

---

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО**

**ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ**

---



**НАЦИОНАЛЬНЫЙ  
СТАНДАРТ  
РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

**ГОСТ Р**

**20**

---

**СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ  
МЕДИЦИНЕ**

**Алгоритмы анализа медицинских изображений**

**Методы испытаний**

**Общие требования**

**Настоящий проект стандарта не подлежит применению до его утверждения**

**Москва**

**20**

## Предисловие

1 РАЗРАБОТАН Федеральным государственным бюджетным учреждением «Российский институт стандартизации» (ФГБУ «РСТ»), Государственным бюджетным учреждением здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ») в соответствии с Программой национальной стандартизации на 2021 г.

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 \_\_\_\_\_ № \_\_\_\_\_

### 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 ([www.rst.gov.ru](http://www.rst.gov.ru))*

© \_\_\_\_\_, оформление, 20\_\_\_\_

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.

# НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Системы искусственного интеллекта в клинической медицине.

Алгоритмы анализа медицинских изображений

Методы испытаний

Общие требования

---

Дата введения — 20 — —

## 1 Область применения и цель

### 1.1 Область применения

Использование систем искусственного интеллекта для принятия врачебных решений в клинических условиях сопряжено с необходимостью соответствовать требованиям, выдвигаемым к программному обеспечению как медицинскому изделию. Должны быть проведены испытания для подтверждения воспроизводимости, надёжности, безопасности использования программного обеспечения и его эффективности в соответствии с предназначенным применением.

Настоящий стандарт устанавливает методы испытаний алгоритмов анализа медицинских изображений в системах искусственного интеллекта.

В область применения настоящего стандарта не входят испытания системы искусственного интеллекта как программного обеспечения.

### 1.2 Цель

Настоящий стандарт устанавливает требования к методам испытаний алгоритмов анализа медицинских изображений в системах искусственного интеллекта в клинической медицине.

---

Издание официальное

## 2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие документы:

ГОСТ Р 59276-2020 Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения

ИСО/МЭК Руководство 2 Стандартизация и связанные виды деятельности. Общий словарь (ISO/IEC Guide 2, Standardization and related kinds of activity - General vocabulary)

ИСО/МЭК Руководство 99 Международный метрологический словарь. Основные и общие понятия и соответствующие термины (ISO/IEC Guide 99, International vocabulary of metrology - Basic and general concepts and associated terms (VIM))

ИСО/МЭК 17000 Оценка соответствия. Словарь и общие требования (ISO/IEC 17000, Conformity assessment - Vocabulary and general requirements)

ГОСТ Р ИСО/МЭК 12207-2010. Информационная технология. Системная и программная инженерия. Процессы жизненного цикла программных средств.

ГОСТ Р ИСО 14155-2014. Клинические исследования. Надлежащая клиническая практика.

ГОСТ ISO 14971-2011. Изделия медицинские. Применение менеджмента риска к медицинским изделиям.

ГОСТ 57668-2017. Пространственные данные. Метаданные. Часть 1. Основные положения.

ГОСТ 58451-2019. Изделия медицинские. Обслуживание техническое. Основные положения.

ГОСТ Р 52379-2005. Надлежащая клиническая практика.

ГОСТ Р 56429-2015 (CNTF/SG5/N2R8:2007). Изделия медицинские. Клиническая оценка.

ГОСТ Р 57193-2016. Системная и программная инженерия. Процессы жизненного цикла систем.

ГОСТ Р 57501-2017. Техническое обслуживание медицинских изделий. Требования для государственных закупок.

ГОСТ Р 59277-2020. Системы искусственного интеллекта. Классификация систем искусственного интеллекта.

ГОСТ Р МЭК 62304-2013. Изделия медицинские. Программное обеспечение. Процессы жизненного цикла.

ГОСТ Р ИСО/МЭК 90003–2014. Разработка программных продуктов. Руководящие указания по применению ИСО 9001:2008 при разработке программных продуктов

**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.

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

**3.3 Ответственный исследователь:** Квалифицированное лицо, ответственное за проведение клинического исследования в исследовательском центре.

