ПРИМЕНЕНИЕ И ЭФФЕКТИВНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЭКСТРЕННОЙ МЕДИЦИНСКОЙ ПОМОЩИ: АНАЛИТИЧЕСКИЙ ОБЗОР

Ф.Т. АДЫЛОВА[1], Р.Р. ДАВРОНОВ[1], Х.М. КАСИМОВ[2]

[1]Институт математики им. В.И. Романовского Академии наук Республики Узбекистан, Ташкент, Узбекистан
[2]Республиканский научный центр экстренной медицинской помощи

APPLICATION AND EFFECTIVENESS OF ARTIFICIAL INTELLIGENCE IN EMERGENCY MEDICINE: AN ANALYTICAL REVIEW

F.T. ADYLOVA[1], R.R. DAVRONOV[1], KH.M. KASIMOV[2]

[1]V.I. Romanovsky Institute of Mathematics, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan
[2]Republican Research Center of Emergency Medicine, Tashkent, Uzbekistan

Цель данного обзора – показать тренд использования и эффективность моделей искусственного интеллекта (ИИ) в экстренной медицине. Результаты показывают, что медицинские системы визуализации на основе ИИ обеспечивают обнаружение заболеваний с точностью 85–90% в рентгенографии и компьютерной томографии. Системы сортировки (триажа) пациентов с поддержкой ИИ успешно классифицируют пациентов с низкой и высокой степенью срочности. Это демонстрирует потенциал обновленных моделей ИИ, но необходимы более масштабные исследования, поскольку остаются проблемы с интеграцией ИИ в клинические рабочие процессы и способностью моделей к обобщению. Обзор частично базируется на корректно выполненной работе за 2024–2025 гг. (https://DOI:10.4103/tjem.tjem_45_251 и статьи 2026 г. https://DOI:10.1126/science. adz4433).

Ключевые слова: *искусственный интеллект, неотложная медицина, обработка изображений, большие языковые модели, машинное обучение.*

The purpose of this review is to demonstrate the trend in the use and effectiveness of artificial intelligence (AI) models in emergency medicine. Results show that AI-based medical imaging systems provide 85–90% accuracy in X-ray and CT imaging. AI-enabled triage systems successfully classify patients with low and high urgency. This demonstrates the potential of updated AI models, but larger-scale studies are needed, as challenges remain with integrating AI into clinical workflows and the generalization ability of models. The review is based in part on work completed correctly in 2024–2025https://DOI:10.4103/tjem.tjem_45_251) and articles for 2026 https://DOI:10.1126/science. adz4433.

Keywords: *artificial intelligence, emergency medicine, image processing, large language models, machine learning.*

https://doi.org/10.54185/TBEM/vol19_iss2/a11

Введение

Искусственный интеллект – быстро развивающаяся, революционная технология в здравоохранении. Экстренная медицина, как быстро обновляющаяся область, открыта для применения искусственного интеллекта (ИИ). Исследования в области ИИ в последние годы растут в геометрической прогрессии во многих областях экстренной медицинской помощи. Использование ИИ расширяется в сортировке пациентов, диагностике, выборе тактики лечения, прогнозировании исходов. Эффективность применения моде-

лей ИИ, как правило, варьируется в зависимости от моделей и областей применения. Известно большое количество обзоров, посвященных конкретным областям использования ИИ в экстренной медицине, но большинство существующих исследований остаются ограниченными по масштабу и со временем неизбежно устаревают. Цель данного обзорного исследования – изучить текущие области применения ИИ в экстренной медицине и оценить их эффективность на основе двух публикаций: систематического обзора за 2024–2025 гг. [143] и статьи 2026 г. [127].

Создание обзора литературы проводилось в соответствии с рекомендациями PRISMA по проведению обзорных исследований. Сегодня традиционные модели ML (Machine Learning, ML) заменяются ансамблевыми методами и глубоким обучением (DL), а версии больших языковых моделей (Large Learning Model, LLM) быстро обновляются. Поэтому были проанализированы статьи, опубликованные в период с 1 января 2024 года по 1 января 2025 года с тем, чтобы представить актуальную подборку. Поиск проводился в базах данных PubMed и Web of Science (WoS) с использованием булевых операторов поиска. Запрос на поиск: «В каких областях (сортировка пациентов, диагностика, прогнозирование и т. д.) системы с поддержкой ИИ наиболее эффективны в сфере экстренной медицины?»

Результаты обзора

Проанализировали в общей сложности 1360 исследований по использованию ИИ в экстренной медицине. В области оказания экстренной помощи модели ИИ оценивались по следующим подкатегориям: обработка изображений (*n*=36), анализ текста (*n*=43), интеллектуальный анализ данных со структурированными большими данными (*n*=85).

Обработка изображений с использованием ИИ

Эффективность анализа изображений на основе ИИ зависит как от методов предварительной обработки изображений, так и от эффективности выбранного метода. Медицинские изображения могут различаться в зависимости от типа используемого оборудования, способа получения изображения и различий между пациентами [1–3]. Модели глубокого обучения достигают более высокой точности при работе с большими наборами данных, что побуждает исследователей интегрировать несколько наборов данных для повышения производительности. Однако объединение данных из разных источников может затруднить обучение моделей ИИ [4, 5], поэтому стандартизация наборов данных и минимизация вариаций изображений необходимы для точного изучения моделями ИИ определенных закономерностей [6]. На этапе предварительной обработки используют методы нормализации изображений, шумоподавления и аугментации данных [7]. Кроме того, точная маркировка и аннотирование являются фундаментальными факторами, определяющими успех обучения модели. Поскольку ручная разметка – трудоемкий процесс, полуавтоматическая разметка и аннотирование с использованием ИИ дают значительные преимущества на этом этапе [8, 9].

Одним из наиболее важных этапов обработки изображений является применение методов улучшения изображений [10]. Необработанные изображения часто имеют низкий контраст, шум или артефакты, что затрудняет прямой анализ. Для улучшения качества изображения обычно используют методы повышения контраста, обнаружения границ, фильтрации, пороговой обработки и сегментации [11–13]. После этого начинается процесс разработки и обучения модели. Производительность моделей ИИ зависит от выбора алгоритмов, качества обучающего набора данных и процесса обучения модели [14]. В медицинской обработке изображений широко используются как методы обучения с учителем (например, сверточные нейронные сети [CNN], ResNet, VGG), так и методы обучения без учителя (например, автокодировщики, GAN).

В процессе обучения модели ошибки минимизируются за счет вычислений функции потерь, параметры модели оптимизируются с помощью алгоритма обратного распространения ошибки, а эффективность процесса повышается за счет вычислений с ускорением на графическом процессоре [15, 16]. После процесса обучения модель должна быть проверена и протестирована для обеспечения ее надежности. Для этого обычно используются такие показатели производительности, как точность, чувствительность, специфичность, F1-мера, индекс структурного сходства для повышения четкости КТ/МРТ (SSIM), и AUC, методы перекрестной проверки [17].

На заключительном этапе модель проходит оптимизацию и развертывание. Для повышения эффективности в реальных приложениях применяются такие методы, как корректировка скорости обучения, методы регуляризации и сжатие модели (например, квантование, обрезка) [18].

Процедуры предварительной обработки изображений на основе ИИ

Предварительная обработка повышает точность и надежность алгоритмов анализа за счет уменьшения вариаций изображения и улучшения общего качества изображения [7, 17]. Рассматриваемые изображения: рентгенография, компьютерная томография (КТ) и магнитно-резонансная томография (МРТ).

В то время как базовые методы предварительной обработки применяются к отдельным кадрам рентгеновских снимков, они реализуются для каждого участка в КТ и МРТ, а также для каждого кадра в видео УЗИ. Однако из-за уникальных характеристик каждого метода визуализации были разработаны специализированные методы предварительной обработки для различных типов изображений, включая рентген, КТ, МРТ и УЗИ [13]. Рентгеновские снимки часто страдают от низкого контраста, низкого разрешения, шума и артефактов. Для повышения эффективности моделей ИИ применяются методы предварительной обработки, – шумоподавление, повышение контраста и обнаружение границ. К методам шумоподавления относятся размытие по Гауссу, медианная фильтрация и вейвлет-шумоподавление [19]. Адаптивное выравнивание гистограммы с ограничением контраста улучшает видимость костных структур [20], а методы нормализации способствуют повышению производительности и надежности модели [17].

На КТ- и МРТ-изображениях каждый срез должен обрабатываться индивидуально для обеспечения согласованности данных. Изменение размера обычно выполняется для стандартизации размеров изображения в разных наборах данных [15, 21]. Методы нормализации и 3D CNN повышают эффективность обработки данных и точность модели [22–24].

Видеозаписи УЗИ требуют специализированной предварительной обработки из-за высокого уровня шума и переменного контраста. Для идентификации ключевых кадров используются алгоритмы анализа движения и оптического потока [25]. Спекл-шум, распространенная проблема в УЗИ-изображениях, уменьшается с помощью таких методов, как фильтр Винера и анизотропная диффузия [26]. Качество изображения можно дополнительно улучшить с помощью методов сверхвысокого разрешения и выравнивания гистограммы [27]. Для анализа временных рядов видео УЗИ часто используются сети долговременной кратковременной памяти и подходы на основе 3D CNN [28].

ИИ-анализ рентгенограмм проведен в 9 научных исследованиях 2024–2025 гг., которые были сосредоточены на ИИ-анализе рентгеновских снимков грудной клетки и костных структур [29–37]. В частности, в исследовании Wang et al. была разработана модель на основе EfficientNetV2 с использованием 3498 рентгенограмм грудной клетки вместе с внешними наборами данных, достигшая AUC=0,878 при выявлении туберкулеза легких в тестовой выборке [31]. Наилучшие диагностические показатели были отмечены в исследовании Ghatak et al., [37], где модель Annalise Enterprise CXR использовалась для выявления компрессионных переломов позвонков на 596 рентгенограммах грудной клетки (272 положительных и 323 отрицательных случая). Эта модель ИИ продемонстрировала высокую эффективность в автоматической диагностике компрессионных переломов позвонков, достигнув AUC = 0,955. Наихудшие показатели имели место при валидации внешнего набора данных в работе Wang et al [31]. Снижение эффективности модели наблюдали при тестировании на наборах данных Montgomery (AUC:0,838) и Shenzhen (AUC:0,806), что подчеркивает ограничение использования данных одного центра с точки зрения масштабируемости [31].

В целом, анализ рентгеновских снимков на основе ИИ повышает точность диагностики, но искусственное расширение набора данных, проводимое в ограниченных клинических условиях, и проблемы, связанные с обобщаемостью модели, остаются ключевыми ограничениями [38].

Эффективность применения ИИ в широком спектре клинических состояний, представленных КТ, включая острый панкреатит, камни в мочеточнике, переломы черепа, внутричерепные гематомы, переломы шейного отдела позвоночника и расслоение аорты, представлена в 11 статьях [39–49]. Исследование с самым большим набором данных было проведено Ruitenbeek et al.; оно включало изображения КТ шейного отдела позвоночника от 2973 пациентов и оценивало влияние ИИ-алгоритма AIDoc Medical на обнаружение переломов шейного отдела позвоночника [48]. Рабочий процесс с использованием ИИ повысил эффективность диагностики, достигнув точности 94,8% и сократив среднее время диагностики переломов на 16 минут. Что касается производительности модели, то показатели AUC и точности варьировались от 0,788 до 0,993. Наибольшее значение AUC (0,993) было зафиксировано в исследовании Zhang et al., посвященном классификации и оценке тяжести острого панкреатита [44]. Модель, обученная на данных

190 пациентов, продемонстрировала высокую точность сегментации поджелудочной железы и успешно выявила такие осложнения, как перипанкреатический некроз и отек.

Наименее эффективной оказалась модель, разработанная Choi et al. для обнаружения внутримозгового кровоизлияния. Алгоритм DLHD, оцененный на 111 изображениях КТ головного мозга, достиг наименьшего значения AUROC= =0,788. Исследование показало, что хотя модель улучшила чувствительность, но снизила специфичность [45].

Таким образом, анализ изображений КТ на основе ИИ обеспечивает значительные успехи в ранней диагностике и быстром вмешательстве, но отсутствие крупномасштабной многоцентровой валидации и трудности адаптации моделей к различным протоколам визуализации остаются критически важными факторами для клинической интеграции.

Три научных исследования МРТ [50–52], которые, в основном, изучали диагностическую эффективность выявления острого ишемического инсульта (ОИИ), прогнозирования смертности и протоколов сверхбыстрой МРТ головного мозга. Размеры выборок варьировались, при этом самый большой набор данных принадлежал исследованию, в котором разработали модель на основе глубокого обучения для прогнозирования смертности у пациентов с ишемическим инсультом, с использованием данных 2710 человек [50]. Производительность, показатели AUC и точности варьировались от 0,852 до 0,95. Самое высокое значение AUC (0,95) для выявления ОИИ было зарегистрировано в исследовании Kim et al., которые разработали 3D-модель CNN [50]. Кроме того, Lang et al. оценили 2-минутный протокол сверхбыстрой МРТ головного мозга, разработанный для быстрой визуализации в экстренных ситуациях, продемонстрировав диагностическое соответствие на уровне 98,5% [51].

