

Методика адаптации алгоритмов глубокого обучения для скрининга диабетической ретинопатии в условиях реальной клинической практики

Е.В. Козлов¹, Д.Д. Лысухин¹, А.П. Першина-Милюткина¹, Е.В. Ковалева¹, А.В. Аредов¹, В.К. Александрова¹,
Т.А. Чистяков¹, А.А. Толкачева¹, Н.Г. Шебардина¹, Е.А. Корчуганова¹, Н.Г. Мокрышева¹

¹ ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России,
117292, Россия, г. Москва, ул. Дмитрия Ульянова, д. 11

Аннотация

Применение систем искусственного интеллекта для скрининга диабетической ретинопатии в реальной клинической практике ограничено проблемой снижения точности алгоритмов при переносе на новые источники данных. Целью данной работы является разработка и валидация методики адаптации предобученных моделей к условиям гетерогенности медицинских изображений. В исследовании проанализировано влияние различий в характеристиках наборов данных на качество классификации. Показано, что при прямом переносе модели с открытых баз данных на локальный клинический набор (1123 снимка, 302 пациента), отличающийся техническими параметрами снимков и выраженным дисбалансом классов, метрика maseo-F1 снижается с 0,86 до 0,38.

Для обеспечения корректного переноса знаний модели и минимизации шума при обучении ключевым этапом работы стало формирование качественного эталонного локального набора данных. Для этого был применен многоступенчатый протокол аннотирования и оценки согласованности врачей-экспертов. На основе подготовленных данных была реализована двухэтапная стратегия трансферного обучения (архитектура EfficientNet-B0). Несмотря на ограниченный объем выборки, данная схема разметки в сочетании с предложенным подходом к дообучению позволила повысить метрику maseo-F1 на валидационной выборке до 0,63. Результаты подтверждают, что создание качественно размеченных локальных датасетов является критически важным шагом для успешной адаптации диагностических систем к условиям практического здравоохранения.

Ключевые слова: диабетическая ретинопатия, глубокое обучение, перенос обучения, ограниченные данные, автоматический скрининг.

Цитирование: Козлов, Е.В. Методика адаптации алгоритмов глубокого обучения для скрининга диабетической ретинопатии в условиях реальной клинической практики / Е.В. Козлов, Д.Д. Лысухин, А.П. Першина-Милюткина, Е.В. Ковалева, А.В. Аредов, В.К. Александрова, Т.А. Чистяков, А.А. Толкачева, Н.Г. Шебардина, Е.А. Корчуганова, Н.Г. Мокрышева // Компьютерная оптика. – 2026. – Т. 50, № 4. – 1839. – doi: 10.18287/COJ1839.

Citation: Kozlov EV, Lysukhin DD, Pershina-Milyutina AP, Kovaleva EV, Aredov AV, Aleksandrova VK, Chistyakov TA, Tolкачева AA, Shebardina NG, Korchuganova EA, Mokrysheva NG. Deep Learning for Diabetic Retinopathy Screening: An Adaptation Methodology for Real-World Clinical Environments. Computer Optics 2026; 50(4): 1839. doi: 10.18287/COJ1839.

Введение

Сахарный диабет (СД) признан глобальной эпидемией: в РФ число пациентов выросло с 3,94 млн в 2013 г. до 5,55 млн в 2023 г. (+40 %) [1].

Диабетическая ретинопатия (ДР) – частое осложнение СД и ведущая причина слепоты. Распространённость ДР в РФ составляет 28,9 % при СД 1 типа и 12,3 % при СД 2 типа, что соответствует более 700 тыс. пациентов [2].

Доступ к офтальмологическому скринингу в РФ неравномерен: плотность кабинетов варьируется от 1–2 до 10–15 на 100 000 населения, из-за чего до 40 % пациентов пропускают ежегодные осмотры и ДР выявляется на поздних стадиях. Отсутствие единых платформ и дефицит кадров усугубляют ситуацию [2].

На сегодняшний день для автоматизированного выявления и стадирования ДР создано множество систем, применяющих различные методы компьютерного зрения и искусственного интеллекта. Среди наиболее известных международных решений – RetmarkerDR [3], EyeArt [4], IDx-DR [5] (первая система, одобренная FDA для автономной диагностики ДР), система от команды Google DeepMind [6], а также крупные коммерческие платформы IRIS [7] и SERI-NUS [8], успешно внедрённые в рутинную клиническую практику за рубежом. Современные алгоритмы этих систем построены на глубоких сверточных нейронных сетях, что обеспечивает им высокие значения чувствительности на тестовых датасетах (до 96–97 %) и специфичности (до 90–95 %) [9, 10].

Однако при переносе в реальную клиническую практику точность моделей падает из-за смещения доменов (domain shift), вызванного различиями в оборудовании, протоколах съёмки и популяциях [11 – 13]. Это требует разработки методов адаптации моделей к локальным условиям [14 – 17].

