальном сайте <http://www.frccsc.ru>.

Автореферат разослан «____» 2025 г.

Учёный секретарь
диссертационного совета 24.1.224.03,
к.т.н.



И.А. Рейер

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИИ

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.

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 интрапетинальные 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 (интрапетинальные 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 здравоохранения. 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.

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 экспертизы 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 экспертизы 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, отличающаяся нейросетевой оценкой качества изображений с применением метода объяснительного ИИ, интеграцией базовой модели на основе визуальных промптов, а также применением перекрёстных 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 «Дубна» (2024); XII Международная 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 результатов подтверждается корректным использованием методов, обоснованием постановки задач, а также экспериментальными исследованиями на открытых данных и данных индустриальных 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 к.ф.-м.н., доц. Аверкина А.Н.

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.

КРАТКОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность темы исследования, сформулированы цель и задачи работы, раскрыты научная новизна и практическая значимость, приведены положения, выносимые на защиту, описаны методы и средства исследования, представлена структура и объём диссертации, а также сведения об апробации и публикациях по теме работы.

В первой главе изложена методика полуавтоматического формирования проблемно-ориентированных коллекций медицинских данных, предназначенных для обучения и объективной валидации нейросетевых моделей анализа ретинальных изображений. Показано, что качество и воспроизводимость алгоритмов глубокого обучения напрямую зависят от корректности исходных данных, стандартизации протоколов их подготовки и соблюдения требований к структуре и целостности коллекций.

Определено, что коллекция медицинских данных представляет собой организованное множество взаимосвязанных изображений и аннотаций, формируемое по единым протоколам сбора, предобработки, хранения и доступа: $D = \{(I_i, A_i)\}_{i=1}^N$, $I_i \in R^{H \times W}$, I_i – изображение, A_i – аннотационная маска или набор меток.

Рассмотрены принципы формирования медицинских коллекций данных в соответствии с нормативными требованиями Российской Федерации. Подчёркнута необходимость обеспечения репрезентативности, полноты, сохранения метаданных, контроля качества изображений и стандартизации форматов хранения. Репрезентативность коллекции достигается охватом различных патологических состояний и типов устройств. Целостность обеспечивается включением контрольных сумм и версионного журнала обработки.

Для автоматизации отбора изображений по техническому качеству предложена модель бинарной классификации $f_\theta(I)$, минимизирующая функцию потерь бинарной кросс-энтропии:

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

где $y_i = 1$ для качественных изображений и $y_i = 0$ для некачественных.

В качестве архитектуры используется свёрточная нейронная сеть с модулем пространственного внимания, что обеспечивает фокусировку на анатомически значимых областях. Данный модуль

визуализирует зоны, оказавшие наибольшее влияние на классификацию, обеспечивая интерпретируемость результатов.

После автоматического отбора выполняется предварительная обработка изображений. Для приведения данных к единому формату применяется нормализация интенсивностей: $I_{\text{norm}}(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}}$, а для устранения спектрального шума используется медианная фильтрация: $I_{\text{median}}(x, y) = \text{median}\{I(u, v) | (u, v) \in N_{k \times k}(x, y)\}$.

Масштабирование изображений до фиксированного размера выполняется с сохранением пропорций: $I_{\text{resized}}(x', y') = I(\alpha x', \beta y')$, $\alpha = \frac{m}{M}$, $\beta = \frac{n}{N}$, что обеспечивает сопоставимость изображений, полученных на различных устройствах.

На следующем этапе описана процедура предварительной аннотации с использованием базовой модели сегментации SAM (Segment Anything Model), способной выделять объекты по визуальным промптам (точкам или рамкам). Формально процесс определяется как отображение: $M = f_{\text{SAM}}(I, P)$, где I – изображение, P – множество промптов, M – множество бинарных масок, соответствующих предполагаемым объектам. Полученные маски служат черновыми аннотациями для последующей экспертной проверки.

Экспертная аннотация выполняется врачами-офтальмологами и включает проверку соответствия разметки клиническим критериям, уточнение границ патологических структур, разделение пересекающихся областей и финальную верификацию. Второй эксперт проводит независимую перекрёстную проверку, после чего формируется согласованная аннотация, принимаемая как эталонная. Такой подход обеспечивает клиническую достоверность и воспроизводимость данных при значительном сокращении времени ручной разметки.

