

<https://doi.org/10.33878/2073-7556-2025-24-4-125-137>



# Искусственный интеллект для прогнозирования отдаленных результатов лечения больных колоректальным раком (систематический обзор и метаанализ)

Абдулаева Р.Ш.<sup>1</sup>, Павлова В.И.<sup>1,2</sup>, Геворкян Т.Г.<sup>1</sup>, Беленькая Я.В.<sup>1</sup>, Манукян М.Ш.<sup>1</sup>, Гордеев С.С.<sup>1,2</sup>

<sup>1</sup>ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России (Каширское ш., д. 23, г. Москва, 115522, Россия)

<sup>2</sup>ФГБУ ВО «Тюменский государственный медицинский университет» Минздрава России (ул. Одесская, д. 54, г. Тюмень, 625023, Россия)

**РЕЗЮМЕ** *ЦЕЛЬ:* оценить эффективность алгоритмов искусственного интеллекта для прогнозирования отдаленных результатов лечения больных колоректальным раком (КРР) на основе клинических данных. *МАТЕРИАЛЫ И МЕТОДЫ:* проведен систематический поиск научных публикаций за 2015–2024 гг. в базах данных PubMed, ScienceDirect, MedRxiv, BioRxiv и Google Scholar. Включены оригинальные исследования, применявшие методы машинного обучения и глубокого обучения исключительно на основе клинических данных для прогнозирования рецидива КРР. Из 657106 выявленных публикаций критериям включения соответствовали 43 исследования, из которых 12 вошли в метаанализ. Оценивались общая площадь под ROC-кривой (AUC), показатели гетерогенности ( $I^2$ ,  $\tau^2$ , Q-тест), наличие публикационного смещения и чувствительность результатов. Чувствительность результатов метаанализа была подтверждена методом leave-one-out. *РЕЗУЛЬТАТЫ:* по результатам проведенного анализа было установлено, что наиболее часто применяемыми алгоритмами были Random Forest (67%), Support Vector Machine (51%) и XGBoost (37%). Общая объединенная прогностическая точность моделей машинного обучения в прогнозировании общей выживаемости КРР показала очень хорошие результаты — AUC = 0,86 (95% ДИ: 0,82–0,89). Вместе с тем, выявлена значительная межисследовательская гетерогенность ( $I^2 = 97,6\%$ ,  $p < 0,001$ ) и умеренное публикационное смещение. *ЗАКЛЮЧЕНИЕ:* высокая прогностическая точность моделей ИИ подтверждает их потенциал для интеграции в клиническую практику при прогнозировании рецидива КРР. Однако существенная гетерогенность между исследованиями ограничивает возможность прямого сравнения эффективности различных алгоритмов и требует осторожности в интерпретации результатов.

**КЛЮЧЕВЫЕ СЛОВА:** колоректальный рак, искусственный интеллект, машинное обучение, прогнозирование выживаемости, прогностическая модель

**КОНФЛИКТ ИНТЕРЕСОВ:** авторы заявляют об отсутствии конфликта интересов

**ФИНАНСИРОВАНИЕ:** Исследование спонсировано Аналитическим Центром правительства Российской Федерации (соглашение № 70-2024-000121 от 29.03.2024. ICG 000000D730324P540002)

**ДЛЯ ЦИТИРОВАНИЯ:** Абдулаева Р.Ш., Павлова В.И., Геворкян Т.Г., Беленькая Я.В., Манукян М.Ш., Гордеев С.С. Искусственный интеллект для прогнозирования отдаленных результатов лечения больных колоректальным раком (систематический обзор и метаанализ). *Колопроктология*. 2025; т. 24, № 4, с. 125–137. <https://doi.org/10.33878/2073-7556-2025-24-4-125-137>

## Artificial intelligence for predicting long-term outcomes in patients with colorectal cancer (a systematic review and meta-analysis)

Rukiyat Sh. Abdulaeva<sup>1</sup>, Valeria I. Pavlova<sup>1,2</sup>, Tigran G. Gevorkyan<sup>1</sup>, Yana V. Belenkaya<sup>1</sup>, Mariam Sh. Manukyan<sup>1</sup>, Sergey S. Gordeev<sup>1,2</sup>

<sup>1</sup>N.N. Blokhin National Medical Research Center of Oncology (Kashirskoye shosse, 23, Moscow, 115522, Russia)

<sup>2</sup>Tyumen State Medical University (Odesskaya Street, 54, Tyumen, 625023, Russia)

**ABSTRACT** Evaluate the performance of artificial-intelligence algorithms in predicting long-term treatment outcomes in patients with colorectal cancer using clinical data alone to evaluate the performance of artificial-intelligence algorithms in predicting long-term treatment outcomes in patients with colorectal cancer (CRC) using clinical data alone.

**MATERIALS AND METHODS:** a systematic search (2015–2024) was conducted in PubMed, Science Direct, MedRxiv, BioRxiv and Google Scholar. Original studies that applied machine-learning or deep-learning techniques exclusively to clinical variables for predicting CRC recurrence were included. Of 657106 records screened, 43 met the eligibility criteria; 12 were entered into a meta-analysis. Pooled area under the ROC curve (AUC), heterogeneity metrics ( $I^2$ ,  $\tau^2$ , Q-test), publication bias and sensitivity were assessed. Robustness was examined with a leave-one-out analysis.

**RESULTS:** a systematic search (2015–2024) in PubMed, Science Direct, MedRxiv, BioRxiv and Google Scholar. Original studies that applied machine-learning or deep-learning techniques exclusively to clinical variables for predicting CRC recurrence were included. Of 657106 records screened, 43 met the eligibility criteria; 12 were entered into a meta-analysis. Pooled area under the ROC curve (AUC), heterogeneity metrics ( $I^2$ ,  $\tau^2$ , Q-test), publication bias and sensitivity were assessed. Robustness was examined with a leave-one-out analysis.

**CONCLUSION:** AI models show promising accuracy in predicting colorectal cancer recurrence, supporting their potential utility in clinical decision-making. Nevertheless, further validation in large-scale, prospective studies is required before widespread clinical implementation.

**KEYWORDS:** colorectal cancer; artificial intelligence; survival prediction; machine learning; deep learning; meta-analysis

**CONFLICT OF INTEREST:** the authors declare no conflict of interest

**FUNDING:** this research was funded by the Analytical Center for the Government of the Russian Federation under agreement No. 70-2024-000121 dated March 29, 2024 (IGK 000000D730324P540002).

**FOR CITATION:** Abdulaeva R.Sh., Pavlova V.I., Gevorkyan T.G., Belenkaya Ya.V., Manukyan M.Sh., Gordeev S.S. Artificial intelligence for predicting long-term outcomes in patients with colorectal cancer (a systematic review and meta-analysis). *Koloproktologia*. 2025;24(4):125–137. (in Russ.). <https://doi.org/10.33878/2073-7556-2025-24-4-125-137>

**АДРЕС ДЛЯ ПЕРЕПИСКИ:** Абдулаева Рукият Шамильевна, ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России, ул. Каширское шоссе, д. 23, Москва, e-mail: ruutlevi@gmail.com

**ADDRESS FOR CORRESPONDENCE:** Rukiyat Shamiliyeva Abdulaeva, N.N. Blokhin National Medical Research Center of Oncology, Kashirskoe шоссе, 23, Moscow, Russia; e-mail: ruutlevi@gmail.com

Дата поступления — 16.06.2025

Received — 16.06.2025

После доработки — 02.09.2025

Revised — 02.09.2025

Принято к публикации — 10.11.2025

Accepted for publication — 10.11.2025

## ВВЕДЕНИЕ

Прогнозирование отдаленных результатов лечения больных колоректальным раком (КРР) представляет собой одну из ключевых задач онкологии, поскольку позволяет персонализировать терапевтические стратегии и оптимизировать послеоперационный мониторинг.

Современные подходы на основе искусственного интеллекта (ИИ), включающие алгоритмы машинного обучения и глубокого обучения, позволяют эффективно анализировать большие массивы данных и улучшать точность прогноза. Среди различных возможных источников информации для анализа, клинические данные являются наиболее доступными и стандартизированными для прогнозирования онкологических исходов. В отличие от молекулярно-генетических маркеров, их сбор не требует высоких затрат и специализированного лабораторного оборудования, что делает их широко применяемыми в реальной практике. Кроме того, использование исключительно клинических параметров обеспечивает

лучшую интерпретируемость моделей и способствует их интеграции в существующие протоколы ведения пациентов. На сегодняшний день применяются различные алгоритмы машинного и глубокого обучения, такие как градиентный бустинг (XGBoost, LightGBM), Random Forest и ансамблевые методы. Однако остаётся нерешённым вопрос о том, какой из этих алгоритмов наиболее эффективен для анализа клинических данных и насколько их применение позволит оптимизировать тактику ведения пациентов с КРР.

