

ГБУЗ «НАУЧНО-ПРАКТИЧЕСКИЙ КЛИНИЧЕСКИЙ ЦЕНТР ДИАГНОСТИКИ И
ТЕЛЕМЕДИЦИНСКИХ ТЕХНОЛОГИЙ ДЕПАРТАМЕНТА ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ»

ЛУЧШИЕ ПРАКТИКИ ЛУЧЕВОЙ И ИНСТРУМЕНТАЛЬНОЙ ДИАГНОСТИКИ



КЛИНИЧЕСКИЕ ИСПЫТАНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ (ЛУЧЕВАЯ ДИАГНОСТИКА)

Москва
2019



**ПРАВИТЕЛЬСТВО МОСКВЫ
ДЕПАРТАМЕНТ ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ**

**ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ «НАУЧНО-ПРАКТИЧЕСКИЙ КЛИНИЧЕСКИЙ ЦЕНТР
ДИАГНОСТИКИ И ТЕЛЕМЕДИЦИНСКИХ ТЕХНОЛОГИЙ ДЕПАРТАМЕНТА
ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ»**

СОГЛАСОВАНО

Главный внештатный специалист
Департамента здравоохранения
города Москвы по лучевой и
инструментальной диагностике


С.П. Морозов

«24» _____ 2019 года

РЕКОМЕНДОВАНО

Экспертным советом по науке
Департамента здравоохранения
города Москвы № 8



_____ 2019 года

**КЛИНИЧЕСКИЕ ИСПЫТАНИЯ
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ
ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ
(ЛУЧЕВАЯ ДИАГНОСТИКА)**

Методические рекомендации № 43

Организация-разработчик:

Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»

Составители:

Морозов С.П. – д.м.н., профессор, главный внештатный специалист по лучевой и инструментальной диагностике Минздрава России в ЦФО РФ, главный внештатный специалист по лучевой и инструментальной диагностике ДЗМ, директор ГБУЗ «НПКЦ ДиТ ДЗМ»

Владимирский А.В. – д.м.н., заместитель директора по научной работе ГБУЗ «НПКЦ ДиТ ДЗМ»

Кляшторный В.Г. – к.б.н., аналитик отдела координации научной деятельности ГБУЗ «НПКЦ ДиТ ДЗМ»

Андрейченко А.Е. – к.ф.-м.н., старший научный сотрудник отдела разработки средств контроля и технического мониторинга ГБУЗ «НПКЦ ДиТ ДЗМ»

Кульберг Н.С. – к.ф.-м.н., руководитель отдела разработки средств медицинской визуализации ГБУЗ «НПКЦ ДиТ ДЗМ»

Гомбеловский В.А. – к.м.н., руководитель отдела развития качества радиологии ГБУЗ «НПКЦ ДиТ ДЗМ»

Сергунова К.А. – к.т.н., руководитель отдела разработки средств контроля и технического мониторинга ГБУЗ «НПКЦ ДиТ ДЗМ»

К–49 Морозов, С.П., Владимирский, А.В., Кляшторный, В.Г. [и др.] Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) / Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 57 – М., 2019. – 51 с.

Рецензенты:

Гусев Александр Владимирович – к.т.н., эксперт К-МИС, член наблюдательного совета ассоциации разработчиков и пользователей искусственного интеллекта для медицины «Национальная база медицинских знаний», член Экспертного совета Минздрава России по вопросам использования ИКТ в системе здравоохранения

Лебедев Георгий Станиславович – д.т.н., заведующий кафедрой информационных и интернет-технологий Первого МГМУ им. И.М. Сеченова, советник директора ФГБУ «ЦНИИОИЗ» Минздрава России по информационным технологиям

Беляев Михаил Геннадьевич – к.ф.-м.н., заведующий сектором «Анализ данных в нейронауках» Института проблем передачи информации им. А.А. Харкевича РАН

Кутичев Владимир Анатольевич – заведующий лабораторией испытаний программного обеспечения медицинских изделий ФГБУ «ВНИИИИМТ» Росздравнадзора

Методические рекомендации прошли публичное обсуждение в формате препринта (arXiv:1908.00381) и были рекомендованы European Society of Medical Imaging Informatics (EuSoMI)

Методические рекомендации предназначены обеспечить методологическую базу для проведения клинических испытаний, клинической апробации и научной оценки алгоритмов и программного обеспечения, разработанных на основе интеллектуальных технологий (искусственного интеллекта). Клинические испытания рассматриваются как этап подготовки к регистрации программного решения в качестве медицинского изделия. Предложены подходы к оценке точности и эффективности разработок на основе искусственного интеллекта для лучевой диагностики.

Данный документ является собственностью Департамента здравоохранения города Москвы, не подлежит тиражированию и распространению без соответствующего разрешения



СОДЕРЖАНИЕ

Нормативные ссылки.....	4
Определения.....	6
Обозначения и сокращения.....	8
Введение.....	9
Основная часть. Общие положения.....	11
Клинические испытания.....	15
Метрики качества.....	21
Требования к эталонным наборам данных.....	26
Оформление результатов.....	28
Заключение.....	29
Список использованных источников.....	30
Приложения.....	31

НОРМАТИВНЫЕ ССЫЛКИ

В настоящем документе использованы ссылки на следующие нормативные документы (стандарты):

1. Федеральный закон от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации».
2. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных».
3. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».
4. Федеральный закон от 27.12.2002 № 184-ФЗ «О техническом регулировании».
5. Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с Национальной стратегией развития искусственного интеллекта на период до 2030 года).
6. Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы».
7. Постановление Правительства Российской Федерации от 27.12.2012 № 1416 «Об утверждении Правил государственной регистрации медицинских изделий».
8. Постановление Правительства Российской Федерации от 05.05.2018 № 555 «О единой государственной информационной системе в сфере здравоохранения».
9. Приказ Министерства здравоохранения Российской Федерации от 09.01.2014 № 2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».
10. Приказ Министерства здравоохранения Российской Федерации от 01.04.2016. № 200н «Об утверждении правил надлежащей клинической практики».
11. Приказ Департамента здравоохранения города Москвы от 25.12.2017 № 918 «О регламенте регистрации данных в системе «Единый радиологический информационный сервис» в медицинских организациях государственной системы здравоохранения города Москвы».
12. ГОСТ Р ИСО 14155-2014 «Клинические исследования. Надлежащая клиническая практика» (приказ Ростехрегулирования от 04.06.2014 № 497-ст).
13. Письмо Федеральной службы по надзору в сфере здравоохранения от 30.12.2015 № 01И-2358/15 «О регистрации программного обеспечения».



14. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы. – М.: ФГБУ «ВНИИИИМТ» Росздравнадзора, 2018. – 31 с.

ОПРЕДЕЛЕНИЯ

В настоящем документе применены следующие термины с соответствующими определениями:

Искусственный интеллект (ИИ) – 1. Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений¹. 2. Научное направление, в рамках которого изучаются проблемы аппаратного и программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными; входит в комплекс компьютерных наук; создаваемые технологии относятся к информационным технологиям.

Математическая модель – это абстракция реального явления, в которой интересующие исследователя отношения между реальными элементами заменены подходящими, зачастую упрощенными, отношениями между математическими объектами. Иными словами, это представление реальной ситуации с помощью математического языка.

Машинное обучение – 1. Раздел компьютерных наук, изучающий применение математических и статистических методов для создания обучающихся информационных систем. Такие системы обучаются решать основную задачу путем предварительного решения множества сходных задач. Является частью искусственного интеллекта. 2. Технологии автоматического обучения алгоритмов ИИ распознаванию и классификации на тестовых выборках объектов для повышения качества распознавания, обработки и анализа данных, прогнозирования.

Набор данных – совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на

¹ Национальная стратегия развития искусственного интеллекта на период до 2030 года (утверждена Указом Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»).

основе искусственного интеллекта². В медицине и здравоохранении – это упорядоченная совокупность биомедицинских данных, включая диагностические изображения и сведения о наличии, характере и локализации патологических изменений на изображениях; структурированные случаи оказания медицинской помощи и связанные с ними электронные медицинские документы из электронных медицинских карт; библиотеки ключевых слов, словосочетаний и их критичных сочетаний, данные генетических тестов, а также комбинации различных медицинских данных, объединенные вокруг одного деперсонифицированного пациента (т.н. «цифровой двойник» – «digital twin»). Если набор данных содержит подтвержденные сведения об окончательном диагнозе для каждого случая или подтвержденном патологическом процессе в изображениях, то он называется «верифицированным». Синонимы: датасет (от англ. «data-set»), эталонные данные, размеченная выборка.

Порог активации («cut-off», точка отсечения) – значение оцениваемого параметра, при достижении которого принимается решение о наличии и/или отсутствии патологии.

Разметка данных – этап обработки структурированных и неструктурированных данных, в процессе которого данным (в том числе текстовым документам, фото- и видеоизображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и (или) осуществляется интерпретация данных для решения конкретной задачи, в том числе с использованием методов машинного обучения³. В медицине и здравоохранении – процесс внесения в набор данных стандартизированных сведений о наличии патологических изменений на диагностических изображениях, сведений о локализации и характере патологии, информации о верификации, извлеченных из неструктурированных текстовых записей признаков, иных связанных данных; выполняется для каждого изображения (исследования) или структурированного электронного медицинского документа независимо несколькими квалифицированными врачами-специалистами. Синоним: тегирование (от англ. «tagging»).

