

АВТОРЫ

С.О. Самохин¹, А.В. Патрушев^{1,2}, Г.К. Леонтьев³, И.В. Карпов⁴, А.М. Шутко⁵

ОРГАНИЗАЦИИ

¹Федеральное государственное бюджетное военное образовательное учреждение высшего профессионального образования «Военно-медицинская академия имени С.М. Кирова», Россия, г. Санкт-Петербург, улица Академика Лебедева лит. Ж, 194044

²Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр имени В.А. Алмазова» Министерства здравоохранения Российской Федерации, Россия, г. Санкт-Петербург, ул. Аккуратова, д. 2, 197341

³Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», Россия, г. Санкт-Петербург, Кронверкский проспект, д. 49, 197101

⁴Компания «web-bee.ru»

⁵Федеральное государственное бюджетное образовательное учреждение высшего образования «Первый Санкт-Петербургский государственный медицинский университет имени академика И.П. Павлова», Россия, г. Санкт-Петербург, ул. Льва Толстого, д. 6-8, 197022

НАЗВАНИЕ СТАТЬИ

Применение программного комплекса СС-2 в работе врача дерматолога для диагностики новообразований кожи.

Структурированная аннотация

Обоснование: В настоящее время проблема своевременной и точной диагностики новообразований кожи остается одной из наиболее актуальных в дерматологии и онкологии. Злокачественные новообразования кожи, включая меланому, являются одним из наиболее распространенных видов онкологических заболеваний с тенденцией к росту заболеваемости во всем мире. Раннее выявление злокачественных новообразований кожи существенно увеличивает шансы пациента на успешное лечение и выживаемость.

Цель исследования: Провести сравнительный анализ между автоматизированной диагностикой меланомы и экспертной оценкой врача.

Методы: наша команда врачей совместно со специалистами машинного обучения (Machine Learning – ML) создали собственный программный комплекс и внедрили его во врачебную практику. Разработанный нами запатентованный алгоритм СС-2 был обучен для выявления новообразований кожи, особенно меланомы и уже показал успешные результаты.

Результаты: Проведенные сравнения комплекса СС-2 с диагностической точностью опытного специалиста показали заметное превосходство над врачами-дерматологами. Модель ResNeXt-50 + СВМ (Convolutional Block Attention Module) показала более высокую эффективность распознавания меланомы по сравнению с опытным дерматологом:

F1-score: 0,745 (на 31% выше, чем у врача — 0,567)

F2-score: 0,754 (на 20% выше, чем у врача — 0,630, $p = 0,015$)

Интеграция СВМ усилила внимание к ключевым признакам, что повысило чувствительность и снизило риск пропуска опасных случаев.

Заключение: применение программного комплекса СС-2 в работе врачей-дерматологов

представляет высокую научную и практическую значимость и отвечает современным тенденциям развития медицины и информационных технологий.

Ключевые слова: Новообразования кожи; меланома; искусственный интеллект; нейронные сети; автоматизация.

AUTHORS

Simon O. Samokhin¹, Alexander V. Patrushev^{1,2}, German K. Leontiev³, Igor V. Karpov⁴, Anfisa M. Shutko⁵

AFFILIATION

¹S.M. Kirov Military Medical Academy, St. Petersburg, Russia, Akademika Lebedeva Street, Lit. Zh, 194044

²V.A. Almazov National Medical Research Centre, St. Petersburg, Russia, Akkuratova, 2, 197341

³ITMO University, St. Petersburg, Russia, Kronverkskiy Prospekt, 49, 197101

⁴Company «web-bee.ru»

⁵Pavlov First State Medical University of St. Petersburg, Russia, St. Petersburg, L'va Tolstogo street, 6-8, 197022

TITLE

Application of the CC-2 program complex in the work of a dermatologist for diagnostics of skin neoplasms.

ABSTRACT

Background: The timely and accurate diagnosis of skin neoplasms is currently one of the most pressing issues in dermatology and oncology. Malignant skin neoplasms, including melanoma, are among the most prevalent cancers, with increasing incidence rates worldwide. Early detection significantly increases patients' chances of successful treatment and survival.

Aims: to conduct a comparative analysis between automated melanoma diagnosis and expert medical assessment.

Materials and methods: Our team of doctors and machine learning specialists created our software and implemented it in medical practice. We developed the patented SS-2 algorithm to detect skin neoplasms, especially melanoma, and it has already shown successful results.

Results: Comparisons of the SS-2 complex with the diagnostic accuracy of experienced specialists have shown it to be notably superior to dermatologists. The ResNeXt-50 + CBAM (Convolutional Block Attention Module) model showed higher melanoma recognition efficiency compared to an experienced dermatologist:

F1-score: 0.745 (31% higher than the doctor's score of 0.567)

F2-score: 0.754 (20% higher than the doctor's score of 0.630, $p = 0.015$)

The integration of CBAM has increased attention to key indicators, which has improved sensitivity and reduced the risk of missing dangerous cases.

Conclusions: The use of the CC-2 software package in the work of dermatologists is of great scientific and practical importance and is in line with current trends in the development of medicine and information technology.

Keywords: Skin neoplasms; melanoma; artificial intelligence; neural networks; automation.

Обоснование

Заболеваемость меланомой, наиболее агрессивной формой рака кожи, продолжает расти во всем мире, особенно среди светлогокожего населения (тип I и II). Раннее выявление имеет решающее значение для улучшения состояния пациентов, и последние достижения в области искусственного интеллекта (ИИ) показали перспективность повышения точности и эффективности диагностики и лечения меланомы [1]. В данном обзоре рассматривается роль ИИ в диагностике поражений кожи, выделяются два основных подхода: машинное обучение, в частности конволюционные нейронные сети (CNN), и экспертные системы. Методы искусственного интеллекта продемонстрировали высокую точность при классификации дерматоскопических изображений, часто соответствующую или превосходящую дерматологов. Внедрение ИИ в дерматологию позволило улучшить такие задачи, как классификация, сегментация и прогнозирование риска, что способствует более раннему и точному вмешательству [2].

Несмотря на эти достижения, остаются нерешенные проблемы, включая предвзятость обучающих данных, проблемы интерпретации, и интеграция ИИ в клинические рабочие процессы. Обеспечение разнообразного представления данных и поддержание высоких стандартов качества изображений являются важнейшими условиями надежной работы ИИ. Будущие направления включают разработку более сложных моделей, таких как зрительно-языковые и мультимодальные модели, и объединенное обучение для решения проблем конфиденциальности и обобщения данных. Постоянная валидация и этическая интеграция ИИ в клиническую практику жизненно важны для реализации его полного потенциала в улучшении диагностики меланомы и лечения пациентов [3].

