



Сравнение эффективности и методологического качества моделей машинного обучения для прогнозирования исходов при сердечной недостаточности с сохраненной и сниженной фракцией выброса: систематический обзор

Буренков Ю. В.¹, Шевцова В. И.¹, Кутилова Д. Е.², Тарасова А. О.²

Цель. Провести сравнительную оценку эффективности и методологического качества моделей машинного обучения (machine learning, ML) для прогнозирования исходов у пациентов с хронической сердечной недостаточностью с сохраненной (ХСНсФВ) и сниженной (ХСНнФВ) фракцией выброса.

Материал и методы. Проведен систематический поиск оригинальных работ в базах данных PubMed и eLibrary (2015-2025гг). В обзор было включено 20 исследований. Методологическое качество оценивалось по PROBAST.

Результаты. Эффективность ML-моделей была сопоставимой для ХСНсФВ и ХСНнФВ (медиана AUC 0,812 в обеих группах, $p=0,48$). 70% исследований имели высокий риск смещения, преимущественно из-за отсутствия внешней валидации. Модели кластеризации демонстрировали клинически значимую стратификацию, выявляя фенотипы со статистически значимым повышенным риском (например, отношение рисков 2,99 (95% доверительный интервал: 2,41-3,7, $p<0,001$)).

Заключение. ML-модели показывают умеренно-высокую эффективность, однако отсутствие внешней валидации ограничивает их готовность к клиническому применению. Необходимы стандартизация методологии для будущих исследований и проспективная валидация.

Ключевые слова: машинное обучение, сердечная недостаточность, прогнозирование исходов, методологическое качество, хроническая сердечная недостаточность с сохраненной фракцией выброса, хроническая сердечная недостаточность со сниженной фракцией выброса.

Отношения и деятельность: нет.

¹ФГБОУ ВО Воронежский государственный медицинский университет им. Н.Н. Бурденко Минздрава России, Воронеж, Россия; ²ФГАУ ВО Белгородский государственный национальный исследовательский университет, Белгород, Россия.

Буренков Ю. В.* — ординатор 1-го года обучения по специальности "Терапия", факультет подготовки кадров высшей квалификации, ORCID: 0009-0007-8705-7438, Шевцова В. И. — к.м.н., доцент, доцент кафедры инфекционных болезней и клинической иммунологии, ORCID: 0000-0002-1707-436X, Кутилова Д. Е. — студентка 5 курса, лечебный факультет, ORCID: нет, Тарасова А. О. — студентка 5 курса, лечебный факультет, ORCID: нет.

*Автор, ответственный за переписку (Corresponding author): ghjd56@bk.ru

ЛЖ — левый желудочек, ФВ — фракция выброса, ХСН — хроническая сердечная недостаточность, ХСНсФВ — хроническая сердечная недостаточность с сохраненной фракцией выброса, ХСНнФВ — хроническая сердечная недостаточность со сниженной фракцией выброса, ХСНнФВ — хроническая сердечная недостаточность с умеренно сниженной фракцией выброса, ЭхоКГ — эхокардиография, AUC — площадь под ROC-кривой, CI — доверительный интервал, HR — отношение рисков, LR — метод логистической регрессии, ML — машинное обучение, RF — модель случайного леса, XGBoost — модель градиентного бустинга.

Рукопись получена 01.12.2025

Рецензия получена 17.02.2026

Принята к публикации 24.03.2026



Для цитирования: Буренков Ю. В., Шевцова В. И., Кутилова Д. Е., Тарасова А. О. Сравнение эффективности и методологического качества моделей машинного обучения для прогнозирования исходов при сердечной недостаточности с сохраненной и сниженной фракцией выброса: систематический обзор. *Российский кардиологический журнал*. 2026;31(2S):6721. doi: 10.15829/1560-4071-2026-6721. EDN: HUTKZB

Comparison of the performance and methodological quality of machine learning models for predicting outcomes in heart failure with preserved and reduced ejection fraction: a systematic review

Burenkov Yu. V.¹, Shevtsova V. I.¹, Kutilova D. E.², Tarasova A. O.²

Aim. To assess performance and methodological quality of machine learning (ML) models for predicting outcomes in patients with heart failure with preserved (HFpEF) and reduced (HFrEF) ejection fraction.

