

Анализ причин ошибочной классификации изображений немеланоцитарных опухолей кожи программами на основе технологий искусственного интеллекта

Корабельников Д.И., Ламоткин А.И.

Автономная некоммерческая организация дополнительного профессионального образования «Московский медико-социальный институт имени Ф.П. Гааза», 123056, Российская Федерация, Москва, ул. 2-я Брестская, 5

Аннотация

Дифференциальная диагностика немеланоцитарных опухолей кожи остается ключевой задачей дерматоонкологии, поскольку своевременное выявление злокачественных форм повышает шансы на успешное лечение. Субъективность традиционных методов стимулирует использование искусственного интеллекта (ИИ), при этом ошибки программ компьютерного зрения требуют анализа их причин.

Цель исследования: анализ причин ошибочной классификации изображений немеланоцитарных опухолей кожи программами на основе технологий ИИ.

Материал и методы. Для ретроспективного анализа фотоизображений опухолей использованы датасеты, обработанные в программах «Derma Onko Check» и «Melanoma Check». В контексте исследования злокачественные опухоли мы считали положительным результатом, доброкачественные – отрицательным. Для визуализации метрик качества изображений (яркость, контраст, энтропия, размытие, RGB-метрики) использованы боксплоты, парные диаграммы рассеяния, карты различий пиксельных значений. Для визуализации областей, наиболее важных для классификационных решений глубокой нейронной сети, применены два метода объяснимого ИИ: взвешенного картирования активации классов (Score-CAM) и определения чувствительности к окклюзии (Occlusion Sensitivity). Для проверки статистических гипотез использованы t-критерий Уэлча и однофакторный дисперсионный анализ; для оценки связи характеристик – корреляционный анализ по Спирмену.

Результаты. Истинно положительные (ИП) результаты характеризовались следующими особенностями: меньшая яркость (медиана – 0,6914 в нормализованной шкале 0–1), что означает естественное равномерное освещение без сильных бликов. Энтропия оказалась высокой (медиана – 4,8584), что указывает на сложную текстуру с многими клинически значимыми деталями: изъязвлением, неровными границами и вариациями пигментации. Размытие было умеренным, обеспечивая приемлемую резкость изображения без сильного смазывания краев и текстуры опухоли. Средние значения красного и зеленого каналов были сбалансированными. Ложноположительные (ЛП) результаты имели повышенную яркость (медиана – 0,7994 – переэкспонированное, слишком светлое фото с бликами), низкую энтропию (медиана – 4,6414 – однородная текстура без сложных паттернов). Значимые различия между ИП и ЛП классами подтверждены для яркости ($F = 5,1848; p < 0,05$), энтропии ($F = 5,2509; p < 0,05$), размытия по FFT ($F = 3,1136; p < 0,05$), среднего значения зеленого канала ($F = 5,3315; p < 0,05$) и среднего значения красного канала ($F = 3,3812; p < 0,05$). Методы объяснимого ИИ и анализ качества фотоизображений показали, что ошибки классификации происходили также из-за артефактов фотосъемки (фон, волосы, тени).

Заключение. Разработчикам для обучения программ «компьютерного зрения» рекомендуется проводить предобработку изображений (автоматический баланс белого, гамма-коррекция, фильтры Sobel для усиления текстуры и Wiener для подавления размытости, онлайн-аугментация яркости и контраста), нормализовать цветовые каналы, мониторить ключевые метрики качества после каждой эпохи обучения, использовать аугментацию, компенсирующую отрицательную корреляцию яркость – энтропия и вариабельность освещения. Пользователям рекомендуется соблюдать стандартные условия съемки: равномерное рассеянное освещение без теней и бликов, яркость $< 0,75$ в нормализованной шкале, отсутствие артефактов в кадре; выполнять съемку в режиме макро с расстояния 8–15 см, центрируя опухоль и обеспечивая энтропию $> 4,8$ и разрешение 2000–3000 пикселей по большей стороне; стабилизировать камеру и активировать функцию автоматического баланса белого на устройстве съемки.

Ключевые слова:	объяснимый искусственный интеллект; новообразования кожи; немеланоцитарные опухоли; ошибки классификации; предобработка изображений; параметры изображений; качество изображений.
Финансирование:	исследование выполнено без финансовой поддержки грантов, общественных, некоммерческих, коммерческих организаций и структур.

Ресурсное обеспечение:	для выполнения вычислительных задач использовались ресурсы облачной платформы Google Colab.
Соответствие принципам этики:	информированное согласие пациентов не требовалось, поскольку использовались анонимизированные фотоизображения из базы данных.
Для цитирования:	Корабельников Д.И., Ламоткин А.И. Анализ причин ошибочной классификации изображений немеланоцитарных опухолей кожи программами на основе технологий искусственного интеллекта. <i>Сибирский журнал клинической и экспериментальной медицины</i> . 2026;41(1):221–231. https://doi.org/10.29001/2073-8552-2026-2881

Analysis of the reasons of misclassification of non-melanocytic skin tumors by artificial intelligence based programs

Korabelnikov D.I., Lamotkin A.I.

Moscow Haass Medical and Social Institute, 5, 2nd Brestskaya str., Moscow, 123056, Russian Federation

Abstract

Introduction. Differential diagnosis of non-melanocytic skin tumors remains a key challenge in dermato-oncology, as timely detection of malignant forms increases the chances of successful treatment. The subjectivity of traditional methods encourages the use of artificial intelligence (AI), but errors in computer vision programs require analysis.

Aim: To analyze the reasons of misclassification of non-melanocytic skin tumor images by AI-based programs. This was accomplished by identifying systematic differences in data characteristics and visualizing regions of interest during image recognition. The results are aimed at improving the efficiency of training and using computer vision programs.

Material and Methods. Datasets processed in the Derma Onko Check and Melanoma Check programs were used for a retrospective analysis of tumor images. For this study, malignant tumors were considered positive, while benign tumors were considered negative. Considering two types of AI decisions (true and false), four standard result classes were considered: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). To visualize image quality metrics (brightness, contrast, entropy, blur, and RGB metrics), boxplots, paired scatterplots, and pixel difference maps were used. To visualize areas that significantly influence image classification, two explainable AI methods were applied: weighted class activation mapping (Score-CAM) and occlusion sensitivity. These methods allow us to understand which image regions are most important for the classification decisions of the deep neural network. To test statistical hypotheses, Welch's t-test and one-way analysis of variance were used; to assess the relationship between characteristics, Spearman's correlation analysis was used.

Results. Significant differences in image characteristics were identified. The IP results were characterized by the following features. Brightness was lower (median 0.6914 on a normalized scale of 0-1), indicating natural, uniform illumination without strong glare. Entropy was high (median 4.8584), indicating a complex texture with many clinically significant details: ulceration, irregular borders, and pigmentation variations. Blurring was moderate, providing acceptable image sharpness without severe blurring of the tumor edges and texture. The mean values of the red and green channels were balanced. The LP results had increased brightness (median 0.7994, indicating an overexposed, overly bright photo with glare, where fine textural details are lost) and low entropy (median 4.6414, indicating a uniform texture without complex patterns). Significant differences between classes were confirmed for brightness ($F = 5.1848$; $p < 0.05$), entropy ($F = 5.2509$; $p < 0.05$), FFT blur ($F = 3.1136$; $p < 0.05$), green channel mean ($F = 5.3315$; $p < 0.05$), and red channel mean ($F = 3.3812$; $p < 0.05$). The Score-CAM and Occlusion Sensitivity explainable AI methods and image quality analysis showed that non-melanocytic tumor classification errors by the Derma Onko Check and Melanoma Check AI programs occurred due to overexposure, low entropy, and photography artifacts; false positives occurred on bright, low-texture images, and false negatives occurred on dark/blurred images. AI models are distracted by the background, hair, shadows.

