

doi: 10.17586/2226-1494-2025-25-5-844-855

УДК 004.05:616

Универсальная модель архитектуры краудсорсинговой системы разметки и подготовки медицинских данных

Лев Алексеевич Коваленко¹✉, Иван Станиславович Блеканов², Федор Валерьевич Ежов³,
Евгений Сергеевич Ларин⁴, Глеб Ирламович Ким⁵

^{1,2,3,4,5} Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

¹ Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

⁵ Санкт-Петербургский государственный университет, Клиника высоких медицинских технологий
им. Н.И. Пирогова, Санкт-Петербург, 190020, Российская Федерация

¹ levozavr1@gmail.com✉, <https://orcid.org/0009-0007-8233-4387>

² i.blekanov@spbu.ru, <https://orcid.org/0000-0002-7305-1429>

³ st056053@student.spbu.ru, <https://orcid.org/0009-0007-1468-0042>

⁴ evgeny.larin@spbu.ru, <https://orcid.org/0009-0007-6199-3607>

⁵ gikim.cor@gmail.com, <https://orcid.org/0000-0002-9344-5724>

Аннотация

Введение. Методы машинного обучения и искусственного интеллекта в целом все больше используются для обработки и интеллектуального анализа медицинских данных. Для применения этих методов требуется наличие специализированных наборов размеченных данных больших размеров. Организация процесса качественной разметки медицинских данных требует привлечения большого числа ассессоров и экспертов в конкретной области медицины, а также наличие специализированных инструментов, упрощающих данный процесс и учитывающих специфику обработки медицинских данных. **Метод.** В работе предложена универсальная архитектурная модель краудсорсинговой системы, специализированной для разметки медицинских данных. Модель поддерживает обработку различных медицинских форматов данных, имеет механизмы анонимизации и многоуровневого контроля качества, позволяет организовать распределенный процесс разметки с привлечением экспертного сообщества. **Основные результаты.** Приведена классификация актуальных проблем процесса сбора и разметки медицинских данных, сформулированы критерии качества и безопасности для сравнительного анализа систем разметки медицинских данных. Предложена схема обобщенного сценария взаимодействия групп пользователей с краудсорсинговой системой в контексте решения задач искусственного интеллекта в области медицины. Построена универсальная модель архитектуры такой системы. На ее основе реализована специализированная краудсорсинговая система разметки медицинских данных на базе Computer Vision Annotation Tool. Проведено тестирование и апробация реализованной системы кардиохирургами Клиники высоких медицинских технологий им. Н.И. Пирогова Санкт-Петербургского государственного университета. **Обсуждение.** Предложенная модель архитектуры краудсорсинговой системы может быть использована для повышения эффективности и безопасности организации и построения процесса разметки медицинских данных пациентов при решении различных прикладных задач машинного обучения/искусственного интеллекта, таких как семантическая сегментация внутренних органов и их патологий, детекция и классификация заболеваний по медицинским снимкам (например, компьютерной томографии). Разработанное решение может использоваться врачами различной специализации, исследователями и разработчиками, направленными на развитие и создание методов и технологий искусственного интеллекта в области медицины.

Ключевые слова

краудсорсинг, разметка медицинских данных, модель программной архитектуры, критерии качества систем краудсорсинга, обработка медицинских данных, пользовательский сценарий взаимодействия

Ссылка для цитирования: Коваленко Л.А., Блеканов И.С., Ежов Ф.В., Ларин Е.С., Ким Г.И. Универсальная модель архитектуры краудсорсинговой системы разметки и подготовки медицинских данных // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 5. С. 844–855. doi: 10.17586/2226-1494-2025-25-5-844-855

A universal architecture model of a crowdsourcing medical data labeling system designed

Lev A. Kovalenko¹✉, Ivan S. Blekanov², Fedor V. Ezhov³, Evgenii S. Larin⁴, Gleb I. Kim⁵

^{1,2,3,4,5} St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation

¹ ITMO University, Saint Petersburg, 197101, Russian Federation

⁵ Saint Petersburg State University Hospital, Saint Petersburg, 190020, Russian Federation

¹ levozavr1@gmail.com✉, <https://orcid.org/0009-0007-8233-4387>

² i.blekanov@spbu.ru, <https://orcid.org/0000-0002-7305-1429>

³ st056053@student.spbu.ru, <https://orcid.org/0009-0007-1468-0042>

⁴ evgeny.larin@spbu.ru, <https://orcid.org/0009-0007-6199-3607>

⁵ gikim.cor@gmail.com, <https://orcid.org/0000-0002-9344-5724>

Abstract

Machine Learning (ML) and Artificial Intelligence (AI) methods are used to process and intelligently analyze medical data. The application of ML/AI methods requires specialized sets of labeled medical data of large dimensions. Process organization of quality medical data labeling requires the involvement of a large number assessors and specialists in a particular field of medicine as well as the availability of specialized tools for labeling process optimization considering the specifics of medical data processing. In this paper a universal architectural model of a crowdsourcing system specifically designed for medical data labeling was proposed. The model supports processing of diverse medical data formats, incorporates data anonymization mechanisms and multi-level quality control, while enabling a distributed annotation process with expert community involvement. As a result, classification of actual problems of the process of medical data labeling and data collection, and a quality and safety criteria for comparative analysis of medical data labeling systems was detected and formulated. The scheme of generalized scenario of users' groups interaction with crowdsourcing system in the context of solving AI problems in the field of medicine was proposed. A universal model of such system architecture was designed and a specialized crowdsourcing system of medical data labeling based on Computer Vision Annotation Tool was implemented on its basis. Testing and approbation of the realized system was carried out at the Pirogov Clinic of High Medical Technologies. The proposed universal model of crowdsourcing system architecture can be used to improve the efficiency and safety of organization and construction of the process of labeling patients' medical data in the context of solving various applied ML/AI tasks, such as semantic segmentation of internal organs and their pathologies, detection and classification of diseases based on medical images (e.g. computed tomography scans). The developed solution can be used by doctors of various specializations, researchers and developers aimed at the development and creation of methods and technologies of AI in the field of medicine.

Keywords

crowdsourcing, medical data annotation, software architecture model, quality criteria for crowdsourcing systems, medical data processing, use case

For citation: Kovalenko L.A., Blekanov I.S., Ezhov F.V., Larin E.S., Kim G.I. A universal architecture model of a crowdsourcing medical data labeling system designed. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 5, pp. 844–855 (in Russian). doi: 10.17586/2226-1494-2025-25-5-844-855

Введение

Сегодня искусственный интеллект (Artificial Intelligence, AI) и машинное обучение (Machine Learning, ML) широко применяются в здравоохранении для диагностики, прогнозирования и персонализированной терапии [1–4]. Эффективность ML/AI-моделей критически зависит от наличия больших объемов качественно размеченных данных, требующих участия экспертов [4–7]. Однако традиционные методы разметки не справляются с растущими объемами данных, что обуславливает необходимость специализированных краудсорсинговых платформ [8–11].

Подготовка медицинских наборов данных связана с рядом уникальных сложностей, обусловленных спецификой медицинских данных. В отличие от стандартных форматов (текст, изображения, видео), медицинские данные (Digital Imaging and Communications in Medicine (DICOM), Neuroimaging Informatics Technology Initiative (NifTI) и другие форматы) требуют специальной обработки перед разметкой, а отсутствие единых стандартов их представления дополнительно усложняет этот процесс [12–15]. Особую проблему представляет обработка конфиденциальных медицинских сведений,

требующая строгого соблюдения нормативных требований (ФЗ-152¹, ФЗ-323², GDPR³, HIPAA⁴), включая обязательную анонимизацию и защиту от утечек, что существенно ограничивает возможности применения краудсорсинга [15, 16]. Дополнительные сложности возникают при контроле качества разметки, поскольку различия в квалификации врачей-экспертов, ассессоров и неоднозначность медицинской информации повышают риск ошибок и требуют разработки специальных механизмов валидации [17–21].

Настоящая работа посвящена решению перечисленных проблем посредством проектирования и разработки системы краудсорсинга, адаптированной под специфику медицинских данных и взаимодей-

¹ Федеральный закон от 27.07.2006 №152-ФЗ «О персональных данных» (ред. от 24.02.2021).

² Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» (ред. от 01.07.2021).

³ Regulation (EU) 2016/679 (General Data Protection Regulation) от 27.04.2016.

⁴ Public Law 104-191 (Health Insurance Portability and Accountability Act) от 21.08.1996.

ствие с врачами-экспертами в удаленном формате. Рассматривается методология разметки медицинских данных врачами-экспертами, основанная на современных методах и инструментах краудсорсинга.