Material and methods. A systematic search of original publications in PubMed and eLibrary (2015-2025) was conducted. Twenty studies were included in the review. Methodological quality was assessed using PROBAST.

Results. The performance of the ML models was comparable for HFpEF and HFrEF (median AUC, 0,812 in both groups, $p=0,48$). The analysis showed that 70% of the studies were at high risk of bias, primarily due to the lack of external validation. The clustering models demonstrated clinically meaningful stratification, identifying phenotypes with significantly increased risk (e.g., hazard ratio 2,99 (95% confidence interval 2,41-3,7, $p<0,001$)).

Conclusion. ML models demonstrate moderate-to-high performance. However, the lack of external validation limits their readiness for clinical application. Standardization of methodology for future studies and prospective validation are needed.

Keywords: machine learning, heart failure, outcome prediction, methodological quality, heart failure with preserved ejection fraction, heart failure with reduced ejection fraction.

Relationships and Activities: none.

¹Burdenko Voronezh State Medical University, Voronezh, Russia; ²Belgorod State National Research University, Belgorod, Russia.

Burenkov Yu. V.* ORCID: 0009-0007-8705-7438, Shevtsova V. I. ORCID: 0000-0002-1707-436X, Kutilova D. E. ORCID: none, Tarasova A. O. ORCID: none.

*Corresponding author:
ghjd56@bk.ru

Received: 01.12.2025 Revision Received: 17.02.2026 Accepted: 24.03.2026

For citation: Burenkov Yu. V., Shevtsova V. I., Kutilova D. E., Tarasova A. O. Comparison of the performance and methodological quality of machine learning models for predicting outcomes in heart failure with preserved and reduced ejection fraction: a systematic review. *Russian Journal of Cardiology*. 2026;31(2S):6721. doi: 10.15829/1560-4071-2026-6721. EDN: HUTKZB

Ключевые моменты

- Модели машинного обучения демонстрируют сопоставимую эффективность при хронической сердечной недостаточности с сохраненной и сниженной фракцией выброса.
- Основной методологический недостаток исследований — отсутствие внешней валидации моделей.
- Необходимо обеспечить соблюдение стандартов отчетности и проведение проспективной валидации для будущих исследовательских работ.

Key messages

- Machine learning models demonstrate comparable efficacy in heart failure with preserved and reduced ejection fraction.
- The main methodological shortcoming is the lack of external validation of the models.
- Adherence to reporting standards and prospective validation is essential for future research.

Хроническая сердечная недостаточность (ХСН) остается глобальной медико-социальной проблемой, характеризующейся неблагоприятным прогнозом и высоким экономическим бременем, где пятилетняя смертность превышает 50%, а затраты на госпитализацию составляют до 75% расходов системы здравоохранения по данному заболеванию [1, 2]. Особую сложность представляет гетерогенность заболевания, требующая дифференцированного подхода к стратификации риска у пациентов с сохраненной и сниженной фракцией выброса (ФВ) левого желудочка (ЛЖ) [3].

Согласно международной классификации ХСН и современным рекомендациям Российского кардиологического общества (2024), выделяют три фенотипа ХСН в зависимости от величины ФВ ЛЖ: со сниженной ФВ (ХСНнФВ) — $\leq 40\%$, с умеренно сниженной ФВ (ХСНунФВ) — 41-49% и со сохраненной ФВ (ХСнсФВ) — $\geq 50\%$ [4, 5]. Однако стоит учитывать, что среди существующих исследований имеется неоднородность по критериям отнесения пациентов с ХСН к тому или иному фенотипу (например, ХСнсФВ определялась как ФВ ЛЖ $\geq 40\%$ или $\geq 45\%$, ХСНнФВ — как $\leq 40\%$ или $\leq 45\%$) [6].

Традиционные прогностические шкалы (MAGGIC, SHFM) демонстрируют умеренную точность и лишь немногие из них разрабатывались для группы с ХСнсФВ [7-9]. Неспособность учитывать комплексные взаимосвязи между многочисленными предикторами ограничивает их использование в персонализированной медицине.