Conclusion. When training computer vision programs, developers are advised to perform image preprocessing (automatic white balance, gamma correction, Sobel filters for texture enhancement and Wiener filters for blur suppression, online brightness and contrast augmentation), normalize color channels, monitor key quality metrics after each training epoch, and use augmentation that compensates for the negative brightness-entropy correlation and illumination variability. Users of the programs are advised to adhere to standard shooting conditions: uniform diffuse lighting without shadows and glare, luminance < 0.75 on the normalized scale, and the absence of artifacts in the frame; shoot in macro mode from a distance of 8-15 cm, centering the tumor and ensuring entropy > 4.8 and a resolution of 2000-3000 pixels on the longest side; stabilize the camera and activate the automatic white balance function on the shooting device.

Keywords:	Explainable Artificial intelligence; Skin lesions; Non-melanocytic tumors; Classification errors; Image preprocessing; Image parameters; Image quality.
Funding:	the study was conducted without financial support from grants, public, non-profit, or commercial organizations or entities.
Resource support:	the resources of the Google Colab cloud platform were used to perform computational tasks.
Compliance with ethical standards:	informed consent from patients was not required, as anonymized images from a database were used.
For citation:	Korabelnikov D.I., Lamotkin A.I. Analysis of the reasons of misclassification of non-melanocytic skin tumors by artificial intelligence based programs. <i>Siberian Journal of Clinical and Experimental Medicine</i> . 2026;41(1):221–231. https://doi.org/10.29001/2073-8552-2026-2881

Введение

Диагностика кожных новообразований, включая немеланоцитарные опухоли (такие как кератиноцитарные / эпидермальные образования), остается одной из ключевых задач современной дерматоонкологии, поскольку своевременное выявление злокачественных форм может значительно повысить шансы на успешное лечение [1]. В 2022 г. зарегистрировано более 331 722 новых случаев рака кожи. Наиболее распространенными злокачественными новообразованиями (ЗНО) являются базальноклеточный и плоскоклеточный рак кожи – среди всех ЗНО кожного покрова они составляют около 99%¹ [2]. Традиционные методы диагностики, основанные на визуальном осмотре и биопсии, подвержены субъективным ошибкам и зависят от опыта специалиста, что приводит к необходимости разработки автоматизированных инструментов на базе искусственного интеллекта (ИИ) [3].

Интеграция программ ИИ (компьютерного зрения) в виде ИИ-ассистированных систем в клиническую онкологию и дерматовенерологию может улучшить раннюю диагностику, повысить точность дифференциальной диагностики. Это также позволит снизить нагрузку на систему здравоохранения, особенно в условиях дефицита специалистов и роста числа обращений по поводу новообразований кожи [4–6].

Программы ИИ, такие как «Derma Onko Check», «Melanoma Check», разработаны для автоматизированной классификации кожных новообразований. Они демонстрируют достаточно высокую эффективность [7, 8], однако ошибки классификации остаются вызовом, требующим детального анализа [9]. Эти ошибки могут быть связаны с вариабельностью визуальных характеристик фотоизображений (яркостью, контрастом, резкостью, энтропией, размытием по FFT, средними и стандартными отклонениями по RGB-каналам, размерами фотоизображений) и недостаточной интерпретируемостью решений моделей. Понимание систематических различий в метриках качества фотоизображений между категориями классификации – истинно положительными (ИП), истин-

но отрицательными (ИО), ложноположительными (ЛП) и ложноотрицательными (ЛО) – становится ключом к улучшению качества данных и алгоритмов. Выявление таких различий позволяет установить, что часть ошибок ИИ обусловлена не только недостатками самой модели, но и систематическими проблемами качества входных фотоизображений. Это дает возможность целенаправленно улучшать как данные (предобработка, стандарты съемки, аугментация), так и алгоритмы. В контексте разработки способа для диагностики кожных новообразований, таких как меланомы и другие опухоли, анализ ошибок моделей глубокого обучения играет ключевую роль в повышении надежности и доверия к автоматизированным инструментам ИИ.

Цель исследования: анализ причин ошибочной классификации изображений немеланоцитарных (кератиноцитарных / эпидермальных) опухолей кожи программами на основе технологий ИИ. Для этого проводилось выявление систематических различий в характеристиках данных и визуализацией областей интереса при распознавании изображений для повышения эффективности обучения и использования программ компьютерного зрения.

Материал и методы

Проведен ретроспективный анализ 151 фотоизображения немеланоцитарных (кератиноцитарных / эпидермальных) опухолей кожи, содержащихся в анонимизированной базе данных², в которую входили доброкачественные и злокачественные опухоли. Каждое фотоизображение сопровождалось структурированными метаданными: ID пациента и фотоизображения, пол, возраст, заключительный клинический диагноз по МКБ-10, локализация, тип и вид опухоли, стадия и TNM (для злокачественных), тип снимка, дата, модель камеры, разрешение, формат и MD5-хэш файла.

Каждое фотоизображение было классифицировано программами ИИ как «злокачественное» или «доброкачественное». По сравнению с окончательным клиническим диагнозом сформированы четыре категории:

¹ Skin cancer statistics. World Cancer Research Fund. URL: <https://www.wcrf.org/cancer-trends/skin-cancer-statistics> (20.11.2025).

² Корабельников Д.И., Ламоткин А.И. База данных новообразований кожи. Свидетельство о государственной регистрации базы данных № 2025624714, 2025.

- ИП – злокачественные опухоли, правильно классифицированные ИИ;
- ИО – доброкачественные опухоли, правильно классифицированные ИИ;
- ЛП – доброкачественные опухоли, ошибочно классифицированные ИИ как злокачественные;
- ЛО – злокачественные опухоли, ошибочно классифицированные ИИ как доброкачественные.

Среди пациентов, фотоизображения опухолей которых были включены в базу данных, было 110 мужчин (72,8%) и 41 женщина (27,2%); возраст пациентов – от 18 до 92 лет (средний возраст составил $63,8 \pm 12,4$ года). Гендерная асимметрия обусловлена спецификой контингента ведомственной медицинской организации, в которой проводился сбор клинического материала.

Фотографирование выполнялось с расстояния 8–15 см от поверхности кожи с использованием камер смартфонов Samsung Galaxy A8 (2018) (SM-A530F, Южная Корея) с основной камерой 16 Мп (f/1.7, фокусное расстояние 4 мм, экв. 26 мм, размер сенсора 1/2.8") и автоматической LED-вспышкой; Vivo X100 Ultra (Китай) с основной камерой 50 Мп (Samsung ISOCELL HP9, 1/0.98", f/1.75, фокусное расстояние 22 мм) и автоматической двойной LED-вспышкой; цифрового фотоаппарата Sony DSLR-A100 с матрицей 10,2 Мп (APS-C CCD, 23.6 × 15.8 мм) и макрообъективом Sony 50 mm f/1.4 (SAL-50F14, Япония).

Все фотоизображения были обработаны двумя программами компьютерного зрения на основе технологий ИИ: «Derma Onko Check» и «Melanoma Check». Для визуализации областей, существенно влияющих на классификацию изображений, применены два метода объяснимого ИИ: метод взвешенного картирования активации классов (Score-CAM), который показывают, на какие участки изображения опирается глубокая нейронная сеть при принятии решения и метод определения чувствительности к окклюзии (Occlusion Sensitivity), который последовательно закрывает небольшие участки изображения и измеряет, как это влияет на уверенность модели в предсказании класса.