В последние годы набирают популярность гибридные решения, сочетающие сегментацию ключевых патологических признаков (микроаневризмы, кровоизлияния, экссудаты) с последующей интерпретацией по экспертным правилам, а также системы с элементами контроля качества снимка и интеллектуальными подсказками для пользователя при съёмке. Российские коллективы также демонстрируют высокий уровень решений. Разработаны прототипы сервисов и сегментационные алгоритмы, достигающие чувствительности 85 – 92 % на ограниченных по объёму и гетерогенных по происхождению выборках, полученных в условиях российских медицинских учреждений. Также исследователи отмечают [18], что все существующие системы критически зависят от используемых датасетов и специфики оборудования: перенос моделей между различными клиниками и аппаратами часто сопровождается падением качества предсказаний из-за смещения целевого домена (англ. domain shift). Это подчёркивает актуальность подходов, учитывающих локальные особенности данных, прозрачную интерпретируемость и возможность масштабируемой адаптации к условиям практического здравоохранения.

Главной задачей разрабатываемой системы является автоматизированное определение степени тяжести ДР по изображениям, полученным с фундус-камеры. Для решения этой задачи были рассмотрены два принципиально разных подхода. Первый подход заключается в использовании end-to-end нейросетевого классификатора, который напрямую определяет стадию ДР на основании исходного снимка. Основным преимуществом этого подхода является его технологическая простота и ускоренное внедрение. Однако очевидным недостатком данного подхода является ограниченная интерпретируемость результатов его работы. Модель функционирует как «чёрный ящик», что снижает доверие к ней со стороны клиницистов, а также затрудняет анализ ошибок. Второй подход декомпозирует задачу на 2 этапа: сначала выполняется сегментация ключевых признаков ДР (микроаневризмы, кровоизлияния, экссудаты и др.), а затем на основе результатов сегментации стадия заболевания определяется с помощью экспертных правил. Такой подход обеспечивает максимальную прозрачность и интерпретируемость, позволяя визуализировать и объяснить, какие именно признаки повлияли на итоговое решение системы.

Поскольку процесс аннотации исходных данных для решения задачи сегментации требует существенно больше времени, а также на порядок сложнее, было принято решение о поэтапной разработке. На первом этапе реализуется система, основанная на нейросетевой классификации степени тяжести ДР. Затем планируется расширение функционала за счет создания модуля, решающего задачу сегментации, и перехода к определению стадии заболевания на основе экспертных правил.

Для предобучения использовались открытые наборы EyePACS и APTOS [19], полученные на фундус-камерах Canon. В РФ распространены иные модели (Kowa, Visucam), что ведёт к различиям в параметрах изображений. Кроме того, в самих открытых данных ограничена согласованность разметки (каппа Коэна 0,71 [6]). Эти факторы вызывают резкое падение качества при прямом переносе модели: на нашем внутреннем наборе macro-F1 упала с 0,86 до 0,38.

Материалы и методы

В обучении использовались открытые (APTOS и EyePACS) и внутренний наборы данных ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России (табл. 1).

Табл. 1. Характеристики открытых наборов данных

Название датасета	Количество снимков	Фундус-камера	Разрешение снимков, сжатие JPEG
APTOS-2019	3662	Canon CR-2	Минимальное: 358×474 Максимальное: 2848×4288 Медианное: 1536×2048
EyePACS	35126	Canon CR-DGi, Canon CR-1	Минимальное: 289×433 Максимальное: 3456×5186 Медианное: 2582×3888

В обоих открытых наборах данных (EyePACS и APTOS) применяется единая пятиуровневая шкала для классификации стадий ДР:

- 0 – No DR (нет ДР);
- 1 – Mild (легкая непролиферативная ДР);
- 2 – Moderate (умеренная непролиферативная ДР);
- 3 – Severe (тяжелая непролиферативная ДР);
- 4 – Proliferative DR (пролиферативная ДР).

Важно упомянуть то, что изображения из открытых наборов данных были соединены в единый набор и заново стратифицированы на обучающую, валидационную и тестовую подвыборки без учета исходного разделения, предусмотренного в соревнованиях Kaggle. Итоговая стратификация представлена на рис. 1.

Во внутреннем же датасете используется разделение из атласа ДР [20] на следующие стадии:

- 0 – отсутствие ДР;
- 1 – непролиферативная стадия ДР;
- 2 – препролиферативная ДР;
- 3 – пролиферативная ДР;
- 4 – непригодная фотография (фотография, на которой невозможно оценить стадию ДР из-за её качества).

