

КИРИНА М.В.,

АНО «Московский центр инновационных технологий в здравоохранении», г. Москва, Россия,
e-mail: KirinaMV@zdrav.mos.ru

БЕЗЫМЯННЫЙ А.С.,

ГБУЗ «ДКЦ №1 ДЗМ», г. Москва, Россия, e-mail: 03moscow@gmail.com

ВАСИЛЬЕВ Ю.А.,

к.м.н., ГБУЗ «НПКЦ ДиТ ДЗМ», г. Москва, Россия, e-mail: npcmr@zdrav.mos.ru

БЛОХИНА Е.В.,

ГКУ ДКД МО ДЗМ, г. Москва, Россия, e-mail: lebedeva488@gmail.com

КАРАМОВ Б.И.,

АНО «Московский центр инновационных технологий в здравоохранении», г. Москва, Россия,
e-mail: KaramovBI@zdrav.mos.ru

АБРОСИМОВ А.С.,

АНО «Московский центр инновационных технологий в здравоохранении», г. Москва, Россия,
e-mail: AbrosimovAS4@zdrav.mos.ru

АРЗАМАСОВ К.М.,

д.м.н., ГБУЗ «НПКЦ ДиТ ДЗМ», г. Москва, Россия, e-mail: ArzamasovKM@zdrav.mos.ru

ПАМОВА А.П.,

к.м.н. ГБУЗ «НПКЦ ДиТ ДЗМ», г. Москва, Россия, e-mail: PamovaAP@zdrav.mos.ru

КАЗАРИНОВА В.Е.,

ГБУЗ «НПКЦ ДиТ ДЗМ», г. Москва, Россия, e-mail: KazarinovaVE@zdrav.mos.ru

ОЦЕНКА РЕЗУЛЬТАТИВНОСТИ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ ДЛЯ ПОСТАНОВКИ ПРЕДВАРИТЕЛЬНОГО ДИАГНОЗА В УСЛОВИЯХ КОНСУЛЬТАТИВНО-ДИАГНОСТИЧЕСКИХ ПОЛИКЛИНИК ГОРОДА МОСКВЫ

DOI: 10.25881/18110193_2025_3_36

Аннотация. Внедрение системы поддержки принятия врачебных решений (СППВР) в клиническую практику требует тщательного контроля для обеспечения безопасности пациентов и оценки эффективности применения технологий искусственного интеллекта.

Целью данной работы является оценка результативности СППВР «ТОП-3» в условиях консультативно-диагностических поликлиник Департамента здравоохранения города Москвы.

Материалы и методы: Мониторинг работы СППВР «ТОП-3» проводился Департаментом здравоохранения города Москвы с 01.10.2020 по 21.03.2024 (n = 63 809 360 чел.). Рассчитывалась метрика Hit-3, на основе которой

принималось решение о необходимости повторного обучения представленной СППВР. Дополнительно было проведено исследование с участием врачей-экспертов: ретроспективный анализ данных на выборке из 3000 пациентов с расчетом согласованности диагнозов от СППВР, врача и эксперта.

Результаты: По результатам мониторинга среднее значение Hit-3 составляло 63,5, 64,5 и 67,7 для первой, второй и третьей версии СППВР соответственно. Экспертиза показала, что в выборке несоответствия диагноза от врача и СППВР ($n = 2000$) в 80,2% случаев эксперт на основе жалоб соглашался с СППВР, в 11,5% случаях - с врачом, а в 8,3% случаев ставил иной диагноз. В выборке соответствия диагноза врача с одним из диагнозов СППВР ($n = 1000$) в 50,4% случаев эксперт соглашался с диагнозом от врача и СППВР, в 37,9% случаев - с одним из двух других альтернативных диагнозов СППВР, в 11,7% случаев ставил иной диагноз.

Заключение: Описанная методика мониторинга, дополненная проведением экспертизы, позволила всесторонне оценить внедряемую в систему здравоохранения СППВР. По итогу оценки результативности «ТОП-3» было принято решение о необходимости расширения анализируемого перечня данных электронных медицинских карт, что будет внедрено в следующей версии СППВР «ТОП-3+».

Ключевые слова: цифровая медицина, искусственный интеллект, системы поддержки принятия врачебных решений, мониторинг эффективности, экспертная оценка.

Для цитирования: Кирина М.В., Безымянный А.С., Васильев Ю.А., Блохина Е.В., Карамов Б.И., Абросимов А.С., Арзамасов К.М., Памова А.П., Казаринова В.Е. Оценка результативности системы поддержки принятия врачебных решений для постановки предварительного диагноза в условиях консультативно-диагностических поликлиник города Москвы. Врач и информационные технологии. 2025; 3: 36-49. DOI: 10.25881/18110193_2025_3_36.

KIRINA M.V.,

Moscow Center for Healthcare Innovations, Moscow, Russia, e-mail: KirinaMV@zdrav.mos.ru

BEZMYANNYY A.S.,

State Budget-Funded Health Care Institution of the City of Moscow "Diagnostic Clinical Center No. 1 of the Moscow Health Care Department", Moscow, Russia, e-mail: 03moscow@gmail.com

VASILEV YU.A.,

PhD, Moscow Center for Diagnostics & Telemedicine, Moscow, Russia, e-mail: npcmr@zdrav.mos.ru

BLOKHINA E.V.,

The State Public Institution of the City of Moscow "The Directorate on Coordination of Activities of Medical Organizations of the Moscow Health Care Department", Moscow, Russia, e-mail: lebedeva488@gmail.com

KARAMOV B.I.,

Moscow Center for Healthcare Innovations, Moscow, Russia, e-mail: KaramovBI@zdrav.mos.ru

ABROSIMOV A.S.,

Moscow Center for Healthcare Innovations, Moscow, Russia, e-mail: AbrosimovAS4@zdrav.mos.ru

ARZAMASOV K.M.,

DSc, Moscow Center for Diagnostics & Telemedicine, Moscow, Russia, e-mail: ArzamasovKM@zdrav.mos.ru

PAMOVA A.P.,

PhD, Moscow Center for Diagnostics & Telemedicine, Moscow, Russia, e-mail: PamovaAP@zdrav.mos.ru

KAZARINOVA V.E.,

Moscow Center for Diagnostics & Telemedicine, Moscow, Russia, e-mail: KazarinovaVE@zdrav.mos.ru

EVALUATION OF THE EFFECTIVENESS OF A MEDICAL DECISION SUPPORT SYSTEM FOR PRELIMINARY DIAGNOSIS IN MOSCOW OUTPATIENT CLINICS

DOI: 10.25881/18110193_2025_3_36

Abstract. *The implementation of the medical decision support system (MDSS) in clinical practice requires careful monitoring to ensure patient safety and track the performance of artificial intelligence technologies.*

Goal: To assess the effectiveness of the TOP-3 MDSS in outpatient clinics of the Moscow Health Care Department.