Технологии искусственного интеллекта – технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта⁴. Синоним: интеллектуальные технологии.

² Ibid.

³ Ibid.

⁴ Ibid.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ЕРИС – Единый радиологический информационный сервис

ИИ – искусственный интеллект

ИТ – информационные технологии

МО – медицинская организация

ОЛД – отделение лучевой диагностики

ПКТИ – предварительные клинико-технические испытания

ПО – программное обеспечение

РИС – радиологическая информационная система

РФ – Российская Федерация

ЦАМИ – централизованный архив медицинских изображений

AUC – Area Under Curve (площадь под характеристической кривой)

CE – Conformité Européenne (маркировка «европейское соответствие»)

CONSORT – Consolidated Standards of Reporting Trials (диаграмма потока участников исследования)

FDA – Food and Drug Administration (Управление по санитарному надзору за качеством пищевых продуктов и медикаментов)

PACS – Picture Archiving and Communication System (система хранения и обмена изображениями)

ROC – Receiver Operating Characteristic (характеристическая кривая)

ВВЕДЕНИЕ

В соответствии с Национальной стратегией развития искусственного интеллекта на период до 2030 года (утверждена Указом Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации») использование технологий искусственного интеллекта в социальной сфере способствует созданию условий для улучшения уровня жизни населения, в том числе за счет повышения качества услуг в сфере здравоохранения (включая профилактические обследования, диагностику, основанную на анализе изображений, прогнозирование возникновения и развития заболеваний, подбор оптимальных дозировок лекарственных препаратов, сокращение угроз пандемий, автоматизацию и точность хирургических вмешательств).

Искусственный интеллект (ИИ) – это научное направление, в рамках которого изучаются проблемы аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными; входит в комплекс компьютерных наук; создаваемые технологии относятся к информационным технологиям. Научной новизны в концепции искусственного интеллекта мало. Автоматизированный анализ медицинской информации – это давно существующий раздел науки и практики. Однако в последние годы наблюдается взрывной рост нового поколения технологий искусственного интеллекта, он связан со значительным инженерным прогрессом вычислительных возможностей компьютерной техники и математики. Искусственный интеллект применим для решения ряда задач в медицинской диагностике (поддержка принятия решений, морфометрия, автоматизация участков производственных процессов, контроль качества и т.д.). Для практического применения системы, созданные на основе технологий искусственного интеллекта, должны пройти клинические испытания, оценку диагностической точности и получить официальный статус медицинского изделия.

В соответствии с Федеральным законом от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» медицинскими изделиями являются любые инструменты, аппараты, приборы, оборудование, материалы и прочие изделия, применяемые в медицинских целях отдельно или в сочетании между собой, а также вместе с другими принадлежностями, необходимыми для применения указанных изделий по назначению, включая специальное программное обеспечение, и предназначенные производителем для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека, проведения медицинских исследований, восстановления, замещения, изменения анатомической структуры или физиологических функций организма, предотвращения или прерывания беременности, функциональное назначение которых не реализуется путем

фармакологического, иммунологического, генетического или метаболического воздействия на организм человека.

Программное обеспечение (ПО), оказывающее влияние на принятие врачом клинически значимых решений либо формирующее новую клинически значимую информацию, несет потенциальный риск применения и тем самым – причинения вреда здоровью пациента. В этой связи такое ПО, включая созданное с использованием технологий искусственного интеллекта, подлежит регистрации в качестве медицинского изделия. Обязательным подготовительным этапом перед направлением такого ПО на регистрацию в качестве медицинского изделия является проведение испытаний клинической эффективности.

Данные методические рекомендации устанавливают единообразные подходы к определению эффективности такого ПО в процессе клинических испытаний.

Причем программное обеспечение на основе ИИ может, как применяться в качестве самостоятельного медицинского изделия, так и быть принадлежностью (составной частью) другого медицинского изделия или предназначаться для установки (инсталляции) на этом медицинском изделии.

ОСНОВНАЯ ЧАСТЬ. ОБЩИЕ ПОЛОЖЕНИЯ

Искусственный интеллект в диагностике представляет мультидисциплинарное (медицина, биология, математика и компьютерные науки) прикладное научное направление по постановке и решению задач моделирования отдельных компонентов аналитических процессов, являющихся частью интеллектуальной деятельности врача (таблица 1).

Таблица 1 – Типовые формы применения искусственного интеллекта в лучевой диагностике

Основные задачи (вопросы)	Действие ИИ	Типовая медицинская задача
Локализация объекта	Обнаружение (детекция)	Приоритизация исследований, поддержка принятия решений, контроль качества
1. Объект присутствует или нет. 2. Определение типа объекта	Классификация	Скрининг, массовые профилактические осмотры. Определение характера патологии
Определение размеров, характеристика объекта	Сегментация, определение радиомических особенностей	Определение анатомических границ. Медицинская морфометрия (мониторинг течения патологического процесса, автоматическое формирование описания диагностического изображения)
Распознавание речи и текста человека	Обработка естественного языка	Протоколирование, контроль качества
Оценка изменений объекта во времени	Обнаружение (детекция), классификация, сегментация	Оценка динамики патологических изменений

Любой алгоритм искусственного интеллекта, успешно прошедший клинические испытания, функционирует исключительно под контролем медицинских работников в качестве системы поддержки принятия врачебных решений. Генерируемые алгоритмом результаты могут использоваться или не использоваться медицинскими работниками. Применение искусственного интеллекта направлено на повышение эффективности системы здравоохранения и производительности труда персонала, снижение рисков и ошибок, стандартизацию представления результатов диагностических исследований.

Программное обеспечение, разработанное на основе технологий искусственного интеллекта, проходит клиническую апробацию на этапе подготовки к регистрации в качестве медицинского изделия (в том числе в соответствии со статьей 36.1. Федерального закона от 21.11.2011 № 323-ФЗ). В процессе испытаний должна быть установлена диагностическая эффективность, подтверждены безопасность использования и соответствие реальных характеристик ПО тем, которые были заявлены изготовителем.

В соответствии со статьей 38 Федерального закона от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» медицинское изделие – это «любые инструменты, аппараты, приборы, оборудование, материалы и прочие изделия, применяемые в медицинских целях отдельно или в сочетании между собой, а также вместе с другими принадлежностями, необходимыми для применения указанных изделий по назначению, включая специальное программное обеспечение, и предназначенные производителем для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека, проведения медицинских исследований, восстановления, замещения, изменения анатомической структуры или физиологических функций организма, предотвращения или прерывания беременности, функциональное назначение которых не реализуется путем фармакологического, иммунологического, генетического или метаболического воздействия на организм человека. Медицинские изделия могут признаваться взаимозаменяемыми, если они сравнимы по функциональному назначению, качественным и техническим характеристикам и способны заменить друг друга».

В соответствии с информационным письмом Федеральной службы по надзору в сфере здравоохранения, медицинским изделием, подлежащим государственной регистрации в установленном порядке, является программное обеспечение для:

- управления работой оборудования и мониторинга за его функционированием;
- получения от оборудования диагностических данных, их накопления и расчета в автоматическом режиме;
- мониторинга функций организма человека и передачи полученных данных (в том числе посредством беспроводных технологий);
- расчета параметров подбора дозы (облучения, лекарственного средства, рентгеноконтрастного вещества и т.д.);
- обработки данных, полученных с диагностического медицинского оборудования, передачи их на системы планирования и терапии;
- обработки медицинских изображений (включая изменение его качества, цветового разрешения и т.д.);
- 3D-моделирования (в том числе, в стоматологии и зуботехнике, протезировании, ортопедии, имплантологии, макетировании органов и костей и т.п. [2]);

- связи диагностического и лечебного оборудования;
- обработки цифровых изображений (в том числе с получением данных от диагностического оборудования в неизменном виде).

Дополнительно следует учитывать, что ПО относится к медицинским изделиям в случае, если оно предназначено для [2]:

- обработки, хранения и передачи медицинской документации;
- оказания медицинской помощи с применением телемедицинских технологий;
- установки в специальной информационной системе оказания медицинской помощи на дому;
- обработки данных для электронной регистрации медицинского оборудования, что позволяет проводить учет применяемых в больнице медицинских изделий;
- создания канала связи между пациентом и медицинским специалистом, чтобы пациент мог проводить самостоятельное лечение в соответствии с правилами, определенными в процессе использования приложения, и/или для обеспечения мониторинга населения;
- установки в информационной системе банка крови или существующих электронных вычислительных машинах, или децентрализованных компьютерах / сетях.

В перспективе требуется проведение рандомизированных исследований, оценивающих отдаленные результаты влияния ПО на основе технологий ИИ на показатели качества медицинской помощи, продолжительность и качество жизни пациентов. Результаты таких исследований могут изменить рекомендации по практическому использованию ПО на основе ИИ. Отдельного изучения требует влияние ИИ на психологию и процесс принятия решений медицинским персоналом. До получения научных результатов следует действовать так осторожно, как если бы вред ИИ мог бы превосходить пользу. Безопасность пациента является абсолютным приоритетом.

Таким образом, применяемые в лучевой диагностике алгоритмы и программное обеспечение на основе искусственного интеллекта относятся к медицинским изделиям и должны проходить клинические испытания. При этом программное решение может применяться самостоятельно, в качестве модуля программного обеспечения, или совместно с другими медицинскими изделиями (как аппаратурой, так и программным обеспечением, например, ЕРИС, PACS, РИС и т.д.), вне зависимости от используемой аппаратной платформы, а также способов размещения программного обеспечения и предоставления доступа к нему.

Изложенная далее методология клинических испытаний (в виде предварительных клиничко-технических испытаний) также может быть использована в рамках пилотных проектов, научно-исследовательских работ и клинических апробаций.

КЛИНИЧЕСКИЕ ИСПЫТАНИЯ

Целью клинических испытаний является подтверждение эффективности, безопасности использования, а также соответствие его характеристик использованию по назначению, указанному изготовителем.

В соответствии с п.8 статьи 38 Федерального закона от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации», в целях государственной регистрации медицинских изделий в порядке, установленном Росздравнадзором, проводится оценка соответствия в форме технических испытаний, клинических испытаний и экспертизы качества, эффективности и безопасности медицинских изделий.

Согласно п. 4 Правил регистрации, утвержденных постановлением №1416, *технические испытания* проводятся с целью определения соответствия характеристик медицинского изделия требованиям нормативной документации, технической и эксплуатационной документации производителя и принятия последующего решения о возможности проведения клинических испытаний. Производитель сам определяет, каким нормативным документам соответствует его изделие.

Клинические испытания представляют собой разработанное и запланированное систематическое исследование, предпринятое для оценки безопасности и эффективности медицинского изделия. Провести клинические испытания можно только при условии успешного проведения технических испытаний.

Критерии допуска к клиническим испытаниям программного обеспечения на основе технологий искусственного интеллекта (список и анкета) – см. Приложения 1, 2. Весомость и значимость каждого критерия может определяться индивидуально для каждой ситуации. Критерии вносятся в техническое задание, описывающее требования к ПО на основе технологий искусственного интеллекта. Также в техническом задании обязательно должны быть указаны *цели и клинические сценарии применения* искусственного интеллекта (например, на основе таблицы 1) по отдельным модальностям, видам исследований, производственным процессам.

Процесс клинических испытаний организуют в соответствии с действующей нормативно-правовой базой и рекомендованной методикой экспертизы качества, эффективности и безопасности медицинских изделий в части программного обеспечения. В ходе испытаний ПО на основе технологий искусственного интеллекта должна быть дана *«оценка клинической связи (научной обоснованности) степени адекватности подбора данных и системы их анализа (концепция, измерения, заключение) целевому назначению»* [2].

Клиническая оценка ПО, разработанного на основе искусственного интеллекта и предназначенного для применения в лучевой диагностике, осуществляется в 2 этапа (рис.1):

1. Аналитическая валидация.
2. Клиническая апробация.

Аналитическая валидация – оценка корректности обработки входных данных программным обеспечением для создания надежных выходных данных; оценивается с применением эталонных наборов размеченных данных. Аналитическая валидация (в рамках клинических испытаний) ПО на основе технологий ИИ включает 6 этапов (таблица 2).

Таблица 2 – Этапы аналитической валидации ПО на основе технологий искусственного интеллекта для лучевой диагностики

Этап	Цель	Краткое описание	Результат этапа
1	2	3	4
I. Анкетирование	(1) сбор информации для рейтинга сервиса (цели, сертификация, доказательность, функциональность, контракт); (2) оценка клинической задачи сервиса и практической применимости в рабочем процессе отделений лучевой диагностики	Юридическому лицу - разработчику ИИ сервиса предлагается заполнить анкету (на основе Приложения 2; практический пример – см. Приложение 3), затем ответы в анкете анализируются комиссией	Заполненная анкета
II. Самостоятельный тест	Проверка технической готовности сервиса (бесшовная интеграция в информационную систему в сфере здравоохранения, обработка данных со всех типов диагностических аппаратов, качественное предоставление результатов медицинским работникам)	Разработчику предоставляется типовой набор данных (обычно от 5 до 10 исследований) для самооценки технической готовности ПО ⁵	Файлы с результатами обработки набора данных ПО

⁵ Пример наборов данных для самостоятельного тестирования – ai.nrcmr.ru

Продолжение таблицы 2

1	2	3	4
III. Интервью	(1) детализация клинической задачи и сценария применения; (2) оценка юридического лица на соответствие корпоративной культуре превосходного качества и организации (КПКО).	С представителями компании проводится интервью согласно протоколу (Приложения 4, 5).	Ответы на вопросы в протоколе
IV. Онлайн-тест	Оценка работы сервиса в режиме реального времени на ограниченном наборе данных (включает выборку показательных случаев)	Анализ данных, интерпретация результатов проводится при непосредственном участии комиссии	Отчет о работе ИИ алгоритма/сервиса по каждому исследованию
V. Доказательный тест	Ретроспективная проверка заявленной точности и остальных параметров работы алгоритмов на эталонном наборе данных	Происходит тестирование ПО на эталонном наборе данных, подготовленном в соответствии с клинической задачей, решаемой ПО.	Оценка диагностической точности (аналитическая валидация – см. далее).
VI. Итоговая оценка	(1) Интегральная оценка работы ПО и готовности к проверке в реальных условиях ОЛД; (2) Интегральная оценка юридического лица-разработчика	Комиссией анализируются все прошедшие этапы тестирования и выставляются промежуточные оценки ПО и юридическому лицу	Решение о допуске ПО сервиса к интеграции с действующей информационной системой в сфере здравоохранения (РИС/РАС) для оценки работы ПО в условиях рутинных производственных процессов отделений лучевой диагностики

Клиническая апробация (оценка эффективности путем использования в рамках стандартного производственного процесса) состоит из двух компонентов (рис.1):

- клинической корреляции (оценки наличия достоверной клинической связи между результатами и целевым клиническим состоянием ПО),
- клинической валидации (подтверждение достижения намеченной цели в целевой популяции в контексте лечебной работы благодаря использованию точных и надежных выходных данных).

На первом этапе определение клинического значения осуществляется путем:

а) оценки надежности юридического лица (разработчика ПО), в том числе посредством проверки на соответствие корпоративной культуре превосходного качества и организации;

б) генерации и анализа новых диагностических данных для конкретного предполагаемого использования ПО:

- формируется эталонный набор подготовленных биомедицинских данных;
- осуществляется автоматизированный анализ набора (как имитация диагностического процесса);
- результаты анализа сопоставляются с эталонным подготовленным набором данных;
- проводится математический и статистический анализ.

На втором этапе испытываемое программное обеспечение применяется по целевому назначению в рамках типового производственного процесса. Процедуры, выполняемые с применением ПО на основе искусственного интеллекта, оцениваются с позиций производительности и качества. Соответственно, проводится сопоставительный хронометраж (при этом измерения должны проводиться многократно, с участием различных сотрудников-исполнителей, в разное время суток и периоды рабочих смен) и ретроспективный пересмотр результатов обследований. Организаторами испытания должны быть предусмотрены меры по минимизации рисков и обеспечению безопасности (безвредности для пациентов и их законных представителей) производственных процессов.

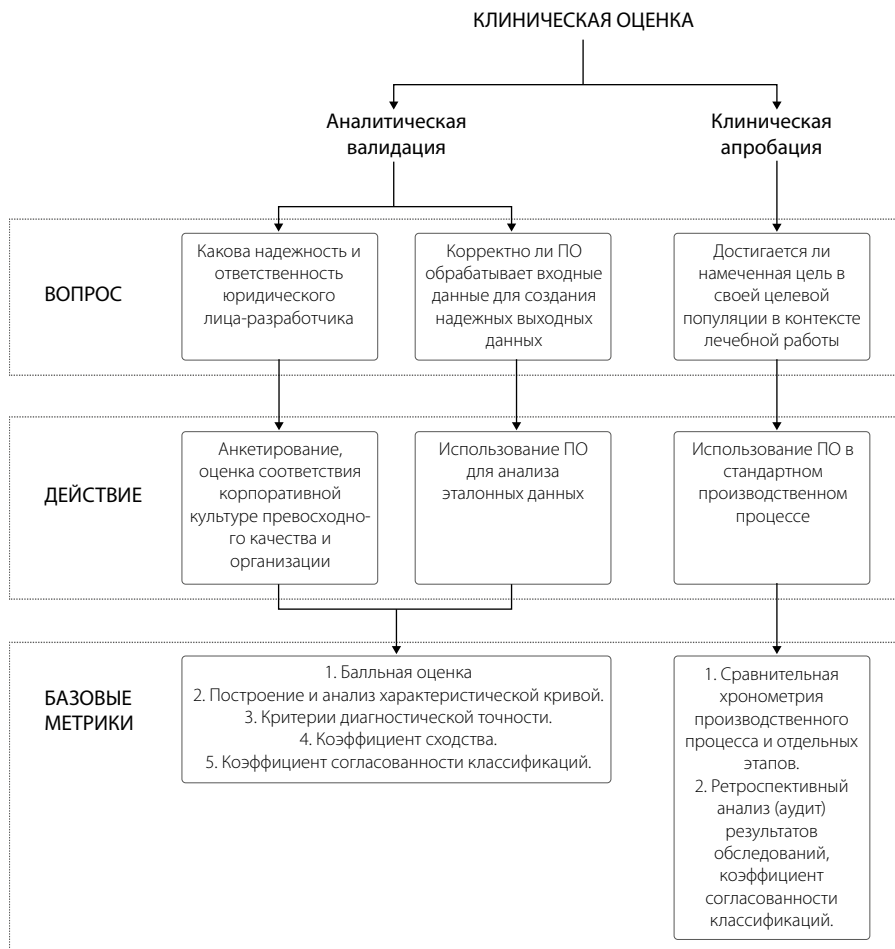


Рисунок 1 – Схема проведения клинической оценки программного обеспечения на основе технологий искусственного интеллекта (для сферы лучевой диагностики)

В программе проведения клинических испытаний должны быть указаны:

- детали сведения о наборе данных (эталонных биомедицинских данных), включая номер свидетельства о государственной регистрации (желательно), количество и характер данных, сведения о разметке (см. далее);
- сведения о необходимости инсталляции ПО с применением технических средств испытательной медицинской организации;

- сведения об использовании вычислительных мощностей, предоставленных разработчиком ПО;
- сведения о деинсталляции ПО после завершения испытаний (в связи с тем, что ПО устанавливается и применяется на соответствующем оборудовании до получения регистрационного удостоверения).

В целом, формирование программы рекомендуется осуществлять в соответствии с чек-листом (см. приложение б).

Предварительные клиничко-технические испытания

Предварительные клиничко-технические испытания (ПКТИ) представляют собой разработанное и запланированное исследование, предпринятое для предварительной, научно-практической оценки безопасности и эффективности программного обеспечения на основе искусственного интеллекта.

ПКТИ проводятся по упрощенной и адаптированной методологии клинических испытаний с целью получения дополнительных знаний о возможностях и ограничениях ПО на основе ИИ. ПКТИ не являются заменой или аналогом клинических испытаний.

ПКТИ могут включать этапы аналитической валидации и клинической апробации либо состоять только из аналитической валидации (с обязательным выполнением 6 этапов – см. таблицу 2). По результатам ПКТИ оформляется отчет (приложение 7).

ПКТИ применимы в контексте научно-исследовательских, опытно-конструкторских работ, а также в качестве инструмента оценки медико-технической готовности ПО на основе ИИ для углубленного изучения в реальных клинических условиях.

МЕТРИКИ КАЧЕСТВА

Оценка диагностической точности

Аналитический этап осуществляется в дизайне диагностического исследования (см. приложение б), при этом используется стандартный набор метрик.

Автоматизированный анализ медицинских данных посредством исследуемого ПО на основе ИИ считают индекс-тестом.

Результаты разметки набора биомедицинских данных в соответствии с известной методологией – референс-тестом.

Для сопоставления результатов индекс- и референс-теста составляют четырехпольную таблицу и определяют абсолютное значение для каждого вида результатов (таблица 3). Далее производят выбор и расчет релевантных метрик (таблица 4) с определением достоверности в рамках 95% доверительного интервала.

Таблица 3 – Виды результатов применения ИИ в соответствии с четырехпольной таблицей (на примере ответа, представляющего собой бинарную переменную⁶)

Вид результата			
Истинно-положительный	Истинно-отрицательный	Ложно-положительный	Ложно-отрицательный
ИП	ИО	ЛП	ЛО
ПО выявило патологию при ее наличии	ПО не выявило патологию при ее отсутствии	ПО выявило патологию при ее отсутствии	ПО не выявило патологию при ее наличии

Таблица 4 – Типовой выбор метрик для аналитического этапа исследования⁷

Действие ИИ	Базовые метрики
Обнаружение (детекция)	Стандартный набор метрик
Классификация	Стандартный набор метрик
Сегментация	Коэффициент сходства
Обработка естественного языка	Коэффициент согласованности классификаций

⁶ Данный тип ответа соответствует большинству типовых задач; вместе с тем большинство иных задач практически всегда могут быть сведены к задаче бинарной классификации.

⁷ Указанные метрики являются базовыми, обязательными для оценки; дополнительно могут быть использованы и другие метрики, применимые к конкретной задаче.

В зависимости от решаемой задачи могут применяться различные сочетания метрик. Например, анализ характеристической кривой и определение коэффициента согласованности классификаций и т.д. Такие основные показатели, как чувствительность, специфичность, площадь под характеристической кривой, должны быть использованы для характеристики каждого типового действия искусственного интеллекта.

Стандартный набор метрик

Для сопоставления эффективности индекс-теста относительно референс-теста применяется классический набор показателей, характеризующих диагностическую результативность (таблица 5).

Таблица 5 – Стандартный набор метрик диагностической ценности

Показатель	Значение	Формула
1	2	3
Чувствительность	Вероятность того, что индекс-тест окажется позитивным при наличии признака	$ИП / (ИП + ЛО)$
Специфичность	Вероятность того, что индекс-тест окажется негативным при отсутствии признака	$ИО / (ИО + ЛП)$
Точность (общая валидность)	Уровень соответствия полученных результатов индекс-тест к действительным значениям референс-теста	$(ИП + ИО) / (ИП + ИО + ЛП + ЛО)$
Отношение правдоподобия положительного результата	Математически отображает во сколько раз при положительном результате индекс-теста вероятность наличия целевой патологии превышает вероятность его отсутствия	$Чувствительность / (1 - специфичность)$
Отношение правдоподобия отрицательного результата	Математически отображает во сколько раз при отрицательном результате индекс-теста вероятность наличия целевой патологии превышает вероятность его отсутствия	$(1 - чувствительность) / специфичность$
Прогностическая ценность положительного результата	Вероятность заболевания при положительном результате индекс-теста	$ИП / (ИП + ЛП)$

Продолжение таблицы 5

1	2	2
Прогностическая ценность отрицательного результата	Вероятность отсутствия заболевания при отрицательном (нормальном) результате индекс-теста	ИО / (ИО + ЛО)
Частота ложных срабатываний	Вероятность ложного срабатывания	1–специфичность

Все метрики, кроме отношения правдоподобия, оцениваются в диапазоне 0–1 или в процентах от 0 до 100%:

- <0,6 – непригодно;
- 0,61 – 0,8 – требуется доработка;
- $\geq 0,81$ – может быть допущено к клинической валидации.

Отношение правдоподобия положительного результата должно быть, как можно выше, тогда как отношение правдоподобия отрицательного результата – как можно ниже.

Характеристическая кривая (ROC-кривая)

Характеристическая кривая – отображение соотношения между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак (то есть чувствительностью), и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак (то есть специфичностью) при варьировании порога активации. Метрика диагностической ценности: площадь под кривой (AUC) – площадь, ограниченная ROC-кривой и абсциссой.

Классическая ROC-кривая: график зависимости чувствительности от 1-специфичности (по оси абсцисс – 1-специфичность, по оси ординат – чувствительность).

При анализе ROC-кривой необходимо провести определение оптимальных значений порога активации (так называемого «cut-off»). Для этого используются различные методики. В качестве примера (но не обязательного требования) можно привести следующие:

1. Минимальное расстояние от верхнего левого угла до ROC-кривой (Минимум d);
2. Индекс Юдена (Youden index), который отображает максимальное расстояние от диагональной линии до ROC-кривой (рис. 2).

Приоритет той или иной методике определения «cut-off» должен отдаваться в зависимости от целей исследования.

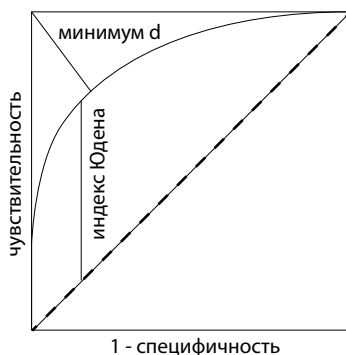


Рисунок 2 – Методика определения оптимальных значений порога активации («cut-off»)

Площадь под кривой (AUC) оценивается в диапазоне 0-1:

- <0,6 – непригодно;
- 0,61 - 0,8 – требуется доработка;
- >=0,81 – может быть допущено к клинической валидации.

Коэффициент согласованности классификаций

Согласованность (конкордантность) – математическое отображение согласованности классификации двух экспертов относительно одного явления.

Формула расчета (1) по методу Коэна (каппа Коэна) (Cohen J., 1960):

$$K = (P_o - P_e) / (1 - P_e), \tag{1}$$

где P_o – доля случаев, когда измерения совпадали, и P_e – ожидаемая доля случаев случайного совпадения. Порядок расчета (2):

		Эксперт №1		
		1	2	Всего
Эксперт №2	1	P_{11}	P_{12}	$P_{1_}$
	2	P_{21}	P_{22}	$P_{2_}$
	Всего	$P_{_1}$	$P_{_2}$	1

$$P_0 = P_{11} + P_{22}$$

$$P_e = P_{-1}P_{1-} + P_{-2}P_{2-} \quad (2)$$

Оценивается в диапазоне 0-1 или в процентах от 0 до 100%:

- <0,6 – непригодно;
- 0,61 - 0,8 – требуется доработка;
- >=0,81 – может быть допущено к клинической валидации.

Коэффициент сходства

Бинарный показатель для количественного определения степени сходства биологических объектов. Формула расчета (3) по методу Дайса-Сёренсена (Dice L., 1945, Sørensen T., 1948):

$$DSC = 2|A \cap B| / (|A| + |B|), \quad (3)$$

где в числителе – удвоенное количество совпавших оценок, в знаменателе – общее количество оценок.

Для множеств с одинаковой мощностью **оценивается** в диапазоне 0–1:

- <0,6 – непригодно;
- 0,61 - 0,8 – требуется доработка;
- >=0,81 – может быть допущено к клинической валидации.

Клиническая апробация

Хронометраж – метод оценки результативности применения ПО, фиксация и замеры продолжительности выполняемых действий (процедур, операций) в рамках производственных процессов. Например, продолжительность описания цифровой рентгенографии легких, выполненной в профилактических целях с и без искусственного интеллекта.

Аудит результатов исследований – ретроспективный анализ качества интерпретации медицинских данных; например, проведения и описания лучевых исследований, выполненных с применением технологий искусственного интеллекта, с целью определения степени соответствия рекомендуемым стандартам [4]. Аудит позволяет оценить качество оказания медицинской помощи, включая проведение и интерпретацию результатов диагностических исследований с и без использования ПО на основе технологий искусственного интеллекта; достоверность различий или их отсутствия определяется статистически.

ТРЕБОВАНИЯ К ЭТАЛОННЫМ НАБОРАМ ДАННЫХ

Для выполнения аналитического этапа клинико-технических и клинических испытаний применяют эталонные данные – наборы биомедицинских данных.

Набор данных – совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта⁸. С точки зрения медицины и здравоохранения набор эталонных данных – это упорядоченная совокупность: диагностических изображений одной модальности и/или однотипных медицинских документов; сведений о наличии, характере и локализации патологических изменений на изображениях; для текстовых документов – библиотеки ключевых слов, словосочетаний и их критичных сочетаний. Если набор содержит сведения об окончательном диагнозе для каждого случая, то он называется «верифицированным».

Эталонный набор данных применяется для аналитической валидации программного обеспечения, разработанного на основе ИИ для решения задач медицинской диагностики.

Размер эталонного набора данных (т.е. размер выборки) для оценки диагностических характеристик алгоритма ИИ устанавливается с помощью методик определения размера выборки для пропорций.

Для каждого конкретного случая (клинического сценария) должны быть заданы желаемые параметры точности оценки диагностических характеристик алгоритма ИИ; необходимо задать ожидаемую пропорцию и желаемую точность (ширину доверительного интервала) ее оценки. Надо учитывать, что чем выше желаемая точность оценки (т.е. чем уже доверительный интервал) и чем ближе пропорция к 50%, тем больше количество размеченных исследований понадобится включить в набор данных.

Эталонный набор данных должен быть проверен профильной медицинской научно-исследовательской организацией на предмет полноты и качества содержащейся в нем информации.

Рекомендуется при проведении клинических испытаний применять эталонные наборы данных, имеющие государственную регистрацию в качестве базы данных.

⁸ Национальная стратегия развития искусственного интеллекта на период до 2030 года (утверждена Указом Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»)

Эталонный набор данных должен содержать такие сведения (описательного характера):

1. Номер свидетельства о государственной регистрации базы данных (рекомендательно);
2. Характеристика популяции (гендерно-возрастные показатели, этнический состав, регионы проживания и т.д.); сведения о деперсонализации; сведения о медицинских организациях, послуживших источниками для формирования набора данных;
3. Характеристика исследований: анатомическая область(и), модальность, проекции, типы медицинских изделий – диагностических приборов, виды и характеристики протоколов исследований;
4. Целевая патология в соответствии с Международной классификацией болезней;
5. Общее количество клинических случаев, исследований, изображений, документов и их распределение по диагностическим группам;
6. Соотношение случаев «норма»:«патология» (случаи «патология» могут быть разделены на несколько подклассов);
7. Сведения о верификации (патогистологическом или ином окончательном диагнозе);
8. Методология разметки (ранее представленной в виде научных публикаций, методических рекомендаций или патентов).

Эталонный набор данных должен отвечать таким требованиям:

1. Соотношение «норма»:«патология» должно соответствовать распространенности целевой патологии в популяции;
2. При формировании эталонного набора должны использоваться данные из нескольких медицинских организаций;
3. Демографические, социально-экономические характеристики и основные показатели здоровья пациентов, чьи данные включаются в эталонный набор данных, должны соответствовать усредненным характеристикам популяции территории, на которой планируется использование ИИ;
4. Планируемый размер эталонного набора данных должен быть обоснован в протоколе исследования, исходя из статистических соображений и желаемой точности оценки основных метрик, указанных выше;
5. Эталонные наборы данных, используемые в процессе клинических испытаний для регистрации ПО в качестве медицинского изделия, не должны быть публично доступны (для исключения возможности обучения алгоритмов ИИ на эталонных наборах данных).

ОФОРМЛЕНИЕ РЕЗУЛЬТАТОВ

При оценке результатов клинико-технических и клинических испытаний следует учитывать следующее:

- ПО проверяется с применением биоматериалов и всего необходимого оборудования (в том числе других медицинских изделий), предназначенного для обеспечения надлежащей работы исследуемого ПО;

- в процессе проведения клинических испытаний должны быть задействованы все модули ПО (при модульной архитектуре), опробованы все функции, указанные производителем в эксплуатационной документации;

- должны быть оценены все позиции доказательных баз ПО (клинические данные, анализ взаимозаменяемых медицинских изделий и т.п.), представленных производителем (изготовителем) в испытательную медицинскую организацию;

- испытательная медицинская организация в обязательном порядке оценивает документацию на программный продукт, включая техническую и эксплуатационную, и делает вывод как о том, что содержание документа доступно и понятно специалистам соответствующего профиля, так и о том, что его содержание обеспечивает надлежащее эффективное и безопасное медицинское применение.

Результаты клинических испытаний Программного обеспечения (как и любого другого медицинского изделия) должны быть оформлены в соответствии с Приложениями к приказу Министерства здравоохранения Российской Федерации от 09.01.2014 № 2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».

При подготовке документации необходимо формировать подробные результаты оценки и анализа данных в соответствии с дополненным чек-листом STARD 2015 (см. Приложение 6).

ЗАКЛЮЧЕНИЕ

Отдельные алгоритмы и инструменты на основе технологий искусственного интеллекта уже успешно применяются в диагностике. Дальнейшая разработка искусственного интеллекта должна быть направлена на решение конкретных медицинских задач.

Внедрение данных рекомендаций позволит унифицировать проведение клинических испытаний программного обеспечения, разработанного на основе технологий искусственного интеллекта для сферы лучевой диагностики. Такие испытания – обязательный компонент для регистрации ПО на основе технологий искусственного интеллекта в качестве медицинского изделия.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Власов, В.В., Реброва, О.Ю. Доказательная медицина: оценка эффективности диагностических вмешательств // Заместитель главного врача. – 2010. – № 4. – С. 50–60.
2. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы. – М.: ФГБУ «ВНИИИИМТ» Росздравнадзора, 2018. – 31 с.
3. Николенко, С., Кадурын, А., Архангельская, Е. Глубокое обучение. – СПб.: Питер, 2018. – 480 с.
4. Морозов, С.П., [и др.]. Оценка качества рентгенорадиологических исследований / С.П. Морозов, Н.Н. Ветшева, Н.В. Ледихова / Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 22. – М., 2019. – 50 с.
5. Allen, B., Dreyer, K. The Role of the ACR Data Science Institute in Advancing Health Equity in Radiology / B. Allen, K. Dreyer // J Am Coll Radiol. – 2019. – Vol. 16, №14. – P. 644–648.
6. Artificial Intelligence in Medical Imaging. Opportunities, Applications and Risks / Erik R. Ranschaert, Sergey Morozov, Paul R. Algra (Eds.). – Springer, 2019.
7. Bossuyt, P.M., et al. For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies / P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, et al. // Radiology. – 2015. – Vol. 277, №3. – P. 826–832.
8. Guidance for Industry and Food and Drug Administration Staff-Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data -Premarket Notification. – 2012. – July. – URL: <https://bit.ly/2tiSybQ>. – Текст: электронный.
9. RCR position statement on artificial intelligence // The Royal College of Radiologists. – 2018. July. – URL: <https://www.rcr.ac.uk/posts/rcr-position-statement-artificial-intelligence>. – Текст: электронный; Tang A., et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology / A. Tang, R. Tam, A. Cadrin-Chênevert, et al. // Can Assoc Radiol J. – 2018. – Vol. 69, №2. – P. 120–135.

Приложение 1

Критерии для допуска ПО на основе ИИ к клиническим испытаниям (предварительным клиничко-техническим испытаниям)

1. Цели:

- ПО обеспечивает предварительный автоматический анализ диагностических изображений для повышения качества и скорости подготовки описания врачом;
- ПО обеспечивает приоритизацию исследований в рабочем списке («worklist») в соответствии с автоматически выявленной патологией;
- ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики);
- ПО обеспечивает поддержку принятия врачебных решений;
- ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ.

2. Сертификация:

- пройдены технические испытания медицинского изделия в аккредитованной лаборатории;
- наличие у ПО на основе технологий искусственного интеллекта сертификации FDA и/или CE (класс II); наличие у юридического лица-разработчика фактического внедрения ПО в медицинские организации: не менее 2-х независимых учреждений; более 6 месяцев эксплуатации; не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ПО решает несколько задач);
- наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых «Scopus» и/или «Web of Science» и входящих в первый-второй квартиль согласно «International Scientific Journal & Country Ranking»;
- должны быть доказаны диагностическая точность – $AUC \geq 0,81$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж).

3. Безопасность:

- соблюдение требований законодательства Российской Федерации в сфере персональных данных, информационной безопасности, охраны здоровья;

- наличие или готовность разместить на территории Российской Федерации серверные мощности, необходимые для работы ПО.

4. Доказательность:

- после окончания разработки оценка точности алгоритмов проводилась на независимых данных⁹;
- диагностическая точность протестирована на популяции, характеристики которой аналогичны населению территории предполагаемого использования ИИ;
- ежегодное обновление информации о диагностической точности.

5. Стандартизация:

- автоматизированный анализ диагностических изображений в стандарте DICOM;
- поддержка стандарта HealthLevel7 (HL7) / FHIR (в частности должен обеспечиваться обмен сообщениями о завершении автоматического анализа изображения, обнаружении и классификации патологии);
- при формировании проектов описаний используются рекомендованные классификации (RADS, MAGNIMS и т.д.).

6. Интеграция:

- наличие или готовность разработать средства для «бесшовной» интеграции с информационными системами в сфере здравоохранения данного субъекта РФ, медицинскими информационными системами;
- наличие средств для интеграции с PACS и ПИС (DICOM «query and retrieve»);
- возможность «бесшовной» интеграции с PACS/ПИС, обеспечиваемой только программными средствами.

7. Функциональность:

- возможность потоковой обработки с последующей отправкой в PACS серий, дополненных результатами анализа ИИ;
- возможность совмещения серий нативных изображений и серий, дополненных результатами анализа ИИ;
- выявление находок (нозологий, патологических признаков, отклонений от нормы) для данной модальности или рабочего процесса в соответствии с техническим заданием;

⁹ Набор медицинских данных для тестирования отличался от набора, использованного для обучения, калибровки и валидации алгоритма (то есть тестирование осуществлено на данных, с которыми ранее алгоритм не работал).

- классификация (определение типа, разновидности) находки, в том числе в соответствии с Международной классификацией болезней (версия 10) если предусмотрено техническим заданием;
- предоставление информации о вероятности наличия целевой патологии;
- автоматическая подготовка и выгрузка в PACS проекта описаний находок (с соблюдением структуры шаблона описаний, предустановленной в PACS);
- результат автоматического анализа отображается стандартными средствами PACS, в том числе графическими возможностями DICOM;
- наличие шаблонов протоколов описаний с их автоматической генерацией и отправкой в РИС/МИС посредством сообщений HL7/FHIR;
- наличие встроенного инструмента оценки точности;
- длительность обработки одного исследования не превышает установленного временного периода (формируется индивидуально, исходя из клинического сценария, характеристик инфраструктуры и проч.; пример – 60 секунд, не учитывая время на пересылку данных; при оценке динамики (сравнении исследований) анализ может занимать более 60 секунд, но не более 60 секунд на одно исследование);
- функция контроля, обеспечивающая старт автоматического анализа только по факту получения из РИС сообщения HL7 «status complete by radiographer»;
- сообщения ПО в стандарте HL7 о завершении автоматического анализа и обнаружении патологии используются для приоритизации исследований в рабочем списке («worklist»);
- возможность автоматизированного поиска в базе PACS/РИС аналогичных исследований для сравнения;
- для каждого проанализированного исследования ПО формирует декларацию, содержащую название ПО (алгоритма), список выявленных находок (нозологий, патологических признаков, отклонений от нормы) с указанием чувствительности и специфичности для каждого типа находок; декларация формируется в DICOM SR, Overlay и т.д. и передается в PACS как отдельная серия;
- результатом работы ПО является сформированная серия изображений в формате DICOM, которая:
 - а) имеет количество срезов, аналогичное исходной серии для возможности синхронного просмотра врачом;
 - б) обладает информацией на каждом срезе о названии ПО, версии, диагностической точности с датой ее проверки и точным временем завершения анализа конкретного исследования;
 - в) допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок).

8. Контракт:

- наличие у юридического лица-разработчика системы менеджмента качества;
- наличие у юридического лица-разработчика политики подготовки и контроля версий;
- регулярные обновления системы, включая обновление информации о диагностической точности;
- обновления ПО включены в стоимость;
- все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика.

Приложение 2

Анкета для допуска программного обеспечения (ПО) на основе технологий искусственного интеллекта / компьютерного зрения к предварительной опытной эксплуатации (предварительным клинико-техническим испытаниям)

Раздел	Метрики	Ответ (1-да/0-нет)	Комментарий
1. Цели	<p>1.1. ПО обеспечивает предварительный автоматический анализ медицинских изображений в формате DICOM для повышения качества и скорости подготовки описания врачом.</p> <p>1.2. ПО обеспечивает приоритизацию исследований в рабочем списке («worklist») в соответствии с автоматически выявленной патологией.</p> <p>1.3. ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ.</p> <p>1.4. ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики).</p>		
2. Сертификация	<p>2.1. Наличие у ПО на основе технологий искусственного интеллекта сертификации FDA и/или CE (класс II). В случае ответа «нет» на пункт 2.1, должны быть ответы «да» на пункты 2.2 и 2.3.</p> <p>2.2. Наличие у юридического лица-разработчика фактического внедрения ПО в медицинские организации:</p> <ul style="list-style-type: none"> - не менее 2-х независимых учреждений; - более 6 месяцев эксплуатации; - не менее 1000 успешно выполненных анализов (подтвержденных 		

	<p>врачом-диагностом) для каждой задачи (если ПО решает несколько задач).</p> <p>2.3. Наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых «Scopus» и/или «Web of Science» и входящих в первый-второй квартиль согласно «International Scientific Journal &Country Ranking; должны быть доказаны диагностическая точность – $AUC \geq 0,81$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж).</p>		
<p>3. Доказательность</p>	<p>3.1. После окончания разработки оценка точности алгоритмов проводилась на независимых данных (набор биомедицинских данных для тестирования отличался от набора, использованного для обучения, калибровки и валидации системы; то есть тестирование осуществлено на данных, с которыми ранее алгоритм не сталкивался).</p> <p>3.2. Диагностическая точность протестирована на данных лиц, относящихся к популяции, характеристики которой аналогичны населению территории предполагаемого использования ИИ.</p> <p>3.3. Ежегодное обновление информации о диагностической точности.</p>		
<p>4. Функциональность</p>	<p>4.1. Наличие встроенного инструмента оценки точности.</p> <p>4.2. Длительность обработки одного исследования не превышает установленного временного периода (формируется индивидуально, исходя из клинического сценария, характеристик инфраструктуры и проч.; пример – 60 секунд, не учитывая время на пересылку данных; при оценке динамики (сравнении исследований) анализ может занимать более 60 секунд, но не более 60</p>		

	<p>секунд на одно исследование).</p> <p>4.3. Результатом работы ПО является сформированная серия изображений в формате DICOM, которая:</p> <ul style="list-style-type: none"> - имеет количество срезов аналогичное исходной серии для возможности синхронного просмотра врачом; - обладает информацией на каждом срезе о названии ПО, версии, диагностической точности с датой ее проверки и точным временем завершения анализа конкретного исследования; - допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок). 		
<p>5. Контракт</p>	<p>5.1. Наличие у юридического лица-работчика системы менеджмента качества.</p> <p>5.2. Регулярные обновления системы, включая обновление информации о диагностической точности.</p> <p>5.3. Обновления ПО включены в стоимость.</p> <p>5.4. Все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика.</p>		

Приложение 3

Радиология Москвы: Анкета для юридического лица-разработчика ПО на основе ИИ для лучевой диагностики



РАДИОЛОГИЯ МОСКВЫ
ДИАГНОСТИКА БУДУЩЕГО

АНКЕТА О СЕРВИСЕ НА ОСНОВЕ «ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» (ИИ) ДЛЯ ЛУЧЕВОЙ ДИАГНОСТИКИ

Раздел	Метрики	Ответ (да/нет)	Комментарий
Название компании			
1. ИИ сервисы	1.1. Перечень ИИ сервисов, к которым применимы ответы в данной анкете	N/A	
2. Клинический сценарий использования ИИ сервиса	2.1. Ситуация, в которой применяется ИИ сервис	Плановая Экстренная	
	2.2. Уровень медицинской помощи	Первичный Вторичный Третичный	
	2.3. Этап медицинской помощи	Догоспитальный Госпитальный Амбулаторный	
	2.4. Этап бизнес-процесса (clinical workflow), на котором врачу предоставляются результаты ИИ сервиса	Оповещение врача: До просмотра рабочего списка В рабочем списке В момент первоначального открытия исследования По запросу врача (второе мнение) Аудит после финализации заключения	
	2.5. Каким образом предоставляются результаты ИИ сервиса врачу	Бинарная оценка наличия патологии Бальная/категориальная оценка наличия патологии Текстовое описание результатов в свободной форме Шаблон протокола описания исследования В форме рекомендаций о дальнейшей тактике Визуализация находок на изображениях исследования Схематичное представление	

	2.6. Какие решения принимают врач-рентгенолог и направляющий врач с учетом результатов ИИ сервиса	Никаких Выбор варианта дальнейшего описания/анализа исследования Маршрутизация к специалисту Дообследование другим методом Дообследование этим же методом через определенный промежуток времени	
	2.7. Конечная практическая цель ИИ сервиса	Сокращение времени между проведением исследования и финализацией протокола Повышение качества работы врача-рентгенолога (сокращение клинически значимых ошибок) Аудит работы врачей	
3. Цели	3.1. ИИ сервис обеспечивает предварительный автоматический анализ медицинских изображений в формате DICOM для повышения качества и скорости подготовки описания врачом	Да Нет	
	3.2. ИИ сервис обеспечивает приоритизацию исследований в рабочем списке («worklist») в соответствии с автоматически выявленной патологией	Да Нет	
	3.3. ИИ сервис обеспечивает автоматическую подготовку шаблона описания исследования по результатам работы	Да Нет	
	3.4. ИИ сервис обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM	Да Нет	
4. Сертификация	4.1. Наличие у ИИ сервиса на основе интеллектуальных технологий сертификации FDA и/или CE (класс II). В случае ответа «нет» на пункт 4.1, должны быть ответы «да» на пункты 4.2 и 4.3	Да Нет В процессе	
	4.2. Наличие у компании-разработчика фактического внедрения ИИ сервиса в медицинские организации: – не менее 2-х независимых учреждений; – более 6 месяцев эксплуатации; – не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ИИ сервис решает несколько задач).	Да Нет В процессе	



	Необходимо предоставить подробный список (включая контактную информацию) больницы/клинических центров, в которых использовался или используется ваш продукт и копии контрактов/соглашений с этими учреждениями (без конфиденциальной информации)		
	4.3. Наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых «Scopus» и/или «Web of Science» и входящих в первый-второй кварталы согласно «International Scientific Journal & Country Ranking»; должны быть доказаны диагностическая точность – AUC \geq 0,8 (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж)	Да Нет В процессе	
5. Доказательность	5.1. После окончания разработки оценка точности алгоритмов проводилась на независимых данных (база медицинских данных для тестирования отличалась от базы, использованной для обучения, калибровки и валидации системы; то есть тестирование осуществлено на данных, с которыми ранее алгоритм не сталкивался)	Да Нет	
	5.2. Диагностическая точность протестирована на данных лиц европеоидной и монголоидной рас	Да Нет	
	5.3. Ежегодное обновление информации о диагностической точности	Да Нет	
	5.4. Критерии оценки клинической эффективности ИИ сервиса <i>Если возможно, пожалуйста, пришлите отзывы или рекомендательные письма, написанные радиологами, которые использовали или в настоящее время используют ваш продукт</i>	N/A	
6. Функциональность	6.1. Наличие встроенного инструмента оценки точности	Да Нет N/A	
	6.2. Длительность обработки одного исследования не превышает 60 секунд, не учитывая время на пересылку данных. Для выполнения цели 3.4 анализ может занимать более 60 секунд, но не более 60 сек на одно исследование <i>Пожалуйста, укажите системные требования. Если ваш продукт не выполняет это условие – добавьте комментарии</i>	Да Нет Другое	

	<p>6.3. Результатом работы ИИИ сервиса является сформированная серия изображений в формате DICOM, которая:</p> <ul style="list-style-type: none"> - имеет количество срезов аналогичное исходной серии для возможности синхронного просмотра врачом; - обладает информацией на каждом срезе о названии ИИИ сервиса, версии, диагностической точности с датой ее проверки и точным временем завершения анализа конкретного исследования; - допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок) <p><i>Пожалуйста, предоставьте списки и примеры выводных файлов, которые ваш продукт предоставляет в клинической практике</i></p>	<p>Да Нет</p>	
7. Контракт	7.1. Регулярные обновления системы, включая обновление информации о диагностической точности	<p>Да Нет</p>	
	7.2. Обновления ИИИ сервиса включены в стоимость	<p>Да Нет</p>	
	7.3. Все медицинские данные, сопутствующие материалы и результаты работы ИИИ сервиса являются собственностью заказчика	<p>Да Нет</p>	
ФИО заполнившего анкету			

Приложение 4

Радиология Москвы: Анкета для юридического лица-разработчика ПО на основе ИИ для лучевой диагностики

Общие требования:

- рабочая станция врача-рентгенолога, поддерживающая визуализацию результатов работы алгоритмов ИИ сервиса;
- возможность во время очной встречи загрузить и обработать ИИ сервисом исследования;
- возможность хронометража обработки ИИ сервисом от загрузки до предоставления результата;
- информация об архитектуре используемых моделей;
- доступ к коду;
- присутствие разработчика, специалиста по анализу больших массивов данных, которые могут проконсультировать по коду и структуре алгоритмов, процессу обучения и валидации алгоритма(-ов);
- присутствие врача, использующего ИИ сервис в своей практике, либо документы, подтверждающие факт успешной клинической апробации с контактной информацией (опционально).

Оценка клинической значимости

- Какую клиническую задачу решает ИИ сервис?
- Является ли применение ИИ сервиса для выбранной клинической задачи обоснованным с точки зрения доказательной медицины?
- Как влияет результат работы ИИ сервиса на тактику ведения пациента?
- Повышается ли качество работы врача?
- Повышается ли скорость работы врача?

Аналитическая валидация

А) Техническая готовность

- Самостоятельный тест.
- Хронометраж, замер времени на обработку 1-го исследования.
- Формат выходных данных.
- Визуализация результатов работы ИИ сервиса на рабочей станции врача-рентгенолога.
- Формат входных данных (DICOM, JPEG, PNG, BMP, AVI (копия)).

Б) Детали создания и применения ИИ модели

- Какие базы данных использовались при построении ИИ моделей (из

открытых ресурсов, от внешних партнеров, собственные, аугментированные)?

- Баланс классов в базах данных.
- Общее количество данных, использованных для построения ИИ модели для каждой клинической задачи.
- Архитектура ИИ модели.
- Библиотеки, ресурсы, использованные для построения модели.
- Подходы к разработке модели (from scratch, transfer learning, retraining of existing architecture).
- Тип машинного обучения (supervised / unsupervised / semi-supervised / weak supervision).
- Валидация модели (independent validation set, cross-validation).
- Длительность создания ИИ моделей. Имеющиеся вычислительные ресурсы.
- Происходит ли дообучение ИИ моделей? Как организована версияность?
- ИИ сервис работает с другими данными, кроме изображений?
- ИИ сервис способен оценить техническое качество проведения исследования?
 - Есть ли предобработка изображений?
 - Способен ли ИИ сервис оценить качество интерпретации исследования (т.е. контроль работы врача-рентгенолога)?
 - Значения метрик оценки точности работы ИИ сервиса для каждой клинической задачи (AUC, sens, spec, etc.)
 - Сравнились ли ИИ модели с существующими аналогами для решения таких же клинических задач?

В) Детали использованных наборов данных и их разметки для разработки ИИ моделей

- Были ли критерии для подборки базы данных для обучения или нет? Какие именно критерии (пол, возраст, болезни, МКБ, соотношение патологии и нормы)?
- Были ли в базе для обучения разные исследования одного пациента (чтобы обучаться распознавать динамику)?
- Что в каждом клиническом сценарии считалось «ground truth» в размеченной базе данных?
- Были ли критерии контроля качества собранной базы данных перед разметкой?
 - Какой бизнес-процесс разметки?
 - Проходила ли подборка разметчиков? Критерии?
 - Был ли контроль качества входных данных и результатов разметки?

- Сколько человек участвовали в разметке данных (сколько человек на 1 исследование; была ли выполнена разметка одного исследования одним специалистом дважды)?
 - Проходили ли разметчики обучение перед разметкой?
 - ПО для разметки свое или внешнее? Какое?
 - Подход в случае несогласованности разметчиков.
 - Бизнес-процесс работы эксперта, подтверждающего находку:
 - рабочий список;
 - доразметка;
 - цифровой след (серия комментариев от эксперта к разметчику).
 - Использовалась ли автоматизация процесса разметки (CAD для оконтуривания)?
 - Мониторинг процесса разметки.

Клиническая валидация

А) Практическое применение в клинической работе

- Происходила ли интеграция с PACS в медицинской организации? Если нет, то каким образом происходила интеграция в рабочий процесс?
- Какой бизнес-процесс работы врача-рентгенолога с ИИ сервисом в практике?
 - Есть ли инструкции для врача по работе с ИИ сервисом?
 - Вопросы к врачу: как он обучался и есть ли инструкции применения ИИ сервиса и возможности для обратной связи?
 - Есть ли триаж исследований?
 - Есть ли локализация находки с помощью ИИ?
 - Есть ли классификация находки?
 - Есть ли сравнение исследований в динамике?
 - Есть ли формирование шаблона описания и/или заключения?
 - На каком этапе бизнес-процесса врач-рентгенолог получает результаты обработки ИИ сервисом?
 - Есть ли обратная связь от врача-рентгенолога по использованию ИИ? Какая? Если есть, то она обязательная?
 - Что происходит, если врач-рентгенолог и ИИ имеют разные мнения (гипердиагностика и гиподиагностика)? Каким образом принимается окончательное решение по исследованию?
 - Есть ли метрики использования ИИ в клинической практике? Какие (например, сокращение количества ошибок врачей-рентгенологов и сокращение времени работы по интерпретации одного исследования)?

Приложение 5

**Оценки составляющих корпоративной культуры
юридического лица-разработчика ПО на основе ИИ
(для лучевой диагностики) на соответствие корпоративной
культуре превосходного качества и организации (КПКО)¹⁰**

Пункт	Наименование	Баллы (20 – удовлетворительно, 15 – наличие не критических замечаний, 5 – наличие критических замечаний, 0 – нет результата)
А.	Безопасность пациентов. Демонстрация обязательств обеспечения безопасности пациента, выделяя данный принцип в качестве основного при принятии любых решений.	
Б.	Качество продукта. Демонстрация обязательств разрабатывать, тестировать и поддерживать медицинское программное обеспечение на самом высоком уровне качества.	
В.	Клиническая ответственность. Демонстрация ответственности в проведении клинических оценок с должным вниманием к пациент-ориентированным вопросам, включая разметку и «человеческий фактор».	
Г.	Обеспечение кибербезопасности. Демонстрация обязательств по обеспечению кибербезопасности, с активным привлечением к данному рода проблемам партнеров и посредников.	
Д.	Активная корпоративная культура. Демонстрация активного подхода к наблюдению и оценке потребностей пользователей, гибкости и непрерывного обучения.	
Итоговый балл (максимум 100)		

¹⁰ Developing a Software Precertification Program: A Working Model.V.1.0 – January 2019. – <https://www.fda.gov/media/119722/download>.

Приложение 6

Чек-лист для детального описания результатов клинических испытаний программного обеспечения, разработанного для лучевой диагностики на основе технологий искусственного интеллекта (на основе STARD 2015 [7] с дополнениями)

Раздел и тема	№	Пункт
НАЗВАНИЕ		
	1	Акцент на оценку диагностической точности программного обеспечения на основе технологий ИИ, предназначенного для лучевой диагностики
РЕЗЮМЕ		
	2	Структурированное изложение дизайна исследования, материалов и методов, результатов и заключения
ВВЕДЕНИЕ		
	3	Клинические и научные сведения, включающие практическое применение результатов нового (исследуемого) метода – индекс-теста
	4	Гипотеза, цели и конечные точки исследования
МЕТОДЫ		
Дизайн исследования	5	Проспективный или ретроспективный характер исследования
Выборка	6	Критерии включения в исследование (включая целевую патологию)
	7	На основании каких критериев проведен отбор пациентов (симптомы, результаты обследований и т.д.)
	8	Где и когда сформирован набор биомедицинских данных с указанием ключевых характеристик (государственная регистрация, характеристика популяции, сведения о деперсонализации, целевая патология и диагностические группы с распределением, характеристика исследований, верификация, методология разметки)
	9	Запланированный размер выборки и способ его определения
	10a	Выборка сформирована последовательно, случайно или другими способами. Обоснование размера выборки

Методы исследования	10b	Информация об индекс-тесте в достаточном для воспроизведения результатов объеме (включая сведения об использованных при разработке математических моделях, нейронных сетях, методах машинного обучения, выборках для обучения и калибровки)
	11	Информация о референс-тесте («золотом стандарте») в достаточном для воспроизведения результатов объеме (референс-тестом является размеченный набор данных (эталонный))
	12a	Обоснование выбора референс-теста (методологии разметки, всего набора данных)
	12b	Порог активации (точка отсечения, «cut-off») для индекс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации («cut-off»): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13a	Порог активации (порог решающего правила, «cut-off») для референс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации: активации (cut-off): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13b	Была ли доступна участвующим в проведении/анализе индекс-теста критичная клинично-лабораторная информация, включая результаты применения референс-теста (было ли исследование слепыми или нет, во втором случае – обоснование причины)
	14	Была ли доступна участвующим в разметке и анализе эталонных данных критичная клинично-лабораторная информация, включая результаты применения индекс-тест (было ли исследование слепыми или нет, во втором случае – обоснование причины)
Анализ	15	Методы оценки и сравнения диагностической точности (метрики)
	16	Тактика при неоднозначных результатах индекс-теста или референс-теста
	17	Тактика при потере данных индекс-теста или референс-теста
	18	Анализ причин вариабельности диагностической точности с разделением на «ожидаемые» и «непредвиденные»

РЕЗУЛЬТАТЫ		
Пациенты	19	Характеристика пациентов: диспозиция пациентов, определение популяции для анализа, число пациентов, включенных в исследование, число досрочно выбывших, число пропущенных или негодных данных. Рекомендуется в дополнение к текстовому описанию представить диаграмму CONSORT
	20	Демографические и клинические характеристики выборки на момент первичного обследования (всех субъектов, а также субъектов, у которых есть результаты обоих тестов)
	21a	Распределение по степени тяжести целевой патологии
	21b	Распределение прочих диагнозов (по Международной классификации болезней) у пациентов без целевой патологии
	22	Любые значимые различия в методиках проведения индекс-теста и референс-теста
Результаты	23	Комбинированная таблица результатов индекс-теста и референс-теста
	24	Подсчет метрик диагностической точности с определенным доверительным интервалом (например, 95%)
	25	Любые побочные эффекты вследствие индекс-теста или референс-теста
ОБСУЖДЕНИЕ		
	26	Ограничения исследования, включая источники систематических ошибок, статистических неточностей и обобщений
	27	Клиническая ценность, включая планируемое применение индекс-теста
ПРОЧЕЕ		
	28	Сведения о регистрации исследования (любого этапа)
	29	Где можно получить полные сведения о разработке (включая информацию о наборе биомедицинских данных для обучения и калибровки)
	30	Источники финансирования и прочее

Приложение 7

Рекомендованная форма отчета о предварительных клинико-технических испытаниях (ПКТИ) программного обеспечения на основе ИИ для лучевой диагностики

№	Пункт	Пояснение
A.	Учреждение	Полное название учреждения, на базе которого проведены ПКТИ
2	Контактная информация	Контактная информация учреждения (адрес, телефон, сайт, электронная почта)
3	Даты проведения ПКТИ	Диапазон дат проведения ПКТИ
4	Резюме	Структурированное изложение дизайна исследования, материалов, методов, результатов и выводов
5	Цель, задачи и конечные точки ПКТИ	Цель, задачи и конечные точки ПКТИ
6	Референс-тест (эталонный набор данных)	Детальная характеристика использованных для ПКТИ эталонных наборов подготовленных данных
6.1	Вид данных	Вид(ы) данных (медицинская документация, результаты исследований и проч.), модальности для исследований, прочие характеристики
6.2	Количество включенных клинических случаев	Количество включенных клинических случаев (пациентов, результатов исследований и проч.)
6.3	Характеристика популяции	Характеристика популяции (расовые, гендерно-демографические, иные характеристики)
6.4	Характеристики набора данных и разметки	Где и когда сформирован эталонный набор данных с указанием ключевых характеристик (государственная регистрация базы данных, сведения о деперсонализации, о наличии информированного добровольного согласия пациентов, критерии включения/исключения, источники клинических случаев). Способ подготовки (разметки) набора данных
6.5	Характеристика патологии в наборе данных	Целевая патология и диагностические группы с распределением по ним. Способ верификации и наличие соответствующей информации в наборе данных

6.6	Способ формирования набора данных	Сформирован последовательно, случайно или другими способами. Обоснование размера выборки.
6.7	Источники данных	Количество и локализация медицинских учреждений, послуживших источниками клинических случаев, включенных в ПКТИ
6.8	Отметка о независимости ПКТИ	Набор данных, использованный для ПКТИ, не должен – как полностью, так и частично – использоваться для обучения или калибровки индекса-теста
7	Индекс-тест (ПО на основе ТИИ)	Информация об использовании, инсталляции, организации доступа и проч. индекс-теста (исследуемого ПО на основе ТИИ)
8	Процесс ПКТИ	Краткое общее описание процесса исследования; возможно использование схем, диаграммы CONSORT
9	Таблица результатов	Комбинированная таблица результатов индекс-теста и референс-теста
10	Порог активации	Точка отсечения («cut-off») для референс-теста и для индекс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные»
11	Показатели диагностической точности	Метрики диагностической точности с 95% доверительным интервалом (чувствительность, специфичность, общая точность, площадь под характеристической кривой и т.д.)
12	Ограничения	Любые ограничения ПКТИ, включая источники систематических ошибок, статистических неточностей и обобщений. Любые значимые различия в методиках проведения индекс-теста и референс-теста
13	Выводы	Краткое обобщение результатов
14	Источники финансирования ПКТИ	Указание источника финансирования работ, связанных с проведением ПКТИ
15	Прочее	Произвольная дополнительная информация
16	Список исполнителей	Список сотрудников учреждения, проводивших ПКТИ (ФИО, должность, ученое звание/степень)
17	Дата подписания отчета	-

18	Подпись ответственного исполнителя	Личная подпись ответственного исполнителя ПКТИ, ФИО
19	Подпись руководителя учреждения	Личная подпись руководителя учреждения, ФИО
20	Печать учреждения	-



ДЛЯ ЗАМЕТОК