Актуальные проблемы краудсорсинга в медицине

Краудсорсинговая разметка медицинских данных — ключевой инструмент для создания наборов данных, необходимых в машинном обучении, однако его внедрение в процесс разработки AI-решений в медицине сталкивается с рядом проблем. Прежде всего, обработка медицинской информации требует строгого соблюдения норм конфиденциальности (ФЗ-152¹, ФЗ-323², GDPR³, HIPAA⁴) и применения надежных методов анонимизации для защиты персональных медицинских данных пациентов [15, 16, 22–24].

Еще одной важной проблемой является обеспечение качества разметки медицинских данных. Различия в квалификации ассессоров и субъективность интерпретации медицинских терминов и концепций могут приводить к ошибкам, что негативно влияет на надежность данных [17–19]. Для минимизации этих рисков предлагаются такие подходы, как обучение ассессоров [25–27], кросс-аннотация с привлечением врачей-экспертов [28, 29], а также автоматизация валидации с использованием специализированных метрик и методов машинного обучения [18, 30–32].

Сложности также связаны с разнообразием форматов медицинских данных (DICOM, NIfTI и др.) и отсутствием единых стандартов их представления [12–14, 33]. Это требует разработки инструментов для преобразования

медицинских данных в удобный для разметки вид без потери метаданных, что особенно важно для обеспечения согласованности наборов данных [34, 35].

Большинство существующих платформ разметки не адаптированы для работы с медицинскими данными. Они часто не поддерживают специализированные форматы, не обеспечивают должного уровня безопасности и не предназначены для использования в закрытых контурах медицинских учреждений [13, 14, 36]. Эти ограничения делают актуальной разработку специализированных решений, учитывающих особенности медицинской отрасли.

В настоящей работе предлагаются классификация проблем и актуальные направления развития краудсорсинговых систем разметки медицинских данных. В табл. 1 выделено три класса проблем: обеспечение безопасности персональных данных, качества данных и качества получаемой разметки.

Учитывая обозначение в табл. 1 проблемы использования инструментов краудсорсинга для разметки медицинских данных, можно составить перечень критериев для оценки и сравнения существующих инструментов и платформ (табл. 2).

Существующие решения

Современные инструменты для разметки медицинских данных существенно различаются по функциональности — от простых аннотаторов изображений до комплексных систем с интеграцией машинного обучения и AI-технологий [37, 38]. По специализации их можно разделить на три категории: специализированные медицинские решения, универсальные платформы

Таблица 1. Классификация актуальных проблем краудсорсинговой разметки медицинских данных

Table 1. Classification of current challenges in crowdsourced medical data annotation

Класс проблем	Название проблемы	Описание проблемы
Безопасность данных пациентов	Соблюдение правовых норм и законов	При обработке персональных медицинских данных необходимо соблюдать законы, например ФЗ-152 ¹ и ФЗ-323 ²
	Анонимизация данных	Не анонимизированные медицинские данные содержат персональную информацию
	Утечка данных	Риски утечки персональных данных пациентов при передаче данных по сети
Обеспечение качества данных	Разнообразие форматов	Существует большое количество специфичных форматов медицинских данных
	Отсутствие единой стандартизации	Различные медучреждения используют свои стандарты и оборудование
Обеспечение качества разметки	Разный уровень квалификации	Разметка ассессоров с разным уровнем квалификации может различаться
	Потребность во врачах-экспертах	Оценка качества полученной разметки требует привлечения врачей-экспертов
	Отсутствие методов валидации разметки	Из-за сложных форматов данных и задач разметки медицинских данных не применимы общеизвестные способы валидации разметки

¹ Федеральный закон от 27.07.2006 №152-ФЗ «О персональных данных» (ред. от 24.02.2021).

² Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» (ред. от 01.07.2021).

Таблица 2. Критерии качества существующих решений с учетом класса проблем

Table 2. Quality assessment criteria for existing solutions by problem category

Критерий качества	Класс проблем	Описание
Соответствие стандартам защиты медицинских данных	Безопасность данных пациентов	Наличие инструментов анонимизации и защиты персональных данных
Режим локального запуска		Возможность запуска платформы разметки в закрытом контуре
Импортозамещенность		Доступность системы в Российской Федерации и соответствие законам
Поддержка разных типов данных для разметки	Обеспечение качества данных	Возможность загружать для разметки данные в специализированных медицинских форматах
Наличие Application Programming Interface		Возможность автоматизировать поставку и предварительную подготовку данных
Удобство и гибкость интерфейса		Возможность настроить способы представления данных для разметки
Поддержка типов аннотаций	Обеспечение качества разметки	Поддержка аннотации текстов, изображений и видео для различных задач
Система автоматической разметки		Возможность подключения машинного обучения для проведения автоматической разметки данных
Управление процессом разметки		Функционал для отслеживания прогресса и статуса задач разметки данных
Ролевая модель		Встроенная ролевая модель для распределения прав пользователей
Система контроля качества разметки		Возможность вводить метрики или инструменты для контроля качества

для разметки стандартных форматов данных и гибридные системы. Вариативность наблюдается и в способах развертывания: open-source решения, коробочные продукты и облачные сервисы. Характеристики ключевых современных платформ представлены в табл. 3, где отражены их тип, направленность и основные функциональные возможности.

Для выбранного набора решений разметки медицинских данных можно провести сравнительный анализ, согласно критериям оценки в табл. 2. Результаты такого анализа приведены в табл. 4.

Выполненный анализ выявил ключевые недостатки рассмотренных решений: отсутствие механизмов анонимизации медицинских данных, слабую поддержку управления процессом разметки (включая ролевые модели и контроль качества), а также ограниченную доступность на территории Российской Федерации. Наиболее перспективными оказались решения CVAT и LabelStudio, которые, несмотря на отсутствие поддержки специализированных медицинских форматов (например, DICOM), в наибольшей степени соответствуют предъявляемым требованиям. На основе сравнитель-

Таблица 3. Современные решения для разметки медицинских данных

Table 3. Existing medical data labeling solutions

Решение	Тип решения	Направленность	Описание	Ссылка
3D Slicer	Коробочное	Медицинские данные	Позволяет сегментировать медицинские снимки	slicer.org
Supervisely	Облачное	Общее решение с поддержкой медицинских данных	Платная платформа разметки изображений, в том числе медицинских	supervisely.com
LabsV7			Платная система разметки данных с поддержкой медицинских снимков	v7labs.com
Encord			Платная платформа разметки с поддержкой DICOM и NifTI снимков	encord.com
AmazonMTurk		Медицинские данные	Платная платформа для создания аннотированных наборов медицинских данных	mturk.com
LabelStudio	Open-source	Общее решение	Открытое решение разметки общеизвестных форматов	labelstud.io
Computer Vision Annotation Tool (CVAT)			Открытое решение разметки общеизвестных форматов	cvat.ai

Таблица 4. Оценка существующих решений согласно критериям (табл. 2)
Table 4. Comparative analysis of existing solutions based on Table 2 criteria

Название критерия	Решения						
	3D Slicer	Super visely	Labs V7	Encord	Label Studio	CVAT	Amazon MTurk
Соответствие стандартам защиты медицинских данных	–	–	–	–	–	–	–
Локальный режим работы	+	–	–	–	+	+	–
Импортозамещенность	+	–	–	–	+	+	–
Поддержка разных типов данных для разметки	+	+	+	+	–	–	–
Наличие Application Programming Interface	–	–	–	–	+	+	+
Удобство и гибкость интерфейса	+	+	+	+	+	+	+
Поддержка различных типов аннотаций	–	+	+	+	+	+	+
Система автоматической разметки	–	+	+	+	+	+	+
Управление процессом разметки	–	+	–	–	+	+	+
Ролевая модель	–	–	–	–	–	+	–
Система контроля качества разметки	–	–	–	–	+	+	–

Примечание. «+» означает соответствие критерию, а «–» — не соответствие. Жирным шрифтом выделены наиболее перспективные решения.

ного анализа была разработана универсальная архитектура краудсорсинговой системы с использованием CVAT в качестве базовой платформы, дополненной функционалом для работы с медицинскими данными.

Проектирование модели и сценария
пользовательского взаимодействия с системой
краудсорсинга разметки медицинских данных

Разработка ML/AI-решений представляет собой сложный итеративный процесс [39–41], где краудсорсинговая система разметки играет ключевую роль в создании аннотированных наборов данных для медицинских задач, включая семантическую сегментацию анатомических структур [42–44] и диагностику по медицинским изображениям [45–47]. Общая схема процесса представлена на рис. 1.

Краудсорсинговая система разметки может применяться на различных стадиях процесса разработки AI-решения (рис. 1): сбор и подготовка данных, интерпретация и анализ работы моделей, а также для демонстрации результатов работы ML/AI-решения и сбора обратной связи от врачей-экспертов.