## ЦЕЛЬ

Анализ современных исследований, использующих методы искусственного интеллекта для прогнозирования отдаленных результатов лечения больных колоректальным раком на основе клинических данных. Основной задачей исследования является сравнение различных алгоритмов искусственного интеллекта, их прогностической точности и потенциальной клинической применимости.

## МАТЕРИАЛЫ И МЕТОДЫ

Поиск научных публикаций осуществлялся с учетом статей, опубликованных в период с 2015 по 2024 гг., в пяти ведущих онлайн-базах данных: PubMed, ScienceDirect, MedRxiv, BioRxiv и Google Scholar. Основной поисковый запрос был «("Artificial intelligence" OR "Machine learning" OR "Deep learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning") AND ("Colorectal Cancer" OR "Rectal Cancer" OR "Colorectal adenocarcinoma" OR "Colon Cancer") AND (diagnos\* OR detect\* OR predict\* OR screen\*)». Данный запрос применялся при поиске в PubMed и Google Scholar. Однако из-за ограничений по количеству символов, он не использовался в других базах данных. Для ScienceDirect, MedRxiv и BioRxiv применялся поисковый запрос: ("Artificial intelligence" OR "Machine learning" OR "Deep learning") AND ("Colorectal Cancer" AND "Rectal Cancer") AND (predict). В этот обзор вошли только исследования, посвященные методам ИИ, используемым для прогнозирования риска рецидива или прогрессирования колоректального рака.

В исследование были включены оригинальные исследования, и исключены клинические наблюдения, обзоры литературы, доклады конференций. Мы также исключили исследования, в которых использовались методы, не связанные с ИИ. Кроме того, были исключены исследования, предоставляющие теоретическую основу для моделей ИИ, применяемых к колоректальному раку. Для данного обзора рассматривались только исследования, опубликованные на английском языке.

Процесс отбора исследований состоял из трех этапов. На первом этапе мы провели поиск литературы в вышеупомянутых базах данных, затем использовали Rayuan для удаления дубликатов всех выявленных исследований. На втором этапе два независимых рецензента проанализировали заголовки и аннотации всех найденных статей, исключив исследования, не соответствующие теме обзора. На заключительном этапе рецензенты независимо рассмотрели полные тексты статей, прошедших предыдущий этап, а все несоответствия между рецензентами разрешались в ходе обсуждения.

Для оценки обобщенной диагностической точности моделей машинного обучения, предназначенных для прогнозирования общей выживаемости у пациентов, был проведен метаанализ 12-ти исследований. В каждом исследовании анализировалась модель с наилучшим значением площади под кривой (AUC).

### **Синтез данных**

После извлечения данных из включенных исследований был использован подход нарративного синтеза.

Синтез суммировал и описывал методы искусственного интеллекта, примененные в исследованиях, сосредоточившись на их целях, характеристиках, источниках данных и алгоритмах (например, Random Forest, Support Vector Machine и др.). Помимо AUC дополнительно оценивали accuracy (общая точность), sensitivity / recall (чувствительность / полноту), specificity (специфичность), precision / PPV (положительную прогностическую ценность), NPV (отрицательную прогностическую ценность), F1-score (среднее точности и полноты), C-index (индекс конкордантности) и HR (отношение рисков). Кроме того, модели искусственного интеллекта были сопоставлены по ключевым параметрам: типу алгоритма, характеристикам исходных данных (размер выборки, вид клинического исхода) и набору показателей качества модели. Управление всеми извлеченными данными на протяжении синтеза осуществляли в Microsoft Excel.

### **Статистический анализ**

Статистическую обработку данных метаанализа выполняли в RStudio (R v. 4.4.2; R Core Team, Vienna, Austria) с использованием пакета metafor v. 4.8-0. Исходя из предположения о межисследовательской гетерогенности, объединенный эффект рассчитывали по модели случайных эффектов. Уровень гетерогенности оценивали по статистикам  $\tau^2$ ,  $I^2$  и Q-тесту Кокрана; значения  $I^2$  свыше 50% трактовали как высокую гетерогенность. Для визуализации индивидуальных и совокупных оценок был построен лесовидный график (forest plots), возможное публикационное смещение проверяли при помощи воронкообразного графика (funnel plots). С целью оценки устойчивости объединенной оценки, был проведен анализ чувствительности с использованием метода leave-one-out: поочередно исключалась каждая публикация для определения её влияния на итоговую AUC и уровень гетерогенности. Также был выполнен модерационный анализ с типом алгоритма в качестве категориального модератора для выяснения вклада каждого метода в общую вариативность результатов.

## РЕЗУЛЬТАТЫ

### **Результаты поиска**

По данным поисковым запросам было идентифицировано 657106 статей из 5-ти баз данных: PubMed ( $n = 2551$ ), Science Direct ( $n = 635180$ ), Google Scholar ( $n = 18900$ ), BioRxiv ( $n = 345$ ), и MedRxiv ( $n = 130$ ). Все статьи из баз данных PubMed, BioRxiv, MedRxiv были проанализированы на основе указанного запроса. Из-за большого объема статей из баз данных, Science Direct и Google Scholar были рассмотрены только

первые 400 (отсортированных по релевантности). Всего для подробного анализа было включено 3526 статей (Рис. 1).

В ходе первоначального поиска было выявлено 657106 статей. Из них 3526 статей были проанализированы, а 3483 были исключены по следующим причинам: 1214 статей не были связаны с искусственным интеллектом, 348 статей не были посвящены колоректальному раку, 455 были обзорами литературы, а 1366 не имели отношения к прогнозированию рецидива. Так же было исключено 100 исследований, в которых изучались модели искусственного интеллекта на основе радиологических, гистологических изображений и геномных данных. В конечном итоге 43 статьи соответствовали критериям включения и были включены в данный обзор.

### Включенные статьи

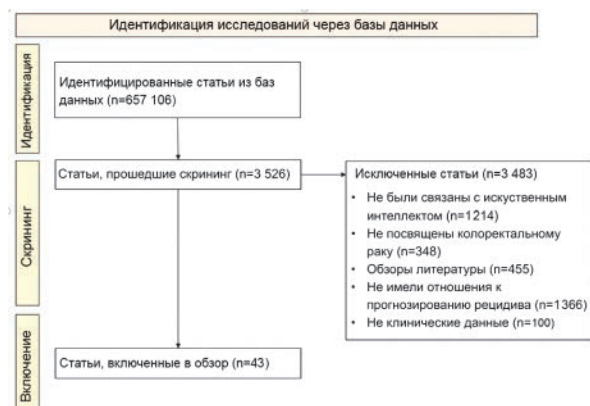
Все включенные исследования были опубликованы в рецензируемых журналах (43 из 43, 100%). Публикации охватывают период с 2018 по 2024 гг.: 2/43 (4,7%) были опубликованы в 2018 году, 2/43 (4,7%) — в 2019 году, 2/43 (4,7%) — в 2020 году, 3/43 (7,0%) — в 2021 году, 9/43 (20,9%) — в 2022 году, 13/43 (30,2%) — в 2023 году и 12/43 (27,9%) — в 2024 году. Количество пациентов в этих исследованиях варьировалось от 164 до 528 060.

