

УДК 616.5-006:004.62:004.932

10.25209/2079-3316-2026-17-3-57-104



Методика формирования клинически верифицированного набора дерматоскопических изображений

Елена Сергеевна Козачок[✉]

Институт системного программирования им. В.П. Иванникова РАН, Москва, Россия

[✉]e.kozachok@ispras.ru

Аннотация. Качество автоматизированных систем диагностической поддержки зависит не только от объема изображений, но и от воспроизводимости процедуры съемки, полноты структурированных метаданных и надежности диагностических меток. К сожалению, международные коллекции формировались в условиях, отличающихся от российской амбулаторной практики и мобильной дерматоскопии. Поэтому обучающей базы недостаточно и требуется независимое (внешнее) оценочное множество для проверки моделей, обученных на открытых коллекциях.

Разработаны трехкомпонентная методика формирования клинически верифицированного набора данных дерматоскопических изображений для задач медицинской информатики и ее программная реализация. Методика объединяет три взаимосвязанных компонента:

- 1) стандартную операционную процедуру (СОП) получения изображения средствами мобильной дерматоскопии,
- 2) информационную модель метаданных из 16 структурированных полей по шести клинически ориентированным блокам в нотации, совместимой с ISIC,
- 3) многоэтапную экспертную верификацию диагностической метки (первичная клиническая разметка, консенсусный пересмотр тремя специалистами, гистологическое подтверждение злокачественных новообразований).

Программная реализация дана связкой «мобильное клиентское приложение – реляционное хранилище – среда коллегиальной разметки Colba – модуль обезличенной выгрузки».

По методике в период с июня 2025 г. по май 2026 г. сформирован набор из 1026 уникальных дерматоскопических изображений от 443 пациентов: на этапе съемки по критериям СОП отбраковано 32 технически непригодных снимка, не попавших в обработку; из поступивших далее 1044 записей дедупликация исключила 18 повторных. Набор включает 9 нозологических категорий, все 39 изображений злокачественных классов (меланома, базалиома, плоскоклеточный рак) гистологически верифицированы; возраст пациентов 2–90 лет (медиана 38), в выборке 279 женщин и 164 мужчины.

Для работы в условиях выраженного дисбаланса определены четыре согласованных уровня агрегации диагностических категорий (от бинарного до девятиклассового) и выполнен анализ совстречаемости нозологий в пределах пациентской записи: у 107 из 443 пациентов зарегистрировано две и более различные категории. Каждое изображение сопровождается экспертно описанными дерматоскопическими структурами и явным индикатором уровня подтверждения диагноза *verification_stage*.

Представленная формализация воспроизводимого подхода к подготовке клинических дерматоскопических наборов и описание поддерживающих его программных средств применимы при создании ресурсов для внешней проверки алгоритмов поддержки принятия врачебных решений в дерматоонкологии.

Ключевые слова и фразы: дерматоскопия, набор данных, стандартная операционная процедура, информационная модель метаданных, экспертная верификация, гистологическая верификация, мобильная дерматоскопия, медицинская информатика, новообразования кожи

Для цитирования: Козачок Е.С. Методика формирования клинически верифицированного набора дерматоскопических изображений // Программные системы: теория и приложения. 2026. Т. 17. № 3(72). С. 57–104. https://psta.psiras.ru/read/psta2026_3_57-104.pdf

Введение

Дерматоскопия является одним из основных инструментов визуальной диагностики пигментных и непигментных новообразований кожи. Точность интерпретации дерматоскопической картины зависит от квалификации специалиста, условий получения изображения, полноты клинического контекста пациента и доступности морфологической верификации диагноза.

Развитие методов глубокого обучения переводит требования к диагностическим коллекциям дерматоскопических изображений из плоскости визуального материала в плоскость структурированных информационных ресурсов: для воспроизводимого обучения и независимой оценки модели необходимы не только сами изображения, но и согласованная информационная модель метаданных, документированный протокол съемки и прозрачная цепочка верификации диагностических меток.

Международные дерматоскопические коллекции – HAM10000, ISIC, BCN20000, PAD-UFES-20 – сыграли определяющую роль в становлении задач автоматической классификации новообразований кожи [1–5]. Накопленный опыт также показал, что переносимость моделей, обученных на таких коллекциях, в локальную клиническую среду ограничивается доменными сдвигами: различиями в распределении фототипов кожи, параметрах съемки, нозологической структуре обращаемости и полноте сопровождающих метаданных [6–10].

Для условий российской амбулаторной практики, в которой дерматоскопия выполняется преимущественно мобильными средствами (оптический дерматоскоп в сочетании со смартфоном), эти ограничения проявляются в полной мере, что обосновывает разработку локальных верифицированных наборов данных.

Ключевая методическая проблема состоит в том, что дерматоскопический набор данных не является простой совокупностью графических файлов. Для пригодности к задачам медицинской информатики необходимо одновременно решить три задачи: обеспечить воспроизводимость условий получения каждого изображения, согласовать структурированный состав метаданных и зафиксировать уровень надежности диагностической метки. Каждый из этих компонентов отдельно описан в публикациях по созданию открытых наборов [3, 5, 11], однако в литературе, ориентированной на медицинскую информатику, реже представлена единая методика, в которой все три компонента интегрированы и согласованы между собой.

Цель работы – разработать методику формирования клинически верифицированного набора дерматоскопических изображений новообразований кожи, объединяющую стандартную операционную процедуру получения изображения, информационную модель метаданных и многоэтапную экспертную верификацию диагностических меток, реализовать ее программно и сформировать по этой методике набор данных, пригодный для задач медицинской информатики.

Целевое назначение формируемого набора данных требует отдельной оговорки, поскольку от него зависит трактовка всех дальнейших проектных решений. Набор проектировался не как обучающая база и не как конкурент открытым коллекциям по объему, а как *независимое (внешнее) оценочное множество* – evaluation-set – для проверки моделей, обученных на открытых источниках, в условиях, отличных от условий их обучения. Из этого назначения следуют три требования, которые и определяют содержание методики:

- (1) *полная независимость от открытых коллекций*: ни одно изображение набора не должно происходить из ISIC, HAM10000 или иного архива, иначе измерение доменного сдвига становится невозможным;
- (2) *однородность протокола получения изображения*, поскольку внешняя оценка имеет смысл только тогда, когда условия съемки целевого домена контролируются и документированы;
- (3) *явная фиксация уровня достоверности каждой метки*, поскольку оценочное множество, в отличие от обучающего, не допускает усреднения ошибок разметки.

Объем набора при таком назначении является не самоцелью, а следствием: оценочное множество должно быть достаточным для получения интервальных оценок метрик, но не обязано быть сопоставимым по размеру с обучающими коллекциями. Ограничения, вытекающие из объема набора, и допустимые уровни агрегации диагностических категорий рассматриваются в разделах 4.2 и 6.

Задачи исследования:

- *сформулировать требования* к информационной модели клинического дерматоскопического набора данных;
- *разработать СОП* получения изображения средствами мобильной дерматоскопии;
- *определить структуру метаданных*, согласованную с практикой международных дерматоскопических коллекций;

- *формализовать многоэтапную процедуру экспертной верификации;*
- *описать программную реализацию методики* – архитектуру комплекса, логическую модель хранения данных и интерфейсы экспертной работы;
- *сформировать по методике набор данных и привести его статистическое описание*, включая уровни агрегации категорий и совстречаемость нозологий в пределах пациентской записи;
- определить ограничения и направления развития.

Научная новизна работы заключается в следующем:

- *Сформулирована трехкомпонентная методика* формирования клинического дерматоскопического набора данных, в которой СОП съемки, информационная модель метаданных и многоэтапная диагностическая верификация рассматриваются как взаимосвязанные элементы единого воспроизводимого информационного ресурса; в литературе по открытым дерматоскопическим коллекциям такая интегрированная формулировка ранее не представлена.
- *Разработана информационная модель метаданных*, организованная по шести клинически ориентированным блокам и согласованная с практикой международных дерматоскопических коллекций по нотации полей, при этом включающая поле экспертной аннотации дерматоскопических структур, создающее основу для количественного сопоставления областей внимания моделей с клинически значимыми зонами изображения.
- *Формализована многоэтапная процедура экспертной верификации* диагностической метки с явной фиксацией уровня надежности записи (первичная разметка – консенсус трех специалистов – гистологическое подтверждение ЗНО), позволяющая использовать набор как пилотный ресурс для независимой оценки, анализа доменного сдвига и интерпретируемости с учетом гетерогенности достоверности меток.
- *Описана программная реализация методики*: архитектура комплекса подготовки данных, логическая модель хранения записей «пациент – исследование – изображение – разметка – экспертное решение», интерфейс структурированного клинического ввода и среда коллегиальной экспертной разметки с журналированием изменений диагностической метки. Тем самым воспроизводимость методики обеспечивается не только регламентом, но и программными ограничениями целостности.
- *Сформирован клинически верифицированный набор* из 1026 изображений от 443 пациентов в условиях российской амбулаторной практики, охватывающий 9 нозологических категорий и обеспеченный гистологической верификацией злокачественных классов; для набора определены четыре согласованных уровня агрегации категорий и выполнен анализ совстречаемости нозологий в пределах пациентской записи.

Практическая значимость состоит в воспроизводимости подхода к подготовке клинического дерматоскопического набора данных: предложенная методика может использоваться при формировании новых медицинских коллекций изображений, независимом тестировании алгоритмов классификации, разработке систем поддержки принятия врачебных решений в дерматоонкологии и сравнительной оценке моделей в условиях, приближенных к российской амбулаторной практике.

1. Обзор существующих наборов дерматоскопических изображений

Сравнение существующих дерматоскопических наборов данных целесообразно проводить не по объему, а по составу информационной модели и характеристикам диагностической верификации, поскольку именно эти параметры определяют пригодность набора для задач воспроизводимого обучения и независимой оценки моделей.

Классическим референсным набором является PH2 [12] – 200 дерматоскопических изображений трех клинических категорий (обычный невус, атипичный невус, меланома) одного центра с гистологическим подтверждением; небольшой объем компенсируется однородностью условий съемки. HAM10000 [1] содержит 10 015 изображений семи диагностических категорий из двух клинических центров и остается одним из наиболее цитируемых бенчмарков.

Соревнование ISIC 2017 [2] стало одной из первых открытых платформ воспроизводимой оценки моделей, а коллекция ISIC 2019, объединившая HAM10000, BCN и часть MSK, расширила объем обучающей выборки до 25 331 изображения восьми классов [13]. ISIC 2020 [3] (33 126 изображений от 2 056 пациентов, полная гистологическая верификация) впервые формально документировал процедуру съемки и расширил клинические метаданные.

BCN20000 [4] описывает 18 946 изображений одного крупного дерматологического центра и фиксирует широкий клинический контекст пациента.

PAD-UFES-20 [5] собран бытовыми средствами съемки и сопровождается анкетой клинических признаков, однако формально относится не к дерматоскопическим, а к клиническим изображениям кожных поражений.

Систематические обзоры публично доступных наборов дерматоскопических изображений [10] и отдельная обзорная работа о недостаточной прозрачности наборов данных и алгоритмов в дерматологическом ИИ [14] фиксируют ряд устойчивых ограничений: непрозрачность процедуры подтверждения диагноза, ограниченную информацию о фототипах кожи, отсутствие документированной СОП получения изображений, неоднородность полей метаданных между коллекциями.

Концепция «паспортов набора данных» (datasheets for datasets) [15] формализует требования к описанию источников, процедуры сбора, состава метаданных и ограничений выборки. В настоящее время она рассматривается как один из стандартных инструментов документирования наборов в медицинском машинном обучении.

В работах, посвященных смещениям и репрезентативности дерматологических алгоритмов [6–8], показано, что недостаточная представленность отдельных фототипов и популяционных групп приводит к снижению устойчивости моделей при применении в среде, отличной от среды формирования обучающей выборки. Применение моделей глубокого обучения к задачам дифференциальной диагностики меланомы и рака кожи сопоставимо по результатам с экспертной интерпретацией только при условии достаточной согласованности обучающего и целевого распределений [11, 16]. Иначе возникают систематические смещения, связанные с составом выборки, источником данных и условиями съемки [16].

Сравнительная характеристика международных наборов и разработанного набора приведена в таблице 1. Таблица не предназначена для ранжирования наборов: ее задача – показать, какие свойства должны быть явно зафиксированы при подготовке клинического дерматоскопического набора данных для задач медицинской информатики.

Из сравнения следует, что в задачи методики формирования клинического дерматоскопического набора данных входят, как минимум, документированное описание условий получения изображения, структурированная информационная модель метаданных и прозрачная процедура подтверждения диагноза. Эти три элемента положены в основу предлагаемой методики.

Хотя разработанный набор существенно меньше по объему, чем ISIC 2019/2020 или BCN20000, однако превосходит сопоставимые по размеру наборы (PH2, PAD-UFES-20) по полноте структурированных метаданных и по уровню формализации цепочки верификации диагностической метки.