Materials and methods: The Moscow Health Care Department monitored the TOP-3 operation between October 1, 2020 and March 03, 2024 ($n = 63,809,360$ people). The Hit-3 metric was used to determine if the MDSS needs retraining. An additional study involving medical experts included a retrospective analysis of data from 3,000 patients and calculation of the diagnostic agreement rate among the MDSS, doctors and the experts.

Results: The monitoring demonstrated mean Hit-3 of 63.5, 64.5 and 67.7 for MDSS ver. 1, 2, and 3, respectively. In cases where the clinicians disagreed with the MDSS ($n = 2000$), based on the patient complaints, experts agreed with the MDSS in

80.2% of cases, with clinicians in 11.5% of cases, and established a different diagnosis in 8.3% of cases. In cases where the clinicians' and the MDSS's conclusions matched ($n = 1000$), experts approved the diagnosis in 50.4% of cases, selected an alternative diagnosis suggested by the MDSS in 37.9% of cases, and established a different diagnosis for 11.7% of patients. Conclusion: The described monitoring methodology, supplemented by expert review, allowed for a comprehensive assessment of the MDSS to be implemented in the health care system. Based on the results of the TOP-3 effectiveness assessment, it was decided that the analyzed list of data from electronic health records should be expanded, which will be implemented in the upcoming version TOP-3+.

Keywords: digital medicine, artificial intelligence, medical decision support systems, efficiency monitoring; expert review.

For citation: Kirina M.V., Bezmyanny A.S., Vasilev YU.A., Blokhina E.V., Karamov B.I., Abrosimov A.S., Arzamasov K.M., Pamova A.P., Kazarinova V.E. Evaluation of the effectiveness of a medical decision support system for preliminary diagnosis in Moscow outpatient clinics. Medical doctor and information technology. 2025; 3: 36-49. DOI: 10.25881/18110193_2025_3_36.

ВВЕДЕНИЕ

В современном мире система здравоохранения переживает глубокую трансформацию. Этот процесс обусловлен активным внедрением искусственного интеллекта (ИИ) в медицинскую практику. В России этому процессу способствовал указ Президента Российской Федерации от 10.10.2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» [1], который определил стратегию развития ИИ и поставил задачу выхода нашей страны на лидирующие позиции в данной сфере.

На текущем этапе внедрения ИИ в медицинской диагностике выделилось два основных направления:

- 1) Компьютерное зрение [2], которое наиболее активно применяется в лучевой диагностике;
- 2) Анализ электронных медицинских карт (ЭМК).

Применение различных инструментов на основе ИИ позволяет повысить точность диагностики, например, за счет применения компьютерного зрения [3], или выстроить правильную врачебную тактику на основании анализа ЭМК с помощью систем поддержки принятия врачебных решений (СППВР) [4–6]. Настоящая публикация посвящена СППВР и ее жизненному циклу в условиях работы в медицинских организациях (МО) города Москвы.

ЭМК выступают как важный инструмент в современной медицине. Они обеспечивают врачу: предоставление полной информации о пациенте (его анамнезе, осмотрах и консультациях, данных лабораторной и инструментальной диагностики и т.д.), быстрое и эффективное заполнение медицинской документации во время приёма пациентов, снижая риск ошибок, связанных с человеческим фактором, и потери данных; аккумуляцию данных пациента из других МО. Пациенту ЭМК предоставляют удобный способ просмотра своей медицинской истории. Помимо этого, ЭМК является источником больших клинических данных, анализ которых с использованием алгоритмов обработки естественного языка открывает возможности для разработки новых технологий в медицине. Примером такой технологии является СППВР, которая использует данные ЭМК для помощи врачу.

СППВР — это программные продукты, которые связывают медицинские знания с данными конкретного пациента и через эту связь

помогают повышать эффективность деятельности врача [7]. Первые СППВР появились в 1970-х годах [8], тогда они требовали много времени для работы и имели плохую системную интеграцию, поэтому не получили широкого распространения. С развитием мировой экономики, ростом вычислительной способности техники и цифровизацией здравоохранения вырос и рынок СППВР [9]. Также расширился спектр функций, предоставляемых СППВР: помощь в лечении — подбор и контроль терапии, мониторинг пациентов, в том числе в удаленном режиме, анализ данных носимых устройств и оборудования с целью поддержки принятия решений медицинским персоналом [7].

В России на данный момент одной из внедрённых СППВР в здравоохранение является «Система поддержки принятия врачебных решений ТОП-3 диагнозов на основе данных электронной истории болезни» (далее «ТОП-3») [10], разработанная командой «СберМедИИ» совместно с Правительством Москвы при поддержке московских врачей. «ТОП-3» — сервис на базе искусственного интеллекта, который помогает ставить предварительный диагноз по жалобам пациента. СППВР была обучена на данных пациентов, собранных в г. Москве. В октябре 2020 года сервис был внедрен в работу во всех взрослых городских поликлиниках Москвы. «ТОП-3» включает в себя 265 групп диагнозов по международной классификации болезней (МКБ-10). Во время приема пациента сервис анализирует жалобы, которые вносит врач-терапевт или врач общей практики (далее — врач) в ЭМК, и предлагает три наиболее вероятных предварительных диагноза. Врач может выбрать один из предложенных вариантов или поставить свой диагноз.

