

ПРОГНОСТИЧЕСКАЯ РОЛЬ МАГНИТНО-РЕЗОНАНСНОЙ ТОМОГРАФИИ НА ОСНОВЕ ПАРАМЕТРИЧЕСКИХ КАРТ ИЗМЕРЯЕМОГО КОЭФФИЦИЕНТА ДИФФУЗИИ И T2-ВЗВЕШЕННЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ВОЗМОЖНОСТЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ДИАГНОСТИКЕ РАКА ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ С ПАТТЕРНОМ ГЛИСОНА 7 (3 + 4 И 4 + 3): СИСТЕМАТИЧЕСКИЙ ОБЗОР И МЕТААНАЛИЗ

О.В. Крючкова^{1*}, А.И. Кузнецов², Е.В. Щепкина^{3,4}, С.В. Епифанова^{1,4},
А.Э. Талышинский⁵, Е.В. Заря¹, Н.А. Рубцова⁶, Б.Я. Алексеев⁶

¹ ФГБУ «Центральная клиническая больница с поликлиникой» Управления делами Президента РФ, Москва

² Московский авиационный институт (национальный исследовательский университет), Москва

³ Российская академия народного хозяйства и государственной службы при Президенте РФ, Москва

⁴ ГБУЗ «Научно-практический клинический центр диагностики и телемедицинских технологий» ДЗМ, Москва

⁵ ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург

⁶ ФГБУ «НМИЦ радиологии» Минздрава России, Москва

A PREDICTIVE ROLE OF PI-RADSV2 AND PARAMETRIC MAP ADC PARAMETERS IN DIAGNOSING PROSTATE CANCER WITH GLEASON PATTERNS 7 (3 + 4 AND 4 + 3): A SYSTEMATIC REVIEW AND META-ANALYSIS

O.V. Kryuchkova^{1*}, A.I. Kuznetsov², E.V. Schepkina^{3,4}, S.V. Epifanova^{1,4},
A.E. Talyshinskiy⁵, E.V. Zarya¹, N.A. Rubtsova⁶, B.Ya. Alekseev⁶

¹ Central Clinical Hospital of of Department of Presidential Affairs, Moscow, Russia

² Moscow Aviation Institute (National Research University), Moscow, Russia

³ Russian Presidential Academy of National Economy and Public Administration – RANEPA, Moscow, Russia

⁴ Research and Practical Clinical Center for Diagnostics and Telemedical Technologies, Moscow, Russia

⁵ Saint Petersburg State University, Saint Petersburg, Russia

⁶ National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, Moscow, Russia

*E-mail: ovk16@bk.ru

Аннотация

Цель настоящего систематического обзора литературы и метаанализа – оценка качества прогностических моделей, построенных на основе радиомических признаков, полученных при бипараметрической МР-томографии для разделения пациентов с раком предстательной железы (РПЖ) на группы с паттерном Глисона 7: $GP \leq 3 + 4$ и $GP \geq 4 + 3$ при первичном обращении. **Материалы и методы.** Систематический поиск публикаций в период с января 2015 г. по декабрь 2024 г. в электронных базах данных eLIBRARY, PubMed, Google Scholar, Web of Science и ResearchGate проводили с использованием установленных методов в соответствии с протоколом PRISMA. Два автора независимо оценили исследования на предмет включения и исключения. **Результаты.** Метаанализ включал пять исследований. Общее число пациентов – 291, из них 41% с РПЖ $GP \geq 4 + 3$ и 59% – с $GP \leq 3 + 4$. Возраст пациентов в среднем составил 67.3 года (от 45 до 84 лет). Исследования были проведены на базе T2-взвешенных изображений и параметрических карт измеряемого коэффициента диффузии. Анализ показал, что наиболее часто исследователи используют следующие алгоритмы машинного обучения: SVM (Support Vector Machine) – 47%, AdaBoost (Adaptive Boosting) – 27%, MLR (Multiple Logistic Regression) – 20% и RF (Random Forest) – 7%. По данным проведенного метаанализа оценки ROC-AUC в 15 прогностических моделях, описанных в изученных нами публикациях с использованием метода случайных эффектов, было получено итоговое значение $ROC - AUC = 0.837$ [95%-ный доверительный интервал (95% ДИ) 0.786–0.887], $I^2 = 91.78\%$, $p < 0.001$. Наиболее точными оказались модели, построенные по принципу «черный ящик» ($ROC - AUC = 0.847$ [95% ДИ 0.794–0.899]), чем «белый ящик» ($ROC - AUC = 0.764$ [95% ДИ 0.655–0.872]). **Заключение.** Результаты исследования оказались многообещающими, но клиническое применение по-прежнему нуждается в более тщательной проверке экспертами в медицинских учреждениях и оценке эффективности в проспективных исследованиях.

Ключевые слова: машинное обучение, магнитно-резонансная томография, рак предстательной железы, диагностические методы и процедуры.

Abstract

Purpose. To assess the quality of prognostic models based on radiomic features obtained at biparametric MR imaging for stratifying patients with prostate cancer (PCa) into groups with Gleason pattern 7: $GP \leq 3 + 4$ and $GP \geq 4 + 3$ at initial examination. **Materials and methods.** A systematic search of publications in the electronic databases eLIBRARY, PubMed, Google Scholar, Web of Science and ResearchGate was performed from January 2015 to December 2024, using the established methods following PRISMA protocol. Two authors independently assessed the found publications for inclusion and exclusion. **Results.** Meta-analysis included five trials. The total number of patients was 291, of which 41% was with PCa $GP \geq 4 + 3$ and 59% – with $GP \leq 3 + 4$. The average age of patients was 67.3 (range 45–84 y.o.). The authors analyzed findings of T2-weighted images and parametric maps of the measured diffusion coefficient.