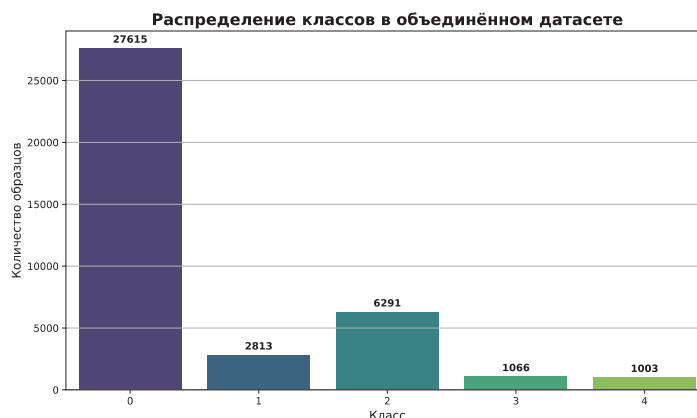


Рис. 1. Статистика классов в открытых датасетах

Внутренний набор исходных данных включал около 13 000 снимков глазного дна с привязкой к уникальному идентификатору пациента. Изображения были получены с помощью фундус-камеры VISUCAM 524 (ZEISS) с 24-мегапиксельной матрицей в разрешении 3608×3608. Исходный набор данных включал изображения, содержащие только один цветовой канал, и многопольные снимки, полученные при объединении изображений отдельных полей глазного дна. В ходе предварительной фильтрации эти изображения были удалены из общего набора, для аннотации осталось около 8 тыс. фотографий. Для достижения более равномерного распределения классов в итоговом датасете с целью дальнейшего аннотирования приоритетно отбирались снимки пациентов, у которых, согласно базе данных медицинской информационной системы, ранее были диагностированы редко встречающиеся стадии ДР. Также для повышения разнообразия пространства признаков количество снимков от одного пациента, включенных в датасет, было ограничено.

Всего в итоговом датасете было 1123 снимка. Распределение классов было следующим: 108 – отсутствие ретинопатии (9,6 %), 452 – непролиферативная стадия (40,2 %), 143 – препролиферативная стадия (12,7 %), 322 – пролиферативная стадия (28,7 %), 98 непригодных фотографий (8,7 %).



Рис. 2. Примеры изображений глазного дна на разных стадиях диабетической ретинопатии:

а) здоровое глазное дно, б) непролиферативная стадия, в) препролиферативная стадия, г) пролиферативная стадия

Протокол аннотирования и оценка согласованности

Аннотирование изображений проходило в развернутом локально веб-приложении CVAT (Computer Vision Annotation Tool). В процессе аннотирования снимков принимало участие 3 врача-офтальмолога ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Для дальнейших действий врачи были разделены на 2 группы, каждый специалист был «ослеплен», то есть не знал о том, какую стадию поставили его коллеги. Перед началом аннотирования каждому врачу была выдана инструкция, включавшая принятое в проекте разделение признаков ДР на глазном дне по классам и иллюстрированный атлас референс-снимков. Процесс разметки изображений строился по следующему протоколу. На первом этапе снимок параллельно назначался двум офтальмологам из разных групп. Если оценки стадии ДР совпадали, данная стадия сохранялась в наборе данных. В случае расхождения снимок назначался третьему специалисту для получения дополнительной оценки. Оценка третьего врача считалась решающей. Если же оценки всех трех специалистов различались, изображение выносилось на очные обсуждения для достижения коллективного консенсуса (рис. 4).