**3.4 Отказ:** Отклонение поведения системы искусственного интеллекта от ожидаемого выполнения предписанных ей функций.

**3.5 Ошибка:** Состояние системы искусственного интеллекта, которое вызывает отказ.

**3.6 Этический комитет:** Независимый комитет, ответственный за анализ клинического исследования в целях защиты прав и обеспечения безопасности и благополучия людей, принимающих участие в клиническом исследовании.

## **4 Обозначения и сокращения**

В настоящем стандарте применены следующие обозначения и сокращения:

ИИ — искусственный интеллект;

СИИ — система искусственного интеллекта.

## **5 Методы испытаний алгоритмов анализа медицинских изображений и основные этапы**

### **5.1 Общие положения**

При проведении испытаний алгоритмов анализа медицинских изображений составляют протокол, в котором должны быть отражены следующие данные:

- а) Название, в котором определяют дизайн испытания, целевую популяцию, тип алгоритма искусственного интеллекта, предназначенное применение СИИ;
- б) Дата испытания и версия программного обеспечения;
- в) Детализированное описание предназначенного применения СИИ: тип/типы медицинских изображений, предполагаемая роль СИИ в клинической практике, конкретная решаемая задача;
- г) 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;
- д) Обоснование выбора компаратора (компараторов);
- е) Описание дизайна испытаний.

1) Используют разные методологии в зависимости от фазы испытаний (Таблица А.1, Приложение А). СИИ в качестве медицинских устройств 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 анализируемая технология ИИ. 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 включённых 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 распространённости заболевания (признака), приводя обоснование (п. 6.12);

- распределение медицинских сотрудников 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 (п. XX).

2) 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 (компараторов).

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

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

ж) 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.

з) Описание критериев включения и исключения на уровне (а) участников исследования (пациентов), (б) участников исследования (медицинских сотрудников) и (в) входных данных.

1) Критерии включения для участников исследования (пациентов) определяют целевую демографию СИИ: тип и (или) тяжесть заболевания и (или) признака, наличие сочетанных заболеваний, диагностические процедуры и др. показатели.

2) Критерии включения для участников исследования (медицинских сотрудников) определяют предполагаемых пользователей СИИ, задавая уровень квалификации, имеющей отношение к изучаемому вмешательству.

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

4) Указывают, как будет оцениваться соответствие критериям включения.

*Пример 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 (п. 9.3), это необходимо расценивать как критерий исключения на уровне входных данных.*

*Пример 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 лёгкого в разных стадиях развития, а также исследования без патологий".*

и) Описание вмешательства для каждой группы сравнения в деталях, необходимых для воспроизведения результатов исследования.

1) Указывают, какая версия алгоритма ИИ будет использована. СИИ, как правило, подвергаются неоднократному изменению и обновлению программного кода течение своего жизненного цикла. В протоколе указывают, какую версию используют в испытаниях, и является ли она той же версией, которую использовали в предыдущих исследованиях, на основании которых обосновывали необходимость проведения

испытаний. Когда это применимо, указывают, какие изменения были внесены в текущую версию и обоснования внесения этих изменений.

2) Указывают процедуру получения, отбора и предварительной обработки входных данных для СИИ.

3) Указывают процедуру оценки качества 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.

4) Указывают, существует ли необходимость участия медицинского сотрудника в обработке (подготовке) входных данных 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 алгоритм ИИ.***

5) Описание выводимых данных СИИ-вмешательства. 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.

к) Описывают ожидаемые первичные, вторичные и другие исходы СИИ-вмешательства, включая конкретный анализируемый признак, метрику анализа, метод объединения данных.

***Пример – анализируемый признак: лёгочный узел. Метрика анализа: линейные размеры. Метод объединения данных: среднее значение со стандартным отклонением.***

л) Диаграмма дизайна исследования, на которой показаны этапы включения участников в исследование, вмешательства (в том числе пауза между этапами в случае использования перекрёстных групп), и анализа данных (Рисунок 1).

