ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 59921.7— 2022

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ МЕДИЦИНЕ. АЛГОРИТМЫ АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Методы испытаний. Общие требования

Издание официальное

Москва
Российский институт стандартизации
2022

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Государственным бюджетным учреждением здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 18 октября 2022 г. № 1142-ст

4 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Оформление. ФГБУ «РСТ», 2022

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Общие положения 5
5 Оценка соответствия системы искусственного интеллекта заявленным техническим требованиям 5
6 Оценка эффективности и безопасности 6
7 Функциональное тестирование 7
8 Уровень производительности 11
9 Надежность 12
10 Тестирование на устойчивость к состязательным атакам 13
11 Тестирование на способность и устойчивость работы с разнородными данными 15
Приложение А (обязательное) Дизайн испытаний 18
Приложение Б (справочное) Содержание программы клинических испытаний системы искусственного интеллекта для анализа медицинских изображений 24
Приложение В (справочное) Пример заполнения протокола по результатам функционального тестирования системы искусственного интеллекта 28
Библиография 29

Введение

Неотъемлемой частью жизненного цикла систем искусственного интеллекта, которые относятся к программному обеспечению с применением технологий искусственного интеллекта, являющемуся медицинским изделием, является подтверждение воспроизводимости, надежности, безопасности использования и его эффективности при применении в соответствии с назначением. Для оценки указанных параметров осуществляются технические испытания программного обеспечения с применением технологий искусственного интеллекта, являющегося медицинским изделием, а также его клиническая оценка.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ МЕДИЦИНЕ. АЛГОРИТМЫ АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Методы испытаний. Общие требования

Artificial Intelligence systems in clinical medicine. Algorithms of medical images analysis. Test methods. General requirements

Дата введения — 2023—01—01

1 Область применения

Настоящий стандарт устанавливает общие требования к методам испытаний системы искусственного интеллекта, функциональным назначением которой является анализ медицинских изображений.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ 28195—89 Оценка качества программных средств. Общие положения

ГОСТ ISO/IEC 17025 Общие требования к компетентности испытательных и калибровочных лабораторий

ГОСТ Р 51275 Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения

ГОСТ Р 51583—2014 Защита информации. Порядок создания автоматизированных систем в защищенном исполнении. Общие положения

ГОСТ Р 55544—2013/IEC/TR 80002-1:2009 Программное обеспечение медицинских изделий. Часть 1. Руководство по применению ИСО 14971 к программному обеспечению медицинских изделий

ГОСТ Р 56429 Изделия медицинские. Клиническая оценка

ГОСТ Р 56939—2016 Защита информации. Разработка безопасного программного обеспечения. Общие требования

ГОСТ Р 59276 Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения

ГОСТ Р 59898—2021 Оценка качества систем искусственного интеллекта. Общие положения

ГОСТ Р 59921.1 Системы искусственного интеллекта в клинической медицине. Клиническая оценка

ГОСТ Р 59921.2 Системы искусственного интеллекта в клинической медицине. Часть 2. Программа и методика технических испытаний

ГОСТ Р 59921.4 Системы искусственного интеллекта в клинической медицине. Часть 4. Оценка и контроль эксплуатационных параметров

ГОСТ Р 59921.5—2022 Системы искусственного интеллекта в клинической медицине. Часть 5. Требования к структуре и порядку применения набора данных для обучения и тестирования алгоритмов

ГОСТ Р ИСО 5725-2 Точность (правильность и прецизионность) методов и результатов измерений. Часть 2. Основной метод определения повторяемости и воспроизводимости стандартного метода измерений

Издание официальное

ГОСТ Р ИСО 9241-161 Эргономика взаимодействия человек—система. Часть 161. Элементы графического пользовательского интерфейса

ГОСТ Р ИСО 14155 Клинические исследования. Надлежащая клиническая практика

ГОСТ Р ИСО/МЭК 9126—93 Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению

ГОСТ Р ИСО/МЭК 12119—2000 Информационная технология. Пакеты программ. Требования к качеству и тестирование

ГОСТ Р ИСО/МЭК 25040 Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Процесс оценки

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1

автоматизированная система: Система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая информационную технологию выполнения установленных функций.

[ГОСТ 34.003—90, статья 1.1]

3.2 **автоматическое тестирование** (benchmarking): Метод, который измеряет производительность автоматизированной системы на тщательно разработанных наборах данных, которые находятся в открытом доступе и/или используются для конкурентного тестирования различных автоматизированных систем.

3.3

восстанавливаемость (recoverability): Атрибуты программного обеспечения, относящиеся к его возможности восстанавливать уровень качества функционирования и восстанавливать данные, непосредственно поврежденные в случае отказа, а также к времени и усилиям, необходимым для этого.

[ГОСТ Р ИСО/МЭК 9126—93, пункт А.2.2.3]

3.4

главный исследователь (principal investigator): Квалифицированное лицо, ответственное за проведение клинического исследования в исследовательском центре.

Примечания

1 Если клинические исследования проводятся группой лиц в исследовательском центре, главный исследователь ответственен за руководство данной группой.

2 Является ли ответственным конкретное лицо или институт, может зависеть от особенностей национальной системы регулирования медицинских изделий.

[ГОСТ Р ИСО 14155—2014, пункт 3.33]

3.5

компаратор (comparator): Медицинское изделие, лечебно-диагностическая процедура (например, активный контроль), плацебо или отсутствие лечения, используемые в контрольной группе при клиническом исследовании.

[ГОСТ Р ИСО 14155—2014, пункт 3.10]

3.6

конечная точка (главная) (endpoint(s)): Важнейший показатель, используемый для проверки основной гипотезы клинического исследования.

[ГОСТ Р ИСО 14155—2014, пункт 3.16]

3.7 **медицинское изображение** (medical image): Информация, получаемая с использованием средств визуализации внутренних структур и функций человеческого тела, представленная в виде изображения, удобного для медицинской диагностики.

Примечания

1 Медицинское изображение может быть получено радиологическими или нерадиологическими методами.

2 Радиологические методы получения медицинского изображения основаны на использовании электромагнитного поля (например, методы рентгеновской диагностики, компьютерной томографии, магнитно-резонансной томографии и т. д.).

3 Нерадиологическими методами получают медицинские изображения, которые отсняты видеокамерой (эндоскопия) или сфотографированные (микроскопические изображения в гистологии, патологии, дерматологические изображения и т.п.).

3.8

отказ: Событие, заключающееся в нарушении работоспособного состояния объекта.

Примечания

1 Отказ может быть полным или частичным.

2 Полный отказ характеризуется переходом объекта в неработоспособное состояние.

Частичный отказ характеризуется переходом объекта в частично неработоспособное состояние.

[ГОСТ Р 27.102—2021, статья 36]

3.9 **робастность** (robustness): Способность системы поддерживать определенный уровень показателей при различных обстоятельствах.

Примечание — См. [1], пункт 76.

3.10

система искусственного интеллекта (artificial intelligence system): Программное обеспечение, в котором используются технологические решения искусственного интеллекта.

[Адаптировано из ГОСТ Р 59276—2020, пункт 3.16]

3.11

согласованность (completeness): Атрибуты программного обеспечения, которые заставляют программу придерживаться соответствующих стандартов или соглашений, или положений законов, или подобных рекомендаций.

[ГОСТ Р ИСО/МЭК 9126—93, пункт А.2.1.4]

3.12 **состязательная атака** (adversarial attack): Применение состязательного примера путем подачи его на вход системы искусственного интеллекта с целью отказа системы искусственного интеллекта.

Примечания

1 Обычно применяется к моделям искусственного интеллекта в форме нейронной сети.

2 См. [2], пункт 3.1.6.

3.13 **состязательный пример** (adversarial example): Входные данные для модели искусственного интеллекта, созданные путем добавления универсального случайного возмущения к примеру входных данных, что приводит к тому, что система искусственного интеллекта выдает неверный результат с высокой степенью достоверности.

Примечания

1 Обычно применяется к моделям искусственного интеллекта в форме нейронной сети.

2 См. [2], пункт 3.1.7.

3.14 **способность к непрерывному обучению (постоянное обучение, обучение на протяжении всей жизни, самообучение)** (continuous learning, continual learning, lifelong learning): Атрибут системы искусственного интеллекта, относящий к последовательному обучению системы искусственного интеллекта, которое происходит на постоянной основе на всем этапе эксплуатации жизненного цикла системы искусственного интеллекта.

3.15

стабильность (maturity): Атрибуты программного обеспечения, относящиеся к частоте отказов при ошибках в программном обеспечении.

[ГОСТ Р ИСО/МЭК 9126—93, пункт А.2.2.1]

3.16 **универсальное случайное возмущение** (universal random perturbation): Шум, который при добавлении к входным данным системы искусственного интеллекта может серьезно снизить ее показатели эффективности.

3.17

уровень производительности (performance efficiency): Производительность относительно суммы использованных при определенных условиях ресурсов.

Примечание — Ресурсы могут включать в себя другие программные продукты, конфигурацию программного и аппаратного обеспечения системы и материалы (например, бумагу для печати, носители).

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.2]

3.18

устойчивость к ошибке (fault tolerance): Атрибуты программного обеспечения, относящиеся к его способности поддерживать определенный уровень качества функционирования в случаях программных ошибок или нарушения определенного интерфейса.

[ГОСТ Р ИСО/МЭК 9126—93, пункт А.2.2.2]

3.19

функциональная корректность (functional correctness): Степень обеспечения продуктом или системой необходимой степени точности корректных результатов.