Таким образом, анализ МРТ с использованием ИИ повышает точность диагностики в экстренных ситуациях, ускоряет ведение пациентов и поддерживает принятие клинических решений. Однако проведение исследований в одном центре, демографический дисбаланс и ограничения в масштабируемости моделей остаются ключевыми факторами, которые необходимо учитывать при их более широком внедрении.

Четыре научные статьи представили результаты ультразвуковых исследований, в которых изучалась эффективность применения ИИ в оценке функции сердца, анализе сжимаемости сонной артерии, острых патологий желчного пузыря, обнаружении инородных тел и определении восстановления спонтанного кровообращения (ROSC) во время сердечно-легочной реанимации (СЛР) [53–56]. Производительность моделей оценена по AUC и точности, которые варьировались от 0,81 до 0,99. Самое высокое значение точности (99,1%) было отмечено в исследовании Holland et al., где использовались модели ИИ на основе U-Net и YOLOv7 для обнаружения инородных тел на изображениях УЗИ, содержащих 12 144 ультразвуковых изображения [56]. В исследовании подчеркивалось, что эти модели могут ускорить принятие решений, особенно в отдаленных районах или условиях с ограниченным доступом к экспертам. Однако процесс маркировки является ресурсоемким, а валидация in vivo остается ограниченной.

Самое низкое значение AUROC (0,81) было зафиксировано в исследовании He et al., где оценивали функцию сердца на эхокардиографии у постели больного (POCUS) [53]. Модель EchoNet-POCUS достигла AUROC 0,92 для оценки функции сердца, но только 0,81 – для качества видео. Park et al. представили ИИ-модель RealCAC-Net, предназначенную для определения восстановления спонтанного кровообращения при СЛР путем анализа сжимаемости сонной артерии [54]. Обученная на 11 958 изображениях и 15 080 для тестирования, модель продемонстрировала превосходные результаты по сравнению с традиционной ручной пальпацией, достигнув 96% точности и 97% показателя F1. Эта система потенциально может улучшить управление реанимацией в стационаре, поддерживая принятие решений при СЛР. Однако были отмечены опасения относительно ее применимости к различным устройствам и группам пациентов. Ge et al. исследовали ИИ в диагностике острых патологий желчного пузыря [55]. Модель глубокого обучения, обученная на 266 изображениях УЗИ от 186 пациентов, различала нормальные и патологические случаи с точностью 91% и классифицировала срочные и несрочные случаи с точностью 82%. Целью исследования была быстрая сортировка пациентов с патологиями желчного пузыря, что потенциально снижает зависимость от врачей-радиологов.

Таким образом, хотя анализ изображений УЗИ на основе ИИ способствует быстрой диагностике и ведению пациентов, для более широкой клинической интеграции необходимо решить проблемы отсутствия многоцентровой валидации, зависимость от устройства и чувствительность к качеству изображения.

Альтернативные методы анализа изображений на базе ИИ вне стандартных методов медицинской визуализации

Этому вопросу посвятили свои исследования авторы из [57–65]. В них изучалась эффективность применения ИИ в различных клинических условиях, включая обнаружение заболеваний сетчатки с помощью фотографий глазного дна, идентификацию переломов с помощью инфракрасных тепловых изображений, скрининг анемии с помощью фотографий конъюнктивы, обнаружение инсульта по изображениям лица и анализ функции сердца с использованием ЭКГ. Исследование с самым большим набором данных было проведено Song et al., сосредоточенном на автоматическом обнаружении патологий заднего сегмента с использованием 90250 изображений оптической когерентной томографии с роботизированным выравниванием [62]. Модель RobOCTNet показала высокую эффективность в качестве инструмента сортировки в отделениях экстренной офтальмологической помощи, достигнув AUC=1,00 при внутренней валидации и 0,91 при внешнем тестировании. Однако исследование выявило ограничения – обучение модели на относительно небольшом объеме данных и отсутствие ее интеграции в реальную клиническую практику. Показатели AUC и точности варьировались от 0,75 до 1,00. При этом самая низкая точность (75,4%) наблюдалась в исследовании Zhao et al., которое было посвящено выявлению анемии с использованием фотографий конъюнктивы [61]. Для выявления анемии путем анализа изображений конъюнктивы использовали приложение для смартфонов eMoglobin, достигшее AUC 0,92 при пороговом значении HBc 7 г/дл. Однако в исследовании отмечалась ограниченная чувствительность модели при выявлении случаев легкой анемии. Biousse et al. сообщили о показателе AUC 0,97 при обнаружении отека диска зрительного нерва с использованием модели BONSAI-DLS с немидриатическими фотографиями глазного дна [57].

Wang et al. разработали модель ИИ для диагностики острого ишемического инсульта по изображениям лица пациентов с инсультом. Эта модель, основанная на EfficientNet и ResNet50, достигла показателя AUC 0,91 при перекрестной валидации и 0,82 в независимых тестах [60]. Более того, Maxin et al. представили модель, направленную на различение ишемического и геморрагического инсульта с помощью комбинации пупиллометрии (современный бесконтактный метод исследования зрачков) и ML. Эта модель, показавшая точность 91,5%, потенциально может служить ценным инструментом поддержки принятия решений при оказании помощи при инсульте на догоспитальном этапе [64].

Хотя альтернативные методы визуализации и анализы с использованием ИИ обещают улучшить быструю диагностику и ведение пациентов, необходимость более широкой многоцентровой валидации, балансировка наборов данных и клиническая адаптация остаются ключевыми факторами для их широкого внедрения.

Анализ текста

Анализ текста используется для обработки неструктурированных медицинских записей, таких как выписные эпикризы, с целью выявления важных закономерностей [66]. Для прогнозирования и извлечения признаков требуется определенная предварительная обработка и анализ с использованием NLP (обработки естественного языка) – метода ИИ, который помогает компьютерам анализировать и понимать письменный текст [66]. В последние годы две концепции, LLM (большие языковые модели) и NLP (обработка естественного языка), быстро приобрели популярность, что привело к резкому увеличению числа публикаций и приложений.

Учитывая обилие вербальных и неструктурированных данных в экстренной медицине, эти концепции нашли широкое применение в этой области. В частности, при составлении отчетов всё чаще обращаются к моделям обобщения эпикризов, извлечения признаков из записей анамнеза, а также к прогнозному анализу. В обзор включены 43 исследования, проведенные по NLP (*n*=26) [67–92] и LLM [76, 91, 93–107] (*n*=17) в экстренной медицине.

Обработка естественного языка (NLP)

Большинство исследований, основанных на обработке естественного языка, предназначено для ретроспективного анализа неструктурированных текстовых данных, включая записи о сортировке пациентов, историю болезни и записи отделения экстренной помощи из электронных медицинских карт, для прогнозирования сортировки пациентов [71], диагноза [69, 74, 83], необходимости вмешательства [70, 80, 82] и исхода [67, 69, 71–73, 75, 77, 79, 81, 82, 83]. В основном используются методы глубокого обучения на основе трансформеров, двунаправленные представления кодировщиков на основе трансформеров (BERT) [67, 70–72, 76, 79, 81], частота терминов – обратная частота документов (Term Frequency-InverseDocumentFrequency, IDF)

[74, 75, 77, 78, 80], мешок слов [75, 78]. Модели ML представляют собой ансамблевые методы, включая алгоритмы бустинга, – категориального бустинга, машины легкого градиентного бустинга (Light Gradient Boosting Machine), экстремального градиентного бустинга (Extreme Gradient Boosting (XGB), логистической регрессии и глубоких нейронных сетей.

Использование структурированных данных вместе с неструктурированными данными в NLP оказывает значительное влияние на значения AUC. Так, исследование показало увеличение значений AUC при совместном использовании этих данных для прогнозирования исходов с учетом основной жалобы, жизненно важных показателей и демографических данных [78].

Наибольшее количество пациентов было в исследовании Patel et al., где использовалась программа BioClinical BERT для принятия решений о госпитализации на основе записей сортировки пациентов [75]. Исследования NLP применялись для диагностических прогнозов, таких как обнаружение обмороков (AUC = 0,95), прогнозирование фебрильных судорог (F1=0,921), прогнозирование серьезных инфекций (AUC=0,913) и прогнозирование COVID-19 (F1= 0,796). Оценивая эффективность прогностического анализа с точки зрения необходимости вмешательства, Chai et al. обнаружили наивысшее значение AUC 0,89 у 38 214 пациентов для прогнозирования показаний к хирургическому вмешательству, в то время как у Weidman et al. наивысший показатель эффективности с AUC 0,79 был достигнут при использовании градиентного бустинга гистограммы с TD-IDF для прогнозирования необходимости проведения срочных вмешательств, лабораторных исследований и визуализации у 12 913 пациентов только на догоспитальном этапе [80, 82]. Эти данные показывают, что методы NLP демонстрируют умеренную и высокую эффективность в прогнозировании диагноза и исхода. Большие различия в данных между исследованиями указывают на то, что методы и сравнение эффективности проведены на разных наборах данных и при структурированной интеграции данных.

Таким образом, методы NLP оказались эффективными во многих областях, от диагностики в отделении экстренной помощи до прогнозирования социально-демографических процессов. Однако для клинического применения требуется дальнейшая оптимизация и прозрачные процессы настройки, тестирование и оптимизация систем поддержки принятия клинических решений на основе NLP.

Большие языковые модели

Развитие больших языковых моделей (LLM) значительно ускорилось за последние 2 года. Эти модели стали способны выполнять различные задачи, основанные на языке, и приобрели такие навыки, как обучение с малым количеством примеров. LLM могут обобщать медицинские записи, предлагать возможные диагнозы и продемонстрировали высокую эффективность при медицинских обследованиях. Однако проблемы безопасности, риск генерации дезинформации и галлюцинации по-прежнему остаются актуальными [108].

GPT-1, одна из первых версий Open AI 2018 года, работала с ограниченным объемом обучающих данных, показав хорошие результаты во многих задачах обработки естественного языка [109]. GPT-2 была выпущена в 2019 году с 1,5 миллиарда параметров, что сделало ее гораздо более мощной и успешной в задачах обработки общего языка. Затем, в 2020 году, была выпущена GPT-3 со 175 миллиардами параметров, которая решала множество задач обработки естественного языка.

Наконец, GPT-4, с гораздо более мощными функциями, привлек внимание способностью принимать многомодальные входные данные. Разработка этих моделей стала возможной, в частности, благодаря сочетанию больших наборов данных и мощных вычислительных ресурсов. За короткое время показатели их точности выросли, однако использование этих технологий в критически важных областях, таких как медицина, по-прежнему сталкивается со многими проблемами в получении точных и надежных результатов [109].

В исследованиях использовались различные модели LLM, такие как GPT, Bard (модель Bard официально переименована в Gemini в феврале 2024 года), и изучалось, как эти модели работают в процессах поддержки принятия клинических решений. Большинство исследований LLM были сосредоточены на GPT-3.5 и GPT-4. Среди используемых методов наиболее заметным было применение моделей LLM в системах поддержки принятия решений врачами, сортировке пациентов и диагностике заболеваний. При изучении исследований обнаруживается больше проспективных наблюдательных и проспективных когортных исследований, которые способствуют решению различных клинических проблем, – сортировка пациентов [91, 92, 94–96, 101], диагностика [92, 95, 98, 110], выбор соответствующих тестов и прогнозирование исхода

(госпитализация [105] и смертность). Наиболее часто используемой LLM была GPT 4 (*n*=10), за ней следовали GPT 3.5 (*n*=6), BERT, Copilot (интеллектуальный ассистент от компании Microsoft; разработана на базе передовых больших языковых моделей (LLM) от OpenAI) и Llama2. Производительность LLM оказалась выше, чем в исследованиях NLP. При количестве 45 пациентов [94] для прогнозирования амбулаторной сортировки (AUC 0,87) и 864089 для прогнозирования госпитализации с помощью BERT с XGBoost заметно вырос AUC 0,87 [105]. Эти различия в размере данных указывают на то, что тестирование моделей на основе LLM с большими наборами данных может дать надежные результаты в клинической практике. LLM необходимо дополнительно тестировать и оптимизировать, прежде чем их можно будет полностью интегрировать в клиническую практику. В целом, LLM превосходят традиционные методы NLP, но доступность данных и размеры выборок сильно различаются. Например, исследования с использованием LLM на основе 864089 записей для прогнозирования госпитализации с помощью BioClinicalBERT и XGBoost показали AUC 0,87, в то время как другое крупное исследование с 484 094 пациентами, использовавшее NLP с GB, показало AUC 0,92 при поступлении в отделение интенсивной терапии [89, 105]. Эти различия указывают на то, что модели на основе LLM требуют больших наборов данных для надежной клинической интеграции. Точность меняется в зависимости от диагноза, что подчеркивает проблемы надежности в процессах скрининга инсульта с GPT-3.5 [100], которая является одной из более старых версий моделей GPT.