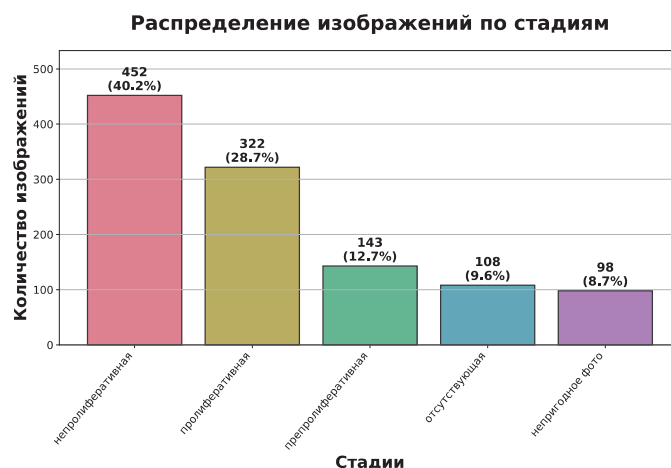


Рис. 3. Статистика классов во внутреннем наборе

Администрирование процесса разметки и проверки снимков

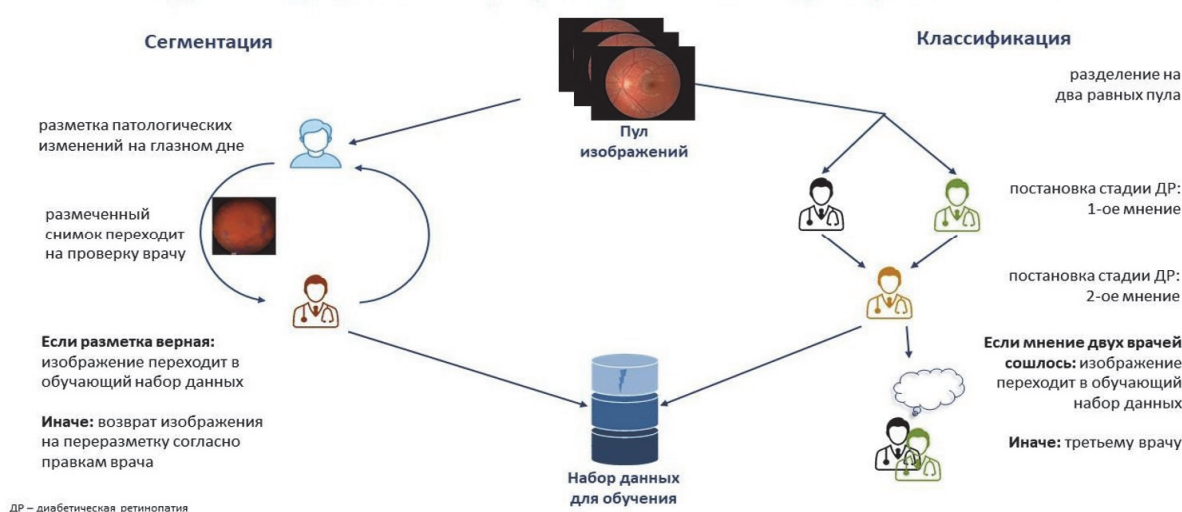


Рис. 4. Схема аннотации и валидации снимков в CVAT

Для количественной оценки качества и надежности разметки данных был проведен анализ внутренней и межэкспертной согласованности работы всех офтальмологов.

Каждый врач вслепую повторно аннотировал 50 изображений, случайным образом выбранных из ранее размеченного им же массива снимков, при этом офтальмологи не были проинформированы, что эти изображения они аннотируют повторно. Этот этап фундаментален для создания собственного набора данных, свободного от экспертной предвзятости и пригодного для воспроизводимого, надежного обучения. Степень внутренней и межэкспертной согласованности оценивалась с помощью коэффициента каппа Коэна.

Анализ внутренней согласованности заключался в повторном слепом аннотировании каждым из трех врачей-офтальмологов выборки из 50 изображений, которые они размечали ранее. Полученные значения показали достаточно серьезные проблемы в разметке и плохую повторяемость разметки экспертами: $k_1 = 0,56$; $k_2 = 0,24$; $k_3 = 0,50$.

Оценка межэкспертной согласованности проводилась на отдельной выборке из 100 снимков, которые были независимо размечены тремя специалистами. Результаты межэкспертной согласованности, так же как и внутренней согласованности, выявили серьезные проблемы в разметке: сильную субъективность экспертов при разметке и плохо составленные инструкции для размечающих специалистов. Значения для пар врачей составили: $k_1 = 0,34$; $k_2 = 0,10$; $k_3 = 0,45$.

Для повышения однородности и согласованности разметки были организованы коллективные обсуждения. На них анализировались случаи с максимальным расхождением в оценках (например, оценки специалистов различались более чем на одну стадию ДР). Целью обсуждения было выработать и формализовать единые, более строгие критерии аннотирования для повышения качества аннотирования изображений в дальнейшем.

Первые эксперименты по обучению модели показали, что классическое деление датасета на обучающую, валидационную и тестовую выборки при небольшом объеме данных приводит к сильному зашумлению на кривых обучения и затрудняет интерпретацию результатов. В связи с этим в финальной схеме датасет был разделен только на обучающую и валидационную выборки.

Архитектура модели и предобработка

Для увеличения вариативности набора данных и повышения обобщающей способности модели использовались стандартные аугментации: случайные повороты, вертикальные и горизонтальные отражения, эластичные деформации изображения. Реализация данных преобразований была основана на библиотеке Albumentations [21].

В качестве архитектурной основы для решения задачи классификации стадий ДР и последующего применения в качестве энкодера сегментирующей модели была выбрана EfficientNet-B0 [22]. Данная модель содержит 6,5 миллионов обучаемых параметров и является самой компактной в семействе EfficientNet. Такой выбор обусловлен минимизацией риска переобучения, возникающего при обучении более сложных моделей на весьма ограниченном внутреннем наборе данных.

Обучение модели проводилось на сервере с графическим ускорителем NVIDIA RTX 4070 (12 Гб видеопамети). При размере изображения 1024×1024 пикселя и размере обучающего батча, равном 16, среднее время прохождения одной эпохи дообучения составляло ~30 секунд. Данный показатель включает как этап инференса, так и этап обратного распространения ошибки. Важно отметить, что разрабатываемая система ориентирована на офлайн-обработку и не требует выполнения инференса в реальном времени, поэтому к вычислительным задержкам не предъявляются жесткие требования.

Протокол обучения и оценка качества

Процесс обучения модели был разделен на 2 этапа в соответствии со стратегией transfer learning. На первом этапе проводилось предварительное обучение модели на объединенном наборе открытых данных с инициализацией весов на базе ImageNet. Задачей данного этапа являлось научить модель выделять признаки, характерные для ДР.

Затем на внутреннем наборе выполнялось дообучение: последний полносвязный слой заменялся на новый (5 классов, инициализация Xavier [23]), веса остальных слоёв фиксировались, кроме нескольких последних, число которых подбиралось экспериментально.

На обоих этапах применялся оптимизатор AdamW и механизм ранней остановки (по значению loss на предобучении, по F1 на дообучении). Для компенсации дисбаланса классов использовался focal loss.

В связи с ограниченностью вычислительных ресурсов обучение проводилось на кропах изображений глазного дна, в то время как валидация результатов осуществлялась на полноразмерных изображениях. В результате экспериментов оптимальным было признано разрешение изображений 1024×1024 пикселя по критерию баланса между качеством и производительностью обучения. Размер батча составлял 2 на обучающей выборке и 1 – на валидационной.