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1.2]

3.20

функциональная полнота (functional completeness): Степень покрытия совокупностью функций всех определенных задач и целей пользователя.

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1.1]

3.21

функциональная целесообразность (functional appropriateness): Степень функционального упрощения выполнения определенных задач и достижения целей.

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1.3]

4 Общие положения

4.1 Контроль качества, эффективности и безопасности системы искусственного интеллекта (СИИ), являющейся медицинским изделием, следует проводить в течение всего жизненного цикла СИИ, в том числе путем проведения его технических испытаний и клинической оценки. Общая методология и описание испытаний СИИ приведены в таблице А.1 приложения А.

4.2 На этапе разработки СИИ выполняют доказательство концепции (proof of concept), необходимое для определения применимости разрабатываемой СИИ для предполагаемого назначения (см. таблицу А.1 приложения А).

4.3 В рамках технических испытаний СИИ должны быть оценены ее метрики качества как программного продукта. Данные метрики и методы испытаний основаны на ГОСТ Р ИСО/МЭК 9126, ГОСТ Р ИСО/МЭК 25040, ГОСТ 28195, ГОСТ Р ИСО/МЭК 12119. Общие требования к номенклатуре метрик качества СИИ приведены в ГОСТ Р 59898 и ГОСТ Р 59276. Общие требования к техническим испытаниям СИИ, функциональным назначением которой является анализ медицинских изображений, приведены в разделе 5.

4.4 В отношении СИИ проводят также оценку ее клинической эффективности, безопасности и качества. Общие требования клинической оценки СИИ приведены в разделе 6.

4.5 Для целей выполнения испытаний производитель СИИ должен, в том числе, предоставлять пользователям информацию о назначении СИИ, включая следующие данные:

- назначение СИИ;
- целевую популяцию и условия, для которых предназначена СИИ;
- предполагаемые результаты использования СИИ для пациентов и системы здравоохранения.

4.6 Предоставленной производителем информации должно быть достаточно для того, чтобы пользователи могли оценить применимость СИИ для целевого назначения. В описании должны содержаться следующие данные:

- заявленные значения метрики, используемой для различения субъектов с наличием и отсутствием целевого признака. Метрика должна быть выбрана в согласии с медицинскими экспертами в целевой области и соответствовать назначению СИИ;

Пример — Если СИИ предназначена для оценки вероятности рака груди, она должна обладать более высокой чувствительностью, чем специфичностью.

- характеристики наборов данных, используемых для обучения и тестирования СИИ;
- тип и формат входных данных;
- известные ограничения СИИ (например, тип входных данных, требования к сетевым ресурсам, масштабируемость системы при вводе СИИ в эксплуатацию, требования к среде функционирования при вводе СИИ в эксплуатацию);
- целевой метод медицинской визуализации (рентгенография, магнитно-резонансная томография, ультразвук, компьютерная томография и т.д.). При этом элементы пользовательского интерфейса должны соответствовать требованиям ГОСТ Р 55544, ГОСТ Р ИСО 9241-161 (см. также [3]).

5 Оценка соответствия системы искусственного интеллекта заявленным техническим требованиям

5.1 Для оценки соответствия характеристик СИИ требованиям технической и эксплуатационной документации изготовителя, заявленным стандартам, применимым регулирующим требованиям и т.д. выполняют технические испытания. Результаты технических испытаний должны быть воспроизводимы на различном совместимом оборудовании и представлены с использованием стандартных метрик оценки соответствующих характеристик.

5.2 С целью выполнения технических испытаний разрабатывают программу испытаний (см. ГОСТ Р 59921.2), включающую методику проведения испытаний существенных характеристик СИИ, определенных в ГОСТ Р 59276 и ГОСТ Р 59898.

5.3 СИИ, применяемая в клинической медицине для анализа медицинских изображений, имеет отличия от общих СИИ. Например, качество данных медицинской визуализации может существенно варьироваться в рамках системы здравоохранения: иметь разное соотношение сигнал/шум, дефекты различной природы, изменяться со временем из-за износа и (или) обновления оборудования. В связи с этим были введены дополнительные характеристики, которые уточнены в разделах 10 и 11. Соответ-

ствие существенных характеристик СИИ и методов испытаний, приведенных в настоящем стандарте, перечислены в таблице 1.

Испытания некоторых характеристик СИИ будут аналогичными испытаниям общих программных продуктов, поэтому для данных пунктов приведены нормативные ссылки.

5.4 Испытания проверки защиты информации СИИ («безопасность» в таблице 1) выполняют путем анализа программного кода, выполнением или проверкой квалификационных испытаний СИИ, выполненных изготовителем, в соответствии с ГОСТ Р 51583, ГОСТ Р 56939. Объективные и субъективные внутренние и внешние факторы, воздействующие на защищаемую информацию, определяют по ГОСТ Р 51275. Необходимо также проверить и убедиться в защите данных, которые применялись на этапе разработки СИИ (обучения модели искусственного интеллекта).

В случае применения сторонних средств защиты информации (например, программы шифрования, антивирусы и т. д.) должны быть предоставлены установленные действующими нормативными правовыми актами документы по сертификации данных программных продуктов.

Таблица 1 — Соответствие существенных характеристик системы искусственного интеллекта и методов испытаний

Группа характеристик	Характеристика по ГОСТ Р 59276	Существенная характеристика	Пункт настоящего стандарта/ нормативная ссылка на метод испытаний
Функциональность	Функциональные возможности	Функциональные возможности (functionality). Способность к взаимодействию (compatibility)	Раздел 8
	Эффективность (производительность)	Уровень производительности (performance efficiency)	Раздел 9
	Мобильность	Мобильность (portability)	ГОСТ Р ИСО/МЭК 12119—2000, раздел 4
	Практичность	Практичность (usability)	ГОСТ Р ИСО/МЭК 12119—2000, раздел 4
	Сопровождаемость	Сопровождаемость (maintainability)	ГОСТ Р ИСО/МЭК 12119—2000, раздел 4
Надежность	Надежность	Надежность (reliability)	Разделы 9—11
Безопасность	—	Защищенность (security)	ГОСТ Р 56939—2016, подраздел 5.4; ГОСТ Р 51583—2014, раздел 6; [14]

5.5 Перед выполнением тестирования необходимо осуществить подготовительные работы (см. ГОСТ Р 59898—2021, пункт 7.2.1). Подготовительные работы включают следующие этапы: установление целей и задач тестирования; определение критериев к набору данных для тестирования; выбор и обоснование существенных характеристик и метрик их оценки; определение допустимого диапазона изменений метрик, а также наиболее существенных факторов, оказывающих влияние на работу СИИ; утверждение состава экспертной группы, составление методики тестирования и подготовка программы тестирования.

6 Оценка эффективности и безопасности

6.1 Анализ эффективности, безопасности и качества использования СИИ при ее применении по назначению изготовителя проводят в процессе клинической оценки. Клинические испытания (как часть клинической оценки) СИИ проводят для проверки эффективности СИИ, обеспечения доверия к СИИ со стороны пользователей, в целях оценки возможности дальнейшего практического применения и должны соответствовать требованиям соответствующих нормативных правовых актов. Клинические испы-

тания включают оценку клинической связи и клиническую валидацию, как этапы клинической оценки (см. ГОСТ Р 59921.1).

6.2 Уровень интеграции СИИ в клиническую практику в течение клинических испытаний может варьировать от работы в фоновом режиме, когда СИИ функционирует одновременно и параллельно с существующими методами оказания медицинских услуг, до полноценных испытаний СИИ в соответствии с ее целевым назначением (см. ГОСТ Р 56429).

6.3 Метрики эффективности и безопасности СИИ должны включать в себя конечные точки (см. ГОСТ Р ИСО 14155), имеющие значение для пациентов (например, детекцию и классификацию патологии, оценку клинического исхода и т.п.).

6.4 Программа проведения клинических испытаний должна соответствовать критериям прозрачности. Также рекомендовано включать в программу испытаний пункты, приведенные в приложении Б.

6.5 Представление результатов клинических испытаний должно отвечать требованиям соответствующих нормативных правовых актов. Также рекомендуется в представление результатов клинических испытаний СИИ включать данные в соответствии с рекомендациями, такими как CONSORT-AI для клинических контролируемых испытаний вмешательств с использованием СИИ (см. [4]), эквивалентных инструментов для исследований диагностической точности СИИ (STARD-AI) (см. [5]), и испытаний предсказательных моделей (TRIPOD-AI) (см. [6]).

7 Функциональное тестирование

7.1 Назначение

Функциональное тестирование выполняют с целью оценки соответствия функциональных возможностей СИИ требованиям, указанным в техническом задании на создание этого продукта. Согласно ГОСТ Р 59898 в рамках функционального тестирования рекомендовано оценивать следующие метрики: функциональная пригодность (functional appropriateness), функциональная корректность (functional correctness), согласованность (compliance), функциональная полнота (functional completeness) и способность к самообучению (ability to learn).

Согласно ГОСТ Р ИСО/МЭК 25040 функциональное тестирование СИИ проводят методом «черного ящика» с контролем документации изготовителя.

Данный вид тестирования соответствует понятию аналитической валидации СИИ как программного обеспечения, являющегося медицинским изделием, в рамках клинической оценки СИИ (см. [7]).

7.2 Требования к квалификации персонала

Для выбора номенклатуры характеристик, подготовки набора данных, а также для проведения тестирования и оценки результатов создают экспертную группу, которая должна удовлетворять требованиям ГОСТ Р 59898—2021 (пункт 7.1.2).