Таблица 1. Сравнительная характеристика дерматоскопических наборов данных

Набор	Год	Изображений	Классов	Способ съемки	Метаданные	Подтверждение диагноза	СОП
PH2 [12]	2013	200	3	Дерматоскопия 1 центра	Мин. набор клин. признаков	Гистология	Не унифицирована
HAM10000 [1]	2018	10 015	7	Дерматоскопия, 2 центра	Возраст, пол, локализация, диагноз	Гистология, клиника или консенсус	Не описана
ISIC 2017 [2]	2018	2 750	3	Дерматоскопия, многоисточник	Минимальный	Разнородные режимы	Не унифицирована
ISIC 2019 [13]	2019	25 331	8	HAM10000 + BCN + MSK	Зависит от подколлекции	Разнородные режимы	Не унифицирована
ISIC 2020 [3]	2020	33 126 (2 056 пациентов)	9 + бин.	Дерматоскопия, многоцентровая	Расширенные клин. + контекст	Гистология	Документирована
BCN20000 [4]	2024	18 946	8	Дерматоскопия 1 центра	Клинические и служ. сведения	Эксперт + морфология части наблюдений	Описана частично
PAD-UFES-20 [5]	2020	2 298 (1 373 пациента)	6	Смартфон, клиническая	Анкета пациента и поражения	Часть гистологий + клиника	Описана для смартфона
Разработанный	2026	1 026 (443 пациента)	9	Мобильная дерматоскопия	16 полей, ISIC-совместима	Многоэтапная: разметка → консенсус 3 → гистол. ЗНО	Документирована

2. Материалы и методы

2.1. Архитектура методики

Разработанная методика рассматривает формирование клинического дерматоскопического набора как воспроизводимый информационный процесс, в котором каждое изображение приобретает интерпретируемое описание и метку с указанием уровня надежности. Методика структурно состоит из трех взаимосвязанных компонентов: стандартной операционной процедуры получения изображения, информационной модели метаданных и многоэтапной экспертной верификации диагностической метки. Сквозными процедурами являются обезличивание пациентской записи, дедупликация и нормализация диагностических категорий. Общая архитектура методики представлена на рисунке 1.



Рисунок 1. Архитектура методики формирования клинически верифицированного набора дерматоскопических изображений

Объектом обработки методики является пара «изображение – пациент». Входом служит дерматоскопический снимок и связанная с ним клиническая запись; выходом – обезличенная запись набора, включающая изображение, диагностическую метку, структурированные метаданные и атрибут уровня верификации. Контроль качества встроен в каждый компонент: на этапе получения – через оценку технической пригодности изображения, на этапе описания – через формальную схему метаданных, на этапе верификации – через явный учет уровня подтверждения диагноза.

Формализованное описание полного pipeline методики в виде псевдокода представлено в алгоритме 1.

АЛГОРИТМ 1. Формирование клинически верифицированного набора дерматоскопических изображений

Вход: Множество пар $\{(i_k, r_k)\}_{k=1}^N$: дерматоскопическое изображение i_k , клиническая запись r_k

Выход: Набор $\mathcal{D}$ обезличенных верифицированных записей

- 1: $\mathcal{D} \leftarrow \emptyset$; $\mathcal{H} \leftarrow \emptyset$ ▷ множество встреченных хэшей
 - 2: **для** каждой пары (i_k, r_k) , $k = 1, \dots, N$ **выполнять**
 - 3: $q \leftarrow \text{ОЦЕНКАКАЧЕСТВА}(i_k)$ ▷ СОП: шкала 1–5, раздел 2.4
 - 4: **если** $q \leq 2$, **то** перейти к следующей паре ▷ отклонение по критерию СОП
 - 5: $h \leftarrow \text{КРИПТОХЭШ}(i_k)$ ▷ ГОСТ Р 34.11-2012, 256 бит; раздел 2.7
 - 6: **если** $h \in \mathcal{H}$, **то** перейти к следующей паре ▷ побайтовый дубликат, раздел 2.7
 - 7: $\mathcal{H} \leftarrow \mathcal{H} \cup \{h\}$
 - 8: $m \leftarrow \text{ЗАПОЛНИТЬМЕТАДАННЫЕ}(r_k)$ ▷ 16 полей по 6 блокам: раздел 2.5, таблица 2
 - 9: $\ell \leftarrow \text{ПЕРВИЧНАЯРАЗМЕТКА}(i_k, r_k)$ ▷ событие журнала «этап 1»: раздел 2.6, таблица 3
 - 10: $\ell \leftarrow \text{КОНСЕНСУСНАЯРАЗМЕТКА}(\ell, \text{3 специалиста})$ ▷ событие журнала «этап 2»: раздел 2.6; Colba, раздел 3.4
 - 11: **если** $\ell \in \{\text{меланома, базалиома, плоскоклеточный рак}\}$ **то**
 - 12: $\ell \leftarrow \text{ГИСТОЛОГИЧЕСКАЯВЕРИФИКАЦИЯ}(r_k)$ ▷ событие журнала «этап 3»: раздел 2.6
 - 13: **конец если**
 - 14: $m.\text{verification_stage} \leftarrow \text{УРОВЕНЬВЕРИФИКАЦИИ}(\text{ЖУРНАЛ}(i_k))$ ▷ раздел 3.4
 - 15: $m.\text{hash} \leftarrow h$ ▷ хэш содержимого файла, раздел 2.7
 - 16: $m.\text{clinical_diagnosis} \leftarrow \text{НОРМАЛИЗАЦИЯКАТЕГОРИИ}(\ell)$ ▷ 9 классов, раздел 2.7
 - 17: $\mathcal{D} \leftarrow \mathcal{D} \cup \{(i_k, m, \ell)\}$
 - 18: **конец для**
 - 19: **вернуть** $\mathcal{D}$
-

Алгоритм отражает детерминированную последовательность операций, применяемую к каждой входящей паре «изображение – клиническая запись», и однозначно определяет условия включения, дедупликации, нормализации и назначения уровня верификации.

Строки алгоритма соотносятся с разделами статьи следующим образом:

- оценка технической пригодности снимка и отбраковка по ее результату (строки 3–4) выполняются по правилам СОП (раздел 2.4);

- заполнение метаданных (строка 8) – по информационной модели (раздел 2.5, таблица 2);
- первичная разметка, консенсусный пересмотр и гистологическая верификация (строки 9–13) – по процедуре многоэтапной верификации (раздел 2.6, таблица 3) в среде разметки, описанной в разделе 3.4;
- вычисление уровня верификации по журналу изменений метки (строка 14) – по разделу 3.4;
- вычисление хэша, проверка наличия дубликата и сохранение хэша в записи (строки 5–7 и 15), а также нормализация категорий (строка 16) – по сквозным процедурам раздела 2.7.

Этапы разметки (строки 9–13) показаны как события, фиксируемые в журнале изменений метки; значение поля `verification_stage` не присваивается в ходе этих этапов вручную, а выводится из журнала на строке 14, что соответствует реализации, описанной в разделе 3.4.

Порядок операций существен: оценка качества предшествует хэшированию, чтобы отбракованные снимки не занимали позиции в множестве $\mathcal{H}$; консенсусный пересмотр предшествует гистологической верификации, поскольку именно консенсусная метка определяет, направляется ли запись на морфологическое подтверждение.

Выбор криптографической хэш-функции (строка 5) требует пояснения, поскольку он определяется не алгоритмическими соображениями, а внутренними требованиями к закрытому исследовательскому контуру организации. Дедупликация выполняется по хэшу побайтового содержимого файла, поэтому от функции требуется единственное свойство – практическая невозможность коллизий; требований к производительности не предъявляется, так как обрабатываемый массив составляет 1044 записи. Использована функция ГОСТ Р 34.11-2012 («Стрибог», длина выхода 256 бит), что обусловлено принятым в организации требованием использовать отечественные криптографические алгоритмы при обработке медицинских данных в закрытом исследовательском контуре. Результат процедуры от конкретной криптостойкой функции не зависит: замена на SHA-256 или BLAKE2b дала бы то же множество из 18 выявленных дубликатов, поскольку совпадение устанавливается по точному равенству содержимого файлов. Длина выхода 256 бит выбрана как минимальная из предусмотренных стандартом; при объеме массива порядка 10^3 записей вероятность случайной коллизии пренебрежимо мала (оценка по парадоксу дней рождения – порядка 10^{-71}). Ограничение подхода состоит в том, что побайтовое хэширование не выявляет визуально идентичные изображения, различающиеся сжатием или кадрированием; соответствующее направление развития рассмотрено в разделе 5.

2.2. Период и источники данных

Набор сформирован проспективно в период с июня 2025 г. по май 2026 г. Учетной единицей изображения служит уникальный файл дерматоскопического снимка; учетной единицей пациента – обезличенная пациентская запись. Включались дерматоскопические изображения кожных новообразований, полученные в соответствии с разработанной СОП и сопровождаемые клиническим заключением и минимальным набором структурированных метаданных. Наличие нескольких изображений у одного пациента допускалось, поскольку в амбулаторной практике один пациент может иметь несколько новообразований, требующих отдельной оценки. Из обработки исключались записи без файла изображения, дубли по хэшу содержимого файла, а также записи с непригодными для статистической обработки значениями ключевых полей.

2.3. Этические аспекты работы с данными

Исследование проводилось в соответствии с принципами Хельсинкской декларации. Сбор дерматоскопических изображений и клинических сведений осуществлялся в рамках плановой амбулаторной дерматоскопической практики; от каждого пациента получено письменное информированное согласие на использование обезличенных клинических данных и дерматоскопических изображений в научно-исследовательских целях.

Обезличивание записей выполнено до включения в набор данных: персональный идентификатор пациента заменен устойчивым обезличенным кодом; изображения не содержат идентифицирующих признаков пациента (лицо, татуировки, иные уникальные кожные маркеры в поле зрения отсутствуют или исключены при кадрировании).

Набор данных является непубличным и используется исключительно в закрытом исследовательском контуре ФГБУН ИСП РАН. Публикация изображений в открытом доступе не предусмотрена.

2.4. Компонент 1. Стандартная операционная процедура получения изображения

Стандартная операционная процедура (СОП) разработана для условий мобильной дерматоскопии и фиксирует требования к оборудованию, условиям съемки и техническому контролю качества каждого изображения. Получение снимка выполняется оптическим дерматоскопом, сопрягаемым со смартфоном; такая конфигурация соответствует амбулаторной практике

и позволяет проводить съемку на различных клинических площадках без специализированного стационарного оборудования. Указанная конфигурация в литературе ассоциируется с мобильной дерматоскопией и описана как самостоятельный класс средств визуализации [5, 11]. При формировании набора использовались следующие модели дерматоскопов: Heine Delta 20T (контактная поляризованная дерматоскопия) и Melanoscope (поляризованный дерматоскоп собственной разработки ФГБУН ИСП РАН).

Базовые требования СОП организованы по четырем группам:

- 1) **Оптические условия.** Используется поляризованная дерматоскопия с контактным или иммерсионным режимом; стандартизируются углы наклона дерматоскопа относительно поверхности кожи и плотность контакта во избежание артефактов компрессии.
- 2) **Параметры съемки.** Кадрирование выполняется так, чтобы новообразование располагалось в центре поля зрения и занимало основную часть кадра; фокус контролируется перед сохранением изображения; минимальное разрешение и формат файла регламентированы.
- 3) **Условия освещения.** Использование собственного источника поляризованного света дерматоскопа; запрет на смешанное освещение, изменяющее цветовой баланс изображения.
- 4) **Технический контроль качества.** Каждое изображение оценивается по сводному показателю качества `image_quality_score` в порядковой шкале от 1 до 5 баллов: 1 – снимок непригоден (расфокус, существенные блики, обширное закрытие новообразования артефактами); 2 – выраженные дефекты, но различим объект интереса; 3 – приемлемое качество с локальными артефактами; 4 – хорошее качество, минимальные артефакты; 5 – эталонное качество. Снимки с оценкой 1–2 исключаются специалистом непосредственно на этапе съемки и не попадают в дальнейшую обработку; снимки с оценкой 3–5 включаются в массив. Значение `image_quality_score` сохраняется в метаданных для возможности дальнейшей стратификации. За период формирования набора на этапе СОП исключено 32 снимка по критерию `image_quality_score` ≤ 2 .

СОП также регламентирует порядок действий специалиста: первичный визуальный осмотр кожного покрова, выбор объекта съемки, подготовка кожи (при необходимости), серия из нескольких кадров и выбор репрезентативного изображения. Документирование действий выполняется в структурированной форме, что обеспечивает воспроизводимость съемки на разных клинических площадках и согласованность изображений между специалистами.

2.5. Компонент 2. Информационная модель метаданных

Метаданные организованы не как произвольный перечень полей, а как клинически ориентированная информационная модель, состоящая из шести блоков. Состав модели приведен в таблице 2. Структура и нотация полей согласованы с практикой международных дерматоскопических коллекций, что упрощает совместную обработку набора с открытыми источниками и независимую оценку моделей.