Методы объяснительного искусственного интеллекта (ОИИ) применяются для визуальной оценки релевантности работы нейросетевых моделей. Рассмотрен метод Class Activation Mapping (CAM), формирующий тепловую карту значимости признаков. Для каждого класса с карты вычисляется как:

$$L_{\text{CAM}}^{(c)}(i, j) = \sum_k w_k^{(c)} A_{k,ij},$$

где $A_{k,ij}$ – карты активаций последнего свёрточного слоя, $w_k^{(c)}$ – веса линейного классификатора. Эта карта масштабируется до размеров

исходного изображения и используется для визуализации областей, наиболее влияющих на решение модели. САМ-анализ интегрирован в контур оценки качества изображений.

В разделе о формировании коллекции данных описано использование существующих открытых наборов оптических когерентных томограмм (OCTDL, OCTID, Kermany OCT2017) как источников изображений. Из них отбираются снимки, удовлетворяющие критериям анатомической релевантности и технического качества, после чего выполняется экспертная разметка в формате сегментации экземпляров.

Формирование стандартизированной коллекции данных завершается объединением оригинальных изображений, масок и метаданных в единую структуру, обеспечивающую воспроизводимость экспериментов и документированную историю обработки. Полученный набор включает 900 ОКТ-срезов с разрешением 512×1024 пикселя и содержит 6752 экземпляра биомаркеров восьми классов (HE, HF, IRC, EZD, SRF, VMT, DRIL, ERM). Маски представляют точные границы патологических очагов и могут использоваться для обучения, тестирования и объективной оценки сегментационных моделей.

Разработанная методика объединяет нейросетевые и экспертные подходы, включая оценку качества изображений, использование базовых моделей с визуальными промптами, ОИИ и перекрёстную аннотацию, формируя основу для создания достоверных и масштабируемых медицинских наборов данных.

Во второй главе изложены теоретические основы и разработан интерпретируемый метод автоматического анализа ОКТ-снимков для диагностики диабетического макулярного отёка (ДМО), объединяющий этапы предварительной классификации, сегментации экземпляров биомаркеров и логического вывода. Дано краткое описание принципов ОКТ и клинико-морфологической специфики ДМО, сформулирована структура конвейера обработки, описаны используемые коллекции данных, процедуры предподготовки изображений, архитектуры и функции потерь нейросетевых модулей, метрики качества, правила извлечения количественных признаков и схема нечёткой классификации.

Показано, что ОКТ обеспечивает послойную визуализацию сетчатки с микрометровым аксиальным разрешением и является стандартом верификации ДМО; перечислены диагностически

значимые ОКТ-биомаркеры: дезорганизация внутренних слоёв (DRIL), эпиретинальная мембрана (ERM), утрата эллипсоидной зоны (EZD), твёрдые экссудаты (HE), гиперрефлективные фокусы (HF), интрапретинальные кисты (IRC), интрапретинальная и субретинальная жидкость (IRF, SRF), витреомакулярная тракция (VMT). Их совместный анализ лежит в основе дальнейшего автоматического выделения и интерпретации признаков.

Предложен метод автоматического анализа ОКТ-снимков для диагностики ДМО, включающий следующие этапы: (1) предобработка и фильтрация изображений по техническому качеству – автоматическое исключение снимков с низкой резкостью, контрастом или наличием артефактов; (2) бинарная классификация «ДМО/нет» – определение наличия признаков отёка на основе нейросетевой модели EfficientNetB0; (3) визуальное объяснение решения – построение тепловой карты внимания методом CAM для подтверждения корректности классификации; (4) сегментация экземпляров биомаркеров – автоматическое выделение областей патологических структур с использованием модифицированной модели YOLOv8-BiFPN-CA; (5) количественный анализ масок – вычисление числа, площади и соотношений сегментированных биомаркеров для получения объективных диагностических показателей; (6) нечёткая классификация состояния – определение степени выраженности диабетического макулярного отёка с использованием системы лингвистических правил на основе количественных признаков.

Для этапа предварительной классификации использована открытая клиническая коллекция OCTDL; для задач сегментации сформирована собственная выборка ОКТ-срезов с полной ручной аннотацией экземпляров указанных биомаркеров. Разбиение при обучении/валидации/тестировании – 70%/15%/15%. Приведены суммарные мощности классов и подчёркнута пригодность набора для объективной оценки алгоритмов.