Внедрение СППВР требует контроля со стороны органов исполнительной власти в сфере охраны здоровья для обеспечения безопасности пациентов и оценки эффективности применения технологий ИИ.

Целью данной работы является оценка результативности СППВР «ТОП-3» в условиях консультативно-диагностических поликлиник Департамента здравоохранения города Москвы.

МАТЕРИАЛЫ И МЕТОДЫ

Мониторинг работы СППВР «ТОП-3» [11] проводился Департаментом здравоохранения

города Москвы (далее ДЗМ) на протяжении 3,5 лет с 01.10.2020 по 21.03.2024 ($n = 63\,809\,360$ чел.) и охватывал весь период работы данного программного обеспечения. По результатам оценки рассчитывалась метрика Hit-3, давалась обратная связь разработчикам, проводилось дообучение сервиса и смена версии. Этапы проведения мониторинга схематично представлены на рисунке 1.

Врач на приеме оформлял протокол осмотра в Единой медицинской информационно-аналитической системе (ЕМИАС). СППВР «ТОП-3» анализировала информацию, указанную в разделе «Жалобы». Если заполненных данных было недостаточно для работы сервиса, выводилось уведомление об этом (рис. 2А), если достаточно — отображались три наиболее вероятных диагноза (рис. 2Б).

Информационное сообщение о вероятном предварительном диагнозе также отображалось у врача в блоке «Диагноз», если у пациента был установлен диагностический статус — «Предварительный» (рис. 3).

Далее врач мог согласиться с диагнозом, предложенным сервисом, или выставить свой.

Для оценки согласованности между двумя оценщиками применяется показатель Каппа Коэна. При расчете по этому методу используется классификация элемента со стороны оценщиков по взаимоисключающим категориям (бинарные исходы). В нашем случае множество классов патологий, бинарная оценка невозможна. Поэтому для расчета согласованности был выбран экстенсивный показатель, с учетом области применения и объема выборки он был модифицирован и назван Hit-3. Фактически,

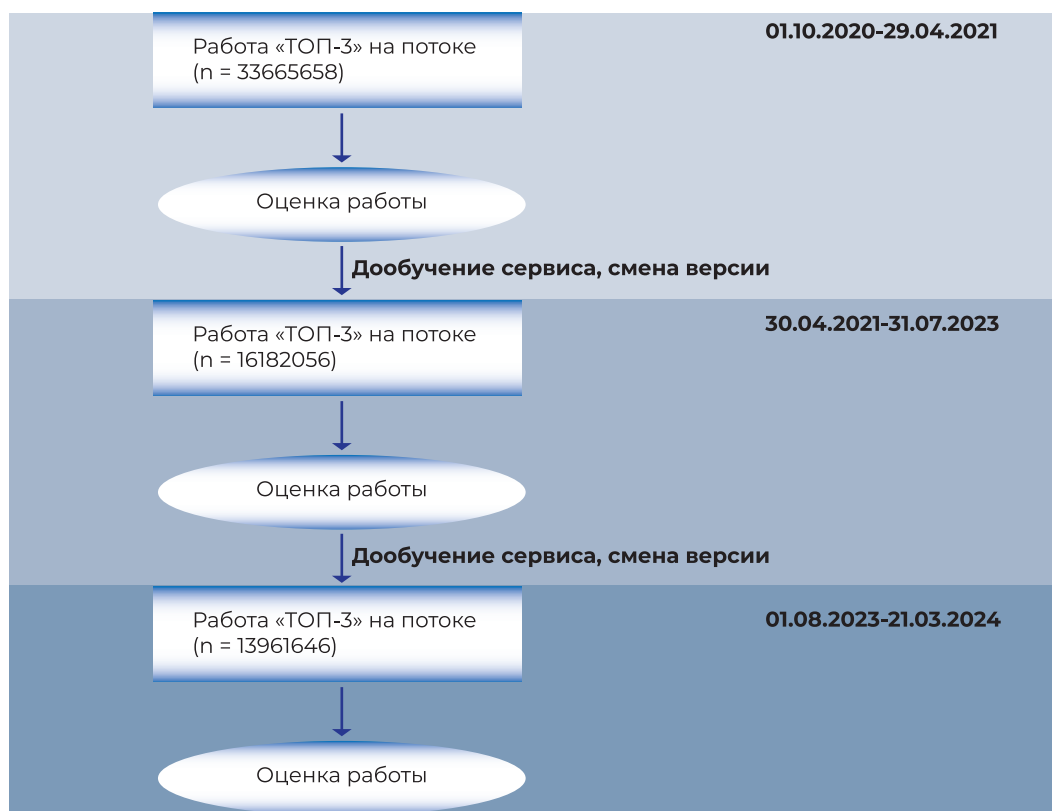


Рисунок 1 — Схема этапов мониторинга работы системы поддержки принятия врачебных решений «ТОП-3».

На момент осмотра жалобы не предъявляет предъявляет

Жалобы

головная боль

сухость во рту

тяжесть в нижних конечностях

В протоколе недостаточно данных для точной постановки диагноза цифровым помощником

На момент осмотра жалобы не предъявляет предъявляет

Жалобы

сухость во рту

боль в горле

затруднительное дыхание

увеличены железы

По данным цифрового помощника наиболее вероятные предварительные диагнозы:

- J31.0 — Хронический ринит
- J35.0 — Хронические болезни миндалин и аденоидов
- J39.0 — Другие болезни верхних дыхательных путей

Рисунок 2 — Рабочий экран врача с внедрённой СППВР «ТОП-3». А — уведомление, если данных для поставки предварительного диагноза недостаточно. Б — три вероятных диагноза от «ТОП-3».