The performed analysis showed that the following machine learning algorithms are most often used by researchers: SVM (Support Vector Machine) – 47%, AdaBoost (Adaptive Boosting) – 27%, MLR (Multiple Logistic Regression) – 20% and RF (Random Forest) – 7%. Due to the performed meta-analysis and due to the assessment of ROC – AUC values in 15 prognostic models described in the studied publications, in which the random effect method was used, the authors have found the final value of ROC – AUC = 0.837 [95% confidence interval (95% CI 0.786–0.887), I² = 91.78%, $p < 0.001$]. The most accurate models were those constructed by the “black box” model principle (ROC – AUC = 0.847 [95% CI 0.794–0.899]) than by the “white box” (ROC – AUC = 0.764 [95% CI 0.655–0.872]). **Conclusion.** The results of the study were promising, but its clinical application still needs more rigorous validation by experts in medical institutions and evaluation of its effectiveness in prospective trials.

Keywords: machine learning, magnetic resonance imaging, prostate cancer, diagnostic methods and procedures.

Ссылка для цитирования: Крючкова О.В., Кузнецов А.И., Щепкина Е.В., Епифанова С.В., Талышинский А.Э., Заря Е.В., Рубцова Н.А., Алексеев Б.Я. Прогностическая роль магнитно-резонансной томографии на основе параметрических карт измеряемого коэффициента диффузии и T2-взвешенных изображений с использованием возможностей искусственного интеллекта в диагностике рака предстательной железы с паттерном Глисона 7 (3 + 4 и 4 + 3): систематический обзор и метаанализ. Кремлевская медицина. Клинический вестник. 2025; 2: 108–114.

Рак предстательной железы (РПЖ) является одним из самых распространенных злокачественных новообразований у мужчин, с пятилетней выживаемостью от 100% для локализованных форм заболевания до 29% для распространенных метастатических вариантов [1]. Индекс Глисона (GS), введенный в 1974 г. [2] для дифференцировки степени опухоли при акцинарной аденокарциноме предстательной железы, служит важным предиктором исхода, однако для РПЖ с GS = 7 частота рецидивов варьирует в зависимости от первичного паттерна Глисона: пятилетняя биохимическая выживаемость без рецидива для GP 3 + 4 составила 81–88% и 60–63% для GP 4 + 3 [3, 4]. Также сообщалось о том, что смертность при GP 4 + 3 РПЖ в три раза выше, чем при GP 3 + 4 [5] в связи с чем РПЖ с GP 4 + 3 принято называть РПЖ с высокой степенью злокачественности. «Золотым стандартом» оценки степени риска по шкале Глисона является биопсия [6]. Гистологические данные (GP) получают по результатам биопсии под контролем трансректального ультразвукового исследования (TRUS-biopsy). В исследовании C. Nan и соавт. [7] была проведена оценка качества применения TRUS-biopsy для разделения групп пациентов с GP 3 + 4 и GP 4 + 3. Выявлено, что этот метод эффективен с точностью ROC – AUC = 0.793 [0.659; 0.892]. Также известно, что результаты GP, полученные с помощью TRUS-biopsy, могут отличаться от результатов, полученных после радикальной простатэктомии, точность составляет 85.7% для опухолей с GS = 7. Таким образом, метод TRUS-biopsy для диагностики РПЖ является неидеальной инвазивной процедурой и сопряжен с дискомфортом и риском осложнений для пациента.

В 2014 г. была предложена новая классификация, разделяющая РПЖ с GS = 7 на градирующие группы второй и третьей степени злокачественности.

Нами было найдено два исследования, в которых авторы попытались сопоставить МР-разделение по группам PI-RADS-3 и PI-RADS-4 по системе PI-RADS v2 с гистологическим результатом по Глисон GP 3 + 4 и GP 4 + 3. В исследовании S.G. Gugliandolo и соавт. [8] исследовали 65 пациентов, из них семеро с GP 4 + 3, и установили, что PI-RADS v2 справляется с задачей с точностью ROC – AUC = 0.9 [0.79; 1.00]. В исследовании C. Nan и соавт. [9] на основе обследования 176 пациентов, из них 82 с GP 4 + 3, было установлено, что ROC – AUC = 0.68 [0.538; 0.801]. Это свидетельствует о неоднозначной и, скорее, низкой степени корреляции МР-признаков системы PI-RADS v2 и данных гистологии.

В последние десятилетия технологии медицинской визуализации значительно продвинулись благодаря радиомике, которая позволяет извлекать количественную информацию из медицинских изображений и демонстрирует корреляцию между гетерогенностью гистологии опухоли и ее радиомической структурой. Радиомика довольно хорошо справляется с задачей выявления кзРПЖ

(GS ≥ 7): ROC – AUC = 0.863 с помощью мультипараметрической магнитно-резонансной томографии мпМРТ и ROC – AUC = 0.853 с помощью бипараметрической МР-томографии (бпМРТ) [10]. Причем результаты многих исследований доказывают, что эти показатели не сильно зависят от опыта врачей: ROC – AUC = 0.767 для менее опытных врачей и ROC-AUC = 0.813 для более опытных врачей [11]. В другом исследовании было показано, что когда оценка PI-RADS v2 ≥ 4 считалась положительной, то согласие врачей было отличным как для поражений периферической зоны ($\kappa = 0.91$), так и для поражений переходной зоны ($\kappa = 0.87$).

Показано, мпМРТ и бпМРТ имеют сопоставимую диагностическую значимость с чувствительностью 94,6 и 93,9% соответственно и специфичностью 84,8% в сравнении с 87,3% [12]. С учетом этих данных предпочтительным методом является бпМРТ, которая не нуждается в динамическом контрастном усилении, что значительно сокращает время обследования и может стать более безопасной и экономически выгодной альтернативой мпМРТ для диагностики рака предстательной железы.