Таблица 2. Информационная модель метаданных разработанного набора данных (16 полей)

Блок	Поля	Назначение
Демографический	age_at_exam, sex, fitzpatrick_type	Базовые характеристики пациента и фототип кожи (шкала Фицпатрика I–VI; в наборе представлены типы I–IV)
Анамнестический	sunburn_history, personal_ca_history	Факторы риска: солнечные ожоги в анамнезе, онкологический анамнез
Морфологический	anatomical_site, lesion_diameter_mm, dominant_colors, elevation, border_regularity	Локализация, размер, цвет и геометрические признаки новообразования
Дерматоскопический	dermoscopic_structures	Экспертная аннотация паттернов (сетка, глобулы, псевдоподы, голубая вуаль, сосудистые структуры)
Диагностический	clinical_diagnosis, histopathology_result, verification_stage	Клинический диагноз, гистологический результат, этап верификации
Служебный	examiner_id, image_quality_score	Обезличенный идентификатор специалиста и техническая оценка изображения

Выделение дерматоскопического блока отражает методическую значимость поля `dermoscopic_structures`. Экспертно описанные паттерны представляют собой клинически интерпретируемые признаки, на которые опирается врач при дифференциальной диагностике. Их фиксация позволяет рассматривать набор данных не только как источник меток для классификации, но и как ресурс для количественного сопоставления

областей внимания моделей машинного обучения с клинически значимыми зонами изображения [13, 17], что создает основу для исследований интерпретируемости в последующих работах.

Связь нескольких изображений с одним пациентом сохраняется через устойчивый обезличенный идентификатор. Исходные персональные данные в набор не включаются; при этом структура записи позволяет выполнять разбиение выборки по пациентам и анализ повторных обращений.

2.6. Компонент 3. Многоэтапная экспертная верификация

Диагностическая верификация построена как многоэтапная процедура, повышающая надежность диагностической метки и фиксирующая ее уровень в явном поле `verification_stage`. Состав этапов и их назначение приведены в таблице 3.

Таблица 3. Этапы экспертной верификации диагностической метки

Этап	Содержание	Результат
Первичная клиническая разметка	Дерматолог фиксирует клинический диагноз и описание дерматоскопических структур по форме модели метаданных	Первичная метка с указанием эксперта
Консенсусный пересмотр	Независимое рассмотрение записи тремя специалистами; разногласия разрешаются обсуждением и фиксацией согласованной формулировки	Подтвержденная или скорректированная метка
Гистологическая верификация	Для случаев с подозрением на ЗНО запись дополняется результатом гистопатологического исследования	Метка с морфологическим подтверждением

Первичная разметка и консенсусный пересмотр выполняются в системе разметки изображений Colba, разработанной в ИСП РАН. Использование специализированной среды разметки, а не универсальных средств аннотирования общего назначения, позволяет ограничить вводимые значения предопределенной номенклатурой и снижает риск рассогласования терминологии между специалистами. Программная реализация всех трех компонентов методики – архитектура комплекса, логическая модель хранения данных, интерфейсы клинического ввода и экспертной разметки, механизм журналирования изменений метки – рассматривается отдельно в разделе 3.

Межэкспертная согласованность на этапе независимого рассмотрения записей рассматривается как отдельный контрольный показатель качества разметки. Для последующих версий набора, сохраняющих первичные

решения каждого эксперта до консенсусного обсуждения, она может быть количественно оценена с использованием коэффициента согласия для нескольких экспертов (например, κ Флейса) и включена в паспорт набора данных.

Такой подход позволяет отделить исходную клиническую гипотезу от подтвержденной диагностической метки и сохранить уровень надежности записи. Записи разного уровня верификации не отбраковываются, а маркируются: этап подтверждения становится свойством записи и может явно использоваться при подборе подмножеств для обучения, валидации и независимого тестирования. В клинических исследованиях принято различать «золотой стандарт» (гистология) и «клиническое подтверждение», и предлагаемая методика реализует это различие непосредственно в информационной модели набора данных.

В сформированном наборе все 1026 уникальных изображений получили $\text{verification_stage} \geq 2$ в результате консенсусного пересмотра тремя специалистами. Все 39 изображений злокачественных категорий (меланома – 18, базалиома – 15, плоскоклеточный рак – 6) дополнительно сопровождаются результатом гистопатологического исследования и имеют $\text{verification_stage} = 3$. Доброкачественные и пограничные образования сохраняют уровень $\text{verification_stage} = 2$, поскольку направление на биопсию для них в клинической практике не выполняется. Такая схема повышает надежность диагностических меток для злокачественных классов и позволяет использовать набор при независимой оценке алгоритмов с явным учетом уровня верификации каждой записи.

2.7. Сквозные процедуры: дедупликация и нормализация диагностических категорий

Первичный массив включал 1044 записи. Дедупликация выполнялась по криптографическому хэшу содержимого файла изображения: для каждого исходного снимка рассчитывался хэш побайтового содержимого файла, и записи с совпадающими значениями рассматривались как один графический объект независимо от имени файла и от того, сколько раз запись выгружалась. Метод выявляет только побайтово идентичные файлы: изображения, различающиеся форматом контейнера, метаданными EXIF, параметрами повторного сжатия или кадрированием, при таком сравнении дубликатами не признаются (см. раздел 5). После исключения 18 повторных записей итоговый массив составил 1026 уникальных изображений; доля исключенных дублей – 1,7 % от первичного числа записей. Хэш сохраняется в служебных полях записи и обеспечивает воспроизводимость процедуры дедупликации на последующих этапах ведения набора.

Диагностические категории нормализованы до единой номенклатуры, включающей 9 классов: невус, диспластический невус, меланома, себорейный кератоз, гемангиома, дерматофиброма, базалиома, плоскоклеточный рак, папилома. Записи с неопределенной служебной категорией доброкачественного невусоподобного образования относились к классу невусов при отсутствии клинических признаков злокачественного процесса. Для доброкачественных образований итоговая метка определялась по клинической оценке и экспертному консенсусу; морфологическое заключение учитывалось только в тех случаях, когда оно было доступно в клинической записи.

Нозологические категории описаны по двум независимым признакам (таблица 4 раздела 4): происхождению образования (меланоцитарное или немеланоцитарное) и клиническому статусу (доброкачественное, пограничное или злокачественное). Диспластические невусы выделены в самостоятельную пограничную группу: морфологически они принадлежат к меланоцитарным образованиям и могут демонстрировать признаки атипии, требующие дифференциальной диагностики с меланомой, но не являются злокачественными по определению. При формировании 2-классовой задачи «доброкачественные / злокачественные» (например, для скринингового сценария) диспластические невусы по умолчанию относятся к группе доброкачественных, поскольку их направление на биопсию определяется клинически. Пользователь набора может изменить эту политику и отнести их к злокачественному классу или сформировать отдельный класс — поля `clinical_diagnosis`, происхождения и клинического статуса в таблице 4 раздела 4 сохраняют исходную нозологическую разметку без потери информации.

2.8. Статистическая обработка

Статистическое описание набора включало подсчет числа изображений и пациентов, распределения по нозологическим категориям, полу, возрасту и количеству изображений на пациента. Для возраста рассчитывались минимальное и максимальное значения, среднее значение и медиана. Для категориальных признаков рассчитывались абсолютные и относительные частоты. Расчеты выполнялись по обезличенным итоговым таблицам набора; графики построены по агрегированным количественным данным.

3. Программная реализация методики

Методика формирования набора данных является воспроизводимой в той мере, в какой ее ограничения закреплены не в регламенте, а в программных средствах, через которые проходит каждая запись. Ниже описан программный комплекс, поддерживающий все три компонента методики, логическая модель хранения данных и интерфейсы, с которыми работают клиницист и эксперт-разметчик. Комплекс развернут в закрытом исследовательском контуре ФГБУН ИСП РАН.

3.1. Архитектура программного комплекса

Комплекс построен как последовательность из четырех подсистем, соответствующих стадиям жизненного цикла записи набора данных (рисунок 2).

Клиентское приложение Melanoscope AI [26] обеспечивает работу врача на приеме: ведение обезличенной пациентской карточки, съемку через мобильный дерматоскоп, структурированный ввод полей информационной модели и предварительный вывод системы поддержки принятия решений. Приложение реализует ту часть СОП (раздел 2.4), которая может быть проконтролирована программно: обязательность заполнения ключевых полей, фиксация модели дерматоскопа, выставление `image_quality_score` до сохранения снимка и блокировка сохранения при оценке ≤ 2 .

Серверная часть и хранилище принимают записи по защищенному каналу и разделяют их на две составляющие: структурированные метаданные размещаются в реляционной СУБД, файлы изображений – в объектном хранилище. При приеме файла вычисляется криптографический хэш его содержимого (раздел 2.7), который сохраняется как атрибут записи и обеспечивается ограничением уникальности на уровне схемы БД; тем самым дедупликация выполняется не постобработкой, а средствами СУБД в момент загрузки.

Среда разметки Colba [25] представляет собой надстройку ИСП РАН над платформой Label Studio, адаптированную для задач медицинской аннотации. Colba формирует из записей хранилища проекты разметки, распределяет задания между экспертами, поддерживает коллегиальный режим и ведет журнал изменений диагностической метки. Консенсусная метка и уровень верификации возвращаются в основное хранилище и становятся частью записи.

Модуль выгрузки формирует итоговый набор данных: выполняет обезличивание, применяет нормализацию диагностических категорий (раздел 2.7), вычисляет значение `verification_stage` по журналу изменений метки и экспортирует результат в виде таблицы метаданных и соответствующего ей каталога изображений. Именно выход этого модуля описан в разделе 4.



Рисунок 2. Архитектура программного комплекса подготовки клинически верифицированного набора данных. Пунктирной стрелкой показан возврат согласованной диагностической метки и уровня верификации в основное хранилище

3.2. Логическая модель хранения данных

Логическая модель (рисунок 3) построена вокруг того обстоятельства, что учетной единицей набора является изображение, а учетной единицей клинического контекста – пациент, причем одному пациенту может соответствовать несколько исследований, а одному исследованию – несколько изображений.

Модель включает семь сущностей. Цепочка «Пациент → Исследование → Изображение → Разметка» реализует связи типа 1:N и обеспечивает

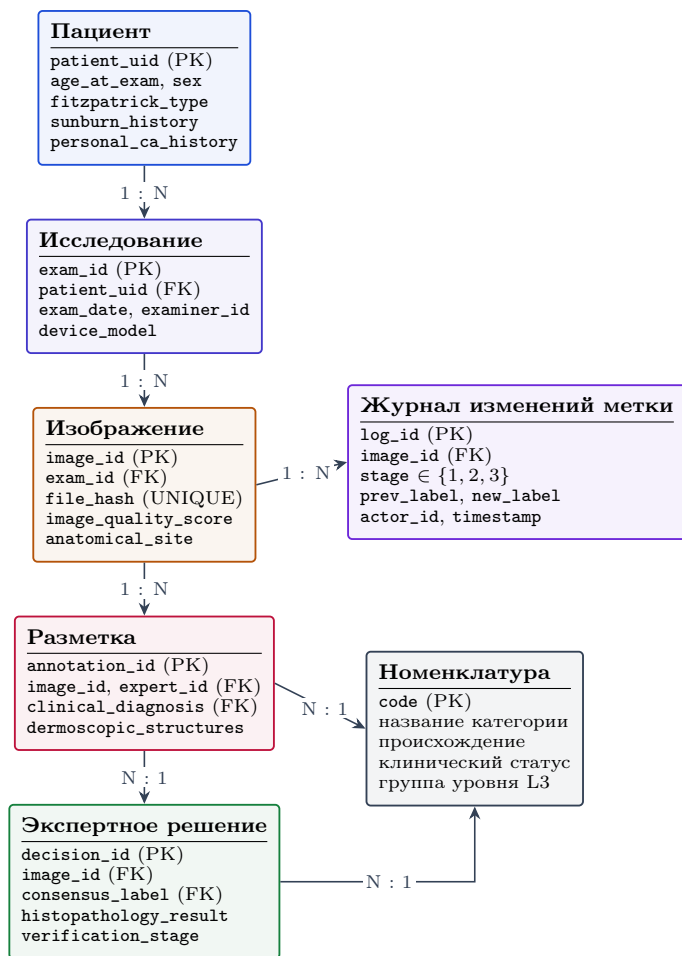


Рисунок 3. Логическая модель хранения данных программного комплекса. Указаны первичные (PK) и внешние (FK) ключи и мощности связей

возможность разбиения выборки по пациентам, необходимость которого обоснована в разделах 4.4 и 4.5. Сущность «Пациент» хранит обезличенный идентификатор **patient_uid** и поля демографического и анамнестического блоков; исходные персональные данные хранятся вне контура набора и в модель не входят. Сущность «Изображение» несет поля морфологического и служебного блоков, включая **file_hash** с ограничением **UNIQUE**—

программную гарантию отсутствия побайтовых дубликатов.

Сущность «Номенклатура» является справочником диагностических категорий и хранит для каждой категории не только наименование, но и классификационные признаки: происхождение образования и клинический статус (см. таблицу 4 раздела 4), а также группу укрупненного уровня агрегации L3 (раздел 4.2). Ссылочная целостность на этот справочник исключает ввод произвольной диагностической формулировки: эксперт может выбрать значение только из предопределенного списка. Именно это ограничение делает возможной последующую однозначную агрегацию категорий (раздел 4.2).

Сущности «Разметка» и «Экспертное решение» намеренно разделены. «Разметка» хранит индивидуальное решение конкретного эксперта с указанием `expert_id` и времени, «Экспертное решение» – итоговую консенсусную метку, результат гистопатологического исследования и вычисленное значение `verification_stage`. Такое разделение позволяет сохранить первичные независимые решения экспертов и делает возможной последующую количественную оценку межэкспертной согласованности (раздел 2.6), не смешивая ее с итоговой меткой записи.