Рак кожи - одна из самых распространенных форм рака, которая отличается высоким уровнем злокачественности. У светлогокожих людей его заболеваемость ежегодно увеличивается на 3–7 % за последние несколько десятилетий, превышая заболеваемость почти всеми другими злокачественными новообразованиями [4]. Поэтому для выявления рака кожи на ранних стадиях очень важны тщательные скрининговые программы и тесты. Автоматизация идентификации меланоцитарных невусов на коже с помощью вычислительного искусственного интеллекта (ИИ) и методов анализа изображений необходима для достижения этой цели [5]. Самая серьезная форма рака кожи, меланома, может быть выявлена на ранних стадиях, и это существенно влияет на результаты лечения пациентов. Это снижает вероятность опасных для жизни последствий, замедляет прогрессирование заболевания и повышает вероятность успешного лечения. Кроме того, раннее выявление позволяет сократить время восстановления, провести менее инвазивные хирургические процедуры и успешно использовать вспомогательные препараты. Улучшение результатов лечения пациентов в основном зависит от своевременного принятия мер и рутинного осмотра кожи [6].

ИИ становится ключевым инструментом в медицине, включая дерматологию, где он облегчает диагностику и лечение кожных заболеваний. Точная и своевременная идентификация поражений, таких как меланома, имеет решающее значение для раннего вмешательства и улучшения прогноза пациентов. Для диагностики поражений кожи используются две основные методологии ИИ. Первый метод использует машинное обучение (МОО), в частности нейронные сети и различные сверточные нейронные сети (CNN) [7], которые требуют больших и надежных наборов обучающих данных, характеризующихся сложными внутренними механизмами, которые нелегко интерпретировать человеческим экспертам. Эта непрозрачность, часто называемая проблемой «черного ящика», возникает, когда модели машинного обучения, в частности алгоритмы глубокого обучения (DL), обрабатывают огромные объемы данных и генерируют предсказания, не предоставляя четких объяснений того, как конкретные входные данные

приводят к определенным результатам. В результате люди-эксперты сталкиваются с проблемой подтверждения, доверия или понимания процесса принятия решений, что ограничивает применимость системы в таких важных областях, как здравоохранение или принятие юридических решений, где интерпретируемость имеет решающее значение [8,9]. Второй метод предполагает использование экспертных систем, которые опираются на информационные онтологии и используют четкие правила и зависимости для кодирования экспертных знаний. Хотя эти системы обеспечивают прозрачность и простоту интерпретации, их разработка и обслуживание требуют больших затрат [10].

Внедрение методов искусственного интеллекта в дерматологию привело к заметному прогрессу в автоматизированном анализе изображений поражений кожи. Эти алгоритмы ИИ продемонстрировали значительный успех в решении различных задач, таких как классификация поражений кожи, сегментация изображений, извлечение признаков и прогнозирование рисков. Несмотря на эти достижения, недавние исследования выявили ряд проблем и препятствий на пути применения ИИ в диагностике рака кожи [11,12].

Традиционные методы диагностики новообразований кожи во многом зависят от квалификации и опыта врача-дерматолога, что вносит элемент субъективности в процесс постановки диагноза. Даже высококвалифицированные специалисты могут сталкиваться с диагностическими трудностями при оценке пограничных и атипичных случаев. Кроме того, существующий дефицит врачей-дерматологов, особенно в отдаленных регионах, ограничивает доступность качественной медицинской помощи для значительной части населения [13].

Развитие технологий искусственного интеллекта и компьютерного зрения, в частности сверточных нейронных сетей, открывает новые перспективы для создания систем автоматизированной диагностики в медицине. Созданный нами и испытанный на практике программный комплекс СС-2, основанный на сверточных нейронных сетях и технологиях искусственного интеллекта, представляет собой инновационное решение, потенциально способное повысить точность, скорость и доступность диагностики новообразований кожи.

Цель исследования: провести сравнительный анализ между автоматизированной диагностикой меланомы и экспертной оценкой врача, проверить, может ли созданная нами автоматизированная система на основе CNN (нейронных сетей) распознавать меланому по дерматоскопическим снимкам **лучше опытного (>5 лет работы в дерматоскопии) дерматолога**, и показать статистически значимое преимущество по клинически важным метрикам программного комплекса.

МЕТОДЫ

ДИЗАЙН ИССЛЕДОВАНИЯ

Открытое, нерандомизированное, контролируемое исследование диагностической точности, выполненное по типу поперечного (кросс-секционного) анализа с использованием госпитальной выборки дерматоскопических изображений, где роль контроля выполняла экспертная оценка врача-дерматолога, а сравнение проводилось на одной и той же когорте изображений с последующей гистологической верификацией («золотой стандарт»).

УСЛОВИЯ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЯ

Исследование проводилось на базе данных ISIC (International Skin Imaging Collaboration) dataset (открытый источник данных) – 63932 снимка, а также фотоснимках дерматоскопии пациентов с кожными новообразованиями (архив кафедры кожных и венерических

болезней Военно-медицинской академии) – 5000 снимков.

Проанализировав преимущества методов искусственного интеллекта перед дерматологами [14], наша команда врачей совместно со специалистами машинного обучения (Machine Learning – ML) создали собственный программный комплекс и внедрили его во врачебную практику. Разработанный нами запатентованный алгоритм CC-2 был обучен для выявления новообразований кожи, особенно меланомы и уже показал успешные результаты.

Изначально нашей задачей было проверить, может ли автоматизированная система на основе CNN (нейронных сетей) распознавать меланому по дерматоскопическим снимкам лучше опытного (>5 лет работы в дерматоскопии) дерматолога, и показать статистически значимое преимущество по клинически важным метрикам.

В рамках данного исследования проведён сравнительный анализ автоматизированной диагностики меланомы с экспертной оценкой врача. Ключевым нововведением является внедрение CBAM (Convolutional Block Attention Module) – модуля внимания, который усиливает обучение нейросети, в архитектуру ResNeXt-50 (архитектура нейросети, используемая для классификации изображений. Она расширяет идеи ResNet, добавляя групповые сверточные операции, которые улучшают представление признаков), что позволяет усиливать релевантные признаки и повышать чувствительность модели к злокачественным образованиям.

Обучение модели для CC-2 проводилось на базах данных ISIC (International Skin Imaging Collaboration) dataset, содержащих вместе 63 932 дерматоскопических изображения, среди которых 54 486 не относятся к меланоме, а также на 5000 архивных снимках кафедры. Важно отметить, что датасет содержит несколько классов кожных заболеваний, включая актинический кератоз, доброкачественные опухоли и другие поражения кожи. Однако, в рамках данной работы была сформирована бинарная классификация: "меланома" vs. "не меланома" [15].

Для обеспечения честного сравнения из тренировочного набора были изъяты изображения, использованные при тестировании врачей, чтобы исключить утечку данных и предвзятость результатов.

Основным нововведением исследования является интеграция механизма внимания CBAM в архитектуру ResNeXt-50.

- ResNeXt-50 — мощная архитектура, обеспечивающая групповые сверточные операции, что делает её эффективной для классификации изображений.
- CBAM (Convolutional Block Attention Module) — позволяет динамически перераспределять внимание как на уровне каналов, так и на уровне пространственных признаков.
 - Channel Attention усиливает наиболее значимые спектральные характеристики изображения.
 - Spatial Attention позволяет фокусироваться на ключевых областях кожи, где могут находиться патологические изменения.