На рис. 2 отображен сценарий взаимодействия ключевых пользователей с краудсорсинговой системой разметки медицинских данных в процессе разработки ML/AI-решения. К ключевым пользователям системы разметки данных можно отнести четыре основные роли: врач, врач-эксперт, ассессор и разработчик ML/AI. Обобщенный сценарий их взаимодействия с системой описан в алгоритме.

Алгоритм. Алгоритм обобщенного сценария взаимодействия различных групп пользователей с краудсорсинговой системой разметки медицинских данных состоит из следующих этапов.

Этап 1. Врач собирает и загружает медицинские данные пациентов в систему разметки.
Этап 2. Врач-эксперт создает задачи разметки на основе собранных данных и распределяет их на ассессоров.

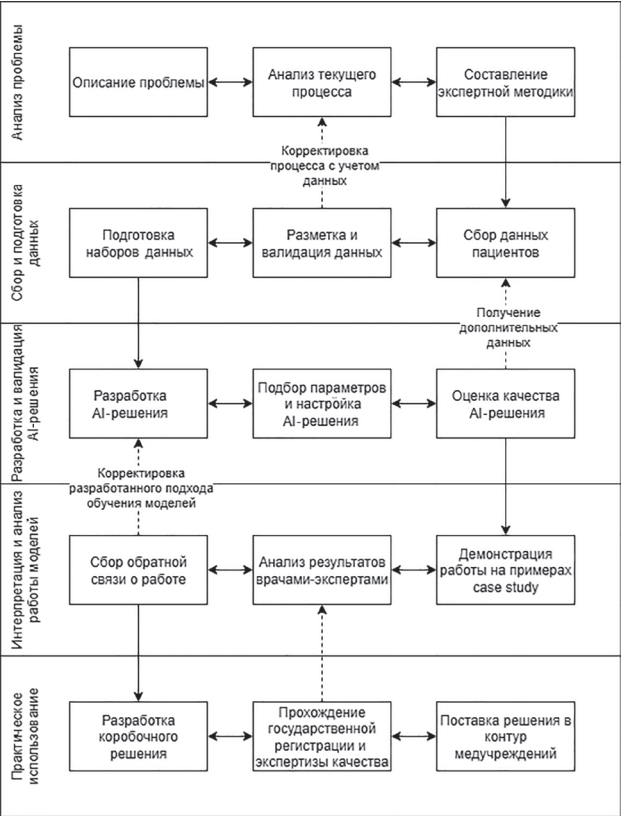


Рис. 1. Схема разработки AI-решения в области медицины
Fig. 1. AI solution development framework in medical domain

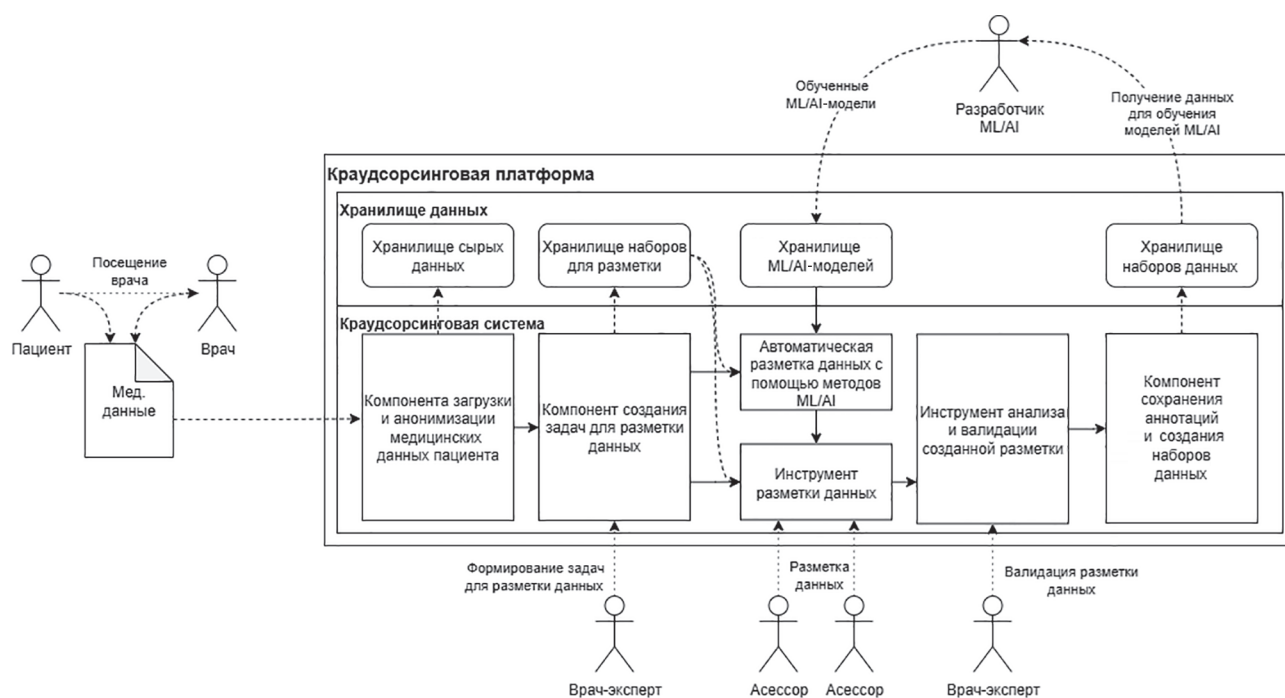


Рис. 2. Модель пользовательского взаимодействия с краудсорсинговой системой разметки медицинских данных в контексте разработки ML/AI-решения по алгоритму

Fig. 2. User interaction model with the crowdsourcing medical data annotation system within ML/AI solution development (Algorithm)

Этап 3. Ассессоры размечают данные самостоятельно или корректируют разметку ML/AI-решения в назначенных им задачах.

Этап 4. Врач-эксперт проводит валидацию разметки ассессоров и ее финальное уточнение.

Этап 5. Разработчик ML/AI получает подготовленные и размеченные наборы медицинских данных.

Этап 6. Разработчик ML/AI поставяет обученную ML/AI-модель для автоматизации разметки.

Результатом данного взаимодействия с системой является итеративное развитие и улучшение разрабатываемого ML/AI-решения. Для работы алгоритма необходимо обеспечить не только сбор данных для последующего увеличения качества и обобщающей способности ML/AI-моделей, но и получение экспертной обратной связи по качеству работы моделей при разметке данных, что позволит корректировать процесс обучения моделей для учета текущих ошибок и недостатков их работы.

Проектирование универсальной модели архитектуры краудсорсинговой системы разметки медицинских данных

Предлагаемая универсальная архитектура краудсорсинговой системы (рис. 3) интегрирует четыре взаимосвязанных компонента, обеспечивающих комплексное решение задач медицинской разметки. Основу системы составляют пользовательские сервисы на базе CVAT, предназначенные для взаимодействия с ассессорами и проведения разметки. Архитектура дополнена специализированными функциональными модулями, выполняющими пред- и постобработку данных, включая

операции анонимизации, автоматическую разметку с использованием ML-моделей и формирование итоговых наборов данных. Для хранения информации предусмотрено надежное хранилище, поддерживающее консистентность данных и сохранение связей между исходными медицинскими снимками и их PNG-представлениями. Система оркестрации гарантирует стабильность работы, горизонтальную масштабируемость и отказоустойчивость за счет централизованного управления всеми этапами обработки данных.

Сервис разметки CVAT (рис. 3) представляет собой гибкий компонент архитектуры, допускающий замену на другие системы (LabelStudio или собственные решения). Ключевые преимущества предложенной архитектуры включают: стандартизацию медицинских данных различных форматов, механизмы анонимизации и защиты информации, отказоустойчивость за счет системы оркестрации, а также поддержку распределенной обработки для повышения производительности. Реализация выполнена на Python 3.11 с использованием инструмента Celery для оркестрации процессов и системы MinIO в качестве объектного хранилища медицинских данных.