Одним из наиболее успешных методов диагностики оказалась многоязычная BERT от Levra et al., которая прогнозирует обморок по записям отделения экстренной помощи с извлечением симптомов и показателем F1 0,98 [69].

Обработка сигналов

Обработка сигналов и методы анализа с использованием ИИ играют всё более важную роль в процессах принятия решений в экстренной медицинской помощи. Исследования, рассмотренные в обзоре, сосредоточены на обработке физиологических сигналов, таких как ЭКГ и изображения головного мозга, с использованием алгоритмов ИИ для повышения точности клинической диагностики. При рассмотрении географического распределения рассмотренных исследований Южная Корея представлена 50,0% [70–75], Китай – 16,7% [76, 77], международные исследования – 8,3%, Тайвань – 8,3% [119=78], Европа (Франция и Испания, Германия) – 16,7% [79, 80]. При оценке с точки зрения особенностей вмешательства одно из рассмотренных исследований (8,3%) было посвящено выявлению инсульта [79]. Количество исследований, включающих анализ ЭКГ, составило 3 (25,0%), два из которых (16,7%) были непосредственно направлены на диагностику инфаркта миокарда (ИМ) [111–113]. Количество исследований по выявлению аритмии (*n*=1; 8,3%) [119], оптимизации процесса СЛР составило (*n*=2; 16,7%) [73, 74] прогнозированию поступления в отделение интенсивной терапии и системам раннего предупреждения (n = 2, 16,7%) [75] и выявлению SARS-CoV-2 (*n*=1, 8,3%) [121].

При анализе исследований с точки зрения методов ИИ сверточные нейронные сети (CNN) являются наиболее широко используемым методом в системах обработки сигналов и системах поддержки принятия медицинских решений. Четыре исследования (33,3%) использовали подходы на основе CNN. Кроме того, два исследования (16,7%) проводили анализ сигналов с использованием архитектуры трансформера. Методы глубокого обучения применялись в двух (16,7%) исследованиях. Передовые методы моделирования, такие как LightGBM, также использовались в 3 (25,0%) исследованиях.

Электрокардиография

Herman et al. разработали модель на основе глубокого обучения для оценки эффективности обнаружения инфаркта миокарда (ИМ) по данным 12-канальной ЭКГ у пациентов с подозрением на острый коронарный синдром. Исследование показало, что модель продемонстрировала в два раза более высокую чувствительность по сравнению с критериями STEMI (инфаркт миокарда с подъемом сегмента ST), но более низкую специфичность [81]. Было высказано предположение, что модель потенциально может улучшить результаты лечения пациентов, поддерживая раннюю диагностику и решения о реваскуляризации в отделениях экстренной помощи. Lee et al. разработали модель глубокого обучения, которая может извлекать цифровые биомаркеры STEMI из распечатанных результатов ЭКГ для улучшения догоспитальной телекардиологии. Было установлено, что модель достигла аналогичных уровней чувствительности и специфичности, что подтверждается консенсусом экспертов [71]. Lee et al. предложили модель анализа ЭКГ с поддержкой ИИ для определения этиологии одышки. Модель обеспечила более

высокую диагностическую точность, чем тест NT-proBNP [70].

Park et al. оценили систему количественной ЭКГ на основе ИИ для выявления острой коронарной окклюзии после остановки сердца вне стационара [72]. Диагностическая эффективность модели была сопоставлена с экспертной оценкой и показала не меньшую эффективность. Liu et al. разработали модель CNN, которая классифицирует аритмии с помощью одноканальной ЭКГ; модель достигла высокой точности при кратковременной записи ЭКГ.

Сердечно-легочная реанимация

Han et al. разработали неинвазивную модель прогнозирования артериального давления при СЛР. Модель достигла высоких коэффициентов корреляции при оценке систолического, диастолического и среднего артериального давления. Это исследование вносит значительный вклад в оценку процесса СЛР в режиме реального времени [73]. Kim et al. показали, что робот для СЛР с использованием ИИ обеспечил аналогичные гемодинамические результаты, как и LUCAS3 (третье поколение портативных автоматических систем для проведения непрямого массажа сердца). Исследование показало, что ИИ может создавать индивидуализированное управление СЛР [74].

Инсульт

Ou et al. создали мультимодальную модель глубокого обучения, которая объединяет видеоизображения и клинические данные для ранней диагностики пациентов с инсультом. Модель достигла более высокой точности по сравнению с отдельными модальностями [76]. Sen et al. разработали модель ML, которая анализирует гемодинамические волновые формы, полученные из сонных артерий, для выявления окклюзии крупных сосудов у пациентов с острым ишемическим инсультом [79]. Это исследование потенциально может способствовать разработке недорогого, быстрого догоспитального скринингового инструмента, который может быть интегрирован с такими устройствами, как портативный допплеровский УЗИ.

Системы оповещения

Zhang et al. разработали модель ML, использующую только неинвазивные параметры для прогнозирования необходимости инвазивной механической вентиляции (ИМВ) [77]. Модель достигла более высокого значения AUC (0,935), чем традиционные методы оценки риска. Эти результаты лежат в основе системы, которая может позволить прогнозировать потребность в ИМВ с помощью систем раннего предупреждения в догоспитальных условиях и в отделениях экстренной помощи. Choi et al. разработали модель на основе ML для повышения эффективности систем раннего предупреждения у пациентов в отделениях интенсивной терапии [75]. Было показано, что модель обладает более высокой чувствительностью, чем традиционные системы оценки, и может ускорить процессы экстренного вмешательства.

Диагностика и лечение

Методы ML применяются для улучшения и ускорения диагностических процессов в отделении экстренной помощи, улучшения лечения заболеваний. Были проанализированы 17 статей, посвященных диагностике и лечению. Выявили, что в исследованиях оценивались модели ML для прогнозирования различных диагнозов в отделении экстренной помощи, а также сепсиса, распознавания ритма и сложных диагнозов [82–98].

В исследовании сепсиса и инфекций общий размер выборки исследований был достаточным. Наибольший размер выборки был в исследовании Song et al., в котором оценивалась эффективность моделей ML в диагностике сепсиса [92] на крупномасштабном наборе данных. Их модели ML продемонстрировали значения AUC от 0,68 до 0,93, при этом XGBoost превзошел другие модели. Aygun et al. представили интерпретируемую модель риска сепсиса на основе XGBoost, включающую значения Shapley для объяснения признаков [89]. Помимо прогнозирования сепсиса, Chiu et al. разработали наиболее успешную модель для прогнозирования бактериемии на основе лабораторных результатов, сочетающую ансамблевое обучение, что привело к самому высокому зарегистрированному значению AUC (0,93) в этом процессе диагностики [98]. Flores et al. применили случайный лес и нейронные сети для диагностики инфекций мочевыводящих путей, показав, что системы поддержки принятия клинических решений, улучшенные с помощью ML, повысили точность диагностики по сравнению с традиционными методами (AUC 0,81–0,88) [83].

Toprak et al. разработали модель ARTEMIS-POC AI, которая использует данные о высокочувствительном сердечном тропонине I для исключения инфаркта миокарда, достигнув высокой отрицательной прогностической ценности (99,96%) и чувствительности (99,68%) [91]. Holmstrom et al. внедрили модели XGB для дифференциации бессосудистой электрической активности от фибрилляции желудочков, что помогло в

диагностике внезапной остановки сердца (AUC 0,68–0,72) [88]. Chang et al. использовали методы балансировки классов в обучающей выборке (SMOTE) и модели ML (RF, SVM, KNN, LR) для прогнозирования риска острого инфаркта миокарда у пациентов с болью в груди, повысив диагностическую чувствительность (AUC 0,63–0,82) [86]. Помимо острого инфаркта миокарда, Yilmaz et al. использовали объяснимые модели ИИ (XGBoost, LASSO, SHAP) для оценки гематологических показателей при диагностике острой сердечной недостаточности, достигнув высокой интерпретируемости и точности [87]. В рассмотренных исследованиях сообщалось о значениях AUC в диапазоне от 0,68 до 0,93, при этом модели XGBoost и случайного леса часто превосходили традиционные статистические модели.

Однако стоит отметить значительную вариативность в производительности моделей в зависимости от характеристик набора данных, методов отбора признаков и методов валидации.

Прогнозирование исходов и рисков

Помимо прогнозирования в догоспитальных процессах, сортировке и диагностике, еще одна важная роль ИИ в экстренной медицине – ведение пациентов и обеспечение их выживания. Модели прогнозирования могут помочь врачам экстренной помощи в принятии клинических решений, улучшении результатов лечения пациентов и оптимизации использования ресурсов. Однако их эффективность зависит от тщательной разработки модели, ее валидации и учета методологических проблем для обеспечения точных и клинически полезных прогнозов. Эффекты лечения могут повлиять на способность выявлять пациентов высокого риска и управлять лечением [99].

В обзоре были рассмотрены 30 статей, посвященных прогнозированию исходов и рисков. Большинство исследований оценивали эффективность моделей ML в прогнозировании смертности и стратификации риска, достигая значений AUC от 0,75 до 0,97 в отношении госпитализации и поступления в отделение интенсивной терапии, а также долгосрочного прогнозирования риска и раннего клинического ухудшения [93, 100–126]. Несколько исследований – Rahmatinejad et al. and Jawad et al. – продемонстрировали превосходство моделей ансамблевого обучения над традиционной логистической регрессией в прогнозировании смертности, достигнув значений AUROC выше 0,83 [100, 122]. Ding et al. and Shashikumar et al. успешно внедрили модели XGBoost и DL для обнаружения интубации и физиологического ухудшения, показав высокую чувствительность и специфичность [103,107]. Кроме того, Richards et al. разработали индекс риска коагуляции на основе ML, превосходящий традиционные оценки на основе INR с AUROC 0,97 [102].

Несмотря на эти достижения, остаются проблемы дисбаланса данных, как это наблюдалось в работе Park et al., которая потребовала внешней валидации из-за изменчивости набора данных [110]. Аналогично, Hinson et al. подчеркнули необходимость проспективной валидации, поскольку большинство моделей обучались на ретроспективных наборах данных, что ограничивает их применение в реальных условиях [117]. Gauss et al. дополнительно подчеркнули проблемы интерпретируемости, отметив, что хотя объяснения признаков на основе SHAP улучшили прозрачность модели ML, но в глубоком обучении при прогнозировании кровотечений у пациентов с травмами [120] это ещё не решено.

В этих исследованиях значения AUC варьировались от 0,75 до 0,97, при этом модели ансамблевого обучения (XGBoost, Random Forest AdaBoost) и методы глубокого обучения превосходили традиционные модели на основе логистической регрессии. Однако наблюдались проблемы дисбаланса наборов данных, требующие расширения данных (например, SMOTE) [115, 123] и многоцентровой валидации для повышения надежности, в то время как некоторые модели использовали методы объяснимого ИИ (SHAP, LIME) [93, 103, 110, 117, 120]. Модели глубокого обучения остаются системами типа «черный ящик», что создает препятствия для их внедрения врачами, но, в целом, разработанные модели ML достигли более успешных результатов, чем классические методы.

Обсуждение

В данном обзорном исследовании были рассмотрены работы, проведенные за последний год, посвященные оказанию помощи пациентам в отделениях экстренной помощи и образованию в области экстренной медицины. Рассматривается использование таких областей ИИ, как обработка изображений, обработка естественного языка и анализ текста в различных областях экстренной медицины, включая сортировку пациентов, диагностику, оценку результатов лечения, анализ рисков.

Результаты показывают, что исследования по применению ИИ в экстренной медицине демонстрируют многообещающий потенциал. Однако каждый метод имеет свои уникальные характе-

ристики, специфические области применения и присущие ему ограничения.

ИИ обладает потенциалом для улучшения процессов медицинской визуализации в экстренной медицине. Модели ИИ могут автоматизировать рутинные задачи, способствовать раннему выявлению заболеваний и ускорять принятие решений, помогая рентгенологам и врачам без формальной подготовки в области радиологии. Инструменты визуализации с поддержкой ИИ значительно сокращают время интерпретации и повышают эффективность принятия решений в отделениях экстренной помощи. Системы автоматизированного обнаружения (CADe) и диагностики (CADx) автоматически выделяют такие патологии, как переломы, заболевания легких и неврологические расстройства, тем самым экономя ценное время врачей.