CBAM был интегрирован в слои layer3 и layer4 модели ResNeXt-50, где обрабатываются высокоуровневые признаки, что позволило усилить важные паттерны перед пропуском через skip-connection [17]

Таким образом, использование CBAM было не просто экспериментом, а необходимым условием для обеспечения новизны исследования.

Обучение проводилось с использованием стратифицированной 5-фолдовой кросс-валидации, а в качестве функции потерь применялся BCEWithLogitsLoss с оптимизатором Adam и планировщиком CosineAnnealingLR.

Для оценки качества модели были использованы F1-score и F2-score (в медицине, особенно при диагностике рака кожи, метрики F1 и F2 важнее, чем просто Accuracy, потому что они учитывают баланс между чувствительностью (Recall) и точностью (Precision)).

КРИТЕРИИ СООТВЕТСТВИЯ (ОТБОРА)

Перед началом исследования были сформулированы критерии, определяющие отбор дерматоскопических изображений для анализа программным комплексом CC-2 и врачом-дерматологом.

1. Критерии включения

В исследование включались изображения, удовлетворяющие следующим условиям:

1. дерматоскопический формат снимка;
2. наличие подтверждённого диагноза («меланома» или «немеланома»);
3. возможность однозначного отнесения к двум исследуемым классам (бинарная классификация);
4. техническая пригодность изображения для визуальной оценки (стандартное разрешение, отсутствие критических артефактов);
5. принадлежность к одному из двух источников данных (ISIC dataset или архив ВМедА).

Эти критерии распространялись на материалы обеих диагностических стратегий (CC-2 и врач).

2. Критерии невключения (exclusion criteria)

Предварительно критерии невключения не были запланированы, так как все изображения, отвечающие критериям включения, подлежали включению в исследование.

3. Критерии исключения (withdrawal criteria)

Критерии исключения (т.е. удаления данных после включения) формально определялись, но ни один из них не сработал, и исключения не проводились. Плановые критерии исключения включали:

1. выявление дубликатов среди уже включённых изображений;
2. обнаружение ошибки в верификации диагноза;
3. невозможность обработки изображения программным комплексом CC-2 вследствие технического сбоя;
4. невозможность оценки врачом (например, повреждение файла);
5. выявление несоответствия формату дерматоскопии на позднем этапе.

По факту в ходе исследования исключений не было ($n = 0$).

4. Специфичные критерии для группы контроля

Поскольку группа контроля (врач-дерматолог) использовала те же изображения, что и СС-2, никаких дополнительных или специфичных критериев соответствия для группы контроля не предусматривалось.

Описание критериев соответствия

Критерии соответствия были определены до начала исследования и использовались для отбора дерматоскопических изображений, включаемых в анализ.

Критерии включения:

1. дерматоскопический формат изображения, соответствующий стандартам ISIC;
2. наличие подтверждённого диагноза («меланома» или «немеланома») на основании гистологии или достоверного клинико-дерматоскопического заключения;
3. возможность однозначного отнесения изображения к бинарной классификации;
4. техническая пригодность снимка (достаточное разрешение, отсутствие выраженных артефактов).

Эти критерии основаны на общепринятых принципах исследований диагностической точности и рекомендациях ISIC по стандартизации изображений.

Критерии невключения:

Отдельные критерии невключения не формулировались, так как все изображения, удовлетворяющие критериям включения, подлежали включению.

Критерии исключения:

Были определены до начала исследования (дубликаты, ошибки верификации диагноза, техническая невозможность оценки), однако фактических исключений не было ($n = 0$).

Все диагностические стратегии (СС-2 и врач-дерматолог) работали с одной и той же выборкой изображений; специфичных критериев соответствия для группы контроля не требовалось.

Подбор участников в группы

В данном исследовании подбор пар (matching) не проводился, поскольку диагностические стратегии (программный комплекс СС-2 и врач-дерматолог) оценивали одну и ту же совокупность дерматоскопических изображений, что обеспечивает полное внутриобъектное соответствие групп и исключает необходимость сопоставления по каким-либо переменным.

ОПИСАНИЕ ВМЕШАТЕЛЬСТВА

Вмешательством в данном исследовании являлась диагностическая процедура, заключающаяся в оценке каждого дерматоскопического изображения двумя независимыми методами: программным комплексом СС-2 и врачом-дерматологом. Эти действия выполнялись специально в рамках исследования и не происходили бы без включения изображений в исследуемую выборку.

1. Оценка изображений программным комплексом СС-2

Каждое изображение проходило автоматизированный анализ, включающий предварительную обработку и классификацию моделями ResNeXt-50 с модулем СВМ. Алгоритм формировал бинарное заключение о принадлежности к классу «меланома» или «немеланома». Обработка осуществлялась однократно, без повторных процедур и без возможности воздействия на результат.

2. Оценка изображений врачом-дерматологом

Врач-дерматолог независимо просматривал те же изображения в стандартизированных условиях (единый монитор, нейтральное освещение) и выносил диагностическое

закключение на основе визуального анализа дерматоскопических структур. Данные оценки также выполнялись однократно.

3. Организационные действия

Исследователи обеспечили стандартизированное предъявление изображений и параллельную независимую оценку двумя диагностическими стратегиями, что и составляло вмешательство в рамках исследования.

Исходы исследования

Основной исход исследования.

Основным исходом исследования была диагностическая точность определения меланомы на дерматоскопических изображениях, оцененная с помощью F2-score как ключевого показателя, отражающего клинически значимое преимущество чувствительности. Дополнительно анализировался F1-score как вторичный показатель согласованности классификации. Бинарные категории исхода («меланома»/«немеланома») определялись на основании подтверждённого диагноза.

Дополнительные исходы исследования

Не изучали

Методы регистрации исходов

Регистрация исходов осуществлялась путём сравнения бинарных заключений программного комплекса CC-2 и врача-дерматолога с подтверждённым диагнозом, установленным по гистологии или стандартам ISIC. Прогноз CC-2 формировался моделью ResNeXt-50 с модулем CBAM (PyTorch, США) после стандартной предобработки изображений; результат представлял собой вероятностную оценку, преобразованную в бинарный исход по фиксированному порогу. Врач-дерматолог оценивал те же изображения в стандартизированных условиях, не имея доступа к результатам CC-2. На основании полученных решений автоматически рассчитывались F1-score и F2-score.

АНАЛИЗ В ПОДГРУППАХ

Нет

СТАТИСТИЧЕСКИЕ ПРОЦЕДУРЫ

Запланированный размер выборки

Размер выборки в данном исследовании **предварительно не рассчитывался**, поскольку анализ выполнялся на основе всей доступной совокупности дерматоскопических изображений, представленных в открытой базе ISIC (63 932 снимка) и архиве BMA (5 000 снимков). Такой подход соответствует практике исследований диагностической точности, в которых использование максимально полной выборки обеспечивает более устойчивые оценки показателей модели и снижает риск выборочного смещения.