В качестве основной метрики использовался maseo-F1, так как он одинаково учитывает все классы и не завышает оценку за счёт часто встречающихся стадий.

Результаты и обсуждение

В результате проведённого обучения оптимальная версия модели EfficientNet-B0, дообученная на внутреннем наборе данных, продемонстрировала следующие показатели качества на валидационной выборке полноразмерных изображений глазного дна. Значение основной метрики качества – maseo-F1-score – составило 0,63. Это свидетельствует об умеренной способности модели корректно классифицировать все стадии ДР. Однако результаты выявляют неравномерность качества предсказания по отдельным классам.

Табл. 2. Лучшие результаты обучения модели на валидационной выборке

Название класса	F1-метрика	Количество примеров в наборе данных
Отсутствие ДР	0,71	108 изображений
Непролиферативная ДР	0,77	452 изображения
Препролиферативная ДР	0,14	143 изображения
Пролиферативная ДР	0,70	322 изображения
Непригодное изображение	0,91	98 изображений
Среднее для всех классов	0,63	1123 изображения (302 пациента)

Модель уверенно справилась с определением классов: «Отсутствие ДР», «Непролиферативная ДР», «Пролиферативная ДР», а также отлично отбраковывала снимки низкого качества («Непригодное изображение»). Однако на классе «Препролиферативная ДР» её F1-метрика оказалась крайне низкой. Основная причина, которую хочется выделить, – размытые диагностические границы препролиферативной стадии. По одиночному изображению характерные признаки данной стадии трудноотличимы от соседних стадий, что усиливает субъективность оценки характерных признаков и только усугубляет проблему выделения данной стадии ДР. Данная проблема неоднократно обсуждалась на клинических консилиумах, но чёткий консенсус так и не был достигнут, что напрямую отразилось на F1-метрике модели.

Для оценки эффективности предложенной двухэтапной стратегии проведено сравнение с базовыми подходами (baseline). Прямой перенос модели, обученной только на открытых данных, на внутренний набор дал $\text{macro-F1} = 0,38$, что подтверждает выраженный domain shift. Обучение «с нуля» только на внутренних данных привело к переобучению и $\text{macro-F1} \approx 0,61$.

Дополнительные эксперименты показали, что обучение на подвыборках, размеченных лишь одним врачом, ухудшает обобщающую способность по сравнению с использованием согласованной разметки. Наилучший результат достигнут при обучении на 522 снимках с консенсусной оценкой; модели, обученные на 427, 319 и 285 снимках от отдельных специалистов, показали более низкие значения macro-F1 (рис. 5).

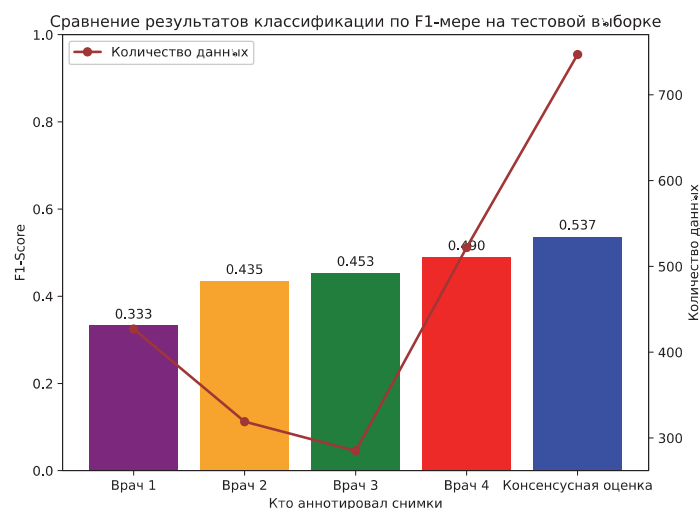


Рис. 5. График результатов моделей, обученных на отдельных выборках каждого размечающего врача

Заключение

В работе представлен полный цикл разработки и валидации системы автоматизированной классификации стадий ДР, адаптированной к условиям российской клинической практики. Научная новизна исследования заключается в реализации жесткого трехэтапного протокола слепого консенсусного аннотирования данных. Показано, что без применения данного подхода формирование эталонного набора данных на основе локальной выборки (1123 изображения) невозможно из-за высокой межэкспертной вариабельности и зашумленности разметки. Созданный датасет позволил реализовать двухэтапную стратегию трансферного обучения на архитектуре EfficientNet-B0, обеспечившую устойчивость модели к техническим особенностям оборудования.

Ключевым этапом работы стала оценка эффекта «смещения домена». При сравнении полученных результатов с коммерческими системами (IDx-DR, EyeArt) и отечественными прототипами (например, работы Нероева В.В. и соавт.), декларирующими чувствительность на уровне 85 – 92 %, следует учитывать различия в условиях тестирования. Высокие метрики аналогов зачастую достигаются на гомогенных, предварительно отфильтрованных выборках. В то же время полученное значение macro-F1 является верифицированной оценкой работы системы в жестких условиях реальных снимков клинического потока с выраженным дисбалансом классов и вариативным качеством снимков. Таким образом, разработанное решение представляет собой валидированный инструмент, готовый к опытной эксплуатации для задач офлайн-скрининга.