Цель исследования – оценка качества прогностических моделей, построенных на основе радиомических признаков бпМРТ, для разделения РПЖ на GP < 3 + 4 и GP > 4 + 3 при первичном обращении.

Материалы и методы

Протокол систематического обзора не был зарегистрирован ни в одной базе данных до начала данного исследования.

Процесс поиска был осуществлен в соответствии с протоколом, предназначенным для систематических обзоров и метаанализов (PRISMA). Литература была исследована в пяти электронных базах данных: eLIBRARY, PubMed, Google Scholar, Web of Science и ResearchGate с целью выявления статей, опубликованных в период с 2015 по 2024 г. Завершающий поиск был проведен 27 декабря 2024 г.

Поисковый запрос состоял из четырех основных компонентов. Первый компонент касался рака предстательной железы и включал два ключевых слова: «простата» и «рак простаты». Второй компонент был связан с использованием МРТ и содержал также два ключевых слова: «МРТ» и «магнитный». Третий компонент охватывал радиомику и включал три ключевых слова: «радиомика», «машинное обучение» и «модель». Четвертый компонент уточнял GP с ключевым словом «4 + 3». Стратегия поиска была сформулирована как #1 AND #2 AND #3 AND #4. Более подробная информация представлена в табл. 1.

Критерии включения/исключения

В соответствии с поставленной целью поиска в исследование были отобраны оригинальные статьи и полные тексты на русском или английском языке, опубликованных в 2019–2024 гг. Монографии, обзоры, метаанализы, описания отдельных случаев,

Таблица 1

Ключевые слова для поиска в электронных базах данных: eLIBRARY, PubMed, Google Scholar, Web of Science и ResearchGate

Номер	Русский язык	Английский язык
1	«Простата» или «рак простаты»	"Prostate" or "prostate cancer"
2	«МРТ» или «магнитный»	"MRI" or "magnetic"
3	«Радиомика», или «машинное обучение», или «модель»	"Radiomic" or "machine learning" or "model"
4	4 + 3	4 + 3

Таблица 2

Критерии включения и исключения статей из исследования

Критерии включения	Критерии исключения
Оригинальная статья	Монография
Наличие полного текста статьи	Обзор, систематический обзор или метаанализ
Статья на русском или английском языке	Тезисы на конференции
Публикация в 2019–2024 гг.	Описание отдельных случаев
–	Исследования на животных
–	Исследования эффективности диагностики

мнения экспертов, а также исследования, результаты которых были опубликованы только в виде тезисов, а также исследования на животных и работы, посвященные оценке эффективности диагностики, были исключены из анализа. Кроме того, из процесса скрининга были исключены статьи, в которых недостаточно внимания уделялось применению алгоритмов машинного обучения (МО), раку предстательной железы и результатам моделирования. Важно отметить, что количество пациентов, включенных в исследование, не было решающим критерием для отбора. Все критерии включения и исключения представлены в табл. 2.

Извлечение и синтез данных исследований

В ходе первичного отбора с использованием описанных выше поисковых запросов было выявлено 32 публикации. Собранные данные были организованы в базе данных (Microsoft Excel 365). Два исследователя проверили названия и аннотации потенциально подходящих статей на соответствие критериям поиска, а возникающие разногласия обсуждались с третьим рецензентом для достижения согласия.

После этого исследователи проанализировали полные тексты статей. Для каждого исследования были зафиксированы следующие данные: имя первого автора, год публикации, количество случаев, возраст участников, тип анализа, а также количество положительных и отрицательных случаев. Также была собрана информация о радиомике и методах МО, используемых в МРТ-исследованиях, деталях сегментации поражений, этапах предварительной обработки данных, количестве извлеченных признаков, процессе выбора признаков, моделях прогнозирования и стратегиях валидации. Минимальный размер выборки пациентов для включения в систематический обзор не был установлен.

Исследование включалось в анализ при выполнении всех следующих условий:

- 1) представлен анализ, направленный на разделение пациентов с раком предстательной железы GP 3 + 4 и GP 4 + 3 и основанный на данных биопсии или гистологии после радикальной простатэктомии;
- 2) предоставлена информация о последовательностях, использованных в МРТ-исследовании; при этом последовательность динамического контрастного усиления не входила в окончательную модель;
- 3) описаны алгоритмы МО, примененные для разделения групп пациентов с РПЖ GP 3 + 4 и GP 4 + 3;
- 4) указаны данные о площади под кривой ROC (ROC – AUC), а также об общем числе пациентов, количестве положительных

и отрицательных результатов, где GP 4 + 3 считали положительным, а GP 3 + 4 – отрицательным.

Исследования, в которых не применяли методы МО или которые содержали недостаточные данные, не были включены в метаанализ.

Статистический анализ

Точность прогнозирования (выявление РПЖ GP 3 + 4 в противовес GP 4 + 3) была количественно оценена с использованием ROC – AUC [13]. Для каждого из включенных исследований ROC – AUC была получена с соответствующими 95%-ми доверительными интервалами (95% ДИ). Стандартную ошибку (SE; standard error) для ROC – AUC рассчитывали на основе формул, представленных в Cochrane Handbook for Systematic Reviews (handbook-5-1. cochrane.org) [14]. Результаты метаанализа представлены в виде блочнограммы (forestplot).

Значение I², полученное на основании метода Хиггинса – Томпсона (Higgins – Thompson), использовали для оценки статистической гетерогенности, которая обеспечивала оценку процента вариабельности включенных в метаанализ исследований. Значения I² 0–25%, 26–50%, 51–75% и > 75% представляют собой очень низкую, низкую, среднюю и высокую гетерогенность соответственно. Статистика I² описывает процент вариаций между исследованиями, обусловленный скорее гетерогенностью, чем случайностью [15]. Анализ подгрупп проводили при наличии высокой гетерогенности.