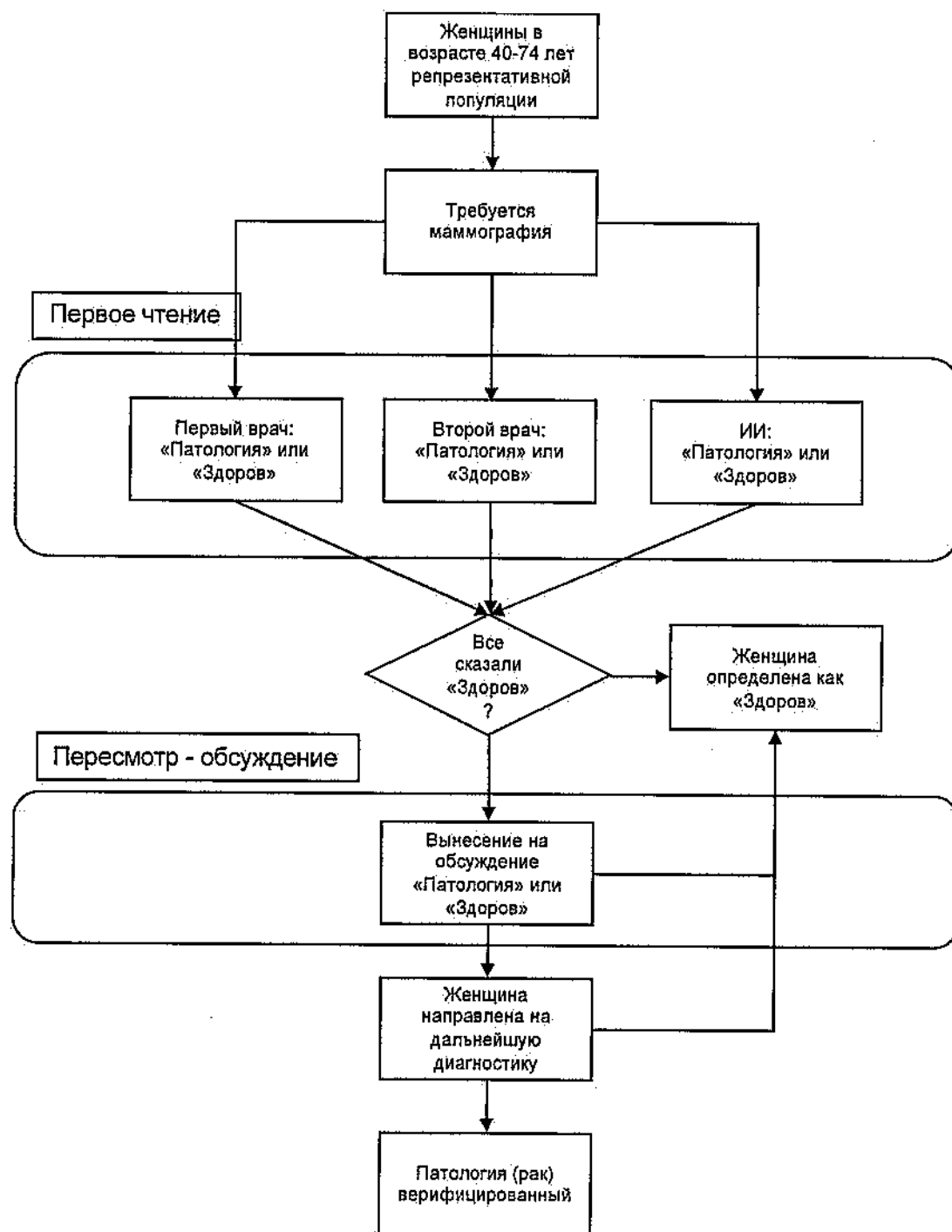


Рисунок 1. Пример диаграммы дизайна исследования (клинические испытания WHO ID STGKS001)

м) Оценка объема выборки пациентов (размера набора данных) согласно проверяемой гипотезе исследования. 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.