3.3. Интерфейс клинического ввода и ведения пациентской записи

Интерфейс клиентского приложения (рисунок 4) организован вокруг пациентской карточки. В левой части размещены поля клинической карточки, включающие как демографические и анамнестические поля информационной модели (тип кожи по Фицпатрику с развернутой формулировкой каждого варианта, наличие солнечных ожогов в анамнезе), так и дополнительные сведения, не включаемые в обезличенную выгрузку. Развернутые формулировки вариантов (вместо кодов I–VI) введены сознательно: они снижают вариативность интерпретации шкалы Фицпатрика разными специалистами, что является известным источником неоднородности метаданных в дерматологических коллекциях [7, 8].

Состав полей интерфейса шире состава информационной модели набора данных, и это различие является намеренным. Клиническая карточка ведется для нужд приема и содержит дополнительные сведения, которые врач использует при оценке риска, но которые не включаются в обезличенную выгрузку: наследственность, прием иммуносупрессоров, количество и размер невусов, преимущественная локализация. В набор данных экспортируются только 16 полей, перечисленных в таблице 2; остальные поля остаются в клиническом контуре. Шкала фототипа поддерживается интерфейсом в полном диапазоне I–VI, тогда как в сформированном наборе фактически представлены типы I–IV, что отражает состав обращающейся популяции и отдельно отмечено в таблице 2.

Дополнительная информация

Тип волос: Русский

Тип кожи: Тип III: иногда легкий ожог, равномерный загар (золотисто-медовый или оливковый)

Расположение родинок: Голова/шея

Количество родинок: Почти нет

Наследственность: Нет

Иммуносупрессоры и заболевания: Нет

Размер родинок: Есть 1-2 крупных родинок более 5 мм

Солнечные ожоги в анамнезе: Наличие волдырей

Отчет

Удалить

Анализ №189

ML-модель: base 29.06.2026, 13:38 Завершен

ДИАГНОЗ
Кератопапиллома

Вероятности по классам enetv2 · 9 классов

Вероятность меланомы	0%
Вероятность невуса	6%
Вероятность актинического кератоза	0%
Вероятность базальноклеточной карциномы	0%
Вероятность кератоанного поражения	63%
Вероятность дерматофибромы	0%
Вероятность рака кожи	0%
Вероятность сосудистых поражений	0%

Рисунок 4. Интерфейс ведения пациентской записи в клиентском приложении Melanoscope AI

В правой части карточки размещен перечень исследований пациента, что непосредственно отражает связь 1:N сущностей «Пациент» и «Исследование»: на приведенном примере одной пациентской записи соответствуют четыре исследования, выполненные в один день. Карточка отдельного исследования содержит дерматоскопический снимок, предварительный диагноз и распределение вероятностей по девяти классам, полученное встроенной моделью. Существенно, что вывод модели носит вспомогательный характер и *не* используется в качестве диагностической метки набора данных: метка формируется только по процедуре раздела 2.6. Отображение вероятностей сохранено, поскольку в клиническом сценарии оно служит подсказкой при отборе образований для дополнительного рассмотрения, однако на этапе подготовки набора данных соответствующее поле в выгрузку не включается во избежание циркулярности разметки.

3.4. Среда экспертной разметки Colba

Экспертная разметка и консенсусный пересмотр выполняются в среде Colba (рисунок 5). Интерфейс эксперта включает список заданий проекта с миниатюрами и счетчиком выполненных аннотаций, область просмотра изображения с инструментами масштабирования и панорамирования и панель выбора диагностической категории.

Три особенности интерфейса имеют методическое значение.

Во-первых, выбор диагностической категории реализован закрытым списком, соответствующим справочнику «Номенклатура» (раздел 3.2); свободный ввод формулировки диагноза не предусмотрен. Номенклатура разметки содержит одиннадцать категорий и служебное значение «Другое»; в сформированном наборе фактически представлены девять категорий (таблица 4 раздела 4), а служебное значение и не встретившиеся категории устраняются при нормализации (раздел 2.7). Наличие в списке категорий, отсутствующих в итоговом наборе, отражает то, что номенклатура задавалась заранее по клиническому охвату, а не подгонялась под собранный материал.

Во-вторых, в заголовке проекта отображается индикатор вида « (k/n) », означающий число полученных независимых аннотаций k из требуемого числа n . Требуемое число n является параметром проекта разметки и реализует коллегиальный режим: задание считается выполненным только после получения решений от всех назначенных экспертов, после чего запись переводится в режим согласования расхождений.

Именно этот механизм реализует этап консенсусного пересмотра таблицы 3 программно, а не организационно. На приведенном снимке экрана показан проект с $n = 2$; при формировании описываемого набора данных значение этого параметра было установлено равным трем в соответствии с процедурой раздела 2.6, поэтому все 1026 записей имеют $\text{verification_stage} \geq 2$ по результатам рассмотрения тремя специалистами.

В-третьих, среда фиксирует каждое изменение диагностической метки в журнале с указанием предшествующего и нового значения, автора и времени. Журнал обеспечивает прослеживаемость перехода «первичная разметка $\rightarrow$ консенсус $\rightarrow$ гистологическое подтверждение», и



Рисунок 5. Интерфейс эксперта в среде разметки Colba (на снимке – проект с двумя требуемыми аннотациями)

поле автоматически получает значение

$$\text{verification_stage} = \begin{cases} 3, & \text{при наличии в журнале события} \\ & \text{внесения результата гистопато-} \\ & \text{логического исследования,} \\ 2, & \text{при наличии зафиксированного} \\ & \text{консенсусного решения,} \\ 1, & \text{в остальных случаях.} \end{cases}$$

Это исключает расхождение между заявленным и фактическим уровнем верификации записи.

3.5. Обезличивание и выгрузка

Обезличивание выполняется на границе между хранилищем и модулем выгрузки. Персональный идентификатор пациента заменяется устойчивым кодом `patient_uid`, сохраняющим связь изображений одного пациента, но не допускающим обратного сопоставления без таблицы соответствия, которая хранится отдельно и в контур набора данных не передается. Из метаданных исключаются дата и время съемки с точностью выше месяца, сведения о клинической площадке и идентификатор врача заменяется обезличенным `examiner_id`. Изображения проверяются на отсутствие в кадре идентифицирующих признаков.

Выгрузка формируется в виде таблицы метаданных, где строка соответствует изображению, и каталога файлов изображений, связанных с таблицей через `image_id`. Формат полей соответствует нотации, принятой в открытых дерматоскопических коллекциях (раздел 2.5), что позволяет использовать набор в существующих конвейерах обработки без переписывания загрузчиков данных.

3.6. Соотношение с ранее опубликованными программными решениями

Программные средства, описанные в настоящем разделе, следует отличать от системы поддержки принятия врачебных решений, описанной в предшествующих работах [20, 21]. В указанных публикациях предметом описания являлась *диагностическая* часть системы: архитектура вывода, каскадная схема классификации, сценарии применения на приеме и результаты клинической апробации.

Настоящая работа описывает *подготовительную* часть контура – программные средства сбора, хранения, экспертной верификации и выгрузки данных, то есть ту подсистему, выход которой служит входом для диагностических моделей. Общей у двух описаний является лишь клиентская часть Melanoscope AI, рассматриваемая здесь исключительно как средство структурированного ввода и контроля СОП, а не как средство диагностики.

4. Результаты

4.1. Общая характеристика набора

В результате применения методики сформирован набор из 1026 уникальных дерматоскопических изображений, полученных от 443 пациентов. На одного пациента приходится от 1 до 10 изображений; медиана – 2 изображения, среднее – 2,32 изображения; у 190 пациентов зарегистрировано одно изображение. В наборе представлены меланоцитарные и немеланоцитарные образования: меланоцитарные – 767 изображений (74,8 %), немеланоцитарные – 259 изображений (25,2 %). Злокачественные новообразования суммарно составили 39 изображений (3,8 %): 18 – меланома, 15 – базалиома, 6 – плоскоклеточный рак.

ТАБЛИЦА 4. Распределение изображений по нозологическим категориям

Нозологическая категория	Кол-во	Доля, %	Происхождение	Клинический статус
Невус	733	71,4	Меланоцитарное	Доброкачественное
Себорейный кератоз	118	11,5	Немеланоцитарное	Доброкачественное
Гемангиома	87	8,5	Немеланоцитарное	Доброкачественное
Дерматофиброма	31	3,0	Немеланоцитарное	Доброкачественное
Меланома	18	1,8	Меланоцитарное	Злокачественное
Диспластический невус	16	1,6	Меланоцитарное	Пограничное
Базалиома	15	1,5	Немеланоцитарное	Злокачественное
Плоскоклеточный рак	6	0,6	Немеланоцитарное	Злокачественное
Папилома	2	0,2	Немеланоцитарное	Доброкачественное
Итого	1026	100,0	—	—

Преобладание невусов является ожидаемым для амбулаторной дерматоскопической практики, поскольку доброкачественные пигментные образования часто становятся причиной профилактического обращения и требуют дифференциальной диагностики с меланомой. Одновременно такая структура демонстрирует выраженный дисбаланс классов, который должен учитываться при обучении и оценке алгоритмов машинного анализа. Графическое представление распределения приведено на рисунке 6.

Примеры дерматоскопических изображений из сформированного набора приведены на рисунке 7. Снимки иллюстрируют типичный визуальный диапазон изображений мобильной дерматоскопии: различные нозологические категории, цветовой и структурный полиморфизм, характерный для амбулаторной выборки.

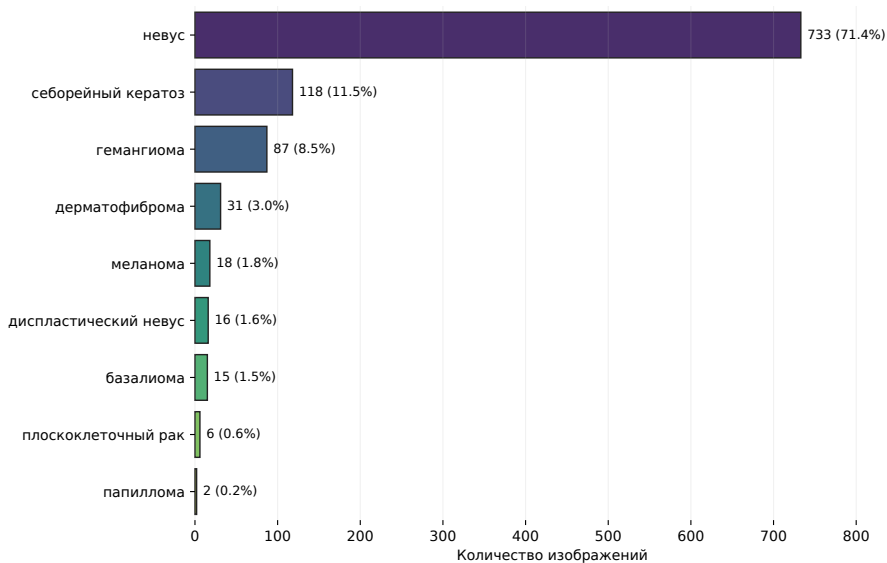


Рисунок 6. Распределение изображений по нозологическим категориям

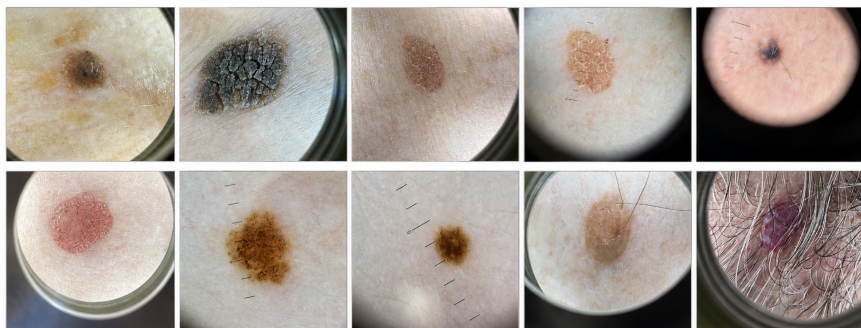


Рисунок 7. Примеры изображений из собранного датасета

4.2. Уровни агрегации диагностических категорий

Распределение, приведенное в таблице 4, характеризуется выраженным дисбалансом: на четыре наиболее частые категории приходится 94,4 % изображений, тогда как четыре редкие категории представлены менее чем двадцатью изображениями каждая. При таком распределении девятиклассовая постановка не обеспечивает интервальных оценок метрик приемлемой ширины для редких классов, что необходимо учитывать при

любом использовании набора.

Поскольку справочник диагностических категорий содержит два независимых классификационных признака – происхождение и клинический статус (раздел 3.2) – агрегация категорий выполняется однозначно и без потери информации. В таблице 5 приведены четыре согласованных уровня агрегации, определенные для набора, с указанием численности каждого класса и области применимости уровня.