Кроме того, системы визуализации на основе ИИ оказывают существенную поддержку в регионах с нехваткой опытных рентгенологов. Исследования показали, что рентгенография с использованием ИИ повышает чувствительность и специфичность при выявлении таких состояний, как переломы, узлы в легких и ишемические инсульты. Эти системы повышают эффективность медицинских услуг за счет повышения точности диагностики, особенно в условиях ограниченных ресурсов. Модели сегментации и классификации на основе ИИ оптимизируют диагностический процесс, минимизируя человеческие ошибки при интерпретации изображений. Например, применение ИИ в ультразвуковой диагностике позволяет быстро оценивать функцию сердца, помогая в лечении тяжелобольных пациентов. Благодаря всё большей интеграции автоматизации, врачи могут принимать более быстрые и точные решения, оптимизируя пути оказания медицинской помощи пациентам.

ИИ интегрирует медицинские изображения с данными пациентов, предоставляя всестороннюю диагностическую информацию. Системы ИИ, работающие совместно с электронными медицинскими картами (ЭМК), могут выявлять такие состояния, как острая сердечная недостаточность и сепсис, на ранней стадии, что позволяет разрабатывать персонализированные планы лечения. Эти мультимодальные подходы ИИ играют решающую роль в будущем медицины, предлагая более целостную оценку состояния здоровья пациентов. ИИ обладает огромным потенциалом для обработки медицинских изображений, но в этой области остается ряд существенных проблем. Медицинские изображения различаются из-за таких факторов, как низкое разрешение, артефакты и различия в устройствах визуализации. Хотя большие высококачественные наборы данных необходимы для достижения высокой точности моделями ИИ, отсутствие стандартизации данных из разных учреждений представляет собой серьезное препятствие.

Более того, модели ИИ, обученные на определенных наборах данных, могут не показывать ожидаемых результатов при применении к различным группам пациентов и методам визуализации. Различия в устройствах визуализации и демографических характеристиках пациентов могут влиять на точность и надежность модели. Проблемы, связанные с устойчивостью и обобщаемостью модели, остаются ключевыми препятствиями на пути широкого внедрения ИИ в клиническую практику.

Еще одна критическая проблема – это интерпретируемость систем на основе глубокого обучения, которые часто воспринимаются как «черные ящики». Непрозрачность процессов принятия решений в ИИ затрудняет для врачей полное понимание и доверие к этим системам. Повышение интерпретируемости имеет важное значение для повышения уверенности врачей и облегчения интеграции ИИ в рутинную медицинскую практику.

Интеграция ИИ в существующие клинические рабочие процессы также создает логистические проблемы. Инструменты ИИ, которые не предназначены для беспрепятственного взаимодействия с информационными системами больниц, часто требуют дополнительной инфраструктуры и значительных вычислительных ресурсов, что ограничивает их использование, особенно в небольших медицинских учреждениях.

Это один из факторов, замедляющих широкое внедрение технологий искусственного интеллекта в медицину. Кроме того, внедрение медицинских инструментов визуализации на основе ИИ вызывает этические опасения, связанные с конфиденциальностью пациентов, безопасностью данных и предвзятостью алгоритмов. Регулирующие органы, такие как FDA, требуют тщательной проверки, прежде чем одобрить диагностические инструменты на основе ИИ для клинического использования. Хотя эти регулирующие меры повышают надежность, они также замедляют переход инноваций в области ИИ от исследований к клинической практике.

Обзор за 1 год особенно отражает быстрый прогресс и конкуренцию в области обработки естественного языка (NLP) и языкового моделирования (LLM). Хотя наиболее важной проблемой NLP является сложная структура самого

языка, эта проблема в значительной степени решена благодаря развитию концепции онтологии в медицинских данных. Однако с появлением языковых моделей, доступных для распространения в последние годы, эта проблема также обеспечила извлечение признаков и рассуждения с помощью более совершенных и быстрых моделей.

Хотя исследования в области LLM, особенно в экстренной медицине, в основном проводятся на примерах сценариев, определенных экспертами для определения точности LLM, сегодня также растет число исследований с использованием реальных данных пациентов. Эта ситуация привела к необходимости корректного ввода данных в электронные медицинские карты.

Хотя минимизация необходимости структурирования данных кажется выгодной, тот факт, что правила обучения больших баз данных LLM могут зависеть от внешних факторов, обуславливает необходимость включения специализированных инструментов для обработки медицинских данных. В целом, исследования на основе LLM обладают значительным потенциалом в условиях отделения экстренной помощи. Было установлено, что такие модели, как GPT-4 и BioClinicalBERT, демонстрируют более высокую производительность, чем исследования в области обработки естественного языка (NLP).

Большинство исследований проводилось с использованием ретроспективного анализа данных, однако разработка систем, смоделированных с использованием потоков данных в реальном времени, остаётся важной для повышения надежности клинических приложений.

Для понимания того, как модели ML могут быть более эффективно использованы в системах поддержки принятия медицинских решений, необходимы более перспективные и масштабные исследования. Методы ML используются для преодоления существующих стандартных систем поддержки принятия клинических решений и разработки новых моделей прогнозирования. Эти модели прогнозирования потенциально могут помочь врачам скорой помощи в принятии решений. Учитывая широту и разнообразие области экстренной медицины, использование моделей ML в практике экстренной медицины представляет собой возможность, которую нельзя игнорировать. Скрытые закономерности, которые будут способствовать оказанию помощи пациентам в экстренной медицине, в тщательно собранных наборах данных могут быть выявлены с помощью моделей ML и использованы в лечении пациентов. Помимо упомянутых проблем обработки изображений и текста, в результате изучения моделей ML со структурированными данными было установлено, что большинство исследований было направлено на прогнозирование различных исходов и диагнозов. Однако не следует забывать, что все эти модели прогнозирования были созданы на основе данных, полученных из существующих наборов данных. В конечном итоге, производительность модели ML также зависит от набора данных. Таким образом, точность структурированных данных, полнота и безошибочность записи, а также тщательная предварительная обработка являются основными факторами успеха или неудачи моделей.

Хотя системы сортировки пациентов на основе ИИ демонстрируют высокую прогностическую способность, остаются опасения относительно предвзятости моделей и проблем интеграции. Многие модели не обладают достаточной адаптивностью к условиям реального времени, что ограничивает их применение в качестве систем сортировки в отделениях экстренной помощи. Будущие исследования должны сосредоточиться на внешней валидации в различных популяциях и интерпретируемых моделях ИИ для повышения принятия их врачами, поскольку каждая популяция уникальна, а полученные результаты действительны только для этой популяции. Валидность для разных популяций должна быть подтверждена исследованиями внешней валидации.

Распространенность мелкомасштабных, нерандомизированных исследований, проводимых в отдельных учреждениях, ограничивает возможность делать обоснованные выводы, что затрудняет определение реальной роли ИИ за пределами первоначального тестирования осуществимости. Как и многие новые технологии, ИИ часто представляется как революционное решение множества проблем, в том числе и в медицинском образовании. Хотя инструменты на основе ИИ, включая LLM и другие подходы ML, потенциально могут улучшить определенные аспекты, их внедрение должно основываться на убедительных доказательствах и реальных потребностях, а не на стремлении по умолчанию внедрять ИИ во все возможные области.

Заключение

Для того чтобы понять эффективность внедрения в клинику методов ИИ, логично было иметь сравнительную оценку моделей и врачей по существующим в этом отношении правилам. Потому представим новые результаты исследо-

вания, которое сегодня широко обсуждается в научных кругах. В исследовании, проведенном группой врачей-ученых из 20 известных научных центров США, включая Harvard, Stanford, Cambridge, Massachusetts [127], показали результаты реального исследования, сравнивающего экспертные заключения человека и вторые мнения искусственного интеллекта (ИИ) у случайно выбранных пациентов в отделении экстренной помощи крупного академического медицинского центра третьего уровня. Во всех экспериментах LLM превзошла базовые показатели врачей и продемонстрировала постоянное улучшение по сравнению с предыдущими поколениями систем поддержки принятия клинических решений на основе ИИ.

Недавние исследования LLM в медицине сосредоточены на узких диагностических задачах или на тщательно отобранных и ограниченных клинических примерах [128, 129, 130]. Что еще более важно, в большинстве исследований LLM для диагностики и лечения до настоящего времени отсутствовали базовые показатели, полученные от врачей-исследователей. Учитывая быстрое улучшение LLM и растущее «насыщение эталонными тестами», необходимо установить базовые показатели, полученные от людей, и изучить задачи, имеющие клиническую основу. В исследовании всесторонне оценили возможности диагностики и лечения усовершенствованной LLM (серия OpenAI o1) в нескольких задачах, используя базовые показатели, полученные от сотен врачей. Также изучили второе мнение LLM вслепую, сравнивая) его с базовыми показателями врачей-экспертов на случайно выбранных пациентах в крупном академическом отделении экстренной помощи в Бостоне [131].

Сначала оценили o1-preview, используя клинико-патологические конференции (КПК), опубликованные New England Journal of Medicine (NEJM), как стандарт для оценки генераторов дифференциальных диагнозов с 1950-х годов (1). Клинико-патологические конференции (КПК), официально известные как «Истории болезни Массачусетской больницы общего профиля» (МГБ), – всемирно известная образовательная серия, публикуемая еженедельно в журнале New England Journal of Medicine (NEJM). В этих случаях представлены сложные, реальные клинические ситуации из МГБ, служащие эталоном для клинического мышления и решения диагностических задач.

Между двумя врачами, оценивающими качество дифференциального диагноза o1-preview, наблюдалось существенное совпадение [совпадение по 120/143 случаям (84%), межэкспертная надежность по коэффициенту каппа Коэна (к), который измеряет согласие между двумя экспертами в категориальных (номинальных) шкалах; к корректирует процент согласия с учетом случайности. (к)=0,66]. o1-preview включал правильный диагноз в свой дифференциальный диагноз в 78,3% случаев [95% – доверительный интервал (ДИ), 70,7–84,8%] (рис. 1).

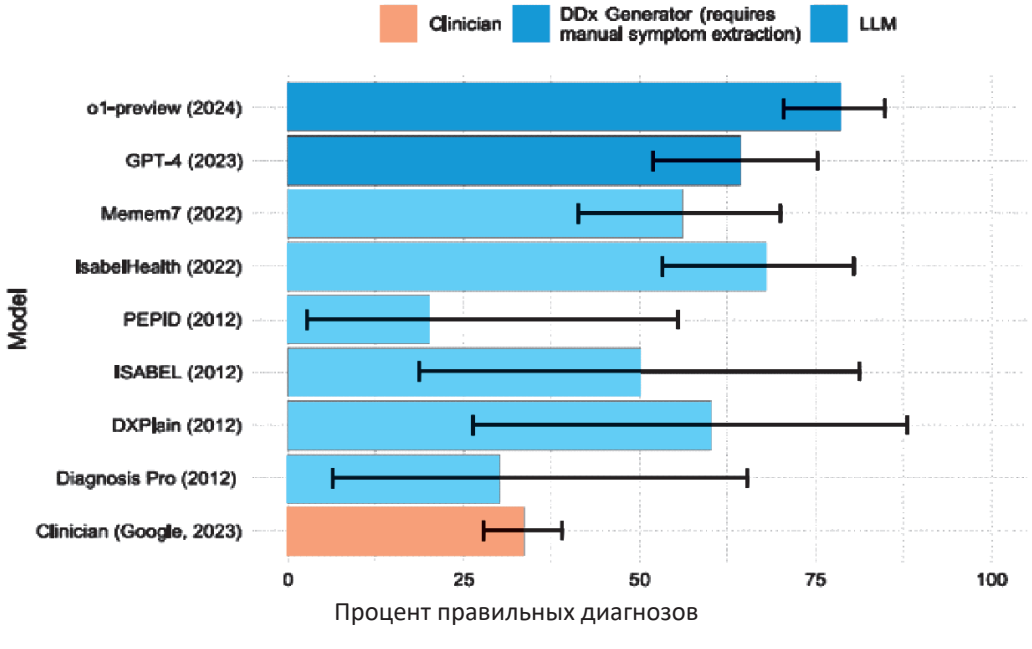

Рис. 1. Эффективность генераторов дифференциальной диагностики и линейных моделей диагностики на клинико-патологических конференциях NEJM (КПК), 2012–2024 гг.

Систематически оценивали способности модели LLM к медицинскому мышлению в шести различных экспериментах, сравнивая модель с сотнями врачей-экспертов. В целом, модель превзошла врачей во всех экспериментах, в том числе в случаях использования реальных и неструктурированных клинических данных, полученных непосредственно из медицинской карты в отделении экстренной помощи. Эти диагностические точки отражают важные решения, принимаемые в отделениях экстренной медицинской помощи, где медсестры и врачи принимают решения в условиях ограниченного времени и с ограниченной информацией. Результаты показали, что люди, GPT-4o и o1 улучшали свои диагностические способности по мере увеличения объема доступной информации; o1 превзошла людей в нескольких точках взаимодействия, с наибольшим разрывом на этапе первичной сортировки в отделении экстренной помощи, где доступно наименьшее количество информации.