Статистическая обработка выполнена в среде Python 3.11 (Google Colab). Используются библиотеки pandas, seaborn, scipy, statsmodels. Применены описательная статистика и визуализация с помощью следующих методов: «ящик с усами», он же боксплот и парные диаграммы рассеяния корреляций. Для выявления систематических различий параметров изображений (яркость, контраст, энтропия, размытие, RGB-метрики) проведены однофакторный дисперсионный анализ (ANOVA) и попарное сравнение характеристик классов (t-критерий Уэлча).

Результаты

Визуализации метрик качества фотоизображений

Ящики с усами позволяют выявить центральные тенденции и разброс метрик в категориях ошибок. Например, для яркости (англ. brightness) медианные значения варьируются от 0,6914 (ИП) до 0,7994 (ЛП), с межквартильным размахом от 0,0682 до 0,0935. Энтропия показывает более высокие медианные значения для ИП (4,8584) по сравнению с ЛП (4,6414), что может указывать на сложность текстур в правильно классифицированных случаях. Выбросы (отдельные точки за пределами «усов») наблюдаются, например, в метрике доли пере-

экспонированных пикселей (англ. clipped_high) (медиана – 0,0123–0,0839), что указывает на аномалии освещения в отдельных снимках.

Статистически значимые различия ($p < 0,05$ по однофакторному дисперсионному анализу) наблюдаются для ряда ключевых метрик: медиана яркости ниже в ИП случаях (0,6914) по сравнению с ЛП (0,7994); медиана энтропии выше в ИП (4,8584), чем в ЛП (4,6414). Эти различия в медианных значениях и разбросе метрик между группами могут быть связаны с особенностями алгоритма ИИ, который лучше распознает сложные текстуры при естественном освещении. Выбросы среди значений доли переэкспонированных пикселей (англ. clipped_high) и среди значений доли недоэкспонированных пикселей (англ. clipped_low) могут указывать на проблемы с качеством фотоизображений (переэкспозиция или недодержка), что требует дополнительной предобработки данных (рис. 1).

Парные диаграммы рассеяния корреляций между метриками качества фотоизображений показывают, что яркость имеет сильную отрицательную корреляцию с энтропией ($-0,7117$), т. е. фотоизображения с более высокой энтропией (сложностью текстуры) имеют меньшую яркость. Контраст положительно коррелирует с энтропией (0,6431), что логично, так как высокая контрастность часто сопровождается сложными паттернами. Резкость демонстрирует слабую положительную корреляцию с энтропией (0,2736) и слабую отрицательную с яркостью ($-0,2215$).

Сильная отрицательная корреляция между яркостью и энтропией может указывать на то, что ИИ имеет трудности с классификацией фотоизображений с низкой яркостью и высокой текстурной сложностью (например, меланомы с неравномерной пигментацией). Эти данные могут быть использованы для корректировки весов признаков в модели или для дополнительной нормализации данных перед классификацией (рис. 2).

Карты пространственного распределения различий

Классифицировать разницу истинных и ложных заключений моделей ИИ можно с помощью метода статистического сравнения карт пространственного распределения (карты различий) пикселей фотоизображений для RGB-каналов. Этот метод позволяет исследовать различия в значениях пикселей на карте и использовать эту информацию для классификации результатов. Карты различий показывают расхождение в центральных областях фотоизображений, связанные с сосудистыми паттернами меланом.

Сравнение ИО и ЛО результатов демонстрирует зоны с высоким различием ($\sim 0,7-1,6$) – желтые пятна в верхней-центральной области. Бирюзовые зоны ($\sim 0,3-0,5$) образуют нерегулярные пятна. Темные области ($\sim 0,1-0,2$) покрывают большую часть сетки, особенно верхнюю половину, показывая высокую однородность. Топовые интервалы (80–107, 26–53, 53–80) указывают на смещение к более высоким и низким тонам, что может отражать различия в интенсивности краев и фона. Ошибки ЛО могут быть связаны с шумом меток или недостаточной чувствительностью модели к слабым сигналам, особенно если обучающие данные содержат артефакты или недостаток положительных примеров (рис. 3а).

При сравнении ИО и ЛП результатов показаны высокие различия (до $\sim 1,6$) в центральной области, с яркими

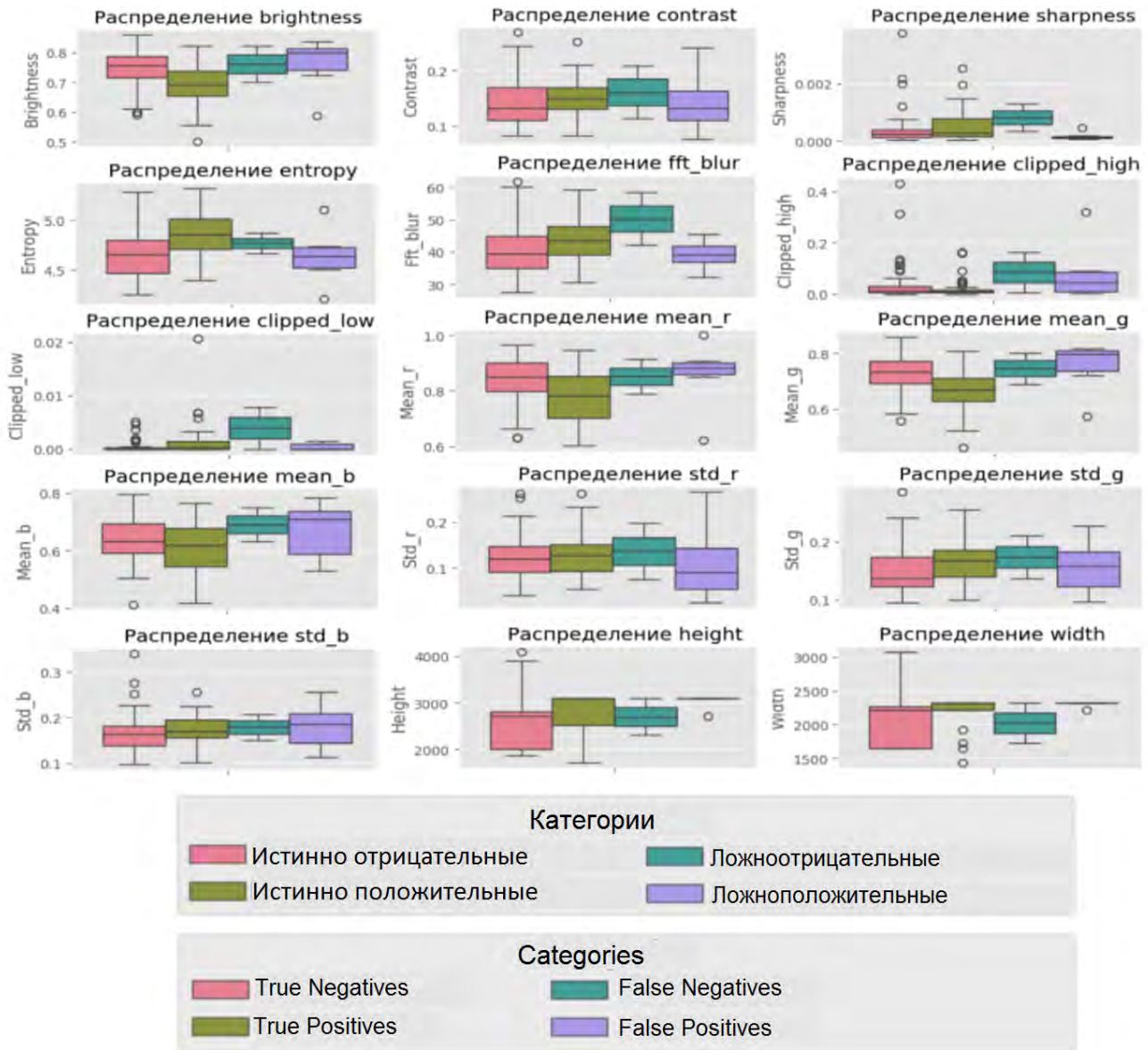


Рис. 1. Визуализация распределения метрик качества фотоизображений по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Примечание: brightness – яркость; contrast – контраст; sharpness – резкость; entropy – энтропия (мера информационной насыщенности); fft_blur – размытость по быстрому преобразованию Фурье; clipped_high – доля пересвеченных пикселей; clipped_low – доля недосвеченных пикселей; mean_r, mean_g, mean_b – средние значения красного (R – Red), зеленого (G – Green) и синего (B – Blue) каналов; std_r, std_g, std_b – стандартное отклонение красного, зеленого и синего каналов; height – высота изображения (пиксели); width – ширина изображения (пиксели).