Допущения о требуемой мощности исследования и ожидаемом эффекте заранее не формулировались; условия остановки исследования (включая достижение статистически значимого результата) **не планировались**. Исследование завершено после полного анализа всех доступных изображений, удовлетворяющих критериям включения.

Статистические методы

Статистический анализ выполняли с использованием программного пакета Python 3.10 (PSF, USA) с библиотеками NumPy, SciPy, Pandas, scikit-learn и PyTorch. Количественные показатели описывали в формате медиана, 25-й и 75-й процентиля (Me [Q1; Q3]), так как распределения показателей диагностической точности были непараметрическими.

Для сравнения диагностических метрик между программным комплексом CC-2 и врачом-дерматологом использовали бутстраповую выборку (1000 перезапусков) с последующим вычислением эмпирического распределения разностей F1- и F2-показателей. Статистическая значимость различий оценивалась с помощью двустороннего пермутационного теста, что соответствует современным рекомендациям для оценки моделей машинного обучения при отсутствии нормальности распределений. (Бутстраповая выборка – статистическая процедура, которая перевыбирает данные из одного набора

данных для создания множества смоделированных выборок. Этот процесс позволяет вычислять стандартные ошибки, строить доверительные интервалы и проводить проверку гипотез для различных типов выборочных статистик).

Основным показателем был F2-score, дополнительным — F1-score; для обоих рассчитывались медианные значения и бутстрап-интервальные оценки. Пропущенные данные, выбросы или нарушения полноты наблюдения отсутствовали, поскольку анализ выполнялся на статичных изображениях. Трансформация данных не применялась.

Выбор статистических методов и формата представления результатов основан на рекомендациях SAMPL и применяемых в литературе подходах к оценке диагностической точности моделей глубокого обучения.

РЕЗУЛЬТАТЫ

- Почему не Ассигасу (метрика качества работы модели, которая показывает, какая доля всех предсказаний оказалась правильной)?
 - В датасете наблюдается дисбаланс классов (~5.8:1 в пользу "не меланом").
 - Ассигасу не учитывает разницу между ложноположительными (FP) и ложноотрицательными (FN) случаями, что делает её малоинформативной в медицинских задачах.
- Почему F1 и F2 (рис. 1)?
 - F1-score балансирует Precision (долю правильных диагнозов среди положительных) и Recall (насколько модель способна находить все случаи меланомы).
 - F2-score делает более сильный акцент на Recall, что критично для задач медицинской диагностики, где важно минимизировать пропущенные случаи заболевания.

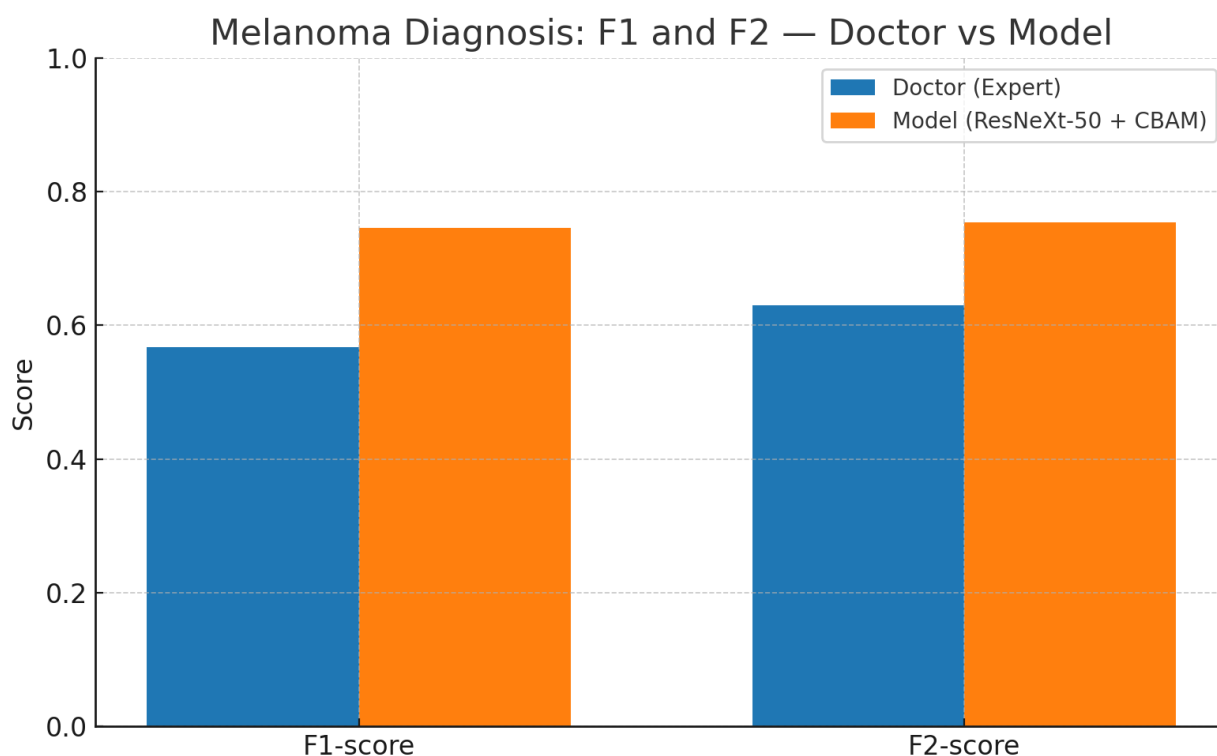


Рис. 1 – Сравнение F1-score и F2-score между врачом и моделью программы.

Для задач, где присутствует дисбаланс классов, использовалась метрика F-beta (рис. 2). Чем больше beta, тем строже алгоритм следил за тем, чтобы не пропустить ни одного больного. Сама формула учитывает верные положительные (True Positive = TP), ложные положительные (False Positive = FP) и ложные отрицательные (False Negative = FN) предсказания. Это отражено в формуле:

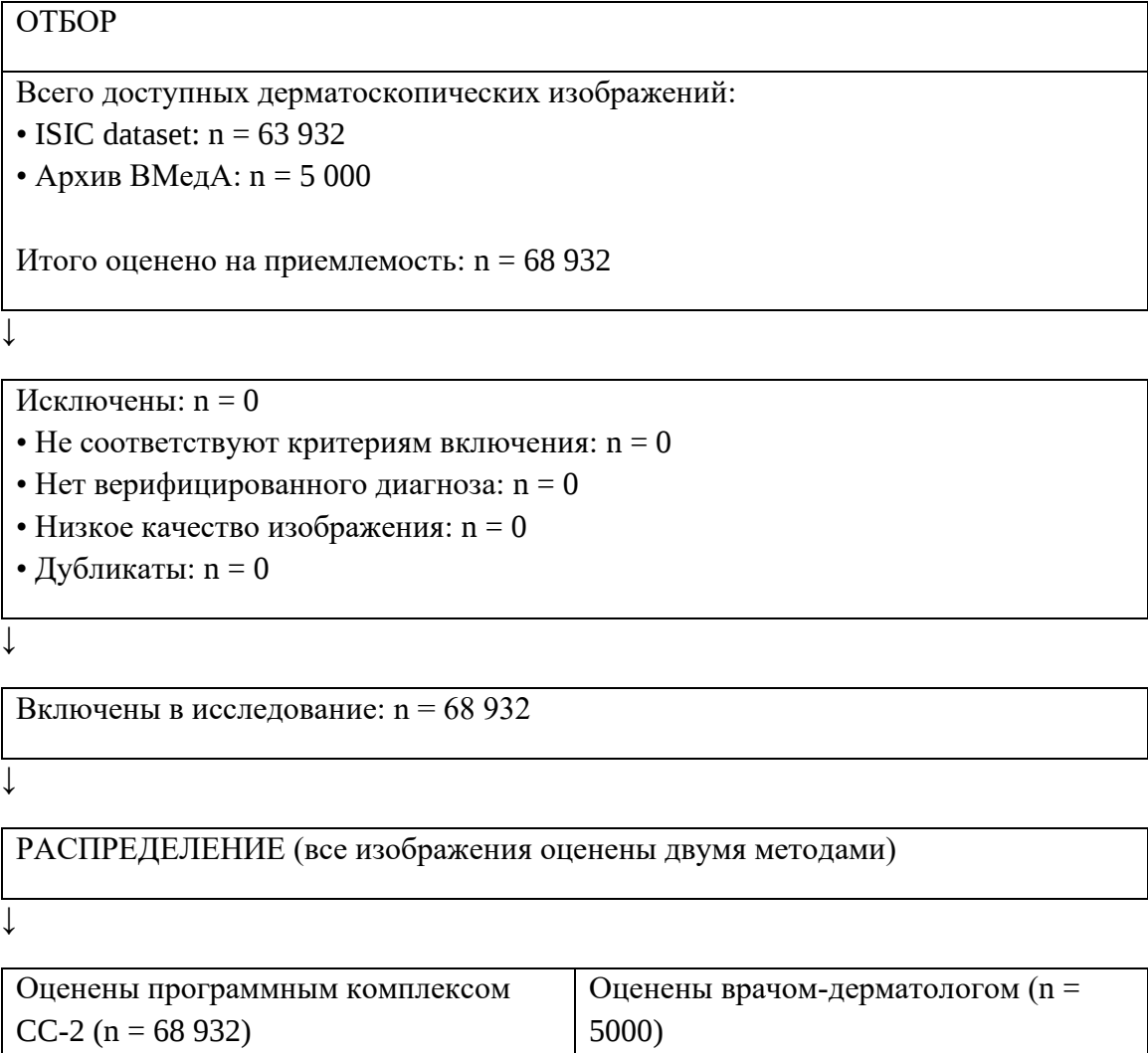
$$(\beta^2 + 1) \frac{TP}{\beta^2 (FN + FP) + (1 - \beta^2) TP}$$

Рис. 2 – Метрика F-beta

Интеграция механизма внимания в архитектуру сверточной сети позволила существенно повысить качество автоматизированной диагностики меланомы. Полученные результаты демонстрируют потенциал использования подобных систем как инструмента поддержки принятия решений в клинической практике дерматологов.

ФОРМИРОВАНИЕ ВЫБОРКИ

Потоковая диаграмма формирования выборки



• Получена оценка: $n = 68\,932$ • Не удалось оценить: $n = 0$	• Получена оценка: $n = 5000$ • Не удалось оценить: $n = 0$
---	--

↓

АНАЛИЗ • Проанализировано СС-2: $n = 68\,932$ • Проанализировано врачом: $n = 5000$ • Исключены из анализа: $n = 0$

ХАРАКТЕРИСТИКИ ВЫБОРКИ

В исследование была включена совокупная выборка дерматоскопических изображений, состоящая из снимков пациентов с кожными новообразованиями, полученных из базы данных ISIC и архива кафедры кожных и венерических болезней Военно-медицинской академии. Каждое изображение оценивалось двумя диагностическими стратегиями: врачом-дерматологом (контроль) и программным комплексом СС-2 (основная группа). Таким образом, группы сравнения не являются независимыми, поскольку оба метода анализировали одни и те же клинические объекты.

Так как сравниваемые диагностические стратегии (врач и СС-2) работали с одной и той же когортой дерматоскопических изображений, группы не являются независимыми в статистическом смысле. Следовательно, сравнение их демографических и клинических характеристик не имеет методологического смысла: различий быть не может.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для осуществления анализа эффективности автоматизированной диагностики меланомы были проведены сравнения с экспертной оценкой врача, используя данные из научного исследования.

Результаты диагностики опытного врача:

- F1-score = 0.567
- F2-score = 0.630

Результаты модели (ResNeXt-50 + CBAM):

- **F1-score:** 0,745 (**на 31% выше**, чем у врача — 0,567)
- **F2-score:** 0,754 (**на 20% выше**, чем у врача — 0,630, $p = 0,015$)

Для проверки статистической значимости различий был проведён бутстраповый анализ ($n=1000$).

- Разница в F2-score оказалась статистически значимой ($p = 0.015$), что подтверждает, что модель действительно снижает число пропущенных диагнозов.
- Разница в F1-score не достигла статистической значимости ($p = 0.124$), что может свидетельствовать о влиянии случайных факторов [8,14,16].

Для исследования использовалась база дерматоскопических изображений ISIC, а также фотоснимки дерматоскопии пациентов с кожными новообразованиями (архив кафедры кожных и венерических болезней Военно-медицинской академии), на основе которых была сформулирована задача бинарной классификации: «меланома» против «немеланома». Для построения диагностической системы применялась сверточная нейронная сеть ResNeXt-50,

в которую был интегрирован модуль внимания CBAM (Convolutional Block Attention Module). Этот механизм позволяет модели автоматически выделять наиболее значимые признаки и области изображения, усиливая информативные паттерны и подавляя фоновые [16].

Реализация алгоритма выполнена на языке Python с использованием библиотек PyTorch и torchvision. Обучение модели проводилось методом стратифицированной пятикратной кросс-валидации. В качестве функции потерь использовалась BCEWithLogitsLoss, оптимизация осуществлялась с помощью алгоритма Adam с планировщиком CosineAnnealingLR. Основными метриками эффективности выбраны F1 и F2, что особенно важно при дисбалансе классов и клинической значимости снижения числа пропущенных случаев меланомы.

Сравнение разработанной модели с диагностической точностью опытного врача-дерматолога показало превосходство нейросети по ключевым метрикам. Так, F1-показатель модели составил 0,745 против 0,567 у врача, а показатель F2 достиг 0,754 против 0,630. При этом улучшение по F2 оказалось статистически значимым ($p = 0,015$).