Между двумя врачами наблюдалось существенное совпадение оценок по пересмотренной шкале IDEA (R-IDEA) – валидированной 10-балльной шкале для оценки четырех основных областей документирования клинического мышления [132] [совпадение по 79/80 (99%) случаям, к=0,89]. В 78 из 80 случаев o1-preview достиг идеального показателя R-IDEA, значительно превзойдя GPT-4 (47 из 80, P<0,0001), лечащих врачей (28 из 80, P<0,0001) и врачей-резидентов (16 из 72, P<0,0001), как показано на рис. 2A. Доля диагнозов, которые нельзя пропустить, выявленных o1-preview во время первичного осмотра, показана на рисунке 2B.

Далее были взяты пять клинических примеров, основанных на реальных случаях, разработанных и оцененных с использованием методов консенсуса 25 врачами-экспертами [133]. Каждый клинический пример был представлен модели, после чего следовал ряд вопросов относительно дальнейших шагов в лечении. Два врача оценили ответы с помощью o1-preview для пяти случаев, показав существенное совпадение (к=0,71). Медианный балл для o1-preview на случай составил 89% (межквартильный размах от 79 до 91%) (рис. 3A), что выгодно отличается от GPT-4.

Использовали шесть клинических примеров, где сравнивали GPT-4 с 50 врачами общей практики [134]. Эти примеры содержат историю заболевания, анамнез, физический осмотр и диагностические исследования [135]. Примеры никогда не публиковались, специально для защиты достоверности оценки от запоминания. Два врача оценили ответы с помощью o1-preview на шесть примеров диагностического мышления, показав умеренное совпадение по общему баллу (к=0,42). Медианный балл для модели o1-preview на один пример составил 97% (межквартильный размах от 95 до 100%) (рис. 3B).

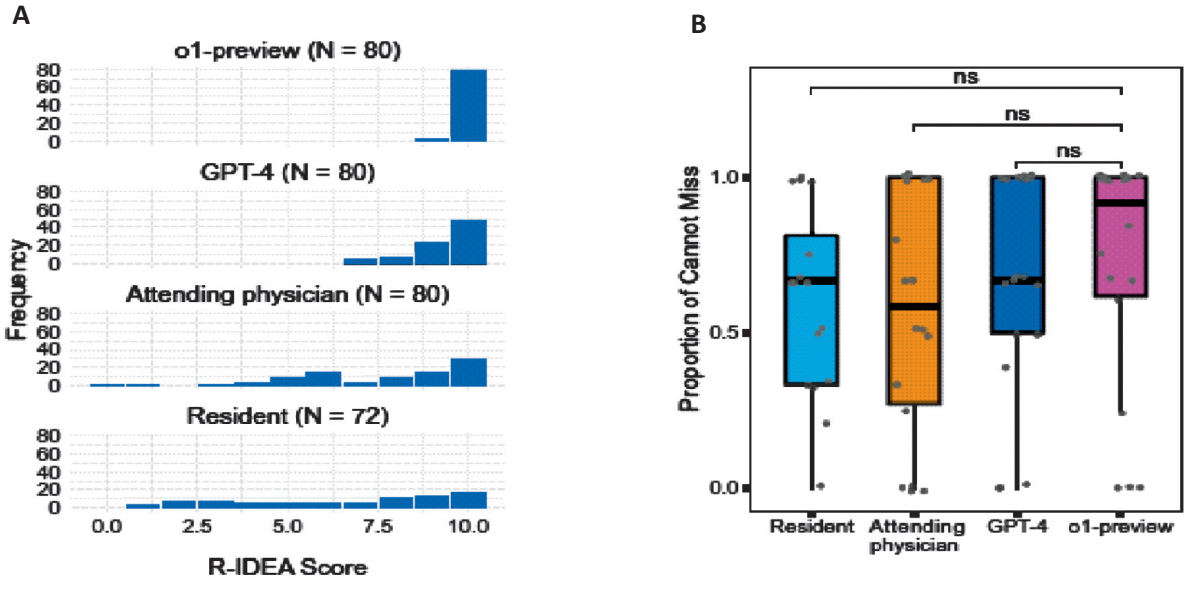

Рис. 2. Сравнение o1-preview, GPT-4, и врачей по клинической диагностической обоснованности.
(A) – распределение 312 баллов R-IDEA, рассортированных по респондентам на основе 20 случаев NEJM Healer;
(B) – диаграмма размаха доли диагнозов, которые нельзя пропустить, включенных в дифференциальный диагноз при первичном осмотре. Общий размер выборки на этом рисунке составляет 70, из которых 18 – ответы от лечащих врачей, GPT-4, и o1-preview, и 16 – ответы от ординаторов

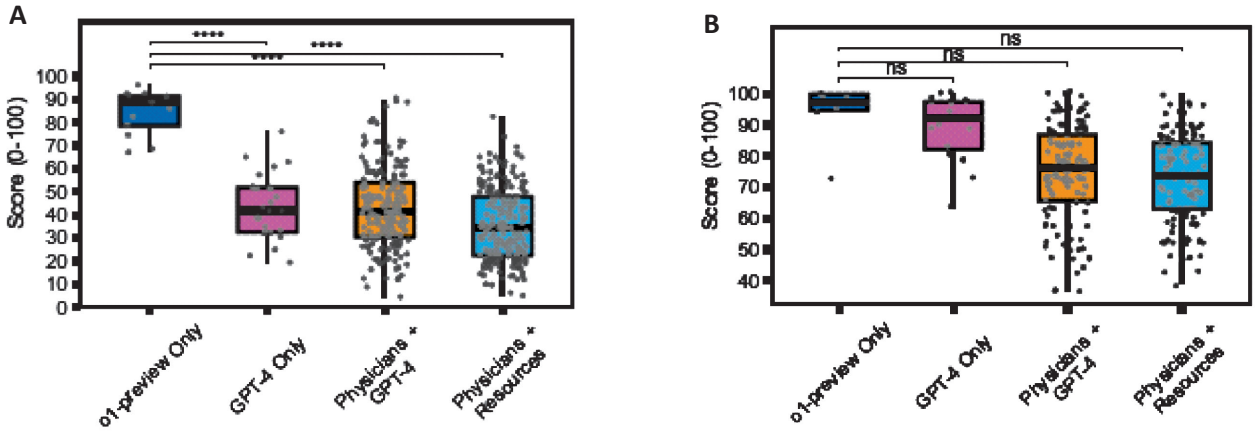

Рис. 3. Сравнение o1-preview, GPT-4, и врачей для управления заболеваниями и диагностического мышления. (A) – диаграмма размаха нормализованных баллов управленческого мышления магистров и врачей по кейсам «Управление серыми веществами». Включено пять кейсов. Сгенерировали три ответа на o1-preview для каждого кейса. (B) – диаграмма размаха нормализованных баллов диагностического мышления магистров и врачей. Было включено шесть диагностических задач. Для каждого случая сгенерировали один предварительный ответ

Случаи в отделении экстренной помощи

Сравнили способность врачей o1, 4o и двух лечащих врачей ставить дифференциальные диагнозы по 76 случаям из Медицинского центра Бет Израэль Диконесс, разделенным на три клинически значимых диагностических этапа [первичная сортировка в отделении экстренной помощи (ОНП), врач ОНП и поступление в терапевтическое отделение или отделение интенсивной терапии (ОИТ)]. В целом, o1 превзошел как 4o, так и двух опытных лечащих врачей, по оценке двух других лечащих врачей, которые не знали источника дифференциального диагноза (человек или модель ИИ) (рис. 4); однако обе модели продемонстрировали значительную неопределенность. Врачи-эксперты показали умеренное согласие в оценках качества [согласие по 496/911 (54%), к = 0,51]. Точность врачей в определении, является ли пациент ИИ или человеком, составила 15,2% для одного врача (83,6% «Не могу определить») и 3,1% для другого (94,4% «Не могу определить»). На каждом этапе диагностики o1 либо демонстрировал номинально лучшие результаты, либо был на одном уровне с двумя лечащими врачами и 4o, при этом значительные различия были обнаружены на первых двух этапах (рис. 4). Различия в результатах были особенно выражены на первом этапе диагностики (первичная сортировка в приемном отделении), где обычно наименьшее количество информации о пациенте и наибольшая срочность в принятии правильного решения. Модель o1 выявила точный или очень близкий диагноз (оценки по шкале Бонда от 4 до 5) в 67,1% случаев во время первоначальной сортировки в приемном отделении, в 72,4% случаев во время консультации с врачом приемного отделения и в 81,6% случаев при поступлении в терапевтическое отделение или отделение интенсивной терапии – превзойдя результаты двух врачей (55,3%, 61,8% и 78,9% для врача 1; 50,0%, 52,6% и 69,7% для врача 2) на каждом этапе.

Из этого видно, что быстрые темпы совершенствования LLM имеют существенные последствия для науки и практики клинической медицины. Хотя применение ИИ для поддержки принятия клинических решений иногда рассматривается как рискованное предприятие [136, 137], более широкое использование этих инструментов может помочь снизить человеческие и финансовые издержки, связанные с диагностическими ошибками, задержками и отсутствием доступа [138, 139]. Результаты указывают на острую необходимость проведения проспективных исследований для оценки этих технологий в реальных условиях оказания медицинской помощи пациентам, а также на необходимость подготовки систем здравоохранения к инвестициям в вычислительную инфраструктуру и разработке взаимодействия врача и ИИ, которое может способствовать безопасной интеграции инструментов ИИ в рабочие процессы оказания медицинской помощи пациентам.

Существующие исследования показывают, что современные базовые модели более ограничены в рассуждениях о нетекстовых входных данных [140, 141]; необходимы дальнейшие ис-

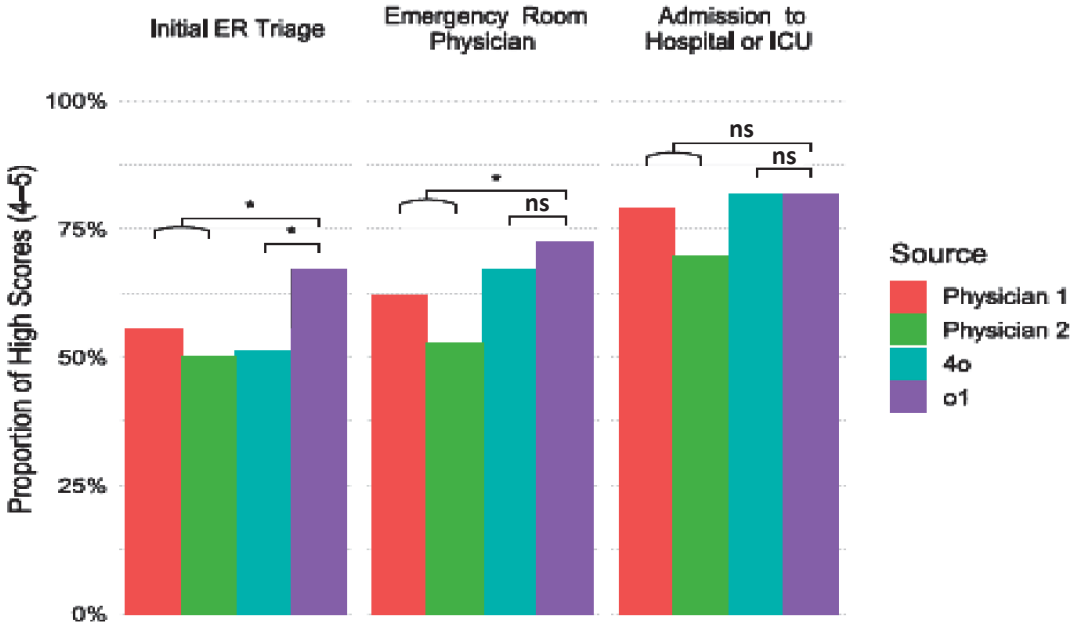

Рис. 4. Оценка второго мнения ИИ и эксперта-человека по реальным случаям из отделения экстренной помощи. Гистограмма, сравнивающая эффективность диагностики двух врачей-терапевтов, o1, и GPT-4o на 76 клинических случаях в трех диагностических точках (триаж в отделении экстренной помощи, первичная оценка врачом и госпитализация в больницу или отделение интенсивной терапии). Дифференциальные диагнозы были ограничены пятью диагнозами для всех участников. Источник дифференциального диагноза был скрыт и оценен двумя независимыми врачами-терапевтами по шкале Бонда. Показана доля ответов, оцененных на 4 или 5 баллов, что указывает на ответ, содержащий что-то точное или очень близкое к истинному диагнозу

следования для оценки того, как люди и машины могут эффективно сотрудничать [142] при использовании нетекстовых сигналов. Это требует новых критериев, испытаний и технологических решений для более точного измерения клинических взаимодействий.

