

Набор дерматоскопических изображений для ранней диагностики злокачественных новообразований кожи

Е. С. Козачок¹ ✉

¹ Институт системного программирования им. В. П. Иванникова РАН
ул. Александра Солженицына, д. 25, г. Москва 109004, Российская Федерация

✉ e-mail: e.kozachok@ispras.ru

Резюме

Цель исследований – формирование набора данных дерматоскопических изображений, содержащих высококачественную разметку клинически значимых признаков кожных новообразований фототипов кожи населения России, предназначенного для ранней диагностики и выявления злокачественных кожных новообразований. Формирование и внедрение наборов дерматоскопических изображений в автоматизированные системы и подходы раннего обнаружения злокачественных новообразований кожи при проведении медицинских обследований пациентов являются актуальными направлениями исследований.

Методы. Предложен подход к формированию набора дерматоскопических изображений с высокой высококачественной разметкой клинически значимых признаков. Основу сформированного набора данных составляют дерматоскопические изображения кожных новообразований с подтвержденными диагнозами, в том числе с применением клинических методов исследования, согласно существующей нозологии пациентов дерматоневрологического профиля населения Российской Федерации врачами дерматологами и онкологами. Отличительной особенностью разработанного набора данных помимо принадлежности к фототипу кожи населения России относится высококачественная разметка клинически значимых признаков, позволяющая применять разработанный набор в методах и алгоритмах машинного обучения и распознавания образов.

Результаты. Сформированный набор дерматоскопических изображений содержит 657 дерматоскопических изображений, сопровождаемых расширенными метаданными и предварительными клиническими заключениями, меланоцитарных (меланома и невус) и не меланоцитарных (плоскоклеточный рак, дерматофиброма, сосудистые поражения, кератоз и др.) новообразований. В основу данного набора заложено распределение как по возрастному критерию, так и по принадлежности и системности течения болезни у пациентов.

Заключение. Практическая направленность сформированного набора дерматоскопических изображений с высококачественной разметкой клинически значимых признаков позволяет применять сформированные изображения как в системах поддержки принятия решений врачей при осуществлении медицинской практики, так и в системах, основанных на применении методов и алгоритмов машинного обучения для ранней диагностики злокачественных новообразований кожи.

Ключевые слова: набор данных; дерматоскопические изображения; кожные новообразования; машинное обучение; распознавание образов.

Конфликт интересов: Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Козачок Е. С. Набор дерматоскопических изображений с высококачественной разметкой клинически значимых признаков для диагностики меланоцитарных новообразований // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2025. Т. 15, № 3. С. 93–111. [https://doi.org/ 10.21869/2223-1536-2025-15-3-93-111](https://doi.org/10.21869/2223-1536-2025-15-3-93-111)

Поступила в редакцию 15.07.2025

Подписана в печать 13.08.2025

Опубликована 30.09.2025

Dermatoscopic image dataset for early diagnosis of malignant skin neoplasms

Elena S. Kozachok¹✉

¹ Ivannikov Institute for System Programming of the Russian Academy of Sciences
25 Alexander Solzhenitsyn Str., Moscow 109004, Russian Federation

✉ e-mail: e.kozachok@ispras.ru

Abstract

The purpose of the research is to develop a dermatoscopic images containing high-quality labeling of clinically significant signs of skin neoplasms of the Russian population skin phototypes, intended for early diagnostics and detection of malignant skin neoplasms. The formation and implementation of sets of dermatoscopic images in automated systems and approaches to the early malignant skin neoplasms detection during medical examinations of patients is a relevant research area.

Methods. An approach to the formation of a dermatoscopic images data set with high-quality labeling of clinically significant features is proposed. The basis of the formed data set is dermatoscopic skin neoplasms images with confirmed diagnoses, including using clinical research methods, according to the existing nosology of the dermatovenereological profile patients of the Russian Federation population by dermatologists and oncologists. A distinctive feature of the developed data set, in addition to belonging to the skin phototype of the Russian population, is the high-quality labeling of clinically significant features, which allows the developed set to be used in methods and algorithms of machine learning and pattern recognition.

Results. The generated data set of dermatoscopic images contains 657 dermatoscopic images, accompanied by extended metadata and preliminary clinical conclusions, of melanocytic (melanoma and nevus) and non-melanocytic (squamous cell carcinoma, dermatofibroma, vascular lesions, keratosis, etc.) neoplasms. This data set is based on the distribution both by age criterion and by the affiliation and course systemic nature of the disease in patients.

Conclusion. The practical focus of the developed dermatoscopic images data set with high-quality marking of clinically significant features allows the use of the generated images both in decision support systems for doctors in medical practice and in systems based on the machine learning methods and algorithms usage for the early malignant skin neoplasms diagnosis.

Keywords: dataset; dermatoscopy images; skin lesions; machine learning; pattern recognition.

Conflict of interest: The Authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Kozachok E.S. Dermatoscopic images data set with high-quality labeling of clinically significant features for the melanocytic neoplasm's diagnosis. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika, informatika. Meditsinskoe priborostroenie* = *Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering*. 2025;15(3):93–111. (In Russ.) <https://doi.org/10.21869/2223-1536-2025-15-3-93-111>

Received 15.07.2025

Accepted 13.08.2025

Published 30.09.2025

Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2025;15(3):93–111

Введение

Меланоцитарные новообразования кожи, включая меланому, представляют собой значительную проблему в современной онкологии. Меланома, несмотря на сравнительно низкую распространенность среди злокачественных опухолей кожи, характеризуется высокой агрессивностью и значительным вкладом в общую смертность от онкологических заболеваний. Ранняя диагностика новообразований данного типа является ключевым фактором, определяющим успешность лечения и общую продолжительность жизни пациентов.

Статистические данные свидетельствуют о неуклонном росте заболеваемости меланомой как в России, так и в мире. В Российской Федерации за период с 2001 по 2021 гг. распространенность меланомы на 100 тысяч населения увеличилась в 1,5 раза [1]. Согласно официальным данным Министерства здравоохранения Российской Федерации (Минздрав России), заболеваемость меланомой кожи в стране демонстрирует тенденцию к росту. В 2021 г. отмечено увеличение числа случаев меланомы кожи на 11,5% по сравнению с 2020 г. [2]. При этом в 2022 г. в России было зарегистрировано 11923 новых случая меланомы, что составляет 14,8% от всех злокачественных новообразований кожи [3].