Непосредственно выполнение тестирования в рамках функционального тестирования проводят силами квалифицированных технических специалистов, имеющих опыт тестирования СИИ, обработки изображений и анализа данных результатов. Для анализа функциональных характеристик СИИ и подготовки набора данных возможно привлечение медицинских специалистов, которые имеют опыт работы и квалификацию в областях, соответствующих решаемым задачам СИИ.

Тестирование проводит не зависимый от изготовителя СИИ коллектив исследователей; в противном случае в протоколе указывают источники конфликта интересов.

7.3 Описание метода тестирования

7.3.1 Тестирование выполняют на наборе данных, достаточного объема для получения статистически значимого результата, путем расчета выбранного набора метрик. Возможно использование автоматических методов тестирования (см. [1], пункт 79).

7.3.2 Метрики функциональной пригодности и способности к самообучению определяют согласно общим требованиям ГОСТ Р 59898—2021 (подраздел 8.2).

7.3.3 Оценку метрик функциональной полноты выполняют регистрационным методом (см. ГОСТ 28195—89, пункт 1.5), путем расчета отношения количества недостающих или неправильно реализованных функций к общему количеству функций СИИ, указанных в технической и эксплуатационной документации (см. ГОСТ Р 59898—2021, пункт 8.2.2).

Пример — Может быть произведен анализ следующих параметров:
- ***визуальная оценка выходных данных СИИ;***
- ***тип принимаемого решения;***
- ***набор заключений СИИ;***
- ***формат входных данных;***
- ***тип анализируемого объекта;***
- ***и т. д.***

7.3.4 Метрику согласованности оценивают путем анализа документации на соответствие стандартам или нормативным правовым актам, или другим рекомендациям (см. ГОСТ Р ИСО/МЭК 9126—93, пункт А.2.1.4). Техническая и эксплуатационная документация на СИИ должна включать описание эксплуатационных параметров, приведенных в ГОСТ Р 59921.4.

7.3.5 Метрики функциональной корректности (правильности) используют для подтверждения того, что СИИ генерирует выходные данные с надлежащим уровнем точности (accuracy), а также повторяемости (repeatability) и воспроизводимости (reproducibility) (см. ГОСТ Р 59921.2). Оценку данной метрики выполняют расчетным методом (см. ГОСТ 28195—89, пункт 1.5), на небольшом наборе данных с последующим определением соответствующих метрик (см. ГОСТ Р 59898—2021, пункт 8.2.3). Набор метрик определяют на основании вида решаемой задачи СИИ и устанавливают решением экспертной группы.

7.4 Требования к наборам данных

Функциональное тестирование проводят на наборе данных достаточного объема для получения статистически значимого результата и включающем изображения с отсутствием целевой патологии, а также изображения с целевой патологией. Набор данных должен быть верифицирован, а также должен включать данные из разных медицинских организаций и разных моделей/производителей оборудования, обработку данных с которых изготовитель СИИ включает в функциональное назначение.

7.5 Требования к оборудованию для проведения тестирования

Функциональное тестирование необходимо проводить с применением испытательного стенда (может быть как отдельным автоматизированным рабочим местом, так и виртуальной тестовой средой, которая удовлетворяет техническим требованиям СИИ (см. [1], статья 10).

Возможно использование виртуальной тестовой среды, позволяющей выполнять автоматические тестирования (benchmarking) (см. [1], статья 79).

Перед выполнением тестирования необходимо удостовериться в отсутствии существенных различий между средой проведения тестирования и средой эксплуатации СИИ (см. ГОСТ Р 59898—2021, подраздел 7.2).

7.6 Порядок выполнения тестирования

7.6.1 Тестирование проводят в соответствии с программой испытаний после выполнения подготовительных работ (см. 5.4).

7.6.2 Набор данных загружают на испытательный стенд, в котором инсталлирована СИИ или через который установлен доступ к СИИ с целью выполнения тестирования.

7.6.3 Обрабатывают элементы набора данных с использованием СИИ и фиксируют выходные данные СИИ.

7.6.4 При проведении тестирования документируют все программные и прочие ошибки, возникающие при эксплуатации СИИ.

7.6.5 При тестировании функциональной полноты необходимо предварительно определить номенклатуру функций СИИ, описанных в технической и эксплуатационной документации. Рекомендовано включить в перечень параметры, указанные в ГОСТ Р 59921.4:

- визуальная оценка выходных данных СИИ;
- типы принимаемого решения [тип (бинарное, вероятностное, локализация, классификация) и формат возвращаемого СИИ заключения];
- набор заключений СИИ [СИИ может формировать одно или больше заключений (многофакторное заключение)];
- формат входных данных и тип анализируемого объекта (входные данные для систем анализа медицинских изображений могут быть представлены статичными и динамическими изображениями

анализируемой области, сериями двумерных изображений, соответствующих сечениям анализируемой области, а также могут содержать метаданные).

При выполнении тестирования каждой функции результаты оформляют в табличном виде, пример представлен в таблице 2.

Таблица 2 — Анализ и оценка результатов тестирования на функциональную полноту

Наименование параметра	Оценочные данные	Результат
Визуальная оценка выходных данных СИИ	Соответствие заявленным в технической документации	Соответствует/не соответствует
Тип принимаемого решения	Соответствие заявленным в технической документации	Соответствует/не соответствует
Формат возвращаемого СИИ заключения	Соответствие заявленным в технической документации	Соответствует/не соответствует
Набор заключений СИИ	Соответствие заявленным в технической документации	Соответствует/не соответствует
Формат входных данных	Соответствие заявленным в технической документации	Соответствует/не соответствует
Тип анализируемого объекта	Соответствие заявленным в технической документации	Соответствует/не соответствует

7.6.6 При тестировании функциональной корректности выполняют тестирование СИИ на наборе данных, объем которых достаточен для получения статистически значимого результата.

Примечание — Нижнюю границу объема набора данных оценивают из расчета 10 образцов (экземпляров) на каждый оцениваемый параметр. Данный объем набора данных будет обеспечивать получение статистически значимого результата.

Порядок тестирования определяют исходя из установленного набора метрик:

- для оценки метрик точности выполняют регистрацию выходных данных СИИ при обработке набора данных;
- для оценки повторяемости выполняют повторную обработку набора данных в одинаковых условиях тестирования и регистрацию выходных данных СИИ;
- для оценки воспроизводимости выполняют повторную обработку одного набора данных в разных условиях испытаний (разные испытательные стенды, операторы, распределение во времени и т. д.) и регистрацию выходных данных СИИ.

Пример — Есть задача сегментации ткани на изображении магнитно-резонансной томографии, которая требует от пользователя ввода начальных точек или выбора срезов для анализа СИИ. Для данной задачи определение повторяемости будет заключаться в повторном анализе одного изображения семь раз одним и тем же оператором (в случае необходимости получения степени свободы не менее 6). Определение воспроизводимости может быть выполнено путем анализа данного изображения двумя операторами или одним оператором на разных испытательных стендах.

7.7 Анализ и оценка результатов тестирования

7.7.1 Оценка функциональной полноты возможна путем расчета метрики полноты реализации функций. Допустимо использовать другие показатели.

Полноту реализации функций определяют по следующей формуле (см. ГОСТ Р 59898—2021, пункт 8.2.2):

$$M_1 = 1 - \frac{A}{B}, \quad (1)$$

где A — количество недостающих или неправильно реализованных функций, обнаруженных при тестировании;

B — общее количество функций, описанных в технической и эксплуатационной документации СИИ.

В случае целесообразности можно вводить весовые коэффициенты для отдельных функций.

7.7.2 Оценка функциональной корректности возможна методом расчета показателей точности, воспроизводимости и повторяемости.

Показатель точности в зависимости от решаемой задачи СИИ и формулы расчета приведены в ГОСТ Р 59898—2021 (пункт 8.2.3):

- общая метрика — результативность;
- для задач регрессии — средняя квадратичная ошибка, средняя абсолютная ошибка;
- в задачах ранжирования — приведенная суммарная эффективность;
- для задач классификации и обнаружения — доля правильных исходов, точность, чувствительность, специфичность, F-мера, площадь под кривой ROC, площадь под кривой PRC. Возможно расширение набора метрик (см. [8], [9]) на основании решения экспертной группы.

Методы расчета показателей воспроизводимости и повторяемости приведены в ГОСТ Р ИСО 5725-2.

7.8 Критерии оценки и представление результатов

Все метрики нормируют, чтобы их значения были в интервале от 0 до 1. Весовые коэффициенты метрик и допустимый интервал определяют члены экспертной группы.

Критерии оценки определяют в соответствии с предназначением СИИ и заявленными характеристиками.

По завершении тестирования составляют отчет, в котором должны быть указаны, в том числе, перечень метрик, методы испытаний, характеристики наборов данных, допустимые интервалы, а также определенные значения параметров (см. таблицу 3).

Результаты функционального тестирования оформляют в виде протокола испытаний в соответствии с ГОСТ ISO/IEC 17025.

Таблица 3 — Анализ и оценка результатов функционального тестирования

Наименование характеристики	Нормативное значение	Определенное значение	Результат
Функциональная полнота	От … до …		Соответствует/ не соответствует
Функциональная корректность			
- точность	От … до …		Соответствует/ не соответствует
- воспроизводимость	От … до …		Соответствует/ не соответствует
- повторяемость	От … до …		Соответствует/ не соответствует
…			
Функциональная пригодность	От … до …		Соответствует/ не соответствует
Способность к самообучению	От … до …		Соответствует/ не соответствует

Оценку согласованности выполняют описательным методом, в связи с этим заполняют таблицу по примеру 4.