Вес каждого исследования рассчитывали с помощью метода обратной дисперсии, в котором вес, присвоенный каждому исследованию, выбирается равным обратному значению дисперсии оценки эффекта, что минимизирует неопределенность оценки объединенного эффекта [14]. Было выполнено объединение исследований и оценена величина эффекта с использованием модели случайных эффектов, которая позволяет оценить распределение истинных эффектов между исследованиями с учетом гетерогенности [16].

Различия считали статистически значимыми при p < 0.05. Метаанализ данных проводили с использованием свободного программного обеспечения OpenMeta[Analyst] (BrownUniversity, Providence, RhodeIsland, USA).

Результаты

Поиск литературы и извлечение данных

Первоначальный поиск выявил 276 потенциально релевантных исследований. После удаления 127 дубликатов исследователи

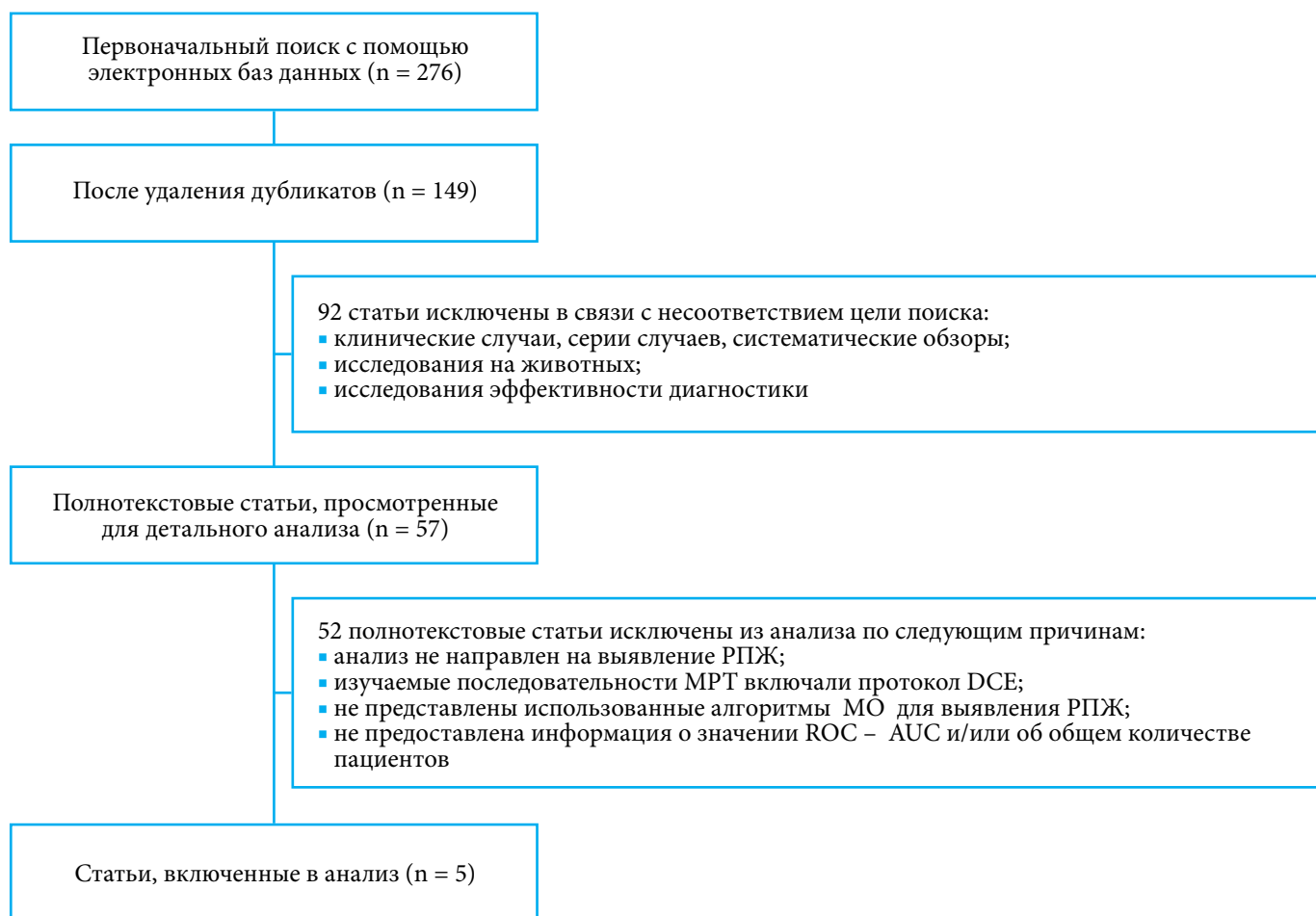


Рис. 1. Блок-схема процесса обзора публикаций и отбора исследований

Таблица 3

Исходная характеристика включенных исследований

Номер	Автор	Год	Название журнала	Страна	Количество	GP ≤ 3 + 4 (n)	GP ≥ 4 + 3 (n)	Возраст	Протоколы МРТ	Алгоритм МО
1	Fehr [17]	2015	Proceedings of the National Academy of Sciences	США	140	114	26	–	ADC	SVM, AdaBoost
2	Chaddad [18]	2018	Frontiers in Oncology	Канада	99	69	30	–	T2WI, ADC	RF
3	Maas [19]	2019	Investigative Radiology	Великобритания	50	20	30	62 (46–73)	ADC	MLR
4	Gugliandolo [8]	2020	Magnetic Resonance Imaging	Италия	65	58	7	74 (45–84)	T2WI	MLR
5	Han [9]	2021	Magnetic Resonance Imaging	Китай	176	94	82	66 (62–71)	ADC	SVM, MLR

проанализировали 149 публикаций. В результате оценки названий и аннотаций 92 статьи были исключены, так как не соответствовали целям поиска. В итоге 57 полнотекстовых статей были оценены каждым исследователем на предмет их соответствия критериям. После дополнительной обработки было исключено еще 52 статьи, и в результате осталось пять статей. Полный алгоритм поиска литературы представлен на рис. 1.