Особенно тревожным является тот факт, что, несмотря на усилия по раннему выявлению, значительная часть случаев диагностируется на поздних стадиях, что существенно ухудшает

прогноз для пациентов. С 2018 по 2022 гг. одногодичная летальность пациентов со злокачественными новообразованиями снизилась с 22,2% до 19,2%. При этом доля пациентов, состоящих на учете 5 лет и более, увеличилась с 54,4% до 58%, а выявление злокачественных новообразований на I-II стадиях возросло с 56,4% до 59,3% [4].

В условиях растущей заболеваемости и значительной смертности от меланомы особое внимание следует уделять совершенствованию методов ранней диагностики. Дерматоскопия зарекомендовала себя как эффективный инструмент для выявления меланоцитарных новообразований на ранних стадиях [5]. Однако точность диагностики во многом зависит от опыта и квалификации специалиста [6]. В этой связи разработка автоматизированных систем анализа дерматоскопических изображений, основанных на алгоритмах машинного обучения, становится актуальной задачей современной медицины.

Для реализации указанной задачи могут быть использованы методы медицинской информатики, получившие широкое распространение и позволяющие осуществить интеграцию информационных технологий, технологий искусственного интеллекта и методов анализа данных в медицинскую практику [7]. Так алгоритмы глубокого обучения активно применяются для анализа медицинских изображений (МРТ, КТ, рентген и др.), что позволяет осуществлять раннее выявление патологий легких, сердца, головного мозга и т. д. [8].

Применение указанных технологий в онкологической практике может позволить с высокой точностью выявлять ранние стадии рака, анализируя огромные объемы медицинских данных, включая изображения, генетическую информацию и анамнез пациента [9]. При этом системы, обученные на больших наборах данных, способны как распознавать опухолевые образования на изображениях, так и прогнозировать их развитие, что значительно улучшает планирование лечения и персонализацию медицинской помощи [10].

Ключевым элементом в создании таких систем является наличие наборов данных, характеризующихся точной разметкой и выделением клинически значимых признаков, учитывающих особенности фототипов кожи, позволяющих с высокой долей вероятности правильно идентифицировать злокачественные новообразования. Существующие базы данных имеют достаточно большой объем изображений и детальную метаинформацию, но характеризуются недостаточно детальным описанием клинически значимых признаков, основанных на особенностях фототипов кожи. Наличие указанных ограничений позволяет сформировать цель проводимого исследования, состоящую в разработке набора дерматоскопических изображений с высококачественной разметкой, учитывающих особенности фототипов кожи населения России, направленного на повышение точности обнаружения меланоцитарных новообразований в процессе проведения диагностики новообразований.

Материалы и методы

В настоящее время в процессе диагностики онкологических новообразований широкое применение получили структурированные наборы данных, содержащих признаковое пространство кожных новообразований. К указанным наборам дерматоскопических изображений относятся:

- набор данных ISIC (International Skin Imaging Collaboration): ISIC-2016, ISIC-2017, ISIC-2018, ISIC-2019 и ISIC-2020 [11];

- набор данных HAM10000 (Human Against Machine with 10000 training images);

- набор данных PH2;

- набор данных IAD;

- набор данных IRMA;

- набор данных Dermaquest.

Набор данных ISIC содержит более 23000 изображений различных кожных новообразований, включая невусы, меланомы, себорейные кератозы, базальноклеточные карциномы (базалиомы), актинические кератозы, дерматофибромы и другие пигментные поражения [12]. Так набор ISIC-16 состоит из 1279 изображений (900 для обучения и 379 для тестирования методами машинного обучения) невусов и меланом с разрешением от 556×679 до 2848×4828 пикселей для осуществления бинарной классификации или сегментации кожных новообразований.

Набор ISIC-17 содержит 2750 изображений с разрешением от 453×679 до 4499×6748 пикселей следующих кожных новообразований: невус (1372 изображения для обучения, 78 – валидации и

393 – тестирования), себорейный кератоз (254 – обучение, 42 – валидация и 90 – тестирование) и меланома (374 – обучение, 30 – валидация и 117 – тестирование) [13].

Набор изображений ISIC-18 сформирован из набора данных HAM10000 и содержит 10015 изображений с разрешением 450×600 пикселей для реализации многоклассовой классификации на семь классов: невусы (6705 изображений), себорейные кератозы (1099 изображений), базальноклеточные карциномы (514 изображений), актинические кератозы (327 изображений), дерматофибромы (115 изображений), сосудистые поражения кожи (142 изображения) и меланомы (1113 изображений). Кроме того, в набор входят 12500 изображений для реализации бинарной сегментации новообразований [14].

В отличие от ISIC-18 набор ISIC-19 дополнен еще одним классом кожных новообразований – плоскоклеточной карциномой. При этом объем выборки увеличен практически в два раза по сравнению с ISIC-18 и составляет 24765 изображений с разрешением от 600×450 до 1024×1024 пикселей, основанных на изображениях из таких наборов данных, как HAM10000, BCN2000 и др. Набор ISIC-19 состоит из следующих изображений: невусы (12875), себорейные кератозы (2624), базальноклеточные карциномы (3323), актинические кератозы (867), дерматофибромы (239), сосудистые поражения кожи (253), меланомы (4522) и плоскоклеточные карциномы (62).

ISIC-20 содержит наибольшее количество изображений из всех наборов

ISIC, составляющее 33090 дерматоскопических изображений с разрешением 1024×1024 пикселей. Стоит отметить, что указанный набор включает два класса изображений: невусы (32542 изображения) и меланы (548 изображений).

Отличительной особенностью всех наборов ISIC является то, что каждое изображение сопровождается подробной разметкой клинически значимых признаков, таких как асимметрия, границы, цветовые вариации и дерматоскопические структуры, что позволяет использовать указанный набор в процессе проведения диагностики новообразований методами машинного обучения.

Набор данных HAM10000 включает 10 015 дерматоскопических изображений, охватывающих семь классов кожных новообразований [15]:

- 1) меланома: 1113 изображений;
- 2) невус: 6705 изображений;
- 3) базальноклеточная карцинома: 514 изображений;
- 4) актинический кератоз: 327 изображений;
- 5) себорейный кератоз: 1099 изображений;
- 6) дерматофиброма: 115 изображений;
- 7) сосудистые поражения: 142 изображения.