Таблица 4 — Анализ и оценка результатов функционального тестирования — согласованность

Наименование параметра	Оценочные данные	Результат
Согласованность	Соответствие нормативным требованиям	Соответствует/не соответствует

Пример заполнения протокола по результатам функционального тестирования СИИ приведен в приложении В.

8 Уровень производительности

8.1 Назначение

В рамках оценки уровня производительности (performance efficiency) рекомендуют оценивать следующие характеристики: характер изменения во времени, характер изменения ресурсов и производительные возможности (см. ГОСТ Р 59898—2021, подраздел 8.3).

8.2 Общие требования к проведению испытания

Для выбора номенклатуры характеристик, подготовки набора данных, а также для проведения тестирования и оценки результатов создают экспертную группу, которая должна удовлетворять требованиям ГОСТ Р 59898—2021 (пункт 7.1.2).

Непосредственно выполнение испытания уровня производительности проводят силами квалифицированных технических специалистов, имеющих опыт тестирования программного обеспечения, в том числе СИИ.

Набор данных для тестирования уровня производительности должен включать такой объем, чтобы выполнить тестирование на максимальном уровне загрузки СИИ. Наборы данных для данного тестирования не должны включать примеры медицинских изображений, отличающиеся от нормы и (или) патологии, применительно как к качеству изображения, так и к его содержанию (см. [1], статья 12).

Тестирование выполняют на испытательном стенде (может быть как отдельным автоматизированным рабочим местом, так и виртуальной тестовой средой, которая удовлетворяет техническим требованиям СИИ (см. [1], статья 10).

Возможно использование виртуальной тестовой среды, позволяющей выполнять автоматические тестирования (benchmarking) (см. [1], статья 79).

Перед выполнением тестирования необходимо удостовериться в отсутствии существенных различий между средой проведения тестирования и средой эксплуатации СИИ (см. ГОСТ Р 59898—2021, подраздел 7.2).

8.3 Метрики

Метрики характера изменения во времени (time behaviour) характеризуют соответствие требованиям временных ресурсов на выполнение операции, а также производительность — количество однотипных задач за определенное время. Методика расчета приведена в ГОСТ Р 59898—2021 (пункт 8.3.4).

Метрики характера изменения ресурсов (resource utilization) относят к объему используемых ресурсов и продолжительности такого использования при выполнении функции (см. ГОСТ Р 59898—2021, пункт 8.3.5).

Метрики производительных возможностей (capacity) характеризуют степень соответствия требованиям таких параметров СИИ, как количество параллельно обрабатываемых наборов данных, количество параллельно работающих пользователей, емкость канала, пропускная способность по транзакциям.

8.4 Обработка результатов и оформление протокола

В протокол испытания включают параметры, влияющие на уровень производительности: параметры центрального процессора, объем памяти, объем хранения, сетевой трафик и прочее.

Результаты тестирования оформляют в виде протокола испытаний в соответствии с ГОСТ ISO/IEC 17025 (см. таблицу 5).

Таблица 5 — Анализ и оценка результатов тестирования уровня производительности

Наименование характеристики	Нормативное значение	Определенное значение	Результат
Характер изменения во времени	От … до …		Соответствует/ не соответствует
Характер изменения ресурсов	От … до …		Соответствует/ не соответствует
Производительная возможность	От … до …		Соответствует/ не соответствует

9 Надежность

9.1 Назначение

Надежность СИИ определяют как способность СИИ сохранять свой уровень качества функционирования при установленных условиях за установленный период времени (см. ГОСТ Р ИСО/МЭК 9126).

К метрикам надежности СИИ относят общие характеристики, такие как стабильность (maturity), устойчивость к ошибке (fault tolerance), восстанавливаемость (recoverability), робастность (robustness). Также к метрикам робастности СИИ относят устойчивость к состязательным атакам (см. раздел 10) и показатели способности и устойчивости работы с разнородными данными (см. раздел 11).

9.2 Общие требования

Для выбора номенклатуры характеристик, подготовки набора данных, а также для проведения тестирования и оценки результатов создают экспертную группу, которая должна удовлетворять требованиям ГОСТ Р 59898—2021 (пункт 7.1.2).

Непосредственно выполнение тестирования в рамках тестирования надежности проводят силами квалифицированных технических специалистов, имеющих опыт тестирования СИИ.

Наборы данных для данного тестирования не должны включать примеры медицинских изображений, резко отличающиеся от нормы и (или) патологии, применительно как к качеству изображения, так и к его содержанию.

Тестирование выполняют на испытательном стенде (может быть как отдельным автоматизированным рабочим местом, так и виртуальной тестовой средой, которая удовлетворяет техническим требованиям СИИ (см. [1], статья 10).

Возможно использование виртуальной тестовой среды, позволяющей выполнять автоматические тестирования (benchmarking) (см. [1], статья 79).

Перед выполнением тестирования необходимо удостовериться в отсутствии существенных различий между средой проведения тестирования и средой эксплуатации СИИ (см. ГОСТ Р 59898—2021, подраздел 7.2).

Для тестирования надежности многократно повторяют обработку тестовых наборов, объем которых определяется экспертной группой, но должен быть не менее 1000 циклов.

В процессе тестирования регистрируют выходные данные СИИ с указанием типа ошибки, фактора, вызывающего ошибку, влияние ошибки на выходные данные, а также определяют время работы СИИ до момента возникновения ошибки.

Проводят оценку ущерба, связанного с каждой возникающей ошибкой. Важным фактором, влияющим на оценку надежности СИИ, являются ошибки, приводящие к снижению безопасности системы.

9.3 Метрики

9.3.1 Метрики стабильности и робастности приведены в ГОСТ Р 59898—2021 (подраздел 8.8). Метрика робастности включает также тестирование на устойчивость к состязательным атакам (см. раздел 10) и тестирование на способность и устойчивость работы с разнородными данными (см. раздел 11).

9.3.2 Устойчивость к ошибке, показывающая способность СИИ выполнять функции при аномальных условиях (сбой аппаратуры, некорректные действия оператора и т.д.). Определяется как отношение количества разных типов отказов, для которых предусмотрены средства восстановления, к общему типу отказов в испытаниях.

9.3.3 Восстанавливаемость, показывающая способность СИИ возобновить работу после возникновения ошибки. Вычисляется как сумма временных интервалов восстановления работоспособности системы после ошибки к количеству ошибок и отказов, зафиксированному в испытаниях.

9.4 Обработка результатов и оформление протокола

Результаты тестирования надежности оформляют в виде протокола испытаний в соответствии с ГОСТ ISO/IEC 17025 (см. таблицу 6). Данные по результатам тестирования на устойчивость к состязательным атакам и к работе с разнородными данными приводят отдельно.

Таблица 6 — Анализ и оценка результатов тестирования надежности

Наименование характеристики	Нормативное значение	Определенное значение	Результат
Метрики стабильности	От … до …		Соответствует/ не соответствует
Метрики устойчивости к ошибке	От … до …		Соответствует/ не соответствует
Метрики восстанавливаемости	От … до …		Соответствует/ не соответствует
Метрики робастности	От … до …		Соответствует/ не соответствует

10 Тестирование на устойчивость к состязательным атакам

10.1 Назначение

Состязательной атакой называют такой состязательный пример, который приводит к некорректной работе СИИ. Состязательный пример возникает в случае крайне небольшого изменения, внесенного во входные данные алгоритма ИИ, который приводит к неожиданному (и неправильному) большому изменению выходных данных, т. е. к совершенно другому результату, чем при неизменных входных данных. Состязательные примеры могут возникать, например, при изменении нескольких пикселей на медицинском изображении, искажение которых незаметно человеческому глазу.

Тестирование на устойчивость к состязательным атакам основано на формировании и обработке состязательных примеров для выявления дефектов в СИИ. Выполняя данное тестирование, проводят оценку устойчивости и надежности СИИ к наличию на входе системы данных, приводящих к возникновению состязательного примера (см. [1], [10]).

10.2 Требования к квалификации персонала

Тестирование на устойчивость к состязательным атакам проводят силами квалифицированных технических специалистов, имеющих опыт тестирования СИИ, обработки изображений и анализа результатов.

10.3 Описание метода тестирования

Оценка точности предсказаний алгоритма ИИ и (или) других метрик в условиях состязательных атак и в нормальных условиях.

10.4 Требования к оборудованию

Тестирование на устойчивость СИИ к состязательным атакам должно проводиться с применением стенда и/или в виртуальной тестовой среде, которая удовлетворяет техническим требованиям СИИ (см. [1]).

10.5 Требования к набору данных

Набор данных для выполнения данного тестирования должен состоять из изображений, которые соответствуют функциональному назначению СИИ. При этом половину набора данных составляют дан-

ные, полученные в нормальных условиях, а вторую — данные с внесенными искажениями для формирования состязательных примеров.

Методы формирования состязательных примеров могут являться добавлением универсального случайного возмущения (universal random perturbation), шума (imperceptible noise). Также могут быть использованы другие методы (FGSM, inverse FGSM, JSMA) (см. [11]).

Специалисты, выполняющие данное тестирование, устанавливают виды состязательных примеров, влияние которых будет оценено в соответствии с указанными в технической документации производителя.

10.6 Порядок выполнения тестирования

Формируют набор данных для выполнения тестирования на устойчивость к состязательным атакам согласно установленным требованиям.

Определяют выходные данные для каждого элемента набора данных, поданного на вход СИИ.

10.7 Анализ и оценка результатов тестирования

Сопоставляют выходные данные СИИ с заданными значениями набора данных. Вычисляют метрики качества, которые определяются типом принимаемого решения СИИ и ее функциональным назначением (см. [12]).