На момент осмотра жалобы* не предъявляет предъявляет

Жалобы

повышенное давление

отеочность нижних конечностей

фасцикуляция мышц языка

По данным цифрового помощника наиболее вероятные предварительные диагнозы:

- I10 — Гипертензивная болезнь сердца (гипертоническая болезнь с преимущественным поражением сердца)
- E11 — Сахарный диабет 2 типа
- I83 — Варикозное расширение вен нижних конечностей

Диагноз

Основной диагноз

Диагностический статус Предварительный Подтвержден

Код по МКБ-10

Введите или выберите значение из списка

На основе данных о жалобах цифровым помощником сформированы наиболее вероятные предварительные диагнозы:

- E11 — Сахарный диабет 2 типа
- I10 — Гипертензивная болезнь сердца (гипертоническая болезнь с преимущественным поражением сердца)
- I83 — Варикозное расширение вен нижних конечностей

Развернутый клинический диагноз

Динамика заболевания ☐ Не выбрано ☐ Положительная

Рисунок 3 — Интерфейс СППВР «ТОП-3». А — информационное сообщение о вероятном предварительном диагнозе в блоке заполнения жалоб пациента; Б — блок «Диагноз», раздел для вынесения предварительного диагноза с предложением вариантов от «ТОП-3».

Hit-3 (Формула 1) отображает уровень совпадения предварительных диагнозов, которые поставил ИИ, с предварительными диагнозами, которые поставил врач на приёме:

$$\text{Hit-3} = \frac{n}{N} \times 100 \quad (1)$$

Формула 1 — Расчёт метрики Hit-3.

N — общее количество событий, где врач, а также «ТОП-3» поставили предварительные диагнозы;
n — количество успешных событий, где диагноз от врача на приеме (код по МКБ-10) совпал с одним из трёх диагнозов, предложенных «ТОП-3».

Дополнительно в 2023 году из-за ухудшения метрик работы сервиса было проведено исследование с участием врачей-экспертов (далее - экспертов), чтобы принять решение о дообучении сервиса. Оно представляло собой ретроспективный анализ работы «ТОП-3» на выборке из 3000 человек за период с 11 по 17 сентября 2023 года включительно. Выборка была разделена на две группы:

1. Группа несоответствия: 2000 человек, у которых диагноз врача не совпадал ни с одним из трех диагнозов, предложенных сервисом.
2. Группа соответствия: 1000 человек, у которых диагноз врача совпадал с одним из трёх диагнозов сервиса.

Выборка была отобрана из базы данных ЕМИАС.

В рамках исследования два эксперта анализировали жалобы пациентов во всех 3000 случаях. Если они расходились во мнениях, то случай дополнительно рассматривался врачебной комиссией, которая принимала окончательное решение о предварительном диагнозе. Затем были проведены сравнения:

- Диагнозы, поставленные экспертами на основе жалоб пациентов,
- Диагнозы «ТОП-3»,
- Диагнозы, поставленные врачом на приеме.

Одновременно с этим, согласно формуле расчёта статистической мощности для нашего объёма генеральной совокупности 3000 исследований достаточно (Формула 2):

$$SS = \frac{Z^2 \times (p) \times (1 - p)}{C^2} \quad (2)$$

Формула 2 — формула расчёта размера выборки. Z — фактор (1,96 для 95% доверительного интервала), p — процент интересующих нас пациентов из генеральной совокупности (0,5), C — доверительный интервал (1.79).

Критерии отбора врачей в качестве экспертов были следующие:

- 1) врач — сотрудник консультативно-диагностической поликлиники ДЗМ;
- 2) специальность — общая врачебная практика (семейная медицина)/терапия;
- 3) опыт работы — не менее 5 лет;
- 4) подтвержденный уровень квалификации (сданы специализированные квалификационные тесты).

Эксперты получали файл с жалобами пациента, тремя наиболее вероятными диагнозами по решению «ТОП-3» и диагнозом врача с приёма (для чистоты эксперимента диагнозы были расположены в случайном порядке без обозначения какой диагноз поставил врач, а какой предложила СППВР). На основании жалоб пациента эксперты выставляли предварительный диагноз из предложенных СППВР и врачом или, если ни один не подходил, ставили свой.

Анализ, визуализация и обработка результатов исследования проводилась с помощью программного обеспечения MS Excel и языка программирования Python в Jupyter Notebook, визуализация осуществлялась с использованием библиотек Python (pandas, matplotlib, plotly.express).

РЕЗУЛЬТАТЫ

Проведен мониторинг работы и экспертиза СППВР «ТОП-3». В результате дообучения ИИ «ТОП-3» произошла смена трёх версий:

- 1) 01.10.2020 — 29.04.2021;
- 2) 30.04.2021 — 31.07.2023;
- 3) 01.08.2023 — 21.03.2024.

Результаты мониторинга работы.

Общее число пациентов, которые посетили терапевта или врача общей практики и которым был поставлен предварительный диагноз со стороны врача и «ТОП-3», представлено на схеме ниже (рис. 4) с указанием версии СППВР.

Не все приёмы пациентов оканчиваются постановкой предварительного диагноза, это только от 30 до 53% приёмов за проанализированный период (от более чем 33 млн. посещений врача), в остальных случаях врач ставит заключительный диагноз. Поэтому ещё меньшее количество приёмов оканчиваются постановкой диагноза со стороны СППВР «ТОП-3» (от 27 до 45% от изначального числа пациентов) — часть диагнозов являются подтвержденными или данных для постановки предварительного диагноза недостаточно.

В качестве анализа работы СППВР были построены графики с линией тренда метрики Hit-3 и рассчитан коэффициент детерминации, показывающий долю дисперсии метрики Hit-3, которая измеряет степень соответствия регрессионной модели. Чем выше коэффициент детерминации, тем лучше регрессионная модель. Графически это изображено на рисунке 5. Если во время проведения мониторинга значение метрики Hit-3 снижалось до 60%, и линия тренда имела постоянную тенденцию на снижение, то принималось решение о необходимости дообучения СППВР «ТОП-3» со стороны ДЗМ, о чём сообщалось разработчику «СберМедИИ». Коэффициент детерминации (R²) линии тренда в версии 1 имел значение: 0,54, в версии 2: 0,57, в версии 3: 0,01.

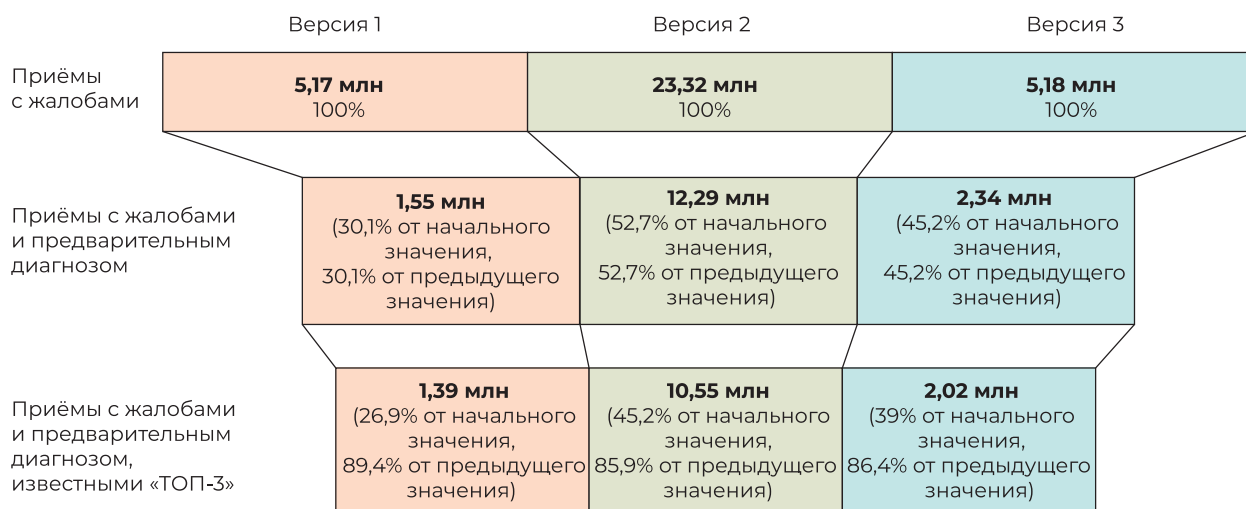


Рисунок 4 — Воронка конверсии пациентов от врача до постановки предварительного диагноза со стороны «ТОП-3». Отображает количество приёмов у врача с зафиксированными жалобами пациента, количество приёмов с жалобами и с предварительным диагнозом от врача и затем от «ТОП-3» в зависимости от версии «ТОП-3».

Таблица 1 — Среднее значение и медиана метрики Hit-3 в зависимости от версии «ТОП-3»

Hit-3	Версия 1	Версия 2	Версия 3
Среднее (Mean)	63,5	64,5	67,7
Медиана (Median)	63,4	63,1	67,7

Исходя из полученных значений коэффициента детерминации, можно сделать вывод, что метрика Hit-3 изменялась нелинейно и зависела от множества факторов, которые сложно идентифицировать. Среднее значение и медиана метрики Hit-3 приведены в таблице 1.

РЕЗУЛЬТАТЫ ЭКСПЕРТИЗЫ

Как уже говорилось выше, в 2023 году из-за снижения метрики Hit-3 было проведено исследование с участием экспертов, чтобы принять решение о дообучении сервиса. Результаты работы СППВР «ТОП-3» сравнивались с диагнозами экспертов на случайной выборке из 3000 пациентов (2000 случаев несоответствия и 1000 соответствия диагнозов от врача и от «ТОП-3»). СППВР и эксперт находились

в одинаковых условиях: постановка предварительного диагноза осуществлялась только на основе жалоб пациента. Таким образом, эксперты помимо основной задачи — определить, кто поставил правильный предварительный диагноз: врач, СППВР или оба, также оценивали достаточно ли только жалоб, чтобы поставить корректный диагноз. Результаты проведенной экспертизы отображены на рисунке 6.

В группе несоответствия диагнозов врача и СППВР (2000 случаев) в 80,2% случаев эксперт на основе жалоб соглашается только с СППВР и в 11,5% случаев — только с врачом. В 8,3% случаев эксперт ставит свой собственный диагноз.

В группе соответствия диагнозов врача и СППВР (1000 случаев) распределение

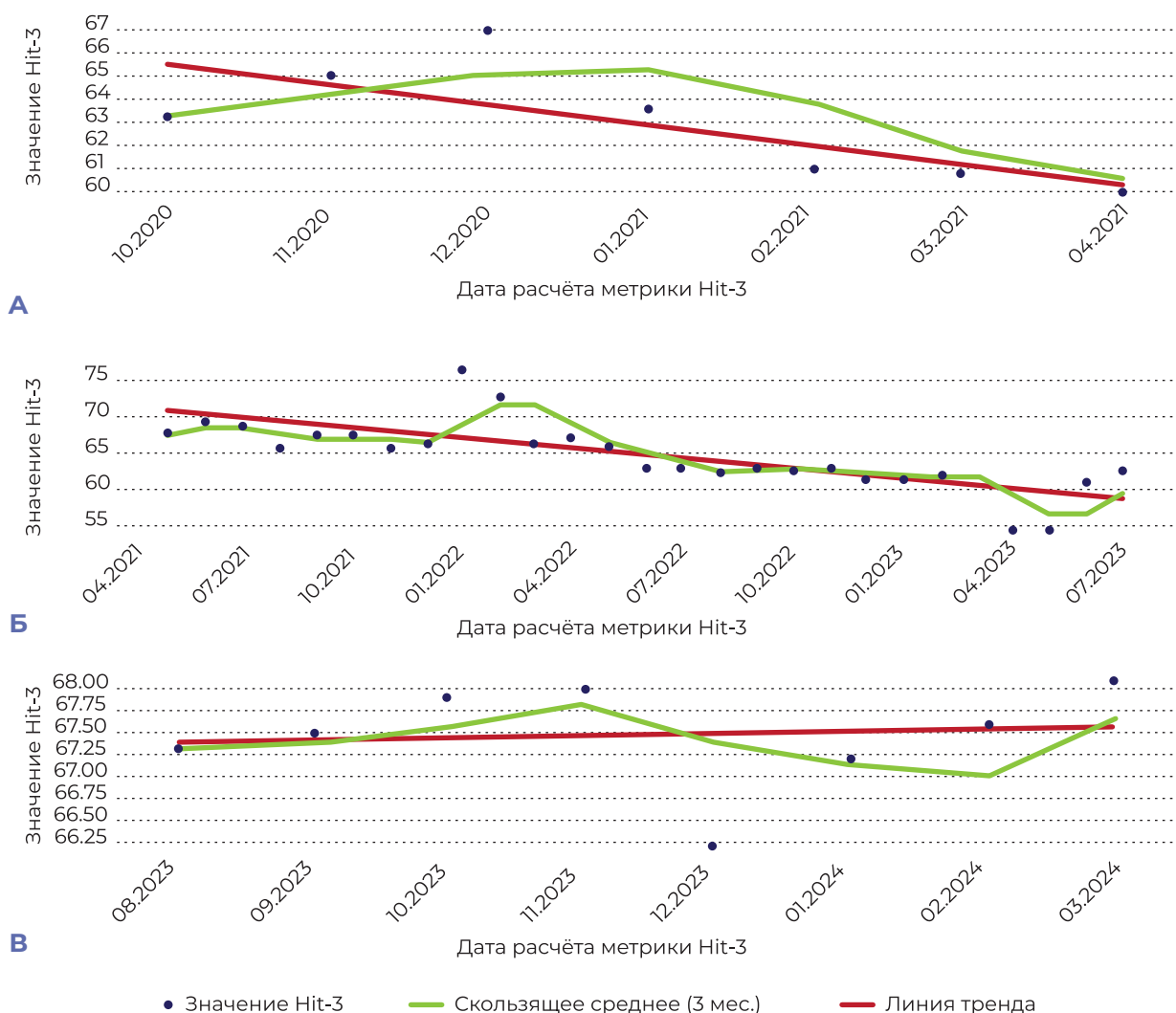


Рисунок 5 — Изменения метрики Hit — 3: фактические значения, скользящее среднее за 3 месяца, линия тренда. А — версия 1 «ТОП-3», Б — версия 2 «ТОП-3», В — версия 3 «ТОП-3».

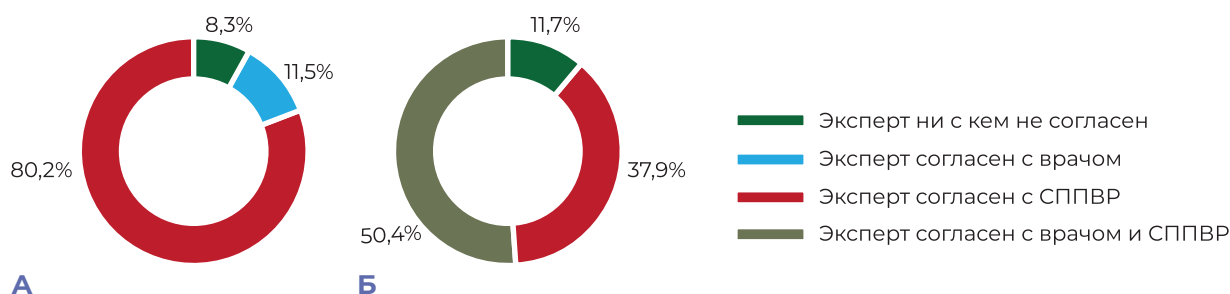


Рисунок 6 — Результаты экспертизы: А — анализ ответов экспертов в группе несоответствия (n = 2000); Б — анализ ответов экспертов в группе соответствия (n = 1000).

согласованности эксперта следующее: в 50,4% случаев эксперт соглашается с диагнозом врача и СППВР (которые также совпадают). В 37,9% случаев эксперт соглашается с одним из двух других альтернативных диагнозов СППВР. В 11,7% случаев эксперт не соглашается ни с врачом, ни с СППВР и ставит иной диагноз.

Таким образом, экспертиза показала, что одних жалоб недостаточно для выставления корректного конечного диагноза. Эксперты чаще соглашались с СППВР именно потому, что не имели возможности ознакомиться с другими данными ЭМК, помимо жалоб. В связи со снижением метрик сервис был отправлен на дообучение. При этом с целью повышения точности сервиса было принято решение расширить анализируемый перечень данных ЭМК. Так появилась новая СППВР «ТОП-3+», которая учитывает не только жалобы, но и пол, возраст, анамнез пациентов.

ОБСУЖДЕНИЕ

Из результатов, полученных выше, может сложиться впечатление рассогласованности врача, эксперта и «ТОП-3». Однако необходим объективный взгляд на эту ситуацию. Врач на приеме имеет преимущество: он видит пациента, проводит опрос и физикальный осмотр пациента, может дополнительно ознакомиться с анамнезом (историей болезни), либо знает пациента уже много лет.

Тем не менее, можно сказать и о том, что эксперт и СППВР принимают решение на основании ограниченной информации, и оно объективно относительно этих данных. Например, врач знает, что у пациента сахарный диабет (СД) II типа. При жалобах на боли в ногах он может автоматически списать это на СД, который приводит к разрушению микроциркуляторного русла в нижних конечностях и игнорировать тот факт, что это может быть началом развития другой сопутствующей патологии. Таким образом, СППВР, предоставляя дополнительное мнение о вероятном диагнозе пациента, обращает внимание врача на возможную патологию и тем самым помогает принять решение. Исследования показывают, что такая помощь особенно полезна специалистам с небольшим опытом работы, например, ординаторам [12]. Следует подчеркнуть, что в конечном счете именно врач принимает решение о диагнозе пациента.

Было проведено дообучение «ТОП-3» до версии «ТОП-3+», где у сервиса появился доступ к дополнительной клинической информации о пациенте (пол, возраст, анамнез), чтобы СППВР обладала входными данными, аналогичными тем, что доступны врачу на приеме. Экспертный анализ усовершенствованной СППВР будет проведён в дальнейшем.

Как указано в разделе «результаты», не все пациенты, посетившие первичный приём врача, получили предварительные диагнозы со стороны «ТОП-3». Случаи, где «жалобы» были заполнены, но СППВР не поставила предварительный диагноз, объясняются тем, что «ТОП-3» обучена ставить только 265 групп диагнозов по коду МКБ-10. Дополнительно к этому, отсутствие диагноза со стороны СППВР могло означать, что данных недостаточно. Это свидетельствует о том, что правильное и чёткое заполнение ЭМК врачом имеет высокое значение.

Hit-3 является метрикой оценки работы СППВР, которая позволяет сравнивать решение СППВР с диагнозом, поставленным врачом, фактически отражая их согласованность. На основе использования Hit-3 и полученных данных, можно принимать решение об отправке СППВР разработчику на дообучение. Дополнительное сравнение работы СППВР с экспертами и с врачами также позволило принять решение о необходимости усовершенствования текущей версии «ТОП-3» до «ТОП-3+», таким образом улучшая качество оказания медицинской помощи.

Оценка согласованности диагнозов, выставленных СППВР и врачом, использовалась в аналогичных исследованиях по изучению эффективности СППВР. Например, в работе Benditz A. и др. [13] совпадение диагнозов, поставленных спинальными хирургами и СППВР, в случае жалоб пациентов на боли в спине равнялось 67,4%, а в исследовании Купеевой И.А. и др. [14] соответствие решения СППВР на основании анализа симптомов дерматологических заболеваний и диагноза врача-эксперта составило 78%. В целом, для оценки эффективности и безопасности применения СППВР в клинической практике исследователи используют различные метрики: оценку согласованности, диагностической точности, клинических исходов и другие, что подтверждает отсутствие стандартного подхода

и создает трудности при сравнении различных СППВР между собой.

Жизненный цикл — непрерывный циклический процесс развития продукта от замысла до вывода из эксплуатации. На примере работы «ТОП-3» в медицинских организациях города Москвы было показано как после внедрения СППВР проводился мониторинг работы сервиса, по результатам которого СППВР дорабатывалась и происходила смена версии, а затем цикл повторялся. Происходила постоянная оценка работы СППВР, и давалась обратная связь разработчикам. Таким образом, постоянное дообучение моделей СППВР на основе нейронных сетей, взаимодействие конечных пользователей (врачей) и разработчиков важно для получения продукта наивысшего качества [15]. Уже существуют определенная методология жизненного цикла ИИ-сервиса [16] и критерии оценки зрелости программного обеспечения на основе ИИ [17], однако они применимы для компьютерного зрения, а для анализа ЭМК с помощью СППВР четких рекомендаций нет. Были проведены аналогичные исследования [18, 19], в которых оценивалась эффективность СППВР по различным метрикам. Однако ценность настоящей публикации заключается в масштабе и длительности исследования, проведенного в условиях реальной клинической практики. Данная работа является ступенью к созданию комплексной методологии оценки качества и эффективности СППВР.

ЗАКЛЮЧЕНИЕ

Средняя точность Hit-3 оцениваемой СППВР составила 63,5, 64,5, 67,7 для первой, второй и третьей версии «ТОП-3», соответственно. Для усовершенствования модели и повышения ее диагностических показателей в дальнейшем требуется учитывать большее количество признаков: пол, возраст и данные анамнеза пациента, что будет внедрено в следующей версии «ТОП-3+».

Ограничением исследования являлся тот факт, что СППВР обучена ставить 265 групп диагнозов, поэтому часть приёмов у врача оканчивались отсутствием диагноза со стороны «ТОП-3».

Описанный дизайн мониторинга работы СППВР на приеме врача общей практики или врача-терапевта, дополнительное сравнение работы сервиса с экспертом и сравнение врача с экспертом может быть использовано повсеместно. Проведенная экспертиза позволяет всесторонне оценить внедряемую в систему здравоохранения СППВР и принять решение о её доработке и дообучении. Что, в конечном счете, приводит к более эффективному распределению рабочего времени врача на приеме (так как он может опираться на один из диагнозов со стороны СППВР), где больше времени будет уделено непосредственно пациенту.

Конфликт интересов. Авторы декларируют отсутствие конфликта интересов.

ЛИТЕРАТУРА/REFERENCES

1. Указ Президента Российской Федерации от 10.10.2019 г. №490 «О развитии искусственного интеллекта в Российской Федерации». <http://www.kremlin.ru/acts/bank/44731>.
2. Васильев Ю.А., Владимирский А.В., Арзамасов К.М. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента. — М.: Издательские решения, 2023. 376 с. [Vasiliev YuA, Vladimirovskiy AV, Arzamasov KM. Komp'yuternoe zrenie v luchevoj diagnostike: pervy`j e`tap Moskovskogo e`ksperimenta. M.: Izdatel'skie resheniya. 2023. 376 p. (In Russ.)]
3. Васильев Ю.А., Тыров И.А., Владимирский А.В., Арзамасов К.М. и др. Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований // Digital Diagnostics. — 2023. — Т.4. — №2. — С.93-104. [Vasilev YA, Tyrov IA, Vladzimirskyy AV, Arzamasov KM, et al. Double-reading mammograms using artificial intelligence technologies: A new model of mass preventive examination organization. Digital Diagnostics. 2023; 4(2): 93-104. (In Russ.)] doi: 10.17816/DD321423.
4. Комков А.А., Рязанова С.В., Мазаев В.П. Эффективные системы поддержки принятия решений в клинической практике и профилактике: обзор литературы // CardioСоматика. — 2023. — Т.14. — №3. — С.177-185. [Komkov AA, Ryazanova SV, Mazaev VP. Effective decision support systems in Clinical Practice and Prevention: Literature Review. CardioSomatics. 2023; 14(3): 177-185. (In Russ.)] doi: 10.17816/CS569263.

5. Гусев А.В., Токарев С.А., Гаврилов Д.В., Кузнецова Т.Ю. Применение автоматизированной системы поддержки принятия врачебных решений при диспансеризации взрослого населения для контроля правильности оценки уровня сердечно-сосудистого риска // Менеджмент качества в медицине. — 2022. — №1. — С.72-79. [Gusev AV, Tokarev SA, Gavrilov DV, Kuznecova TYu. Primenenie avtomatizirovannoj sistemy` podderzhki prinyatiya vrachebny`x reshenij pri dispanserizacii vzroslogo naseleniyadlya kontrolya pravil`nosti ocenki urovnya serdechno-sosudistogo riska. Menedzhment kachestva v medicine. 2022; 1: 177-185. (In Russ.)]
6. Taheri Moghadam S, Sadoughi F, Velayati F, Ehsanzadeh SJ, Poursharif S. The effects of clinical decision support system for prescribing medication on patient outcomes and physician practice performance: a systematic review and meta-analysis. BMC Med Inform Decis Mak. 2021; 21(1): 98. doi: 10.1186/s12911-020-01376-8.
7. Васильев Ю.А., Лантух З.А., Шатёнок М.П., Солдатов И.В. и др. Методика выбора системы поддержки принятия врачебных решений в лучевой диагностике для обеспечения радиационной безопасности: методические рекомендации // М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2023. — 20 с. [Vasiliev YA, Lantux ZA, Shatyonok MP, Soldatov IV, et al. Metodika vy`bora sistemy` podderzhki prinyatiya vrachebny`x reshenij v luchevoj diagnostike dlya obespecheniya radiacionnoj bezopasnosti: metodicheskie rekomendacii. M.: Moscow Center for Diagnostics and Telemedicine; 2023. (In Russ.)]
8. Edward HS, Bruce GB. A Model of Inexact Reasoning in Medicine. Mathematical Biosciences. 1975; 23: 351-379. doi: 10.1016/0025-5564(75)90047-4.
9. Пугачев П.С., Гусев А.В., Кобякова О.С., Кадыров Ф.Н. и др. Мировые тренды цифровой трансформации отрасли здравоохранения // Национальное здравоохранение. — 2021. — Т.2. — №2. — С.5-12. [Pugachev PS, Gusev AV, Kobyakova OS, Kady`rov FN, et al. Mirovy`e trendy` cifrovoj transformacii otrasli zdravooxraneniya. Nacional`noe zdravooxranenie. 2021; 2(2): 5-12. (In Russ.)]
10. Регистрационное удостоверение на медицинское изделие от 24.05.2022 №ПЗН 2022/17272. <https://goszdravnadzor.ru/ru-roszdravnadzora-№-rzn-2022-17272-ot-24-maya-2022-go/>
11. Blinov P, Avetisian M, Kokh V, Umerenkov D, Tuzhilin A. Predicting Clinical Diagnosis from Patients Electronic Health Records Using BERT-Based Neural Networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2020: 111-121. doi: 10.1007/978-3-030-59137-3_11.
12. Martinez-Franco AI, Sanchez-Mendiola M, Mazon-Ramirez JJ, et al. Diagnostic accuracy in Family Medicine residents using a clinical decision support system (DXplain): a randomized-controlled trial. Diagnosis (Berl). 2018; 5(2): 71-76. doi: 10.1515/dx-2017-0045.
13. Benditz A, Pulido LC, Grifka J, Ripke F, Jansen P. A clinical decision support system in back pain helps to find the diagnosis: a prospective correlation study. Arch Orthop Trauma Surg. 2023; 143(2): 621-625. doi: 10.1007/s00402-021-04080-y.

14. Купеева И.А., Разнатовский К.И., Раводин Р.А., Карелин В.В. и др. Оценка эффективности интеллектуальной системы поддержки принятия врачебных решений // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. — 2016. — №2. — С.62-68. [Kupeeva IA, Raznatovskij KI, Ravodin RA, Karelin VV, et al. Ocenka effektivnosti intellektual'noj sistemy podderzhki prinyatiya vrachebny'x reshenij. Vestnik Sankt-Peterburgskogo universiteta. Prikladnaya matematika. Informatika. Processy upravleniya. 2016; 2: 62-68. (In Russ.)) doi: 10.21638/11701/spbu10.2016.206.
15. Заюнчковский С.Ю., Коновалов С.А., Зинченко В.В., Шарова Д.Е. и др. Система менеджмента качества: инструмент развития организации или дополнительная нагрузка? // Digital Diagnostics. — 2023. — Т.4. — №3. — С.439-447. [Zayunchkovskij SYU, Konovalov SA, Zinchenko VV, SHarova DE, et al. Sistema menedzhmenta kachestva: instrument razvitiya organizacii ili dopolnitel'naya nagruzka? Digital Diagnostics. 2023; 4(3): 439-447. (In Russ.)) doi: 10.17816/DD321971.
16. Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Арзамасов К.М. и др. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики // Digital Diagnostics. — 2023. — Т.4. — №3. — С.252-267. [Vasiliev YA, Vlazimirsky AV, Omelyanskaya OV, Arzamasov KM, et al. Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics. Digital Diagnostics. 2023; 4(3): 252-267. (In Russ.)) doi: 10.17816/dd321971.
17. Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Шулькин И.М. и др. Оценка зрелости технологий искусственного интеллекта для здравоохранения: методические рекомендации. — М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2023. — 28 с. [Vasiliev YA, Vladzimirskij AV, Omelyanskaya OV, Shul'kin IM, et al. Ocenka zrelosti tehnologij iskusstvennogo intellekta dlya zdravooxraneniya: metodicheskie rekomendacii. M.: Moscow Center for Diagnostics and Telemedicine; 2023. (In Russ.))
18. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. J Am Med Inform Assoc. 2023; 30(10): 1730-1740. doi: 10.1093/jamia/ocad120.
19. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. J Am Med Inform Assoc. 2024; 31(5): 1172-1183. doi: 10.1093/jamia/ocae060.