н) Оценка числа медицинских сотрудников, принимающих участие в исследовании, с описанием ролей и указанием квалификации и специализации. Описывают предусмотренные меры в случае отказа или невозможности дальнейшего участия медицинского сотрудника в исследовании.

о) Описание подхода к распределению анализируемых данных между медицинскими сотрудниками в группах сравнения. Как правило, необходимые размеры выборок являются слишком большими, чтобы их мог полностью обработать один медицинский сотрудник. 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 – сгенерированное компьютером случайное распределение 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.

п) Если предполагается использование референсного стандарта, указывают, проводится ли ослепление участников испытания к этим результатам с конкретным указанием, для кого и как проводят ослепление: для пациентов, СИИ, медицинских сотрудников, 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.***

р) Методика сбора 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 лёгочные очаги размерами от 4 до 30 мм, сохраняя такую информацию о находках, как локализация лёгочного очага (положение центра находки по двум измерениям на изображении и номеру среза); диаметр находки; тип лёгочного очага (солидный, полусолидный или очаг по типу матового стекла) с помощью программного обеспечения FAnTom. Рекомендовано не отмечать кальцинированные и перифиссуральные очаги в лёгких, а также не отмечать более пяти крупнейших лёгочных очагов на одном КТ-снимке."*

с) Управление данными. Описывают планы по хранению данных исследования, обеспечению их безопасности и защищённости, а также любые мероприятия, направленные на обеспечение качества данных.

*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.*

т) Статистические методы. В протоколе должны быть указаны и описаны планируемые методы статистического анализа в случае, если их возможно предусмотреть [1]. 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 протоколе указывают основной метод статистического анализа первичного исхода.

Указывают планы по дополнительному статистическому анализу, предназначенному для подгрупп пациентов. Анализ подгрупп позволяет определить статистически значимые различия испытываемой технологии ИИ 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.

у) Мероприятия, направленные на согласование исследования в этическом комитете. В протоколе 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 протоколу.

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

х) Протокол заверяют в установленном порядке в организации коллектива, проводящего испытания, с указанием даты заверения. Любые изменения в протоколе испытаний отражают в поправках к протоколу, которые также должны быть согласованы с этическим комитетом.

5.2 Испытания алгоритмов анализа медицинских изображений проводят в соответствии с протоколом испытаний.

## **6. Функциональный тест**

### **6.1 Требования к наборам данных**

Функциональное тестирование проводится на небольшом наборе данных, включающим изображения с отсутствием целевой патологии, изображения с целевой патологией, а также изображения с дефектами. Под дефектами подразумеваются нестандартное для СИИ содержимое результатов диагностических исследований: некорректная или обрезанная анатомическая область, зашумленные изображения и т.п. Это необходимо для проверки надежности и стабильности работы СИИ.

## **6.2 Порядок выполнения тестирования**

6.2.1 Тестирование проводится независимым коллективом исследователей; в противном случае в протоколе указывают источники конфликта интересов (см. пункт 6.19).

6.2.2 Тестирование проводят в согласии с протоколом испытаний. При проведении тестирования документируют все программные и прочие ошибки, возникшие при работе алгоритма ИИ.

6.2.3 Критерии оценки определяют в соответствии с предназначением автоматизированной системы и заявленными характеристиками.

## **6.3 Анализ и оценка результатов тестирования**

6.3.1 По завершении тестирования составляют отчёт, в котором заключают, готова ли СИИ к следующей фазе испытаний. Анализ и оценку результатов тестирования проводят совместно с разработчиками СИИ для того, чтобы определить, какие из найденных ошибок необходимо исправить незамедлительно, какие не способны повлиять на клиническую эффективность алгоритма, и какие были обнаружены некорректно (автоматизированная система сработала правильно).

6.3.2 При исправлении отмеченных дефектов тестирование проводят повторно.

6.3.3 В случае прохождения СИИ этапа функционального тестирования вся документация, имеющая отношение к тестированию, архивируется для использования в качестве справочных данных при последующих испытаниях.