Исходные характеристики включенных исследований показаны в табл. 3.

В итоге в систематический обзор и метаанализ было включено пять исследований [8, 9, 17–19]. Количество пациентов в исследованиях варьировало от 50 до 176. Таким образом, всего в исследовании участвовал 291 пациент, из них 41% с РПЖ GP ≥ 4 + 3

и 59% с GP ≤ 3 + 4. Возраст пациентов в среднем составил 67,3 года (от 45 до 84 лет). Два исследования были проведены по протоколу T2-взвешенных изображений и пять – с использованием параметрических карт измеряемого коэффициента диффузии.

В двух исследованиях (40%) была проведена процедура выбора наиболее значимых переменных [9, 17]. Для этого использовали три метода: RFE (Recursive Feature Elimination), ANOVA (Analysis of Variance) и тест Краскела – Уоллиса (Kruskal – Wallis test).

В одном исследовании (20%) была выделена тестовая выборка, на которой проводили тестирование разработанной прогностической модели [9]. Тестовая выборка составила 30%. В четырех других исследованиях (80%) применяли кросс-валидацию для проверки созданной модели [8, 9, 17–19].

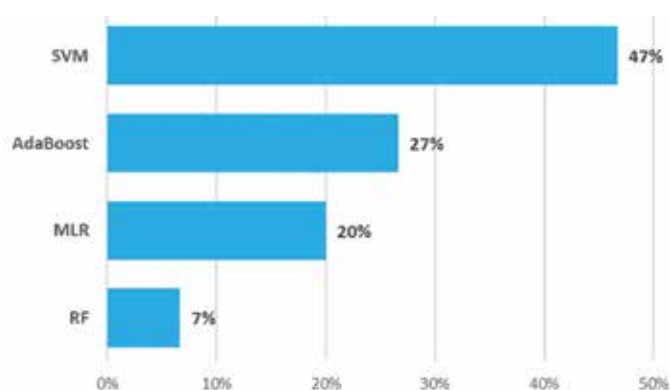


Рис. 2. Доля использованных алгоритмов МО в исследованиях

В одном исследовании (20%) использовали тест Де Лонга (DeLong test) для оценки различий в ROC – AUC между построенными моделями [9].

Анализ показал, что наиболее часто в своих работах авторы применяли следующие алгоритмы МО: SVM (Support Vector Machine) – 47%, AdaBoost – 27%, MLR (Multiple Logistic Regression) – 20% и RF (Random Forest) – 7%. Более подробная информация представлена на рис. 2.

Общий анализ

Согласно результатам метаанализа оценки ROC – AUC в 15 прогностических моделях, представленных в пяти исследованных публикациях с применением метода случайных эффектов, было установлено, что общее значение ROC – AUC составило 0.837 [95% ДИ 0.786–0.887], $I^2=91.78\%$, $p<0.001$ (рис. 3).

Анализ в подгруппах

Поскольку результаты анализа прогностических моделей оказались достаточно гетерогенными ($I^2=91.78\%$, $p<0.001$), было решено провести анализ в подгруппах. Исследование данных по подгруппам дало следующие результаты.

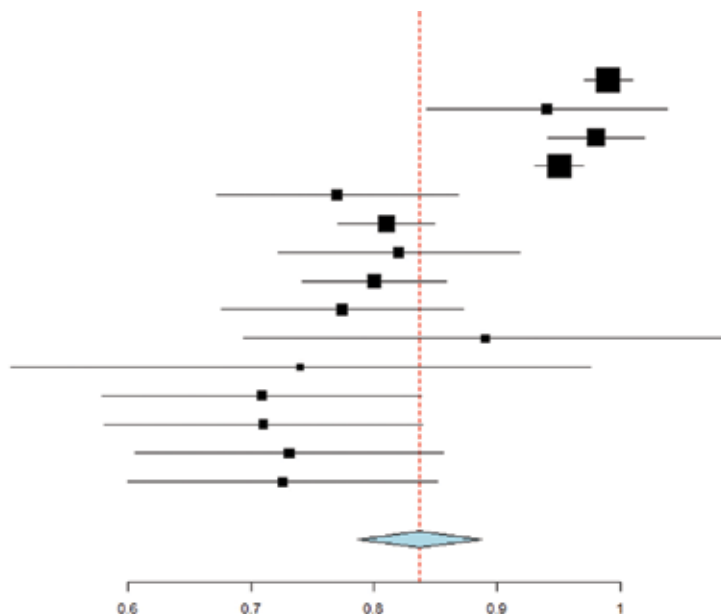
При оценке прогностических моделей в зависимости от используемого алгоритма выяснилось, что модели, основанные на методе МО SVM, продемонстрировали наивысшую точность (ROC – AUC=0.901 [95% ДИ 0.850–0.951]). На втором месте оказались модели, построенные с использованием алгоритма AdaBoost (ROC – AUC=0.805 [95% ДИ 0.775–0.834]). Третье место заняли модели, разработанные с помощью метода MLR (ROC – AUC=0.764 [95% ДИ 0.655–0.872]). Однако наиболее точные модели, основанные на SVM, оказались менее стабильными и более гетерогенными ($I^2=88.34\%$, $p<0.001$) по сравнению с моделями, созданными с использованием AdaBoost ($I^2=0\%$, $p=0.880$) и MLR ($I^2=13.6\%$, $p=0.314$) (рис. 4).