Предварительная классификация осуществляется с помощью модели EfficientNetB0. В данной архитектуре используется каскад MBConv-блоков с механизмом Squeeze-and-Excitation и глобальным усреднением. Композиция стадий представлена как $N = F_1 \circ F_2 \circ \dots \circ F_K(X)$, а ключевые операции внутри MBConv записаны в виде расширения каналов: $X_{\text{exp}} = \text{Conv}_{1 \times 1}(X)$, $\dim(X_{\text{exp}}) = H \times W \times tC$,

глубинной свёртки: $X_{\text{dw}} = \text{DWConv}_{k \times k}(X_{\text{exp}})$, SE-калибровки по каналам: $z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X_{\text{dw}}^{(c)}(i, j)$ и финальной проекции с возможной остаточной связью: $X_{\text{proj}} = \text{Conv}_{1 \times 1}(X_{\text{se}}), Y = X_{\text{proj}} + X$ (при $s = 1$). На выходе применяется $\text{softmax}(W_{\text{FC}} \cdot \text{GAP}(X_{\text{final}}) + b)$, а в качестве активации – $\text{SiLU}(x) = x \cdot \sigma(x)$. Результаты классификации на датасете OCTDL представлены в Табл. 1.

Таблица 1. 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

Выбор модели YOLOv8 в качестве базовой архитектуры был обусловлен её высокой скоростью инференса, модульной архитектурой и поддержкой одновременного решения задач детекции и сегментации, что делает её оптимальной основой для разработки интерпретируемого метода анализа медицинских изображений. Кроме того, структура модели может быть адаптирована к интеграции дополнительных механизмов агрегации и внимания, необходимых для точного выделения мелких биомаркеров на ОКТ-снимках.

Сегментация реализуется через комплексную трёхкомпонентную функцию потерь:

$$\begin{aligned} \mathcal{L}_{\text{total}} &= \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{bbox}} L_{\text{bbox}} + \lambda_{\text{mask}} L_{\text{mask}}, \text{ где:} \\ \mathcal{L}_{\text{cls}} &= -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C [y_{i,c} \log \hat{p}_{i,c} + (1 - y_{i,c}) \log (1 - \hat{p}_{i,c})], \\ \mathcal{L}_{\text{bbox}} &= \frac{1}{N} \sum_{i=1}^N (1 - \text{IoU}(B_i, \hat{B}_i)), \\ \mathcal{L}_{\text{mask}} &= -\frac{1}{M} \sum_{j=1}^M [m_j \log \hat{m}_j + (1 - m_j) \log (1 - \hat{m}_j)]. \end{aligned}$$

Для повышения чувствительности к мелким структурам базовую архитектуру дополнительно усиливают модулями BiFPN и Coordinate Attention (CA). В BiFPN обучаемые неотрицательные веса нормируются: $\widetilde{w}_l = \frac{\max(0, w_i)}{\varepsilon + \sum_{j=1}^m \max(0, w_j)}$, $Y = \phi(\sum_{i=1}^m \widetilde{w}_i X_i)$, что даёт устойчивое взвешенное объединение многоуровневых признаков. В Coordinate Attention выполняется раздельное усреднение по осям: $p_h(c, y) = \frac{1}{W} \sum_{x=1}^W X(c, y, x)$, $p_w(c, x) = \frac{1}{H} \sum_{y=1}^H X(c, y, x)$, полученные

представления проходят через свёртки и сигмоиду: $a_h = \sigma(\text{Conv}_{1 \times 1}^{(h)}(f_h))$, $a_w = \sigma(\text{Conv}_{1 \times 1}^{(w)}(f_w))$, после чего исходные признаки масштабируются: $Y(c, y, x) = X(c, y, x) \cdot a_h(c, y) \cdot a_w(c, x)$.

Интеграция модулей BiFPN и Coordinate Attention в архитектуру YOLOv8-seg направлена на повышение точности сегментации мелких деталей за счёт усовершенствования путей агрегации признаков и внедрения пространственно-чувствительного внимания. BiFPN обеспечивает двунаправленное распространение информации и адаптивное объединение признаков разных уровней разрешения, а Coordinate Attention позволяет учитывать пространственную ориентацию и положение объектов в изображении. Такая комбинация обеспечивает прирост точности сегментации без существенного увеличения вычислительной нагрузки, что особенно важно при работе с изображениями, насыщенными мелкими деталями в задачах медицинской диагностики. Результаты, проведённых вычислительных экспериментов, представленные в Табл. 2 показывают преимущество модифицированной архитектуры YOLOv8-BiFPN-CA. Средний прирост метрики IoU составил 2-3 п.п., особенно заметный на классах с мелкими и слабоконтрастными объектами (HF, HE, DRIL, EZD), что подтверждает эффективность предложенной модели.

