ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 72313— 2025

СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ИЗВЛЕЧЕНИЯ ДАННЫХ ИЗ НЕСТРУКТУРИРОВАННЫХ МЕДИЦИНСКИХ ЗАПИСЕЙ

Методы формирования набора данных для обучения и тестирования. Метрики оценки качества

Издание официальное

Москва
Российский институт стандартизации
2025

gost.e-ecolog.ru

Предисловие

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

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

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 29 октября 2025 г. № 1299-ст

4 ВВЕДЕН ВПЕРВЫЕ

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

© Оформление. ФГБУ «Институт стандартизации», 2025

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

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Сокращения 3
5 Извлечение данных из неструктурированных медицинских записей 4
6 Обучение и тестирование систем поддержки принятия врачебных решений 7
7 Метрики оценки качества работы систем поддержки принятия врачебных решений 9
8 Фрагменты тестовых наборов данных (демонстрационные тестовые наборы данных) 13
Библиография 14

Введение

Внедрение систем поддержки принятия врачебных решений в реальную клиническую практику является одним из важнейших направлений цифровой трансформации здравоохранения (см. [1]). Применение различных систем поддержки принятия врачебных решений помогает улучшить сбор и анализ данных, сокращает количество ошибок в медицинской документации, а также когнитивную нагрузку на медицинский персонал в части повседневных задач анализа имеющейся медицинской информации о пациенте и, как следствие, способствует повышению качества принятия клинических решений и эффективности лечебно-диагностических процессов (см. [2], [3]).

Несмотря на доказанную перспективность внедрения систем поддержки принятия врачебных решений реальная эффективность данного направления существенным образом зависит от качества работы конкретных продуктов, выводимых разработчиками и производителями на рынок. В свою очередь это качество зависит от качества исходных материалов, используемых в процессах создания и тестирования продуктов, включая создание отдельных алгоритмов и моделей машинного обучения.

Полученные в условиях национального здравоохранения результаты применения различных систем искусственного интеллекта убедительно доказывают, насколько важны вопросы стандартизации и тестирования продуктов и моделей машинного обучения (см. [4]). В этой связи стандартизация различных процессов создания систем поддержки принятия врачебных решений является одним из важных этапов обеспечения приемлемого качества, эффективности и безопасности работы таких систем.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ИЗВЛЕЧЕНИЯ ДАННЫХ ИЗ НЕСТРУКТУРИРОВАННЫХ МЕДИЦИНСКИХ ЗАПИСЕЙ

Методы формирования набора данных для обучения и тестирования. Метрики оценки качества

Artificial intelligence-powered clinical decision support systems for data extraction from unstructured medical records. Curation of training and validation datasets. Quality metrics

Дата введения — 2026—01—01

1 Область применения

Настоящий стандарт устанавливает требования к разработке и тестированию систем поддержки принятия врачебных решений в части:

- процесса извлечения данных из неструктурированных медицинских записей;
- методов формирования наборов данных для целей обучения и тестирования систем поддержки принятия врачебных решений;
- метрик оценки качества работы систем поддержки принятия врачебных решений.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 59921.0 Системы искусственного интеллекта в клинической медицине. Основные положения

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

ГОСТ Р 59921.11—2025 Системы искусственного интеллекта в клинической медицине. Наборы данных для тестирования алгоритмов. Методы контроля набора данных на универсальность и структурированность

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.

Издание официальное

3 Термины и определения

В настоящем стандарте применены термины по ГОСТ Р 59921.0, а также следующие термины с соответствующими определениями:

3.1

алгоритм системы поддержки принятия врачебных решений; алгоритм СППВР: Конечно-упорядоченное множество точно определенных правил для решения конкретной задачи в системе поддержки принятия врачебных решений.

Примечание — Алгоритм СППВР может быть представлен моделью машинного обучения, математической моделью (формулой), последовательностью инструкций по обработке входных данных или иной программной реализацией. Данное определение основано на общем определении алгоритма, предусмотренном ГОСТ 33707—2016 (пункт 4.39)[1)].

[ГОСТ Р 71671—2024, пункт 2.2]

3.2

данные: Предоставление информации в формальном виде, пригодном для передачи, интерпретации или обработки людьми или компьютерами.

[ГОСТ 33707—2016, пункт 4.259]

3.3 **жизненный цикл:** Развитие системы, продукции, услуги, проекта или другой создаваемой человеком сущности от замысла до списания.

3.4

искусственный интеллект: Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение, поиск решений без заранее заданного алгоритма и достижение инсайта) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

Примечание — 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, анализу и синтезу решений.

[ГОСТ Р 59277—2020, пункт 3.18]

3.5

машинное обучение: Процесс автоматического обучения и совершенствования поведения системы искусственного интеллекта на основе обработки массива обучающих данных без явного программирования.

[ГОСТ Р 59895—2021, пункт 2.1.7]

3.6 **медицинская информационная система медицинской организации;** МИС МО: Программное обеспечение, предназначенное для сбора, хранения, обработки и предоставления информации, необходимой для автоматизации процессов оказания и учета медицинской помощи и информационной поддержки медицинских работников, включая информацию о пациентах, об оказываемой им медицинской помощи и о медицинской деятельности медицинских организаций.

Примечание — Термин адаптирован из [5].

[1)] ГОСТ 33707—2016 (ISO/IEC 2382:2015) «Информационные технологии. Словарь».

3.7

метрика (в подходах к качеству): Материальная мера некоторых аспектов характеристик качества.

Примечания

1 Другими словами, это способ назначения определенного значения с использованием методов измерения или тестирования для количественной оценки объекта качества, как шкала критерий, степень, вес, магнитуда, отношение или утвержденная форма и т. д.

2 В [1][1)] «материальная мера» определяется как устройство, постоянно воспроизводящее или поставляющее во время своего использования количество заданных видов, каждое с назначенным значением.

3 Допускается применение синонима «метрика качества» для унификации с другими терминами и определениями настоящего стандарта.

[Адаптировано из ГОСТ Р 54837—2011, пункт 4.4]

3.8

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 интеллекта.

Примечание — Слово «необходимости» заменено на «необходимых» для унификации с другими терминами и определениями настоящего стандарта.

[Адаптировано из ГОСТ Р 59921.4—2021, пункт 3.5]

3.9

программное обеспечение (программа; программное средство): Упорядоченная последовательность инструкций (кодов) для вычислительного средства, находящаяся в памяти этого средства и представляющая собой описание алгоритма управления вычислительными средствами и действий с данными.

[ГОСТ Р 53622—2009, пункт 3.8]

3.10

система поддержки принятия врачебных решений; СППВР: Программное обеспечение, позволяющее путем обработки и интерпретации собираемой информации на основе алгоритмов поддерживать принятие врачом решения на всех этапах лечебно-диагностического процесса с целью снижения ошибок и повышения качества оказываемой медицинской помощи.

[ГОСТ Р 71671—2024, пункт 2.1]

4 Сокращения

В настоящем стандарте использованы следующие сокращения:

БД — база данных;

ИИ — искусственный интеллект;

ИС — информационная система;

НД — набор данных;

НПА — нормативный правовой акт;

ЭДО — электронный документооборот;

ЭМК — электронная медицинская карта;

NLP — обработка естественного языка (natural language processing).

[1)] Руководство ИСО/МЭК 99:2007, Международный словарь по метрологии. Основные и общие понятия и соответствующие термины (VIM) (International vocabulary of metrology — Basic and general concepts and associated terms (VIM)).

5 Извлечение данных из неструктурированных медицинских записей

5.1 Общее описание процессов использования электронных медицинских карт как источника данных для работы систем поддержки принятия врачебных решений

5.1.1 Системы ведения ЭМК, включая входящие в них медицинские документы, являются основным источником данных для работы СППВР.

5.1.2 Интерпретация ЭМК путем использования алгоритмов СППВР, включая модели машинного обучения, является сложной научной и технологической задачей в силу следующих особенностей ЭМК как источника данных (см. [6]):

- применение разработчиками функций повторного использования (копирования) однажды внесенных данных в новые записи, приводящее к наличию дублированной и непроверенной врачом информации;
- децентрализованный характер систем ведения ЭМК, многие из которых созданы на устаревших технологиях и используют локально-устанавливаемые в медицинских организациях сервера без возможности ведения единой общей БД, приводящий к фрагментированному хранению медицинской информации в таких ЭМК;
- отсутствие единой нормативно-справочной информации для кодирования клинической информации в медицинских документах, включая врачебные осмотры и протоколы инструментального исследования;
- некачественное заполнение экранных форм пользователями;
- отсутствие форматно-логического контроля при вводе данных пользователями для формирования немедленной обратной связи и принудительное исправление пользовательских ошибок;
- преимущественно текстовый способ внесения информации в медицинские документы, приводящий к хранению неструктурированных медицинских записей, требующих предварительной обработки для получения машинно-читаемых и машинно-понимаемых признаков.

5.1.3 Для обеспечения приемлемого качества работы СППВР необходимо, чтобы в СППВР был реализован непрерывный контроль качества входных данных, используемых для анализа и интерпретации, включая форматно-логический контроль заполненности, непротиворечивости, наличия ошибочных или аномальных значений и т. д.

5.1.4 Для анализа и интерпретации данных СППВР получает их:

- из пользовательского интерфейса или системы ведения ЭМК, которая, в свою очередь, обеспечивает интерфейс с пользователем и ввод изначально структурированных или частично-структурированных данных;
- из неструктурированных записей, которые также могут быть получены из собственного пользовательского интерфейса или через интеграцию системы ведения ЭМК с СППВР.

5.1.5 В зависимости от принятой архитектуры и перечня функциональных возможностей, реализованных в СППВР, извлечение данных из неструктурированных медицинских записей возможно путем онлайн-извлечения, анализа и обработки, а также отложенного извлечения, анализа и обработки.

При онлайн-извлечении, анализе и обработке пользователь вносит информацию в соответствующий интерфейс ЭМК или СППВР, введенные сведения сразу же отправляются для анализа, полученный результат возвращается и визуализируется в интерфейсе пользователя.

При отложенном извлечении, анализе и обработке накопленная в ЭМК информация первоначально загружается в СППВР для обработки, в рамках которой осуществляется извлечение и анализ данных, формирование и сохранение результатов интерпретации. Пользователь в интерфейсе ЭМК или СППВР дает команду для получения результатов работы СППВР, которые были перед этим предварительно подготовлены.

5.1.6 При применении отложенного извлечения, анализа и обработки данных из неструктурированных медицинских записей, входящих в состав ЭМК, следует придерживаться трехуровневой архитектуры СППВР, условно представленной на рисунке 1. Каждый уровень предназначен для своей конкретной задачи. Последующие уровни в качестве входных данных используют результаты работы предыдущего уровня.



Рисунок 1— Архитектура сбора, извлечения и хранения данных в СППВР

5.1.7 Данные, собранные в итоге в цифровом профиле пациента, могут быть исходным источником для работы моделей машинного обучения и иных алгоритмов СППВР, которые на выходе выдают результат своей работы — клинически значимые новые знания о пациенте для врача.

5.1.8 Программная реализация предложенной архитектуры СППВР может включать несколько последовательных процессов работы с данными, приведенных на рисунке 2.



Рисунок 2 — Процессы сбора, извлечения и подготовки данных в СППВР

5.2 Порядок сбора медицинских записей от поставщиков

5.2.1 Источником медицинских записей для работы СППВР могут быть различные информационные системы (поставщики записей), включая:

а) МИС МО, в которых осуществляется ведение ЭМК;

б) иные информационные системы в сфере здравоохранения, в которых накапливаются медицинские записи:

1) телемедицинские системы;

2) системы дистанционного мониторинга;

3) решения для предрейсовых и предсменных осмотров и т. д.;

в) прочие информационные системы, такие как персональные медицинские помощники и т. д.

5.2.2 Все полученные от поставщиков медицинские записи принимаются строго в обезличенном виде. Процесс обезличивания медицинских записей должен осуществляться в защищенной инфраструктуре соответствующего оператора персональных данных (медицинской организации или иной организации, если данные были собраны не в рамках медицинской организации).

5.2.3 При получении записей от поставщиков СППВР должна выполнить проверку на минимальную полноту и достоверность предоставленных сведений. Например, СППВР не должна регистрировать медицинскую запись и извлекать из нее данные, если не известен какой-либо из обязательных параметров, например возраст и пол.

5.2.4 Прошедшие проверку медицинские записи сохраняются в СППВР в неизменном виде. Все данные, которые СППВР извлечет из полученных записей, должны быть сохранены в базе данных СППВР в отдельном разделе.

5.2.5 Запрещается внесение изменений со стороны СППВР в полученные для обработки медицинские записи, они должны оставаться неизменными в течение всего жизненного цикла СППВР. Полученные в результате работы СППВР данные, включая извлеченные признаки и итоговые рекомендации для врача, должны быть сохранены отдельно от исходных медицинских записей.

5.2.6 Полученные от поставщиков медицинские записи могут содержать данные, необходимые для работы СППВР, в виде структурированных и неструктурированных данных.

В случае структурированных данных каждый признак передается отдельным полем, содержащим значение данного признака, а также дополнительными метаданными, включая клинически значимые атрибуты, например дата извлечения признака.

В случае неструктурированных данных медицинская запись содержит одно или несколько полей, в которых хранится текстовая информация. Для получения признаков из такой информации требуется использование дополнительных алгоритмов извлечения данных.

5.3 Порядок извлечения данных из медицинских записей

5.3.1 Основным объектом для работы алгоритмов СППВР является признак — достаточное условие для принадлежности объекта некоторому классу. Каждый признак содержит данные, позволяющие сделать вывод о наличии интересующего явления.

5.3.2 Для извлечения признаков из неструктурированных медицинских записей могут быть использованы следующие методы:

- программное сопоставление кода признака со специальным справочником кодов, встроенным в СППВР и настраиваемым специально под различные особенности работы систем ведения ЭМК и иных источников медицинских записей, в том числе с учетом особенностей конкретного продукта, используемых форматов кодирования и хранения записей и других технических особенностей;
- извлечение с помощью технологий NLP и соответствующих моделей машинного обучения;
- расчет значения признака [используется в случае, если СППВР имеет какие-то значения признаков, на основании которых она самостоятельно рассчитывает значение другого признака, например СППВР может извлечь признак «Индекс массы тела» (ИМТ) путем расчета по соответствующей формуле на основании числовых данных о росте и весе];
- извлечение программной функцией (СППВР особым алгоритмом, реализованным в коде, извлекает значение признака из исходной информации, например таким образом извлекаются признаки из ответов пациентов на вопросы анкеты диспансеризации);
- иные методы, включая гибридный подход.

5.3.3 Извлеченные для работы СППВР признаки могут быть отнесены к одному из следующих видов:

- демографические данные, например пол, возраст, регион проживания, данные о месте работы, социальный статус и т. д.;
- данные об обращении в медицинские организации, включая даты и типы случаев обращения, исходы и результаты лечения, коды диагнозов и т. д.;
- данные о зарегистрированных заболеваниях, включая коды МКБ-10 [7], даты первой регистрации заболевания и последнего обращения, тип заболевания (острое или хроническое) и т. д.;
- клинические данные, включая физиологические данные и данные объективного осмотра пациента (например, артериальное давление, рост, вес, частота сердечных сокращений и т. д.), а также симптомы и синдромы, извлеченные из врачебных осмотров (например, курение, головная боль, отягощенный анамнез и т. д.);

- лабораторные данные, например общий холестерин, глюкоза крови, креатинин и т. д.;
- данные инструментальных исследований, которые извлекаются из протоколов обследований, включая ультразвуковое исследование, электрокардиограмму, компьютерную томографию, рентгенографию, маммографию и т. д.;
- данные о назначенном и проведенном лекарственном лечении, включая наименование лекарственных препаратов, кратность приема, лекарственную форму, дозу и т. д.;
- данные об эффективности и безопасности назначенных лекарственных препаратов и прочие.

5.3.4 Извлеченные признаки могут иметь следующие типы значений:

- бинарные, например курение, прием антигипертензивных препаратов и т. д.;
- числовые признаки, например температура, частота сердечных сокращений, артериальное давление, рост, вес, значения лабораторных тестов и т. д.;
- текстовые, например формулировка диагноза, название лекарственного средства, название симптома и т. д.;
- даты, например дата рождения, дата регистрации пациента, дата смерти и т. д.;
- время, например время госпитализации, время измерения какого-то признака;
- категориальные, представленные кодами значений из справочников, например код заболевания по МКБ-10 [7], код симптома, код пола и т. д.;
- онтологические, представленные кодами значений из онтологической базы знаний.

5.3.5 Извлеченные в ходе работы СППВР признаки должны быть сохранены в БД СППВР отдельно от хранилища полученных для обработки исходных медицинских записей.

6 Обучение и тестирование систем поддержки принятия врачебных решений

6.1 Общие требования к наборам данных для обучения систем поддержки принятия врачебных решений

6.1.1 Основная задача создания СППВР состоит в получении отражающего уровня потребности врачей программного средства для решения конкретных задач, предусмотренных спецификацией производителя.

6.1.2 Обработка входных данных, полученных СППВР во время эксплуатации, осуществляется с помощью соответствующих алгоритмов СППВР, разработанных для решения конкретной клинической задачи (например, выявления подозрения на заболевание, получение прогноза, оценка риска, контроль данных и т. д.).

6.1.3 Одной из главных задач создания алгоритма СППВР является выбор и подготовка материала, отражающего современный уровень знаний и опыта специалистов (врачей, экспертов, ученых, педагогов). Такими материалами могут быть:

- утвержденные клинические рекомендации;
- научные публикации в специализированных журналах;
- отчеты Всемирной организации здравоохранения по программе контроля безопасности и эффективности медицинских изделий («The WHO prequalification project»), размещенными на сайте Всемирной организации здравоохранения в информационно-телекоммуникационной сети Интернет;
- учебные материалы медицинских университетов;
- базы знаний;
- наборы данных, отражающие процессы диагностики и лечения (заболевания и выздоровления) реальных пациентов.

6.1.4 При использовании технологий машинного обучения для создания алгоритмов СППВР важнейшим этапом с точки зрения безопасности, эффективности и качества работы СППВР является подготовка соответствующих НД. Метрики работы моделей машинного обучения, используемых в СППВР, будут главным образом зависеть от качества используемых для обучения НД.

6.1.5 Процессы подготовки НД и машинного обучения на их основе (и других способов имитировать когнитивные функции врача) должны осуществляться подготовленным персоналом, имеющим представление о технологиях ИИ, методах и алгоритмах машинного обучения, методах работы с данными, а также детальное представление о создаваемой СППВР, включая ее предназначение, условия применения и решаемые данным продуктом задачи.

6.2 Общие требования к подготовке наборов данных для тестирования систем поддержки принятия врачебных решений

6.2.1 Для обеспечения требуемого уровня доверия к работе СППВР необходимо применение комплексного подхода к тестированию и контролю качества, включающего процессы, перечисленные в 6.2.1.1—6.2.1.3.

6.2.1.1 Внутреннее тестирование СППВР, предназначенное для проверки соответствия продукта технической документации производителя, проводят на тестовых НД. Подтвержденное прохождение внутреннего тестирования свидетельствует о готовности продукта к проведению клинических исследований или к следующему этапу жизненного цикла.

6.2.1.2 Внешнее тестирование, включая внешнюю клиническую валидацию, предназначено для независимой проверки достижения заявленных производителем метрик работы СППВР, в том числе в рамках процедур регистрации СППВР в качестве медицинского изделия. Для этого испытания необходимо использовать верифицированный НД в медицинской организации, проводящей клинические испытания.

6.2.1.3 Валидация во время эксплуатации (мониторинг) предназначена для периодической проверки соблюдения заявленных производителем метрик работы СППВР во время применения в реальной клинической практике. В этом испытании допускается использовать верифицированный НД организации, проводящей мониторинг, либо результаты проверки доступности работы СППВР, например количество положительных ответов на обращения к СППВР, количество сбоев и т. д.

6.2.2 Помимо процессов клинических испытаний, включая тестирование и валидацию, целесообразно поддерживать процессы независимых клинических исследований, направленных на сбор дополнительных данных эффективности, безопасности и качества работы СППВР, в том числе с целью оценки клинической, организационной и финансово-экономической целесообразности внедрения и развития СППВР.

6.2.3 Клинические испытания СППВР, включая внутреннее тестирование, должен осуществлять подготовленный персонал, имеющий представление как о методологических основах проведения клинических испытаний, используемых методах и показателях качества работы СППВР, так и общее представление о предназначении и особенностях реализации тестируемой СППВР.

6.2.4 Внутреннее тестирование СППВР выполняется путем использования заранее подготовленного НД, содержащего входные признаки и сопоставленные им выходные признаки (эталонные значения). Внутреннее тестирование необходимо проводить как при разработке СППВР, так и при подготовке к выпуску новых версий СППВР и/или отдельных алгоритмов, входящих в состав СППВР. Требования к процессу внутреннего тестирования должны быть определены таким образом, чтобы можно было убедиться в достижении заявленных в технической документации метрик.

6.2.5 Подготовку НД выполняют по ГОСТ Р 59921.5.

6.2.6 Валидационный НД для внутреннего и внешнего тестирования должен включать тщательно подобранные и задокументированные клинические случаи.

6.2.7 Отдельный клинический случай (запись НД) должен содержать:

а) входные признаки, которые могут быть разными для различных этапов тестируемого клинического процесса;

б) выходные события, представленные соответствующими признаками.

Примечание — Например, для тестирования функции выявления подозрения на заболевание выходным событием НД будет код диагноза.

6.2.8 В зависимости от тестируемой задачи и проверяемых знаний нозологической группы на этапе подбора наборов клинических случаев для испытаний потребуется разное общее количество и разное процентное соотношение простых, «стертых» и коморбидных случаев.

6.2.9 При подготовке валидационных НД необходимо обеспечить участие экспертов или материалов из разных медицинских школ и/или разных регионов, способных поставлять качественно задокументированные случаи оказания медицинской помощи, в противном случае потребуется перепроверка правильности работы СППВР в каждой медицинской организации, которая приступает к внедрению СППВР. Например, у людей в особых климатических зонах могут быть отличия в проявлениях по некоторым болезням, в ответе на лечение и т. п.

6.2.10 Для внутреннего тестирования СППВР НД подготавливает производитель соответствующего продукта. Записи НД для тестирования не могут содержать данные, использованные в машинном обучении.

6.2.11 Для внешней клинической валидации НД подготавливает независимая от производителя медицинская организация. Должно быть исключено использование НД или отдельных записей, доступных и использованных на этапе создания алгоритмов СППВР, а также на этапе внутреннего тестирования.

6.2.12 При подготовке НД для тестирования учитывают требования ГОСТ Р 59921.11—2025 (пункт 5.2.2). Основным является определение числа записей НД, необходимого и достаточного для формирования надежных выводов о качестве работы СППВР. При подготовке НД должна быть обеспечена репрезентативность выборки. В связи с тем, что обеспечение репрезентативности выборки вероятностным способом ее формирования (случайный, систематический и кластерный отбор) сложно достигаемо, используют такие невероятностные способы формирования НД, как выборка удобства (произвольная), последовательная (сплошная) выборка, выборка добровольцев, квотный отбор и т. д. Необходимо учитывать относительную частоту состояния в целевой популяции (долю, частоту) диагностируемого (или прогнозируемого) состояния в используемом для проведения испытаний НД, т. к. именно от него зависят наиболее значимые метрики качества СППВР.

6.2.13 Размер выборки для тестирования определяется и обосновывается исходя из заданных параметров и особенностей работы алгоритма СППВР или всей системы в целом, а также с учетом целевой статистической мощности и обеспечения необходимого уровня безопасности, эффективности и качества работы СППВР в условиях реальной клинической практики.

6.3 Требования к использованию наборов данных при тестировании систем поддержки принятия врачебных решений

6.3.1 Каждый этап процесса тестирования должен состоять в том, что записи из НД подают на вход алгоритма СППВР или всей СППВР в целом. Далее СППВР или алгоритм СППВР обрабатывает полученные данные и выдает результат, который затем сохраняется в используемом НД.

6.3.2 Для каждой использованной строки данных из НД при проведении испытания фиксируют полученный результат СППВР. Полученные данные используют для расчета матрицы ошибок, которая далее используется для расчета соответствующих метрик качества.

6.3.3 В ходе обработки полученной записи из НД СППВР может осуществлять форматно-логический контроль полученных данных на предмет выявления несоответствий (например, полученные записи не соответствуют критериям работы СППВР, в данных содержатся пропуски или аномалии). При выявлении несоответствия СППВР также возвращает результат обработки данных, однако этот результат должен содержать не результат интерпретации, а сообщение об ошибке. В ходе тестирования такие ответы СППВР также записываются в тестовый НД, но не участвуют в дальнейшем в расчете матрицы ошибок работы СППВР и, соответственно, расчете точности работы алгоритмов СППВР.

7 Метрики оценки качества работы систем поддержки принятия врачебных решений

7.1 Требования к определению метрик качества работы систем поддержки принятия врачебных решений

7.1.1 Метрики определяются для каждого алгоритма СППВР в отдельности. Результаты тестирования СППВР представляют в виде итоговой таблицы достигнутых метрик для всех алгоритмов, включенных в состав СППВР.

7.1.2 По итогам проведения тестирования для каждой записи в тестовом НД, которая была отправлена для интерпретации в СППВР и для которой получен ответ, определятся результат сравнения ответа с правильным значением, который может принимать одно из следующих значений:

- истинно положительный (ИП) — алгоритм СППВР выявил патологию при ее наличии (для диагностической задачи) или правильно предсказал событие при его наступлении (для прогностической задачи);
- ложноположительный (ЛП) — алгоритм СППВР выявил патологию при ее отсутствии (для диагностической задачи) или ошибочно предсказал событие, хотя оно не наступило (для прогностической задачи);

- ложноотрицательный (ЛО) — алгоритм СППВР не выявил патологию при ее наличии (для диагностической задачи) или не предсказал событие при его наступлении (для прогностической задачи);
- истинно отрицательный (ИО) — алгоритм СППВР не выявил патологию при ее отсутствии (для диагностической задачи) или не предсказал событие, которое не наступило (для прогностической задачи).

7.1.3 После определения результатов тестирования в записях НД рассчитывается матрица ошибок, содержащая абсолютные значения истинно положительных, ложноположительных, ложноотрицательных и истинно отрицательных результатов.

7.1.4 На основании матрицы ошибок для каждого алгоритма СППВР в обязательном порядке должны быть определены (на этапе создания) и оценены (на этапах тестирования) следующие метрики:

а) площадь под характеристической кривой (AUC ROC), показывающая общие классификационные способности алгоритма СППВР;

б) чувствительность Чв (sensitivity, Se, recall), показывающая вероятность того, что алгоритм СППВР будет правильно выявлять пациентов с патологией или правильно предсказывать прогнозируемое событие, вычисляемая по формуле

$$Чв = ИП / (ИП + ЛО); \quad (1)$$

в) специфичность Сп (specificity, Sp), показывающая вероятность того, что алгоритм СППВР будет правильно выявлять пациентов без патологии или правильно предсказывать отсутствие прогнозируемого события, вычисляемая по формуле

$$Сп = ИО / (ИО + ЛП); \quad (2)$$

г) общая точность алгоритма (accuracy), показывающая общую вероятность правильной работы алгоритма СППВР как при выявлении пациентов с патологией, так и без нее, а также правильность прогнозирования событий, вычисляемая по формуле

$$Общая\ точность = (ИП + ИО) / (ИП + ИО + ЛП + ЛО). \quad (3)$$

7.1.5 Кроме обязательных метрик целесообразно для каждого алгоритма СППВР определить (на этапе создания) и оценить (на этапах тестирования) дополнительные метрики, приведенные в 7.1.5.1—7.1.5.7.

7.1.5.1 Прогностическую ценность положительного результата ПЦПР (positive predictive value, PPV, precision), показывающую вероятность наличия патологии при ее выявлении алгоритмом СППВР или вероятность наступления события, если оно предсказано моделью прогнозирования, вычисляют по формуле

$$ПЦПР = ИП / (ИП + ЛП). \quad (4)$$

7.1.5.2 Прогностическую ценность отрицательного результата ПЦОР (negative predictive value, NPV), показывающую вероятность отсутствия патологии, если ее не выявил алгоритм СППВР, или вероятность не наступления события, если оно не было предсказано моделью прогнозирования, вычисляют по формуле

$$ПЦОР = ИО / (ИО + ЛО). \quad (5)$$

7.1.5.3 Отношение правдоподобия положительного результата ОППР, показывающее, во сколько раз при положительном результате алгоритма СППВР вероятность правильного ответа превышает вероятность его отсутствия, вычисляют по формуле

$$ОППР = Чв / (1 - Сп). \quad (6)$$

7.1.5.4 Отношение правдоподобия отрицательного результата ОПОР, показывающее, во сколько раз при отрицательном результате алгоритма СППВР вероятность правильного ответа превышает вероятность его отсутствия, вычисляют по формуле

$$ОПОР = (1 - Чв) / Сп. \quad (7)$$

7.1.5.5 Частоту ложных срабатываний ЧЛС, показывающую долю неверных результатов к общему числу записей за пределами класса, вычисляют по формуле

$$ЧЛС = 1 - Сп. \quad (8)$$

7.1.5.6 *F*-меру (совместную оценку чувствительности и ПЦПР) вычисляют по формуле

$$F\text{-мера} = 2 \cdot ПЦПР \cdot Чв / (ПЦПР + Чв). \quad (9)$$

7.1.5.7 Уровень совпадения диагнозов определяют по выражению

$$(n \cdot 100)/N, \quad (10)$$

где n — количество успешных событий, где результат СППВР совпал с основным диагнозом;

N — общее количество истинных диагнозов.

7.1.6 Для всех используемых метрик следует рассчитывать не только точечные оценки, но и интервальные — доверительные интервалы с доверительной вероятностью от 95 % до 99 %, т. е. те диапазоны, в которых фактически с той или иной вероятностью могут варьировать соответствующие метрики. При этом важно учитывать, что чувствительность и специфичность, так же, как и прогностические ценности положительного и отрицательного результатов, изменяются зависимо друг от друга, в связи с чем, оптимизируя один показатель, разработчик алгоритма СППВР, как правило, ухудшает другой.

7.1.7

Площадь под характеристической кривой не является бинарным показателем. ROC-анализ может быть проведен как в координатах ((1 − Сп); Чв), так и в координатах ((1 − ПЦОР); ПЦПР). Последний анализ предпочтительнее, т. к. ориентирован на врача — лицо, принимающее решение в отношении конкретного пациента. ROC-анализ используют для предварительного сравнения точности изучаемых алгоритмов СППВР, особенно если их много. Однако наличие AUC ROC недостаточно для оценки эффективности работы алгоритма СППВР. В связи с чем для каждого алгоритма СППВР должен быть определен порог активации бинарного классификатора (cut-off threshold) для классификации случаев, критериями оптимальности которой могут быть:

- минимум ошибки I рода (гипердиагностики) при приемлемой ошибке II рода (гиподиагностике);
- минимум ошибки II рода (гиподиагностики) при приемлемой ошибке I рода (гипердиагностике);
- баланс чувствительности и специфичности;
- максимизация суммы чувствительности и специфичности (критерий Юдена) и т. д.

[[8], пункт 5.4]

7.1.8 Если в ходе тестирования отдельно набирались позитивные и негативные случаи, необходима поправка на отдельную частоту состояния в целевой популяции. Вычисления проводят по формулам:

$$ПЦПР = Чв \cdot ОЧ / (Чв \cdot ОЧ + (1 - Сп) \cdot (1 - ОЧ)), \quad (11)$$

$$ПЦОР = Сп \cdot (1 - ОЧ) / (Сп \cdot (1 - ОЧ) + (1 - Чв) \cdot ОЧ). \quad (12)$$

7.1.9 Расчет общей точности алгоритма СППВР основан на отношении суммы истинно положительных и истинно отрицательных результатов к общему числу наблюдений в выборке. Для точности должны быть рассчитаны доверительные интервалы (99 %, 95 %). Точность зависит от относительной частоты состояния в целевой популяции и не может быть рассчитана для нерепрезентативной выборки без поправки на относительную частоту состояния в целевой популяции.

7.2 Требования к оценке метрик качества работы систем поддержки принятия врачебных решений

7.2.1 Показатель точности работы алгоритма СППВР учитывает чувствительность и специфичность, поэтому является довольно общим. В этой связи при принятии решений о качестве реализации алгоритма СППВР не допускается опираться только на этот показатель, необходимо всегда рассматривать его комплексно с учетом отдельных показателей чувствительности и специфичности.

7.2.2 Показатели чувствительности и специфичности устанавливаются и оцениваются для принятия решений следующим образом:

а) при создании СППВР — минимальные значения, достижение которых требуется обеспечить на этапах создания алгоритма СППВР для его последующего включения в продукт;

б) при внутреннем тестировании алгоритма СППВР — если предусмотренные технической документацией производителя метрики были достигнуты на данном этапе, алгоритм СППВР может быть включен в СППВР, если заявленные метрики чувствительности и специфичности не были достигнуты, необходима доработка алгоритма или принятие решения о невозможности его использования в СППВР;

в) при внешнем тестировании — если заявленные производителем метрики были достигнуты, принимается решение об успешном прохождении испытания, если заявленные метрики не были достигнуты, тестируемый алгоритм признается некачественным, в целом результаты испытания СППВР признаются неудовлетворительными.

7.2.3 Показатели общей точности, чувствительности и специфичности алгоритма СППВР устанавливаются и обосновываются на этапе разработки СППВР, исходя из особенностей решаемой клинической задачи и обеспечения необходимого уровня безопасности, эффективности и качества работы СППВР в условиях реальной клинической практики. При этом рекомендуется руководствоваться следующими критериями:

- значения общей точности, чувствительности или специфичности менее 0,6 — алгоритм СППВР имеет неприемлемо низкое значение, его использование в реальной клинической практике недопустимо;

- значения общей точности, чувствительности или специфичности находятся в пределах 0,61—0,8 — алгоритм СППВР имеет перспективное значение, но целесообразно его улучшение до более высокого значения;

- значения общей точности, чувствительности или специфичности более 0,81 — алгоритм СППВР имеет приемлемое значение, его можно использовать в условиях реальной клинической практики.

7.2.4 Для алгоритмов СППВР, предназначенных для выявления больных пациентов (диагностические алгоритмы, прогностические алгоритмы для выявления пациентов высокого риска), необходимо оптимизировать и оценивать показатель чувствительности, а также прогностической ценности положительного результата при соблюдении минимально-приемлемых значений специфичности и прогностической ценности отрицательного результата.

7.2.5 Для алгоритмов СППВР, предназначенных для выявления здоровых пациентов (скрининговые диагностические алгоритмы, прогностические алгоритмы для выявления пациентов низкого риска), необходимо оптимизировать и оценивать показатель специфичности и прогностической ценности отрицательного результата при соблюдении минимально-приемлемых значений чувствительности и прогностической ценности положительного результата.

Примечание — Возможно одностороннее использование алгоритма СППВР, например использование только для подтверждения искомого состояния (т. е. для диагностики). В этом случае значения чувствительности и прогностической ценности должны быть максимизированы и при этом цена ошибок II рода (гиподиагностики) должна быть приемлемой. Также возможна обратная ситуация — алгоритм СППВР может быть предназначен исключительно для скрининга, например оппортунистического. В этом случае необходимо максимизировать специфичность и прогностическую ценность отрицательного результата, допуская низкие значения чувствительности и прогностической ценности положительного результата, но при условии, что цена ошибок гипердиагностики приемлема.

7.2.6 Рекомендуется оценка и включение в техническую документацию на СППВР показателей прогностической ценности положительного и отрицательного результата, т. к. это позволяет оценивать результаты работы алгоритма СППВР при применении системы. Например, вывод в интерфейсе СППВР не только результатов работы алгоритмов СППВР, но показателей прогностической ценности позволяет предоставить врачу информацию о вероятности правильного результата работы алгоритма СППВР на данных конкретного больного, что может быть использовано для понимания врачом вероятности гипер- и гиподиагностики.

8 Фрагменты тестовых наборов данных (демонстрационные тестовые наборы данных)

Основной демонстрационный тестовый набор данных для испытания системы поддержки принятия врачебных решений, содержащий примеры записей с входными данными и результатами их интерпретации с использованием алгоритма СППВР на примере «Модели оценки риска наличия атеросклеротических бляшек брахиоцефальных артерий при избыточной массе тела и ожирении у пациентов без сердечно-сосудистых заболеваний», приведен в [9].

Библиография

[1] Sutton R.T., Pincock D., Baumgart D.C., Sadowski D.C., Fedorak R.N., Kroeker K.I. An overview of clinical decision support systems: benefits, risks, and strategies for success // NPJ Digit Med — 2020 — Feb 6. — 3:17

[2] Muhiyaddin R., Abd-Alrazaq A.A., Househ M., Alam T., Shah Z. The Impact of Clinical Decision Support Systems (CDSS) on Physicians: A Scoping Review // Stud Health Technol Inform — 2020 — Jun 26. — Vol. 272. — P. 470—473

[3] Borges do Nascimento I., Abdulazeem H., Vasanthan L., Martinez E., Zucoloto M., Østen-gaard L., Azzopardi-Muscat N., Zapata T., Novillo-Ortiz D. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. The Lancet Digital Health, Volume 5, Issue 8, 2023, P. e534—e544

[4] Тыров И.А., Васильев Ю.А., Арзамасов К.М. и др. Оценка зрелости технологий искусственного интеллекта для здравоохранения: методология и ее применение на материалах Московского Эксперимента по компьютерному зрению в лучевой диагностике // Врач и информационные технологии, 2022. № 4 — С. 76—92

[5] Требования к государственным информационным системам в сфере здравоохранения субъектов Российской Федерации, медицинским информационным системам медицинских организаций и информационным системам фармацевтических организаций (утверждены приказом Министерства здравоохранения Российской Федерации от 24 декабря 2018 г. № 911н)

[6] Гусев А.В., Зингерман Б.В., Тюфилин Д.С., Зинченко В.В. Электронные медицинские карты как источник данных реальной клинической практики. Реальная клиническая практика: данные и доказательства. 2022. 2(2). С. 8—20

[7] МКБ-10 Международная статистическая классификация болезней и проблем, связанных со здоровьем (10-й пересмотр)

[8] ПНСТ 872—2023 Системы поддержки принятия врачебных решений с применением искусственного интеллекта. Методы клинических испытаний

[9] Демонстрационный тестовый набор данных для алгоритмов системы поддержки принятия врачебных решений / ГБУЗ «НПКЦ ДиТ ДЗМ». — М., 2024

УДК 615.841:006.354 ОКС 11.040.01

Ключевые слова: системы поддержки принятия врачебных решений, методы испытаний, набор данных, неструктурированные медицинские записи

Редактор *Н.В. Таланова*
Технический редактор *И.Е. Черепкова*
Корректор *С.И. Фирсова*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 30.10.2025. Подписано в печать 18.11.2025. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 2,32. Уч.-изд. л. 1,97.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

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

gost.e-ecolog.ru