Методы машинного обучения (ML) открывают новые перспективы за счет возможности анализа больших объемов многомерных данных. Способность алгоритмов ML выявлять сложные паттерны и создавать индивидуализированные прогностические модели позволяет преодолеть ограничения традиционных методов [8]. Многие методы обучения с учителем ("традиционное машинное обучение"), не основан-

ные на нейронных сетях, например, дерево решений или градиентный бустинг, применяются в основном для решения таких задач, как прогнозирование исходов [10].

Однако, несмотря на растущее число публикаций, отсутствуют данные о сравнительной эффективности ML при разных фенотипах ХСН [11]. Существующие обзоры не учитывают методологическое качество исследований и фенотипические различия, что может приводить к некорректным выводам [12].

Цель систематического обзора — сравнительная оценка эффективности и методологического качества моделей ML для прогнозирования исходов у пациентов с сердечной недостаточностью с сохраненной и сниженной ФВ.

Материал и методы

Стратегия поиска была разработана с использованием комбинаций ключевых слов. Поиск проводился в базах данных PubMed и eLibrary за период с 1 января 2015г по 20 октября 2025г. Для PubMed использовался запрос с применением медицинских предметных рубрик (MeSH) и поиск по заголовкам/аннотациям. Аналогичная стратегия была адаптирована для eLibrary с использованием русскоязычных терминов. Использовались следующие поисковые запросы: heart failure, heart failure with reduced ejection fraction, HFrEF, heart failure with preserved ejection fraction, HFpEF, machine learning, prognosis, prediction, ХСН, хроническая сердечная недостаточность, ML, прогностическая модель, прогнозирование госпитализации, прогнозирование смертности.

Для уточнения цели систематического обзора и улучшения точности при отборе исследований использовались следующие критерии (PICO):

- Р (участники): взрослые пациенты с ХСнсФВ или ХСНнФВ;
- I (вмешательство): модели ML для прогнозирования госпитализации/смертности или выявления прогностических кластеров;
- С (сравнение): сравнение с традиционными шкалами (например, MAGGIC, SHFM) или с другими методами ML;

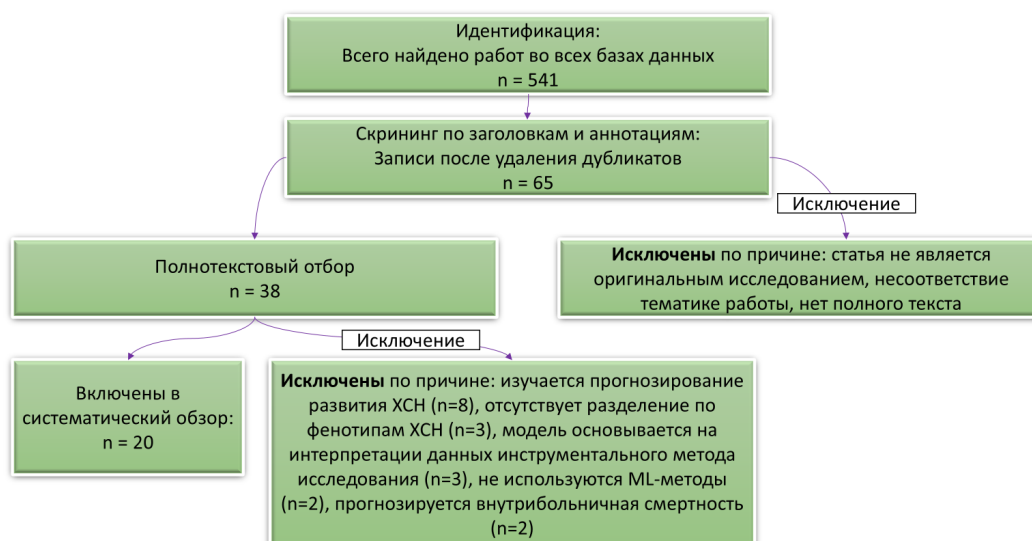


Рис. 1. Блок-схема PRISMA.

Сокращения: ФВ — фракция выброса, ХСН — хроническая сердечная недостаточность.

• О (исходы): госпитализация, смертность; метрики AUC-ROC, C-statistic, hazard ratio (HR).

После скрининга и оценки полнотекстовых статей в итоговый обзор было включено 20 исследований, соответствующих критериям. Процесс отбора детализирован на блок-схеме PRISMA (рис. 1). Основные причины исключения: изучается прогнозирование развития ХСН (n=8), отсутствует разделение по фенотипам ХСН (n=3), модель основывается на интерпретации данных инструментального метода исследования (n=3), не используются ML-методы (n=2), прогнозируется внутрибольничная смертность (n=2).