Таблица 2. Метрики сегментации на YOLOv8 и YOLOv8-BiFPN-CA

Класс	IoU (base)	IoU (mod)	Sens. (base)	Sens. (mod)	Spec. (base)	Spec. (mod)
DRIL	0.86 ± 0.021	0.91 ± 0.017	0.87 ± 0.020	0.92 ± 0.015	0.985 ± 0.014	0.990 ± 0.010
ERM	0.88 ± 0.018	0.90 ± 0.015	0.89 ± 0.017	0.91 ± 0.014	0.98 ± 0.012	0.98 ± 0.009
EZD	0.85 ± 0.024	0.91 ± 0.018	0.86 ± 0.023	0.92 ± 0.015	0.97 ± 0.015	0.98 ± 0.010
HE	0.85 ± 0.026	0.90 ± 0.020	0.87 ± 0.024	0.91 ± 0.017	0.97 ± 0.014	0.98 ± 0.011
HF	0.85 ± 0.025	0.91 ± 0.019	0.86 ± 0.023	0.92 ± 0.016	0.97 ± 0.016	0.98 ± 0.012
IRC	0.89 ± 0.017	0.91 ± 0.014	0.90 ± 0.016	0.92 ± 0.013	0.98 ± 0.012	0.98 ± 0.010

IRF	0.88 ± 0.018	0.90 ± 0.015	0.89 ± 0.017	0.91 ± 0.014	0.98 ± 0.012	0.98 ± 0.010
SRF	0.90 ± 0.015	0.92 ± 0.013	0.91 ± 0.014	0.93 ± 0.011	0.98 ± 0.011	0.98 ± 0.009
VMT	0.87 ± 0.019	0.89 ± 0.017	0.88 ± 0.018	0.90 ± 0.015	0.98 ± 0.013	0.98 ± 0.011

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.

Определение количественных характеристик экземпляров осуществляется путём вычисления площади каждого сегментированного биомаркера: $A = s_x s_y \sum_{i,j} M_{ij}$, где s_x, s_y – физические размеры пикселя по осям, M_{ij} – бинарная маска экземпляра. В отсутствие калибровки площадь выражается в пикселях: $A_{px} = \sum_{i,j} M_{ij}$. Полученные численные характеристики используются для дальнейшей интерпретации и построения нечётких правил классификации.

Для интеграции количественных признаков в клиническую интерпретацию применяется нечёткий классификатор с лингвистическими правилами вида «если $A_1 = a_{1j}, \dots, A_n = a_{nj}$, то $\mu_{r_j}(O) = \pi_{ij}$ ». Агрегированные степени принадлежности классу C_k вычисляются как: $S_k(O) = \sum_{r=1}^R \pi_r \eta_{r,k}$, $M_k^*(O) = \frac{S_k(O)}{\sum_{q=1}^K S_q(O)}$, что обеспечивает переход от численных характеристик экземпляров биомаркеров к диагностическим заключениям о стадии патологического процесса.

В третьей главе изложен метод анализа фундус-снимков, ориентированный на автоматическое выделение анатомических зон глазного дна и семантическую сегментацию биомаркеров диабетической ретинопатии с последующей интерпретацией на основе логических правил. Предложен метод автоматического анализа фундус-снимков, реализованный в виде последовательного конвейера, включающего следующие этапы: (1) ввод изображения и проверка качества – контроль технических параметров и исключение неинформативных снимков; (2) морфологическая коррекция –

компенсация неравномерного фона и выделение оптического диска и фовеа для анатомической привязки изображения; (3) нормализация изображения – выравнивание контраста и подавление шума с сохранением сосудистой структуры; (4) семантическая сегментация биомаркеров – автоматическое выделение патологических областей с использованием модифицированной архитектуры U-Net, дополненной обучаемыми габоровскими фильтрами для повышения чувствительности к мелким и ориентированным структурам; (5) постобработка и количественный анализ – подсчёт экземпляров биомаркеров, определение их площади и распределения по анатомическим квадрантам сетчатки; (6) логическая классификация стадии – определение стадии диабетической ретинопатии (NPDR, PPDR, PDR) на основе количества и пространственного расположения выявленных биомаркеров.