Апробация использования краудсорсинговой системы разметки медицинских данных

Экспериментальная апробация системы проводилась кардиохирургами Клиники высоких медицинских технологий им. Н.И. Пирогова Санкт-Петербургского государственного университета для разметки снимков компьютерной томографии (КТ) грудной аорты. Процесс включал:

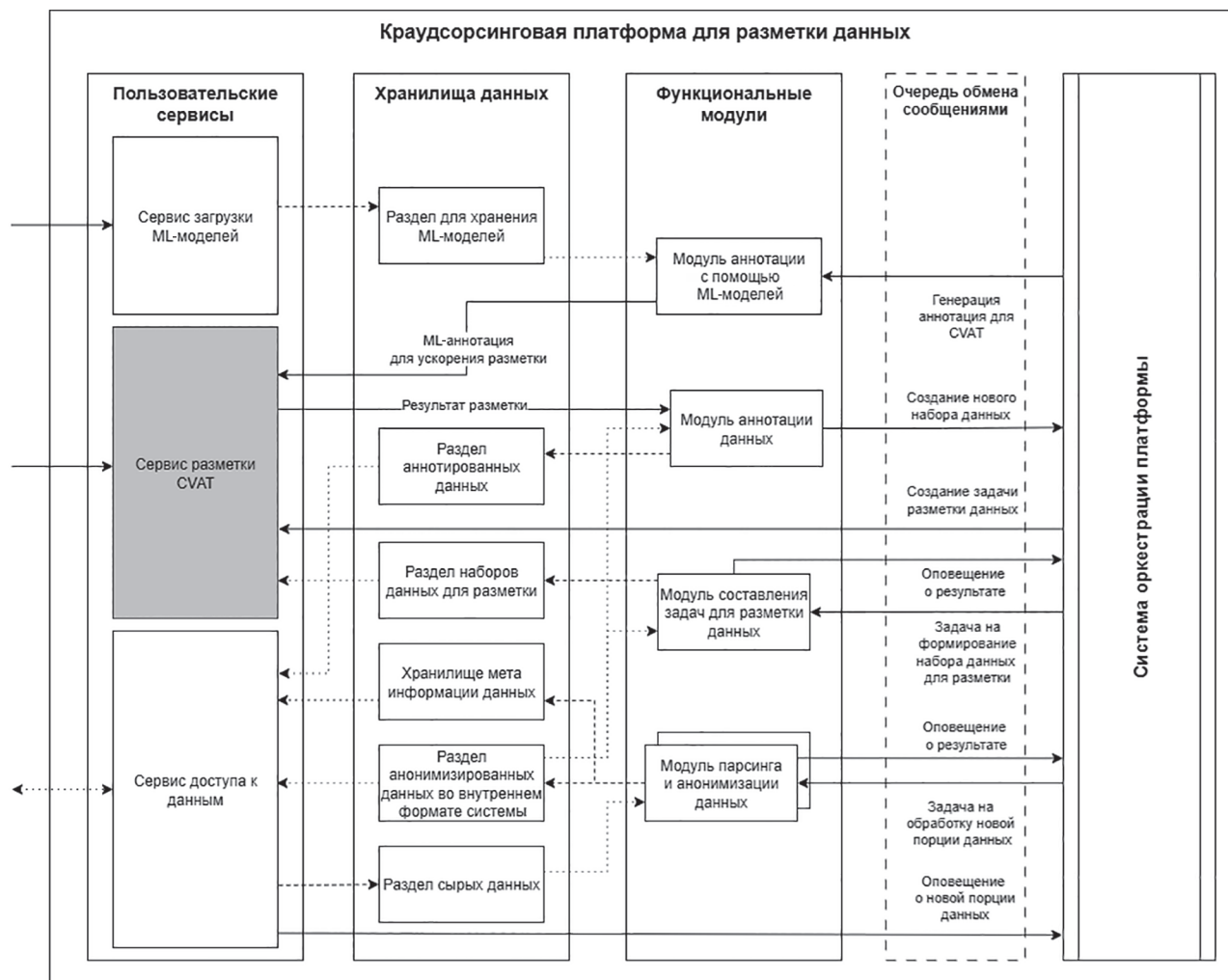


Рис. 3. Универсальная модель архитектуры краудсорсинговой системы разметки медицинских данных

Fig. 3. Universal architecture model of the crowdsourcing medical data annotation platform

— загрузку и настройку отображения DICOM-данных (рис. 4);

— анонимизацию через MD5-шифрование персональных данных;

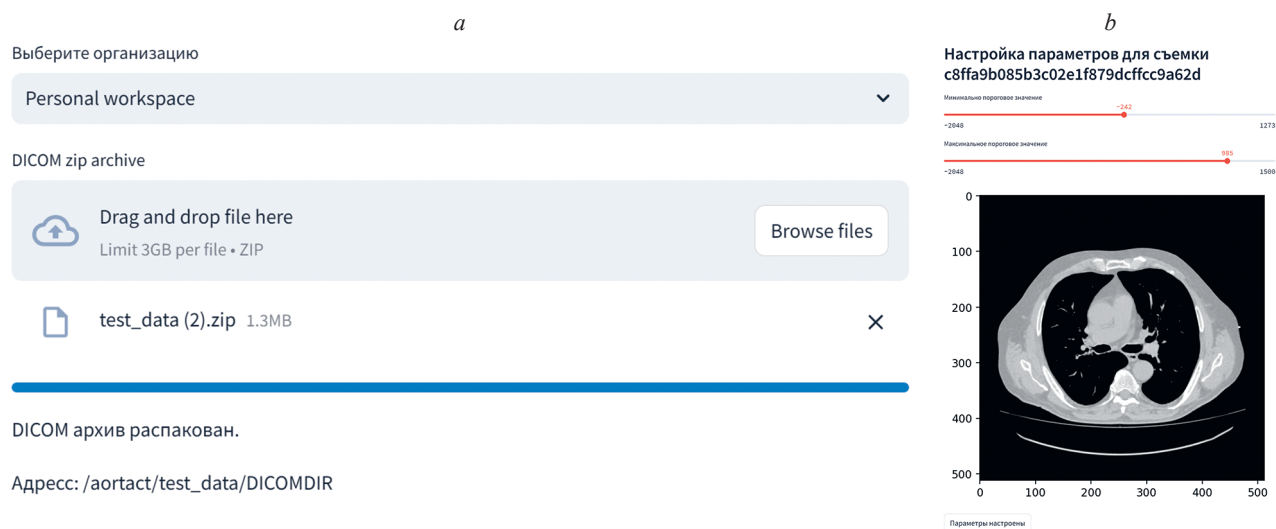


Рис. 4. Примеры интерфейсов сервиса доступа к данным: загрузки DICOM снимка (a); настройки параметров отображения КТ-изображений (b)

Fig. 4. Data access service interface example: DICOM image upload interface (a); CT image display settings interface (b)

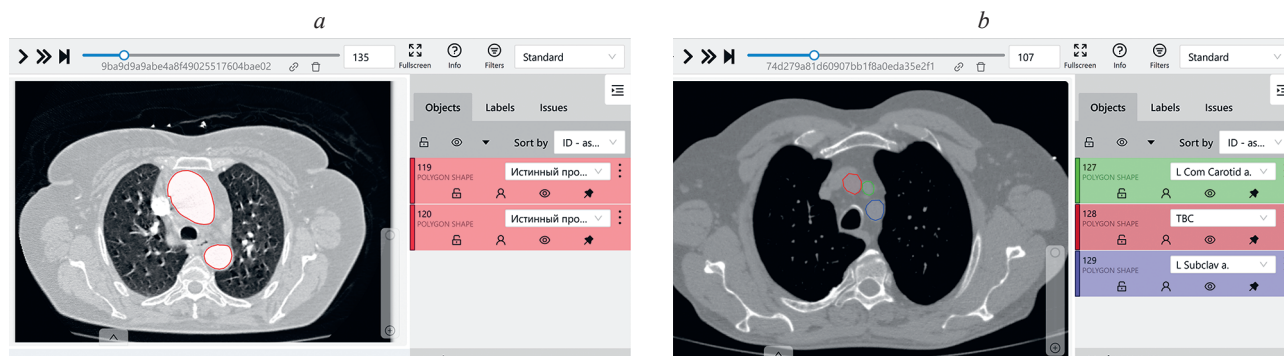


Рис. 5. Интерфейсы разметки КТ-снимков: грудной аорты (а); коронарных артерий (б)

Fig. 5. Interface for marking CT images: thoracic aorta (a), coronary arteries (b)

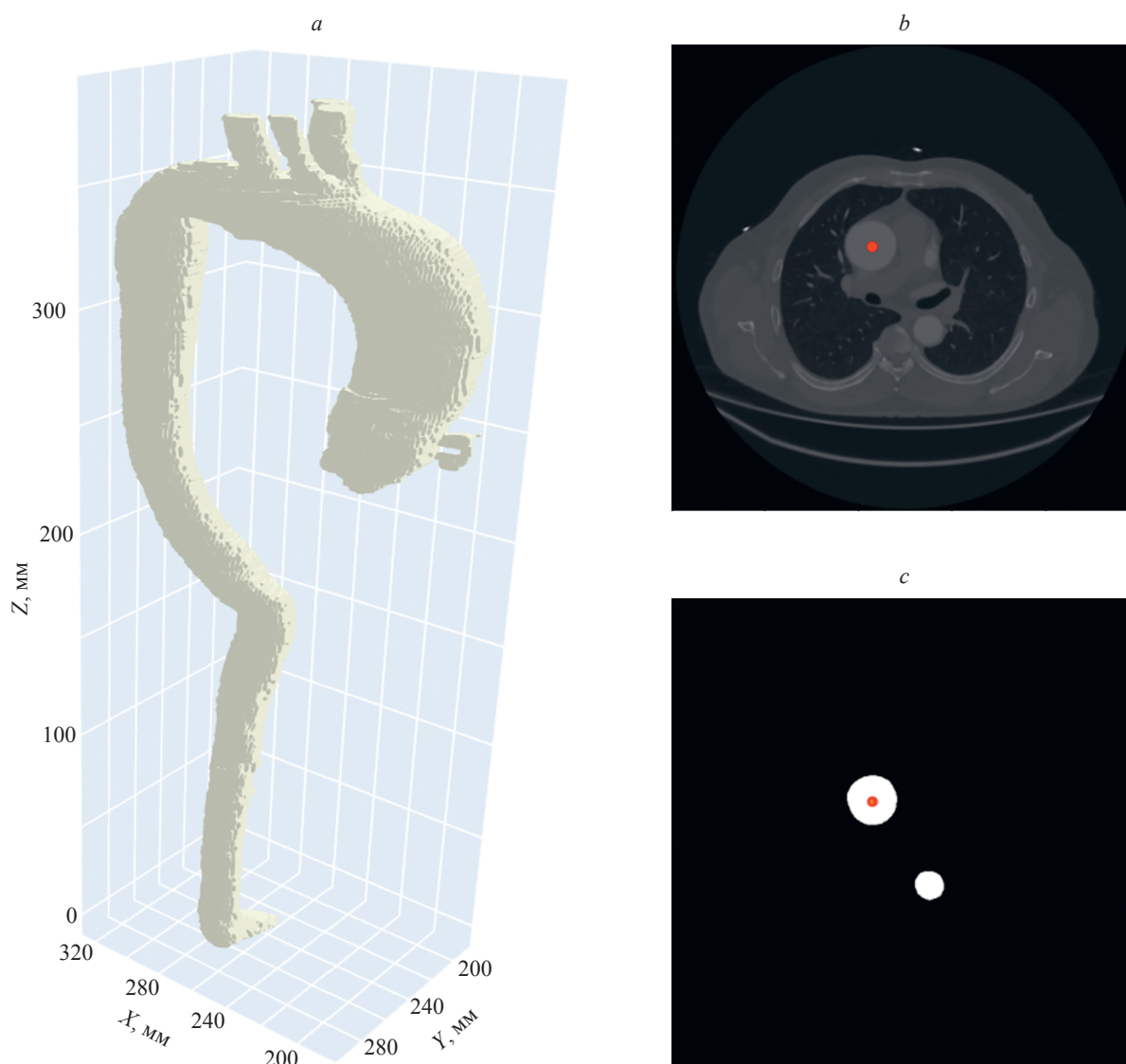


Рис. 6. Примеры размеченных данных аорты: 3D-модель аорты на основе полученной разметки (а); поперечный срез КТ-снимка аорты (б); разметка среза аорты в виде маски (с)

Fig. 6. Examples of annotated aorta data: 3D aorta model based on annotations (a); transverse CT slice of aorta (b); aorta slice annotation mask (c)

— автоматическое преобразование медицинских снимков в PNG-формат и создание задач разметки в CVAT.

В ходе апробации системы 30 ассессоров и два врача-эксперта выполнили разметку 50 КТ-снимков аорты сердца (в среднем по 500 срезов на снимок) с исполь-

Таблица 5. Временные характеристики, затраченные на обработку одной аорты при создании набора данных из 50 аорт, с
Table 5. Time metrics for single aorta processing when creating a dataset of 50 aortas, s

Операция	Минимальное время	Среднее время	Максимальное время
Анонимизация входных данных одного КТ-снимка	не более 0,01	Менее 0,01	Менее 0,01
Формирование задачи разметки аорты в CVAT	32,6	71,3	117,4
Формирование связей между исходными данными КТ-снимка и изображениями для разметки в CVAT	не более 0,01	0,02	0,04
Проведение автоматической разметки загруженных снимков	16,7	17,4	21,7
Построение набора данных на основе полученной разметки	9,0	12,3	17,2
Разметка одной аорты ассессорами вручную	7200	18 000	28 800

зованием 20 уникальных классов. Система автоматически сформировала набор данных для обучения ML-моделей, содержащий 25 655 размеченных срезов. Как показано на рис. 5, процесс разметки включал выделение областей аорты и сосудов. Результаты работы предложенной системы (рис. 6) демонстрируют возможность преобразования размеченных данных (рис. 6, *b*) в наборы масок (рис. 6, *c*) и 3D-модели (рис. 6, *a*) сердечно-сосудистых структур.

В процессе использования разработанной системы краудсорсинговой разметки медицинских данных были проведены замеры скорости автоматического построения набора данных на основе разметки пользователей в CVAT и исходных данных DICOM-изображений. Для обучения моделей требуется использовать не преобразованные к PNG-формату данные, а исходные данные медицинского снимка в более широком диапазоне, поэтому необходимо автоматически сопоставлять разметку с данными КТ-снимка, связь между которыми храниться в системе. В реализованном решении на базе универсальной модели архитектуры такие связи хранятся в формате JSON. В табл. 5 приведены замеры значений временных характеристик выполнения различных операций в процессе построения размеченного набора данных, состоящего из 50 аорт (25 655 срезов), для задач машинного и глубокого обучений.

Заключение

В работе спроектирована универсальная модель архитектуры краудсорсинговых систем для разметки медицинских данных (на примере семантической сегментации компьютерной томографии грудной аорты) и построения специализированных наборов данных для решения задач машинного и глубокого обучений в области медицины и кардиохирургии, в частности.

Предложенная универсальная модель удовлетворяет критериям для оценки эффективности и безопасности существующих решений в этой области (табл. 2), а также позволяет решать основные актуальные проблемы организации процесса разметки медицинских данных в соответствии с обобщенной в настоящей работе классификацией проблем краудсорсинговой разметки данных (табл. 1). По сравнению с существующими решениями предложенная модель ориентирована на специфику медицинских данных и может быть адаптирована к любому формату данных в зависимости от решаемой задачи по интеллектуальному анализу таких данных. По результатам сравнительного анализа современные системы разметки показали низкую адаптивность к разметке медицинских данных, в частности, выявлены проблемы с безопасностью медицинских данных пациентов, организацией процесса и контроля качества разметки.

Предложена модель взаимодействия различных групп пользователей (как отдельных врачей, так и разработчиков машинного обучения) при разработке систем искусственного интеллекта в области медицины. Описан в виде алгоритма обобщенный сценарий взаимодействия различных групп пользователей с системой краудсорсинга на основе представленной универсальной модели архитектуры.

Разработанная краудсорсинговая система на основе предложенной модели архитектуры была успешно протестирована и апробирована в задаче сбора данных для семантической сегментации аорты сердца врачами-экспертами Клиники высоких медицинских технологий им. Н.И. Пирогова Санкт-Петербургского государственного университета. Временные характеристики работы новой функциональности, замеренные в ходе тестирования системы, незначительны в сравнении со временем, затраченным на саму разметку аорты.

Литература

1. Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019. 341 p.
2. Obermeyer Z., Emanuel E.J. Predicting the future — big data, machine learning, and clinical medicine // *New England Journal of Medicine*. 2016. V. 375. N 13. P. 1216–1219. <https://doi.org/10.1056/nejmp1606181>
3. Jiang F., Jiang Y., Zhi H., Dong Y., Li H., Ma S., et al. Artificial intelligence in healthcare: past, present and future // *Stroke and*

References

1. Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019. 341 p.
2. Obermeyer Z., Emanuel E.J. Predicting the future — big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 2016, vol. 375, no. 13, pp. 1216–1219. <https://doi.org/10.1056/nejmp1606181>
3. Jiang F., Jiang Y., Zhi H., Dong Y., Li H., Ma S., et al. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular*