Таким образом, модели ИИ развиваются и приобретают значительный потенциал в различных областях экстренной медицины, таких как сортировка пациентов, диагностика и прогнозирование исходов. Однако в основном они сталкиваются с такими проблемами, как изменчивость данных, обобщаемость моделей и интеграция в клинические рабочие процессы. Быстрое обновление версий требует, чтобы результаты, представленные в литературе, развивались с той же скоростью. Постоянное совершенствование моделей и повышение качества данных демонстрируют многообещающие результаты в практике оказания экстренной помощи.

Литература

1. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering (Basel). 2023; 10:1435.
2. Hosny A., Parmar C., Quackenbush J., Schwartz L.H., Aerts H.J. Artificial intelligence in radiology. Nat Rev Cancer. 2018; 18:500–510.
3. Drukker K., Chen W., Gichoya J., Gruszauskas N., Kalpathy-Cramer J., Koyejo S., et al. Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. J Med Imaging (Bellingham). 2023; 10:061104.
4. Babar M., Qureshi B., Koubaa A. Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging. PLoS One. 2024; 19:e0302539.
5. Chang Q., Yan Z., Zhou M., Qu H., He X., Zhang H., et al. Mining multi-center heterogeneous medical data with distributed synthetic learning. Nat Commun. 2023; 14:5510.
6. Masoudi S., Harmon S.A., Mehralivand S., Walker S.M., Raviprakash H., Bagci U., et al. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. J Med Imaging (Bellingham). 2021;8:010901.
7. Singh N.T., Kaur C., Chaudhary A., Goyal S. Preprocessing of medical images using deep learning: a comprehensive review. In: 2023 Second International Conference on Augmented Intelligence

and Sustainable Systems (ICAISS). IEEE; 2023. p. 521–527.
8. Krenzer A., Makowski K., Hekalo A., Fitting D., Troya J., Zoller W.G., et al. Fast machine learning annotation in the medical domain: a semi-automated video annotation tool for gastroenterologists. Biomed Eng Online. 2022; 21:33.
9. Galbusera F., Cina A. Image annotation and curation in radiology: an overview for machine learning practitioners. Eur Radiol Exp. 2024; 8:11.
10. Jawdekar A., Dixit M. A review of image enhancement techniques in medical imaging. In: Agrawal S., Gupta K.K., Chan J.H., Agrawal J., Gupta M., editors. Machine Intelligence and Smart Systems. Algorithms for Intelligent Systems. Singapore: Springer Nature; 2021. p. 25-33.
11. Fu S., Zhang M., Mu C., Shen X. Advancements of medical image enhancement in healthcare applications. J Healthc Eng. 2018; 2018:7035264.
12. Islam S.M., Mondal H.S. Image enhancement based medical image analysis. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE; 2019. p. 1–5.
13. Razmjooy N., Rajinikanth V., editors. Frontiers of Artificial Intelligence in Medical Imaging. Bristol (UK): IOP Publishing; 2022.
14. Goodfellow I., Bengio Y., Courville A. Deep learning. Cambridge (MA): MIT Press; 2016.
15. Zhou S.K., Greenspan H., Davatzikos C., Duncan J.S., van Ginneken B., Madabhushi A., et al. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proc IEEE Inst Electr Electron Eng. 2021; 109:820–838.
16. Zhang A., Lipton Z.C., Li M., Smola A.J. Dive into deep learning. Cambridge: Cambridge University Press; 2023.
17. Litjens G., Kooi T., Bejnordi B.E., Setio A.A., Ciompi F., Ghafoorian M., et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017; 42:60-88.
18. Han S., Pool J., Tran J., Dally W.J. Learning both weights and connections for efficient neural network. In: Advances in Neural Information Processing Systems 28 (NIPS 2015). Cambridge (MA): MIT Press; 2015.
19. Chu K.C., Yeh C.H., Lin J.M., Chen C.Y., Cheng C.Y., Yeh Y.Q., et al. Using convolutional neural network denoising to reduce ambiguity in X-ray coherent diffraction imaging. J Synchrotron Radiat. 2024;31:1340-1345.
20. Ait Nasser A., Akhloufi M.A. A review of recent advances in deep learning models for chest disease detection using radiography. Diagnostics (Basel). 2023; 13:159.
21. Shen D., Wu G., Suk H.I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017; 19:221–248.
22. Zhang Y., Chu Y., Yu H. Reduction of metal artifacts in X-ray CT images using a convolutional neural network. In: Müller B., Wang G., editors. Developments in X-Ray Tomography XI. Proc SPIE 10391. Bellingham (WA): SPIE; 2017. p. 103910V.
23. Tustison N.J., Avants B.B., Cook P.A., Zheng Y., Egan A., Yushkevich P.A., et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010; 29:1310–1320.
24. Kamnitsas K., Ledig C., Newcombe V.F., Simpson J.P., Kane A.D., Menon D.K., et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017; 36:61–78.
25. Zhang Q., Yang D., Zhu Y., Liu Y., Ye X. An optimized optical-flow-based method for quantitative tracking of ultrasound-guided right diaphragm deformation. BMC Med Imaging. 2023; 23:108.
26. Yu Y., Acton S.T. Speckle reducing anisotropic diffusion. IEEE Trans Image Process. 2002; 11:1260-1270.
27. Wang X., Yi J., Guo J., Song Y., Lyu J., Xu J., et al. A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sens. 2022; 14:5423.
28. Barros B., Lacerda P., Albuquerque C., Conci A. Pulmonary COVID-19: learning spatiotemporal features combining CNN and LSTM networks for lung ultrasound video classification. Sensors (Basel). 2021; 21:5486.
29. Prinster D., Mahmood A., Saria S., Jeudy J., Lin C.T., Yi P.H., et al. Care to explain? AI explanation types differentially impact chest radiograph diagnostic performance and physician trust in AI. Radiology. 2024; 313(2):e233261.
30. Ye Q., Wang Z., Lou Y., Yang Y., Hou J., Liu Z., et al. Deep learning approach based on a patch residual for pediatric supracondylar subtle fracture detection. Biomol Biomed. 2025;25(7):1631-1646. doi:10.17305/bb.2024.11341.
31. Wang C.H., Chang W., Lee M.R., Tay J., Wu C.Y., Wu M.C., et al. Deep learning-based diagnosis of pulmonary tuberculosis on chest X-ray in the emergency department: a retrospective study. J Imaging Inform Med. 2024; 37(2):589–600.
32. Quek J.J.X., Nickalls O.J., Wong B.S.S., Tan M.O. Deploying artificial intelligence in the detection of adult appendicular and pelvic fractures in the Singapore emergency department after hours: efficacy, cost savings and non-monetary benefits. Singapore Med J. 2025; 66(4):202–207. doi:10.4103/singaporemedj.SMJ-2023-170.
33. Kavak N., Kavak R.P., Güngörer B., Turhan B., Kaymak S.D., Duman E., et al. Detecting pediatric appendicular fractures using artificial intelligence. Rev Assoc Med Bras (1992). 2024; 70(9):e20240523.
34. López Alcolea J., Fernández Alfonso A., Cano Alonso R., Álvarez Vázquez A., Díaz Moreno A., García Castellanos D., et al. Diagnostic performance of artificial intelligence in chest radiographs referred from the emergency department. Diagnostics (Basel). 2024; 14(22):2592.

35. Novak A., Ather S., Gill A., Aylward P., Maskell G., Cowell G.W., et al. Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying pneumothoraces on plain chest X-ray: a multi-case multi-reader study. Emerg Med J. 2024; 41(10):602–609.
36. Lee C.K., Chen T.L., Wu J.E., Liao M.T., Wang C., Wang W., et al. Multimodal deep learning models utilizing chest X-ray and electronic health record data for predictive screening of acute heart failure in emergency department. Comput Methods Programs Biomed. 2024; 255:108357.
37. Ghatak A., Hillis J.M., Mercaldo S.F., Newbury-Chaet I., Chin J.K., Digumarthy S.R., et al. The potential clinical utility of an artificial intelligence model for identification of vertebral compression fractures in chest radiographs. J Am Coll Radiol. 2025; 22(2):220–229.
38. Kumari J., Kumar E., Kumar D. A structured analysis to study the role of machine learning and deep learning in the healthcare sector with big data analytics. Arch Comput Methods Eng. 2023;30(6):3673-3701.
39. Laletin V., Ayobi A., Chang P.D., Chow D.S., Soun J.E., Junn J.C., et al. Diagnostic performance of a deep learning-powered application for aortic dissection triage prioritization and classification. Diagnostics (Basel). 2024; 14(17):1877.
40. Kim J., Kwak C.W., Uhmn S., Lee J., Yoo S., Cho M.C., et al. A novel deep learning-based artificial intelligence system for interpreting urolithiasis in computed tomography. Eur Urol Focus. 2024; 10(6):1049–1054.
41. Lu C.Y., Wang Y.H., Chen H.L., Goh Y.X., Chiu I.M., Hou Y.Y., et al. Artificial intelligence application in skull bone fracture with segmentation approach. J Imaging Inform Med. 2025; 38(1):31–46.
42. Xu Y., Fu Q., Qu M., Chen J., Fan J., Hou S., et al. Automated hematoma detection and outcome prediction in patients with traumatic brain injury. CNS Neurosci Ther. 2024; 30:e70119.
43. Liu H.H., Chang C.B., Chen Y.S., Kuo C.F., Lin C.Y., Ma C.Y., et al. Automated detection and differentiation of Stanford type A and type B aortic dissections in CTA scans using deep learning. Diagnostics (Basel). 2024; 15(1):12.
44. Zhang C., Peng J., Wang L., Wang Y., Chen W., Sun M.W., et al. A deep learning-powered diagnostic model for acute pancreatitis. BMC Med Imaging. 2024; 24:154.
45. Choi S.Y., Kim J.H., Chung H.S., Lim S., Kim E.H., Choi A. Impact of a deep learning-based brain CT interpretation algorithm on clinical decision-making for intracranial hemorrhage in the emergency department. Sci Rep. 2024; 14:22292.
46. Kim J.Y., Choi H.J., Kim S.H., Ju H. Improved differentiation of cavernous malformation and acute intraparenchymal hemorrhage on CT using an AI algorithm. Sci Rep. 2024; 14:11818.
47. Wang Y., Zhang J., Li M., Miao Z., Wang J., He K., et al. SMART: development and application of a multimodal multi-organ trauma screening model for abdominal injuries in emergency settings. Acad Radiol. 2025; 32(5):2655–2666.
48. Ruitenbeek H.C., Oei E.H., Schmahl B.L., Bos E.M., Verdonschot R.J., Visser J.J. Towards clinical implementation of an AI algorithm for detection of cervical spine fractures on computed tomography. Eur J Radiol. 2024; 173:111375.
49. Warman P., Warman A., Warman R., Degnan A., Blickman J., Smith D., et al. Using an artificial intelligence software improves emergency medicine physician intracranial haemorrhage detection to radiologist levels. Emerg Med J. 2024; 41(5):298–303.
50. Kim J., Oh S.W., Lee H.Y., Choi M.H., Meyer H., Huwer S., et al. Assessment of deep learning-based triage application for acute ischemic stroke on brain MRI in the ER. Acad Radiol. 2024; 31(11):4621–4628.
51. Lang M., Clifford B., Lo W.C., Applewhite B.P., Tabari A., Filho A.L., et al. Clinical evaluation of a 2-minute ultrafast brain MR protocol for evaluation of acute pathology in the emergency and inpatient settings. AJNR Am J Neuroradiol. 2024; 45(4):379–385.
52. Kim C., Kwon J.M., Lee J., Jo H., Gwon D., Jang J.H., et al. Deep learning model integrating radiologic and clinical data to predict mortality after ischemic stroke. Heliyon. 2024; 10(10):e31000.
53. He B., Dash D., Duanmu Y., Tan T.X., Ouyang D., Zou J. AI-enabled assessment of cardiac function and video quality in emergency department point-of-care echocardiograms. J Emerg Med. 2024; 66(2):184–191.
54. Park S., Yoon H., Kang S.Y., Jo I.J., Heo S., Chang H., et al. Artificial intelligence-based evaluation of carotid artery compressibility via point-of-care ultrasound in determining the return of spontaneous circulation during cardiopulmonary resuscitation. Resuscitation. 2024; 202:110302.
55. Ge C., Jang J., Svrcek P., Fleming V., Kim Y.H. Exploring deep learning applications using ultrasound single view cines in acute gallbladder pathologies: preliminary results. Acad Radiol. 2025; 32(2):770–775.
56. Holland L., Hernandez Torres S.I., Snider E.J. Using AI segmentation models to improve foreign body detection and triage from ultrasound images. Bioengineering (Basel). 2024; 11(2):128.
57. Biousse V., Najjar R.P., Tang Z., Lin M.Y., Wright D.W., Keadey M.T., et al. Application of a deep learning system to detect papilledema on nonmydriatic ocular fundus photographs in an emergency department. Am J Ophthalmol. 2024; 261:199–207.
58. Li H., Cao J., You K., Zhang Y., Ye J. Artificial intelligence-assisted management of retinal detachment from ultra-widefield fundus images based