1. Ещё до внедрения CBAM нейросеть на основе ResNeXt-50 превосходила врача по F1-score и F2-score, однако использование CBAM позволило улучшить эффективность выявления меланомы.
2. Интеграция CBAM обеспечила усиление ключевых признаков изображения, что особенно важно в медицинской диагностике, где необходимо учитывать мелкие изменения на коже.
3. Применение бутстрапа позволило подтвердить статистическую значимость улучшения модели, особенно по метрике F2-score ($p = 0.015$).
4. Таким образом, использование CBAM в ResNeXt-50 стало необходимым условием новизны исследования, а статистический анализ доказал его эффективность в задаче диагностики меланомы.

ДОПОЛНИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

НЕЖЕЛАТЕЛЬНЫЕ ЯВЛЕНИЯ

Отмечено отсутствие нежелательных явлений.

ОБСУЖДЕНИЕ

РЕЗЮМЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Внедрение разработанной системы в клиническую практику обладает рядом значимых преимуществ. В частности, она позволяет:

- Сократить время, необходимое для постановки диагноза
- Повысить точность и объективность диагностики
- Снизить количество ошибочных диагнозов и необоснованных биопсий
- Обеспечить раннее выявление злокачественных новообразований
- Сделать качественную диагностику более доступной в регионах с дефицитом специалистов

Модель ResNeXt-50 + CBAM показала более высокие результаты, чем врач, и бутстраповый анализ подтвердил, что внедрение CBAM действительно улучшает способность нейросети к выявлению меланомы.

ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Результаты проведённого анализа указывают на то, что интеграция модуля внимания СВМ в архитектуру ResNeXt-50 приводит к статистически значимому повышению диагностической точности при выявлении меланомы по сравнению как с исходной моделью, так и с экспертной оценкой дерматолога. Это наблюдение особенно важно в контексте цели исследования — оценить способность программного комплекса СС-2 улучшать диагностические решения в клинической практике.

(а) Новые результаты, показанные впервые

Полученные данные демонстрируют, что добавление пространственно-канального модуля внимания СВМ обеспечивает прирост F1- и особенно F2-показателей по сравнению с экспертной оценкой врача (F2: 0,754 против 0,630, $p = 0,015$). В доступной литературе применение СВМ исследовалось преимущественно на общих задачах компьютерного зрения, но его эффективность в дерматоскопии изучена ограниченно. Таким образом, данное исследование впервые показывает, что использование СВМ в контексте дерматоскопической диагностики позволяет улучшить чувствительность модели к минимальным визуальным признакам меланомы — в том числе слабовыраженным структурным паттернам. Это наблюдение имеет клиническую значимость, поскольку ранние этапы меланомы часто характеризуются «размытыми» или неявными признаками, на которые опытный врач может не обратить внимания.

(б) Результаты, согласующиеся с ранее опубликованными данными

Полученные данные согласуются с исследованиями, демонстрирующими превосходство современных сверточных нейросетей над клинической оценкой в задачах диагностики меланомы. Ранее было показано, что архитектуры ResNet-/ResNeXt-класса способны достигать и превышать уровень дерматологов при анализе изображений из ISIC-репозитория, что подтверждает общий тренд: глубокие модели, обученные на больших массивах дерматоскопических данных, способны выявлять сложные или незаметные визуальные закономерности. Дополнительное согласование наблюдается с данными о том, что методы внимания улучшают устойчивость модели при наличии неоднородного фона, артефактов и вариативности качества снимков. Таким образом, выявленное повышение точности в нашем исследовании не противоречит общей тенденции, а расширяет её, подтверждая применимость механизмов внимания в медицинской визуализации.

(в) Результаты, частично противоречащие ранее опубликованным данным

Отдельные исследования указывают, что преимущество нейросетевых моделей над врачами сокращается при использовании клинических, а не дерматоскопических изображений, либо при наличии сложных клинических случаев, где учитываются не только визуальные признаки, но и данные анамнеза. В настоящем исследовании модель демонстрирует уверенное превосходство над врачом по всем ключевым метрикам, что может противоречить выводам работ, подчёркивающих ограниченность ИИ в клиническом контексте. Однако это противоречие объяснимо:

(1) в настоящем исследовании сравнение проводилось на стандартизированных дерматоскопических изображениях, где нейросети имеют максимальное преимущество;

- (2) врач не имел доступа к дополнительной клинической информации, что могло ограничить его эффективность;
- (3) модель обучалась на большем массиве данных, чем реально доступно одному врачу в течение профессиональной карьеры.

Таким образом, выявленное преимущество ИИ не следует интерпретировать как универсальное превосходство в реальной клинической практике, но как демонстрацию специфического преимущества технологий глубокого обучения в условиях изолированного анализа изображений.

Общее значение результатов и вклад в научную литературу

Настоящее исследование вносит вклад в область дерматологической диагностики, показывая, что интеграция механизмов внимания, таких как СВМ, может существенно улучшить качество классификации меланомы. Эти результаты дополнительно подчеркивают важность разработки методов, направленных на усиление локальных структурных признаков — ключевого фактора в ранней онкодиагностике. Несмотря на полученное улучшение, данные следует рассматривать как этап формирования доказательной базы, а не как самостоятельное основание для изменения клинических рекомендаций. Требуются дальнейшие исследования, включающие (а) многоцентровые выборки, (б) клинические данные пациентов, (в) оценку эффективности в реальной практике.

ОГРАНИЧЕНИЯ ИССЛЕДОВАНИЯ

1. Ограничения, связанные с выборкой исследования

Выборка исследования формировалась из открытого дерматоскопического ресурса ISIC и архива (кафедра кожных и венерических болезней ВМедА). Такая стратегия, с методологической точки зрения, ближе к «госпитальной выборке» с элементами удобной (а не случайной) выборки и, очевидно, отличается от идеальной ситуации, когда в исследование включаются все пациенты целевой популяции в режиме сплошного учета. Это создаёт риск селекционного (выборочного) смещения:

- открытые базы, как правило, содержат более «качественные» и часто более типичные изображения, чем рутинная практика;
- госпитальные случаи нередко представляют более тяжёлый или клинически «подозрительный» спектр поражений.

В совокупности это может завышать оценку диагностической эффективности модели (и частично врача) по сравнению с реальной популяцией пациентов, где доля трудноинтерпретируемых, нетипичных и технически несовершенных изображений выше. Размер выборки был достаточен для получения статистически значимых различий по F2-score, однако остаётся вероятность того, что доверительные интервалы для отдельных подгрупп (по локализации, фототипу кожи, размеру новообразований) шире, чем для совокупной выборки. Следовательно, обобщаемость (генерализуемость) результатов на более разнородные популяции (другие регионы, уровни медпомощи, типы оборудования) ограничена и, вероятнее всего, несколько переоценивает верхнюю границу возможной точности алгоритма в рутинной практике.