Из каждого исследования извлекались данные по стандартизированной форме, включавшей характеристики исследования, популяции, моделей ML и показателей эффективности. Методологическое качество оценивалось с использованием инструмента PROBAST, фокусирующегося на четырех доменах: участники, предикторы, исход, анализ. Исследования с отсутствием внешней валидации автоматически получали высокий риск смещения.

Вследствие значительной методологической гетерогенности проведение метаанализа было признано нецелесообразным. Сравнительный анализ с качественным синтезом данных выполнен по трем направлениям:

- 1) эффективность моделей между фенотипами ХСН,
- 2) сравнение подходов supervised (прямое прогнозирование) и unsupervised learning (анализ кластеров),
- 3) анализ методологического качества.

Результаты

Характеристики исследования

Первоначальный поиск выявил 541 потенциально подходящую статью. На этапе прочтения аннота-

ций и названий были исключены 503 работы. После разбора полнотекстовых вариантов для включения в обзор отобраны 20 оригинальных исследований. Процесс отбора детально отображен на блок-схеме PRISMA (рис. 1).

Характеристика включенных в обзор исследований

Из 20 включенных исследований 6 работ было посвящено ХСНнФВ, 8 — ХСНсФВ и 3 сфокусированы на ХСНунФВ. Решение о включении в обзор исследований, посвященных ХСНунФВ, было принято из-за отсутствия единых критериев для ХСНнФВ среди рассматриваемых работ. Дополнительно было рассмотрено 4 исследования, изучающих роль ML в прогнозировании исхода ХСН посредством выделения кластеров в рамках существующего деления на фенотипы ХСН в зависимости от ФВ ЛЖ. Наиболее часто используемыми типами моделей ML были случайный лес (RF), градиентный бустинг (XGBoost) и логистическая регрессия (LR), использовавшиеся в 13, 11 и 9 исследованиях, соответственно. Размеры когорт варьировались от 116 до 47498 пациентов (табл. 1), сравнение эффективности предложенных в рассматриваемых работах моделей приведены в таблице 2.

Эффективность моделей для ХСНнФВ

В 7 исследованиях, посвященных ХСНнФВ, эффективность моделей варьировалась в широких пределах: AUC от 0,649 до 0,96 (медиана 0,812, $p < 0,05$). В работах, где проводилось сравнение предложенных моделей с традиционными шкалами, методы ML показали превосходство [13, 14]. Наибольшую эффективность (AUC > 0,85) показала модель RF, медиана AUC которой составила 0,915 [14, 15], и модель, основанная на глубоком обучении, с диапазоном AUC 0,938-0,977 [16]. При этом наилучшие результаты были показаны в работах, не учитывающих данные эхокардиографии (ЭхоКГ) пациентов [14, 16], в отличие