Таблица 5. Уровни агрегации диагностических категорий сформированного набора (выделен наименьшая численность)

Уровень	Классы и численность	Область применимости
L1	1) не-злокачественные – 987; 2) злокачественные – 39	Скрининговый сценарий «направить / наблюдать»
L2	1) Доброкачественные – 971; 2) пограничные – 16 ; 3) злокачественные – 39	Оценка поведения модели на пограничной группе
L3	1) Меланоцитарные незлокачественные – 749; 2) кератиноцитарные незлокачественные – 120; 3) прочие немеланоцитарные незлокачественные – 118; 4) злокачественные – 39	Многоклассовая оценка; три класса из четырех содержат ≥ 100 изображений
L4	Полная нозологическая номенклатура (таблица 4)	Качественный анализ, интерпретируемость, отбор подвыборок; для оценки метрик по редким классам не рекомендуется

На уровне L3 меланоцитарный класс объединяет невусы и диспластические невусы и потому обозначен как «незлокачественный», а не «доброкачественный»: пограничный статус диспластического невуса, выделенный на уровне L2 и в таблице 4, при таком укрупнении не утрачивается, а лишь не различается. Уровень L3 отвечает требованию наполненности классов лишь частично: три из четырех классов содержат не менее 100 изображений (749, 120 и 118), тогда как класс злокачественных образований объемом 39 изображений остается наименьшим – при том что

именно он полностью гистологически верифицирован. Статистическая точность оценок для этого класса, таким образом, ограничена, и это ограничение не устраняется агрегацией.

Дальнейшее укрупнение злокачественного класса за счет объединения с пограничной группой методически неприемлемо, поскольку диспластический невус не является злокачественным образованием по определению; наращивание же его объема средствами заимствования из открытых коллекций противоречит целевому назначению набора как независимого оценочного множества (см. введение и раздел 6). Численность злокачественного класса увеличивается по мере продолжения сбора по той же СОП, что сохраняет однородность ресурса.

Уровни L1–L3 определены как соответствие категорий уровня L4 и заданы атрибутами справочника «Номенклатура» (раздел 3.2), а не вычисляются пользователем набора. Уровни L1 и L2 выводятся непосредственно из атрибута клинического статуса; уровень L3 требует дополнительного признака гистогенетической группы, поскольку по происхождению и статусу кератиноцитарные и сосудисто-фиброзные доброкачественные образования неразличимы, и этот признак хранится в справочнике отдельным полем. Тем самым любые две работы, использующие набор на одном уровне агрегации, оперируют идентичным разбиением.

4.3. Демографическая структура

После исключения некорректных возрастных значений возрастной анализ включал 439 пациентов. Возраст находился в диапазоне от 2 до 90 лет; медиана составила 38 лет, среднее значение – 41,1 года. Возрастное распределение представлено на рисунке 8.

В выборке зафиксировано 279 женщин (63,0 %) и 164 мужчины (37,0 %). Преобладание женщин может быть связано с особенностями обращаемости за профилактическим дерматоскопическим осмотром и не должно интерпретироваться как популяционная оценка распространенности новообразований кожи. Распределение по полу представлено на рисунке 9.

4.4. Количество изображений на пациента

Поскольку один пациент может иметь несколько кожных образований, отдельной характеристикой набора является распределение количества изображений на пациента (рисунок 10). Наличие нескольких снимков у части пациентов отражает реальный клинический сценарий дерматоскопического осмотра, при котором врач оценивает не одно изолированное образование, а несколько клинически значимых очагов. Это свойство имеет методическое значение при разбиении набора на обучающую и тестовую части: для предотвращения переоценки качества алгоритмов необходима стратификация по пациентам.

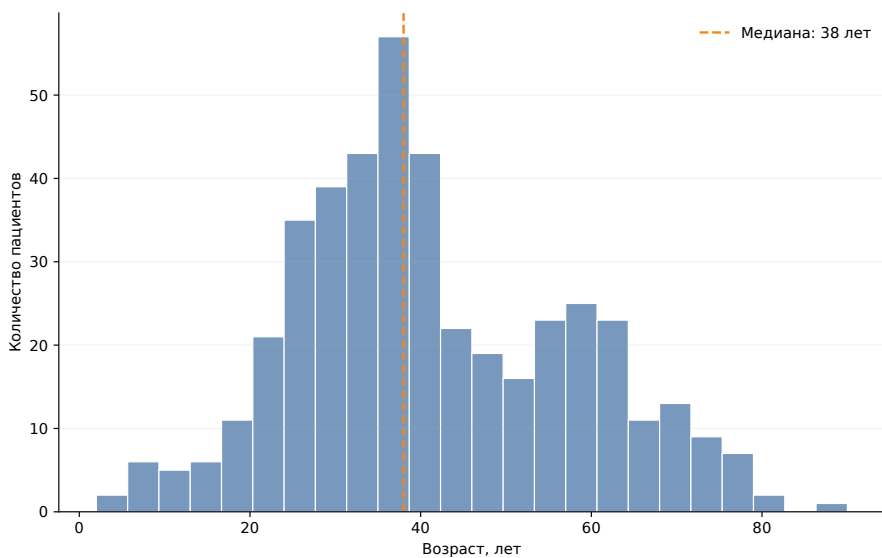


Рисунок 8. Возрастное распределение пациентов

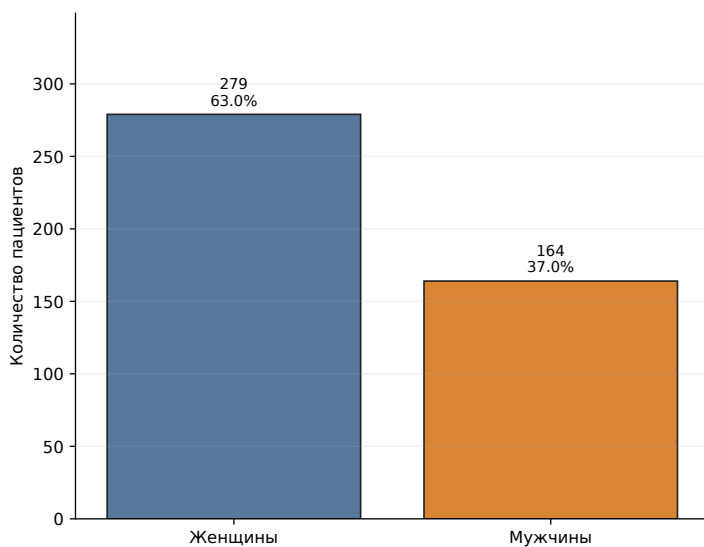


Рисунок 9. Распределение пациентов по полу

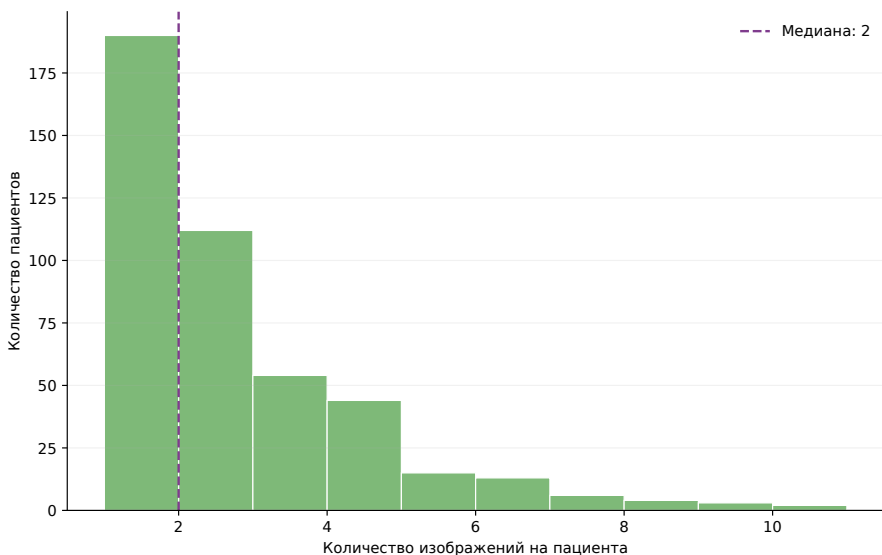


Рисунок 10. Количество изображений на пациента

4.5. Совстречаемость нозологических категорий в пределах пациентской записи

Наличие у 253 пациентов более одного изображения позволяет анализировать не только распределение категорий по изображениям, но и их сочетания в пределах одной пациентской записи. Такой анализ возможен именно благодаря сохранению связи «пациент – изображение» в информационной модели (раздел 3.2) и в большинстве открытых дерматоскопических коллекций невыполним из-за отсутствия устойчивого пациентского идентификатора.

Из 443 пациентов у 336 все изображения относятся к одной нозологической категории, у 89 представлены две различные категории, у 16 – три, у 2 – четыре; таким образом, две и более различные категории зарегистрированы у 107 пациентов (24,2 % выборки, или 42,3 % пациентов с несколькими изображениями). Матрица совстречаемости приведена на рисунке 11: элемент вне главной диагонали равен числу пациентов, у которых обе категории представлены одновременно, элемент главной диагонали – числу пациентов, у которых данная категория встречается совместно хотя бы с одной другой.

Наиболее частые сочетания приведены в таблице 6. Доли рассчитаны от 253 пациентов, имеющих более одного изображения.

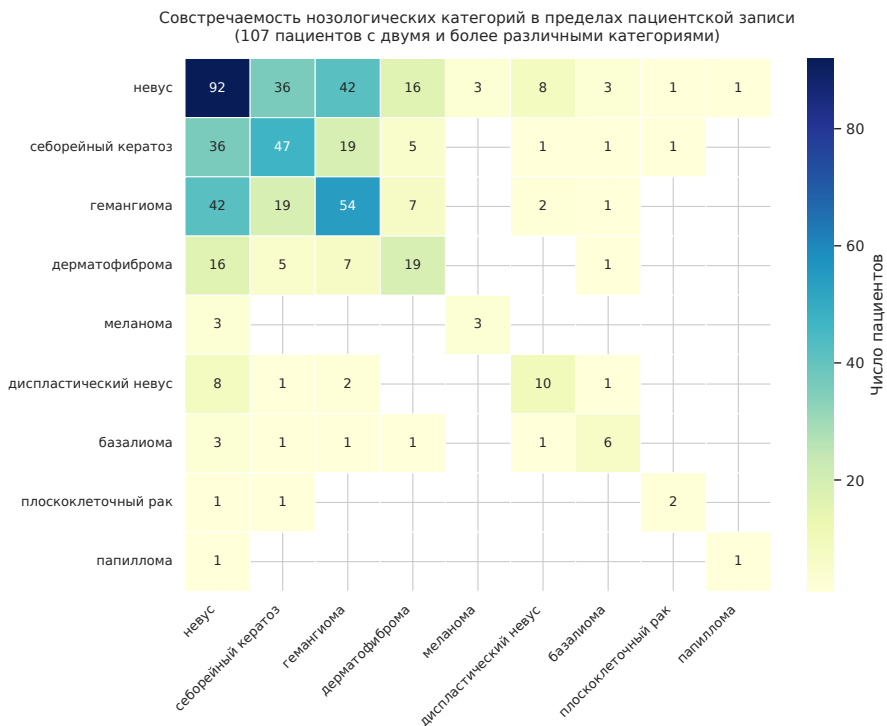


Рисунок 11. Матрица совстречаемости нозологических категорий в пределах пациентской записи

Полученная структура сочетаний допускает содержательную клиническую интерпретацию. Преобладающие пары образованы наиболее распространенными доброкачественными образованиями и отражают типичную картину профилактического дерматоскопического осмотра, при котором врач оценивает несколько образований разной природы у одного пациента. Сочетание «невус + диспластический невус» (8 пациентов) согласуется с представлением о множественном характере диспластических изменений меланоцитарных образований.

Клинически наиболее значимы сочетания злокачественных категорий с доброкачественными: «невус + меланома» (3 пациента) и «невус + базалиома» (3 пациента) соответствуют сценарию, в котором злокачественное образование выявляется в ходе осмотра, поводом для которого послужило другое образование. Всего злокачественные образования зарегистрированы у 31 пациента; у 11 из них в пациентской записи присутствуют также

ТАБЛИЦА 6. Наиболее частые сочетания нозологических категорий в пределах пациентской записи

Сочетание категорий	Пациентов	Доля, %
<i>Пары</i>		
Невус + гемангиома	42	16,6
Невус + себорейный кератоз	36	14,2
Гемангиома + себорейный кератоз	19	7,5
Невус + дерматофиброма	16	6,3
Невус + диспластический невус	8	3,2
Гемангиома + дерматофиброма	7	2,8
Себорейный кератоз + дерматофиброма	5	2,0
Невус + меланома	3	1,2
Невус + базалиома	3	1,2
<i>Тройки</i>		
Невус + гемангиома + себорейный кератоз	10	4,0
Невус + гемангиома + дерматофиброма	5	2,0
Невус + себорейный кератоз + дерматофиброма	5	2,0
Гемангиома + себорейный кератоз + дерматофиброма	2	0,8

доброкачественные категории, у остальных 20 представлены только злокачественные образования, причем в 17 случаях запись содержит единственное изображение. Сочетаний двух различных злокачественных категорий у одного пациента в наборе не зафиксировано.