## **7 Нагрузочные тестирования**

### **7.1 Требования к наборам данных**

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

### **7.2 Порядок выполнения тестирования**

7.2.1 Надёжность автоматизированной системы определяют как вероятность ошибочного срабатывания при N циклов непрерывной обработки тестового набора данных (медицинских изображений):

Вероятность == (Число ошибочных срабатываний)/(общее число примеров в тестовом наборе данных)

Число циклов  $N$  определяется разработчиком автоматизированной системы с приведением обоснования, но не может быть меньше 1000 циклов.

7.2.2 Для каждой выявленной ошибки определяют параметры: тип ошибки, триггер ошибки, влияние ошибки. Определяют время работы СИИ до момента возникновения ошибки и её воспроизводимость.

### **7.3 Анализ и оценка результатов тестирования**

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

### **7.4 Критерии оценки (метрики)**

7.4.1 Среднее время наработки на отказ, определяемое как отношение суммы интервалов безотказной работы вне зависимости от типа ошибки к количеству отказов в системе.

7.4.2 Устойчивость к ошибкам и отказам, 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.

7.4.3 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.

## **8. Тестирования на защищенность**

### **8.1 Порядок выполнения тестирования**

8.1.1 Тестирование на защищённость проводят силами квалифицированных специалистов и (или) с использованием специализированного программного обеспечения.

### **8.2 Анализ и оценка результатов тестирования**

8.2.1 Оценку рисков проводят для идентификации, анализа и классификации рисков, связанных с защищённостью СИИ. Оценка рисков позволяет понять, какие из



угроз являются наиболее вредоносными и требующими обязательного исправления программного кода.

## **9 Тестирование на разнородность входных данных**

### **9.1 Требования к наборам данных**

9.1.1 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.

9.1.2 Наборы данных должны включать данные, полученные на разной аппаратуре, исследования с которой СИИ потенциально способна обработать.

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

9.1.4 Допускается включение в набор данных зашумлённых или искажённых данных (перевернутых снимков, снимков, содержащих посторонние объекты).

9.1.5 Наборы данных должны содержать анатомическую область, соответствующую назначению СИИ с точки зрения решаемых клинических задач.

### **9.2 Порядок выполнения тестирования**

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

### **9.3 Анализ и оценка результатов тестирования**

9.3.1 Составляют отчёт об испытаниях, в котором перечисляют формат и тип входных данных, наименования медицинского оборудования, размеры и разрешения изображений, типы зашумлённых и искажённых изображений, которые СИИ смогла и не смогла обработать.

Приложение А  
(обязательное)

Дизайн испытаний

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

Фаза испытаний	Дизайн испытания	Исследование	Метод	Задачи	Комментарии
Фаза 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)
Фаза 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 сотрудников при выполнении данной задачи; Приемлемые значения измеряемых метрик зависят от клинических последствий возможных ошибок.
Фаза 0, 1		Популяционное исследование	Наблюдения и опросы; Анализ последовательности операций (workflow analysis).	Понимание причин, влияющих на принятие клинических решений; Определение и детальное изучение процессов, изучение процессов,	Проводят оценку необходимых ресурсов на реализацию и обеспечение функционирования вариантов

				<p>требующих автоматизации (требований пользователя);  Определение полезных функциональных возможностей и вариантов их реализации.</p>	<p>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</p>
Фаза 1, 2	Доказательство работоспособности (аналитическая валидация)	Функциональное тестирование	Описательный анализ	<p>Проверка выводимых данных СИИ</p>	<p>Выводимые данные СИИ должны соответствовать заявленным в технической документации</p>
				<p>Проверка типа принимаемого решения</p>	<p>Тип (бинарное, вероятностное, локализация, классификация) и формат возвращаемого СИИ заключения должны соответствовать назначению СИИ с точки зрения решаемых клинических задач</p>
				<p>Проверка количества заключений СИИ</p>	<p>СИИ может формировать одно или больше заключений (многофакторное заключение); набор заключений должен соответствовать назначению СИИ с точки зрения</p>