Интенсивности каждого цветового канала $c \in \{R, G, B\}$ приводятся к $[0,1]$ методом Min-Max: $I_c^{(n)}(x) = \frac{I_c(x) - \min_{\Omega} I_c}{\max_{\Omega} I_c - \min_{\Omega} I_c}$, $x \in \Omega$, далее выделяется зелёный канал $I_g(x) = I_G^{(n)}(x)$ и применяется билатеральная фильтрация. Такой конвейер нормализует яркость и контраст изображения, а также подавляет шум без размывания границ сосудов. Дополнительно используется метод CLAHE.

Для предварительной классификации (норма/патология) используется модель VGG16 (трансферное обучение и регуляризация). По итогам экспериментов (Табл. 3) высокая чувствительность патологического класса подтверждает пригодность этапа в качестве предфильтра перед детальной обработкой.

Таблица 3.

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

В разделе представлена методика выделения анатомических ориентиров глазного дна – оптического диска и зоны фовеа на основе морфологических преобразований и анализа гистограмм яркости цветовых каналов изображения. Метод разработан как универсальный и не требующий предварительного обучения, что обеспечивает его применение к изображениям, полученным на различных фундус-камерах и при различной освещённости.

На первом этапе выполняется компенсация неравномерного фона и ослабление влияния глобальной освещённости с использованием морфологического открытия.

Для каждого цветового канала I_c вычисляется корректированное изображение: $I_c^* = I_c - (I_c \circ b_r)$, где символ \circ обозначает операцию морфологического открытия, а b_r – структурирующий элемент круглой формы радиуса r . Такое преобразование устраниет плавные перепады яркости и подчёркивает локальные яркие структуры, к числу которых относится оптический диск.

Затем проводится анализ гистограмм яркости каналов I_R, I_G, I_B для определения наиболее информативного канала при поиске каждого из ориентиров. Для оптического диска, как наиболее яркой области, используется канал с наибольшей суммой значений в правом хвосте гистограммы: $c_{OD} = \arg \max_c \sum_{k=\lceil \mu_c + \alpha \sigma_c \rceil}^{255} H_c(k)$, где

$H_c(k)$ – гистограмма яркости канала c , μ_c, σ_c – среднее и среднеквадратическое отклонение яркости, α – коэффициент, определяющий зону насыщенных значений. Для фовеа, напротив, выбирается канал с выраженным минимумом яркости в левом хвосте гистограммы: $c_F = \arg \max_c \sum_{k=0}^{\lfloor \mu_c - \alpha \sigma_c \rfloor} H_c(k)$. Такой выбор обеспечивает надёжную дифференциацию между яркой зоной диска и тёмной зоной фовеа.

После определения соответствующих каналов производится поиск координат максимумов и минимумов интенсивности для локализации ориентиров. Положение оптического диска определяется как пик яркости на морфологически скорректированном изображении $I_{c_{OD}}^*$, а фовеа – как локальный минимум яркости в канале I_{c_F} с учётом ожидаемого анатомического расположения (фовеа обычно смещена по горизонтали от центра диска).

На заключительном этапе изображение делится на квадранты относительно найденных ориентиров, что позволяет выполнить топографическую систематизацию признаков и учесть пространственное распределение патологических структур. Разделение поля зрения на зоны необходимо для последующего анализа локализации биомаркеров и построения правил логической классификации стадий диабетической ретинопатии.

Разработанный метод устойчив к артефактам освещённости, варьированию контраста и различиям между устройствами съёмки. Отсутствие процедуры обучения делает его независимым от параметров конкретных датасетов, а использование только базовых морфологических операций обеспечивает воспроизводимость и низкую вычислительную сложность при высоком качестве локализации ориентиров.

Задача семантической сегментации формулируется как отображение $f: R^{H \times W \times 3} \rightarrow \{1, \dots, K\}^{H \times W}$, при котором каждому пикселью входного изображения $I \in R^{H \times W \times 3}$ ставится в соответствие метка класса $k \in \{1, \dots, K\}$, где K – число целевых биомаркеров (например, микроаневризмы, кровоизлияния, твёрдые и мягкие экссудаты).

Для решения задачи выделения патологических областей на фундус-снимках в работе использована архитектура U-Net, специально разработанная для задач плотной пиксельной сегментации медицинских изображений. Её структура основана на симметричной форме типа «энкодер-декодер» с прямыми пропускными связями между соответствующими уровнями свёртывающего и восстанавливающего путей.