Это распределение отражает различие клинических маршрутов: доброкачественные образования фиксируются преимущественно в ходе профилактического осмотра, тогда как злокачественные – при целевом обращении по поводу конкретного образования.

Из выявленных сочетаний следует методическое требование, усиливающее сформулированное в разделе 4.4. Наличие внутрипациентских корреляций диагнозов означает, что случайное разбиение по изображениям приводит не только к утечке визуальных признаков конкретного пациента между обучающей и тестовой частями, но и к утечке информации о совместном распределении категорий. Поэтому разбиение выборки при обучении и тестировании должно выполняться *по пациентам*, а не по изображениям; для сохранения нозологического баланса разбиение целесообразно стратифицировать по набору категорий пациентской записи, а не по категории отдельного изображения. Соответствующее поле пациентского идентификатора включено в выгрузку набора (раздел 3.5).

4.6. Сводные характеристики набора

Основные количественные характеристики набора приведены в таблице 7. Они рассчитаны по обезличенным итоговым таблицам и описывают полноту и согласованность сформированного массива.

Таблица 7. Сводные характеристики сформированного набора данных

Показатель	Значение
Исключено на этапе СОП по <code>image_quality_score</code> ≤ 2	32
Исходные записи (после СОП-контроля качества)	1 044
Исключено дублей (по криптографическому хэшу)	18
Доля исключенных дублей, %	1,7
Уникальные изображения	1 026
Пациенты с изображениями	443
Минимум / медиана / среднее / максимум изображений на пациента	1 / 2 / 2,32 / 10
Диагностические категории (уровень L4)	9
Уровни агрегации категорий (L1 / L2 / L3 / L4)	2 / 3 / 4 / 9
Пациентов с двумя и более различными категориями	107
Пациентов с тремя и более различными категориями	18
Полей в информационной модели метаданных	16
Изображений с <code>verification_stage</code> = 2 (консенсус)	987
Изображений с <code>verification_stage</code> = 3 (гистология)	39
Доля гистологически верифицированных от всех ЗНО, %	100,0

5. Обсуждение

Предложенная методика рассматривает формирование клинического дерматоскопического набора данных как самостоятельную задачу медицинской информатики. Ее новизна состоит не в отдельных технических приемах, а в одновременной формализации трех компонентов – стандартной операционной процедуры съемки, информационной модели метаданных и многоэтапной экспертной верификации, – рассматриваемых как взаимосвязанные элементы единого информационного ресурса. В сравнении с подходом, в котором набор данных описывается только

числом изображений и перечнем классов, такая структура фиксирует происхождение каждой записи, связь с обезличенной пациентской записью, состав структурированных метаданных и уровень подтверждения диагноза. Это особенно важно для последующего применения набора в задачах глубокого обучения, где ошибки разметки, дубли и неполнота клинического контекста могут существенно искажать оценку качества моделей [10–16].

Описание стандартной операционной процедуры получения изображения переводит требования к съемке из области «клинического умения специалиста» в область проверяемых технологических условий. Это важно по двум причинам. Во-первых, описанная СОП повышает воспроизводимость съемки на нескольких клинических площадках и снижает вариативность входных изображений между специалистами. Во-вторых, явная регламентация параметров мобильной дерматоскопии (поляризация, фокусировка, центрирование, оценка качества) позволяет аналитически отделить искажения, связанные с условиями съемки, от искажений, обусловленных свойствами модели. В международных дерматоскопических коллекциях такая регламентация представлена неоднородно [10], что усложняет независимое сравнение моделей.

Информационная модель метаданных из 16 полей расширяет область применимости набора. Демографический и анамнестический блоки позволяют оценивать устойчивость алгоритмов в подгруппах пациентов, морфологический и дерматоскопический блоки – анализировать клинические факторы дифференциальной диагностики, диагностический блок – корректно разделять клиническую гипотезу, морфологическое подтверждение и уровень верификации. Поле `dermoscopic_structures` имеет особое значение: оно создает количественную основу для анализа интерпретируемости моделей, обеспечивая возможность сопоставления областей высокой активации классификатора с экспертно описанными паттернами изображения [13, 17]. Эта возможность реализована количественно в работе по клинической валидации системы Melanoscope на материале данного набора [22]: IoU между картами внимания и аннотациями поля `dermoscopic_structures` составил 0.51–0.69 для различных архитектур, согласованность системы с экспертным заключением – 88.6 %. Набор таким образом создает условия для последующих исследований объяснимого искусственного интеллекта в дерматоонкологии.

Многоэтапная экспертная верификация формирует прозрачную цепочку надежности диагностической метки. Этап `verification_stage` сохраняется в записи и используется как полноценный признак при отборе подвыборок для обучения и тестирования. Гистологическая верификация

злокачественных классов соответствует представлению о «золотом стандарте» в клинической диагностике и существенно снижает риск систематических ошибок при оценке чувствительности классификатора к злокачественным образованиям.

5.1. Аугментация изображений набора

Ограниченный объем набора и дисбаланс классов делают вопрос о допустимой аугментации изображений частью методики, а не частностью конкретного эксперимента. Особенность дерматоскопии состоит в том, что диагностически значимыми являются как раз те свойства изображения, которые в задачах общего компьютерного зрения принято считать несущественными, поэтому набор преобразований здесь не может быть заимствован без ограничений.

Геометрические преобразования применимы к дерматоскопическому кадру шире, чем к большинству медицинских изображений. Дерматоскопический кадр не имеет выделенной ориентации: положение образования в поле зрения определяется тем, как врач расположил дерматоскоп, а не анатомией, поэтому горизонтальное и вертикальное отражение, транспонирование и повороты на произвольный угол не порождают клинически невозможных изображений. Умеренные аффинные преобразования, сдвиг, масштабирование и слабые эластические деформации допустимы и дают регуляризующий эффект. Ограничение накладывается на интенсивность: агрессивные деформации и окклюзии способны разрушить тонкие структуры (пигментную сеть, точки и глобулы, псевдоподы), несущие диагностическую информацию, поэтому эффект таких преобразований подлежит контролю раздельно на внутридоменной и внешней выборках, а не по агрегированной метрике.

Фотометрические преобразования представляют для рассматриваемого набора наибольший интерес, поскольку моделируют именно тот источник доменного сдвига, для измерения которого набор и предназначен: различия дерматоскопов, спектральных характеристик подсветки, настроек баланса белого и постобработки изображения в камере смартфона. Цветовые джиттеры, преобразования в пространстве оттенков – насыщенность – яркость, гамма-коррекция, локальное выравнивание гистограммы, моделирование хроматической аберрации и методы цветовой константности заставляют модель опираться на структуру образования, а не на абсолютные значения RGB, которые в значительной мере определяются источником данных. В исследовании поиска устойчивых политик аугментации [24], показано, что при переносе бинарного классификатора на источники, не представленные в обучении, наибольший прирост качества дают именно комбинированные политики с преобладанием фотометрических операций, тогда как максимизация внутридоменной точности не является для этого надежным ориентиром.

Моделирование артефактов съемки составляет третью группу. Волосы, покров, пузырьки иммерсионной среды, блики и размыточные следы маркера являются регулярными артефактами мобильной дерматоскопии; известно, что нанесенная врачом разметка может систематически смещать решение классификатора [16]. Соответственно, осмысленными являются как синтетическое добавление таких артефактов, так и их программное удаление; и то и другое повышает устойчивость модели к неинформативным корреляциям.

Отдельно следует зафиксировать преобразования, *недопустимые* для данной предметной области. Интенсивные цветовые искажения, изменяющие соотношение оттенков внутри образования, модифицируют сам классифицируемый признак: голубовато-белая вуаль, цветовой полиморфизм и характер сосудистых структур входят в диагностические алгоритмы дерматоскопии и описаны в поле `dermoscopic_structures` информационной модели. Аугментация, изменяющая эти признаки, порождает изображения с некорректной меткой. По той же причине неприемлемы преобразования, инвертирующие или существенно смещающие цветовой тон, и агрессивное локальное контрастирование, создающее структуры, отсутствовавшие в исходном кадре.

Информационная модель набора позволяет управлять аугментацией адресно. Поле `image_quality_score` дает возможность стратифицировать интенсивность преобразований по исходному качеству снимка: применять артефактные и размывающие преобразования преимущественно к изображениям высокого качества, не ухудшая дополнительно снимки с оценкой 3. Поле `fitzpatrick_type` позволяет ограничивать цветовые преобразования так, чтобы они не выводили изображение за пределы физически достижимого диапазона тона кожи. Наконец, при использовании набора как оценочного множества аугментация к нему не применяется вовсе: описанные преобразования относятся к обучающим выборкам, тогда как внешняя проверка должна выполняться на неизмененных изображениях, иначе измеряемый доменный сдвиг оказывается искажен самой процедурой аугментации.

5.2. Соотношение с предшествующими публикациями

Ранее опубликованный пилотный материал по созданию клинического дерматоскопического набора [18] включал 657 дерматоскопических изображений с разметкой клинически значимых признаков, учитывающих фототипы кожи населения Российской Федерации. На этом материале

выполнены прикладные исследования, в которых датасет использовался как источник данных, а не как объект методической формализации: оценка алгоритмов многоклассовой и каскадной двухэтапной классификации дерматоскопических изображений на собственном наборе и репозитории ISIC-2019 [19], архитектурное описание интеллектуальной системы поддержки принятия врачебных решений, реализующей разработанные алгоритмы [20], а также клиническая валидация скринингового сценария с использованием мобильной дерматоскопии и оценкой согласованности с экспертным заключением [21].

Настоящая работа представляет новый клинически верифицированный набор, сформированный по формализованной методике, а не продолжение сбора пилотного набора, описанного ранее [18]. По сравнению с предшествующим этапом исследования изменен сам принцип подготовки данных:

- (i) набор сформирован заново и включает 1026 уникальных изображений от 443 пациентов с явной фиксацией распределения количества изображений на пациента;
- (ii) сформулирована стандартная операционная процедура мобильной дерматоскопической съемки, регламентирующая оптические условия, параметры съемки, условия освещения и технический контроль качества изображения;
- (iii) перечень метаданных организован в информационную модель из 16 структурированных полей по шести клинически ориентированным блокам в нотации, совместимой с ISIC, с явным выделением поля экспертной аннотации дерматоскопических структур;
- (iv) диагностическая метка задается через многоэтапную процедуру экспертной верификации (первичная разметка → консенсус трех специалистов → гистологическое подтверждение злокачественных новообразований) с фиксацией уровня надежности в служебном поле `verification_stage`;
- (v) дедупликация выполнена по криптографическому хэшу содержимого файла изображения и удалила 18 повторных записей из 1044 исходных.

Тем самым настоящая статья представляет самостоятельный исследовательский ресурс и формализует трехкомпонентную методику его формирования, выделяя СОП съемки, информационную модель метаданных и многоэтапную верификацию в качестве самостоятельных взаимосвязанных компонентов. Прикладные исследования [19–21], выполненные ранее на пилотном материале, рассматриваются как предшествующий этап и не определяют состав настоящего набора.

Примененная дедупликация по криптографическому хэшу содержимого файла устойчива к различиям в именах файлов и к повторным выгрузкам и сохраняет воспроизводимость процедуры через служебный хэш записи. Ее областью действия являются побайтово идентичные файлы; изображения, подвергшиеся перекодированию, изменению метаданных контейнера или кадрированию, этим способом не выявляются. В перспективе процедура может быть дополнена перцептивным хэшированием для выявления визуально близких изображений, не совпадающих побайтно.

Разработанный набор не заменяет крупные открытые коллекции, а дополняет их как локальный клинический ресурс с явно документированной информационной моделью. Его ценность состоит в связке мобильной дерматоскопии, расширенных пациентских и дерматоскопических метаданных и многоэтапной верификации; такой ресурс может применяться для независимого тестирования моделей, обученных на открытых источниках, и для исследования методов адаптации алгоритмов к российской амбулаторной практике. Необходимость локально верифицированного ресурса для независимого тестирования подтверждается количественно: при переносе моделей, обученных на ISIC, на клинические данные разработанного набора ROC-AUC снижается с 0.952–0.966 до 0.797–0.893, а ожидаемая ошибка калибровки (ECE) возрастает с 0.02 до 0.27–0.39 [23].

6. Ограничения

Сформированный набор данных имеет ряд ограничений, которые должны учитываться при его использовании:

- 1) **Дисбаланс классов и объем.** Распределение нозологических категорий отражает структуру амбулаторной обращаемости и характеризуется выраженным преобладанием невусов; малое число изображений отдельных злокачественных категорий ограничивает прямую оценку метрик для редких классов и требует применения стратифицированных схем валидации и разбиения по пациентам.

Для работы в этих условиях определены уровни агрегации категорий (раздел 4.2): на уровнях L1–L3 наполненность классов допускает расчет интервальных оценок, точность которых для малочисленных классов (16 изображений в пограничной группе уровня L2 и 39 в злокачественном классе уровней L1 и L3) остается ограниченной, тогда как полная девятиклассовая постановка L4 для оценки метрик по редким классам не рекомендуется.

Устранение дисбаланса за счет заимствования изображений из открытых коллекций сознательно отвергнуто: оно несовместимо с целевым назначением набора как независимого оценочного множества (см. введение), поскольку делает невозможным измерение доменного сдвига относительно этих же коллекций и вносит неучитываемую неоднородность процедуры подтверждения диагноза (таблица 1).