Каждое изображение снабжено метаданными, содержащими информацию о диагнозе, локализации и клинических признаках. Указанный набор изображений является базовым, включенным во многие наборы дерматоскопических изображений, и широко применяемым в процессе обучения и тестирования

алгоритмов классификации методов машинного обучения при решении задач обнаружения кожных новообразований в процессе дерматологических исследований.

PH2 Database содержит 200 дерматоскопических изображений меланоцитарных поражений с разрешением 768×560 пикселей, из которых 80 – доброкачественные невусы, 80 – атипичные невусы и 40 – меланомы [16]. Каждое изображение из набора сопровождается экспертной разметкой, включающей идентификацию границ поражения, цветовых и текстурных характеристик, а также диагностических признаков, таких как наличие сетки, глобул и гомогенных областей. Каждый из представленных параметров оценен экспертами дерматологами.

Рассмотренные наборы данных дерматоскопических изображений являются свободно распространяемыми и представлены в открытом доступе, что позволяет использовать указанные наборы не только во врачебной практике, но и в процессе обучения алгоритмов классификации и моделей машинного обучения для распознавания злокачественных кожных новообразований. Наряду с рассмотренными наборами данных во врачебной практике применяются следующие наборы дерматоскопических изображений: IAD, IRMA, Dermaquest, Dermofit Image Library и DERM12345, которые не доступны для общего пользования, что не позволяет использовать их в процессе обнаружения кожных новообразований методами машинного обучения.

Набор данных IAD содержит 100 изображений, из которых 70 изображений невусов и 30 меланом, с разрешением 700×447 пикселей в цветовой схеме RGB без соответствующих масок сегментации. Однако по отдельному запросу разработчиками набора данных такие маски могут быть предоставлены только для проведения исследований.

Так же, как и набор данных IAD, набор данных IRMA доступен только по специальному запросу к Рейнско-Вестфальском техническому университету Ахена. Он содержит 747 дерматоскопических изображений с разрешением 512×512 пикселей, из которых 187 изображений относятся к меланомам, а 560 – к невусам. Набор данных Dermaquest [17] характеризуется наличием 137 изображений, 76 из которых являются меланомами и 61 – невусами.

Dermofit Image Library содержит 1300 высококачественных изображений различных кожных заболеваний, включая: меланомы (76 изображений), невусы (331 изображение), кератозы (302 изображения), базальноклеточные карциномы (239 изображений), дерматофибромы (65 изображений), сосудистые поражения кожи (199 изображений) и плоскоклеточный рак (88 изображений) [18]. Каждое изображение сопровождается подробным описанием и разметкой клинически значимых признаков, осуществленными врачами-дерматологами.

DERM12345 – набор данных дерматоскопических изображений, содержащий 12345 изображений кожных новообразований фототипов кожи жителей переходной зоны Европы и Азии, преимущественно жителей Турции [19].

Набор DERM12345 состоит из 5 суперклассов, 15 основных классов и 40 подклассов изображений с высоким разрешением (до 3840×2160 пикселей) и экспертными аннотациями, включая классы новообразований:

1) меланоцитарные доброкачественные: 10043 изображения;

2) меланоцитарные злокачественные (меланома): 400 изображений;

3) немеланоцитарные доброкачественные (кератоз, дерматофиброма, сосудистые поражения): 1077 изображений;

4) немеланоцитарные неопределенные (актинический кератоз): 58 изображений;

5) немеланоцитарные злокачественные (саркома Капоши, болезни Боуэна и Педжета молочной железы, базально-клеточный и плоскоклеточный рак и др.): 767 изображений.

Распределение изображений в представленных наборах дерматоскопических изображений приведено ниже (табл. 1).

Таблица 1. Распределение изображений по типам кожных новообразований в наборах данных дерматоскопических изображений

Table 1. Image distribution by skin lesion types in dermatoscopic image data sets

Набор данных Dataset	Кол-во изображений Images	Невус Nev	Себорейный кератоз SK	Базально-клеточная карцинома BCC	Актинический кератоз АК	Дерматофиброма DF	Сосудистые поражения VL	Плоскоклеточный рак SC	Меланома Mel
ISIC-16	1279	727	—	—	—	—	—	—	173
ISIC-17	2750	1843	386	—	—	—	—	—	521
ISIC-18	10015	6705	1099	514	327	115	142	—	1113
ISIC-19	24765	12875	2624	3323	867	239	253	62	4522
ISIC-20	33090	32542	—	—	—	—	—	—	548
HAM 10000	10015	6705	1099	514	327	115	142	—	1113
PH2	200	160	—	—	—	—	—	—	40
IAD	100	70	—	—	—	—	—	—	30
IRMA	747	560	—	—	—	—	—	—	187
Dermaquest	137	61	—	—	—	—	—	—	76
Dermofit Image Library	1300	331	257	239	45	65	199	88	76
DERM 12345	12345	8278	607	423	58	184	282	303	400

Проведенный анализ наборов данных дерматоскопических изображений позволяет сделать вывод о наличии достаточно большого объема изображений, содержащих злокачественные новообразования кожи, разработанных с учетом специфики и требований методов и алгоритмов машинного обучения, а также фототипов кожи различных стран и этнических групп. В то же время существенным недостатком рассмотренных наборов является то, что в процессе их формирования не учитываются особенности фототипов кожи населения России, что, в свою очередь, не позволяет использовать данные наборы в процессе диагностики меланоцитарных новообразований на территории Российской Федерации.

Для формирования набора данных дерматоскопических изображений с высококачественной разметкой клинически значимых признаков, учитывающего особенности фототипов кожи населения России (I–IV фототипы по классификации Фицпатрика), реализованы следующие этапы:

1) получение информированного согласия у пациентов дерматовенерологического профиля на исследование кожных новообразований;

2) дерматоскопия пораженных областей кожи врачом-дерматологом и/или онкологом с целью постановки первичного диагноза;

3) создание снимка высокой четкости новообразования и сбор метаданных, содержащих обезличенную информацию

о пациенте (возраст, пол, информация о типе новообразования, предварительном диагнозе и пр.);

4) проведение клинических исследований (биопсия и гистологический анализ) для подтверждения (опровержения) первичного диагноза (в случае необходимости по решению врача);

5) добавление дерматоскопического изображения и метаданных в набор данных дерматоскопических изображений;

6) разметка клинически значимых признаков (асимметрия, границы, цветовые вариации, диаметр и динамика изменений);

7) разметка дерматоскопических изображений клинически значимыми признаками, выполняемая независимо не менее чем тремя врачами-дерматологами и/или онкологами.