2. Ограничения, связанные с сопоставимостью групп

Сравнение проводилось между двумя «диагностическими стратегиями» — нейросетевой моделью и экспертом-дерматологом — на одном и том же наборе изображений. С методологической точки зрения это обеспечивает внутриобъектное сопоставление (каждое изображение оценивается обеими сторонами), что снижает вариабельность и повышает мощность анализа. Однако ряд условий идеального сравнительного исследования не был выполнен:

Отсутствие случайного распределения и ослепления. Поскольку речь идёт о диагностическом испытании, классическая рандомизация пациентов в группы невозможна, но можно рассматривать «ослепление» эксперта относительно гипотез исследования и распределения классов. Если врач знал, что исследование посвящено бинарной задаче «меланома/немеланома» и мог предполагать повышенную долю злокачественных случаев, это могло сместить его порог настороженности и, соответственно, изменить баланс чувствительности и специфичности. Вектор смещения в этом случае неочевиден: усиление настороженности увеличивает чувствительность, но снижает специфичность и F1/F2; наоборот, избыточная осторожность в постановке диагноза меланомы может приводить к необнаружению.

3. Ограничения, связанные с показателями исследования

В исследовании использовалась бинарная классификация («меланома» против «немеланома»), что методологически упрощает реальную клиническую задачу, в которой врач имеет дело с широким спектром как злокачественных, так и доброкачественных дерматологических нозологий. Такое укрупнение классов:

- снижает сложность задачи для нейросети (и для врача)
- может завышать значения F1- и F2-показателей по сравнению с многоклассовыми сценариями, где требуется дифференцировать несколько типов новообразований, зачастую с частично пересекающимися визуальными признаками.

Источником «золотого стандарта» в идеальном исследовании служила бы гистологическая верификация для всех случаев. На практике часть диагнозов может быть основана на клинко-дерматоскопическом заключении и динамическом наблюдении, что создаёт риск информационного смещения: единичные ранние меланомы могли быть ошибочно отнесены к доброкачественным поражениям. В этом случае чувствительность и врача, и модели относительно истинной популяции скорее занижена, однако алгоритм, более чувствительный к слабо выраженным признакам, теоретически может либо выигрывать (если «видит» такие случаи), либо демонстрировать ложную уверенность, основанную на неточном референс-лейбле.

Кроме того, оценка эффективности преимущественно агрегирована в виде сводных показателей (F1, F2) для всей выборки без подробного анализа подгрупп (по фототипу, возрасту, локализации). Это ограничивает возможность выявить систематические ошибки модели в отдельных клинических контекстах (например, у пациентов с тёмным фототипом кожи) и может скрывать смещение, направленное в сторону переоценки одних подгрупп и недооценки других.

4. Ограничения, связанные с методами измерения

Методы измерения в идеальном исследовании предполагают высокую конструктивную валидность, воспроизводимость и полное соответствие условиям реальной клинической практики. В нашем случае:

Дерматоскопические изображения получены на конкретных моделях дерматоскопов и камер, при относительно стандартизированных условиях. В повседневной практике устройства и настройки могут существенно варьировать, увеличивая долю артефактов (блики, волосы, неоднородное освещение). Это означает, что оценка точности модели в условиях исследования, вероятнее всего, выше, чем в неконтролируемой реальной среде.

Применялись методы предобработки изображений и аугментации, оптимизированные для обучения модели. Эти процедуры улучшают устойчивость алгоритма, но в повседневной работе могут быть недоступны или реализованы иначе, что также может снизить фактическую эффективность при внедрении.

С точки зрения конструктивной валидности, модель оценивает исключительно визуальный компонент дерматоскопического изображения. В идеальной клинической системе измерения исхода включали бы как визуальные, так и клинические параметры (анамнез, факторы риска, данные осмотра), что могло бы изменить как абсолютную, так и относительную эффективность алгоритма и врача.

Наконец, экономическая и техническая доступность предлагаемого решения (необходимость вычислительных ресурсов, сопровождения программного обеспечения, интеграции с медицинскими информационными системами) может ограничивать его применение в учреждениях с ограниченными ресурсами. Вектор влияния здесь таков: даже при высоких методологических показателях точности реальная «эффективность» для системы здравоохранения может быть ниже из-за барьеров внедрения.

В совокупности перечисленные ограничения указывают на то, что результаты исследования следует рассматривать как оптимистичную, но не окончательную оценку диагностического потенциала модели ResNeXt-50 + CBAM. Наиболее вероятное направление смещения — в сторону завышения обобщаемой диагностической эффективности и степени превосходства модели над клиницистом по сравнению с многообразной и менее контролируемой реальной практикой. Это требует подтверждения полученных выводов в более широких, многоцентровых исследованиях с расширенным спектром пациентов, устройств и клинических сценариев.

ЗАКЛЮЧЕНИЕ

Разработанная модель ResNeXt-50 с интеграцией CBAM показала статистически значимое превосходство в выявлении меланомы по сравнению с опытным дерматологом. Усиление внимания к ключевым признакам изображений позволило повысить чувствительность и снизить риск пропуска опасных случаев. Полученные результаты подтверждают эффективность применения модулей внимания в медицинской диагностике и перспективность внедрения подобных решений в клиническую практику.

Участие авторов

Самохин Семён Олегович – поисково-аналитическая работа, написание статьи, организация исследования, работа с данными.

Патрушев Александр Владимирович – организация исследования, работа с данными.

Леонтьев Герман Константинович – разработка программы.

Карпов Игорь Владимирович – разработка программы.

Шутко Анфиса Максимовна – поисково-аналитическая работа.

Все авторы одобрили рукопись (версию для публикации), а также согласились нести ответственность за все аспекты настоящей работы, гарантируют надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой её части.

БЛАГОДАРНОСТИ

СОГЛАСИЕ НА ПУБЛИКАЦИЮ

Неприменимо.

ЭТИЧЕСКАЯ ЭКСПЕРТИЗА

Экспертиза в данной работе не требуется.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Рукопись и исследование подготовлены на личные средства авторского коллектива.

РАСКРЫТИЕ ИНТЕРЕСОВ

Авторы данной статьи сообщают об отсутствии отношений, деятельности и интересов за последние 36 месяцев с третьими лицами (физическими и юридическими), интересы которых могут быть затронуты содержанием статьи.

ЗАЯВЛЕНИЕ ОБ ОРИГИНАЛЬНОСТИ

В работе использованы как впервые собранные авторами сведения (архив снимков кафедры кожных и венерических болезней Военно-медицинской академии), так и открытая база данных ISIC (International Skin Imaging Collaboration) dataset (открытый источник данных).

ДОСТУП К ДАННЫМ

Авторы предоставляют полный открытый (неограниченный) доступ к данным, размещённым на внешнем ресурсе (репозитории).

ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Авторы заявляют о не использовании генеративного ИИ с целью создания рукописи.

РАССМОТРЕНИЕ И РЕЦЕНЗИРОВАНИЕ

Рукопись направлена в редакцию журнала в инициативном порядке.

ДИСКЛЕЙМЕР

ПРИЛОЖЕНИЯ

СПИСОК ЛИТЕРАТУРЫ

1. Sivodedova NA, Karyakin NN, Shlivko IL. Early diagnosis of malignant skin neoplasms using an artificial intelligence-based software product and its medical-economic effect. *Menedzher Zdravookhraneniya*. 2024;6:108-18. doi:10.21045/1811-0185-2024-6-108-118
2. Kaur R, GholamHosseini H, Sinha R, Linden M. Melanoma classification using a novel deep convolutional neural network with dermoscopic images. *Sensors (Basel)*. 2022;22(3):1134. doi:10.3390/s22031134
3. Sinelnikov IE, Utyashev IA, Nazarova VV. Features of dermatoscopy in the diagnosis of skin melanoma: literature review. *Effektivnaya Farmakoterapiya*. 2024;20(5):10-7. doi:10.33978/2307-3586-2024-20-5-10-1713
4. Sivodedova NA, Karyakin NN, Shlivko IL. Modern methods of detecting malignant skin neoplasms, including the use of mobile applications and artificial intelligence: literature review. *ORGZDRAV: Novosti, Mneniya, Obuchenie. Vestn VShOUZ*. 2024;10(2):78-93. doi:10.33029/2411-8621-2024-10-2-78-93
5. Smalyuk AF, Deshko MS, Kupchikova ID. Segmentation of dermoscopic images of skin neoplasms: comparison of methods. *Syst Anal Prikl Inform*. 2024;1:50-8. doi:10.21122/2309-4923-2024-1-50-58
6. Hagggenmüller S, Maron RC, Hekler A, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *Eur J Cancer*. 2021;156:202-16. doi:10.1016/j.ejca.2021.06.049
7. Jojoa Acosta MF, Caballero Tovar LY, Garcia-Zapirain MB, Percybrooks WS. Melanoma diagnosis using deep learning techniques on dermoscopic images. *BMC Med Imaging*. 2021;21(1):6. doi:10.1186/s12880-020-00534-8
8. Cerminara SE, Cheng P, Kostner L, et al. Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: a new era of skin cancer screening? *Eur J Cancer*. 2023;190:112954. doi:10.1016/j.ejca.2023.112954
9. Ali AR, Li J, Yang G, O'Shea SJ. A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images. *PeerJ Comput Sci*. 2020;6:e268. doi:10.7717/peerj-cs.268
10. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8. doi:10.1038/nature21056
11. Tertipi N, Sfyri E, Grech VS, Kefala V, Rallis E. Optimizing skin quality via AI-enhanced physical activity. *Cosmetics*. 2025;12(3):104. doi:10.3390/cosmetics12030104
12. Janda M, Soyer HP. Can clinical decision making be enhanced by artificial intelligence? *Br J Dermatol*. 2019;180(2):247-8. doi:10.1111/bjd.17110
13. Mabrouk MS, Sayed AY, Afifi HM, Sheha MA, Sharwy A. Fully automated approach for early detection of pigmented skin lesion diagnosis using ABCD. *J Healthc Inform Res*. 2020;4(2):151-73. doi:10.1007/s41666-020-00067-3
14. Kalidindi S. The role of artificial intelligence in the diagnosis of melanoma. *Cureus*. 2024;16(9):e69818. doi:10.7759/cureus.69818
15. SIIM-ISIC 2020 challenge dataset [dataset]. International Skin Imaging Collaboration;

2020. doi:10.34970/2020-ds01.
 16. Chollet F. Deep learning with Python. 2nd ed. St. Petersburg: Piter; 2023. 576 p.
 17. <https://github.com/german-leontiev/melanoma>, 2025. Герман Леонтьев, Игорь Карпов.

ТАБЛИЦЫ

РИСУНКИ

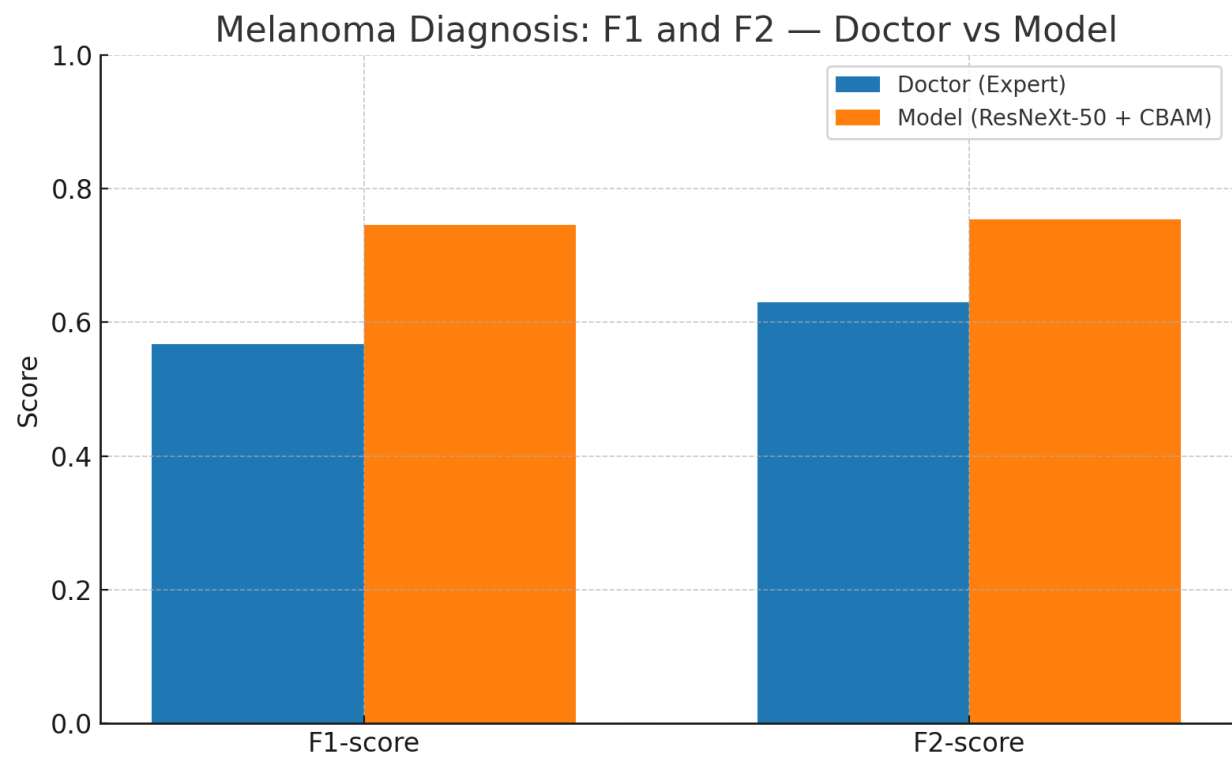


Рис. 1 – Сравнение F1-score и F2-score между врачом и моделью программы.

$$(\beta^2 + 1) \frac{TP}{\beta^2(FN + FP) + (1 - \beta^2)TP}$$

Рис. 2 – Метрика F-beta