Составляют ранжированный список состязательных атак, упорядоченный по степени влияния на алгоритм.

10.8 Критерии оценки (метрики) и оформление результатов

Критерии оценки устойчивости СИИ к состязательным атакам определяют исходя из требований, приведенных в технической документации СИИ.

Пример — СИИ анализирует оптические изображения глаза для сегментации области глаукомы. На исходном изображении метрика точности сегментации (индекс Jaccard) составляла 71,1 %. После внесения изменений в исходное изображение, чтобы получить состязательный пример, значение метрики снизилось до 16,4 % (см. [13]).

Результаты оценки устойчивости СИИ к состязательным атакам оформляют в виде протокола испытаний в соответствии с ГОСТ ISO/IEC 17025 (см. таблицу 7). В таблице приводят данные метрик качества для используемых видов состязательных примеров (от 1 до N) (см. 10.5).

Таблица 7 — Анализ и оценка тестирования на устойчивость СИИ к состязательным атакам

Наименование характеристики	Нормативное значение	Определенное значение	Результат
Метрика качества на исходных изображениях	От … до …	…	Соответствует/ не соответствует
Метрика качества на изображениях — состязательных примерах (1-й вид)	От … до …	…	Соответствует/ не соответствует
Метрика качества на изображениях — состязательных примерах (2-й вид)	От … до …	…	Соответствует/ не соответствует
…			
Метрика качества на изображениях — состязательных примерах (N-й вид)	От … до …	…	Соответствует/ не соответствует

11 Тестирование на способность и устойчивость работы с разнородными данными

11.1 Назначение

Тестирование проводят для проверки соответствия СИИ установленным требованиям к способности и устойчивости работы с разнородными входными данными. В результате данного тестирования проверяют:

- работоспособность СИИ при работе с входными данными, соответствующими технической документации изготовителя;
- ограничения СИИ при работе с входными данными, не соответствующими технической документации изготовителя;
- корректность работы системы обнаружения и оповещения пользователя о подаче на вход СИИ изображений низкого качества, наличия инородных предметов, артефактов и др.

С целью выполнения данного тестирования формируют соответствующий набор данных. Разнородные входные данные включают:

- данные, определенные технической документацией изготовителя;
- данные, которые потенциально могут быть направлены на вход СИИ [полученные от оборудования, не указанного в технической документации, изображения, содержащие отличные от указанных в технической документации области сканирования (анатомической области)];
- изображения, подверженные разного рода трансформациям;
- изображения с визуальными артефактами;
- изображения, полученные от нецелевой популяции (другая раса, возраст, пол и т.д.);
- изображения, содержащие инородные предметы;
- другое.

11.2 Требования к квалификации персонала

Тестирование на способность и устойчивость работы с разнородными данными проводят силами квалифицированных технических специалистов, имеющих опыт тестирования СИИ, обработки изображений и анализа результатов.

11.3 Описание метода тестирования

Моделирование типовых задач СИИ с различными наборами данных и диагностическими устройствами.

Данное тестирование оценивает способность СИИ успешно обрабатывать входные данные, содержащие разного рода искажения путем расчета метрик качества.

11.4 Требования к наборам данных

11.4.1 Наборы данных для тестирования СИИ на способность и устойчивость работы с разнородными данными могут включать изображения из клинической практики и искусственно сгенерированные изображения путем применения методов трансформации. Параметры формирования набора (включая источник изображений) данных должен быть подробно описан в методах тестирования и протоколе с результатами.

11.4.2 Наборы данных должны содержать входные данные в формате, описанном в технической документации СИИ. Допускается включение в набор примеров данных в ином формате для исследования способности СИИ распознавать и анализировать альтернативно организованную информацию.

11.4.3 Наборы данных должны включать данные, полученные на разной аппаратуре, исследования с которой СИИ потенциально способна обработать (в том числе не указанные в технической документации).

11.4.4 Наборы данных должны включать весь возможный диапазон размеров и разрешений изображений, который можно получить на медицинском оборудовании, актуальном на момент проведения тестирования.

11.4.5 Необходимо включение в набор данных зашумленных или искаженных данных (изображения, подверженные разного рода трансформациям, изображения, содержащие артефакты). Данные изображения можно генерировать путем применения различных трансформаций к исходному изображению.

11.4.6 Наборы данных должны содержать изображения, включающие объекты, которые не соответствуют функциональному назначению СИИ, но могут быть в результате ошибки маршрутизации поданы на вход СИИ [например, изображения с анатомическими областями, отличными от заявленных в технической документации, белый шум, инородные предметы (имплантируемые изделия, металлоконструкции и иные объекты, которые не наблюдаются на данном медицинском изображении при обычных условиях) и др.].

11.4.7 Наборы данных должны включать также сложные для интерпретации экспертами изображения.

11.5 Требования к оборудованию

Тестирование должно быть проведено с применением стенда либо в виртуальной тестовой среде, которая удовлетворяет техническим требованиям СИИ (см. [1]).

11.6 Порядок выполнения тестирования

Для каждого элемента сформированного набора данных получают выходные данные СИИ с использованием испытательного оборудования.

Регистрируют выходные данные СИИ и сформированные ошибки и уведомления СИИ.

Перечисляют формат и тип входных данных, наименования диагностического устройства, с которого получено изображение, размеры и разрешения изображений, типы зашумленных и искаженных изображений, которые были поданы на вход системы (см. таблицу 8).

Таблица 8 — Характеристики входных данных

Наименование	Характеристика
Формат входных данных	
Тип входных данных	
Наименование диагностического устройства	
Размер изображений	
Разрешение изображений	
Характеристики изображений (зашумленных, искаженных, артефакты и др.)	Шум; искажение изображения (описание); другая анатомическая область (описание); инородные предметы (описание); др.

11.7 Анализ и оценка результатов тестирования

Рассчитывают показатели качества СИИ для соответствующего набора данных.

Для каждой выявленной ошибки СИИ определяют параметры: тип ошибки и триггер ошибки.

11.8 Критерии оценки (метрики)

Критерии оценки способности и устойчивости работы СИИ с разнородными данными определяют на основе анализа изменения заявленных значений метрик качества при подаче на вход СИИ разнородного набора данных.

Возможно использование также метрик точности определения некорректных изображений, которые не соответствуют функциональному назначению СИИ.

Результаты тестирования на способность и устойчивость работы с разнородными данными оформляют в виде протокола испытаний в соответствии с ГОСТ ISO/IEC 17025 (см. таблицы 9, 10). Таблица 10 формируется для всех смоделированных видов изображений низкого качества, которые могут быть поданы на вход СИИ, изображений с разных диагностических устройств и т.д. Набор проверяемых разнородных данных определяется технической документацией изготовителя.

Таблица 9 — Анализ и оценка результатов тестирования на способность и устойчивость работы с разнородными входными данными с оценочными результатами

Наименование параметра	Оценочные данные	Результат
Визуальная оценка выходных данных СИИ	Соответствие заявленным в технической документации	Соответствует/не соответствует
Уведомление пользователя о поступлении на вход изображения инородного предмета	Имеется	Соответствует/не соответствует
Уведомление пользователя о поступлении на вход изображения не соответствующей назначению анатомической области	Имеется	Соответствует/не соответствует
....		

Таблица 10 — Анализ и оценка результатов тестирования на способность и устойчивость работы с разнородными входными данными с количественными результатами

Наименование	Нормативное значение	Определенное значение	Результат
Метрика качества на исходных изображениях	От ... до ...	...	Соответствует/ не соответствует

Приложение А
(обязательное)

Дизайн испытаний

Таблица А.1 — Дизайн испытаний в зависимости от фазы испытаний

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
0	Доказательство концепции (proof of concept)	Контроль качества данных	Описательный анализ	Убедиться, что качество данных отвечает необходимым стандартам (требованиям) и что состав (диапазон) данных соответствует целевой популяции	Качество и диапазон данных могут изменяться в зависимости от цели исследования (анализируемых параметров)
0		Тестирование алгоритма	Статистический анализ	Оценка точности предсказаний алгоритма ИИ и (или) других метрик	Использование раздельных обучающего и тестирующего наборов данных; в качестве эталона выступает среднестатистический показатель профильных медицинских специалистов при выполнении данной задачи; приемлемые значения измеряемых метрик зависят от клинических последствий возможных ошибок
0, 1		Популяционное исследование	Наблюдения и опросы; анализ последовательности операций (workflow analysis)	Понимание причин, влияющих на принятие клинических решений; определение и детальное изучение процессов, требующих автоматизации (требований пользователя); определение полезных функциональных возможностей и вариантов их реализации	Проводят оценку необходимых ресурсов на реализацию и обеспечение функционирования вариантов автоматизированной системы; Оценку преимуществ и недостатков каждого варианта; сопоставление требований пользователя и характеристик предлагаемой системы и выбор оптимального варианта; определение порядка оценки качества и условий приемки системы; оценку эффектов, получаемых от автоматизированной системы