				решаемых клинических задач	
				Входные данные для систем анализа медицинских изображений могут быть представлены статическими изображениями анализируемой области; сериями двумерных изображений, соответствующих сечениям анализируемой области; содержать метаданные. Формат входных данных, организация метаданных и анализируемая область (анатомическая структура) должны соответствовать технической документации.	Проверка формата входных данных и типа анализируемого объекта
				Потребление ресурсов центрального и (или) графического процессора; Возможность многопоточной работы; Потребление оперативной памяти; Использование дискового пространства для создания и хранения постоянных и временных файлов; Потребление сетевых ресурсов	Проверка производительности СИИ
				Время, затрачиваемое на каждую операцию в последовательности (workflow) на рекомендуемой	Проверка времени обработки запроса
			Нагрузочные тестирования	Моделирование типовых задач СИИ при ожидаемой и повышенной нагрузке системы (большие объёмы обрабатываемых данных, одновременная работа нескольких пользователей)	

				конфигурации аппаратного оборудования (см. Пункт 6.7)
			Проверка надёжности	Вероятность ошибочного срабатывания СИИ при N циклов непрерывной обработки тестировочного набора медицинских изображений (см Пункт 9)
Тестирования на устойчивость к состязательным атакам (adversarial attacks)	Сравнительный статистический анализ	Оценка точности предсказаний алгоритма ИИ и (или) других метрик для каждой модальности и для всех модальностей вместе в условиях состязательных атак и в нормальных условиях	Применимо только для алгоритмов, использующих глубокое обучение; Составляют ранжированный список состязательных атак, упорядоченный по степени влияния на алгоритм; Составляют ранжированный список модальностей, упорядоченный по устойчивости к состязательным атакам	
Тестирование на защищённость	Анализ программного кода	Проверка защиты от взлома, несанкционированного доступа и прочих внешних воздействий, а также приватности 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
	Тестирование на разнородность входных данных	Моделирование типовых задач СИИ с различными наборами данных и диагностическими устройствами	Проверка соответствия технической документации, если требования к качеству входных данных определены	Проверка способности обработки некорректных данных производится два последовательно два индекс-теста на выборке с наличием искаженных данных и на выборке без искаженных данных	
				Проверка обработки данных с устройств, указанных в эксплуатационной документации; Тестирование на корректную работу с примерами, сложными к интерпретации экспертами; Адаптируемость к новым типам данных при вводе СИИ в эксплуатацию.	
Фаза 3,4*	Клинические испытания (клиническая валидация)	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 перекрёстными группами (crossover design)	Испытание разбивают на два этапа. Группы сравнения: Интерпретация медицинских изображений с использованием СИИ (первый этап), интерпретация медицинских изображений без использования ИИ (второй этап); Обратный порядок.	Количественная оценка эффекта СИИ на принятие медицинских решений; Количественная оценка диагностической точности СИИ и (или) других метрик	Требуют меньшего объёма 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.
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 объёма выборки, чем испытания с параллельными группами;	
Пре- и пост-имплементационные сравнения (pre-post comparison)	Статистический анализ	Мультицентровая скорректированная с учётом 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 со временной зависимостью (сезонные заболевания и др.), неочевидные тенденции развития и распространения заболеваний.	

\*Аналогичные типы исследований допускается проводить на Фазах испытаний 1 и 2. В этом случае они несут название пилотных проектов, требуют значительно меньших размеров выборок, 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.

## Библиография

[1] Методические рекомендации по оценке качества статистического анализа в клинических исследованиях.

---

УДК 615.841:006.354

ОКС 11.040.01

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

---

Руководитель организации-разработчика  
ГБУЗ «НПКЦ ДиТ ДЗМ»

Руководитель  
разработки

Директор \_\_\_\_\_

A handwritten signature in blue ink, consisting of several loops and a long vertical stroke, positioned over the signature line.

С.П. Морозов