Модели, созданные с использованием множественной линейной регрессии (MLR), являются открытыми, что подразумевает возможность для исследователей и пользователей видеть и интерпретировать все параметры и коэффициенты, влияющие на предсказания, что делает такие модели более прозрачными и понятными для анализа. В отличие от них, модели, построенные с помощью алгоритмов МО SVM, AdaBoost и RF, считаются закрытыми, так как они часто представляют собой сложные математические конструкции, которые труднее интерпретировать и понять, что может затруднить анализ влияния отдельных переменных на результаты.

Метод опорных векторов (SVM) является мощным алгоритмом МО, который используется для классификации. Он работает, находя гиперплоскость, которая наилучшим образом разделяет данные на разные классы, максимизируя расстояние между ближайшими точками данных (опорными векторами) и гиперплоскостью. SVM может эффективно обрабатывать как линейно разделяемые, так и нелинейные данные с использованием ядерных функций для преобразования входных данных в более высокие размерности, что позволяет находить сложные границы между классами.

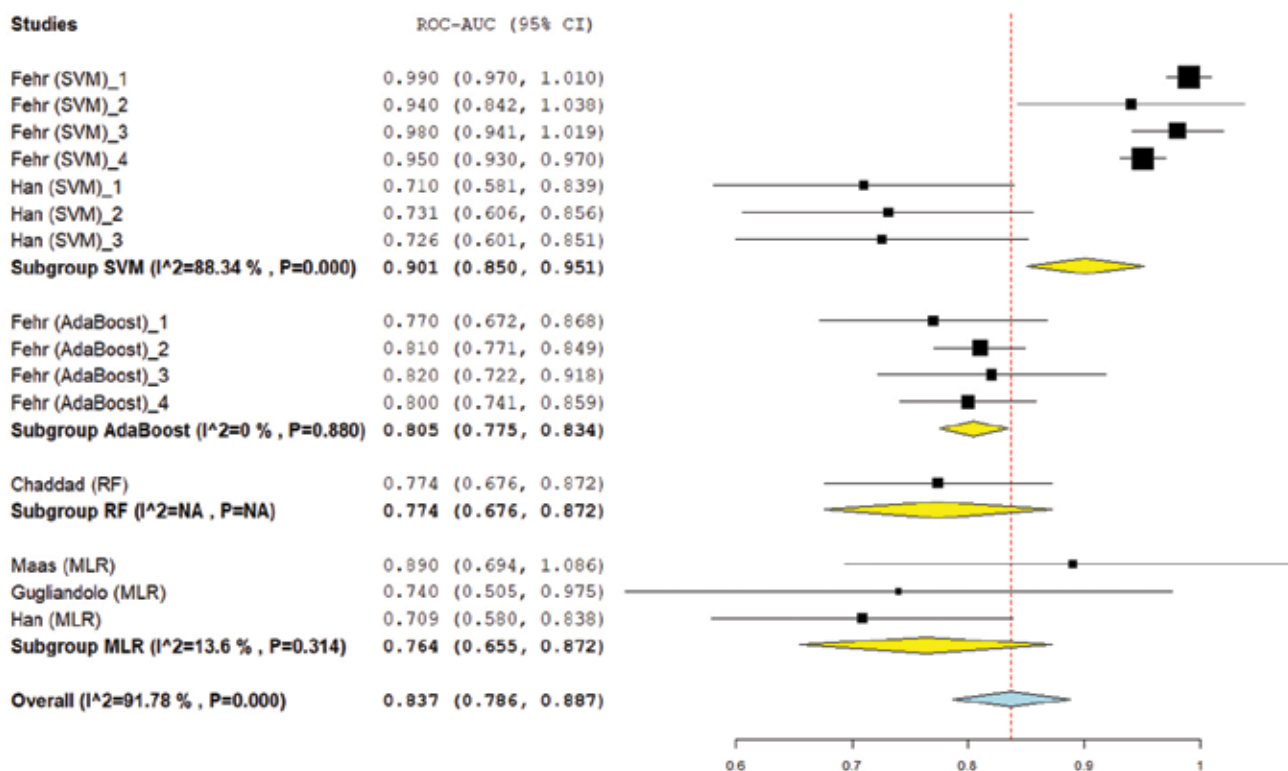
Алгоритм AdaBoost, или адаптивный бустинг, работает путем комбинирования нескольких слабых классификаторов для создания одного сильного. Он последовательно обучает классификаторы, придавая больший вес ошибочно классифици-

Studies	ROC-AUC (95% CI)
Fehr (SVM)_1 2015	0.990 (0.970, 1.010)
Fehr (SVM)_2 2015	0.940 (0.842, 1.038)
Fehr (SVM)_3 2015	0.980 (0.941, 1.019)
Fehr (SVM)_4 2015	0.950 (0.930, 0.970)
Fehr (AdaBoost)_1 2015	0.770 (0.672, 0.868)
Fehr (AdaBoost)_2 2015	0.810 (0.771, 0.849)
Fehr (AdaBoost)_3 2015	0.820 (0.722, 0.918)
Fehr (AdaBoost)_4 2015	0.800 (0.741, 0.859)
Chaddad (RF) 2018	0.774 (0.676, 0.872)
Maas (MLR) 2019	0.890 (0.694, 1.086)
Gugliandolo (MLR) 2020	0.740 (0.505, 0.975)
Han (MLR) 2021	0.709 (0.580, 0.838)
Han (SVM)_1 2021	0.710 (0.581, 0.839)
Han (SVM)_2 2021	0.731 (0.606, 0.856)
Han (SVM)_3 2021	0.726 (0.601, 0.851)
Overall ($I^2=91.78\%$, $P<0.001$)	0.837 (0.786, 0.887)



Примечание. Горизонтальные линии представляют 95% ДИ точечных оценок. Каждый сплошной прямоугольник представляет значение ROC – AUC отдельных моделей, а размер прямоугольника указывает на вес исследования. Ромб означает объединенное значение ROC – AUC всех 15 моделей в пяти исследованиях. Пунктирная линия обозначает среднее значение ROC – AUC.