Продолжение таблицы А.1

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
1, 2	Технические испытания	Функциональное тестирование (аналитическая валидация) (раздел 7)	Описательный анализ и расчетные методы	Функциональная полнота	В СИИ присутствует и соответствует документации набор функций
				Функциональная корректность	СИИ генерирует выходные данные с надлежащим уровнем точности (accuracy), а также повторяемости (repeatability) и воспроизводимости (reproducibility)
				Функциональная пригодность	Оценка степени функционального упрощения выполнения определенных задач
				Способность к самообучению	Оценка способности СИИ извлекать знания из накопленного опыта и применять их для улучшения качества поставленных задач
				Согласованность	СИИ соответствует стандартам или соглашениям, или положениям законов или подобных рекомендаций
		Уровень производительности (раздел 8)	Регистрационный метод (моделирование типовых задач СИИ)	Характер изменения во времени	Соответствие требованиям временных ресурсов на выполнение операции, а также производительность — количество однотипных задач за определенное время: время, затрачиваемое на каждую операцию в последовательности (workflow) на рекомендуемой конфигурации аппаратного оборудования
				Характер изменения ресурсов	Объем используемых ресурсов и продолжительности такого использования при выполнении функций: потребление ресурсов центрального и (или) графического процессора; возможность многопоточной работы; потребление оперативной памяти
				Производительные возможности	Степень соответствия требованиям таких параметров СИИ, как количество параллельно обрабатываемых наборов данных, количество параллельно работающих пользователей, емкость канала, пропускная способность по транзакциям

Продолжение таблицы А.1

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
1, 2	Технические испытания	Надежность (раздел 9)	Регистрационный метод (моделирование типовых задач), способность СИИ сохранять свой уровень качества функционирования при установленных условиях за установленный период времени	Стабильность	Вероятность ошибочного срабатывания СИИ при N циклах непрерывной обработки тестового набора медицинских изображений
				Устойчивость к ошибке	Способность СИИ поддерживать определенный уровень качества функционирования в случаях программных ошибок или нарушения определенного интерфейса
				Восстанавливаемость	Способность СИИ возобновить работу после возникновения ошибки
				Робастность	Способность СИИ демонстрировать требуемую точность выходных данных при наличии разного рода выбросов (соответствует тестированию на устойчивость к состязательным атакам и тестированию на способность и устойчивость работы с разнородными данными)
		Тестирования на устойчивость к состязательным атакам (adversarial attacks) (относится к метрике надежности)	Сравнительный статистический анализ	Оценка точности предсказаний алгоритма ИИ и (или) других метрик для каждой модальности в отдельности и для всех модальностей вместе в условиях состязательных примеров и в нормальных условиях	Моделируют состязательные примеры разных видов; проверяют их влияние на выходные данные СИИ; составляют ранжированный список видов состязательных примеров, упорядоченный по степени влияния на выходные данные СИИ

Продолжение таблицы А.1

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
1, 2	Технические испытания	Тестирование на способность и устойчивость работы с разнородными данными (относится к метрике надежности)	Моделирование типовых задач СИИ с различными наборами данных и диагностическими устройствами	Проверка соответствия технической документации, если требования к качеству входных данных определены	Формирование набора данных с изображениями с качеством, соответствующим документации, и низкого качества, наличием инородных предметов, артефактов и др.; тестирование СИИ путем подачи на вход набора данных; выполнение визуальной оценки выходных данных СИИ; если применимо, оценка метрик качества
				Тестирование на совместимость с диагностическим устройством	Проверка обработки данных с устройств, указанных в эксплуатационной документации производителя; тестирование на корректную работу с примерами, сложными к интерпретации экспертами; адаптируемость к новым типам данных при вводе СИИ в эксплуатацию; если применимо, оценка метрик качества
				Проверка корректности работы системы обнаружения и оповещения пользователя о подаче на вход СИИ изображений низкого качества	Формирование набора данных с изображениями низкого качества, наличием инородных предметов, артефактов и др.; тестирование СИИ путем подачи на вход набора данных; фиксирование наличия уведомления пользователя о поступлении на вход СИИ изображений низкого качества
		Тестирование на защищенность	Анализ программного кода	Проверка защиты от взлома, несанкционированного доступа и прочих внешних воздействий, а также защиты персональных данных	Проверяют разграничение пользовательского доступа, организацию пользовательской авторизации; проверяют, записывает ли СИИ персональные данные пользователей (пол, возраст, место работы и должность, клинические данные, семейное положение и др.) для постоянного или временного хранения; проверяют, требуется ли СИИ потенциально вредоносное дополнительное (стороннее) программное обеспечение для корректной работы; проверяют, требуется ли СИИ постоянное подключение к сети Интернет для корректной работы

Продолжение таблицы А.1

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
3, 4*	Клинические испытания (оценка клинической связи)	Анализ данных	Анализ имеющихся данных или получение новой информации	Обоснование наличия связи между выходными данными СИИ и его функциональным назначением	Примеры: литературные источники, документация, рекомендации профессионального сообщества, проведенные предварительные клинико-технические испытания, анализ выходных данных после апробации СИИ, выполнение клинического испытания (исследования)
		Неконтролируемые испытания (uncontrolled trials)	Анализ медицинских изображений, содержащих целевую патологию (анатомическую структуру)	Количественная оценка способности СИИ распознавать целевую анатомическую структуру	Используют для алгоритмов, предназначенных для высокоточной сегментации целевых анатомических структур; не рекомендуется для оценки диагностической точности алгоритма, поскольку данный тип исследований ассоциирован с существенной переоценкой этой метрики
		Рандомизированные контролируемые испытания с параллельными группами (parallel group design)	Группы сравнения: интерпретация медицинских изображений с использованием СИИ; интерпретация медицинских изображений без использования СИИ	Количественная оценка эффекта СИИ на принятие медицинских решений; количественная оценка диагностической точности СИИ и (или) других метрик	Участники исследования (медицинские сотрудники) остаются в группе сравнения на протяжении всего исследования; распределение участников между группами сравнения проводят случайным образом; необъяснимую вариабельность результатов ассоциируют с разницей между участниками исследования; групп сравнения может быть больше двух
		Рандомизированные контролируемые испытания с перекрестными группами (crossover design)	Испытание разбивают на два этапа. Группы сравнения: интерпретация медицинских изображений с использованием СИИ (первый этап); интерпретация медицинских изображений без использования СИИ (второй этап); обратный порядок	Количественная оценка эффекта СИИ на принятие медицинских решений; количественная оценка диагностической точности СИИ и (или) других метрик	Требуют меньшего объема выборки, чем испытания с параллельными группами; каждый участник выступает собственным контролем; меньшая вариабельность результатов по сравнению с испытаниями с параллельными группами; возможно включение периода паузы между этапами для снижения психологического эффекта использования СИИ; требуют более длительного времени проведения, чем испытания с параллельными группами

Окончание таблицы А.1

Этап	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
		Рандомизированные контролируемые факториальные испытания (factorial design)	Группы сравнения: интерпретация медицинских изображений с использованием СИИ 1; интерпретация медицинских изображений с использованием СИИ 2; интерпретация медицинских изображений с использованием СИИ 1 и СИИ 2; интерпретация медицинских изображений без использования СИИ	Сравнение двух алгоритмов искусственного интеллекта, предназначенных для решения одной и той же задачи	Позволяют оценить комбинированный эффект альтернативных СИИ; позволяют получить ответ сразу на несколько исследовательских вопросов; требуют меньшего объема выборки, чем испытания с параллельными группами
		Пре- и постимплементационные сравнения (pre-post comparison)	Статистический анализ	Мультицентровая скорректированная с учетом времени (time-adjusted) оценка эффекта внедрения СИИ на частоту детекции целевой патологии и (или) другие метрики	В качестве контроля используют ретроспективные данные для той же когорты пациентов до внедрения СИИ в медицинской организации, на базе которой проходят клинические испытания; потенциальные источники систематической ошибки: факторы с временной зависимостью (сезонные заболевания и др.), неочевидные тенденции развития и распространения заболевания
*Аналогичные типы исследований допускается проводить на фазах испытаний 1 и 2. В этом случае они носят название пилотных проектов, требуют значительно меньших размеров выборки и направлены на доказательство потенциальной работоспособности алгоритма в клинической медицине. Успешное завершение пилотного проекта (статистически значимый эффект от внедрения) не является доказательством клинической эффективности алгоритма и служит только обоснованием для проведения полноценных испытаний.					

Приложение Б
(справочное)

Содержание программы клинических испытаний системы искусственного интеллекта для анализа медицинских изображений

При проведении клинических испытаний СИИ, назначением которой является анализ медицинских изображений, составляют программу. Программа клинических испытаний должна удовлетворять требованиям действующих нормативных правовых актов, ГОСТ Р 59921.1.