Fig. 1. Visualization of the distribution of image quality metrics by classification error categories in the «Derma Onko Check» and «Melanoma Check» programs

Note: fft_blur – blurriness according to the fast Fourier transform; clipped_high – proportion of overexposed pixels; clipped_low – proportion of underexposed pixels; mean_r, mean_g, mean_b – average values of the red (R – Red), green (G – Green) and blue (B – Blue) channels; std_r, std_g, std_b – standard deviation of the red, green and blue channels; height – image height (pixels); width – image width (pixels).

желтыми ядрами и зелеными структурами (~1,0–1,4). Темные зоны (~0,2–0,4) преобладают на краях и в центральных областях, указывая на согласованность вне центральной области. Топовые интервалы (13–26, 39–52, 26–39) показывают низкие интенсивности, что может указывать на шум или слабые артефакты. Ошибки ЛП могут возникать из-за избыточной чувствительности к шуму, особенно если модель переобучена на положительные образцы с артефактными особенностями, имитирующими структуры ИО (рис. 3b).

Большие различия (до ~1,4) при сравнении ИП и ЛО результатов сосредоточены в центральной области с

ярким желтым ядром особенно в верхних и нижних квадрантах. Зеленые / бирюзовые градиенты (~0,8–1,2) простираются наружу, тогда как темные зоны (~0,2–0,4) видны в середине периферии. Топовые интервалы (23–47, 141–165, 117–141) охватывают широкий диапазон, включая высокие интенсивности, что может указывать на значительные вариации в сигнальных областях. Ошибки ЛО могут быть вызваны недостаточной чувствительностью к центральным пикам, особенно если фотоизображения ЛО имеют ослабленные сигналы из-за размытия или низкого разрешения, что требует улучшения обработки фотоизображений в обучении (рис. 4a).

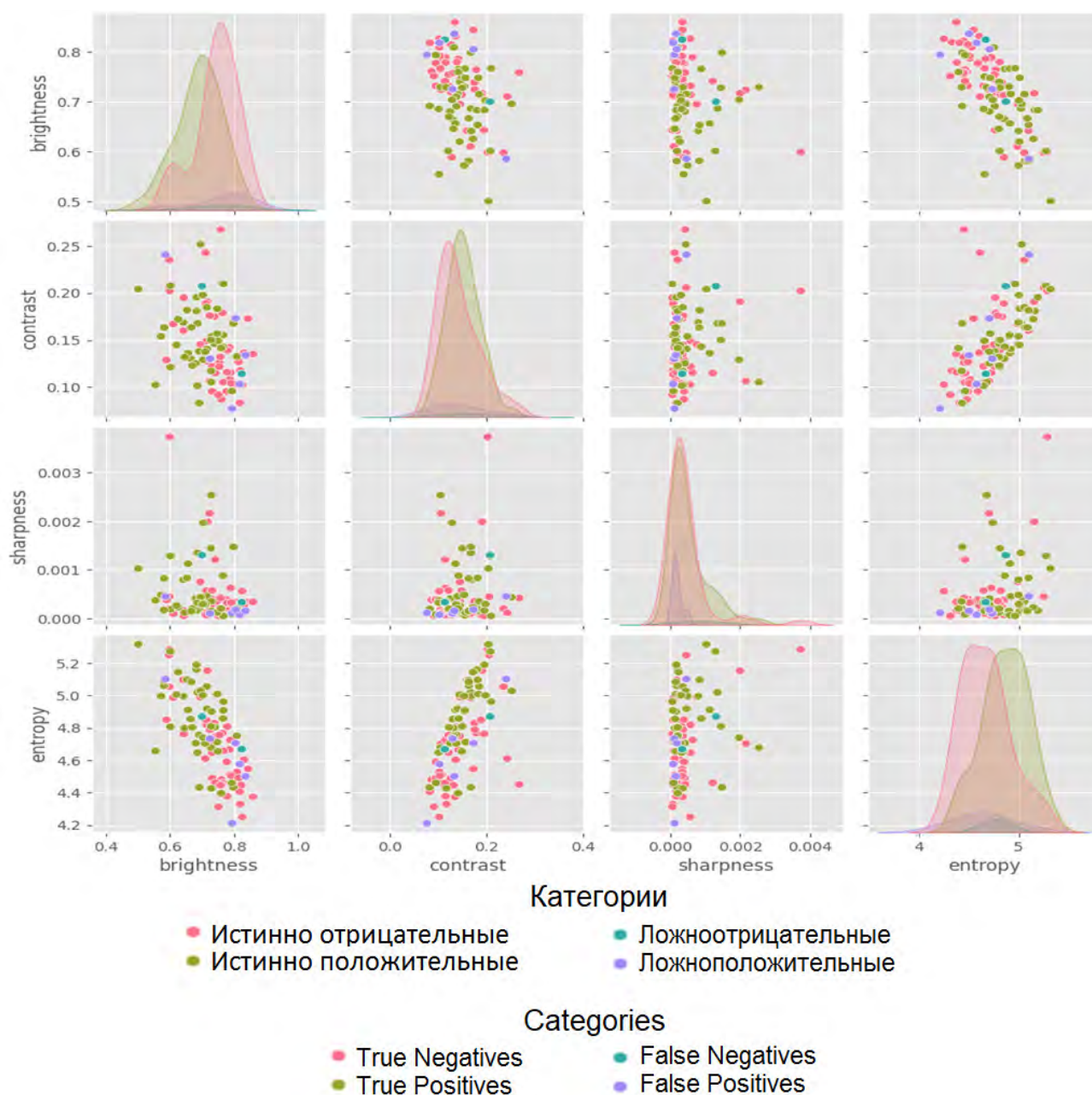


Рис. 2. Парные диаграммы рассеяния (англ. pairplot) корреляций между метриками качества фотоизображений (brightness – яркость, contrast – контраст, sharpness – резкость, entropy – энтропия) по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Fig. 2. Pairplot of correlations between photo image quality metrics (brightness, contrast, sharpness, entropy) by classification error categories in the «Derma Onko Check» and «Melanoma Check» programs

Области с наибольшими различиями ИП и ЛП результатов (желтый, ~0,8–1,0) находятся по периферии, с рассеянными средними различиями в центре (~0,4–0,6) в спиралевидном стиле. Темные полосы (~0,1–0,3) на внешних кольцах и между рукавами указывают на перекрытие периферийных элементов. Топовые интервалы (30–45, 15–30, 0–15) сосредоточены на низких тонах, что может отражать шум или слабые границы. Ошибки ЛП могут быть результатом переобобщения модели на аугментированных данных, где шум или тени имитируют структуры ИП фотоизображений, требуя более строгой регуляризации (рис. 4b).