on weakly supervised approach. Front Med (Lausanne). 2024; 11:1326004.
59. Shobayo O., Saatchi R., Ramlakhan S. Convolutional neural network to classify infrared thermal images of fractured wrists in pediatrics. Healthcare (Basel). 2024; 12(10):994.
60. Wang Y., Ye Y., Shi S., Mao K., Zheng H., Chen X., et al. Prediagnosis recognition of acute ischemic stroke by artificial intelligence from facial images. Aging Cell. 2024; 23(8):e14196.
61. Zhao L., Vidwans A., Bearnot C.J., Rayner J., Lin T., Baird J., et al. Prediction of anemia in real-time using a smartphone camera processing conjunctival images. PLoS One. 2024; 19(5):e0302883.
62. Song A., Lusk J.B., Roh K.M., Hsu S.T., Valikodath N.G., Lad E.M., et al. RobOCTNet: robotics and deep learning for referable posterior segment pathology detection in an emergency department population. Transl Vis Sci Technol. 2024; 13(3):12.
63. Choi Y.J., Park M.J., Cho Y., Kim J., Lee E., Son D., et al. Screening for RV dysfunction using smartphone ECG analysis app: validation study with acute pulmonary embolism patients. J Clin Med. 2024; 13(16):4792.
64. Maxin A.J., Gulek B.G., Lim D.H., Kim S., Shaibani R., Winston G.M., et al. Smartphone pupillometry with machine learning differentiates ischemic from hemorrhagic stroke: a pilot study. J Stroke Cerebrovasc Dis. 2025; 34(1):108198.
65. Choi J., Kim J., Spaccarotella C., Esposito G., Oh I.Y., Cho Y., et al. Smartwatch ECG and artificial intelligence in detecting acute coronary syndrome compared to traditional 12-lead ECG. Int J Cardiol Heart Vasc. 2025; 56:101573.
66. Gharehchopogh F.S., Khalifelu Z.A. Analysis and evaluation of unstructured data: text mining versus natural language processing. In: 2011 5th International Conference on Application of Information and Communication Technologies (AICT); 2011 Oct 12–14; Baku, Azerbaijan. IEEE; 2011. P. 1–4.
67. Lee S.J., Alzeen M., Ahmed A. Estimation of racial and language disparities in pediatric emergency department triage using statistical modeling and natural language processing. J Am Med Inform Assoc. 2024; 31(4):958–967.
68. Farhat H., Alinier G., Tluli R., Chakif M., Rekik F.B., Alcantara M.C., et al. Enhancing patient safety in prehospital environment: analyzing patient perspectives on non-transport decisions with natural language processing and machine learning. J Patient Saf. 2024;20(5):330-339.
69. Levra A.G., Gatti M., Mene R., Shiffer D., Costantino G., Solbiati M., et al. A large language model-based clinical decision support system for syncope recognition in the emergency department: a framework for clinical workflow integration. Eur J Intern Med. 2025; 131:113–120.
70. Jang J.H., Lee S.W., Kim D.Y., Shin S.H., Lee S.C., Kim D.H., et al. Use of artificial intelligence-powered ECG to differentiate between cardiac and pulmonary pathologies in patients with acute dyspnoea in the emergency department. Open Heart. 2024; 11(2):e002924.
71. Lee S.H., Hong W.P., Kim J., Cho Y., Lee E. Smartphone AI versus medical experts: a comparative study in prehospital STEMI diagnosis. Yonsei Med J. 2024; 65(3):174–180.
72. Park M.J., Choi Y.J., Shim M., Cho Y., Park J., Choi J., et al. Performance of ECG-derived digital biomarker for screening coronary occlusion in resuscitated out-of-hospital cardiac arrest patients: a comparative study between artificial intelligence and a group of experts. J Clin Med. 2024; 13(5):1354.
73. Han J., Ahn K., Cha K., Kim S.J., Jung W.J., Roh Y.I., et al. Prediction of blood pressure using chest compression waveform during cardiopulmonary resuscitation. Resuscitation. 2024; 202:110331.
74. Kim T., Suh G.J., Kim K.S., Kim H., Park H., Kwon W.Y., et al. Development of artificial intelligence-driven biosignal-sensitive cardiopulmonary resuscitation robot. Resuscitation. 2024; 202:110354.
75. Choi D.H., Lee H., Joo H., Kong H.J., Lee S.B., Kim S., et al. Development of prediction model for intensive care unit admission based on heart rate variability: a case-control matched analysis. Diagnostics (Basel). 2024; 14(8):816.
76. Ou Z., Wang H., Zhang B., Liang H., Hu B., Ren L., et al. Early identification of stroke through deep learning with multi-modal human speech and movement data. Neural Regen Res. 2025; 20(1):234–241.
77. Zhang G., Xie Q., Wang C., Xu J., Liu G., Su C. Intelligent alert system for predicting invasive mechanical ventilation needs via noninvasive parameters: employing an integrated machine learning method with integration of multicenter databases. Med Biol Eng Comput. 2024; 62(11):3445–3458.
78. Liu L.R., Huang M.Y., Huang S.T., Kung L.C., Lee C.H., Yao W.T., et al. An arrhythmia classification approach via deep learning using single-lead ECG without QRS wave detection. Heliyon. 2024; 10:e27200.
79. Sen A., Navarro L., Avril S., Aguirre M. A data-driven computational methodology towards a prehospital acute ischaemic stroke screening tool using haemodynamics waveforms. Comput Methods Programs Biomed. 2024; 244:107982.
80. Woehrle T., Pfeiffer F., Mandl M.M., Sobtzick W., Heitzer J., Krstova A., et al. Point-of-care breath sample analysis by semiconductor-based E-Nose technology discriminates non-infected subjects from SARS-CoV-2 pneumonia patients: a multi-analyst experiment. MedComm (2020). 2024; 5(11):e726.
81. Herman R., Meyers H.P., Smith S.W., Bertolone D.T., Leone A., Bermpeis K., et al. International

evaluation of an artificial intelligence-powered electrocardiogram model detecting acute coronary occlusion myocardial infarction. Eur Heart J Digit Health. 2024; 5(2):123–133.
82. Heyman E.T., Ashfaq A., Ekelund U., Ohlsson M., Björk J., Khoshnood A.M., et al. A novel interpretable deep learning model for diagnosis in emergency department dyspnoea patients based on complete data from an entire health care system. PLoS One. 2024; 19(12):e0311081.
83. Flores E., Martínez-Racaj L., Blasco Á., Diaz E., Esteban P., López-Garrigós M., et al. A step forward in the diagnosis of urinary tract infections: from machine learning to clinical practice. Comput Struct Biotechnol J. 2024; 24:533–541.
84. Roshanaei G., Salimi R., Mahjub H., Faradmal J., Yamini A., Tarokhian A. Accurate diagnosis of acute appendicitis in the emergency department: an artificial intelligence-based approach. Intern Emerg Med. 2024; 19(8):2347–2357.
85. Saboorifar H., Rahimi M., Babaahmadi P., Farokhzadeh A., Behjat M., Tarokhian A. Acute cholecystitis diagnosis in the emergency department: an artificial intelligence-based approach. Langenbecks Arch Surg. 2024; 409(1):288.
86. Chang C.H., Nguyen P.A., Huang C.C., Liu C.F., Melisa S., Chen C.J., et al. Acute myocardial infarction risk prediction in emergency chest pain patients: an external validation study. Int J Med Inform. 2025; 193:105683.
87. Yilmaz R., Yagin F.H., Colak C., Toprak K., Abdel Samee N., Mahmoud N.F., et al. Analysis of hematological indicators via explainable artificial intelligence in the diagnosis of acute heart failure: a retrospective study. Front Med (Lausanne). 2024; 11:1285067.
88. Holmstrom L., Bednarski B., Chugh H., Aziz H., Pham H.N., Sargsyan A., et al. Artificial intelligence model predicts sudden cardiac arrest manifesting with pulseless electric activity versus ventricular fibrillation. Circ Arrhythm Electrophysiol. 2024; 17(2):e012338.
89. Aygun U., Yagin F.H., Yagin B., Yasar S., Colak C., Ozkan A.S., et al. Assessment of sepsis risk at admission to the emergency department: clinical interpretable prediction model. Diagnostics (Basel). 2024; 14(5):457.
90. Ben-Haim G., Yosef M., Rowand E., Ben-Yosef J., Berman A., Sina S., et al. Combination of machine learning algorithms with natural language processing may increase the probability of bacteremia detection in the emergency department: a retrospective, big-data analysis of 94,482 patients. Digit Health. 2024; 10:20552076241277673.
91. Toprak B., Solleder H., Di Carluccio E., Greenslade J.H., Parsonage W.A., Schulz K., et al. Diagnostic accuracy of a machine learning algorithm using point-of-care high-sensitivity cardiac troponin I for rapid rule-out of myocardial infarction: a retrospective study. Lancet Digit Health. 2024; 6(10):e729-e738.
92. Song Y.F., Huang H.N., Ma J.J., Xing R., Song Y.Q., Li L., et al. Early prediction of sepsis in emergency department patients using various methods and scoring systems. Nurs Crit Care. 2025; 30(3):e13201.
93. Li H., Liu Z., Sun W., Li T., Dong X. Interpretable machine learning for the prediction of death risk in patients with acute diquat poisoning. Sci Rep. 2024; 14:16101.
94. Brasen C.L., Andersen E.S., Madsen J.B., Hastrup J., Christensen H., Andersen D.P., et al. Machine learning in diagnostic support in medical emergency departments. Sci Rep. 2024; 14:17889.
95. Rahadian R.E., Okada Y., Shahidah N., Hong D., Ng Y.Y., Chia M.Y., et al. Machine learning prediction of refractory ventricular fibrillation in out-of-hospital cardiac arrest using features available to EMS. Resusc Plus. 2024; 18:100606.
96. Schipper A., Belgers P., O'Connor R., Jie K.E., Dooijes R., Bosma J.S., et al. Machine-learning based prediction of appendicitis for patients presenting with acute abdominal pain at the emergency department. World J Emerg Surg. 2024; 19(1):40.
97. Kijpaisalratana N., Saoraya J., Nhuboonkaew P., Vongkulbhisan K., Musikatavorn K. Real-time machine learning-assisted sepsis alert enhances the timeliness of antibiotic administration and diagnostic accuracy in emergency department patients with sepsis: a cluster-randomized trial. Intern Emerg Med. 2024; 19(5):1415–1424.
98. Chiu C.P., Chou H.H., Lin P.C., Lee C.C., Hsieh S.Y. Using machine learning to predict bacteremia in urgent care patients on the basis of triage data and laboratory results. Am J Emerg Med. 2024; 85:80–85.
99. Goodacre S. Using clinical risk models to predict outcomes: what are we predicting and why? Emerg Med J. 2023; 40(10):728–730.
100. Rahmatinejad Z., Dehghani T., Hoseini B., Rahmatinejad F., Lotfata A., Reihani H., et al. A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital mortality in the emergency department. Sci Rep. 2024; 14:3406.
101. Lee S., Lee K.S., Park S.H., Lee S.W., Kim S.J. A machine learning-based decision support system for the prognostication of neurological outcomes in successfully resuscitated out-of-hospital cardiac arrest patients. J Clin Med. 2024; 13(24):7600.
102. Richards J.E., Yang S., Kozar R.A., Scalea T.M., Hu P. A machine learning-based Coagulation Risk Index predicts acute traumatic coagulopathy in bleeding trauma patients. J Trauma Acute Care Surg. 2025; 98(4):614–620. doi:10.1097/TA.0000000000004463.
103. Ding H., Feng X., Yang Q., Yang Y., Zhu S., Ji X., et al. A risk prediction model for efficient intubation in the emergency department: a 4-year single-center retrospective analysis. J Am Coll Emerg Physicians Open. 2024; 5(3):e13190.