На нисходящей (кодирующей) части сети выполняется последовательное извлечение признаков и уменьшение пространственного разрешения изображения. Каждый уровень энкодера включает два последовательных свёрточных слоя с активацией ReLU и слой подвыборки (max-pooling), что позволяет сети постепенно выделять высокоуровневые признаки — сосудистую структуру, зоны отёка, экссудаты и другие морфологические элементы.

В восходящей (декодирующей) части выполняется обратная реконструкция изображения с использованием транспонированных свёрток (upsampling), восстанавливающих пространственные детали. На каждом уровне декодера результаты восстановления объединяются (конкатенируются) с соответствующими картами признаков из энкодера. Эти пропускные связи позволяют сохранять пространственный контекст и улучшать точность локализации границ патологических областей.

На выходе модели располагается свёрточный слой с функцией активации softmax, который присваивает каждому пикселью вероятность принадлежности одному из классов

биомаркеров. Таким образом, U-Net формирует карту сегментации, на которой для каждого пикселя определён соответствующий тип патологического образования.

Для повышения чувствительности модели к ориентированным и полосовым структурам сосудистой сети и мелким патологическим элементам в архитектуру введён блок габоровых фильтров с обучаемыми параметрами. Двумерный фильтр Габора определяется выражением:

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

$$\text{где } x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta.$$

Признак ориентации выбирается как индекс фильтра с минимальным откликом из банка:

$$F(x, y) = \arg \min_j [(I * G_R)(x, y; \lambda_j, \theta_j, \phi, \sigma, \gamma)].$$

Параметры банка фильтров дискретизируются по частотам и ориентациям: $\omega_n = \frac{\pi}{2} 2^{-\frac{n-1}{2}}$, $\theta_m = \frac{\pi}{8}(m-1)$. Модифицированный первый блок энкодера U-Net включает последовательность: GaborConv($k \times k$) \rightarrow BatchNorm \rightarrow ReLU \rightarrow Conv(3×3) с обучаемыми параметрами фильтров $\lambda, \phi, \sigma, \gamma, \theta$. Для них приведены аналитические выражения градиентов, обеспечивающие устойчивое обучение и адаптацию фильтров к геометрии сосудов и текстурным особенностям биомаркеров.

Такое расширение архитектуры позволило повысить чувствительность модели к мелким ориентированным структурам и улучшить качество сегментации по метрике IoU без увеличения числа параметров (Табл. 4, 5).

Таблица 4. Результаты по метрике IoU (%),
± стандартное отклонение

Класс	U-Net	Gabor-U-Net	Прирост
MA	85.6 ± 1.6	88.1 ± 1.5	+2.5
HEM	87.0 ± 1.4	90.0 ± 1.3	+3.0
EX	89.1 ± 1.2	91.6 ± 1.1	+2.5
SE	86.2 ± 1.8	88.8 ± 1.6	+2.6
NV	85.9 ± 1.9	88.3 ± 1.8	+2.4

Таблица 5. Результаты по метрике Dice (%),
± стандартное отклонение

Класс	U-Net	Gabor-U-Net	Прирост
MA	89.0 ± 1.3	91.4 ± 1.0	+2.4

HEM	90.2 ± 1.1	92.0 ± 0.9	+1.8
EX	90.7 ± 1.0	92.0 ± 0.8	+1.3
SE	89.3 ± 1.4	91.1 ± 1.1	+1.7
NV	89.0 ± 1.6	90.9 ± 1.3	+1.9

Согласно данным таблиц 4 и 5, во всех классах наблюдается устойчивое повышение точности по метрикам IoU и Dice, что указывает на более точное совпадение предсказанных масок с эталонными разметками. Средний прирост по метрике IoU составил от 2,4 % до 3,0 %, а по метрике Dice — от 1,3 % до 2,4 %. Наибольшее улучшение достигнуто для классов MA (микроаневризмы) и HEM (кровоизлияния) — структур, характеризующихся малыми размерами и низкой контрастностью.

Для визуализации вкладов признаков используется метод ОИИ Grad-CAM. При активациях $A_k \in R^{h \times w}$ и логите класса y^c :

$$\alpha_k^c = \frac{1}{HW} \sum_{i=1}^h \sum_{j=1}^w \frac{\partial y^c}{\partial A_{k,ij}},$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\sum_k \alpha_k^c A_k).$$

Для попиксельной интерпретации семантической сегментации целевая величина задаётся как: $T = y^c(p^*)$, где $p^* = (x, y)$ — рассматриваемый пиксель. Выбор последнего блока энкодера в качестве целевого слоя обеспечивает баланс между локализацией и устойчивостью объяснений.