Во всех 43 исследованиях использовались алгоритмы машинного обучения, при этом в 3 (7%) исследованиях дополнительно применялись методы глубокого обучения. Random Forest оказался самым часто используемым алгоритмом, задействованным в 29/43 (67%) исследованиях. Далее следует Support Vector Machine, применявшийся в 22/43 (51%) исследованиях, затем логистическая регрессия и XGBoost, каждый из которых встречался в 16/43 (37%) исследованиях. Среди прочих методов можно отметить Naïve

**Таблица 1.** Характеристики используемых методов искусственного интеллекта

**Table 1.** Characteristics of the artificial intelligence methods used

Типы	Исследования N = 43 (100%)	Источник
Тип ИИ		
Машинное обучение (МО)	43/43 (100%)	[1–42]
Глубокое обучение (ГО)	3/43 (7%)	[6,18,43]
Алгоритмы/модели/методы ИИ		
Random Forest	29/43 (67%)	[1, 2, 4–12, 14, 18, 20, 21, 23–28, 30, 32–35, 40, 42, 43]
Support Vector Machine	22/43 (51%)	[1–5, 7–10, 12, 15, 18, 20, 21, 23, 27, 30, 32, 35, 41–43]
Logistic Regression	16/43 (37%)	[1, 2, 4–6, 8–12, 18, 20, 26, 34, 41, 42]
XG Boost	16/43 (37%)	[8, 10–12, 14, 17, 24, 26, 27, 29, 30, 32, 34–37]
Naive Bayes	7/43 (16%)	[4, 8, 14, 18, 21, 25, 41]
Artificial Neural Networks (ANN/DNN)	11/43 (26%)	[4, 6, 9, 12, 18, 19, 23, 34, 41, 43]
Decision Tree	15/43 (35%)	[5, 6, 8, 10, 12, 15, 20, 21, 24, 25, 30, 32, 41–43]
Cat Boost	3/43 (7%)	[7, 13, 24]
Auto-AI (автоматизированный перебор моделей)	1/43 (2%)	[22]
K-Nearest Neighbors Algorithm	14/43 (33%)	[8, 9, 12, 15, 21, 23, 24, 26, 27, 34, 35, 42, 43]
Extreme gradient boosting model	4/43 (9%)	[1, 2, 13, 20]
Gradient Boosting	9/43 (21%)	[6, 8, 16, 20, 21, 24, 37, 38, 40]
Light GBM	8/43 (19%)	[8, 16, 20, 21, 24, 26, 39, 41]
mboost	1/43 (2%)	[6]
Ada Boost	3/43 (7%)	[7, 9, 25]



**Рисунок 1.** Блок-схема PRISMA-ScR (предпочтительные элементы отчетности для систематических обзоров и метаанализов)

**Figure 1.** PRISMA-ScR flowchart (preferred reporting elements for systematic reviews and meta-analyses)

Bayes 7/43 (16%), Artificial Neural Networks 11/43 (26%) и Decision Tree 15/43 (35%). Характеристики методов ИИ, применяемых в каждом исследовании, представлены в таблице 2.

В таблице 2 представлены ключевые характеристики проведенных исследований, включающие общее количество пациентов, число анализируемых параметров, применяемые алгоритмы и используемые методы валидации. Чаще всего в качестве входных переменных фигурировали: возраст, пол, локализация опухоли, стадия опухолевого процесса по классификации TNM, количество удаленных и пораженных лимфоузлов, размер опухоли и степень дифференцировки, наличие лимфоваскулярной и периневральной инвазии, статус краёв резекции, уровень онкомаркёров СЕА и СА 19-9, мутационный статус, показатели общего и биохимического анализа крови показатели крови

**Таблица 2.** Подробная характеристика включенных исследований  
**Table 2.** Detailed characteristics of the included studies

Автор, год	Алгоритм	Прогнозируемый клинический исход	Результаты	Метод валидации	Число анализируемых параметров	Количество пациентов
Youngbae Jeon, et al. [1]	SVM, RF, EGB, LR	Рецидив	<p><b>AUC:</b> SVM — 0,79 (95% ДИ: 0,76–0,82); RF — 0,84 (95% ДИ: 0,81–0,87); XGBoost — 0,85 (95% ДИ: 0,82–0,88).</p> <p><b>Accuracy:</b> SVM — 76% (95% ДИ: 70–82); RF — 82%; XGBoost — 83%.</p> <p><b>Sensitivity:</b> SVM — 70% (95% ДИ: 65–75); RF — 78%; XGBoost — 79%;</p> <p><b>Specificity:</b> SVM — 80% (95% ДИ: 75–85); RF — 86%; XGBoost — 88%;</p>	Tomek link	16	961
Mo Tang, et al. [2]	LR, RF, EGB, SVM	Рецидив	<p><b>AUC:</b> LR — 0,73 (95% ДИ: 0,70–0,76); RF — 0,74; XGBoost — 0,79.</p> <p><b>Accuracy:</b> LR — 78% (95% ДИ: 74–82); RF — 79%; XGBoost — 80%.</p> <p><b>Sensitivity:</b> LR — 75% (95% ДИ: 71–79); RF — 76%; XGBoost — 78%.</p> <p><b>Specificity:</b> LR — 80%; RF — 81%; XGBoost — 83%.</p>	Set Aside Method	20	558
Wen-Chien Ting, et al. [3]	SVM, Reduced Error Pruning Tree	Рецидив	<p><b>AUC:</b> SVM — 0,87 (95% ДИ: 0,83–0,91); Reduced Error Pruning Tree — 0,62.</p> <p><b>Accuracy:</b> SVM — 83% (95% ДИ: 79–87); Reduced Error Pruning Tree — 78%.</p> <p><b>Sensitivity:</b> SVM — 79%; Reduced Error Pruning Tree — 72%.</p> <p><b>Specificity:</b> SVM — 86%; Reduced Error Pruning Tree — 80%.</p>	10-кратная кросс-валидация	7	4299
Achilonu OJ, et al. [4]	naïve Bayes (NB), C5.0, LR, RF SVM, ANN	Рецидив	<p><b>AUC:</b> LR — 0,93; NB — 0,92; RF — 0,94; ANN — 0,95 (95% ДИ: 0,92–0,98).</p> <p><b>Accuracy:</b> LR — 87%; NB — 86%; RF — 88%.</p> <p><b>Sensitivity:</b> LR — 85%; NB — 83%; RF — 86%.</p> <p><b>Specificity:</b> LR — 89%; NB — 87%; RF — 90%.</p>	10-кратная кросс-валидация	13	716
Chen PC, et al. [5]	LR, RF, DT, SVM	Рецидив	<p><b>AUC:</b> LR — 0,87; RF — 0,84; CART — 0,83.</p> <p><b>Accuracy:</b> LR — 84%; RF — 83%; CART — 82%.</p> <p><b>Sensitivity:</b> LR — 81%; RF — 80%; CART — 79%.</p> <p><b>Specificity:</b> LR — 86%; RF — 85%; CART — 84%.</p>	5-кратная кросс-валидация	13	1073
Alinia S, Zhou C, et al. [6]	DT, RF, RSF, GB, mboost, DLNN, Cox Regression	Рецидив и смертность	<p><b>Прогнозирование смерти</b>  <b>Accuracy:</b> RF — 96% (0,81–0,99); GB — 71,4% (0,57–0,83); Mboost — 89% (0,78–0,96);  <b>Sensitivity:</b> DT — 78,3% (0,56–0,92); RF — 100% (0,85–1); RSF — 77,5% (0,66–0,86); GB — 75,9% (0,56–0,89); Mboost — 96,9% (0,83–0,99);</p> <p><b>Прогнозирование рецидива</b>  <b>Accuracy:</b> RF — 67,8% (0,47–0,84); RSF — 74% (0,66–0,81); GB — 96% (0,87–0,99); mboost — 76% (0,63–0,87);  <b>Sensitivity:</b> RF — 100% (0,82–1); RSF — 77,8% (0,66–0,86); GB — 100% (0,87–1);  Mboost — 96,8% (0,83–0,99);  <b>Specificity:</b> RF — 0% (0–0,28); RSF — 71% (0,58–0,81); GB — 92,9% (0,76–0,99); Mboost — 53,8% (0,33–0,73); DLNN — 98% (0,97–0,98)</p>	Train-Test Split	7	284
Erkan K., et al., 2024 [7]	CatBoost, RF, AdaBoost, 16 ML models tested	Рецидив	<p><b>AUC:</b> CatBoost — 0,92; RF — 0,78; AdaBoost — 0,80.</p> <p><b>Accuracy:</b> CatBoost — 88%; RF — 79%; AdaBoost — 69%.</p> <p><b>Recall:</b> CatBoost — 70%; RF — 67%; AdaBoost — 69%.</p> <p><b>Precision:</b> CatBoost — 90%; RF — 81%; AdaBoost — 83%.</p> <p><b>F1-score:</b> CatBoost — 77%; RF — 74%; AdaBoost — 75%.</p>	10-кратная кросс-валидация + train-test split	9	396