В дальнейших работах планируется расширение датасета за счёт многоцентрового сбора снимков с редко встречающимися стадиями ДР, внедрение гибридной схемы «сегментация + экспертные правила» для повышения интерпретируемости, а также проведение клинического испытания в реальных условиях.

Таким образом, продемонстрировано, что сочетание открытых и локальных данных при строгом протоколе аннотации позволяет создать надёжный инструмент для скрининга ДР, адаптированный к техническим и клиническим условиям РФ.

References

- [1] Federal State Budgetary Institution “Central Research Institute for Organization and Informatization of Health Care” of the Ministry of Health of the Russian Federation (FSBI “CNIOIZ” of the Ministry of Health of Russia). Medical statistics [in Russian]. 2025. Available at: <https://mednet.ru/medicinskaya-statistika>.
- [2] Dedov II, Shestakova MV, Vikulova OK, Zheleznyakova AV, Isakov MA, Eremina IA, Galstyan GR. Diabetes mellitus in the Russian Federation: Dynamics of epidemiological indicators according to the Federal Register of Diabetes Mellitus for the period 2010–2022 [in Russian]. Diabetes Mellitus 2023; 26(2): 104-123. DOI: 10.14341/DM13035.
- [3] Ribeiro L, Oliveira CM, Neves C, Ramos JD, Ferreira H, Cunha-Vaz J. Screening for diabetic retinopathy in the central region of Portugal: Added value of automated “disease/no disease” grading. Ophthalmologica 2015; 233(2): 96-103. DOI: 10.1159/000368426.

- [4] Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda S, Solanki K. The value of automated diabetic retinopathy screening with the EyeArt system: A study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technology & Therapeutics* 2019; 21(11): 635-643. DOI: 10.1089/dia.2019.0164.
- [5] Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine* 2018; 1: 39. DOI: 10.1038/s41746-018-0040-6.
- [6] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316(22): 2402-2410. DOI: 10.1001/jama.2016.17216.
- [7] Walton OB, Garoon RB, Weng CY, Gross J, Young AK, Camero KA, et al. Evaluation of an automated teleretinal screening program for diabetic retinopathy. *JAMA Ophthalmology* 2016; 134(2): 204-209. DOI: 10.1001/jamaophthalmol.2015.5083.
- [8] Ting DSW, Cheung CYL, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, Yeo IYS, Lee SY, Wong EYL, Sabanayagam C, Baskaran M, Ibrahim F, Tan NCS, Finkelstein EA, Lamoureux EL, Wong IYH, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017; 318(22): 2211-2223. DOI: 10.1001/jama.2017.18152.
- [9] Zhang Y, Wang L, Wu X, Chen H, Li Y, Wang S. A cross-domain weakly supervised diabetic retinopathy lesion identification method based on multiple instance learning and domain adaptation. *Bioengineering* 2023; 10(9): 1100. DOI: 10.3390/bioengineering10091100.
- [10] Wang Z, Yin Y, Shi J, Fang W, Li H. Multi-model domain adaptation for diabetic retinopathy classification. *Frontiers in Physiology* 2022; 13: 918929. DOI: 10.3389/fphys.2022.918929.
- [11] Benouaich X, Soares M, Azrak C, Hajjar C, Weber M, Kodjikian L. Accuracy of autonomous artificial intelligence-based diabetic retinopathy screening in real-life clinical practice. *Journal of Clinical Medicine* 2024; 13(16): 4776. DOI: 10.3390/jcm13164776.
- [12] Li H, Liu P, Guo J, Wang Y, Liang D, Chen H. Cross-modality transfer learning with knowledge infusion for diabetic retinopathy grading. *Frontiers in Medicine (Lausanne)* 2024; 11: 1400137. DOI: 10.3389/fmed.2024.1400137.
- [13] Zhang K, Liu X, Xu J, Wang H, Li Y, Zhao S. Generalizing to unseen domains in diabetic retinopathy with disentangled representations. *arXiv preprint arXiv:2406.06384*; 2024. Available at: <https://arxiv.org/abs/2406.06384>.
- [14] Sarki R, Ahmed K, Wang H, Zhang Y, Wang K. A systematic review of transfer learning-based approaches for diabetic retinopathy detection. *Artificial Intelligence in Medicine* 2022; 134: 102513. DOI: 10.1016/j.artmed.2022.102513.
- [15] Oltu B, Karaca BK, Erdem H, Özgür A. A systematic review of transfer learning-based approaches for diabetic retinopathy detection. *Gazi University Journal of Science* 2023; 36(3): 1140-1157. DOI: 10.35378/gujs.1081546.computeroptics+1
- [16] Sharma S, Gupta S, Gupta D, Arora C, Vyas R. Learning self-supervised representations for label efficient cross-domain knowledge transfer on diabetic retinopathy fundus images. *arXiv preprint arXiv:2304.11168*; 2023. Available at: <https://arxiv.org/abs/2304.11168>.
- [17] Su W, Tang S, Liu X, Zhang Y, Chen H. Domain adaptive diabetic retinopathy grading with model absence and flowing data. *arXiv preprint arXiv:2412.01203*; 2022. Available at: <https://arxiv.org/abs/2412.01203>.
- [18] Neroev VV, Bragin AA, Zaitseva OV. Development of a prototype service for diagnosing diabetic retinopathy based on fundus images using artificial intelligence methods [in Russian]. *National Health Care (Russia)* 2021; 2(2): 64-72. DOI: 10.47093/2713-069X.2021.2.2.64-72.
- [19] APTOS 2019 blindness detection. Kaggle competition; The Asia Pacific Tele-Ophthalmology Society; 2019. Available at: <https://www.kaggle.com/competitions/aptos2019-blindness-detection>.
- [20] Lipatov DV, Bessmertnaya EG, Kuzmin AG, Tolkacheva AA, Chistyakov TA. Atlas of diabetic retinopathy: A practical guide for physicians [in Russian]. In: Dedov II, Shestakova MV, eds. Moscow: Medical Information Agency; 2017. ISBN: 978-5-9986-0275-8.
- [21] Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and flexible image augmentations. *Information* 2020; 11(2): 125. DOI: 10.3390/info11020125.
- [22] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*. Proceedings of Machine Learning Research 2019; 97: 6105-6114.
- [23] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*. Journal of Machine Learning Research – Workshop and Conference Proceedings 2010; 9: 249-256.