104. Ortiz-Barrios M., Petrillo A., Arias-Fonseca S., McClean S., de Felice F., Nugent C., et al. An AI-based multiphase framework for improving the mechanical ventilation availability in emergency departments during respiratory disease seasons: a case study. Int J Emerg Med. 2024; 17:45.
105. Li L., Han X., Zhang Z., Han T., Wu P., Xu Y., et al. Construction of prognosis prediction model and visualization system of acute paraquat poisoning based on improved machine learning model. Digit Health. 2024; 10:20552076241287891.
106. Deng Y.X., Wang J.Y., Ko C.H., Huang C.H., Tsai C.L., Fu L.C. Deep learning-based Emergency Department In-hospital Cardiac Arrest Score (Deep EDICAS) for early prediction of cardiac arrest and cardiopulmonary resuscitation in the emergency department. BioData Min. 2024; 17:52.
107. Shashikumar S.P., Le J.P., Yung N., Ford J., Singh K., Malhotra A., et al. Development and validation of a deep learning model for prediction of adult physiological deterioration. Crit Care Explor. 2024; 6(9):e1151.
108. Jawad B.N., Shaker S.M., Altintas I., Eugen-Olsen J., Nehlin J.O., Andersen O., et al. Development and validation of prognostic machine learning models for short- and long-term mortality among acutely admitted patients based on blood tests. Sci Rep. 2024; 14:5942.
109. Choi H.J., Lee C., Chun J., Seol R., Lee Y.M., Son Y.J. Development of a predictive model for survival over time in patients with out-of-hospital cardiac arrest using ensemble-based machine learning. Comput Inform Nurs. 2024; 42:388-395.
110. Park S.W., Yeo N.Y., Kang S., Ha T., Kim T.H., Lee D., et al. Early prediction of mortality for septic patients visiting emergency room based on explainable machine learning: a real-world multicenter study. J Korean Med Sci. 2024; 39:e53.
111. Siakopoulou S., Billis A., Logaras E., Stelmach V., Zouka M., Fyntanidou V., et al. Experimentation of AI models towards the prediction of medium-risk emergency department cases disposition outcome. Stud Health Technol Inform. 2024; 316:914–918.
112. Wang C.H., Tay J., Wu C.Y., Wu M.C., Su P.I., Fang Y.D., et al. External validation and comparison of statistical and machine learning-based models in predicting outcomes following out-of-hospital cardiac arrest: a multicenter retrospective analysis. J Am Heart Assoc. 2024; 13(20):e037088.
113. Lee S., Kim D.W., Oh N.E., Lee H., Park S., Yon D.K., et al. External validation of an artificial intelligence model using clinical variables, including ICD-10 codes, for predicting in-hospital mortality among trauma patients: a multicenter retrospective cohort study. Sci Rep. 2025; 15:1100.
114. Lin Y.T., Deng Y.X., Tsai C.L., Huang C.H., Fu L.C. Interpretable deep learning system for identifying critical patients through the prediction of triage level, hospitalization, and length of stay: prospective study. JMIR Med Inform. 2024; 12:e48862.
115. Nikouline A., Feng J., Rudzicz F., Nathens A., Nolan B. Machine learning in the prediction of massive transfusion in trauma: a retrospective analysis as a proof-of-concept. Eur J Trauma Emerg Surg. 2024; 50:1073–1081.
116. Tsai S.C., Lin C.H., Chu C.J., Lo H.Y., Ng C.J., Hsu C.C., et al. Machine learning models for predicting mortality in patients with cirrhosis and acute upper gastrointestinal bleeding at an emergency department: a retrospective cohort study. Diagnostics (Basel). 2024; 14(17):1919.
117. Hinson J.S., Zhao X., Klein E., Badaki-Makun O., Rothman R., Copenhaver M., et al. Multisite development and validation of machine learning models to predict severe outcomes and guide decision-making for emergency department patients with influenza. J Am Coll Emerg Physicians Open. 2024; 5:e13117.
118. Mehrpour O., Saeedi F., Vohra V., Hoyte C. Outcome prediction of methadone poisoning in the United States: implications of machine learning in the National Poison Data System (NPDS). Drug Chem Toxicol. 2024; 47:556–563.
119. Chen J.Y., Hsieh C.C., Lee J.T., Lin C.H., Kao C.Y. Patient stratification based on the risk of severe illness in emergency departments through collaborative machine learning models. Am J Emerg Med. 2024; 82:142–152.
120. Gauss T., Moyer J.D., Colas C., Pichon M., Delhaye N., Werner M., et al. Pilot deployment of a machine-learning enhanced prediction of need for hemorrhage resuscitation after trauma: the ShockMatrix pilot study. BMC Med Inform Decis Mak. 2024; 24:315.
121. Simon G.E., Johnson E., Shortreed S.M., Ziebell R.A., Rossom R.C., Ahmedani B.K., et al. Predicting suicide death after emergency department visits with mental health or self-harm diagnoses. Gen Hosp Psychiatry. 2024; 87:13–19.
122. Jawad B.N., Altintas I., Eugen-Olsen J., Niazi S., Mansouri A., Rasmussen L.J.H., et al. Prospective and external validation of machine learning models for short- and long-term mortality in acutely admitted patients using blood tests. J Clin Med. 2024; 13(21):6437.
123. Chang C.H., Chen C.J., Ma Y.S., Shen Y.T., Sung M.I., Hsu C.C., et al. Real-time artificial intelligence predicts adverse outcomes in acute pancreatitis in the emergency department: comparison with clinical decision rule. Acad Emerg Med. 2024; 31(2):149–155.
124. Soundararajan K., Adams D., Nathanson B., Mader T.J., Godwin R.C., Melvin R.L., et al. Use of machine learning models to predict neurologically intact survival for advanced age adults following out-of-hospital cardiac arrest. Acad Emerg Med. 2025; 32(2):169–171. doi:10.1111/acem.15018.

125. Yang S., Hu P., Kalpakis K., Burdette B., Chen H., Parikh G., et al. Utilizing ultra-early continuous physiologic data to develop automated measures of clinical severity in a traumatic brain injury population. Sci Rep. 2024; 14:7618.
126. Shung D.L., Chan C.E., You K., Nakamura S., Saarinen T., Zheng N.S., et al. Validation of an electronic health record-based machine learning model compared with clinical risk scores for gastrointestinal bleeding. Gastroenterology. 2024; 167(6):1198–1212.
127. Brodeur P.G., Buckley T.A., Kanjee Z., Goh E., Ling E.B., Jain P., et al. Performance of a large language model on the reasoning tasks of a physician. Science. 2026; 392(6797):524–527. doi:10.1126/science.adz4433.
128. Kanjee Z., Crowe B., Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA. 2023; 330(1):78–80. doi:10.1001/jama.2023.8288.
129. Johri S., Jeong J., Tran B.A., Schlessinger D.I., Wongvibulsin S., Barnes L.A., et al. An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med. 2025; 31(1):77–86. doi:10.1038/s41591-024-03328-5.
130. Rodman A., Zwaan L., Olson A.P., Manrai A.K. When it comes to benchmarks, humans are the only way. NEJM AI. 2025; 2(4):AIe2500143. doi:10.1056/AIe2500143.
131. Abdulnour R.-E.E., Parsons A.S., Muller D., Drazen J., Rubin E.J., Rencic J. Deliberate practice at the virtual bedside to improve clinical reasoning. N Engl J Med. 2022; 386(20):1946–1947. doi:10.1056/NEJMe2204540.
132. Schaye V., Miller L., Kudlowitz D., Chun J., Burk-Rafel J., Cocks P., et al. Development of a clinical reasoning documentation assessment tool for resident and fellow admission notes: a shared mental model for feedback. J Gen Intern Med. 2022; 37(3): 507–512. doi:10.1007/s11606-021-06805-6.
133. Goh E., Gallo R.J., Strong E., Weng Y., Kerman H., Freed J.A., et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nat Med. 2025; 31(4):1233–1238. doi:10.1038/s41591-024-03456-y.
134. Goh E., Gallo R., Hom J., Strong E., Weng Y., Kerman H., et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open. 2024; 7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969.
135. Berner E.S., Webster G.D., Shugerman A.A., Jackson J.R., Algina J., Baker A.L., et al. Performance of four computer-based diagnostic systems. N Engl J Med. 1994; 330(25):1792–1796. doi:10.1056/NEJM199406233302506.
136. Ratwani R.M., Bates D.W., Classen D.C. Patient safety and artificial intelligence in clinical care. JAMA Health Forum. 2024;5(2):e235514. doi: 10.1001/jamahealthforum.2023.5514.
137. Jin Q., Chen F., Zhou Y., Xu Z., Cheung J.M., Chen R., et al. Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine. NPJ Digit Med. 2024;7(1):190. doi:10.1038/s41746-024-01185-7.
138. Newman-Toker D.E., Peterson S.M., Badihian S., Hassoon A., Nassery N., Parizadeh D., et al. Diagnostic errors in the emergency department: a systematic review. Comparative Effectiveness Review No. 258. Rockville (MD): Agency for Healthcare Research and Quality; 2022. Report No.: 22(23)-EHC043.
139. Auerbach A.D., Lee T.M., Hubbard C.C., et al. Diagnostic errors in hospitalized adults who died or were transferred to intensive care. JAMA Intern Med. 2024; 184(2):164–173. doi:10.1001/jamainternmed.2023.7347.
140. Buckley T., Diao J.A., Rajpurkar P., Rodman A., Manrai A.K. Multimodal foundation models exploit text to make medical image predictions. arXiv:2311.05591 [cs.CV]. 2024.
141. Buckley T.A., Conci R., Brodeur P.G., Gusdorf J., Beltrán S., Behrouzi B., et al. Teaching large language models to reason like expert diagnosticians. arXiv:2509.12194 [cs.AI]. 2025.
142. Yu F., Moehring A., Banerjee O., Salz T., Agarwal N., Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med. 2024; 30(3):837–849. doi:10.1038/s41591-024-02850-w.
143. Berikol G.B., Kanbakan A., Ilhan B., Doğanay F. Mapping artificial intelligence models in emergency medicine: a scoping review on artificial intelligence performance in emergency care and education. Turk J Emerg Med. 2025; 25(2):67–91.

SHOSHILINCH TIBBIY YORDAMDA SUN'IY INTELLEKTNI QO'LLASH VA UNING SAMARADORLIGI: ANALITIK SHARH

F.T. ADİLOVA[1], R.R. DAVRONOV[1], X.M. QOSİMOV[2]

[1]O'zbekiston Respublikasi Fanlar akademiyasi V.I. Romanovskiy nomidagi Matematika instituti, Toshkent, O'zbekiston
[2]Respublika shoshilinch tibbiy yordam ilmiy markazi, Toshkent, O'zbekiston

Ushbu sharhning maqsadi – shoshilinch tibbiyotda sun'iy intellekt (SI) modellaridan foydalanish tendensiyasi va samaradorligini ko'rsatib berishdir. Natijalar shuni ko'rsatadiki, SIga

asoslangan tibbiy vizuallashtirish tizimlari rentgenografiya va kompyuter tomografiyasida kasalliklarni 85–90% aniqlik bilan aniqlashni ta'minlaydi. SI ko'magidagi bemorlarni saralash (triaj) tizimlari past va yuqori darajadagi shoshilinchlikka ega bemorlarni muvaffaqiyatli klassifikatsiya qiladi. Bu yangilangan SI modellarining salohiyatini namoyish etadi, biroq SI ni klinik ish jarayonlariga integratsiya qilish va modellarning umumlashtirish qobiliyati bilan bog'liq muammolar saqlanib qolayotganligi sababli, yanada keng ko'lamli tadqiqotlar o'tkazish zarur. Sharh qisman 2024–2025-yillardagi muvaffaqiyatli bajarilgan ishga (https://DOI:10.4103/tjem.tjem_45_251) va 2026-yilgi maqolaga asoslangan.

Kalit so'zlar: *Sun'iy intellekt, shoshilinch tibbiyot, tasvirlarga ishlov berish, katta til modellari, mashinali o'qitish.*

Сведения об авторах:

Адылова Фатима Туйчиевна – доктор технических наук, профессор, заведующая лабораторией «Биомединформатика» Института математики им. В.И. Романовского Академии наук Республики Узбекистан.
E-mail: fatadilova@gmail.com.
ORCID: 0000-0002-8607-6676

Давронов Рифкат Рахимович – кандидат технических наук, старший научный сотрудник лаборатории «Биомединформатика» Института математики им. В.И. Романовского Академии наук Республики Узбекистан.
E-mail: rifqat@gmail.com.
ORCID:0000-0003-2322-1802

Касимов Хамит Махмудович – кандидат технических наук, эксперт Республиканской ассоциации врачей экстренной медицинской помощи.
E-mal: medkib@mail.ru.
ORCID: 0009-0004-4103-460X

Поступила в редакцию: 26.05.2026

Information about authors:

Adylova Fatima Tuychievna – Doctor of Technical Sciences, Professor, Head of the Laboratory of Biomedinformatics, V.I. Romanovsky Institute of Mathematics, Academy of Sciences of the Republic of Uzbekistan.
E-mail: fatadilova@gmail.com.
ORCID: 0000-0002-8607-6676.

Davronov Rifkat Rakhimovich – Candidate of Technical Sciences, Senior Researcher, Laboratory of Biomedinformatics, V.I. Romanovsky Institute of Mathematics, Academy of Sciences of the Republic of Uzbekistan.
E-mail: rifqat@gmail.com.
ORCID: 0000-0003-2322-1802.

Kasimov Khamit Makhmudovich – Candidate of Technical Sciences, expert of the Republican Association of Emergency Medicine Physicians.
E-mail: medkib@mail.ru.
ORCID: 0009-0004-4103-460X.

Received: 26.05.2026