Для соблюдения принципов анонимизации и защиты персональных данных пациентов, установленных нормативно-правовыми актами Российской Федерации¹, на этапе получения информированного согласия данных каждому пациенту присваивается уникальный идентификатор (patient ID)². В качестве patient ID используется уникальная символьная строка, полученная в результате криптографического хэширования комбинации номера СНИЛС и даты рождения пациента, что позволяет исключить необходимость хранения или обработки конфиденциальной информации и персональных данных пациентов.

На этапе постановки первичного диагноза врачами дерматологами и/или

¹ О персональных данных: Федеральный закон от 27.07.2006 г. № 152-ФЗ: [ред. от 08.08.2024 г.]. URL: https://www.consultant.ru/document/cons_doc_LAW_372682/3d0cac60971a511280cbba229d9b6329c07731f7/ (дата обращения: 11.06.2025).

² Об основах охраны здоровья граждан в Российской Федерации: Федеральный закон от 21.11.2011 г. № 323-ФЗ: [ред. от 13.07.2022 г.]. URL: https://www.consultant.ru/document/cons_doc_LAW_121895/ (дата обращения: 11.06.2025).

онкологами осуществляется дерматоскопия пораженных областей кожи, позволяющая классифицировать изображения согласно классификаторам нозологии

кожных новообразований. Для постановки первичного диагноза применяются неинвазивные методов диагностики (табл. 2).

Таблица 2. Точность постановки первичного диагноза неинвазивными методами

Table 2. Primary diagnosis accuracy using non-invasive methods

Метод диагностирования Diagnostic method	Чувствительность, % Sensitivity, %	Специфичность, % Specificity, %
Правило ABCD (без E и F) ABCD Rule (without E and F)	47,3 (для меланом $d < 3$ мм)	47,3 (для меланом $d < 3$ мм)
Дерматоскопическое правило ABCD Dermatoscopy ABCD rule	84,1	83,5
Алгоритм CASH Algorithm CASH	98	68
7-точечный список Макки Mackey's 7-point list	91,7	53,5
7-точечное правило Арджентиано Argenziano's 7-point rule	97	71
Метод Мензиса Menzies method	92	71
Метод TADA TADA method	94	73,5

Стоит отметить, что полученные результаты экспериментальных оценок (чувствительность и специфичность) неинвазивных методов не позволяют обеспечить гарантированную постановку первичного диагноза о наличии или отсутствии новообразований и зачастую накладывают требование по проведению дополнительных клинических исследований.

На этапе создания снимка высокой четкости новообразования и сбора метаданных осуществляется фотографирование новообразования с итоговым разрешением изображения 4032×3024 пикселей. Примеры изображений из сформированного набора данных представлены ниже (рис. 1).



Рис. 1. Примеры дерматоскопических изображения из сформированного набора данных

Fig. 1. Dermatoscopic images examples from the generated data set

Каждое изображение новообразования содержит метаданную, обеспечивающую высокую степень клинической информативности. В состав метаданных входят следующие компоненты:

- данные о специалисте, проводившем обследование: идентификатор пользователя (username);

- дата и время проведения исследования (date);

- информация о пациенте: возраст (age), пол (sex), тип волос (hair), тип кожи (skin), фототип кожи по Фицпатрику (group), сведения о наследственности (genetic), характеристики иммунного статуса (immune);

- информация о кожном новообразовании: анатомическая локализация (place), размер образования (size), общее количество родинок (count), наличие красных точек (dots), цвет (color);

- предварительный клинический диагноз (diagnosis), установленный по результатам визуального и дерматоскопического осмотра;

- заключение по результатам гистологического исследования (result) при наличии такового.

Расширенная структура метаданных обеспечивает возможность как подробного клинического анализа, так и глубокой многофакторной валидации моделей машинного обучения, включая стратификацию по фототипу кожи, возрасту, полу, типу образования и другим признакам.

Разметка клинически значимых признаков на дерматоскопических изображениях с последующей разметкой дерматоскопических изображений

клинически значимыми признаками состоит из следующих шагов:

- определение клинически значимых признаков: асимметрия – различие формы, структуры и распределения пигментации в различных половинах образования; границы – степень четкости, ровности и непрерывности контура новообразования; цветовая вариабельность – наличие нескольких цветов, включая черный, коричневый, синий, белый и красный и дерматоскопические структуры – сетка, глобулы, псевдоподы, вуаль, сосудистые паттерны и другие характерные элемент;

- создание аннотационной системы и стандартизация. Для обеспечения воспроизводимости и унификации разметки создаются стандартизированные протоколы аннотирования, включающие: алгоритм ABCD, шкалу TDS (Total Dermoscopy Score) и алгоритм CHAOS & CLUES; правила аннотирования каждого признака, включая точные границы и допустимые вариации;

- экспертная разметка и валидация. Процесс разметки проводится в несколько шагов: первичная и консенсусная валидация, а также обратная связь и дообучение.

Сформированные дерматоскопические изображения составляют уникальный набор данных дерматоскопических изображений, включающий 657 изображений, отличающийся высококачественной ручной разметкой, расширенными метаданными (возраст, пол, фототип кожи, локализация, цвет, артефакты и др.) и репрезентативностью в отношении популяции Российской Федерации

(I–IV фототипы кожи по классификации Фицпатрика).

Результаты и их обсуждение

В процессе формирования набора данных дерматоскопических изображений посредством проспективного клинического наблюдения сформировано 657 дерматоскопических изображений от 208 пациентов дерматовенерологического профиля, сопровождаемых расширенными метаданными и предварительными клиническими заключениями, полученными в условиях реальной медицинской практики. Для каждого случая разметка была выполнена опытными дерматологами с учетом визуальных признаков, типичных для широкого спектра кожных новообразований, как

меланоцитарной, так и немеланоцитарной природы.

Диагнозы были установлены на основании клинической и дерматоскопической оценки, в ряде случаев с последующим гистологическим подтверждением. Стоит отметить, что гистология не являлась обязательным критерием включения в процессе медицинского осмотра, что, в свою очередь, позволяет отразить реалии амбулаторной практики и повысить прикладную ценность формируемого набора данных.

На рисунке 2 представлено распределение собранных дерматоскопических изображений по основным нозологическим категориям, представленным в наборе данных.