Статистическая значимость различий

Согласно результатам дисперсионного анализа

(ANOVA), выявлены статистически значимые различия ($p < 0,05$): для яркости ($F = 5,1848$), энтропии ($F = 5,2509$), размытия ($F = 3,1136$), среднего значения красного канала ($F = 3,3812$), среднего значения зеленого канала ($F = 5,3315$), высоты фотоизображения ($F = 2,5645$). t-критерий Уэлча подтвердил значимые различия по яркости между ИО и ИП классами ($t = 3,6619$; $p = 0,0004$); по энтропии между ИО и ИП ($t = -3,8187$; $p = 0,0002$); по размытию между ИО и ИП ($t = -2,4678$; $p = 0,0155$), а также между ИП и ЛП классами ($t = 2,3417$; $p = 0,0450$). Показаны различия по среднему значению красного канала между классами ИО и ИП результатов ($t = 2,9321$; $p = 0,0043$); по среднему значению зелёного канала – между ИО и ИП ($t = 3,5506$; $p = 0,0006$); по дисперсии зеленого канала – между ИО и ИП ($t = -2,0714$; $p = 0,0413$); по резкости меж-

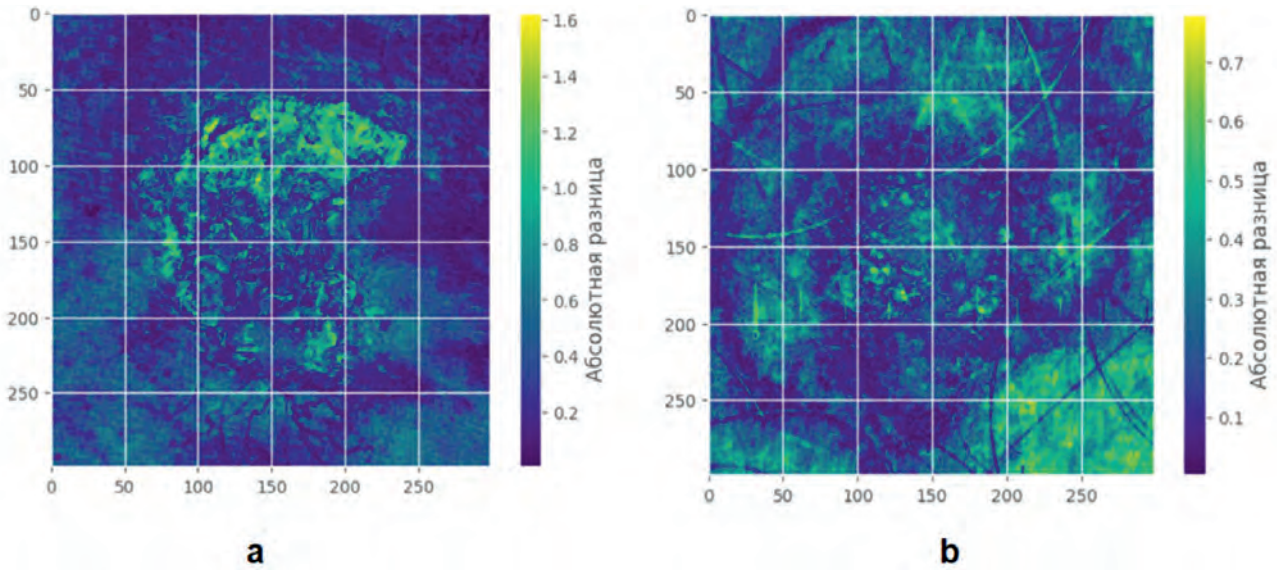


Рис. 3. Карта различий. Истинно отрицательные vs ложноотрицательные (a) и истинно отрицательные vs ложноположительные (b)
 Fig. 3. Difference map. True negatives vs. false negatives (a) and true negatives vs. false positives (b)

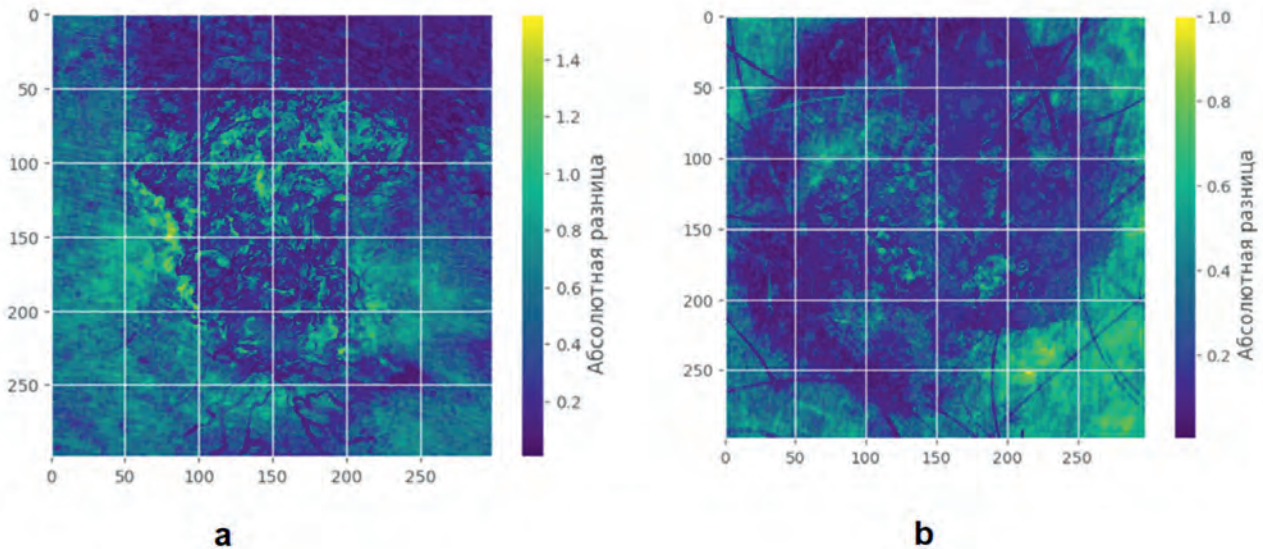


Рис. 4. Карта различий. Истинно положительные vs ложноотрицательные (a) и истинно положительные vs ложноположительные (b)
 Fig. 4. Difference map. True positives vs. false negatives (a) and true positives vs. false positives (b)

ду ИО и ЛП ($t = 2,2982$; $p = 0,0270$), а также между ИО и ЛП результатами ($t = 3,3535$; $p = 0,0021$).

Это подтверждает влияние яркости, энтропии и цветовых средних изображений на появление ошибок с наибольшими различиями между ИП и ЛП результатами и обосновывает необходимость фокуса на этих метриках в доработке моделей. Выбросы среди значений доли переэкспонированных пикселей (0,0123–0,0839) указывают на влияние этого параметра на ЛО и ЛП результаты.

Показано существование разнонаправленных корреляций между метриками качества изображений (яркость – энтропия: $r = -0,7117$; контраст–энтропия: $r = 0,6431$), что указывает на взаимосвязь освещения и текстурной сложности и может косвенно влиять на ошибки классификации.