В текущем объеме набор не рассматривается как обучающая база для многоклассовой диагностики.

- (2) **Региональная специфика.** Набор формировался в условиях российской амбулаторной практики и не претендует на представительство мировой популяции; внешняя валидация на независимых выборках других регионов является направлением дальнейшей работы.
- (3) **Полнота метаданных.** Часть полей информационной модели может быть неполна для отдельных записей, что типично для данных, получаемых в условиях реальной клинической практики; такие записи маркируются и сохраняются в наборе с явным указанием уровня заполненности.
- (4) **Уровень верификации.** Записи доброкачественных и пограничных образований подтверждены консенсусом трех специалистов (`verification_stage` = 2), но не имеют гистологического подтверждения, поскольку биопсия в таких случаях клинически не показана. Это ограничивает применение набора в качестве «золотого стандарта» при оценке специфичности классификатора в доброкачественных классах; при необходимости строгого уровня верификации следует отбирать подвыборку по полю `verification_stage` = 3.
- (5) **Межэкспертная согласованность.** В текущей версии набора сохраняется итоговая консенсусная метка, тогда как первичные независимые решения каждого эксперта в выгрузку не включены. Это не позволяет рассчитать коэффициент согласия для нескольких экспертов по опубликованному ресурсу. Логическая модель хранения (раздел 3.2) разделяет сущности «Разметка» и «Экспертное решение» и допускает такую выгрузку без изменения схемы данных; включение первичных решений и оценки κ Флейса в паспорт набора запланировано в следующей версии.
- (6) **Непубличность ресурса.** Набор используется в закрытом исследовательском контуре и не размещается в открытом доступе, что ограничивает независимую воспроизводимость результатов, полученных на нем, внешними исследовательскими группами. Воспроизводимой при этом остается сама методика: СОП, информационная модель, процедура верификации и описанные программные средства (раздел 3) допускают повторное применение к иному клиническому материалу.

7. Заключение

Разработана трехкомпонентная методика формирования клинически верифицированного набора дерматоскопических изображений, описана ее программная реализация и по этой методике создан набор данных, предназначенный для независимой (внешней) оценки моделей в задачах медицинской информатики.

Основные результаты и вклад работы:

- (1) Формализована трехкомпонентная архитектура методики формирования клинического дерматоскопического набора данных, обеспечивающая воспроизводимость процедуры съемки, проверяемую полноту структурированных метаданных и явный учет уровня подтверждения диагностической метки.
- (2) Разработана стандартная операционная процедура мобильной дерматоскопической съемки, регламентирующая оптические условия, параметры съемки, условия освещения и технический контроль качества изображения; СОП пригодна для применения на различных клинических площадках без специализированного стационарного оборудования.
- (3) Разработана информационная модель метаданных из 16 полей в нотации, совместимой с международными дерматоскопическими коллекциями, организованная по шести клинически ориентированным блокам и содержащая поле экспертной аннотации дерматоскопических структур, создающее основу для последующей количественной оценки интерпретируемости моделей.
- (4) Формализована многоэтапная процедура экспертной верификации диагностической метки с явной фиксацией уровня надежности записи в поле `verification_stage`, обеспечивающая корректное разделение клинической гипотезы и подтвержденной метки.
- (5) Описана программная реализация методики: архитектура комплекса подготовки данных, логическая модель хранения записей «пациент – исследование – изображение – разметка – экспертное решение», интерфейс структурированного клинического ввода и среда коллегиальной экспертной разметки Colba с журналированием изменений диагностической метки.
- (6) Ключевые ограничения методики закреплены программно: уникальность `file_hash` на уровне схемы БД, ссылочная целостность на справочник диагностических категорий и автоматическое вычисление `verification_stage` по журналу изменений.

- (7) По разработанной методике сформирован клинически верифицированный набор данных, включающий 1026 уникальных дерматоскопических изображений от 443 пациентов, 9 нозологических категорий и гистологически подтвержденные злокачественные классы; набор может использоваться как локальный ресурс для независимого тестирования, анализа интерпретируемости и доменной адаптации.
- (8) Для работы в условиях дисбаланса определены четыре согласованных уровня агрегации диагностических категорий, хранимых в справочнике номенклатуры, и выполнен анализ встречаемости нозологий в пределах пациентской записи: две и более различные категории зарегистрированы у 107 из 443 пациентов, что обосновывает требование стратифицированного разбиения выборки по пациентам, а не по изображениям. Раздельно рассмотрены допустимые и недопустимые для дерматоскопии виды аугментации изображений.

Практическая значимость работы состоит в формализации воспроизводимого подхода к подготовке клинических дерматоскопических наборов данных. Предложенная информационная модель и процедура верификации могут применяться при формировании новых медицинских коллекций изображений, внешней проверке моделей машинного анализа, анализе доменного сдвига, методической демонстрации алгоритмов и сравнительной оценке интерпретируемости в условиях, приближенных к амбулаторной практике.

Дальнейшее развитие работы связано с пополнением редких диагностических классов, расширением географии данных и количественной оценкой согласованности областей внимания моделей с экспертно описанными дерматоскопическими структурами на основе поля `dermoscopic_structures`.

Список использованных источников

- [1] Tschandl P., Rosendahl C., Kittler H. *The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions* // Scientific Data. – 2018. – Vol. 5. – id. 180161. – 9 pp.  [↑](#)58, 61, 63
- [2] Codella N. C. F., Gutman D., Celebi M. E., Helba B., Marchetti M. A., Dusza S. W., Kallou A., Liopyris K., Mishra N., Kittler H., Halpern A. *Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging* // 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (Washington, DC, USA, 04–07 April 2018). – IEEE. – 2018. – ISBN 978-1-5386-3636-7. – Pp. 168–172.  [↑](#)58, 61, 63

- [3] Rotemberg V., Kurtansky N., Betz-Stablein B., Caffery L., Chousakos E., Codella N., Combalia M., Dusza S., Guitera P., Gutman D., Halpern A., Helba B., Kittler H., Kose K., Langer S., Lioprys K., Malvey J., Musthaq S., Nanda J., Reiter O., Shih G., Stratigos A., Tschandl Ph., Weber J., Soyer H. P. *A patient-centric dataset of images and metadata for identifying melanomas using clinical context* // Scientific Data.– 2021.– Vol. **8**.– id. 34.– 8 pp.  [↑58, 61, 63](#)
- [4] Hernandez-Perez C., Combalia M., Podlipnik S., Codella N. C. F., Rotemberg V., Halpern A. C., Reiter O., Carrera C., Barreiro A., Helba B., Puig S., Vilaplana V., Malvey J. *BCN20000: Dermoscopic lesions in the wild* // Scientific Data.– 2024.– Vol. **11**.– id. 641.– 9 pp.  [↑58, 61, 63](#)
- [5] Pacheco A. G. C., Lima G. R., Salomao A. S., Krohling B., Biral I. P., de Angelo G. G., Alves F. C. R. Jr, Esgario J. G. M., Simora A. C., Castro P. B. C., Rodrigues F. B., Frasson P. H. L., Krohling R. A., Knidel H., Santos M. C. S., do Espírito Santo R. B., Marcelo T. L. S. G., Canuto T. R. P., de Barros L. F. S. *PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones* // Data in Brief.– 2020.– Vol. **32**.– id. 106221.– 10 pp.  [↑58, 61, 63, 68](#)
- [6] Daneshjou R., Vodrahalli K., Novoa R. A., Jenkins M., Liang W., Rotemberg V., Ko J., Swetter S. M., Bailey E. E., Gevaert O., Mukherjee P., Phung M., Yekrang K., Fong B., Sahasrabudhe R., Allerup J. A. C., Okata-Karigane U., Zou J., Chiou A. S. *Disparities in dermatology AI performance on a diverse, curated clinical image set* // Science Advances.– 2022.– Vol. **8**.– No. 31.– id. eabq6147.  [↑58, 62](#)
- [7] Groh M., Harris C., Soenksen L. R., Lau F., Han R., Kim A., Koochek A., Badri O. *Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset* // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Nashville, TN, USA, 19–25 June 2021).– IEEE.– 2021.– ISBN 978-1-6654-4900-7.– Pp. 1820–1828.  [↑58, 62, 76](#)
- [8] Kinyanjui N. M., Odonga T., Cintas C., Codella N. C. F., Panda R., Sattigeri P., Varshney K. R. *Estimating skin tone and effects on classification performance in dermatology datasets*.– 2019.– 11 pp. [arXiv:1910.13268](#) [cs.CV]  [↑58, 62, 76](#)
- [9] Esteva A., Kuprel B., Novoa R. A., Ko J., Swetter S. M., Blau H. M., Thrun S. *Dermatologist-level classification of skin cancer with deep neural networks* // Nature.– 2017.– Vol. **542**.– Pp. 115–118.  [↑58](#)
- [10] Wen D., Khan S. M., Xu A. J., Ibrahim H., Smith L., Caballero J., Zepeda L., de Blas Perez C., Denniston A. K., Liu X., Matin R. N. *Characteristics of publicly available skin cancer image datasets: A systematic review* // The Lancet Digital Health.– 2022.– Vol. **4**.– No. 1.– Pp. e64–e74.  [↑58, 62, 90](#)
- [11] Brinker T. J., Hekler A., Enk A. H., Berking C., Haferkamp S., Hauschild A., Weichenthal M., Klode J., Schadendorf D., Holland-Letz T., von Kalle C., Fröhling S., Schilling B., Utikal J. S. *Deep neural networks are superior to dermatologists in melanoma image classification* // European Journal of Cancer.– 2019.– Vol. **119**.– Pp. 11–17.  [↑58, 62, 68, 90](#)

- [12] Mendonca T., Ferreira P. M., Marques J. S., Marcal A. R. S., Rozeira J. *PH² – A dermoscopic image database for research and benchmarking* // *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka, Japan, 03–07 July 2013).– IEEE.– 2013.– ISBN 978-1-4577-0216-7.– Pp. 5437–5440.  ↑_{61, 63, 90}
- [13] Combalia M., Codella N., Rotemberg V., Carrera C., Dusza S., Gutman D., Helba B., Kittler H., Kurtansky N. R., Liopyris K., Marchetti M. A., Podlipnik S., Puig S., Rinner Ch., Tschandl Ph., Weber J., Halpern A., Malvey J. *Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: The 2019 International Skin Imaging Collaboration grand challenge* // *The Lancet Digital Health*.– 2022.– Vol. **4**.– No. 5.– Pp. e330–e339.  ↑_{61, 63, 70, 90}
- [14] Daneshjou R., Smith M. P., Sun M. D., Rotemberg V., Zou J. *Lack of transparency and potential bias in artificial intelligence data sets and algorithms: A scoping review* // *JAMA Dermatology*.– 2021.– Vol. **157**.– No. 11.– Pp. 1362–1369.  ↑_{62, 90}
- [15] Gebru T., Morgenstern J., Vecchione B., Vaughan J. W., Wallach H., Daume H. III, Crawford K. *Datasheets for datasets* // *Communications of the ACM*.– 2021.– Vol. **64**.– No. 12.– Pp. 86–92.  ↑_{62, 90}
- [16] Winkler J. K., Fink C., Toberer F., Enk A., Deinlein T., Hofmann-Wellenhof R., Thomas L., Lallas A., Blum A., Stolz W., Haenssle H. A. *Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition* // *JAMA Dermatology*.– 2019.– Vol. **155**.– No. 10.– Pp. 1135–1141.  ↑_{62, 90, 92}
- [17] Tschandl P., Rinner C., Apalla Z., Argenziano G., Codella N., Halpern A., Janda M., Lallas A., Longo C., Malvey J., Paoli J., Puig S., Rosendahl C., Soyer H. P., Zalaudek I., Kittler H. *Human-computer collaboration for skin cancer recognition* // *Nature Medicine*.– 2020.– Vol. **26**.– No. 8.– Pp. 1229–1234.  ↑_{70, 90}
- [18] Козачок Е. С. *Набор дерматоскопических изображений для ранней диагностики злокачественных новообразований кожи* // *Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение*.– 2025.– Т. **15**.– № 3.– С. 93–111.  ↑_{92, 93}
- [19] Козачок Е. С., Серегин С. С., Козачок А. В., Елецкий К. В., Самоваров О. И. *Методика скринингового обследования для ранней дифференциальной диагностики новообразований кожи с использованием мобильной дерматоскопии* // *Врач и информационные технологии*.– 2025.– № 3.– С. 50–64.  ↑₉₃
- [20] Козачок Е. С., Серегин С. С., Козачок А. В., Елецкий К. В., Самоваров О. И. *Интеллектуальная система поддержки принятия решений врача в диагностике новообразований кожи на основе анализа дерматоскопических изображений* // *Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение*.– 2025.– Т. **15**.– № 3.– С. 50–65.  ↑_{80, 93}

- [21] Козачок Е. С., Серегин С. С. *Интеллектуальная система поддержки принятия решений врача в диагностике новообразований кожи на основе мобильной дерматоскопии* // Российский журнал телемедицины и электронного здравоохранения. – 2025. – Т. 11. – № 3. – С. 38–44.  [↑80, 93](#)
- [22] Kozachok E. S., Seregin S. S. *Clinical validation of the melanoscope AI mobile dermoscopy clinical decision support system.* – 2026. – 24 pp. [arXiv](#)  [2605.27561](#) [↑90](#)
- [23] Kozachok E. S., Seregin S. S., Kozachok A. V., Latyshev I. P., Samovarov O. I. *Cascade Classification of Dermoscopic Images of Skin Neoplasms with Controllable Sensitivity and External Clinical Validation.* – 2026. – 28 pp. [arXiv](#)  [2606.13135](#) [↑94](#)
- [24] Kozachok A. V., Latyshev I. P., Karpulevich E. A., Kozachok E. S., Ushakov E. V., Samovarov O. I. *Searching for robust augmentations to improve out-of-domain generalization in dermoscopic skin cancer classification.* – 2026. – 25 pp. [arXiv](#)  [2607.26765](#) [↑91](#)
- [25] *Colba: система поддержки создания наборов данных.* – Институт системного программирования им В.П. Иванникова РАН. – 2 с.  [↑73](#)
- [26] Козачок А. В., Латышев И. П., Козачок Е. С., Хомичук М. В., Самоваров О. И. *MELANOSCOPE AI – программный комплекс мобильной дерматоскопии для выявления злокачественных новообразований кожи*, Свидетельство о государственной регистрации программы для ЭВМ, номер свидетельства: RU 2026664695, номер заявки: 2026664247Д. – Институт системного программирования им В.П. Иванникова РАН. – 2026.  [↑73](#)

Поступила в редакцию	26.05.2026;
одобрена после рецензирования	11.08.2026;
принята к публикации	11.08.2026;
опубликована онлайн	11.09.2026.