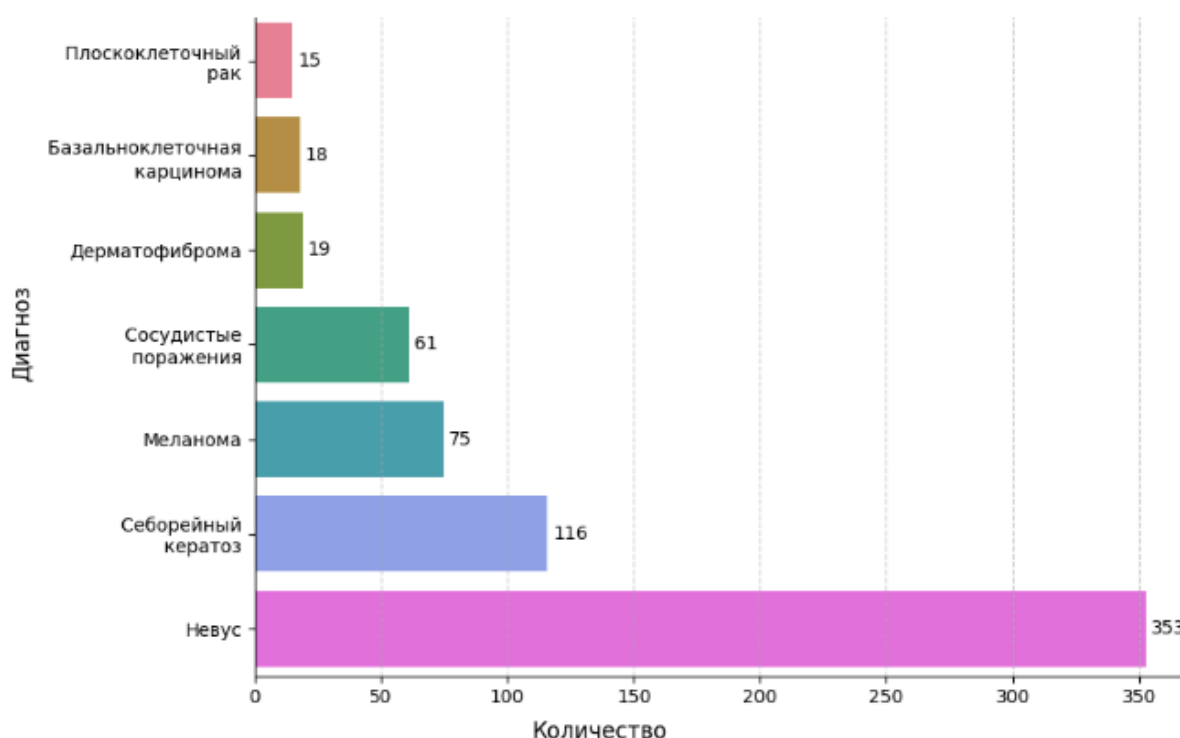


Рис. 2. Распределение изображений в наборе данных по классам нозологии кожных новообразований

Fig. 2. Images distribution in the data set by classes of skin neoplasm nosology

Исходя из графика значительное преобладание по количеству наблюдается в группе невусов (353 изображения), что составляет более половины всех случаев. Такое смещение в сторону невусов является ожидаемым и клинически оправданным, поскольку:

1) пигментные невусы представляют собой наиболее частую причину обращений к дерматологу, особенно при наличии подозрительных признаков (асимметрия, неровные границы, изменение цвета), требующих дифференциальной диагностики с меланомой;

2) пациенты с большим количеством невусов часто находятся под диспансерным наблюдением, в результате чего у одного пациента может быть получено несколько снимков разных образований;

3) в реальной практике невусы часто становятся объектами динамического наблюдения и тем самым обеспечивают большой объем визуальных данных пациента.

Следующими по частоте встречаемости являются:

1. Себорейный кератоз (116 изображений) – доброкачественное образование, которое часто ошибочно принимается за меланому как пациентами, так и врачами, особенно при атипичной локализации или выраженной пигментации.

2. Меланома (75 изображений), характеризуется сравнительно меньшим числом, при этом остается одним из наиболее значимым и распространенным злокачественным кожным новообразованием.

3. Сосудистые поражения (61 изображение), дерматофибромы (19 изображений), а также злокачественные опухоли немеланоцитарного происхождения: базальноклеточная карцинома (18 изображений) и плоскоклеточный рак (15 изображений), представлены в меньшем объеме, что отражает их редко встречаемость при диагностики и в рамках амбулаторного дерматоскопического скрининга, а также потенциальную сложность их визуальной верификации без гистологического подтверждения.

Представленное распределение отражает естественную клиническую структуру обращаемости, а также специфику отбора данных в условиях амбулаторной практики. Кроме того, полученные значения распределения изображений одновременно подчеркивают потребность в балансировке классов при обучении моделей машинного обучения и важность включения большого числа негативных (доброкачественных) случаев для повышения специфичности диагностических алгоритмов.

На рисунке 3 приведено распределение возрастов 208 пациентов на момент обследования. Основная масса наблюдений приходится на возрастной интервал от 25 до 65 лет с пиком около 35 лет. Это отражает активную диагностическую активность указанных возрастных групп, а также высокую онконастороженность и темпы распространения злокачественных кожных новообразований.