Таблица 1

Общая характеристика исследований

Порядковый номер исследования	Автор, год	Фенотип ХСН	Размер когорты (n)	Тип ML-модели	Прогнозируемый исход	Основные предикторы
1	Angraal S, Mortazavi B, Gupta A, et al. 2020	ХСНсФВ	1767	LR, LR with lasso, XGBoost, Support vector machine, RF	Госпитализация, смертность	Данные ЭКГ, ФВ ЛЖ, клинические, опросник KCCQ, демографические и лабораторные (86 переменных)
2	Chang CY, Chen CC, Tsai ML, et al. 2024	ХСНсФВ	6092	RF	Госпитализация, смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (15 переменных в итоговой модели)
3	Hu Y, Wang X, Xiao S, et al. 2022	ХСНнФВ	700	Normogram model	Госпитализация	Данные ЭхоКГ, демографические, лабораторные, принимаемые лекарственные средства (43 переменных)
4	Ru B, Tan X, Liu Y, et al. 2023	ХСНнФВ	30687	Multilayer perceptron neural network model, XGBoost, RF, LR	Госпитализация	Данные ФВ ЛЖ, клинические, демографические, лабораторные, принимаемые лекарственные средства (46 переменных)
5	Ma F, Hu Y, Han P, et al. 2025	ХСНнФВ	733	RF, Decision Tree, XGBoost, Support Vector Machine, LR, LightGBM, and Multilayer Perceptron	Госпитализация	Данные ЭхоКГ, демографические, лабораторные (включая специально отобранные воспалительные и метаболические показатели), принимаемые лекарственные средства (50 переменных)
6	Zheng B, Liang T, Mei J, et al. 2024	ХСНсФВ	746	XGBoost, RF, neural network, LR	Госпитализация	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (12 переменных в итоговой модели)
7	Драгомирецкая Н. А., Толмачева А. В., Ветлужская М. В. и др. 2024	ХСНсФВ, ХСНунФВ	135	RF, LR	Смертность	Данные ЭхоКГ, клинические, демографические, лабораторные (33 переменных)
8	McDowell K, Kondo T, Talebi A, et al. 2024	ХСНсФВ	15187	The Cox model	Госпитализация, смертность	Данные ФВ ЛЖ, клинические, демографические, лабораторные, принимаемые лекарственные средства (11 итоговых переменных)
9	Shin I, Bhatt N, Alashi A, et al. 2024	ХСНсФВ	3235	Support Vector Classifier, LR, Lasso Regression, Elastic Net, RF, Histogram-based Gradient Boosting Classifier, XGBoost	Смертность	Данные клинические, демографические, лабораторные (36 переменных)
10	Hu Y, Ma F, Hu M, et al. 2025	ХСНсФВ	1569	RF, Decision Tree, XGBoost, Support Vector Machine, LR, LightGBM, Multilayer Perceptron	Госпитализация	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (9 переменных в итоговой модели)
11	Gtifi I, Abdelhedi R, Ouarda W, et al. 2022	ХСНнФВ	116	Generalized linear model, RF, XGBoost	Смертность	Данные ЭКГ, ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (9 переменных в итоговой модели)
12	Park J, Zhong X, Babaie Sarjaloo F, Wokhlu A. 2022	ХСНсФВ	1965	LR	Госпитализация, смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (24 переменных в итоговой модели)
13	Park J, Zhong X, Babaie Sarjaloo F, Wokhlu A. 2022	ХСНсФВ	1965	LR	Госпитализация, смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (25 переменных в итоговой модели)

Таблица 1. Продолжение

Порядковый номер исследования	Автор, год	Фенотип ХСН	Размер когорты (n)	Тип ML-модели	Прогнозируемый исход	Основные предикторы
14	Kyodo A, Kanaoka K, Keshi A, et al. 2023	ХСНсФВ	638	A variational Bayesian–Gaussian mixture model, Hierarchical clustering (18 переменных), RF	Госпитализация, смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (24 переменных)
15	Monzo L, Bresso E, Dickstein K, et al. 2024	ХСНнФВ	8591	Latent class model, RF and a decision tree classifier	Госпитализация, смертность	Данные клинические, демографические, лабораторные, принимаемые лекарственные средства (24 переменных)
16	Segar MW, Patel KV, Ayers C, et al. 2019	ХСНсФВ	1767	Cox regression analysis, Cohen's Kappa statistic	Госпитализация, смертность	Данные ЭКГ, ЭхоКГ, клинические, демографические, лабораторные (20 переменных в итоговой модели)
17	Soltani F, Jenkins DA, Kaura A, et al. 2024	ХСНсФВ, ХСНунФВ	2187	Cox Proportional Hazards and Fine-Gray competing risks models, Density-Based Spatial Clustering and Application with NOISE; Gaussian mixture modelling; and k-means clustering	Госпитализация, смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, инвазивная ангиография на момент первичного поступления в стационар (42 переменных в итоговой модели)
18	Wang Z, Chen X, Tan X, et al. 2021	ХСНнФВ	47498	Sequential model architecture based on bi-directional LSTM, Bag-of-words model; LR, RF, XGBoost	Госпитализация	Данные демографические, сведения о заболеваниях, принимаемой терапии (28 переменных)
19	Tian P, Liang L, Zhao X, et al. 2023	ХСНунФВ	424	Cox-Boost, LR; XGBoost, RF, the support vector machine	Смертность	Данные ЭхоКГ, клинические, демографические, лабораторные, принимаемые лекарственные средства (17 переменных в итоговой модели)
20	Feng J, Huang L, Zhao X, et al. 2024	ХСНсФВ	1083	Lasso-Cox	Смертность	Данные ФВ ЛЖ, клинические, демографические, лабораторные, принимаемые лекарственные средства (50 переменных)