Информация об авторе:



Елена Сергеевна Козачок

специалист ФГБУН «Институт системного программирования им. В. П. Иванникова РАН» (ИСП РАН). Научные интересы: медицинская информатика, машинное обучение в анализе медицинских изображений, дерматоскопия, формирование клинических наборов данных для задач компьютерного зрения.



0009-0007-9432-1663

e-mail: e.kozachok@ispras.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*



Methodology for Creating a Clinically Verified Dermoscopic Image Dataset

Elena Sergeevna **Kozachok**

Ivannikov Institute for System Programming of RAS, Moscow, Russia

e.kozachok@ispras.ru

Abstract. The performance of automated diagnostic support systems depends not only on the volume of images, but also on the reproducibility of the image acquisition procedure, the completeness of structured metadata, and the reliability of diagnostic labels. Unfortunately, international collections were primarily created under conditions that differ substantially from routine Russian outpatient practice and mobile dermatoscopy. Therefore, the training base is insufficient and an independent (external) evaluation set is required to test models trained on open collections.

This study presents a methodology for constructing a clinically verified dataset of dermatoscopic images for medical informatics research, together with its software implementation. The proposed methodology integrates three interconnected components:

- (1) a standard operating procedure (SOP) for acquiring images via mobile dermatoscopy,
- (2) an information model comprising 16 structured metadata fields organized into six clinically oriented blocks in ISIC-compatible notation,
- (3) a multi-stage expert verification of diagnostic labels (initial clinical annotation, consensus review by three specialists, and histological confirmation of all malignant neoplasms);

the software implementation couples a mobile client application, a relational store, the Colba collaborative annotation environment, and a de-identified export module.

Using this methodology, a dataset of 1 026 unique dermatoscopic images from 443 patients was collected between June 2025 and May 2026. At acquisition, 32 technically unusable frames were rejected under the SOP criteria and never entered processing; of the 1 044 records that did, 18 duplicates were excluded. The dataset includes nine nosological categories; all 39 malignant lesions (18 melanomas, 15 basal cell carcinomas, and 6 squamous cell carcinomas) were histologically verified. Patient age ranged from 2 to 90 years (median 38), with 279 females (63 %) and 164 males (37 %).

To support work under pronounced class imbalance, four consistent levels of diagnostic-category aggregation are defined (from binary to nine-class), and the co-occurrence of nosological categories within a patient record is analysed: 107 of 443 patients present two or more distinct categories. Each image is accompanied by expert-annotated dermatoscopic structures and an explicit `verification_stage` field indicating the level of diagnostic confirmation.

The presented formalization of a reproducible approach to preparing clinical dermoscopy datasets, along with the description of the supporting software tools, is applicable to the creation of resources for the external validation of clinical decision support algorithms in dermato-oncology. (*In Russian*).

Key words and phrases: dermatoscopy, dataset, standard operating procedure, metadata information model, expert verification, histopathological verification, mobile dermoscopy, medical informatics, skin lesions

2020 *Mathematics Subject Classification:* 68T07; 68T45, 92C50

For citation: Elena S. Kozachok. *Methodology for Creating a Clinically Verified Dermoscopic Image Dataset*. Program Systems: Theory and Applications, 2026, **17**:3(72), pp. 57–104. (*In Russ.*). https://psta.psiras.ru/read/psta2026_3_57-104.pdf

References

- [1] P. Tschandl, C. Rosendahl, H. Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”, *Scientific Data*, **5** (2018), id. 180161, 9 pp. 
- [2] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern. “Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging”, *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC, USA, 04–07 April 2018), IEEE, 2018, ISBN 978-1-5386-3636-7, pp. 168–172. 
- [3] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N., Codella Combalia M., Dusza S., Guitera P., Gutman D., Halpern A., Helba B., Kittler H., Kose K., Langer S., Lioprys K., Malvey J., Musthaq S., Nanda J., Reiter O., Shih G., Stratigos A., Tschandl Ph., Weber J., Soyer H. P.. “A patient-centric dataset of images and metadata for identifying melanomas using clinical context”, *Scientific Data*, **8** (2021), id. 34, 8 pp. 
- [4] C. Hernandez-Perez, M. Combalia, S. Podlipnik, N. C. F. Codella, V. Rotemberg, A. C. Halpern, O. Reiter, C. Carrera, A. Barreiro, B. Helba, S. Puig, V. Vilaplana, J. Malvey. “BCN20000: Dermoscopic lesions in the wild”, *Scientific Data*, **11** (2024), id. 641, 9 pp. 
- [5] A. G. C. Pacheco, G. R. Lima, A. S. Salomao, B. Krohling, I. P. Biral, Angelo G. G. de, F. C. R. Jr Alves, J. G. M. Esgario, A. C. Simora, P. B. C. Castro, F. B. Rodrigues, P. H. L., Frasson Krohling R. A., Knidel H., Santos M. C. S., do Espírito Santo R. B., Marcelo T. L. S. G., Canuto T. R. P., de Barros L. F. S.. “PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones”, *Data in Brief*, **32** (2020), id. 106221, 10 pp. 
- [6] R. Daneshjou, K. Vodrahalli, R. A. Novoa, M. Jenkins, W. Liang, V., Rotemberg Ko J., Swetter S. M., Bailey E. E., Gevaert O., Mukherjee P., Phung M., Yekrang K., Fong B., Sahasrabudhe R., Allerup J. A. C., Okata-Karigane U., Zou J., Chiou A. S.. “Disparities in dermatology AI performance on a diverse, curated clinical image set”, *Science Advances*, **8**:31 (2022), id. eabq6147. 
- [7] M. Groh, C. Harris, L. R. Soenksen, F. Lau, R. Han, A. Kim, A. Koochek, O. Badri. “Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset”, *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Nashville, TN, USA, 19–25 June 2021), IEEE, 2021, ISBN 978-1-6654-4900-7, pp. 1820–1828. 
- [8] N. M. Kinyanjui, T. Odonga, C. Cintas, N. C. F. Codella, R. Panda, P. Sattigeri, K. R. Varshney. *Estimating skin tone and effects on classification performance in dermatology datasets*, 2019, 11 pp. arXiv: 1910.13268 [cs.CV]

- [9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun. “Dermatologist-level classification of skin cancer with deep neural networks”, *Nature*, **542** (2017), pp. 115–118. [doi](#)
- [10] D. Wen, S. M. Khan, A. J. Xu, H. Ibrahim, L. Smith, J. Caballero, L. Zepeda, Blas Perez C. de, A. K. Denniston, X. Liu, R. N. Martin. “Characteristics of publicly available skin cancer image datasets: A systematic review”, *The Lancet Digital Health*, **4**:1 (2022), pp. e64–e74. [doi](#)
- [11] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, Kalle C. von, S. Fröhling, B. Schilling, J. S. Utikal. “Deep neural networks are superior to dermatologists in melanoma image classification”, *European Journal of Cancer*, **119** (2019), pp. 11–17. [doi](#)
- [12] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, J. Rozeira. “PH²– A dermoscopic image database for research and benchmarking”, *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka, Japan, 03–07 July 2013), IEEE, 2013, ISBN 978-1-4577-0216-7, pp. 5437–5440. [doi](#)
- [13] M. Combalia, N. Codella, V. Rotemberg, C. Carrera, S. Dusza, D. Gutman, B. Helba, H. Kittler, N. R. Kurtansky, K. Liopyris, M. A., Marchetti Podlipnik S., Puig S., Rinner Ch., Tschandl Ph., Weber J., , Halpern A., Malvey J.. “Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: The 2019 International Skin Imaging Collaboration grand challenge”, *The Lancet Digital Health*, **4**:5 (2022), pp. e330–e339. [doi](#)
- [14] R. Daneshjou, M. P. Smith, M. D. Sun, V. Rotemberg, J. Zou. “Lack of transparency and potential bias in artificial intelligence data sets and algorithms: A scoping review”, *JAMA Dermatology*, **157**:11 (2021), pp. 1362–1369. [doi](#)
- [15] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. III Daume, K. Crawford. “Datasheets for datasets”, *Communications of the ACM*, **64**:12 (2021), pp. 86–92. [doi](#)
- [16] J. K. Winkler, C. Fink, F. Toberer, A. Enk, T. Deinlein, R. Hofmann-Wellenhof, L. Thomas, A. Lallas, A. Blum, W. Stolz, H. A. Haenssle. “Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition”, *JAMA Dermatology*, **155**:10 (2019), pp. 1135–1141. [doi](#)
- [17] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvey, J. Paoli, S. Puig, C. Rosendahl, H. P. Soyer, I. Zalaudek, H. Kittler. “Human-computer collaboration for skin cancer recognition”, *Nature Medicine*, **26**:8 (2020), pp. 1229–1234. [doi](#)
- [18] E. S. Kozachok. “A dermoscopic image dataset with high-quality annotation of clinically significant features for diagnosing melanocytic neoplasms”, *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel’naya tekhnika, informatika. Medicinskoe priborostroyeniye*, **15**:3 (2025), pp. 93–111 (In Russian). [doi](#)

- [19] E. S. Kozachok, S. S. Seregin, A. V. Kozachok, K. V. Eleckij, O. I. Samovarov. “A screening examination methodology for early differential diagnosis of skin neoplasms using mobile dermoscopy”, *Vrach i informacionnye tehnologii*, 2025, no. 3, pp. 50–64 (In Russian). 
- [20] E. S. Kozachok, S. S. Seregin, A. V. Kozachok, K. V. Eleckij, O. I. Samovarov. “Intelligent clinical decision support system for skin lesion diagnosis based on dermoscopic image analysis”, *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naya tehnika, informatika. Medicinskoe priborostroenie*, **15**:3 (2025), pp. 50–65 (In Russian). 
- [21] E. S. Kozachok, S. S. Seregin. “Intelligent clinical decision support system for skin lesion diagnosis based on mobile dermoscopy”, *Rossiiskij zhurnal telemeditsiny i elektronnoho zdravooxraneniya*, **11**:3 (2025), pp. 38–44 (In Russian). 
- [22] E. S. Kozachok, S. S. Seregin. *Clinical validation of the melanoscope AI mobile dermoscopy clinical decision support system*, 2026, 24 pp. [arXiv:2605.27561](#) 
- [23] E. S. Kozachok, S. S. Seregin, A. V. Kozachok, I. P. Latyshev, O. I. Samovarov. *Cascade Classification of Dermoscopic Images of Skin Neoplasms with Controllable Sensitivity and External Clinical Validation*, 2026, 28 pp. [arXiv:2606.13135](#) 
- [24] A. V. Kozachok, I. P. Latyshev, E. A. Karpulevich, E. S. Kozachok, E. V. Ushakov, O. I. Samovarov. *Searching for robust augmentations to improve out-of-domain generalization in dermoscopic skin cancer classification*, 2026, 25 pp. [arXiv:2607.26765](#) 
- [25] *Colba: sistema podderzhki sozdaniya naborov dannyx*, Institut sistemnogo programmirovaniya im V.P. Ivannikova RAN, 2 pp. (In Russian). 
- [26] A. V. Kozachok, I. P. Latyshev, E. S. Kozachok, M. V. Xomichuk, O. I. Samovarov. *MELANOSCOPE AI — programmnyj kompleks mobil'noj dermatoskopii dlya vyvavleniya zlokachestvennyx novoobrazovaniy kozhi*, Svidetel'stvo o gosudarstvennoj registracii programmy dlya EVM, nomer svidetel'stva: RU 2026664695, nomer zayavki: 2026664247D, Institut sistemnogo programmirovaniya im V.P. Ivannikova RAN, 2026 (In Russian).