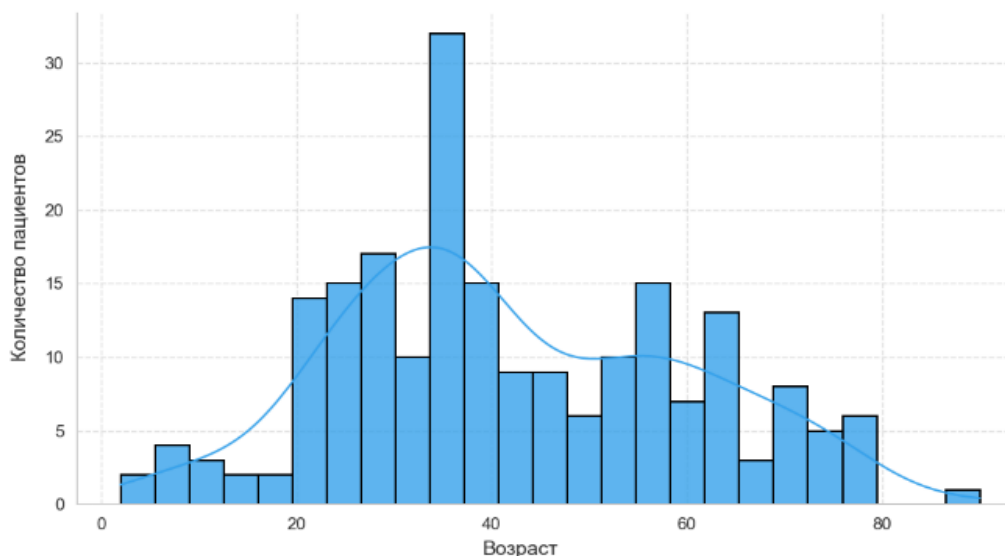


Рис. 3. Распределение пациентов дерматовенерологического профиля по возрасту

Fig. 3. Patients distribution with dermatovenereological profile by age

Ввиду того, что пациенты дерматовенерологического профиля могут иметь более одного заболевания или характеризуются наличием нескольких локализаций, на рисунке 4 представлено распределение числа дерматоскопических изображений, приходящихся на одного пациента в сформированном наборе данных. В большинстве случаев (порядка 25%) присутствует только одно

изображение на пациента, однако в наборе данных присутствуют пациенты с множественными образованиями и/или серийным наблюдением – вплоть до 9 снимков. Данная особенность позволяет формировать профиль выборки, приближенный к амбулаторной клинической практике, где одному пациенту может соответствовать несколько подозрительных очагов новообразований.

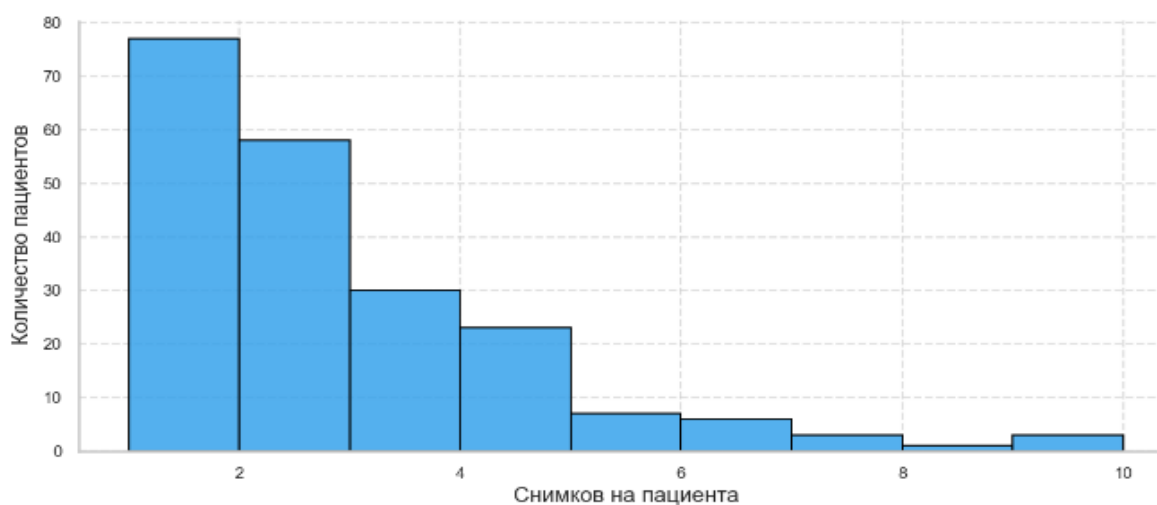


Рис. 4. Распределение количества снимков в расчете на одного пациента

Fig. 4. Distribution of the images numbers per patient

Разработанный набор дерматоскопических изображений состоит из 657 изображений, включая меланоцитарные (содержащие клетки меланоцитов): невус и меланома – 428 – и немеланоцитарные (не содержащие клетки меланоцитов): базальноклеточная карцинома, плоскоклеточный рак, дерматофиброма, сосудистые поражения и себорейный кератоз – 229 изображений.

Стоит отметить, что сформированный набор данных дерматоскопических изображений значительно уступает по количеству изображений таким наборам, как ISIC, HAM1000, IRMA, Dermofit Image Library и DERM12345. При этом разработанный набор данных по нозологическому разнообразию и этнопопуляционной репрезентативности превосходит такие общедоступные наборы, как PH2, IAD, IRMA и DermaQuest, в которых либо ограничено количество изображений, либо представлены лишь одна-две диагностические категории. В отличие от них представленный набор данных охватывает все основные типы кожных новообразований, включая как доброкачественные, так и злокачественные, а также сосудистые образования и дерматофибромы, что делает его более подходящим для обучения универсальных и клинически применимых диагностических моделей. Кроме того, стоит отметить уникальность сформированного набора данных, заключающуюся в том, что он сформирован на основе фототипов кожи пациентов, населяющих территорию Российской Федерации, что либо не представлено в международных наборах, таких как ISIC или HAM10000, либо ориентировано на

другие популяционные группы населения, например Турции, как в наборе данных DERM12345.

Сформированный набор данных дерматоскопических изображений предназначен для многопрофильного применения: он может быть использован как в клинической практике дерматологов и онкологов при проведении дерматоскопических исследований, так и в научных и прикладных целях, включая обучение, валидацию и тестирование алгоритмов машинного обучения, ориентированных на автоматизированную поддержку принятия решений врача.

Выводы

Благодаря качественной разметке, разнообразию нозологий, учету фототипов кожи, характерных для населения России, и наличию расширенных метаданных, набор данных изображений представляет собой значимый ресурс как для улучшения клинической диагностики, так и для разработки интеллектуальных медицинских систем, способных повысить точность и эффективность скрининга кожных новообразований в реальных условиях здравоохранения.

Основным направлением дальнейших исследований является дальнейшее расширение и обновление сформированного набора дерматоскопических изображений, включая:

- дополнение набора данных новыми изображениями диагностированных кожных новообразований;
- добавление новых классов злокачественных кожных новообразований, ранее недостаточно представленных в наборе данных;

– увеличение доли изображений, соответствующих различным фототипам кожи населения России, включая фототипы IV и выше.

Решение указанных задач позволит использовать сформированный набор дерматоскопических изображений в

процессе обучения и применения методов и алгоритмов машинного обучения и систем оптического распознавания, внося вклад в развитие систем телемедицинской диагностики и совершенствования методов ранней диагностики заболеваний.