Рис. 3. Блобограмма отдельных прогностических моделей для объединенной площади под кривой (ROC – AUC) и 95% ДИ характеристики разделения РПЖ на GP 4+3 и GP 3+4



Примечание. Горизонтальные линии представляют 95% ДИ точечных оценок. Каждый сплошной прямоугольник представляет значение ROC – AUC отдельных моделей, а размер прямоугольника указывает на вес исследования. Ромб означает объединенное значение ROC – AUC всех 15 моделей в пяти исследованиях. SVM – метод опорных векторов. AdaBoost – адаптивный бустинг. RF – случайный лес. MLR – множественная линейная регрессия.

Рис. 4. Блобограмма отдельных прогностических моделей для объединенной площади под кривой (ROC – AUC) и 95% ДИ характеристики разделения РПЖ на $GP \geq 4+3$ и $GP \leq 3+4$

рованным образцам, что позволяет улучшить общую точность модели. Алгоритм «случайный лес» (RF) представляет собой ансамблевый метод, в котором используется множество деревьев решений для повышения точности и устойчивости предсказаний. Каждое дерево обучается на случайной подвыборке данных, и итоговый результат формируется путем голосования среди всех деревьев, что снижает вероятность переобучения и повышает надежность модели.

В итоге можно сделать вывод, что закрытые модели продемонстрировали лучшие показатели качества по сравнению с открытыми моделями для разделения рака предстательной железы на группы $GP \geq 4+3$ и $GP \leq 3+4$ (ROC – AUC = 0.847 [95% ДИ 0.794–0.899] и ROC – AUC = 0.764 [95% ДИ 0.655–0.872] соответственно).

Также представляет интерес анализ процесса создания прогностических моделей. Анализировали, использовали ли авторы публикаций следующие приемы:

1. *Проверка признаков на мультиколлинеарность.* Анализ мультиколлинеарности между переменными имеет важное значение, так как включение в модель сильно коррелирующих факторов может привести к их взаимной компенсации, что затрудняет определение независимого влияния каждого из них на итоговый результат. Исключение таких переменных помогает уменьшить размерность модели; вместо того чтобы использовать обе переменные, целесообразно выбрать ту, которая оказывает большее влияние на результат, и исключить другую. Проверка на мультиколлинеарность была проведена только в одном из пяти исследований [8], в котором были удалены переменные с коэффициентом корреляции $R \geq 0.8$.

2. *Тестирование модели.* Выяснилось, что точность моделей, оцененных на тестовых выборках (test validation), была несколько ниже по сравнению с моделями, прошедшими проверку с ис-

пользованием кросс-валидации (cross-validation). Это может быть связано с тем, что кросс-валидация позволяет более эффективно использовать данные для обучения и тестирования, обеспечивая более надежную оценку производительности модели, так как она учитывает различные подмножества данных и минимизирует влияние случайных факторов. В то время как тестовые выборки могут не полностью отражать разнообразие данных, на которых модель была обучена, кросс-валидация помогает выявить устойчивость модели к различным условиям и уменьшает вероятность переобучения. Test validation: 0.719 [95% ДИ 0.656–0.783], $I^2=0\%$, $p=0.993$; cross-validation: 0.874 [95% ДИ 0.822–0.926], $I^2=92.02\%$, $p<0.001$.

Обсуждение

Цель данного исследования заключалась в оценке качества прогностических моделей, разработанных для разделения при первичном обращении пациентов с раком предстательной железы с паттерном Глисона 7 на группы $GP 4+3$ и $GP 3+4$.

В представленном метаанализе, основанном на данных МРТ с применением радиомики и методов МО, были получены многообещающие результаты с совокупным значением ROC – AUC = 0.837 [95% ДИ 0.786–0.887]. Поскольку между исследованиями наблюдалась высокая степень неоднородности ($I^2=91.78\%$, $p<0.001$), было решено провести анализ в подгруппах для выявления источников этой неоднородности. Из-за значительной гетерогенности результатов во всех объединенных анализах использовали модель случайных эффектов. Наивысшее значение ROC – AUC = 0.990 было получено в исследовании [17], в то время как самое низкое составило ROC – AUC = 0.710 [9].

При анализе прогностических моделей в зависимости от используемого алгоритма МО выяснилось, что модели, основан-

ные на SVM, продемонстрировали наивысшую точность (ROC – AUC=0.901 [95% ДИ 0.850–0.951]). Однако эти модели оказались менее стабильными и более гетерогенными ($I^2=88.34\%$, $p<0.001$) по сравнению с моделями, построенными с использованием AdaBoost ($I^2=0\%$, $p=0.880$) и MLR ($I^2=13.6\%$, $p=0.314$).

Важно отметить, что модели, основанные на SVM, AdaBoost и RF, являются «черными ящиками» и не поддаются интерпретации. Тем не менее закрытые модели показали лучшие результаты по сравнению с открытыми моделями для разделения пациентов с раком предстательной железы $GP \geq 4+3$ и $GP \leq 3+4$ (ROC – AUC=0.847 [95% ДИ 0.794–0.899] и ROC – AUC=0.764 [95% ДИ 0.655–0.872] соответственно).