Для перехода от семантической сегментации к сегментации экземпляров биомаркеров диабетической ретинопатии применён морфологический метод водораздела. Метод основан на интерпретации карты яркости или градиента изображения как топографической поверхности, на которой линии водораздела формируются вдоль границ между областями локальных минимумов, соответствующих отдельным объектам. Его использование позволяет разделять соприкасающиеся или частично перекрывающиеся области одного класса, формируя независимые экземпляры патологических структур, что обеспечивает корректное количественное определение числа и площади биомаркеров в пределах анатомических зон сетчатки.

В заключительной части главы представлен метод логической классификации стадий диабетической ретинопатии, основанный на формализованных правилах, отражающих клинические критерии распределения патологических признаков по

зонам сетчатки. Предложенный подход использует результаты сегментации экземпляров биомаркеров и обеспечивает интерпретируемое определение стадии заболевания на основе количественного анализа и пространственного распределения выявленных патологий.

Для каждого изображения формируется матрица $N = [N_{c,i}] \in N_{\emptyset}^{C \times 4}$, где $C = 5$ соответствует числу классов патологий: микроаневризмы (MA), кровоизлияния (HEM), твёрдые экссудаты (EX), мягкие экссудаты (SE) и неоваскуляризация (NV). Каждый столбец матрицы соответствует одному из четырёх анатомических квадрантов сетчатки, а элемент $N_{c,i}$ отражает количество объектов класса c в i -й зоне. На основе этой матрицы вычисляются суммарные и локальные показатели распространённости биомаркеров, которые сравниваются с пороговыми значениями $\Theta = \{\theta_{\text{MA}}, \theta_{\text{HEM}}, \theta_{\text{4Q}}, \theta_{\text{NV}}\}$.

Процесс классификации заключается в последовательной проверке логических условий, определяющих принадлежность изображения к одной из трёх стадий заболевания: непролиферативной (NPDR), препролиферативной (PPDR) или пролиферативной (PDR).

Наиболее тяжёлая стадия (PDR) фиксируется при наличии зон неоваскуляризации, число которых превышает порог θ_{NV} . При отсутствии признаков неоваскуляризации, но при множественных кровоизлияниях и микроаневризмах во всех квадрантах, изображение классифицируется как препролиферативная стадия (PPDR). Если выявлены отдельные микроаневризмы или кровоизлияния, количество которых превышает минимальные пороги θ_{MA} или θ_{HEM} , но не удовлетворяет критериям более тяжёлых форм, определяется непролиферативная стадия (NPDR).

Таким образом, метод использует комбинацию количественного и топографического анализа сегментированных биомаркеров, позволяя объективно оценить степень поражения сетчатки. Его преимуществом является интерпретируемость: каждое решение напрямую объясняется конкретными признаками и их распределением, что обеспечивает клиническую прозрачность и совместимость с экспертной оценкой офтальмолога.

В заключении сформулированы основные результаты диссертационного исследования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ

1. Разработана методика формирования специализированных коллекций данных для автоматического анализа ретинальных изображений, опирающаяся на нейросетевую оценку качества, использование объяснительного ИИ для верификации релевантности признаков, интеграцию моделей с визуальными промптами для ускорения предварительной разметки и перекрёстные экспертные аннотации; такая организация конвейера повышает воспроизводимость экспериментов, снижает трудоёмкость ручной разметки и улучшает согласованность данных.
2. Предложен метод автоматического анализа ретинальных изображений, сочетающий сегментацию экземпляров, формализованные логические правила экспертных знаний и средства объяснительного ИИ; переход к экземплярной постановке обеспечил выделение мелких и пересекающихся структур и агрегирование количественных показателей в диагностически значимые заключения.
3. Разработан метод поиска и сегментации анатомических зон сетчатки на фундус-снимках на основе анализа бинаризованных гистограмм цветового канала, характеризующийся нечувствительностью к артефактам и отсутствием необходимости предварительного обучения.
4. Достижение цели диссертационного исследования – создание интерпретируемых и устойчивых методов анализа ретинальных изображений обеспечено комплексным подходом, включающим формирование качественной коллекции медицинских данных, реализацию нейросетевых архитектур с интерпретируемым поведением и логический вывод на основе клинических знаний; предложенные решения ориентированы на практическое применение в системах поддержки принятия врачебных решений и анализа биомедицинских изображений.