Список литературы

1. Меланома кожи (C43): анализ тенденций заболеваемости и смертности в свете пандемии COVID-19, молекулярная эпидемиология / Ю. О. Бахарева, В.О. Тараканова, М. Ю. Рубаняк, Е. М. Каменских // Вопросы онкологии. 2023. Т. 69, № 4. С. 631–638. <https://doi.org/10.37469/0507-3758-2023-69-4-631-638>

2. Директор Государственного научного центра дерматовенерологии и косметологии Минздрава России, академик РАН Алексей Кубанов о вредном воздействии солнечных лучей и защите от них // Министерство здравоохранения Российской Федерации URL: <https://minzdrav.gov.ru/news/2022/07/04/18987-direktor-gosudarstvennogo-nauchnogo-tsentra-dermatovenerologii-i-kosmetologii-minzdrava-rossii-akademik-ran-aleksey-kubanov-o-vrednom-vozdeystvii-solne-chnyh-luchey-i-zaschite-ot-nih> (дата обращения: 11.06.2025).

3. Современное состояние заболеваемости и диагностики меланомы и других злокачественных новообразований кожи в Российской Федерации и Уральском федеральном округе / Н. П. Малишевская, А. В. Соколова, В. В. Петкау, Л. В. Демидов, В. А. Игликов // Эффективная фармакотерапия. 2024. № 20(5). С. 18–23. <https://doi.org/10.33978/2307-3586-2024-20-5-18-23>

4. Федеральный проект «Борьба с онкологическими заболеваниями» // Министерство здравоохранения Российской Федерации. URL: <https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravoohranenie/onko> (дата обращения: 11.06.2025).

5. Бакулев А. Л., Конопацкова О. М., Станчина Ю. В. Дерматоскопия в диагностике пигментных невусов кожи // Вестник дерматологии и венерологии. 2019. № 95(4). С. 48–56. <https://doi.org/10.25208/0042-4609-2019-95-4-48-56>

6. Хисматуллина З. Р., Чеботарев В. В., Бабенко Е. А. Современные аспекты и перспективы применения дерматоскопии в дерматоонкологии // Креативная хирургия и онкология. 2020. Т. 10, № 3. С. 241–248. <https://doi.org/10.24060/2076-3093-2020-10-3-241-248>

7. Ранняя диагностика злокачественных новообразований кожи с помощью технологий искусственного интеллекта / С. О. Самохин, А. В. Патрушев, Ю. И. Акаева, С. А. Парфенов, Г. Г. Кутелев // Вестник дерматологии и венерологии. 2024. Т. 100, № 1. С. 38–46. <https://doi.org/10.25208/vdv16746>

8. Применение моделей машинного обучения для многоклассовой классификации дерматоскопических снимков новообразований кожи / А. В. Козачок, А. А. Спирин, О. И. Самоваров, Е. С. Козачок // Труды Института системного программирования РАН. 2024. Т. 36, № 5. С. 241–252. [https://doi.org/10.15514/ISPRAS-2024-36\(5\)-17](https://doi.org/10.15514/ISPRAS-2024-36(5)-17)
9. Хабарова Р. И., Кулева С. А. Искусственный интеллект в диагностике доброкачественных новообразований кожи у пациентов детского возраста. Интеграция нейронной сети в мобильное приложение // Вопросы онкологии. 2022. Т. 68, № 6. С. 820–826. <https://doi.org/10.37469/0507-3758-2022-68-6-820-826>
10. Козачок А. В., Спирин А. А., Козачок Е. С. Обзор методов раннего обнаружения меланомы // Труды институт системного программирования РАН. 2022. Т. 34, № 4. С. 241–250. [https://doi.org/10.15514/ISPRAS-2022-34\(4\)-17](https://doi.org/10.15514/ISPRAS-2022-34(4)-17)
11. The International Skin Imaging Collaboration // ISIC Archive. URL: <https://www.isic-archive.com> (дата обращения: 11.06.2025).
12. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC) / D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern // Computer Vision and Pattern Recognition. 2016. URL: <https://arxiv.org/abs/1605.01397> (дата обращения: 11.06.2025). <https://doi.org/10.48550/arXiv.1605.01397>
13. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2017, hosted by the international skin imaging collaboration (ISIC) / N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler // 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). P. 168–172. URL: <https://arxiv.org/abs/1710.05006> (дата обращения: 11.06.2025). <https://doi.org/10.1109/ISBI.2018.8363547>
14. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC) / N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti // Computer Vision and Pattern Recognition. 2019. URL: <https://arxiv.org/abs/1902.03368> (дата обращения: 11.06.2025). <https://doi.org/10.48550/arXiv.1902.03368>
15. Tschandl P., Rosendahl C., Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions // Scientific Data. 2018. Vol. 5. P. 1–9. <https://doi.org/10.1038/sdata.2018.161>
16. Mendonça T., Ferreira P. M., Marques J. S. PH² – A Dermoscopic Image Database for Research and Benchmarking // Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Osaka, Japan: IEEE, 2013. P. 5437–5440. <https://doi.org/10.1109/EMBC.2013.6610779>

17. A survey, review, and future trends of skin lesion segmentation and classification / H. Kamrul, A. Asif, H. Y. Choon, Y. Guang // *Computers in Biology and Medicine*. 2023. Vol. 155. P. 1–36. <https://doi.org/10.1016/j.compbimed.2023.106624>
18. Dermofit Image Library // The University of Edinburgh. URL: <https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library> (дата обращения: 11.06.2025).
19. DERM12345: A Large, Multisource Dermatoscopic Skin Lesion Dataset with 40 Subclasses / A. Yilmaz, S. P. Yasar, G. Gencoglan, B. Temelkuran // *Scientific Data*. 2024. N 11. <https://doi.org/10.1038/s41597-024-04104-3>