Сведения об авторах

Козлов Егор Васильевич, 2002 года рождения, программист, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: применение искусственного интеллекта в медицине, компьютерное зрение. E-mail: kozlov.egor@endocrincentr.ru ORCID: 0009-0008-2206-1505

Лысухин Даниил Дмитриевич, 1993 года рождения, программист, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: машинное обучение, применение ИИ в медицине, компьютерное зрение. E-mail: lyukhin.daniil@endocrincentr.ru ORCID: 0009-0008-2158-2624

Першина-Милютин Анастасия Павловна, 1998 года рождения, руководитель группы управления данными – врач-статистик ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: системы поддержки принятия клинических решений, экспертные системы, медицинская статистика, эндокринология. E-mail: milyutina.anastasiya@endocrincentr.ru ORCID: 0000-0002-9462-8522

Ковалева Елена Владимировна, 1992 года рождения, кандидат медицинских наук, руководитель отдела цифровой трансформации - врач-эндокринолог, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: гипопаратиреоз, первичный гиперпаратиреоз, вторичный гиперпаратиреоз различного генеза, нарушения минерального обмена во время беременности, остеопороз, заболевания щитовидной железы. E-mail: kovaleva.elena@endocrincentr.ru ORCID: 0000-0002-9258-2591

Аредов Алексей Вячеславович, 2004 года рождения, аналитик данных, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: системы поддержки принятия клинических решений, медицинская статистика, машинное обучение. E-mail: aredov.aleksey@endocrincentr.ru ORCID: 0009-0007-1336-4152

Александрова Вера Кирилловна, 1974 года рождения, кандидат медицинских наук, врач-офтальмолог, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: диабетическая ретинопатия, офтальмологические осложнения при эндокринологических заболеваниях. E-mail: aleksandrova.vera@endocrincentr.ru

Чистяков Тимофей Александрович, 1975 года рождения, врач-офтальмолог, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: диабетическая ретинопатия, офтальмологические осложнения при эндокринологических заболеваниях. E-mail: Chistyakov.Timofey@endocrincentr.ru

Толкачева Анна Анатольевна, 1971 года рождения, врач-офтальмолог, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: диабетическая ретинопатия, эндокринная офтальмопатия. E-mail: Tolkacheva.Anna@endocrincentr.ru

Шебардина Наталья Геннадьевна, 1999 года рождения, младший научный сотрудник, Научно-исследовательский институт физико-химической биологии имени А.Н. Белозерского МГУ им. М.В. Ломоносова. Область научных интересов: регенеративная терапия, клеточные технологии, диабетология, офтальмология. E-mail: Shebardina.Natalia@endocrincentr.ru ORCID: 0009-0007-3372-5015

Корчуганова Елена Александровна, 1973 года рождения, доктор медицинских наук, заведующий лечебно-диагностическим отделением диабетической ретинопатии и офтальмохирургии, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: диабетическая ретинопатия, офтальмологические осложнения при эндокринологических заболеваниях, организация здравоохранения. E-mail: korchuganova.elena@endocrincentr.ru ORCID: 0000-0001-5147-0385

Мокрышева Наталья Георгиевна, 1970 года рождения, доктор медицинских наук, директор, ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. Область научных интересов: эндокринология, организация здравоохранения. E-mail: mokrisheva.natalia@endocrincentr.ru ORCID: 0000-0002-9717-9742

Поступила в редакцию 18 сентября 2025 г.



Окончательный вариант – 16 апреля 2026 г.