В программе клинических испытаний, в том числе, должны быть отражены следующие данные:

1) название, в котором определяют дизайн испытания, целевую популяцию, тип алгоритма искусственного интеллекта, назначение СИИ;

2) дата испытания и версия СИИ;

3) детализированное описание назначения СИИ: тип/типы медицинских изображений, предполагаемая роль СИИ в клинической практике, конкретная решаемая задача; типы выходных данных и т. п.;

4) описание цели испытания, обоснование необходимости его проведения, включая анализ опубликованной релевантной литературы (научные статьи, патенты, грантовые заявки, в том числе зарубежные), указание потенциальной пользы и вреда использования СИИ. Данный пункт включает также проведение анализа клинической связи как этапа клинической оценки СИИ (ГОСТ Р 59921.1);

5) обоснование выбора компаратора (компараторов);

6) описание дизайна испытаний:

а) используют разные методологии в зависимости от фазы испытаний (таблица А.1 приложения А). СИИ в качестве медицинских изделий отличаются от лекарственных препаратов и медицинских приборов тем, что предназначены для воздействия на принятие решений медицинским работником. По этой причине оценка эффектов СИИ в клинической медицине не может проводиться независимо от предполагаемых пользователей (медицинских сотрудников), которые являются такими же участниками исследования, как пациенты и СИИ. Таким образом, в каждом испытании должно быть минимум две переменных, определяющих характеристики группы сравнения:

- распределение пациентов между группами сравнения на имеющих целевое заболевание и (или) детектируемый средствами медицинской визуализации признак (категория «случай») и с отсутствием заболевания и (или) детектируемого средствами медицинской визуализации признака (категория «контроль»). В зависимости от фазы испытаний вместо данных специально включенных в испытания пациентов допускается использовать медицинские изображения из специализированных наборов данных, в том числе находящихся в публичном доступе. Соотношение распределения пациентов (исследований) указывают в виде случай/контроль, например 1/1, 1/2 и т. д.; конкретное соотношение выбирают в зависимости от цели испытаний и распространенности заболевания (признака), приводя обоснование (пункт г);

- распределение медицинских специалистов между группами сравнения на проводящих интерпретацию медицинских изображений с использованием СИИ и без использования СИИ. Указывают соотношение численности между этими группами и уровень квалификации медицинских сотрудников (трудовой стаж в годах, если применимо — пройденное обучение по целевому медицинскому вмешательству, медицинской технологии, медицинскую специализацию), которые выбирают в зависимости от цели испытаний, приводя обоснование (пункт г);

б) также указывают гипотезу испытаний: превосходство, неуступающая эффективность, эквивалентность.

Гипотеза превосходства предполагает, что эффективность и (или) другая метрика исследуемой СИИ по выбранному критерию выше, чем у компаратора(ов).

Гипотеза неуступающей эффективности предполагает, что эффективность и (или) другая метрика исследуемой СИИ по выбранному критерию не хуже, чем у компаратора(ов).

Гипотеза эквивалентности предполагает, что эффективность и (или) другая метрика исследуемой СИИ по выбранному критерию является эквивалентной с компаратором(ами);

7) описание условий проведения испытаний (например, амбулаторное учреждение, стационар). Описание требований к аппаратному и программному обеспечению, необходимости дообучения алгоритма на локальных данных для успешной интеграции СИИ в цифровой контур организации, проводящей испытания. Существуют ограничения обобщаемости алгоритмов ИИ при их использовании вне среды разработки и обучения. В протоколе указывают:

- требует ли внедрение СИИ каких-либо устройств от конкретной фирмы-производителя;
- есть ли потребность в локальном вычислительном аппаратном обеспечении;
- существует ли потребность в обеспечении интеграции облачных решений, при необходимости с уточнением конкретной фирмы — поставщика услуг;
- при необходимости внесения любых изменений в алгоритм в рамках процедуры внедрения этот процесс должен быть описан;

8) описание критериев включения и исключения на уровне (а) участников исследования (пациентов), (б) участников исследования (медицинских сотрудников) и (в) входных данных:

а) критерии включения для участников исследования (пациентов) определяют целевую популяцию применения СИИ: тип и (или) тяжесть заболевания и (или) признака, наличие сочетанных заболеваний, выполненных диагностических процедур и др. показатели;

б) критерии включения для участников исследования (медицинских сотрудников) определяют предполагаемых пользователей СИИ, задавая уровень квалификации, специализации, согласно назначению СИИ;

в) под входными данными понимают данные, необходимые СИИ для выполнения предполагаемой функции. В программе указывают, есть ли минимальные требования к входным данным (разрешение изображения, формат данных и др.);

г) указывают, как будет оцениваться соответствие критериям включения.

Примеры

1 Если пациент соответствует критериям включения на уровне участников исследования, но качество полученного исследования компьютерной томографии по каким-либо причинам было неудовлетворительным для использования СИИ, это необходимо расценивать как критерий исключения на уровне входных данных.

2 Участники должны иметь компетенции в интерпретации исследований компьютерной томографии грудной клетки согласно системе LUNG-Rads, версия 1.1. Проверку компетенции проводят тестированием на предварительно размеченном наборе данных компьютерной томографии, содержащем примеры рака легкого в разных стадиях развития, а также исследования без патологий;

9) описание вмешательства для каждой группы сравнения в деталях, необходимых для воспроизведения результатов испытаний:

а) указывают, какая версия СИИ будет использована. СИИ, как правило, подвергаются неоднократному изменению и обновлению программного кода в течение жизненного цикла. В протоколе указывают, какую версию используют в испытаниях, и является ли она той же версией, которую использовали в предыдущих испытаниях, на основании которых обосновывали необходимость проведения испытаний. Когда это применимо, указывают, какие изменения были внесены в текущую версию и обоснования внесения этих изменений;

б) указывают процедуру получения, отбора и предварительной обработки входных данных для СИИ (формирование набора данных, см. ГОСТ Р 59921.5);

в) указывают процедуру оценки качества наборов данных и действия, предпринимаемые с низкокачественными (не соответствующими минимальным требованиям к качеству) или недостающими данными. Низкокачественные данные могут осложнить интерпретацию также медицинским сотрудникам, не использующим СИИ. По этой причине необходимо, в случае применимости, дополнительно указывать такую информацию для контрольного вмешательства;

г) указывают, существует ли необходимость участия медицинского сотрудника в обработке (подготовке) наборов данных для испытания СИИ, и требуемую квалификацию медицинского сотрудника, включая обучение и инструктаж по работе с СИИ.

Пример — Медицинский сотрудник должен отметить исследуемую область на медицинском изображении, которую затем будет анализировать СИИ;

д) описание выходных данных СИИ. СИИ может предоставлять данные о диагностической классификации, вероятности существования либо развития патологии, рекомендуемые мероприятия либо другую информацию. Тип выводимых данных имеет прямую связь с эксплуатационными свойствами СИИ, а также медицинскими решениями, на которые она может повлиять;

10) описание ожидаемых первичных, вторичных и других исходов, включая конкретный анализируемый признак, метрику анализа, метод объединения данных.

Пример — Анализируемый признак: легочный узел. Метрика анализа: линейные размеры. Метод объединения данных: среднее значение со стандартным отклонением;

11) диаграмма дизайна испытаний, на которой показаны этапы включения участников в испытание, вмешательства (в том числе пауза между этапами в случае использования перекрестных групп), и анализа данных (см. рисунок Б.1).

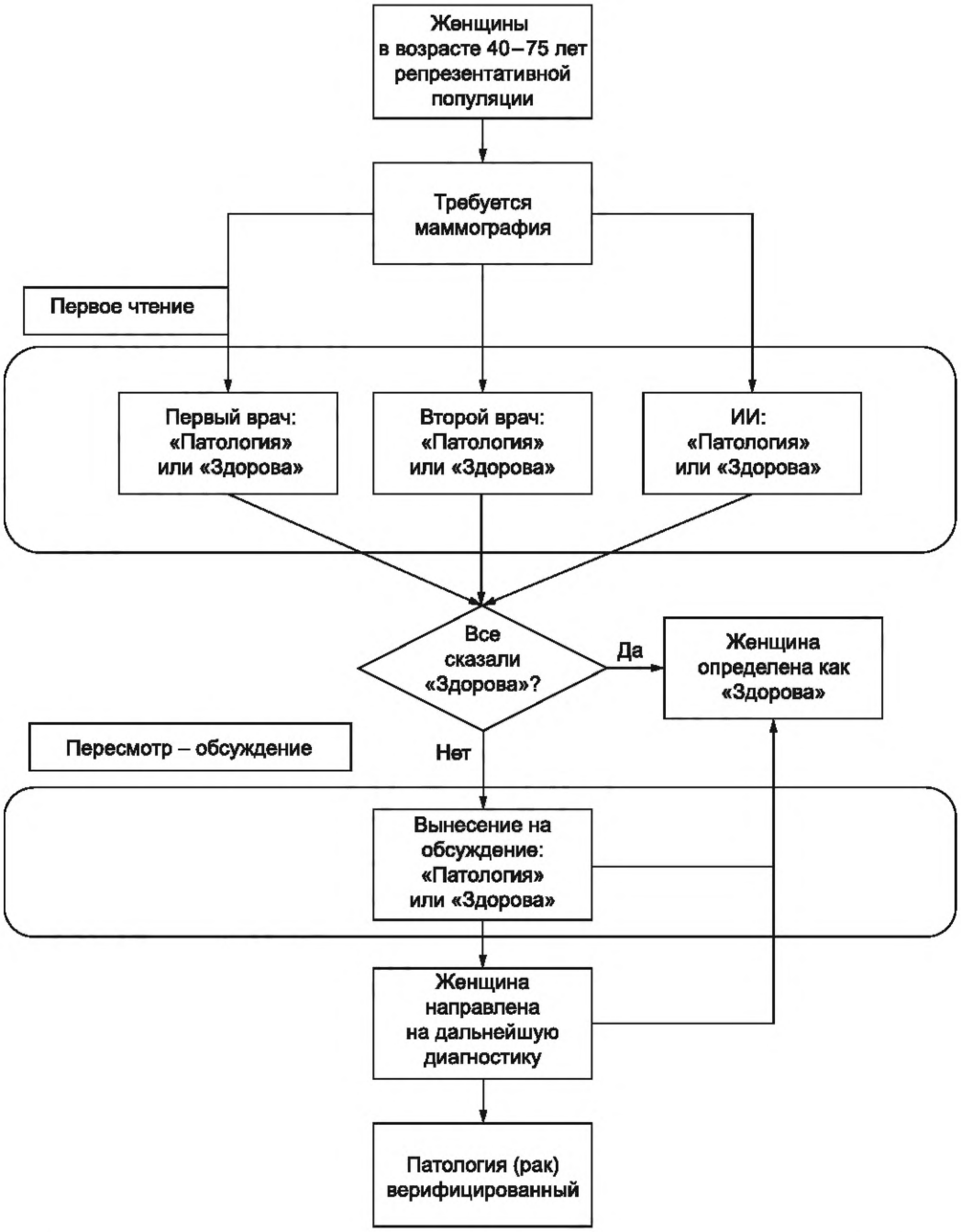


Рисунок Б.1 — Пример диаграммы дизайна испытания (клинические испытания WHO ID STGKS001)