Ошибочные заключения, которые проанализированы методами объяснимого ИИ

Для понимания причин ошибочной классификации

были визуализированы области интереса моделей. На рис. 5a представлено изображение базальноклеточного рака кожи, ЛО результат. Методом Score-Cam сгенерирована тепловая карта (визуализация, в которой цветом показана степень влияния области изображения на предсказание модели; красные зоны наиболее значимые для решения модели), которая правильно выделила основную область, наиболее значимую для оценки ее моделью. Методом Occlusion Sensitivity выявлена центральная область опухоли, которую характеризуют желто-зеленые тона (0,2–0,6) с темно-синим ядром (–0,2). На периферии опухоль окружена красными пятнами (0,8–1,0), а голубые участки доминируют в периферийных зонах изображения (–0,2). Вероятно, модель неправильно классифицировала данные, полагаясь только на центральную область и потеряв контекст краев из-за шума в данных областях.

На рис. 5b представлено изображение себорейного кератоза, ЛП результат. Методом Score-CAM сгенерирована тепловая карта, которая в целом правильно

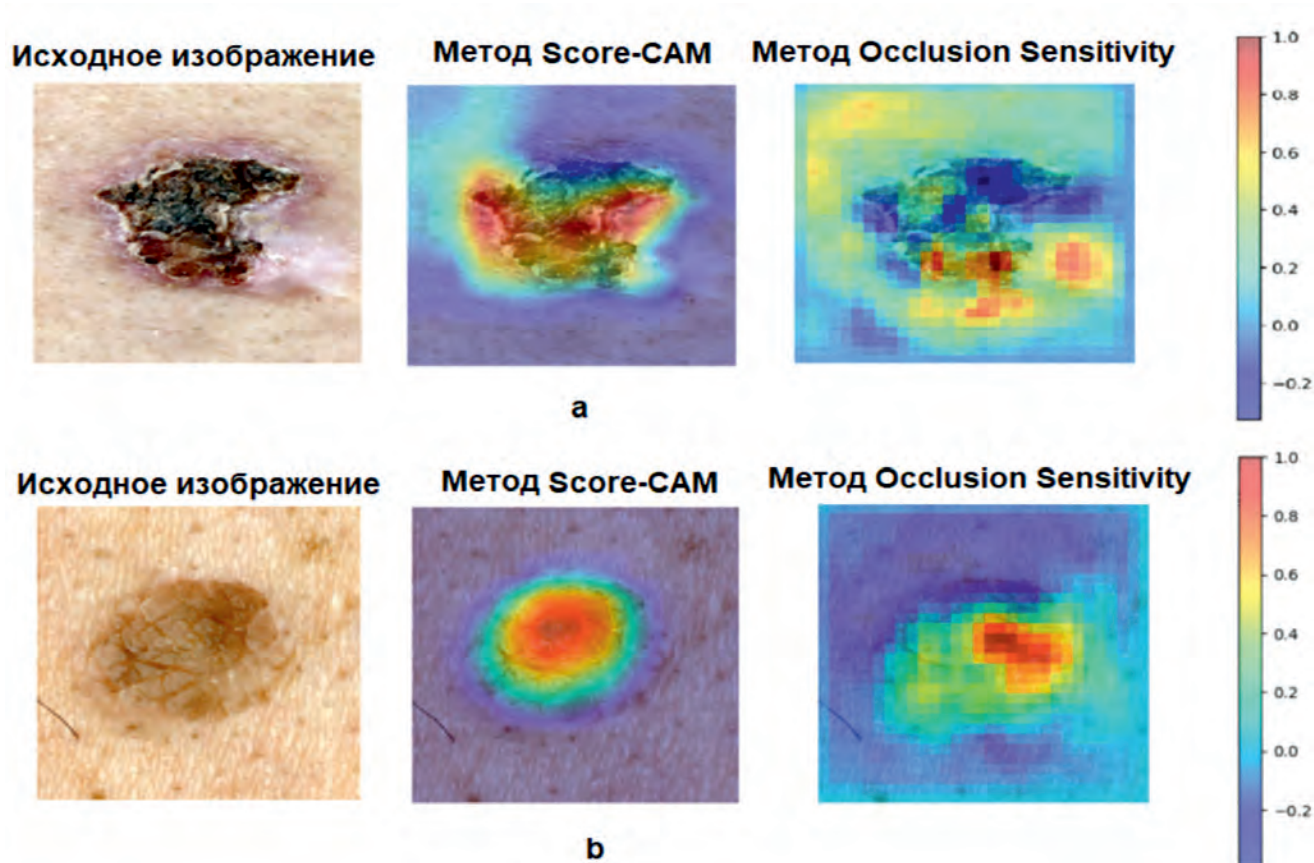


Рис. 5. Ложноотрицательные заключения программ
Fig. 5. False negative conclusions of the programs

выделила центральную область опухоли как наиболее значимую для оценки моделью, с акцентом на текстуру и границы. Методом Occlusion Sensitivity выявлена горизонтально-вытянутая центральная область опухоли, которая демонстрирует ярко-красные тона (0,8–1,0) в ядре, окруженные желтыми и оранжевыми зонами (0,4–0,6), соответствующими текстурным элементам и границам. Голубые и синие участки (–0,2) преобладают в периферийных зонах, указывая на фоновые элементы, которые «мешают» модели. Вероятно, модель неправильно классифицировала доброкачественное образование как злокачественное, чрезмерно полагаясь на центральную область, интерпретируя его текстуру как признак злокачественности и игнорируя периферийные зоны из-за низкого контраста или артефактов. Причина ошибочного ЛП заключения, возможно, обусловлена переобучением на коррелирующих текстурах (например, шероховатость себорейного кератоза, похожая на злокачественные паттерны) с потерей глобального контекста фона.

В данных примерах ХАИ-анализ (объяснимый ИИ) показал, что в ЛО ошибках модель отвлекается на фон (нормальную кожу), а в ЛП – переобучается на артефактах (тени, волосы), игнорируя края опухоли. Выбросы

значений доли переэкспонированных и недоэкспонированных пикселей (0,0123–0,0839) указывают на проблемы качества фотоизображений.

После выявления ключевых факторов ошибок классификации (низкая яркость, высокая энтропия, переэкспозиция и текстурная сложность фотоизображений) было проведено дообучение моделей с учетом этих особенностей и выпущены версии 2.0 программ «Derma Onko Check»³ и «Melanoma Check»⁴. В решениях программы «Derma Onko Check» диагностическая точность возросла с 90,9 до 92,4%, чувствительность практически не изменилась (94,7% вместо 94,4%), специфичность увеличилась с 89,6 до 91,8% (+2,2 п.п.). Соответственно, доля ЛП заключений снизилась с 10,4 до 8,2%, а ЛО незначительно увеличилась (на 0,5%). В решениях программы «Melanoma Check» диагностическая точность возросла с 84,3 до 92,4%, чувствительность увеличилась с 92,1 до 94,9%, а специфичность – с 82,3 до 92,1%. Это означает снижение доли ЛП заключений на 55% (до 7,9%), а ЛО на 35% (до 5,1%).

Обсуждение

Как и в большинстве моделей ИИ для медицинской

³ Ламоткин А.И., Ламоткин И.А., Корабельников Д.И. Программа для автоматизированной визуальной идентификации злокачественных и доброкачественных опухолей кожи «Derma Onko Check» ver. 2.0. Свидетельство о государственной регистрации программы для ЭВМ № 2025661513, 2025.

⁴ Ламоткин А.И., Ламоткин И.А., Корабельников Д.И. Программа для автоматизированной визуальной идентификации меланомы кожи «Melanoma Check» ver. 2.0. Свидетельство о государственной регистрации программы для ЭВМ № 2025661514, 2025.

визуализации, ошибки классификации могут возникать по нескольким причинам. К ним относятся: переобучение на артефактах, недостаточная фокусировка на релевантных признаках или влияние фоновых элементов фотоизображения. Для анализа таких ошибок мы задействовали методы Score-CAM и Occlusion Sensitivity, которые позволяют визуализировать, на каких областях фотоизображения модель фокусируется при проведении классификации.

Для оценки того, правильно ли модель фокусируется на образованиях на фотоизображениях, мы проанализировали все фотоизображения кератиноцитарных / эпидермальных опухолей в датасете⁵ с помощью объяснимого ИИ метода Score-CAM. Этот метод генерирует тепловые карты, где более яркие области (обычно красные) указывают на зоны, наиболее влияющие на заключение модели. В большинстве случаев модель правильно обращала внимание на само образование: пигментация, асимметрию границ, цветовые вариации и другие клинические признаки. Однако в некоторых ошибочных случаях тепловая карта выделяла нерелевантные области, такие как окружающая кожа, волосы или артефакты освещения. Интересно, что при обрезке (маскировке) этих «лишних» областей модель начинала правильно фокусироваться на опухоли, но при этом выдавала неверное заключение. Причины этого феномена могут быть связаны с переобучением на коррелирующих, но не причинных признаках (например, модель «учится» ассоциировать определенные текстуры кожи с злокачественностью, даже если они не относятся к опухоли); потерей контекста – обрезка удаляет вспомогательные признаки, такие как градиенты цвета вокруг опухоли, что приводит к недооценке злокачественности; шумом на фотоизображениях, где модель полагается на артефакты для компенсации слабых сигналов в основной области. Такие случаи подчеркивают необходимость баланса между фокусом на локальных признаках и глобальным контекстом фотоизображения.

Для более детального определения того, на какие текстуры, края и другие элементы обращает внимание модель, мы применили метод объяснимого ИИ Occlusion Sensitivity. Этот подход работает по следующему принципу: входное фотоизображение разделяется на патчи (небольшие блоки), каждый из которых последовательно «закрывается» (например, заменяется на среднее значение или черный цвет), после чего оценивается изменение в заключении модели. Если окклюзия патча приводит к значительному снижению уверенности в заключении модели по определению класса (или изменению класса), этот патч считается важным. Результат визуализируется в виде карты чувствительности: положительные значения (красные области на шкале от 1,0 до 0) указывают на патчи, чье удаление уменьшает вероятность правильного заключения в определении целевого класса (т. е. они важны для правильного решения), а отрицательные (синие, до -1.0) – на патчи, чье удаление увеличивает вероятность, подразумевая, что они «мешают» модели. Интерпретация карты позволяет выявить, фокусируется ли модель на клинически релевантных признаках (например, неровные края опухоли) или на шумовых элементах (волосы, фон). Этот метод особенно полезен в медицинской визу-

ализации, так как он не зависит от внутренних градиентов сети и может выявлять скрытые предвзятости.

При анализе методом Occlusion Sensitivity мы обнаружили, что в случаях ЛО заключения (модель классифицирует злокачественную опухоль как доброкачественную) важные пиксели часто располагались в областях обычной кожи без образований. Это указывает на то, что модель «отвлекается» на фоновые текстуры, возможно, из-за недостаточного контраста опухоли или переобучения на датасетах с преобладанием нормальной кожи. Однако был и ЛП случай решения, когда модель в большей степени обращала внимание на нормальную кожу, но все равно классифицировала опухоль как злокачественную – здесь, вероятно, повлияли артефакты (тени, волосы), которые модель ошибочно интерпретировала как признаки меланомы. Дополнительно, анализируя предоставленные фотоизображения, мы видим конкретные примеры таких ошибок в контексте классификации.

В рамках данного исследования мы интегрируем методы объяснимого искусственного интеллекта (XAI), такие как Score-CAM и Occlusion Sensitivity, для визуализации фокуса моделей на релевантных областях фотоизображений, с комплексным статистическим и визуализационным анализом. Использование метода Score-CAM позволяет оценить, на какие признаки (например, текстуру или края опухоли) модель обращает внимание, в то время как метод Occlusion Sensitivity выявляет критические пиксели, влияющие на заключение. Эти методы дополняются подходами раірplot корреляций, boxplots метрик, картами различий.

Анализ ошибок классификации выявил значимые факторы: яркость, энтропию и цветовые характеристики (ANOVA: $F=3,11-5,33$; $p < 0,05$). ИП имеют более низкую яркость (0,6914) и высокую энтропию (4,8584), тогда как ЛП – более высокую яркость (0,7994) и низкую энтропию (4,6414), что указывает на трудности ИИ с текстурной сложностью и освещением. Корреляции (яркость – энтропия: $-0,7117$; контраст – энтропия: $0,6431$) подтверждают влияние текстур на ошибки. Карты различий показывают расхождения в центральных областях фотоизображений, особенно в каналах G и B (различия до 1,4–1,6), связанные с сосудистыми паттернами меланом. Выбросы среди значений доли переэкспонированных пикселей (англ. clipped_high) (0,0123–0,0839) указывают на аномалии переэкспозиции, влияющие на ЛО и ЛП.

Ограничения исследования заключаются в относительно малой выборке, гендерных и возрастных особенностях пациентов, связанных с контингентом ведомственной медицинской организации. Данные ограничения не снижают практической ценности работы, поскольку была поставлена цель: проанализировать причины ошибок ИИ и разработать рекомендации по уменьшению их количества в дальнейшем.

Таким образом, дообучение с учетом выявленных факторов ошибок (яркость, энтропия, переэкспозиция, текстурная сложность) позволило существенно сократить количество как ЛП, так и ЛО заключений и повысить общую диагностическую точность обеих программ, при этом особенно выраженный прирост показателей наблюдался у программы «Melanoma Check».

⁵ Корабельников Д.И., Ламоткин А.И. База данных новообразований кожи. Свидетельство о государственной регистрации базы данных № 2025624714, 2025.

Заключение

Настоящее исследование демонстрирует эффективность комбинации методов объяснимого ИИ и статистических подходов для анализа ошибок в дифференциальной диагностике немеланокитарных опухолей кожи с использованием программ ИИ «Derma Onko Check» и «Melanoma Check». Выявленные систематические различия в метриках фотоизображений, такие как повышенная яркость и низкая энтропия в ЛП случаях, а также фокус модели на фоновых артефактах в ЛО случаях подчеркивают ключевую роль качества данных для надежности решений ИИ. Значимые корреляции ряда метрик и различия в RGB-каналах разных классов изображений подтверждают влияние освещения, текстуры и размытости на заключения моделей ИИ.

На основании проведенного исследования сформулированы ключевые рекомендации для пользователей приложений, направленные на минимизацию ошибок классификации за счет улучшения качества изображений, нормализации метрик и снижения вариабельности данных. Для решения задачи дифференциальной диагностики кератиноцитарных / эпидермальных опухолей кожи с помощью программ компьютерного зрения пользователям рекомендуется придерживаться следующих правил фотографирования:

- обеспечивать равномерное освещение без теней, поддерживая яркость фотоизображений на уровне ниже 0,75 в нормализованной шкале, поскольку более высокая яркость повышает вероятность ЛП результатов;
- избегать недоэкспозиции или переэкспозиции для снижения вероятности ЛО результатов;
- использовать функцию автоматического баланса белого для нормализации цветовых каналов (RGB), с фокусом на G и B каналах;
- стремиться к контрастности фотоизображения выше 0,20 для четкой визуализации границ опухоли, что уменьшает долю ЛО результатов;
- минимизировать размытие, поддерживая резкость выше среднего уровня и применяя фильтры усиления резкости и текстуры, стабилизируя камеру во время съемки;
- осуществлять съемку крупным планом для достижения энтропии выше 4,8, избегая артефактов; желательно активировать режим макросъемки и использовать аугментацию текстуры в приложении для усиления значимости центральных областей;
- обеспечивать высокое разрешение изображений в диапазоне 2000–3000 пикселей, центрируя опухоль в кадре для минимизации фоновых элементов и снижения вариаций.

Кроме того, сформулированы ключевые рекомендации для разработчиков систем ИИ, предназначенных для дифференциальной диагностики опухолей кожи. Для оптимизации качества данных и уменьшения влияния различного рода выбросов рекомендуется на этапах предобработки, аугментации и обучения:

- нормализовать цветовые каналы с использованием автоматического баланса белого и гамма-коррекции, поскольку вариации в RGB-каналах коррелируют с ошибками классификации, особенно в случаях переэкспозиции или недоэкспозиции;
- применять фильтры усиления текстуры для повышения энтропии фотоизображений выше 4,8, так как

это помогает в распознавании сложных текстур, снижая ЛО ошибки;

- использовать методы коррекции размытости (например, линейный фильтр наблюдаемого зашумленного процесса) для повышения резкости изображения и фокуса модели на релевантных областях;
- внедрять аугментацию яркости и цвета при обучении моделей, чтобы компенсировать отрицательную корреляцию яркости с энтропией и повысить устойчивость к вариабельности данных, особенно в ярких или темных фотоизображениях;
- обеспечивать мониторинг ключевых метрик (яркость, энтропия, контраст, доля недоэкспонированных и переэкспонированных пикселей после каждой эпохи обучения для предотвращения переобучения на артефактах (фон, волосы или тени).

Реализация данных рекомендаций позволяет снизить частоту ошибок классификации, существенно повысить диагностическую точность и обеспечить устойчивость моделей к реальной вариабельности клинических фотографий немеланокитарных опухолей кожи.

Литература / References

1. Stafford H., Buell J., Chiang E., Ramesh U., Migden M., Nagarajan P. et al. Non-melanoma skin cancer detection in the age of advanced technology: a review. *Cancers (Basel)*. 2023;15(12):3094. <https://doi.org/10.3390/cancers15123094>
2. Ferlay J., Colombet M., Soerjomataram I., Dyba T., Randi G., Bettio M. et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer*. 2018;103:356–387. <https://doi.org/10.1016/j.ejca.2018.07.005>
3. Fahim Y.A., Hasani I.W., Kabba S., Ragab W.M. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. *European Journal of Medical Research*. 2025;30(1):848. <https://doi.org/10.1186/s40001-025-03196-w>
4. Behara K., Bhero E., Agee J.T. AI in dermatology: a comprehensive review into skin cancer detection. *PeerJ Computer Science*. 2024;10:e2530. <https://doi.org/10.7717/peerj-cs.2530>
5. Melarkode N., Srinivasan K., Qaisar S.M., Plawiak P. AI-Powered Diagnosis of Skin Cancer: A Contemporary Review, Open Challenges and Future Research Directions. *Cancers*. 2023;15(4):1183. <https://doi.org/10.3390/cancers15041183>
6. Kolla L., Parikh R.B. Uses and limitations of artificial intelligence for oncology. *Cancer*. 2024;130(12):2101–2107. <https://doi.org/10.1002/cncr.35307>
7. Ламоткин А.И., Корабельников Д.И., Олисова О.Ю., Ламоткин И.А. Эффективность предварительной дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с помощью программы искусственного интеллекта Derma Onko Check. *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2025;18(2):261–270. <https://doi.org/10.17749/2070-4909/farmakoeconomika.2025.294>
8. Ламоткин А.И., Корабельников Д.И., Олисова О.Ю., Ламоткин И.А. Effectiveness of preliminary differential diagnosis of benign and malignant skin neoplasms using the artificial intelligence program Derma Onko Check. *Farmakoeconomika. Modern Pharmacoeconomics and Pharmacoepidemiology*. 2025;18(2):261–270. (In Russ.). <https://doi.org/10.17749/2070-4909/farmakoeconomika.2025.294>
9. Ламоткин А.И., Корабельников Д.И., Ламоткин И.А. и др. Точность предварительной диагностики злокачественных меланокитарных опухолей кожи с помощью программы искусственного интеллекта. *Медицинский вестник ГВКГ им. Н.Н. Бурденко*. 2025;(1(19)):42–51. <https://doi.org/10.53652/2782-1730-2025-6-1-42-51>
10. Ламоткин А.И., Корабельников Д.И., Ламоткин И.А. et al. Accuracy of preliminary diagnosis of malignant melanocytic skin tumors using an artificial intelligence program. *Medical Bulletin of the Main Military Clinical Hospital named after N.N. Burdenko*. 2025;(1(19)):42–51. (In Russ.). <https://doi.org/10.53652/2782-1730-2025-6-1-42-51>
11. Mevorach L., Farcomeni A., Pellacani G., Cantisani C. A comparison of skin lesions' diagnoses between AI-based image classification, an expert dermatologist, and a non-expert. *Diagnostics (Basel)*. 2025;15(9):1115. -

Информация о вкладе авторов

Корабельников Д.И. – концепция исследования, интерпретация результатов, написание рукописи, финальная редактура; Ламоткин А.И. – сбор и анализ данных, написание рукописи, статистическая обработка, подготовка иллюстраций.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Сведения об авторах

Корабельников Даниил Иванович, канд. мед. наук, доцент, почетный работник сферы образования Российской Федерации, заведующий кафедрой внутренних болезней с курсами семейной медицины, функциональной диагностики, инфекционных болезней, профессиональных болезней медицинского факультета, ректор АНО ДПО «Московский медико-социальный институт им. Ф.П. Гааза», Москва, Россия, e-mail: dkorabelnikov@mail.ru; <http://orcid.org/0000-0002-0459-0488>.

Ламоткин Андрей Игоревич, канд. мед. наук, ассистент, кафедра внутренних болезней с курсами семейной медицины, функциональной диагностики, инфекционных болезней, профессиональных болезней, медицинский факультет, АНО ДПО «Московский медико-социальный институт им. Ф.П. Гааза», Москва, Россия, e-mail: lamotkin.an@yandex.ru; <http://orcid.org/0000-0001-7930-6018>.

Поступила 28.11.2025;
рецензия получена 23.12.2025;
принята к публикации 25.02.2026.

Information on author contributions

Korabelnikov D.I. – research concept, results interpretation, manuscript writing, final editing; Lamotkin A.I. – data collection and analysis, manuscript writing, statistical processing, illustrations preparation.

Conflict of interest: the authors declare no conflict of interest.

Information about the authors

Korabelnikov Daniil Ivanovich, Cand. Sci. (Med.), Associate Professor, Honorary Worker of Education of the Russian Federation, Head of the Department of Internal Medicine with courses in Family Medicine, Functional Diagnostics, Infectious Diseases, and Occupational Diseases, Medical Faculty of the Moscow Haass Medical and Social Institute, Moscow, Russia, e-mail: dkorabelnikov@mail.ru; <http://orcid.org/0000-0002-0459-0488>.

Lamotkin Andrei Igorevich, Cand. Sci. (Med.), Assistant Professor, Department of Internal Medicine with courses in Family Medicine, Functional Diagnostics, Infectious Diseases, and Occupational Diseases, Medical Faculty of the Moscow Haass Medical and Social Institute, Moscow, Russia, e-mail: lamotkin.an@yandex.ru; <http://orcid.org/0000-0001-7930-6018>.

Received 20.04.2025;
review received 26.06.2026;
accepted for publication 24.08.2026.