References

1. Bakhareva Yu.O., Tarakanova V.O., Rubanyak M.Yu., Kamenskich E.M. Skin Melanoma (C43): Analysing incidence and mortality trends in the context of the COVID-19 pandemic, molecular epidemiology. *Voprosy onkologii = Issues of Oncology*. 2023;69(4):631–638. (In Russ.) <https://doi.org/10.37469/0507-3758-2023-69-4-631-638>
2. Alexey Kubanov, Director of the State Scientific Center for Dermatovenereology and Cosmetology of the Ministry of Health of the Russian Academy of Sciences, on the harmful effects of sunlight and protection from them. Ministry of Health of the Russian Federation. (In Russ.) Available at: <https://minzdrav.gov.ru/news/2022/07/04/18987-direktor-gosudarstvennogo-nauchnogo-tsentra-dermatovenerologii-i-kosmetologii-minzdrava-rossii-akademik-ran-aleksey-kubanov-o-vrednom-vozdeystvii-solne-chnyh-luchey-i-zaschite-ot-nih> (accessed 11.06.2025).
3. Malishevskaya N.P., Sokolova A.V., Petkau V.V., Demidov L.V., Iglikov V.A. The current state of the incidence and diagnosis of melanoma and other malignant neoplasms of the skin in the Russian Federation and the Ural Federal District. *Effective Pharmacotherapy*. 2024; (20):18–23. (In Russ.) <https://doi.org/10.33978/2307-3586-2024-20-5-18-23>
4. Federal project "Fight against oncological diseases". Ministry of Health of the Russian Federation. (In Russ.) Available at: <https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravooхранenie/onko> (accessed 11.06.2025).
5. Bakulev A.L., Konopatskova O.M., Stanchina Yu.V. Dermatoscopy in the diagnosis of pigmented skin nevi. *Vestnik dermatologii i venerologii = Bulletin of Dermatology and Venereology*. 2019;(95):48–56. (In Russ.) <https://doi.org/10.25208/0042-4609-2019-95-4-48-56>
6. Khismatullina Z.R., Chebotarev V.V., Babenko E.A. Modern aspects and prospects of using dermatoscopy in dermatooncology. *Kreativnaya khirurgiya i onkologiya = Creative Surgery and Oncology*. 2020;10(3):241–248. (In Russ.) <https://doi.org/10.24060/2076-3093-2020-10-3-241-248>
7. Samokhin S.O., Patrushev A.V., Akayeva Yu.I., Parfenov S.A., Kutelev G.G. Early diagnosis of skin malignancies using artificial intelligence technologies. *Vestnik dermatologii*

i venerologii = Bulletin of Dermatology and Venereology. 2024;100(1):38–46. (In Russ.) <https://doi.org/10.25208/vdv16746>

8. Kozachok A.V., Spirin A.A., Samovarov O.I., Kozachok E.S. Application of machine learning models for multiclass classification of dermatoscopic images of skin neoplasms. *Trudy instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute of System Programming of the Russian Academy of Sciences*. 2024;36(5):241–252. (In Russ.) [https://doi.org/10.15514/ISPRAS-2024-36\(5\)-17](https://doi.org/10.15514/ISPRAS-2024-36(5)-17)

9. Khabarova R.I., Kuleva S.A. Artificial intelligence in the diagnosis of benign skin neoplasms in pediatric patients. Integration of a neural network into a mobile application. *Voprosy onkologii = Issues of Oncology*. 2022;68(6):820–826. (In Russ.) <https://doi.org/10.37469/0507-3758-2022-68-6-820-826>

10. Kozachok A.V., Spirin A.A., Kozachok E.S. Review of methods for early detection of melanoma. *Trudy institut sistemnogo programmirovaniya RAN = Proceedings of the Institute of System Programming of the Russian Academy of Sciences*. 2022;34(4):241–250. (In Russ.) [https://doi.org/10.15514/ISPRAS-2022-34\(4\)-17](https://doi.org/10.15514/ISPRAS-2022-34(4)-17)

11. The International Skin Imaging Collaboration. ISIC Archive. Available at: <https://www.isic-archive.com> (accessed 11.06.2025)

12. Gutman D., Codella N.C., Celebi E., Helba B., Marchetti M., Mishra N., Halpern A. Skin lesion analysis towards melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *Computer Vision and Pattern Recognition*. 2016. Available at: <https://arxiv.org/abs/1605.01397> (accessed 11.06.2025). <https://doi.org/10.48550/arXiv.1605.01397>

13. Codella N.C., Gutman D., Celebi M.E., Helba B., Marchetti M.A., Dusza S.W., Kalloo A., Liopyris K., Mishra N., Kittler H. Skin lesion analysis towards melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2017, hosted by the international skin imaging collaboration (ISIC). In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. P. 168–172. Available at: <https://arxiv.org/abs/1710.05006> (accessed 11.06.2025). <https://doi.org/10.1109/ISBI.2018.8363547>

14. Codella N., Rotemberg V., Tschandl P., Celebi M.E., Dusza S., Gutman D., Helba B., Kalloo A., Liopyris K., Marchetti M. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *Computer Vision and Pattern Recognition*. 2019. Available at: <https://arxiv.org/abs/1902.03368> (accessed 11.06.2025). <https://doi.org/10.48550/arXiv.1902.03368>

15. Tschandl P., Rosendahl C., Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. 2018;5:1–9. <https://doi.org/10.1038/sdata.2018.161>

16. Mendonça T., Ferreira P.M., Marques J.S. PH² – A Dermoscopic Image Data-base for Research and Benchmarking. In: *Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society (EMBC). Osaka, Japan: IEEE; 2013. P. 5437–5440. <https://doi.org/10.1109/EMBC.2013.6610779>

17. Kamrul H., Asif A., Choon H.Y., Guang Y. A survey, review, and future trends of skin lesion segmentation and classification. *Computers in Biology and Medicine*. 2023;155:1–36. <https://doi.org/10.1016/j.compbiomed.2023.106624>

18. Dermofit Image Library. The University of Edinburgh. Available at: [https:// licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library](https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library) (accessed 11.06.2025).

19. Yilmaz A., Yasar S.P., Gencoglan G., Temelkuran B. DERM12345: A Large, Multi-source Dermatoscopic Skin Lesion Dataset with 40 Subclasses. *Scientific Data*. 2024;(11). <https://doi.org/10.1038/s41597-024-04104-3>

Информация об авторе / Information about the Author

Козачок Елена Сергеевна, специалист,
Институт системного программирования
им. В. П. Иванникова РАН,
г. Москва, Российская Федерация,
e-mail: e.kozachok@ispras.ru,
Researcher ID: rid108085,
ORCID: 0009-0007-9432-1663

Elena S. Kozachok, Specialist,
Ivannikov Institute for System Programming
of the Russian Academy of Sciences,
Moscow, Russian Federation,
e-mail: e.kozachok@ispras.ru,
Researcher ID: rid108085,
ORCID: 0009-0007-9432-1663