- Vascular Neurology. 2017. V. 2. N 4. P. 230–243. <https://doi.org/10.1136/svn-2017-000101>
4. Secinaro S., Calandra D., Secinaro A., Muthurangu V., Biancone P. The role of artificial intelligence in healthcare: a structured literature review // *BMC Medical Informatics and Decision Making*. 2021. V. 21. N 1. P. 125. <https://doi.org/10.1186/s12911-021-01488-9>
 5. Roh Y., Heo G., Whang S.E. A survey on data collection for machine learning: a big data – AI Integration perspective // *IEEE Transactions on Knowledge and Data Engineering*. 2021. V. 33. N 4. P. 1328–1347. <https://doi.org/10.1109/TKDE.2019.2946162>
 6. Апанасович К.С., Махныткина О.В., Кабаров В.И., Далеvская О.П. RuPersonaChat: корпус диалогов для персонификации разговорных агентов // *Научно-технический вестник информационных технологий, механики и оптики*. 2024. Т. 24. № 2. С. 214–221. <https://doi.org/10.17586/2226-1494-2024-24-2-214-221>
 7. Shaheen Z., Mourmtsev D.I., Postny I. RuLegalNER: a new dataset for Russian legal named entities recognition. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2023. V. 23. N 4. P. 854–857. <https://doi.org/10.17586/2226-1494-2023-23-4-854-857>
 8. Sayin B., Krivosheev E., Yang J., Passerini A., Casati F. A review and experimental analysis of active learning over crowdsourced data // *Artificial Intelligence Review*. 2021. V. 54. N 7. P. 5283–5305. <https://doi.org/10.1007/s10462-021-10021-3>
 9. Xintong G., Hongzhi W., Song Y., Hong G. Brief survey of crowdsourcing for data mining // *Expert Systems With Application*. 2014. V. 41. N 17. P. 7987–7994. <https://doi.org/10.1016/j.eswa.2014.06.044>
 10. Hecht R., Kalla M., Krüger T. Crowd-sourced data collection to support automatic classification of building footprint data // *Proc. of the ICA*. 2018. V. 1. P. 54. <https://doi.org/10.5194/ica-proc-1-54-2018>
 11. Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., et al. Human-level control through deep reinforcement learning // *Nature*. 2015. V. 518. N 7540. P. 529–533. <https://doi.org/10.1038/nature14236>
 12. Rahmani A.M., Yousefpoor E., Yousefpoor M.S., Mehmood Z., Haider A., Hosseinzadeh M., Naqvi R.A. Machine learning (ML) in medicine: review, applications, and challenges // *Mathematics*. 2021. V. 9. N 22. P. 2970. <https://doi.org/10.3390/math9222970>
 13. Wang C., Han L., Stein G., Day S., Bien-Gund C., Mathews A., et al. Crowdsourcing in health and medical research: a systematic review // *Infectious Diseases of Poverty*. 2020. V. 9. N 1. P. 8. <https://doi.org/10.1186/s40249-020-0622-9>
 14. Ellis R.J., Sander R.M., Limon A. Twelve key challenges in medical machine learning and solutions // *Intelligence-Based Medicine*. 2022. V. 6. P. 100068. <https://doi.org/10.1016/j.ibmed.2022.100068>
 15. Xia H., McKernan B. Privacy in crowdsourcing: a review of the threats and challenges // *Computer Supported Cooperative Work (CSCW)*. 2020. V. 29. N 3. P. 263–301. <https://doi.org/10.1007/s10606-020-09374-0>
 16. Rother A., Niemann U., Hielscher T., Völzke H., Ittermann T., Spiliopoulou M. Assessing the difficulty of annotating medical data in crowdworking with help of experiments // *PLOS ONE*. 2021. V. 16. N 7. P. e0254764. <https://doi.org/10.1371/journal.pone.0254764>
 17. Ye C., Coco J., Epishova A., Hajaj C., Bogardus H., Novak L., et al. A crowdsourcing framework for medical data sets // *AMIA Joint Summits on Translational Science proceedings*. 2018. P. 273–280.
 18. Kittur A., Nickerson J., Bernstein M., Gerber E., Shaw A., Zimmerman J., et al. The future of crowd work // *Proc. of the Conference on Computer Supported Cooperative Work*. 2013. P. 1301–1318. <https://doi.org/10.1145/2441776.2441923>
 19. Ørting S.N., Doyle A., van Hilten A., Hirth M., Inel O., Madan C.R., et al. A survey of crowdsourcing in medical image analysis // *Human Computation*. 2020. V. 7. N 1. P. 1–26. <https://doi.org/10.15346/hc.v7i1.1>
 20. Lu J., Li W., Wang Q., Zhang Y. Research on data quality control of crowdsourcing annotation: a survey // *Proc. of the IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)*. 2020. P. 201–208. <https://doi.org/10.1109/DASC-PiCom-CBDCCom-CyberSciTech49142.2020.00044>
 21. Lu X., Ratcliffe D., Kao T.-T., Tikhonov A., Litchfield L., Rodger C., Wang K. Rethinking quality assurance for crowdsourced multi-ROI image segmentation // *Proc. of the 11th AAAI Conference on Human Neurology*, 2017, vol. 2, no. 4, pp. 230–243. <https://doi.org/10.1136/svn-2017-000101>
 4. Secinaro S., Calandra D., Secinaro A., Muthurangu V., Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making*, 2021, vol. 21, no. 1, pp. 125. <https://doi.org/10.1186/s12911-021-01488-9>
 5. Roh Y., Heo G., Whang S.E. A survey on data collection for machine learning: a big data — AI Integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 2021, vol. 33, no. 4, pp. 1328–1347. <https://doi.org/10.1109/TKDE.2019.2946162>
 6. Apanasovich K.S., Makhnytkina O.V., Kabarov V.I., Dalevskaya O.P. RuPersonaChat: a dialog corpus for personalizing conversational agents. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 2, pp. 214–221. (in Russian). <https://doi.org/10.17586/2226-1494-2024-24-2-214-221>
 7. Shaheen Z., Mourmtsev D.I., Postny I. RuLegalNER: a new dataset for Russian legal named entities recognition. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 4, pp. 854–857. <https://doi.org/10.17586/2226-1494-2023-23-4-854-857>
 8. Sayin B., Krivosheev E., Yang J., Passerini A., Casati F. A review and experimental analysis of active learning over crowdsourced data. *Artificial Intelligence Review*, 2021, vol. 54, no. 7, pp. 5283–5305. <https://doi.org/10.1007/s10462-021-10021-3>
 9. Xintong G., Hongzhi W., Song Y., Hong G. Brief survey of crowdsourcing for data mining. *Expert Systems With Application*, 2014, vol. 41, no. 17, pp. 7987–7994. <https://doi.org/10.1016/j.eswa.2014.06.044>
 10. Hecht R., Kalla M., Krüger T. Crowd-sourced data collection to support automatic classification of building footprint data. *Proc. of the ICA*, 2018, vol. 1, pp. 54. <https://doi.org/10.5194/ica-proc-1-54-2018>
 11. Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., et al. Human-level control through deep reinforcement learning. *Nature*, 2015, vol. 518, no. 7540, pp. 529–533. <https://doi.org/10.1038/nature14236>
 12. Rahmani A.M., Yousefpoor E., Yousefpoor M.S., Mehmood Z., Haider A., Hosseinzadeh M., Naqvi R.A. Machine learning (ML) in medicine: review, applications, and challenges. *Mathematics*, 2021, vol. 9, no. 22, pp. 2970. <https://doi.org/10.3390/math9222970>
 13. Wang C., Han L., Stein G., Day S., Bien-Gund C., Mathews A., et al. Crowdsourcing in health and medical research: a systematic review. *Infectious Diseases of Poverty*, 2020, vol. 9, no. 1, pp. 8. <https://doi.org/10.1186/s40249-020-0622-9>
 14. Ellis R.J., Sander R.M., Limon A. Twelve key challenges in medical machine learning and solutions. *Intelligence-Based Medicine*, 2022, vol. 6, pp. 100068. <https://doi.org/10.1016/j.ibmed.2022.100068>
 15. Xia H., McKernan B. Privacy in crowdsourcing: a review of the threats and challenges. *Computer Supported Cooperative Work (CSCW)*, 2020, vol. 29, no. 3, pp. 263–301. <https://doi.org/10.1007/s10606-020-09374-0>
 16. Rother A., Niemann U., Hielscher T., Völzke H., Ittermann T., Spiliopoulou M. Assessing the difficulty of annotating medical data in crowdworking with help of experiments. *PLOS ONE*, 2021, vol. 16, no. 7, pp. e0254764. <https://doi.org/10.1371/journal.pone.0254764>
 17. Ye C., Coco J., Epishova A., Hajaj C., Bogardus H., Novak L., et al. A crowdsourcing framework for medical data sets. *AMIA Joint Summits on Translational Science proceedings*, 2018, pp. 273–280.
 18. Kittur A., Nickerson J., Bernstein M., Gerber E., Shaw A., Zimmerman J., et al. The future of crowd work. *Proc. of the Conference on Computer Supported Cooperative Work*, 2013, pp. 1301–1318. <https://doi.org/10.1145/2441776.2441923>
 19. Ørting S.N., Doyle A., van Hilten A., Hirth M., Inel O., Madan C.R., et al. A survey of crowdsourcing in medical image analysis. *Human Computation*, 2020, vol. 7, no. 1, pp. 1–26. <https://doi.org/10.15346/hc.v7i1.1>
 20. Lu J., Li W., Wang Q., Zhang Y. Research on data quality control of crowdsourcing annotation: a survey. *Proc. of the IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)*, 2020, pp. 201–208. <https://doi.org/10.1109/DASC-PiCom-CBDCCom-CyberSciTech49142.2020.00044>
 21. Lu X., Ratcliffe D., Kao T.-T., Tikhonov A., Litchfield L., Rodger C., Wang K. Rethinking quality assurance for crowdsourced multi-ROI

- Computation and Crowdsourcing. 2023. V. 11. N 1. P. 103–114. <https://doi.org/10.1609/hcomp.v11i1.27552>
22. Тесленко Е.В. Искусственный интеллект в медицине. Правовые аспекты // Наука молодых — будущее России: сборник научных статей 8-й Международной научной конференции перспективных разработок молодых ученых. Курск: Университетская книга, 2023. С. 435–438.
23. Hulsen T. Sharing is caring — data sharing initiatives in healthcare // *International Journal of Environmental Research and Public Health*. 2020. V. 17. N 9. P. 3046. <https://doi.org/10.3390/ijerph17093046>
24. Sims M.H., Shaw M.H., Gilbertson S., Storch J., Halterman M.W. Legal and ethical issues surrounding the use of crowdsourcing among healthcare providers // *Health Informatics Journal*. 2019. V. 25. N 4. P. 1618–1630. <https://doi.org/10.1177/1460458218796599>
25. Mason W., Suri S. Conducting behavioral research on Amazon's Mechanical Turk // *Behavior Research Methods*. 2012. V. 44. N 1. P. 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
26. Buecheler T., Sieg J.H., Fuchsli R.M., Pfeifer R. Crowdsourcing, open innovation and collective intelligence in the scientific method: a research agenda and operational framework // *Proc. of the 12th International Conference on the Synthesis and Simulation of Living Systems*. 2010. P. 679–686.
27. Dortheimer J. Collective intelligence in design crowdsourcing // *Mathematics*. 2022. V. 10. N 4. P. 539. <https://doi.org/10.3390/math10040539>
28. Le K.H., Tran T.V., Pham H.H., Nguyen H.T., Le T.T., Nguyen H. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis // *IEEE Access*. 2023. V. 11. P. 14105–14114. <https://doi.org/10.1109/ACCESS.2023.3243845>
29. Petrović N., Moyà-Alcover G., Varona J., Jaume-i-Capó A. Crowdsourcing human-based computation for medical image analysis: a systematic literature review // *Health Informatics Journal*. 2020. V. 26. N 4. P. 2446–2469. <https://doi.org/10.1177/1460458220907435>
30. Vindas Y., Guépié B.K., Almar M., Roux E., Delachartre P. Semi-automatic data annotation based on feature-space projection and local quality metrics: An application to cerebral emboli characterization // *Medical Image Analysis*. 2022. V. 79. P. 102437. <https://doi.org/10.1016/j.media.2022.102437>
31. Philbrick K. A., Weston A.D., Akkus Z., Kline T.L., Korfiatis P., Sakinis T., et al. RIL-Contour: a medical imaging dataset annotation tool for and with deep learning // *Journal of Digital Imaging*. 2019. V. 32. N 4. P. 571–581. <https://doi.org/10.1007/s10278-019-00232-0>
32. Li H., Zhang B., Zhang Y., Liu W.W., Mao Y.J., Huang J.C., Wei L.F. A semi-automated annotation algorithm based on weakly supervised learning for medical images // *Biocybernetics and Biomedical Engineering*. 2020. V. 40. N 2. P. 787–802. <https://doi.org/10.1016/j.bbe.2020.03.005>
33. Larobina M., Murino L. Medical image file formats // *Journal of Digital Imaging*. 2014. V. 27. N 2. P. 200–206. <https://doi.org/10.1007/s10278-013-9657-9>
34. Willemink M.J., Koszek W.A., Hardell C., Wu J., Fleischmann D., Harvey H., et al. Preparing medical imaging data for machine learning // *Radiology*. 2020. V. 295. N 1. P. 4–15. <https://doi.org/10.1148/radiol.2020192224>
35. Pfof A., Lu S.-C., Sidey-Gibbons C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison // *BMC Medical Research Methodology*. 2022. V. 22. N 1. P. 282. <https://doi.org/10.1186/s12874-022-01758-8>
36. Кондратенко С.С., Коржук В.М. Архитектура системы обработки медицинских данных с учетом требований обеспечения целостности. Сборник тезисов докладов конгресса молодых ученых. 2023. [Электронный ресурс]. URL: <https://kmu.itmo.ru/digests/article/11444>
37. Васильев Ю.А., Савкина Е.Ф., Владимирский А.В., Омелянская О.В., Арзамасов К.М. Обзор современных средств разметки цифровых диагностических изображений // *Казанский медицинский журнал*. 2023. Т. 104. № 5. С. 750–760. <https://doi.org/10.17816/KMJ349060>
38. Ежов Ф.В., Коваленко Л.А., Разумилов Е.С., Блеканов И.С. Инструменты краудсорсинга для анализа и обработки медицинских изображений в виде снимков КТ // *Процессы управления и устойчивость*. 2023. Т. 10. № 1. С. 291–297.
39. Saltz J.S., Krasteva I. Current approaches for executing big data science projects — a systematic literature review // *PeerJ Computer Science*. 2022. V. 8. P. e862. <https://doi.org/10.7717/peerj-cs.862>
- image segmentation. *Proc. of the 11th AAAI Conference on Human Computation and Crowdsourcing*, 2023, vol. 11, no. 1, pp. 103–114. <https://doi.org/10.1609/hcomp.v11i1.27552>
22. Teslenko E.V. Artificial intelligence in medicine. Legal aspects. *Proc. of the Science of the young — the future of Russia*. 2023. pp. 435–438. (in Russian)
23. Hulsen T. Sharing is caring — data sharing initiatives in healthcare. *International Journal of Environmental Research and Public Health*, 2020, vol. 17, no. 9, pp. 3046. <https://doi.org/10.3390/ijerph17093046>
24. Sims M.H., Shaw M.H., Gilbertson S., Storch J., Halterman M.W. Legal and ethical issues surrounding the use of crowdsourcing among healthcare providers. *Health Informatics Journal*, 2019, vol. 25, no. 4, pp. 1618–1630. <https://doi.org/10.1177/1460458218796599>
25. Mason W., Suri S. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 2012, vol. 44, no. 1, pp. 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
26. Buecheler T., Sieg J.H., Fuchsli R.M., Pfeifer R. Crowdsourcing, open innovation and collective intelligence in the scientific method: a research agenda and operational framework. *Proc. of the 12th International Conference on the Synthesis and Simulation of Living Systems*, 2010, pp. 679–686.
27. Dortheimer J. Collective intelligence in design crowdsourcing. *Mathematics*, 2022, vol. 10, no. 4, pp. 539. <https://doi.org/10.3390/math10040539>
28. Le K.H., Tran T.V., Pham H.H., Nguyen H.T., Le T.T., Nguyen H. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis. *IEEE Access*, 2023, vol. 11, pp. 14105–14114. <https://doi.org/10.1109/ACCESS.2023.3243845>
29. Petrović N., Moyà-Alcover G., Varona J., Jaume-i-Capó A. Crowdsourcing human-based computation for medical image analysis: a systematic literature review. *Health Informatics Journal*, 2020, vol. 26, no. 4, pp. 2446–2469. <https://doi.org/10.1177/1460458220907435>
30. Vindas Y., Guépié B.K., Almar M., Roux E., Delachartre P. Semi-automatic data annotation based on feature-space projection and local quality metrics: An application to cerebral emboli characterization. *Medical Image Analysis*, 2022, vol. 79, pp. 102437. <https://doi.org/10.1016/j.media.2022.102437>
31. Philbrick K. A., Weston A.D., Akkus Z., Kline T.L., Korfiatis P., Sakinis T., et al. RIL-Contour: a medical imaging dataset annotation tool for and with deep learning. *Journal of Digital Imaging*, 2019, vol. 32, no. 4, pp. 571–581. <https://doi.org/10.1007/s10278-019-00232-0>
32. Li H., Zhang B., Zhang Y., Liu W.W., Mao Y.J., Huang J.C., Wei L.F. A semi-automated annotation algorithm based on weakly supervised learning for medical images. *Biocybernetics and Biomedical Engineering*, 2020, vol. 40, no. 2, pp. 787–802. <https://doi.org/10.1016/j.bbe.2020.03.005>
33. Larobina M., Murino L. Medical image file formats. *Journal of Digital Imaging*, 2014, vol. 27, no. 2, pp. 200–206. <https://doi.org/10.1007/s10278-013-9657-9>
34. Willemink M.J., Koszek W.A., Hardell C., Wu J., Fleischmann D., Harvey H., et al. Preparing medical imaging data for machine learning. *Radiology*, 2020, vol. 295, no. 1, pp. 4–15. <https://doi.org/10.1148/radiol.2020192224>
35. Pfof A., Lu S.-C., Sidey-Gibbons C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Medical Research Methodology*, 2022, vol. 22, no. 1, pp. 282. <https://doi.org/10.1186/s12874-022-01758-8>
36. Kondratenko S.S., Korzhuk V.M. Architecture of a medical data processing system taking into account integrity requirements. *Collection of abstracts from the Congress of Young Scientists*. 2023. Available at: <https://kmu.itmo.ru/digests/article/11444> (in Russian)
37. Vasilev Y.A., Savkina E.F., Vladimirovskii A.V., Omelianskaia O.V., Arzamasov K.M. Overview of modern digital diagnostic image markup tools. *Kazan Medical Journal*, 2023, vol. 104, no. 5, pp. 750–760. (in Russian). <https://doi.org/10.17816/KMJ349060>
38. Ezhov F.V., Kovalenko L.A., Razumilov E.S., Blekanov I.S. Crowdsourcing tools for the analysis and processing of medical CT images. *Processy Upravleniya i Ustoichivost'*, 2023, vol. 10, no. 1, pp. 291–297. (in Russian)
39. Saltz J.S., Krasteva I. Current approaches for executing big data science projects — a systematic literature review. *PeerJ Computer Science*, 2022, vol. 8, pp. e862. <https://doi.org/10.7717/peerj-cs.862>