Автор, год	Алгоритм	Прогнозируемый клинический исход	Результаты	Метод валидации	Число анализируемых параметров	Количество пациентов
Susić et al., 2023 [8]	LR, DT, RF, KNN, Naive Bayes, SVM, GB, LGBM, XGBoost	1-, 2-, 3-, 4-, 5-летняя ОБ	<b>5-летняя ОБ AUC:</b> LR — 0,872; DT — 0,756; RF — 0,810; KNN — 0,800; Naive Bayes — 0,712; SVM — 0,855; GB — 0,854; LGBM — 0,868; XGBoost — 0,855.	5-кратная кросс-валидация + train-test split	7	1236
Gupta et al., 2019 [9]	RF, SVM, LR, Multilayer Perceptron, K-NN, Adaptive Boosting	БРВ	<b>AUC:</b> RF — 0,82 ± 0,10; SVM — 0,77 ± 0,03; Logistic Regression: 0,76 ± 0,02; <b>MLP — 0,78 ± 0,11; KNN — 0,75 ± 0,06; AB — 0,77 ± 0,03.</b> <b>Accuracy:</b> RF — 84%; SVM — 77%; Logistic Regression — 76%; ML — 78%; KNN — 75%; Adaptive Boosting — 77%.	5-кратная кросс-валидация + train-test split	12	4021
Hu et al., 2023 [10]	DT, RF, SVM, XGBoost, LR	Рецидив	<b>AUC:</b> Decision Tree — 0,72; Random Forest — 0,97; SVM — 0,89; XGBoost — 0,94. <b>Sensitivity:</b> (Финальная модель, Тестовая выборка): 94% Specificity (Финальная модель, Тестовая выборка): 86,2% Sensitivity (Финальная модель, Тестовая выборка): 94,4% Specificity (Финальная модель, Тестовая выборка) 66,7%	10-кратная кросс-валидация + train-test split	12	272
Leonard G, et al., 2022 [11]	RF, XGBoost, XGBoost with SMOTE, LR	Смертность	<b>AUC:</b> Logistic Regression- 0,730 (95% ДИ: 0,725–0,735); Random Forest- 0,757 (95% ДИ: 0,752–0,762); XGBoost- 0,756 (95% ДИ: 0,751–0,761); XGBoost with SMOTE- 0,748 (95% ДИ: 0,743–0,753).	Train-Test Split	30	528060
Nopour, 2024 [12]	Random Forest, XGBoost, Bagging, Logistic Regression, SVM, ANN, Decision Tree, KNN	5-летняя ОБ	<b>AUC:</b> XGBoost: 0,906 (internal), 0,813 (external) (лучшая модель) Random Forest: 0,825 Bagging: 0,883 SVM: 0,815 ANN: 0,724 Decision Tree: 0,726 KNN: 0,786 Logistic Regression: 0,727	10-кратная кросс-валидация + external validation	12	1062
Ji et al., 2024 [13]	Linear Regression, Linear Discriminant Analysis, eXtreme GB, CatBoost	ОБ, БРВ и рецидив	<b>Прогнозирование ОБ AUC:</b> LR: 0,75 (0,70–0,80); LDA: 0,76 (0,71–0,81); XGBoost- 0,71 (0,64–0,76); CatBoost- 0,75 (0,70–0,81). <b>Прогнозирование БРВ AUC:</b> LR- 0,71 (0,61–0,80); LDA- 0,71 (0,61–0,80); XGBoost- 0,69 (0,60–0,78); CatBoost- 0,75 (0,69–0,82) <b>Прогнозирование рецидива AUC:</b> LR- 0,80 (0,65–0,92); LDA- 0,83 (0,69–0,93); XGBoost- 0,81 (0,65–0,93); CatBoost- 0,82 (0,67–0,93) .	5-кратная кросс-валидация + bootstrap validation	12	1330
Cardoso et al., 2023 [14]	NB, RF, XGBoost	1-летняя, 3-х летняя, 5-летняя ОБ	<b>5-летняя ОБ AUC:</b> Naïve Bayes — 0,765; Random Forest — 0,844; XGBoost — 0,845.	Train-test split, ROC curve evaluation	10	29670
Kos et al., 2024 [15]	DT, SVM, K-NN, Ensemble Classifier, NN	ОБ	<b>AUC:</b> SVM — 0,84; DT — 0,81; Ensemble — 0,83; NN — 0,76; KNN — 0,76.	10-кратная кросс-валидация	20	498
Osman et al., 2022 [16]	Light GBM	5-летняя ОБ	<b>AUROC</b> External Validation- 0,805; Sensitivity — 68,14%; Positive Predictive Value- 49,88%.	18-кратная кросс-валидация	12	364316
Rodriguez et al., 2023 [17]	XGBoost, Lasso, Ridge, Elastic Net, Superlearner	Рецидив	<b>AUC</b> XGBoost — 0,87 (6 месяцев), 0,94 (54 месяца); Lasso, Ridge, Elastic Net: AUC 0,58–0,69	10-кратная кросс-валидация	18	448
Zhang W. et al., 2023 [42]	LR, KNN, SVM, DT, RF, balanced RF (BRF)	Метастазирование с легкие	BRF model AUC 0,874	Internal and external validation	18	48816
Kudo S.-E. et al., 2021 [19]	ANN	Метастазирование в лимфоузлы	AUC 0,83–0,84	Internal and external validation	8	4079
Wang, Xiaojie et al., 2022 [20]	LASSO, SVM, LR, XGB, LightGB, DT, RF, MLP	Метастазирование в лимфоузлы	MLP model AUC — 0,873, Sensitivity — 81,0%, Specificity — 82,5%	5-кратная кросс-валидация	10	2891
Chi, Shengqiang et al., 2019 [18]	Semi-Supervised LR, LR, SVM, RF, NN, NB	ОБ	AUC 0,73 (LR, NN)	5-кратная кросс-валидация	6	115217

Автор, год	Алгоритм	Прогнозируемый клинический исход	Результаты	Метод валидации	Число анализируемых параметров	Количество пациентов
Han T. et al., 2023 [21]	DT, LGBM, RF, KNN, SVM, Classification and Regression Trees, NB, Bootstrap aggregating	Метастазирование в печень	Stacking Bagging Model AUC 0,9631	10- кратная кросс-валидация	16	17111
Kasai S. et al., 2021 [22]	"Prediction One" (Sony Network Communications)	Метастазирование в лимфоузлы	AUC 0,903, accuracy 80,4%, sensitivity 90,0%, specificity 79,4%	hold-out validation	17	267
Masum S. et al., 2022 [23]	RF, KNN, SVM, MLP, Bidirectional Long Short-Term Memory (BI-LSTM)	Общая смертность, 31-, 91-дневная смертность	Accuracy 80–96%, Sensitivity 84–93%, Specificity 75–100%	10- кратная кросс-валидация	47	4336
Wozniacki A. et al., 2024 [24]	RF, XGBoost, CatBoost, LGBM, GB, Extra Trees, KNN, DT	1-, 3-, 5- летняя ОБ	<b>1-летний показатель</b> (LightGBM): Accuracy 0,8187, F1-score 0,7544, Точность 0,7904, recall 0,7356, AUC 0,855. <b>3-летний показатель</b> (Gradient Boosting): Accuracy 0,7861, F1-score 0,7811, precision 0,7803, recall 0,7821, AUC 0,865. <b>5-летний показатель</b> (CatBoost): accuracy 0,8185, F1-score 0,7615, precision 0,7716, recall 0,7534, AUC 0,885. <b>Общая смертность</b> (RF): accuracy 0,7889, F1-score 0,7746, precision 0,7870, recall 0,7689, AUC 0,861.	10- кратная кросс-валидация	58	72961
Oliveira T. et al., 2018 [25]	Комбинация: Bagging, AdaBoost, Bayesian Boosting, Stacking, Voting + k-NN, NB, DT, RF	1-, 2-, 3-, 4-, 5-летняя ОБ	<b>5-летняя ОБ</b> <b>Ободочная кишка:</b> Accuracy 96,45%, AUC 0,985; <b>Прямая кишка:</b> Accuracy 94,51%, AUC 0,971.	10- кратная кросс-валидация	61	27795
Qiu, Binxu et al., 2022 [26]	RF, LGBM, XGB, multilayer perceptron (MLP), LR, KNN	Метастазирование в печень	XGB: AUC — 0,926, accuracy — 0,919, sensitivity — 0,740, specificity — 0,765	10- кратная кросс-валидация	11	20882
Liu, Yuan et al., 2024 [27]	XGBoost, RF, SVM, KNN	Метастазирование в печень	XGBoost: AUC 0,922 (0,833–0,995), accuracy 0,908 (0,889–0,926), sensitivity 0,924 (0,873–0,975), specificity 0,883 (0,810–0,956)	10- кратная кросс-валидация	38	1151
Li X. et al., 2024 [28]	ggplot2 3.5.1, glmnet 4.1.8, xgboost 1.7.7.1, RF SRC 3.3.0, ggvenn 0.1.10, rms	1-, 3-, 5-летняя ОБ	<b>AUC</b> <b>1-летний показатель</b> 0,874 (95%ДИ 0,809–0,939), <b>3-летний показатель</b> 0,929 (95%ДИ 0,884–0,975), <b>5-летний показатель</b> 0,889 (95%ДИ 0,812–0,966); C-index 0,862 (95%ДИ 0,818–0,906)	Internal and external validation	17	356
Qiu, Binxu et al., 2023 [29]	XGB	Отдаленное метастазирование	AUC — 0,814; AUPRC — 0,609; accuracy — 0,800; precision — 0,810	10- кратная кросс-валидация	12	25045
Xing Y. et al., 2024 [30]	XGB, DT, RF, SVM	Метастазирование в печень	<b>AUC:</b> XGB 0,85; DT — 0,83; RF — 0,84; SVM — 0,71	holdout validation	12	51632
Zhao, B et al., 2020 [31]	Cox proportional hazard regression with lasso regression	3-летняя ОБ	C-index 0,761–0,794	internal and external validation	12	19364
Orafaie A. et al., 2024 [32]	RF, Recursive Feature Elimination, simulated annealing, SVM, DT, XGB	ОВ и БРВ	<b>Прогнозирование рецидива:</b> Sensitivity — 75%, specificity — 75,7%, AUC — 0,666; <b>Прогнозирование ОБ:</b> sensitivity — 80,5%, specificity — 64,8%, AUC — 0,630	K-fold cross-validation	8	200
Tian, Yu et al., 2018 [33]	Cox regression, AFT model, RF	ОВ	C-index 0,898 (0,895–0,902)	Кросс-валидация	9	128061
Wei R. et al., 2024 [34]	RF, LR, XGB, NN, KNN	Отдаленное метастазирование	RF AUC — 0,843	Кросс-валидация	15	200958

Автор, год	Алгоритм	Прогнозируемый клинический исход	Результаты	Метод валидации	Число анализируемых параметров	Количество пациентов
Li J. et al., 2023 [43]	ML models: SVM, KNN, DT, RF, extra trees. NLP models: bidirectional encoder representations from transformer (BERT)	Метастазирование в печень	Accuracy 80,8%; precision 80,3%; recall 80,5%; F1-score 80,8%	Внешняя-валидация	20	1463
Liu Y. et al., 2023 [35]	XGBoost, RF, SVM, KNN	Рецидив	XGBoost: AUC 0,952 KNN: Accuracy — 0,935, Specificity — 0,932 RF: Sensitivity — 0,959	10-кратная кросс-валидация	44	1187
Ishizaki, Tetsuo et al., 2023 [36]	NN, XGBoost	БПВ	AUC 0,775	5-кратная кросс-валидация	14	259
Bibault, Jean-Emmanuel et al., 2021 [37]	XGBoost	10-летняя ОВ	AUC 0,84, accuracy 0,83	Кратная кросс-валидация	29	2359
Amygdalos, I. et al., 2023 [38]	GBDT	ОВ	AUC 0,72; Accuracy 0,69; Sensitivity 0,73; Specificity 0,66	Кросс-валидация	24	487
Piao Z. et al., 2023 [39]	LightGBM	Метастазирование в лимфоузлы	AUC 0,960; specificity 85,8%; accuracy 92,9%; PPV 36,3%; NPV 100%; sensitivity 100%	5-кратная кросс-валидация	12	651
Yang, Xulin et al., 2023 [40]	RF, GBM, DeepSurv, DeepHit, Cox-Time, N-MTLR	5-летняя БПВ	DeepHit: C-index 0,789 RF: Brier score 0,096	5-кратная кросс-валидация	23	2157
Mohammadi G. et al., 2024 [41]	LR, NB, SVM, NN, DT, LGBM	Госпитальная смертность	<b>AUC</b> NB — 0,70; LGBM — 0,70. <b>Sensitivity</b> LR — 100%, <b>Specificity:</b> DT 91%, PPV: DT — 0,53, NPV: LGBM — 0,86, <b>Accuracy:</b> DT — 0,75	10-кратная кросс-валидация	23	1853

Примечание: ОВ — общая выживаемость; БПВ — безрецидивная выживаемость; RF — случайный лес; SVM — машина опорных векторов; LR — логистическая регрессия; NB — наивный байесовский классификатор; DT — дерево решений; KNN — метод k-ближайших соседей; XGBoost / GB / GBM / LightGBM / LGBM / CatBoost — градиентный бустинг; Bagging — бутстраповая агрегация; AdaBoost / AB — адаптивный бустинг; ET — экстремально рандомизированные деревья; SuperLearner — ансамбль SuperLearner; RFE — рекурсивное исключение признаков; SA — метод имитации отжига; ANN / NN — искусственная нейронная сеть; MLP — многослойный перцептрон; DNN — глубокая нейронная сеть; BI-LSTM — двунаправленная LSTM-сеть; DeepSurv — нейронная модель выживаемости DeepSurv; DeepHit — многособытийная нейронная модель DeepHit; Cox-Time — нейронная модель Cox-Time; N-MTLR — нейронная мультизадачная логистическая регрессия; LASSO — L1-регуляризованный отбор признаков (LASSO); EN — эластичная сетка; RSF — случайный лес выживаемости; SSLR — полу-супервизированная логистическая регрессия; AFT — модель ускоренного времени до события; SMOTE — алгоритм SMOTE (синтетическое увеличение малого класса); AUC / AUROC — площадь под ROC-кривой; AUPRC — площадь под PR-кривой; Brier score — ошибка Бриера; Precision / PPV — положительная прогностическая ценность; Recall / Sensitivity — чувствительность (полнота); Specificity — специфичность; Accuracy — общая точность; F1-score — среднее точности и чувствительности; C-index — индекс конкордантности; NPV — отрицательная прогностическая ценность; HR — отношение рисков; OR — отношение шансов; I<sup>2</sup> — статистика гетерогенности I-квадрат; τ<sup>2</sup> — межисследовательская дисперсия τ-квадрат; Q-test — тест Кокрана Q

и характер проведённого лечения. Представленные данные служат обзором исследовательского дизайна и методологических подходов, позволяя оценить масштабы выборок и обоснованность используемых моделей. В таблице 3 представлено распределение исследований по количеству изученных признаков. Из представленных данных следует, что большинство исследований (16 исследований, 37,2%) включали 10–15 признаков для проведения дальнейшего анализа. Показатели эффективности во включенных статьях представлены в таблице 4. Как представлено в таблице 4, наиболее часто применявшимся показателем эффективности явилась площадь под ROC-кривой (AUC), фигурировавшая в 37/43 (86%) исследованиях с диапазоном значений от 0,58 до 0,985. Точность (Accuracy) применялась в 20/43 (46,5%) работах значения варьировались от 47,4% до 96%. Показатель чувствительности

(Sensitivity) применялся в 17/43 (39,5%) исследованиях, специфичности (Specificity) в 16/43 (37,2%) исследованиях, значения чувствительности колебались в пределах 1,75–100%, а специфичности — 0–98%. Положительная прогностическая ценность (PPV) применялась в 4/43 (9,3%) исследованиях и находилась в диапазоне 49,88–85%, тогда как отрицательная прогностическая ценность (NPV) — в 3/43 (7%) с разбросом значений 70–100%. Показатель C-индекса фигурировал в 4/43 (9,3%) исследованиях и варьировал от 0,692 до 0,929. F1-score фигурировал в 2/43 (7%) исследований с диапазоном значений от 75% до 80,8%, Precision 4/43 (9,3%) — 77,16–90% и Recall 3/43 (7%) — 67–80,5%.

### Метаанализ

Результаты оценки площади под кривой (AUC) представлены на основе 12 исследований, в которых

**Таблица 3.** Характеристики и типы данных в рассматриваемых статьях (N = 43 исследований)

**Table 3.** Characteristics and types of data in the articles (N = 43)

Количество включенных в анализ признаков	Число исследований (N = 43)	Источник
< 10	8/43 (18,6%)	[3,6–8,18,19,32,33]
10–15	16/43 (39,2%)	[4,5,9,10,12–14,16,20,26,29–31,34,36,39]
16–20	9/43 (20,9%)	[1,2,15,17,21,22,28,42,43]
> 20	10/43 (23,3%)	[11,23–25,27,35,37,38,40,41]

**Таблица 4.** Показатели эффективности включенных в анализ исследований (N = 43 исследований)

**Table 4.** Performance metrics reported by the studies included in the analysis (N = 43)

Показатель	Исследования (N = 43)	Результаты (диапазон)	Источник
AUC	37/43 (86%)	0,58–0,985	[1–5,7–22,24–30,32,34–37,39,41–43]
Точность (Accuracy)	20/43 (46,5%)	47,4–96%	[1–7,9,22–27,29,35,37,39,41,43]
Чувствительность	17/43 (39,5%)	1,75–100%	[1–6,10,16,20,22,23,26,27,32,35,39,41]
Специфичность	16/43 (37,2%)	0–98%	[1–6,10,16,20,22,23,26,27,32,35,39,41]
Положительная прогностическая ценность (PPV)	4/43 (9,3%)	49,88–85%	[16,39,41,43]
Отрицательная предиктивная ценность (NPV)	3/43 (7%)	70–100%	[39,41,43]
C-index	4/43 (9,3%)	0,692–0,929	[28,31,33,40]
F1-score	3/43 (7%)	75–80,8%	[7,24,43]
Precision	4/43 (9,3%)	77,16–90%	[7,24,29,43]
Recall	3/43 (7%)	67%–80,5%	[7,24,43]

**Таблица 5.** Анализ гетерогенности исследований

**Table 4.** Analyses of research heterogeneity

Метрика	Значение	Интерпретация
$I^2$ <sup>1</sup>	97,6	Высокая гетерогенность
p-value $Q$ <sup>2</sup>	< 0,001	Значимые различия между исследованиями
$\tau^2$ <sup>3</sup>	0,289	Высокая межисследовательская вариабельность

Примечание: <sup>1</sup> $I^2$  — процент дисперсии, объясняемой межисследовательской гетерогенностью, <sup>2</sup>p-value  $Q$  — тест Кокрана, <sup>3</sup> $\tau^2$  — оценка межисследовательской дисперсии.

прогнозировалась общая выживаемость. В каждом из которых была выбрана модель с наилучшим значением AUC. На лесовидном графике (Forest plot) представлены оценки AUC и соответствующие 95% доверительные интервалы по каждому включённому исследованию (Рис. 2).

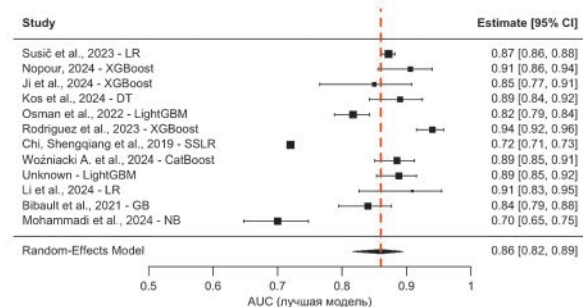
Наибольшее значение AUC было зафиксировано в исследовании Rodríguez et al. (2023) при применении алгоритма XGBoost (0,94) [17]. Наименьшее значение

AUC было представлено в исследовании Mohammadi et al. (2024) — NB (0,70) [41]. Обобщённое значение AUC составило 0,86 (95% ДИ: 0,82–0,89). Для количественной оценки вариабельности между включёнными исследованиями был проведён анализ гетерогенности (Табл. 5).

Анализ гетерогенности показал высокую степень вариабельности между включёнными исследованиями. Значение  $I^2$  — 97,6%, указывает на выраженную межисследовательскую гетерогенность. Критерий Кокрана  $Q = 669,03$ ,  $p < 0,001$ , что указывает на статистически значимую гетерогенность между исследованиями, результат  $\tau^2$  — 0,289 демонстрирует вариабельность истинных эффектов между исследованиями.

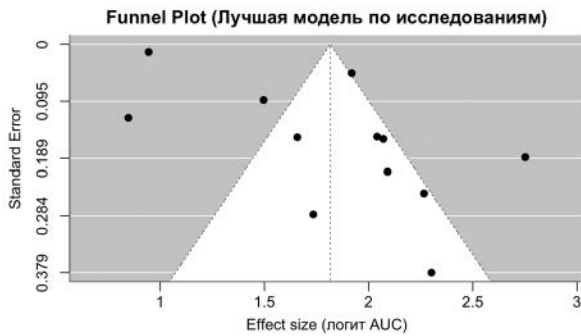
Оценка публикационного смещения представлена на воронкообразной диаграмме (funnel plot), отображающей взаимосвязь между логит-преобразованными значениями AUC и их стандартной ошибкой (Рис. 3).

На графике наблюдается умеренная асимметрия распределения точек в сторону положительных значений эффекта. Визуальная интерпретация проводится на основе симметричности распределения и требует



**Рисунок 2.** График Forest plot объединённых оценок AUC по включённым исследованиям

**Figure 2.** Graph Forest plot of combined estimates AUC for included studies



**Рисунок 3.** График Funnel plot  
**Figure 3.** Graph of Funnel plot

количественного подтверждения. Для оценки устойчивости результатов метаанализа был проведён анализ чувствительности с последовательным исключением каждого исследования (leave-one-out analysis), (Рис. 4).

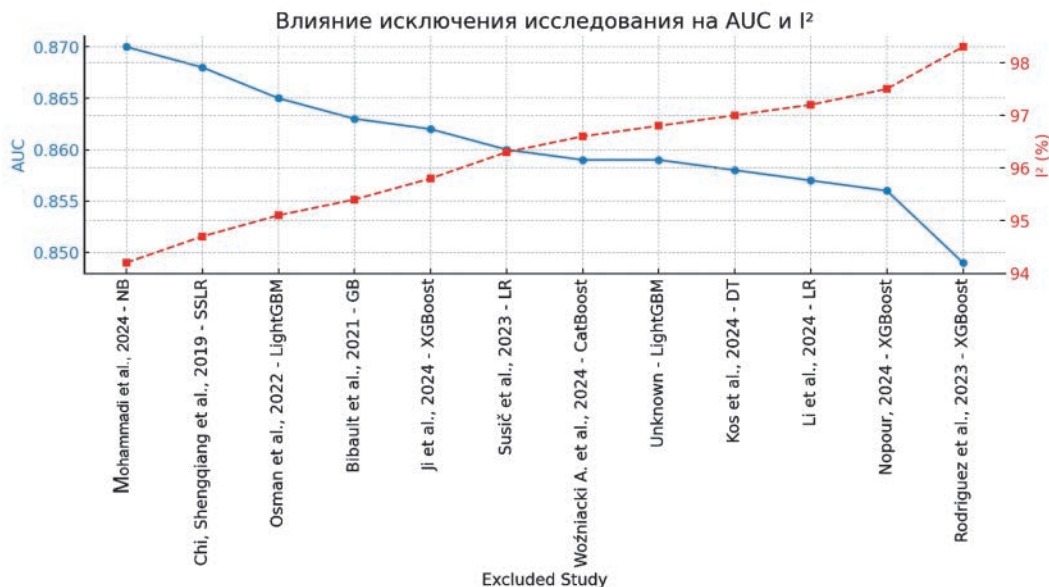
Удаление большинства исследований не привело к существенным изменениям итоговой AUC. Наибольшее снижение AUC (до 0,849) и снижение гетерогенности (до 94%) наблюдалось при исключении исследования Rodriguez et al., 2023 — XGBoost. Для оценки влияния типа алгоритма на вариабельность объединённых оценок был проведён модерационный анализ. В качестве модератора использовалась категориальная переменная — тип алгоритма. Результаты представлены в таблице 6.

Алгоритмы DT и LR продемонстрировали статистически значимое положительное влияние на итоговый

AUC ( $p < 0,001$ ), в то же время, алгоритмы NB и SSLR были связаны со статистически значимым снижением AUC по сравнению с остальными моделями ( $p = 0,034$  и  $p = 0,049$ , соответственно). Общий тест модераторов оказался статистически значимым (QM ( $df = 7$ ) = 15,87,  $p = 0,026$ ).

## ОБСУЖДЕНИЕ

Проведённый метаанализ подтвердил высокую прогностическую эффективность алгоритмов машинного обучения в задаче предсказания общей выживаемости у пациентов с колоректальным раком. Объединённое значение AUC составило 0,86 (95% ДИ: 0,82–0,89), что соответствует хорошей дискриминативной способности моделей. Однако высокая гетерогенность между исследованиями ( $I^2 = 97,6\%$ ) ограничивает обобщаемость полученных результатов. Вероятными источниками гетерогенности являются как методологические различия между исследованиями (подход к отбору признаков, численность выборок, методы валидации моделей), так и клиническая неоднородность включённых пациентов по стадиям заболевания и другим характеристикам. Таким образом, в условиях выраженной методологической и клинической гетерогенности любые прямые сопоставления эффективности алгоритмов могут быть методологически необоснованными.



**Рисунок 4.** Анализ чувствительности: влияние исключения отдельных исследований на объединённую AUC и гетерогенность ( $I^2$ ) при поочередном исключении отдельных публикаций. Синяя линия — объединённая оценка AUC. Красная пунктирная линия — уровень гетерогенности  $I^2$

**Figure 4.** Sensitivity analysis: the effect of exclusion of individual studies on combined AUC and heterogeneity ( $I^2$ ) when individual publications are excluded in turn. The blue line is the combined AUC score. The red dotted line is the heterogeneity level  $I^2$

**Таблица 6.** Модерационный анализ: влияние алгоритма ИИ на вариабельность AUC  
**Table 6.** Moderation analysis: the effect of the artificial intelligence algorithm on variability AUC

Алгоритм	Оценка (логит AUC)	Стандартная ошибка	ДИ (нижняя граница)	ДИ (верхняя граница)	p-value
DT	0,4097	0,4097	1,2377	2,8436	< 0,001
GB	0,5974	0,5974	-1,1207	1,2209	0,273
LightGBM	0,4581	0,5854	-1,2481	2,1643	0,289
LR	0,9487	0,5987	-0,1986	2,0961	0,118
NB	-1,5719	0,5719	-2,3143	-0,0724	0,034
SSLR	-0,9148	0,4585	-1,8125	-0,017	0,049
XGBoost	0,4847	0,4847	-0,7026	1,1974	0,609

Наиболее часто используемым алгоритмом оказался Random Forest, примененный в 65% включённых исследований. Этот метод характеризуется устойчивостью и высокой способностью к обнаружению нелинейных зависимостей между переменными. Высокие показатели эффективности алгоритма Random Forest были продемонстрированы в работах Jeon и соавт., 2023 г. и Chen и соавт., 2022 г., где он достиг значений AUC 0,84 и точности более 80% [1,5]. Наиболее высокие показатели прогностической точности среди рассмотренных исследований были получены при использовании методов градиентного бустинга (XGBoost). Например, в исследованиях Rodriguez и соавт., 2023 г., Erkan и соавт., 2024 г., и Gupta и соавт., 2019 г. алгоритм XGBoost демонстрировал более высокие значения AUC по сравнению с другими подходами, достигая 0,92–0,94, при чувствительности до 95% и специфичности до 94% [7,9,17]. Метод опорных векторов (SVM) также продемонстрировал высокую эффективность. Согласно исследованию Ting и соавт., 2020 г., данный алгоритм достиг AUC 0,87, точности 83% и специфичности 86% [3]. Аналогичные результаты представлены в работе Achilou и соавт., 2021 г., где SVM достиг AUC 0,80 и точности 87,8% [4]. Однако, несмотря на конкурентные показатели, модели SVM имеют ограниченную масштабируемость при увеличении размерности данных и требуют тщательной настройки гиперпараметров, что снижает их практическую применимость. В отличие от ансамблевых методов, таких как Random Forest и XGBoost, модели SVM менее устойчивы к шуму в данных и требуют значительных вычислительных ресурсов для обработки крупных выборок. Наименее стабильные результаты продемонстрировал алгоритм K-Nearest Neighbors (KNN), который существенно зависит от объёма и структуры данных. В исследовании Gupta и соавт., 2019 г., KNN достиг AUC 0,75 [9]. В то же время, точность KNN составила 75%, что уступает Random Forest (84%) и XGBoost (85%) [9]. Также в исследовании Kos и соавт., 2024 г., KNN показал AUC 0,76, что является худшим показателем среди всех рассмотренных методов, включая Support Vector Machine (0,84) и нейронные сети (0,76) [15].

Что касается объёма выборок, заслуживает внимания тот факт, что модели, разработанные на ограниченных данных, так же демонстрируют высокую прогностическую эффективность. В исследовании Leonard и соавт., 2022 г., где было проанализировано ограниченное количество признаков (менее 10 переменных), значение AUC варьировало в диапазоне 0,74–0,75 и точность 76–80%, что свидетельствует о возможности разработки прогностических моделей с высокой степенью предсказательной способности даже при ограниченном объёме исходных данных [11]. Однако в большинстве работ (33,3%) было проанализировано от 10 до 15 параметров, что позволило достичь оптимального баланса между сложностью модели и качеством прогноза.

Однако наблюдаемая эффективность того или иного алгоритма может варьировать под влиянием различных методологических и клинических факторов, таких как размер и репрезентативность исследуемых выборок, используемые подходы к валидации, качество исходных данных и количество включённых клинических параметров. Различия в характеристиках моделей могут отражать не столько внутренние преимущества определённого алгоритма, сколько влияние потенциальных источников систематической ошибки, включая публикационное смещение, неоднородность критериев включения пациентов, вариабельность процедур подготовки данных и неодинаковые стратегии подбора гиперпараметров моделей.

Отсутствие единых критериев оценки и неполное описание параметров обучения снижает воспроизводимость и ограничивает сопоставимость результатов. Включённые модели ИИ различались по числу предикторов (от 7 до 44), что могло так же повлиять на сложность и точность прогнозов. Значительные различия отмечены и в численности выборок (от 164 до 528 060 пациентов), источниках данных и стадиях заболевания. Одни исследования включали, преимущественно, пациентов на ранних стадиях, другие — более клинически разнообразных когорты. Данное предположение подтверждается результатами анализа чувствительности: исключение большинства работ не оказывало существенного влияния на итоговую

AUC, однако удаление одного из исследований привело к снижению объединённой оценки и уровня гетерогенности, что указывает на его значительный вклад в общую вариабельность. Это может быть связано с высоким качеством модели и оптимизированной настройкой параметров в данном исследовании. Дополнительным ограничением может быть потенциальное наличие публикационного смещения, оцененное с помощью funnel plot. Умеренная асимметрия графика указывает на возможную избирательность публикации исследований с положительными результатами, что также может искажать агрегированные оценки.

## ЗАКЛЮЧЕНИЕ

Современные исследования всё чаще подчёркивают необходимость интеграции алгоритмов машинного обучения в клиническую практику, что способствует повышению точности прогнозирования рецидива колоректального рака и разработке более эффективных персонализированных стратегий лечения. Для повышения воспроизводимости и клинической применимости моделей необходима стандартизация подходов к разработке, валидации и отчётности моделей машинного обучения.

### УЧАСТИЕ АВТОРОВ

Концепция и дизайн исследования: Гордеев С.С.

Сбор и обработка материалов: Абдулаева Р.Ш., Беленькая Я.В., Манукян М.Ш., Геворкян Т.Г., Павлова В.И.

Статистическая обработка: Абдулаева Р.Ш., Гордеев С.С.

Написание текста: Абдулаева Р.Ш., Беленькая Я.В.

Редактирование: Гордеев С.С.

### AUTHORS CONTRIBUTION

Concept and design of the study: Sergey S. Gordeev

Collection and processing of material: Rukiyat Sh. Abdulaeva, Yana V. Belenkaya, Mariam Sh. Manukyan, Tigran G. Gevorkyan, Valeria I. Pavlova.

Statistical processing: Rukiyat Sh. Abdulaeva, Sergey S. Gordeev

Text writing: Rukiyat Sh. Abdulaeva, Yana V. Belenkaya

Editing: Sergey S. Gordeev

### СВЕДЕНИЯ ОБ АВТОРАХ (ORCID)

Абдулаева Рукият Шамильевна — ORCID: 0009-0004-6399-963X

Павлова Валерия Игоревна — ORCID: 0000-0002-0899-0809

Геворкян Тигран Гагикович — ORCID: 0009-0008-3486-302X

Беленькая Яна Владимировна — ORCID: 0000-0003-2163-1752

Манукян Мариам Ширавовна — ORCID: 0000-0002-5084-4872

Гордеев Сергей Сергеевич — ORCID: 0000-0002-9303-8379

### INFORMATION ABOUT THE AUTHORS (ORCID)

Abdulaeva Sh. Rukiyat — 0009-0004-6399-963X

Valeria I. Pavlova — 0000-0002-0899-0809

Tigran G. Gevorkyan — 0009-0008-3486-302X

Yana V. Belenkaya — 0000-0003-2163-1752

Mariam Sh. Manukyan — 0000-0002-5084-4872

Sergey S. Gordeev — 0000-0002-9303-8379

## ЛИТЕРАТУРА/REFERENCES

1. Jeon Y, et al. Machine learning based prediction of recurrence after curative resection for rectal cancer. *PLoS One*. 2023;18(12):e0290141.
2. Tang M, et al. Machine learning based prognostic model of Chinese medicine affecting the recurrence and metastasis of I-III stage colorectal cancer: A retrospective study in China. *Front Oncol*. 2022;12:1044344.
3. Ting WC, et al. Machine Learning in Prediction of Second Primary Cancer and Recurrence in Colorectal Cancer. *Int J Med Sci*. 2020;17(3):280–291.
4. Achilonu OJ, et al. Predicting Colorectal Cancer Recurrence and Patient Survival Using Supervised Machine Learning Approach: A South African Population-Based Study. *Front Public Health*. 2021;9:694306.
5. Chen PC, et al. A Prediction Model for Tumor Recurrence in Stage II-III Colorectal Cancer Patients: From a Machine Learning Model to Genomic Profiling. *Biomedicines*. 2022;10(2):340.
6. Alinia S, et al. Predicting mortality and recurrence in colorectal cancer: Comparative assessment of predictive models. *Heliyon*.

2024;10(6):e27854.

7. Kayikcioglu E, et al. Machine learning for predicting colon cancer recurrence. *Surgical Oncology*. 2024;54:102079.

8. Susič D, et al. Artificial intelligence based personalized predictive survival among colorectal cancer patients. *Comput Methods Programs Biomed*. 2023;231:107435.

9. Gupta P, et al. Prediction of Colon Cancer Stages and Survival Period with Machine Learning Approach. *Cancers (Basel)*. 2019; 11(12).

10. Hu J, et al. Construction and validation of a progression prediction model for locally advanced rectal cancer patients received neoadjuvant chemoradiotherapy followed by total mesorectal excision based on machine learning. *Front Oncol*. 2023;13:1231508.

11. Leonard G, et al. Machine Learning Improves Prediction Over Logistic Regression on Resected Colon Cancer Patients. *J Surg Res*. 2022;275:181–193.

12. Nopour R. Development of Prediction Model for 5-year Survival of Colorectal Cancer. *Cancer Inform*. 2024;23:11769351241275889.

13. Ji XL, et al. Prognostic prediction models for postoperative

- patients with stage I to III colorectal cancer based on machine learning. *World J Gastrointest Oncol.* 2024;16(12):4597–4613.
14. Buk Cardoso L, et al. Machine learning for predicting survival of colorectal cancer patients. *Sci Rep.* 2023;13(1):8874.
  15. Kos FT, et al. Comparison of Different Machine Learning Models for Predicting Long-Term Overall Survival in Non-metastatic Colorectal Cancers. *Cureus.* 2024;16(12):e75713.
  16. Osman MH, et al. Machine Learning Model for Predicting Postoperative Survival of Patients with Colorectal Cancer. *Cancer Res Treat.* 2022;54(2):517–524.
  17. Rodriguez PJ, et al. Using Machine Learning to Leverage Biomarker Change and Predict Colorectal Cancer Recurrence. *JCO Clin Cancer Inform.* 2023;7:e2300066.
  18. Chi S, et al. Semi-supervised learning to improve generalizability of risk prediction models. *J Biomed Inform.* 2019;92:103117.
  19. Kudo SE, et al. Artificial Intelligence System to Determine Risk of T1 Colorectal Cancer Metastasis to Lymph Node. *Gastroenterology.* 2021;160(4):1075–1084.e2.
  20. Wang X, et al. Development and validation of artificial intelligence models for preoperative prediction of inferior mesenteric artery lymph nodes metastasis in left colon and rectal cancer. *Eur J Surg Oncol.* 2022;48(12):2475–2486.
  21. Han T, et al. Application of artificial intelligence in a real-world research for predicting the risk of liver metastasis in T1 colorectal cancer. *Cancer Cell Int.* 2022;22(1):28.
  22. Kasai, S, et al. The Effectiveness of Machine Learning in Predicting Lateral Lymph Node Metastasis From Lower Rectal Cancer: A Single Center Development and Validation Study. *Ann Gastroenterol Surg.* 2022;6(1):92–100.
  23. Masum S, et al. Data analytics and artificial intelligence in predicting length of stay, readmission, and mortality: a population-based study of surgical management of colorectal cancer. *Discov Oncol.* 2022;13(1):11.
  24. Woźniack, A, Książek W, Mrowczyk P. A Novel Approach for Predicting the Survival of Colorectal Cancer Patients Using Machine Learning Techniques and Advanced Parameter Optimization Methods. *Cancers (Basel).* 2024. 16(18).
  25. Oliveira T, et al. Survivability Prediction of Colorectal Cancer Patients: A System with Evolving Features for Continuous Improvement. *Sensors (Basel).* 2018; 18(9).
  26. Qiu, B, et al. Application of machine learning techniques in real-world research to predict the risk of liver metastasis in rectal cancer. *Front Oncol.* 2022;12:1065468.
  27. Liu Y, et al. Utilizing machine learning algorithms for predicting risk factors for bone metastasis from right-sided colon carcinoma after complete mesocolic excision: a 10-year retrospective multicenter study. *Discov Oncol.* 2024;15(1):463.
  28. Li X, et al. Preoperative Albumin to Alkaline Phosphatase Ratio and Inflammatory Burden Index for Rectal Cancer Prognostic Nomogram-Construction: Based on Multiple Machine Learning. *J Inflamm Res.* 2024;17:11161–11174.
  29. Qiu B, et al. A machine learning-based model for predicting distant metastasis in patients with rectal cancer. *Front Oncol.* 2023;13:1235121.
  30. Xing Y, et al. Development of prediction models for liver metastasis in colorectal cancer based on machine learning: a population-level study. *Transl Cancer Res.* 2024;13(11):5943–5952.
  31. Zhao B, et al. Using machine learning to construct nomograms for patients with metastatic colon cancer. *Colorectal Dis.* 2020;22(8):914–922.
  32. Orafaie A, et al. The association of preoperative hematologic parameters with short-term clinical outcomes in rectal cancer: A feature importance analysis. *Cancer Med.* 2024;13(10):e7225.
  33. Tian Y, et al. Spatially varying effects of predictors for the survival prediction of nonmetastatic colorectal Cancer. *BMC Cancer.* 2018;18(1):1084.
  34. Wei R, et al. Construction and validation of machine learning models for predicting distant metastases in newly diagnosed colorectal cancer patients: A large-scale and real-world cohort study. *Cancer Med.* 2024;13(5):e6971.
  35. Liu Y, et al. Identification of high-risk factors for recurrence of colon cancer following complete mesocolic excision: An 8-year retrospective study. *PLoS One.* 2023;18(8):e0289621.
  36. Ishizaki T, et al. Predictive modelling for high-risk stage II colon cancer using auto-artificial intelligence. *Tech Coloproctol.* 2023;27(3):183–188.
  37. Bibault JE, Chang DT, Xing L. Development and validation of a model to predict survival in colorectal cancer using a gradient-boosted machine. *Gut.* 2021;70(5):884–889.
  38. Amygdalos I, et al. Novel machine learning algorithm can identify patients at risk of poor overall survival following curative resection for colorectal liver metastases. *J Hepatobiliary Pancreat Sci.* 2023;30(5):602–614.
  39. Piao ZH, Ge R, Lu L. An artificial intelligence prediction model outperforms conventional guidelines in predicting lymph node metastasis of T1 colorectal cancer. *Front Oncol.* 2023;13:1229998.
  40. Yang X, et al. Predicting Colorectal Cancer Survival Using Time-to-Event Machine Learning: Retrospective Cohort Study. *J Med Internet Res.* 2023;25:e44417.
  41. Mohammadi G, et al. Classification and Diagnostic Prediction of Colorectal Cancer Mortality Based on Machine Learning Algorithms: A Multicenter National Study. *Asian Pac J Cancer Prev.* 2024;25(1):333–342.
  42. Zhang W, et al. Development and validation of an artificial intelligence prediction model and a survival risk stratification for lung metastasis in colorectal cancer from highly imbalanced data: A multicenter retrospective study. *Eur J Surg Oncol.* 2023;49(12):107107.
  43. Li J, et al. An interpretable deep learning framework for predicting liver metastases in postoperative colorectal cancer patients using natural language processing and clinical data integration. *Cancer Med.* 2023;12(18):19337–19351.