Наше исследование продемонстрировало наличие широкого спектра первичных исследований, направленных на создание прогностических моделей на основе бпМРТ и методов МО для разделения РПЖ с паттерном Глисона 7: $GP \geq 4+3$ и $GP \leq 3+4$. Однако нам не удалось найти соответствующие метаанализы, посвященные заявленной цели нашего исследования. Таким образом, можно утверждать, что представленный метаанализ является первым в своем роде.

Данный систематический обзор имеет ряд ограничений. Прежде всего, количество исследований, соответствующих критериям отбора, было относительно небольшим. Это частично связано с несоответствием между заголовком статьи и ее содержанием. Мы также обнаружили высокую гетерогенность, что является общей проблемой для других метаанализов, связанных с МРТ предстательной железы и точностью диагностических тестов в целом [20, 21]. Тем не менее общее качество прогностических моделей в отобранных исследованиях было достаточным для проведения метаанализа.

Заключение

Систематический обзор с метаанализом продемонстрировал многообещающие результаты в количественной идентификации РПЖ с паттерном Глисона 7 на группы $GP 4+3$ и $GP 3+4$, основанные на методах МО и радиомическом анализе данных бпМРТ. Однако к этим выводам следует подходить с осторожностью, так как многоцентровые исследования, которые обеспечивают прямое сравнение эффективности алгоритмов МО и работы рентгенологов, все еще отсутствуют. Тем не менее современные методы с использованием искусственного интеллекта могут повысить точность дифференциальной диагностики рака предстательной железы $GP 4+3$ и $GP 3+4$ и улучшить воспроизводимость методик в клинической практике.

Источник финансирования

Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Литература

1. Siegel R.L. et al. Cancer statistics // CA Cancer J Clin. – 2017. – V. 67. – No 1. – P. 7–30.
2. Gleason D.F. et al., Mellinger G.T. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging // Journal of Urology. – 1974. – V. 111. – No 1. – P. 58–64.
3. Stark J.R. et al. Gleason score and lethal prostate cancer: does $3+4=4+3$? // Journal of Clinical Oncology. – 2009. – V. 27. – No 21. – P. 3459–3464.
4. Choy B. et al. Prognostic significance of percentage and architectural types of contemporary Gleason pattern 4 prostate cancer in radical prostatectomy // American Journal of Surgical Pathology. – 2016. – V. 40. – No 10. – P. 1400–1406.
5. Epstein J.I. et al. A contemporary prostate cancer grading system: a validated alternative to the Gleason score // Eur Urol. – 2016. – V. 69. – No 3. – P. 428–435.
6. Epstein J.I. Prostate cancer grading: a decade after the 2005 modified system // Modern Pathology. – 2018. – V. 31. – P. 47–63.
7. Han C. et al., Liu J., Wu X. et al. Radiomics models based on apparent diffusion coefficient maps for the prediction of high-grade prostate cancer at radical prostatectomy: comparison with preoperative biopsy // Journal of Magnetic Resonance Imaging. – 2021. – V. 54. – No 6. – P. 1892–1901.
8. Gugliandolo S.G. et al. MRI-based radiomics signature for localized prostate cancer: a new clinical tool for cancer aggressiveness prediction? Sub-study of prospective phase II trial on ultra-hypofractionated radiotherapy (AIRC IG-13218) // Eur Radiol. – 2021. – V. 31. – No 2. – P. 716–728.
9. Han C. et al. Radiomics models based on apparent diffusion coefficient maps for the prediction of high-grade prostate cancer at radical prostatectomy: comparison with preoperative biopsy // Journal of Magnetic Resonance Imaging. – 2021. – V. 54. – No 6. – P. 1892–1901.
10. Chen T. et al. MRI based radiomics compared with the PI-RADS v2.1 in the prediction of clinically significant prostate cancer: biparametric vs multiparametric MRI // Front Oncol. – 2022. – V. 11. DOI: 10.3389/fonc.2021.821435.
11. Zhang L. et al. Predicting the grade of prostate cancer based on a biparametric MRI radiomics signature // Contrast Media Mol Imaging. – 2021. – V. 2021. – P. 1–10.
12. Kuhl C.K. et al. Abbreviated biparametric prostate MR imaging in men with elevated prostate-specific antigen // Radiology. – 2017. – V. 285. – No 2. – P. 493–505.
13. Steyerberg E.W. et al. Assessing the performance of prediction models // Epidemiology. – 2010. – V. 21. – No 1. – P. 128–138.
14. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions version 6.0. – URL: <http://www.training.cochrane.org/handbook>.
15. Higgins J.P.T. et al. Measuring inconsistency in meta-analyses // BMJ. – 2003. – V. 327. – No 7414. – P. 557–560.
16. DerSimonian R. et al. Meta-analysis in clinical trials // Control Clin Trials. – 1986. – V. 7. – No 3. – P. 177–188.
17. Fehr D. et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images // Proceedings of the National Academy of Sciences. – 2015. – V. 112. – No 46. – P. E6265–E6273.
18. Chaddad A. et al. Predicting Gleason score of prostate cancer patients using radiomic analysis // Front Oncol. – 2018. – V. 8. – P. 630. DOI: 10.3389/fonc.2018.00630.
19. Maas M.C. et al. A single-arm, multicenter validation study of prostate cancer localization and aggressiveness with a quantitative multiparametric magnetic resonance imaging approach // Invest Radiol. – 2019. – V. 54. – No 7. – P. 437–447.
20. Zhang L. et al. A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR imaging for the detection of prostate cancer // Eur Radiol. – 2017. – V. 27. – No 12. – P. 5204–5214.
21. Zhen L. et al. Accuracy of multiparametric magnetic resonance imaging for diagnosing prostate cancer: a systematic review and meta-analysis // BMC Cancer. – 2019. – V. 19. – No 1. – P. 1244.