Deep Learning for Diabetic Retinopathy Screening: An Adaptation Methodology for Real-World Clinical Environments

*E.V. Kozlov¹, D.D. Lysukhin¹, A.P. Pershina-Milyutina¹, E.V. Kovaleva¹, A.V. Aredov¹, V.K. Aleksandrova¹,
T.A. Chistyakov¹, A.A. Tolkacheva¹, N.G. Shebardina¹, E.A. Korchuganova¹, N.G. Mokrysheva¹*
*¹I.I. Dedov National Medical Research Center for Endocrinology, Ministry of Health of the Russian Federation,
Dmitry Ulyanov Str. 11, Moscow, 117292, Russia*

Abstract

The deployment of deep learning systems for diabetic retinopathy screening in real-world clinical settings is limited by performance degradation when algorithms are applied to novel data sources. This study aims to develop and validate a methodology for adapting pre-trained models to the heterogeneity inherent in medical imaging data. The research analyzes the impact of dataset disparities on classification performance. It is demonstrated that the direct transfer of a model trained on public databases to a local clinical dataset (1,123 images, 302 patients)—characterized by distinct technical parameters and pronounced class imbalance—results in a significant decline in the macro-F1 score from 0.86 to 0.38.

To ensure robust knowledge transfer and minimize training noise, a critical stage of this work involved constructing a high-quality reference local dataset. This was achieved through a multi-stage annotation protocol and an assessment of inter-rater agreement among expert physicians. Utilizing the curated data, a two-stage transfer learning strategy was implemented using the EfficientNet-B0 architecture. Despite the limited sample size, this annotation scheme, combined with the proposed fine-tuning approach, improved the macro-F1 score on the validation set to 0.63. The results confirm that the creation of annotated local datasets is a critical step for the successful adaptation of diagnostic systems to practical healthcare settings.

Keywords: diabetic retinopathy, deep learning, transfer learning, limited data, automated screening.

Citation: Kozlov EV, Lysukhin DD, Pershina-Milyutina AP, Kovaleva EV, Aredov AV, Aleksandrova VK, Chistyakov TA, Tolkacheva AA, Shebardina NG, Korchuganova EA, Mokrysheva NG. Deep Learning for Diabetic Retinopathy Screening: An Adaptation Methodology for Real-World Clinical Environments. *Computer Optics* 2026; 50(4): 1839. doi: 10.18287/COJ1839.

Author's information

Egor Vasilievich Kozlov (b. 2002), programmer, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: application of AI in medicine, computer vision. E-mail: kozlov.egor@endocrincentr.ru ORCID: 0009-0008-2206-1505

Daniil Dmitrievich Lysukhin (b. 1993), programmer, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: machine learning, application of AI in medicine, computer vision. E-mail: lysukhin.daniil@endocrincentr.ru ORCID: 0009-0008-2158-2624

Anastasia Pavlovna Pershina-Milyutina (b. 1998), Head of the Data Management Group — physician-statistician, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: clinical decision support systems, expert systems, medical statistics, endocrinology. E-mail: milyutina.anastasiya@endocrincentr.ru ORCID: 0000-0002-9462-8522

Elena Vladimirovna Kovaleva (b. 1992), Candidate of Medical Sciences, Head of the Digital Transformation Department — endocrinologist, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: hypoparathyroidism, primary hyperparathyroidism, secondary hyperparathyroidism of various origins, mineral metabolism disorders during pregnancy, osteoporosis, thyroid diseases. E-mail: kovaleva.elena@endocrincentr.ru ORCID: 0000-0002-9258-2591

Alexey Vyacheslavovich Aredov (b. 2004), data analyst, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: clinical decision support systems, medical statistics, machine learning. E-mail: aredov.aleksey@endocrincentr.ru ORCID: 0009-0007-1336-4152

Vera Kirillovna Aleksandrova (b. 1974), Candidate of Medical Sciences, ophthalmologist, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation.

Research interests: diabetic retinopathy, ophthalmic complications in endocrine diseases.

E-mail: aleksandrova.vera@endocrincentr.ru

Timofey Alexandrovich Chistyakov (b. 1975), ophthalmologist, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: diabetic retinopathy, ophthalmic complications in endocrine diseases.

E-mail: Chistyakov.Timofey@endocrincentr.ru

Anna Anatolyevna Tolkacheva (b. 1971), ophthalmologist, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: diabetic retinopathy, endocrine ophthalmopathy. E-mail: Tolkacheva.Anna@endocrincentr.ru

Natalia Gennadiyevna Shebardina (b. 1999), Junior Researcher, A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University. Research interests: regenerative therapy, cellular technologies, diabetology, ophthalmology. E-mail: Shebardina.Natalia@endocrincentr.ru ORCID: 0009-0007-3372-5015

Elena Alexandrovna Korchuganova (b. 1973), Doctor of Medical Sciences, Head of the Therapeutic and Diagnostic Department of Diabetic Retinopathy and Ophthalmic Surgery, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: diabetic retinopathy, ophthalmic complications in endocrine diseases, healthcare organization.

E-mail: korchuganova.elena@endocrincentr.ru ORCID: 0000-0001-5147-0385

Natalia Georgievna Mokrysheva (b. 1970), Doctor of Medical Sciences, Director, National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Ministry of Health of the Russian Federation. Research interests: endocrinology, healthcare organization.

E-mail: mokrysheva.natalia@endocrincentr.ru ORCID: 0000-0002-9717-9742

Received September 18, 2025. The final version – April 16, 2026.