12) оценка объема выборки пациентов (размера набора данных) согласно проверяемой гипотезе испытания. Расчет объема выборки проводят по первичному исходу исходя из предположения, что недостающих данных не будет. Явным образом указывают клинические и статистические предположения, на которые опирались при вычислениях (если применимо): доля популяции с целевым заболеванием (признаком) либо среднее значение со стандартным отклонением, статистический критерий, величина ошибки первого и второго рода;

13) оценка числа медицинских специалистов, принимающих участие в исследовании, с описанием ролей и указанием квалификации и специализации (согласно ГОСТ Р 59921.5—2022, пункт 7.2.1). Описывают предусмотренные меры в случае отказа или невозможности дальнейшего участия медицинского специалиста в исследовании;

14) описание подхода к распределению анализируемых данных между медицинскими специалистами в группах сравнения. Как правило, необходимые размеры выборок являются слишком большими, чтобы их мог полностью обработать один медицинский специалист. Допускается частичное распределение набора данных между участниками исследования с соблюдением обязательных условий:

- каждое исследование должно быть просмотрено хотя бы дважды разными участниками из группы, проводящей интерпретацию с использованием СИИ, и дважды разными участниками из группы, не использующей СИИ при интерпретации медицинских изображений. Это позволит оценить согласие между экспертами и влияние СИИ на этот показатель;

- распределение исследований между участниками (медицинскими сотрудниками) должно проводиться случайным образом.

Пример — Случайное распределение данных между участниками испытаний выполняют путем генерации компьютером случайного распределения идентификационных номеров исследований в наборе данных.

Если дизайн испытаний подразумевает группировку данных (по медицинской организации, полу и (или) возрасту пациентов, стадии заболевания и др.), указывают список факторов, по которым проводили группировку (стратификацию) данных, с обоснованием этих факторов;

15) клинические испытания проводят с применением верифицированного набора данных (ГОСТ Р 59921.5—2022, пункт 4.2.3). В программе указывают, проводится ли ослепление участников испытания к истинным данным с конкретным указанием, для кого и как проводят ослепление: для пациентов, СИИ, медицинских специалистов, специалистов по обработке данных. Возможно также проводить ослепление участников испытания к проверяемой гипотезе.

Пример — При проверке гипотезы эквивалентности допускается информировать участников, что проводится проверка гипотезы превосходства одного из вмешательств;

16) методика сбора набора данных. Указывают, каким образом будет проводиться оценка целевых признаков на медицинских изображениях для первичного, вторичного и других исходов исследования, с указанием специализированных программных инструментов в случае их использования.

Пример — Медицинские сотрудники ищут на КТ-снимках легочные очаги размерами от 4 до 30 мм, сохраняя такую информацию о находках, как локализация легочного очага (положение центра находки по двум измерениям на изображении и номеру среза); диаметр находки; тип легочного очага (солидный, полусолидный или очаг по типу матового стекла) с помощью программного обеспечения FAnTom. Рекомендовано не отмечать кальцинированные и перифиссуральные очаги в легких, а также не отмечать более пяти крупнейших легочных очагов на одном КТ-снимке;

17) управление данными. Описывают планы по хранению данных исследования, обеспечению их безопасности и защищенности, а также любые мероприятия, направленные на обеспечение качества данных с учетом ГОСТ Р 59921.5—2022, подраздел 7.4.

Пример — Данные будут храниться в цифровом контуре медицинской организации, на базе которой будут проходить испытания. При необходимости для промежуточного или итогового анализа данные будут извлечены для исследовательских целей под ответственность главного исследователя. Перед проведением любого анализа, в том числе статистического, будет проведена анонимизация (псевдонимизация) данных. Каждый параметр данных будет проходить проверку на достоверность по типу и диапазону;

18) статистические методы. В программе должны быть указаны и описаны планируемые методы статистического анализа [15]. Должны быть отражены все предполагаемые методы анализа при сравнении групп исследования. Результаты испытаний могут подвергаться существенному влиянию со стороны методов статистического анализа. Если для конкретного исхода, особенно первичного, предполагают использовать более одной стратегии анализа, возникает потенциальная возможность недопустимого выборочного представления наиболее ярких и интересных результатов. В программе указывают основной метод статистического анализа первичного исхода.

Если применимо, указывают методы дополнительного статистического анализа, предназначенного для подгрупп пациентов. Анализ подгрупп позволяет определить статистически значимые различия получаемых выходных данных испытываемой СИИ для разных категорий пациентов с целевой патологией и (или) признаком, в конечном итоге обеспечивая персонализированный подход в медицине. Однако некорректно проведенный анализ подгрупп, а также выборочное представление результатов, особенно в случае, если обработку данных проводили посредством апостериорного анализа, сопряжены с риском сомнительных (ложных) выводов;

19) мероприятия, направленные на согласование клинического испытания в этическом комитете. В программе необходимо указать, было ли получено согласование этического комитета, с указанием даты и номера согласования и названия этического комитета, либо наметить планы по получению согласования.

В случае если необходимо получение информированного добровольного согласия на сбор и использование данных от пациентов, соответствующую форму прикладывают к программе;

20) декларация интересов. Указывают (при наличии) конфликт финансовых и иных интересов для коллектива, проводящего испытание, и для всех медицинских организаций, в рамках которых оно будет проходить.

Приложение В
(справочное)

Пример заполнения протокола по результатам функционального тестирования системы искусственного интеллекта

Таблица В.1

Наименование параметра	Данные, указанные в технической документации	Результат в соответствии с порядковым номером исследования			Коммен-тарий
		1	2	…	
Возможность приоритизации (триаж)	Имеется	Соответствует	Соответствует		
Наличие графического обозначения	Имеется	Соответствует	Соответствует		
Наличие дополнительной серии от СИИ	Имеется	Соответствует	Соответствует		
Название дополнительной серии	Имеется	Соответствует	Соответствует		
Возможность синхронизации серий	Не применимо	—	—		
Отображение информации о СИИ	Имеется	Соответствует	Соответствует		
Отображение вероятности(ей) находки(ок)	Отсутствует	—	—		
Указание категории находок	Не применимо	—	—		
Возможность отключения маркировки	Отсутствует	—	—		
Наличие DICOM SR 1)	Имеется	Соответствует	Соответствует		
Создание шаблона протокола	Имеется	—	—		
Возможность сравнения исследований в динамике	Не применимо	—	—		
Прочее (указать)	—	N	#		

Библиография

[1] ПНСТ 553—2021 Информационные технологии. Искусственный интеллект. Термины и определения

[2] ISO/IEC TR 29119-11:2020(E) Системная и программная инженерия. Тестирование программного обеспечения. Часть 11. Руководящие указания по тестированию систем искусственного интеллекта (Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems)

[3] ИСО 9241-171:2008 Эргономика взаимодействия человека и системы. Часть 171. Руководство по доступности программного обеспечения (Ergonomics of human-system interaction — Part 171: Guidance on software accessibility)

[4] Liu X., Cruz Rivera S., Moher D. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 26, 1364—1374 (2020)

[5] Sounderajah V., Ashrafian H., Golub R.M. On behalf of the STARD-AI Steering Committee, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11:e047709

[6] Collins G.S., Dhiman P, Andaur Navarro C.L. et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008

[7] IMDRF/SaMD WG/N41 — Software as a Medical Device (SaMD): Clinical Evaluation, 2017

[8] Fenster A., Chiu B. Evaluation of segmentation algorithms for medical imaging. In: Conf Proc IEEE Eng Med Biol Soc. Shanghai; 2005. p. 7186—189

[9] Taha A.A., Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging (2015) 15:29

[10] Hirano H., Minagi A. and Takemoto K. Universal adversarial attacks on deep neural networks for medical image classification. *BMC Med Imaging* 21, 9 (2021)

[11] Apostolidis K.D., Papakostas G.A. A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis. Electronics 2021, 10, 2132

[12] С.П. Морозов и др. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика)/Серия «Лучшие практики лучевой и инструментальной диагностики». — М., 2019. — Вып. 57. — 51 с.

[13] Defending Deep Learning-Based Biomedical Image Segmentation from Adversarial Attacks: A Low-Cost Frequency Refinement Approach. Available online: gost.e-ecolog.ru (accessed on 4 June 2021)

[14] Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы» (ФГБУ «ВНИИИМТ» Росздравнадзора, утверждены 12 февраля 2021 г.). 2021, 33 с.

[15] Методические рекомендации по оценке качества статистического анализа в клинических исследованиях (ФГБУ «ЦЭККМП» Минздрава России, утверждены 29 декабря 2017 г.). 2017, 34 с.

УДК 615.841:006.354 ОКС 11.040.01

Ключевые слова: система искусственного интеллекта, искусственный интеллект, клиническая медицина, алгоритмы обработки медицинских изображений, методы испытаний

Редактор *Л.В. Коретникова*
Технический редактор *И.Е. Черепкова*
Корректор *Е.Д. Дульнева*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 19.10.2022. Подписано в печать 27.10.2022. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 4,18. Уч.-изд. л. 3,76.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «РСТ»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru