

Применение программы интеллектуальной аналитики текста с бумажного носителя и сегментации по заданным параметрам в клинической практике

Комков А. А.^{1,2}, Мазаев В. П.¹, Рязанова С. В.¹, Кобак А. А.³, Базаева Е. В.¹, Самочатов Д. Н.², Кошкина Е. В.², Бушуева Е. В.², Драпкина О. М.¹

¹ФГБУ "Национальный медицинский исследовательский центр терапии и профилактической медицины" Минздрава России, Москва; ²ГБУЗ "ГКБ № 67 им. Л. А. Ворохобова ДЗМ". Москва; ³ООО "Кобак Лаб". Санкт-Петербург, Россия

Развитие новых технологий с применением элементов искусственного интеллекта (ИИ) в медицине обращено к практическому клиническому внедрению и обеспечению ключевых вопросов, включая усовершенствование в использовании рутинных клинических данных, направленных на практическую значимость, стандартизацию, конфиденциальность и безопасность пациентов.

Цель. Оценить эффективность медицинской информационной системы (МИС) RuPatient в реальной клинической практике для извлечения и структурирования медицинских данных.

Материал и методы. Извлечение и распознавание данных с применением МИС из различных источников: амбулаторных карт, выписок, рутинных медицинских заключений, эпикризов и прочей структурированной и неструктурированной медицинской информации на основе разработанной технологии интеллектуальной аналитики текста, оптического распознавания знаков, по заданным словам и фразам, и применения элементов машинного обучения. Частным критерием оценки эффективности МИС использовано время, затраченное на заполнение электронных медицинских карт по сравнению с реальной клинической практикой.

Результаты. Время внесения и обработки информации системой распознавания медицинской документации, входящей в МИС RuPatient, было короче, чем при стандартной практике (20,3±1,4 vs 25,1±1,5 мин, $p < 0,001$), среднее время распознавания документов составило 30±4,3 сек. При ROC-анализе определено, что пороговое значение, которое позволяет с высокой точностью распознавать изображения выписных эпикризов с помощью системы RuPatient, составило 83,5% со значением площади под кривой 0,76.

Заключение. Разработанная МИС RuPatient имеет модуль распознавания медицинской документации для создания структурированных данных на основе элементов технологий ИИ может ис-

пользоваться как необходимый элемент при создании электронной истории болезни и накоплению структурированных данных для реализации задач по практическому и научному использованию больших данных и проектов ИИ в медицине. При использовании системы RuPatient может быть снижена нагрузка на медперсонал при выполнении документооборота и упрощен доступ к первичной медицинской информации.

Ключевые слова: искусственный интеллект, машинное обучение, глубокое обучение, обработка данных, автоматизация, здравоохранение, медицина, RuPatient, медицинская информационная система, оптическое распознавание текста, обработка естественного языка.

Отношения и деятельность. Источник финансирования — федеральный бюджет (гос. задание № АААА-А20-120013090084-6).

Поступила 28/10-2022

Рецензия получена 02/11-2022

Принята к публикации 09/11-2022



Для цитирования: Комков А. А., Мазаев В. П., Рязанова С. В., Кобак А. А., Базаева Е. В., Самочатов Д. Н., Кошкина Е. В., Бушуева Е. В., Драпкина О. М. Применение программы интеллектуальной аналитики текста с бумажного носителя и сегментации по заданным параметрам в клинической практике. *Кардиоваскулярная терапия и профилактика*. 2022;21(12):3458. doi:10.15829/1728-8800-2022-3458. EDN PAJVKC

*Автор, ответственный за переписку (Corresponding author):

e-mail: artemkomkov@gmail.com

[Комков А. А. — к.м.н., с.н.с. лаборатории рентгенэндоваскулярных методов диагностики и лечения, врач по РЭВДил, врач-кардиолог, ORCID: 0000-0001-7159-1790, Мазаев В. П. — д.м.н., профессор, руководитель лаборатории рентгенэндоваскулярных методов диагностики и лечения, ORCID: 0000-0002-9782-0296, Рязанова С. В. — к.м.н., с.н.с. лаборатории рентгенэндоваскулярных методов диагностики и лечения, врач-кардиолог, ORCID: 0000-0001-6776-0694, Кобак А. А. — главный разработчик, ORCID: 0000-0002-9010-4780, Базаева Е. В. — к.м.н., н.с. отдела нарушений сердечного ритма и проводимости, врач-кардиолог, ORCID: 0000-0002-5405-5459, Самочатов Д. Н. — к.м.н., зав. отделением рентгенохирургических методов диагностики и лечения, врач по РЭВДил, ORCID: 0000-0002-5230-2006, Кошкина Е. В. — к.м.н., зав. 3 отделением анестезиологии-реанимации, врач-кардиолог, ORCID: 0000-0001-9040-6757, Бушуева Е. В. — зав. отделением кардиологии № 3, врач-кардиолог, ORCID: 0000-0001-6628-0138, Драпкина О. М. — д.м.н., профессор, академик РАН, директор, ORCID: 0000-0002-4453-8430].

Application of the program for artificial intelligence analytics of paper text and segmentation by specified parameters in clinical practice

Komkov A. A.^{1,2}, Mazaev V. P.¹, Ryazanova S. V.¹, Kobak A. A.³, Bazaeva E. V.¹, Samochatov D. N.², Koshkina E. V.², Bushueva E. V.², Drapkina O. M.¹

¹National Medical Research Center for Therapy and Preventive Medicine. Moscow; ²L. A. Vorokhobov City Clinical Hospital № 67. Moscow;

³Kobak Lab. Saint-Petersburg, Russia

The development of novel technologies using elements of artificial intelligence (AI) in medicine is addressed to practical clinical implementation and provision of key issues, including improvement in the use of routine clinical data, aimed at practical relevance, standardization, confidentiality and patient safety.

Aim. To evaluate the effectiveness of the RuPatient electronic heart record (EHR) system in real clinical practice for extracting and structuring medical data.

Material and methods. Extraction and recognition of data using EHR from various following sources: outpatient records, statements, routine medical reports, epicrisis and other structured and unstructured medical information based on the developed technology of intelligent text analytics, optical character recognition, for specified words and phrases, and the use of machine learning elements. A particular criterion for evaluating the effectiveness of EHR is the time spent on filling out electronic medical records compared to real clinical practice.

Results. The time of entering and processing information by the recognition system of medical documentation included in the RuPatient EHR was shorter than in standard practice (20,3±1,4 minutes, 25,1±1,5 minutes, respectively, $p < 0,001$), the average time of recognition of documents was 30±4,3 seconds. During the ROC analysis, we determined that the threshold value that allows high accuracy to recognize images of discharge epicrisis using the RuPatient system was 83,5% with an area under the curve (AUC) value of 0,76.

Conclusions. The developed RuPatient EHR has a medical documentation recognition module for creating structured data based on AI technology elements and can be used in creating an electronic medical history and accumulation of structured data for the implementation of tasks for the practical and scientific use of big data and AI projects in medicine. When using the RuPatient system, the burden on medical

staff during document management can be reduced and access to primary medical information simplified.

Keywords: AI, machine learning, deep learning, data processing, automation, healthcare, medicine, RuPatient, electronic health record, optical character recognition, natural language processing.

Relationships and Activities. The source of funding is the federal budget (state task no. AAAA20-120013090084-6).

Komkov A. A.* ORCID: 0000-0001-7159-1790, Mazaev V. P. ORCID: 0000-0002-9782-0296, Ryazanova S. V. ORCID: 0000-0001-6776-0694, Kobak A. A. ORCID: 0000-0002-9010-4780, Bazaeva E. V. ORCID: 0000-0002-5405-5459, Samochatov D. N. ORCID: 0000-0002-5230-2006, Koshkina E. V. ORCID: 0000-0001-9040-6757, Bushueva E. V. ORCID: 0000-0001-6628-0138, Drapkina O. M. ORCID: 0000-0002-4453-8430.

*Corresponding author:
artemkomkov@gmail.com

Received: 28/10-2022

Revision Received: 02/11-2022

Accepted: 09/11-2022

For citation: Komkov A. A., Mazaev V. P., Ryazanova S. V., Kobak A. A., Bazaeva E. V., Samochatov D. N., Koshkina E. V., Bushueva E. V., Drapkina O. M. Application of the program for artificial intelligence analytics of paper text and segmentation by specified parameters in clinical practice. *Cardiovascular Therapy and Prevention*. 2022;21(12):3458. doi:10.15829/1728-8800-2022-3458. EDN PAJBKC

ЕМИАС — Единая медицинская информационно-аналитическая система, ИИ — искусственный интеллект, МИС — медицинская информационная система, ОРС — оптическое распознавание символов, ЭМК — электронная медицинская карта, ePAT — electronic patient-authored text data (электронные текстовые данные, созданные пациентом), AUC — площадь под кривой.

Ключевые моменты

Что известно о предмете исследования?

- Медицинская информационная система RuPatient включает в себя модуль распознавания печатной медицинской документации на основе оптического распознавания языка и машинного обучения.

Что добавляют результаты исследования?

- Время, затрачиваемое на внесение необходимой информации в электронную медицинскую карту, было достоверно меньше при использовании системы RuPatient, чем внесение стандартным способом.
- Качество изображения для успешного распознавания системой должно быть не <85% по данным субъективной оценки пользователя.

Key messages

What is already known about the subject?

- RuPatient electronic health record includes a module for recognizing printed medical documentation based on optical character recognition and machine learning.

What might this study add?

- The time spent on entering the necessary information into the electronic medical record was significantly less when using the RuPatient than in the standard way.
- The image quality for successful recognition by the system must be at least 85% according to the user's subjective assessment.

Введение

Развитие новых технологий с внедрением элементов искусственного интеллекта (ИИ) в медицине все больше обращено к практическому клиническому применению и обеспечению ключевых практических вопросов, включая усовершенствование в использовании рутинных клинических данных, для повышения практической значимости, стандартизации, конфиденциальности и безопасности пациентов [1].

В более широком плане улучшение лечения зависит от полноценной информации, необходимой для принятия решений и снижения вероятности медицинских ошибок [2].

Потенциал новых технологий далеко не реализован, и следует рассматривать внедрение простых доступных новых технологий на основе использования наблюдений и медицинских данных и их распознавание. При этом открытые базы данных будут иметь первостепенное значение для разработки и применения в моделях ИИ. Это позволит воспроизводить результаты, сравнивать точность различных методов и подходов и обосновывать научные выводы [3].

Цель — оценить эффективность медицинской информационной системы (МИС) RuPatient в реальной клинической практике для извлечения и структурирования медицинских данных.

Технологическое обеспечение

Веб-сервис RuPatient представляет собой МИС для медицинских учреждений с модулем аналитики на базе ИИ и продвинутым современным интерфейсом. Основной целью использования системы RuPatient является помощь медицинским специалистам в выполнении рутинных задач, в т.ч. с помощью стандартной процедуры анализа и внесения данных. Для реализации данной цели была создана МИС с модулем распознавания документации и функцией обмена данными непосредственно с пациентом при помощи встроенного чата.

Техническая часть разработки выполнена как набор программного обеспечения, состоящего из серверной и клиентской части. Доступ к сервису предоставляется из браузера, установленного на устройстве клиента. Серверная часть сервиса реализовывалась на языке программирования PHP версии 8.1. В качестве Backend (программно-аппаратная часть сервиса) фреймворка использовалась Symfony версии 6.1. На сервер устанавливалась операционная система Ubuntu версии 20.04. В качестве менеджера процессов использовалась PHP-fpm. В качестве веб-сервера использовалась NGINX. В качестве базы данных используется MySQL версии 5.8. Backend предоставляет REST API (программный интерфейс приложения), через которое с ним взаимодействовала клиентская часть сервиса. В качестве способа передачи данных использова-

лась HTTP запросы, данные передаются в формате JSON. Клиентская часть сервиса реализуется при помощи фреймворка ReactJS версии 18 и стандарта ECMAScript 9. Взаимодействие с сервером реализуется при помощи асинхронных запросов к API. В качестве менеджера пакетов используется npm актуальной версии. Для сборки клиентской части веб-сервиса используется WebPack версии 5.74.0.

Аналитический модуль печатной медицинской документации автоматизирует процесс работы врача с документами, распознавая текст и вставляя различные поля заключений в соответствующие поля веб-сервиса. Также добавлена возможность распознавания паспорта, полиса и страхового пенсионного свидетельства. Сервис распознавания работает в три этапа; на первом этапе происходит предобработка изображений при помощи алгоритмов, представленных в библиотеке OpenCV: кадрирование, увеличение кривых яркость-контрастность, подавление шумов. Далее изображение попадает в нейросеть детекции, для которой использовалась модель DBNet, которая определяет блоки текста на изображении. Для распознавания текста используется модель RobustScanner. Все модели представлены в открытом фреймворке MMOCR и обучались на специально подготовленных датасетах.

Разработанный прототип программного обеспечения веб-сервис RuPatient упрощает работу медперсонала в медицинских учреждениях за счет автоматизации и цифровизации работы с данными пациентов^{1,2}.

Материал и методы

Извлечение и распознавание данных с применением МИС проводилось из выписных эпикризов на основе разработанной технологии интеллектуальной аналитики текста, оптического распознавания знаков, по заданным словам и фразам, и применения элементов машинного обучения. В исследование включены две группы врачей стационара ГБУЗ ГКБ № 67 им. Л. А. Ворохобова ДЗМ: группа 1 обрабатывала выписные эпикризы и заполняла электронные медицинские карты (ЭМК) стандартным способом, группа 2 — использовала систему RuPatient (не менее 3 врачей в каждой группе). Врачи были обучены использованию системы распознавания RuPatient, после распознавания (группа 2), либо рутинным методом без использования системы распознавания (группа 1), данные заносятся в ручном режиме в систему ЕМИАС

¹ Комков А. А., Мазаев В. П., Рязанова С. В. и др. Государственная регистрация программы для ЭВМ "Медицинская информационная система RuPatient в формате веб-интерфейса" № 2021664147. Россия; 2021.

² Комков А. А., Мазаев В. П., Рязанова С. В. и др. Государственная регистрация программы для ЭВМ "Медицинская информационная система RuPatient. Система интеллектуальной аналитики текста с бумажного носителя и сегментации по заданным параметрам" № 2021664092. Россия; 2021.

Таблица 1
Исходные данные пациентов
и распознаваемые данные (n=72)

Характеристики	Значение
Возраст, лет	57,9±8,2
Мужской пол, n (%)	37 (51,4%)
Количество страниц, n	112
Время распознавания, сек	30±4,3
Показатель качества изображений при ОРС, %	87,9±8,2

(Единая медицинская информационно-аналитическая система). Использовались данные больных кардиологического профиля. Значительный эффект распознавания был получен при использовании модели оптического распознавания символов (ОРС) и машинного обучения, компьютерного зрения и нейросетей. Оценка качества распознаваемых изображений (наличие артефактов, достаточное разрешение, размытости и т.п.) осуществлялась специалистами после фотографирования/сканирования эпикризов, где за 100% принималось идеальное качество (отсутствие артефактов, высокое разрешение, отсутствие размытости и шума), 0% — плохое качество (много артефактов, шума, размытости, низкое разрешение).

Частным критерием оценки эффективности МИС использовано время, затраченное на заполнение ЭМК по сравнению с реальной клинической практикой [4]. После перераспределения первичного массива данных была проведена корректировка. Окончательная проверка верифицирующей части и предварительная оценка эффективности были проведены с учетом чувствительности и специфичности при сопоставлении машинных и наблюдательных данных по оценке МИС. Положительным результатом работы системы принималось распознавание системой >80% полей диагноз, анамнез, жалобы, состояние, рекомендации.

Примененная система, включающая элементы обработки естественного языка (NLP — Natural Language Processing), и интеллектуальный анализ текста использовались для захвата, извлечения и анализа данных о симптомах, заболеваниях, результатов клинического осмысливания и технологических данных.

Статистическая обработка выполнялась в SPSS 26 (IBM) и Excel (Microsoft). Переменные представлялись как среднее (M) ± стандартное отклонение (SD), для проверки гипотез о средних в двух группах применялся t-критерий Стьюдента. Для выявления предикторов успешного распознавания большинства полей использовался анализ характеристической кривой (ROC) с расчетом площади под кривой (AUC) и порогового значения с его специфичностью и чувствительностью.

Результаты

Первичные данные заносились как при непосредственном контакте с пациентом, так и удаленно через встроенный чат. Программа производила стандартизацию названий из данных, полученных от пациента, корректировку и автоматическое внесение в соответствующие поля. Результат считался позитивным, если система правильно определяет названия ≥80% 5 основных полей данных: диагноз, жалобы, анамнез, состояние, рекомендации.

В исследование были включены данные 72 пациентов, 37 (51,4%) из них были мужчины (таблица 1). Средний возраст пациентов составил 57,9±8,2 лет. Всего было распознано 112 страниц выписных эпикризов, среднее время распознавания каждой страницы при этом составило 30±4,3 сек. Среднее качество изображений при ОРС по данным субъективной оценки врачей составило 87,9±8,2%.

По основным характеристикам вносимых данных 1 и 2 группы не различались (таблица 2). Время, затрачиваемое на внесение необходимой информации в ЭМК, было достоверно меньше при использовании системы RuPatient, чем стандартным способом — 25,1±1,5 и 20,3±1,4 мин для 1 и 2 групп, соответственно (p<0,001). При заполнении в "ручном" режиме, без системы распознавания, врачи определяли все поля с 100%-ной точностью, при этом система давала положительный результат (≥80% полей распознавались правильно) в 87%.

Таблица 2
Исходные данные пациентов и их документов данные по группам (n=72)

Характеристики	Группа 1. Врачи, заполняющие ЭМК стандартным способом (n=41)	Группа 2. Врачи, заполняющие ЭМК с использованием системы RuPateint (n=31)	p
Возраст, лет	59,3±8,6	56,1±7,5	0,107
Мужской пол, n (%)	22 (53,7%)	15 (48,4%)	0,709
Количество страниц, n	63	49	0,714
Время внесения, мин	25,1±1,5	20,3±1,4	<0,001
Время распознавания, сек	29,5±4,8	30,6±3,6	0,280
Показатель качества изображений при ОРС, %	89,0±7,5	86,3±9,0	0,166
Использование чата, n (%)	6 (14,6%)	3 (9,7%)	0,529
Позитивный результат распознавания полей, n (%)	41 (100%)	27 (87,1%)	0,147

Примечание: ОРС — оптическое распознавание символов, ЭМК — электронная медицинская карта.

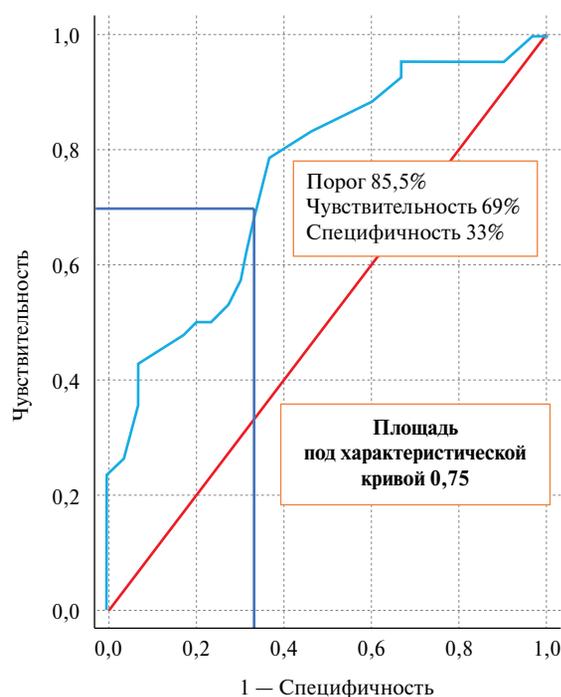
Пороговое значение качества изображений при ОРС, которое позволяет с наиболее высокой точностью распознавать изображения выписных эпикризов с помощью системы RuPatient, составило 85,5% с чувствительностью 69% и специфичностью 33% со значением AUC 0,75 ($p < 0,001$), что указывает на достаточно хорошее качество теста (рисунок 1).

По результатам контрольного исследования и оценкам результативности RuPatient, получена достаточная диагностическая точность с приемлемой частотой ложноположительных результатов, чувствительностью, специфичностью и AUC рабочих характеристик при опросе. Результаты показали, что эффективность анализа данных была выше у сотрудников с продолжительным опытом обработки первичного материала, чем среди лиц, вновь включенных в опрос.

Веб-сервис RuPatient внедряется в работу медицинского учреждения. После проведенного тестирования и устранения недостатков сервиса функционал был улучшен, расширен и применен в работе практикующих врачей ГКБ № 67 им. Л.А. Ворохобова ДЗМ.

Обсуждение

Необходимость в получении медицинских данных в структурированном формате мотивируется несколькими стимулами, важнейшими из которых могут быть сокращение времени для экспертного анализа, использование данных для масштабной автоматизированной обработки и уменьшение времени прочтения текстов свободного изложения [5]. Возможность анализировать большие объемы данных, полученных на протяжении продолжительного периода времени, для построения научных заключений и для лечения, профилактики известных и редких заболеваний, становится очевидной. Несмотря на технологический прогресс в разработке автоматизированных систем при создании и использовании клинической информации, применение системы анализа и генерирования структурированных данных востребовано³. Этим требованиям соответствует система обработки данных RuPatient. Этот проект направлен на повышение эффективности клинически значимых данных в структурированной и кодируемой неструктурированной информации для заполнения определенных форм и шаблонов, востребованных как элементы цифровой трансформации здравоохранения на основе анализа больших данных. Имеющиеся современные системы сократили время скрининга



Источник кривой

- Качество изображения при ОРС %
- Опорная линия

Рис. 1 Характеристическая кривая связи качества изображений с точностью распознавания МИС RuPatient.

Примечание: ОРС — оптическое распознавание символов.

для поиска кандидатов для участия в клинических испытаниях или в выявлении побочных действий лекарств [6].

Стремление к созданию единого контура данных в отечественном здравоохранении может быть реальным путем улучшения здоровья и благополучия населения. Применение алгоритмов машинного обучения с целью адекватного принятия клинических решений при использовании больших данных с новыми коммуникационными технологиями позволит повысить результативность федеральных проектов в области здравоохранения⁴.

Текущая версия системы RuPatient позволяет достичь диагностической эффективности, сравнимой с получаемой медицинскими экспертами, особенно в областях, связанных с распознаванием изображений. Дальнейшие исследования могут быть направлены на другие типы медицинской визуализации, такие как магнитно-резонансная томография и другие медицинские практики, не связанные с изображениями.

³ Аналитический отчет по сегменту рынка ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МЕДИЦИНЕ. <https://academpark.com/upload/medialibrary/362/36244984677a893f2c2d4a0080de0105.pdf>. (2022 Oct 11).

⁴ Федеральный проект "Создание единого цифрового контура в здравоохранении на основе единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ)". <https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravoohranenie/tsifra> (2022 Oct 11).

Применение интерфейсов обмена данных позволит перейти различным клиникам на единый стандарт ЭМК.

Исследования, основанные на клинически значимых потребностях, подкрепленные адекватным дизайном и ориентированные на цели клинической практики дали заметные результаты. Стремление к устойчивым образовательным программам через целенаправленное обучение, ориентированным на медицинских работников в сотрудничестве с заинтересованными сторонами из инженерных областей, позволит создать новые возможности на основе ИИ в медицине [7].

В будущем программы анализа данных могут быть направлены на обработку собственных записей пациентов с переводом их в структурированную форму. Это позволит учитывать потребности пациентов, выраженные с помощью электронных текстовых данных, созданных пациентом (electronic patient-authored text data, ePAT), и даст дополнительные возможности для понимания проблем больных. Новое видение информации о пациенте в системе ePAT может отчетливо проявиться в получении информации о состоянии здоровья. Оценка симптомов в режиме реального времени с помощью обработки естественного языка и интеллектуального анализа текста, может иметь существенное значение для системы здравоохранения, ориентированной на пациента [8].

При том что доступ к ЕМИАС связан с вычислительными трудностями, объем информации, предоставляемой непосредственно пациентами, открывает новые горизонты для точной медицины, описания субклинических симптомов и создания персональных библиотек здоровья, как это предусмотрено национальной программой модернизации здравоохранения⁵.

Анализ обработанных структурированных медицинских записей из социальных сетей — новая область исследований для развития ИИ. Обеспечение однородности данных для инфраструктур, используемых для сбора, хранения и анализа личных данных и клинической информации пациентов, относится к проблеме стандартизации при использовании ИИ в медицине [9].

⁵ Национальные проекты "Здравоохранение" и "Демография". <https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravooxranenie>. (2022 Oct 11).

Литература/References

1. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30. doi:10.1038/S41591-018-0307-0.
2. Ahsani-Estahbanati E, Gordeev VS, Doshmangir L. Interventions to reduce the incidence of medical error and its financial burden in health care systems: A systematic review of systematic reviews. *Front Med (Lausanne)*. 2022;9. doi:10.3389/FMED.2022.875426/FULL.
3. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94. doi:10.7861/FUTUREHOSP.6-2-94.
4. Komkov AA, Mazaev VP, Ryazanova SV, et al. First study of the RuPatient health information system with optical character recognition of medical records based on machine learning. *Cardiovascular Therapy and Prevention*. 2021;20(8):3080. (In Russ.) Комков А. А., Мазаев В. П., Рязанова С. В. и др. Первое исследование

на пути к стандартизации неструктурной медицинской информации

Доказано, что приложения ИИ расширяют наши возможности моделирования, диагностики, классификации и прогнозирования заболеваний в широком диапазоне клинических областей и различных сценариев [10]. Эти доказательства часто ограничиваются лабораторными и тестовыми сценариями. Данные из общедоступных репозиториях, клинических регистров, клинических испытаний и баз данных постоянно используются для разработки и проверки моделей ИИ, демонстрирующих отличные результаты в контексте соответствующих дизайнов исследований [11]. Однако существует огромная потребность в улучшении методологических отчетов и повышении надежности моделей [12]. Объяснимый ИИ — это растущая область исследований, которая будет отвечать потребностям в понимании клинических данных и данных о здоровье [13]. Сочетание моделирования ИИ и объяснимых стратегий будет иметь большую клиническую ценность при диагностике и лечении заболеваний, позволяя системам здравоохранения повышать качество всеобщего охвата услугами здравоохранения, реагировать на чрезвычайные ситуации⁶.

Заключение

Разработанная МИС RuPatient имеет модуль распознавания медицинской документации для создания структурированных данных на основе элементов технологий ИИ может использоваться как необходимый элемент при создании электронной истории болезни и накоплении структурированных данных для реализации задач по практическому и научному использованию больших данных и проектов ИИ в медицине. При использовании системы RuPatient может быть снижена нагрузка на медперсонал при выполнении документооборота и упрощен доступ к первичной медицинской информации.

Отношения и деятельность. Источник финансирования — федеральный бюджет (гос.задание № АААА-А20-120013090084-6).

⁶ Use of artificial intelligence on the rise, but its impact on health still limited, new study finds. <https://www.who.int/europe/news/item/27-09-2022-use-of-artificial-intelligence-on-the-rise--but-its-impact-on-health-still-limited--new-study-finds>. (2022 Oct 11).

- дование медицинской информационной системы RuPatient по автоматическому распознаванию медицинской документации на основе "машинного обучения". Кардиоваскулярная терапия и профилактика. 2021;20(8):3080. doi:10.15829/1728-8800-2021-3080.
5. Kalkman S, van Delden J, Banerjee A, et al. Patients' and public views and attitudes towards the sharing of health data for research: A narrative review of the empirical evidence. *J Med Ethics*. 2022;48(1):3-13. doi:10.1136/MEDETHICS-2019-105651.
 6. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun*. 2018;11:156. doi:10.1016/J.CONCTC.2018.08.001.
 7. Paranjape K, Schinkel M, Panday RN, et al. Introducing Artificial Intelligence Training in Medical Education. *JMIR Med Educ*. 2019;5(2). doi:10.2196/16048.
 8. Dreisbach C, Koleck TA, Bourne PE, et al. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int J Med Inform*. 2019;125:37-46. doi:10.1016/J.IJMEDINF.2019.02.008.
 9. Zinchenko VV, Khoruzhaya AN, Sharova DE, et al. Standardization in regulating artificial intelligence systems in Russian healthcare. *Kazan medical Journal*. 2021;102(6):923-33. (In Russ.) Зинченко В.В., Хоружая А.Н., Шарова Д.Е. и др. Стандартизация в области регулирования технологий искусственного интеллекта в российском здравоохранении. *Казанский медицинский журнал*. 2021;102(6):923-33. doi:10.17816/KMJ2021-923.
 10. Ahsan MM, Luna SA, Siddique Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare*. 2022;10(3):541. doi:10.3390/healthcare10030541.
 11. Weessler EH, Naumann T, Andersson T, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*. 2021;22. doi:10.1186/s13063-021-05489-x.
 12. Tourangeau R. Survey Reliability: Models, Methods, and Findings. *J Surv Stat Methodol*. 2021;9(5):961. doi:10.1093/JSSAM/SMAA021.
 13. Xu F, Uszkoreit H, Du Y, et al. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In: *Natural Language Processing and Chinese Computing*. ISBN 978-3-030-32235-9. 2019:563-574. doi:10.1007/978-3-030-32236-6_51.