ПУБЛИКАЦИИ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

1. Аверкин, А. Н. Объяснительный искусственный интеллект в анализе цифровых изображений на основе нейронных сетей глубокого обучения / А. Н. Аверкин, **Е. Н. Волков**, С. А. Ярушев // Известия Российской академии наук. Теория и системы управления. – 2024. – № 1. – С. 150-178. – DOI: 10.31857/S0002338824010122 (BAK K-1);
Переводная версия статьи: 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. – Vol. 63. – No. 1. – P. 175-203. – DOI: 10.1134/S1064230724700138 (Scopus Q3);
2. Аверкин, А. Н. Гибридный метод анализа изображений на основе технологий искусственного интеллекта и нечетких множеств / А. Н. Аверкин, **Е. Н. Волков**, С. А. Ярушев // Известия Российской академии наук. Теория и системы управления. – 2025. – № 3. – С. 99-112. – DOI: 10.31857/S0002338825030103 (BAK K-1);
Переводная версия статьи: 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. – Vol. 64. – No. 3. – P. 460-473. – DOI: 10.1134/S1064230725700431 (Scopus Q3);
3. **Волков, Е. Н.** Интеграция механизма внимания в свёрточную нейронную сеть для улучшения результатов классификации офтальмологических заболеваний / Е. Н. Волков, А. Н. Аверкин, С. А. Ярушев // Мягкие измерения и вычисления. – 2024. – Т. 80, № 7. – С. 105-114. – DOI: 10.36871/2618-9976.2024.07.009. (BAK K-3);
4. **Волков, Е. Н.** Оценка возможностей применения метода объяснительного ии Grad-CAM в задаче нейросетевой классификации стадий диабетической ретинопатии на снимках фундус-камеры / Е. Н. Волков, А. Н. Аверкин // Мягкие измерения и вычисления. – 2024. – Т. 85, № 12. – С. 72-81. – DOI: 10.36871/2618-9976.2024.12.007 (BAK K-3);
5. Перспективы применения объяснительного искусственного интеллекта в системах поддержки принятия врачебных решений на основе искусственных нейросетей / С. А. Ярушев, А. Н. Аверкин, **Е. Н. Волков**, А. Н. Лукьянов // Мягкие измерения и

- вычисления. – 2023. – Т. 70, № 9-1. – С. 56-67. – DOI: 10.36871/2618-9976.2023.09.006 (BAK K-3);
6. Averkin A. N., **Volkov E. N.**, Yarushev S. A. Possibilities of Application of Neuro-Fuzzy Networks for Ophthalmologic Image Classification //Pattern Recognition and Image Analysis. – 2024. – Vol. 34. – No. 3. – P. 610-616. – DOI: 10.1134/S1054661824700421 (Scopus Q3);
 7. **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. – P. 130-133. – DOI: 10.1109/SCM58628.2023.10159038 (Scopus);
 8. **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. – P. 134-137. – DOI: 10.1109/SCM58628.2023.10159033 (Scopus);
 9. **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. – P. 3-6. – DOI: 10.1109/NeuroNT58640.2023.10175852 (Scopus);
 10. **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. – P. 6-9. – DOI: 10.1109/NeuroNT58640.2023.10175855 (Scopus);
 11. **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. – P. 50-54. – DOI: 10.1109/IPS62349.2024.10499480 (Scopus);
 12. **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. – P. 428-431. – DOI: 10.1109/SCM62608.2024.10554254 (Scopus);
 13. **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. – P. 189-192. – DOI: 10.1109/SCM62608.2024.10554222 (Scopus);
- 14. **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. – P. 1-4. – DOI: 10.1109/NeuroNT62606.2024.10585353 (Scopus);
 - 15. **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. – P. 1-4. – DOI: 10.1109/DSPA64310.2025.10977883 (Scopus);
 - 16. **Volkov E.**, Sechin V., Averkin A. Visual-Language Model Fine-Tuning via LoRA for Structed Medical Reports Generating for Lung X-Ray Skans //2025 XXVIII International Conference on Soft Computing and Measurements (SCM). – IEEE, 2025. – P. 438-442. – DOI: 10.1109/SCM66446.2025.11060140 (Scopus);
 - 17. 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. – P. 171-173. – DOI: 10.1109/SCM66446.2025.11060242 (Scopus);
 - 18. **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. – P. 66-68. – DOI: 10.1109/NeuroNT66873.2025.11049967 (Scopus).