40. Saltz J.S. CRISP-DM for data science: strengths, weaknesses and potential next steps // *Proc. of the IEEE International Conference on Big Data*. 2021. P. 2337–2344. <https://doi.org/10.1109/bigdata52589.2021.9671634>
41. Saltz J., Hotz N. Factors that influence the selection of a data science process management methodology: an exploratory study // *Proc. of the 54th Hawaii International Conference on System Sciences*. 2021. P. 949–958. <https://doi.org/10.24251/hicss.2021.116>
42. Zhao X., Zhang P., Song F., Fan G.D., Sun Y.Y., Wang Y.J., et al. D2A U-Net: Automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution // *Computers in Biology and Medicine*. 2021. V. 135. P. 104526. <https://doi.org/10.1016/j.compbmed.2021.104526>
43. Xie Y., Padgett J., Biancardi A.M., Reeves A.P. Automated aorta segmentation in low-dose chest CT images // *International Journal of Computer Assisted Radiology and Surgery*. 2014. V. 9. N 2. P. 211–219. <https://doi.org/10.1007/s11548-013-0924-5>
44. Ким Г.И., Блеканов И.С., Ежов Ф.В., Коваленко Л.А., Ларин Е.С., Разумилов Е.С. [и др.] Методы искусственного интеллекта в сердечно-сосудистой хирургии и диагностика патологии аорты и аортального клапана (обзор литературы) // *Сибирский журнал клинической и экспериментальной медицины*. 2024. Т. 39. № 2. С. 36–45. <https://doi.org/10.29001/2073-8552-2024-39-2-36-45>
45. Gao R., Zhao S., Aishanjiang K., Cai H., Wei T., Zhang Y.C., et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data // *Journal of Hematology & Oncology*. 2021. V. 14. N 1. P. 154. <https://doi.org/10.1186/s13045-021-01167-2>
46. Chen P.-T., Wu T.H., Wang P.C., Chang D.W., Liu K.L., Wu M.S., et al. Pancreatic cancer detection on CT scans with deep learning: a nationwide population-based study // *Radiology*. 2023. V. 306. N 1. P. 172–182. <https://doi.org/10.1148/radiol.220152>
47. Zhou H., Li L., Liu Z., Zhao K., Chen X., Lu M., et al. Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images // *European Radiology*. 2021. V. 31. N 6. P. 3931–3940. <https://doi.org/10.1007/s00330-020-07454-9>
40. Saltz J.S. CRISP-DM for data science: strengths, weaknesses and potential next steps. *Proc. of the IEEE International Conference on Big Data*, 2021, pp. 2337–2344. <https://doi.org/10.1109/bigdata52589.2021.9671634>
41. Saltz J., Hotz N. Factors that influence the selection of a data science process management methodology: an exploratory study. *Proc. of the 54th Hawaii International Conference on System Sciences*, 2021, pp. 949–958. <https://doi.org/10.24251/hicss.2021.116>
42. Zhao X., Zhang P., Song F., Fan G.D., Sun Y.Y., Wang Y.J., et al. D2A U-Net: Automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution. *Computers in Biology and Medicine*, 2021, vol. 135, pp. 104526. <https://doi.org/10.1016/j.compbmed.2021.104526>
43. Xie Y., Padgett J., Biancardi A.M., Reeves A.P. Automated aorta segmentation in low-dose chest CT images. *International Journal of Computer Assisted Radiology and Surgery*, 2014, vol. 9, no. 2, pp. 211–219. <https://doi.org/10.1007/s11548-013-0924-5>
44. Kim G.I., Blekanov I.S., Ezhov F.V., Kovalenko L.A., Larin E.S., Razumilov E.S., et al. Artificial intelligence methods in cardiovascular surgery and diagnosis of pathology of the aorta and aortic valve (literature review). *Siberian Journal of Clinical and Experimental Medicine*, 2024, vol. 39, no. 2, pp. 36–45. (in Russian). <https://doi.org/10.29001/2073-8552-2024-39-2-36-45>
45. Gao R., Zhao S., Aishanjiang K., Cai H., Wei T., Zhang Y.C., et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. *Journal of Hematology & Oncology*, 2021, vol. 14, no. 1, pp. 154. <https://doi.org/10.1186/s13045-021-01167-2>
46. Chen P.-T., Wu T.H., Wang P.C., Chang D.W., Liu K.L., Wu M.S., et al. Pancreatic cancer detection on CT scans with deep learning: a nationwide population-based study. *Radiology*, 2023, vol. 306, no. 1, pp. 172–182. <https://doi.org/10.1148/radiol.220152>
47. Zhou H., Li L., Liu Z., Zhao K., Chen X., Lu M., et al. Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images. *European Radiology*, 2021, vol. 31, no. 6, pp. 3931–3940. <https://doi.org/10.1007/s00330-020-07454-9>

Авторы

Коваленко Лев Алексеевич — ведущий программист, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация; ведущий программист, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 59225183700](https://orcid.org/0009-0007-8233-4387), <https://orcid.org/0009-0007-8233-4387>, levozavr1@gmail.com

Блеканов Иван Станиславович — кандидат технических наук, доцент, заведующий кафедрой, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, [sc 56149559700](https://orcid.org/0000-0002-7305-1429), <https://orcid.org/0000-0002-7305-1429>, i.blekanov@spbu.ru

Ежов Федор Валерьевич — инженер-программист, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, [sc 59224591300](https://orcid.org/0009-0007-1468-0042), <https://orcid.org/0009-0007-1468-0042>, st056053@student.spbu.ru

Ларин Евгений Сергеевич — ведущий аналитик, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, <https://orcid.org/0009-0007-6199-3607>, evgeny.larin@spbu.ru

Ким Глеб Ирламович — кандидат медицинских наук, врач сердечно-сосудистый хирург, Санкт-Петербургский государственный университет. Клиника высоких медицинских технологий им. Н.И. Пирогова, Санкт-Петербург, 190020, Российская Федерация; доцент, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, [sc 57704764600](https://orcid.org/0000-0002-9344-5724), <https://orcid.org/0000-0002-9344-5724>, gikim.cor@gmail.com

Authors

Lev A. Kovalenko — Leading Software Developer, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation; Leading Software Developer, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 59225183700](https://orcid.org/0009-0007-8233-4387), <https://orcid.org/0009-0007-8233-4387>, levozavr1@gmail.com

Ivan S. Blekanov — PhD, Associate Professor, Head of Department, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, [sc 56149559700](https://orcid.org/0000-0002-7305-1429), <https://orcid.org/0000-0002-7305-1429>, i.blekanov@spbu.ru

Fedor V. Ezhov — Software Developer - Engineer, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, [sc 59224591300](https://orcid.org/0009-0007-1468-0042), <https://orcid.org/0009-0007-1468-0042>, st056053@student.spbu.ru

Evgenii S. Larin — Leading Analyst, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, <https://orcid.org/0009-0007-6199-3607>, evgeny.larin@spbu.ru

Gleb I. Kim — PhD (Medicine), Cardiovascular Surgeon, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation; Associate Professor, Saint Petersburg State University Hospital, Saint Petersburg, 190020, Russian Federation, [sc 57704764600](https://orcid.org/0000-0002-9344-5724), <https://orcid.org/0000-0002-9344-5724>, gikim.cor@gmail.com

Статья поступила в редакцию 15.01.2025

Одобрена после рецензирования 30.07.2025

Принята к печати 30.09.2025

Received 15.01.2025

Approved after reviewing 30.07.2025

Accepted 30.09.2025



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»