Сокращения: ФВ ЛЖ — фракция выброса левого желудочка, ХСН — хроническая сердечная недостаточность, ХСНсФВ — хроническая сердечная недостаточность с сохраненной фракцией выброса, ХСНнФВ — хроническая сердечная недостаточность со сниженной фракцией выброса, ХСНунФВ — хроническая сердечная недостаточность с умеренно сниженной фракцией выброса, LR — метод логистической регрессии, ML — машинное обучение, RF — модель случайного леса, XGBoost — модель градиентного бустинга, ЭКГ — электрокардиограмма, ЭхоКГ — эхокардиография.

от других исследований, включавших эти показатели, но получивших значимо меньшие значения AUC (медиана 0,745) [17–19]. Только работа Monzo L, et al. [14] прошла внешнюю проверку, показав высокую эффективность (AUC 0,94) и готовность к внедрению в практическую деятельность, в связи с ориентиром данной модели на показатели, доступные широкому кругу врачей, при небольшом числе переменных (всего 24 переменных).

Эффективность моделей для ХСНсФВ и ХСНунФВ

Среди 13 исследований с фокусом на ФВ ЛЖ >40% диапазон показателей эффективности составил от 0,69 до 0,98 (медиана 0,812, $p < 0,05$). При сравнении с традиционными шкалами и более старыми моделями предложенные методы показали большую эффективность [8, 13, 20–22] и лишь в одном исследовании этот показатель примерно равный [23]. Наиболее эффективными моделями ($AUC > 0,8$) оказались RF с медианой AUC 0,89 [8, 22], XGBoost с медианой AUC 0,906 [20, 22], Logistic Regression с медианой AUC 0,85 [24, 25] и модель LightGBM с показателем AUC 0,87 [17]. SHAP-анализ для выявления наиболее значимых предикто-

ров для исследуемых моделей был проведен в 4 из 7 вышеуказанных работ с наилучшими результатами [8, 17, 20, 22]. Примечательно, что в исследовании Kyodo A, et al. [25], где проводилось сравнение моделей RF и LR, в качестве предикторов не учитывались данные ЭхоКГ и сведения о принимаемых лекарственных препаратах, что может свидетельствовать о роли этих переменных в повышении прогностической точности модели RF. Внешняя валидация была проведена лишь в 2 исследованиях для моделей RF и LightGBM [8, 17].

Эффективность моделей, основанных на кластеризации

4 представленные в обзоре статьи были посвящены методам кластеризации в рамках деления ХСН на фенотипы в зависимости от ФВ ЛЖ. Работа [14] выделяла кластеры в рамках ХСНнФВ, в то время как исследования [25–27] фокусировались на группе пациентов с ФВ ЛЖ >40%. В 3 работах из 4 была проведена внешняя проверка моделей, что говорит об их высоком методологическом качестве. В 3 исследованиях, предоставивших количественную оценку риска с помощью HR, анализ выявленных кластеров показал значительные различия

Таблица 2

Сравнение эффективности и тип валидации

Порядковый номер исследования	Фенотип ХСН	Метрика эффективности: AUC-ROC, C-statistic	Сравнение с традиционной шкалой	Тип валидации
1	ХСНсФВ	C-статистика для смертности и госпитализации в связи с СН составила 0,70 и 0,69, соответственно, рассчитанная методом RF	Сиэтлская модель сердечной недостаточности, показала C-статистику 0,73 для смертности в сравнении с 0,72 в этом исследовании для смертности	Кросс-валидация
2	ХСНсФВ	AUC — 0,869	Не проводилось	Внешняя валидация
3	ХСНнФВ	AUC — 0,794	Не проводилось	Только тренировка и тест
4	ХСНнФВ	XGBoost с AUC 0,595 для 30-дневной повторной госпитализации, RF с AUC 0,630 для 90-дневной повторной госпитализации, XGBoost с AUC 0,649 для 365-дневной повторной госпитализации	Не проводилось	Только тренировка и тест
5	ХСНнФВ	RF достиг наивысшего AUC 0,89	Не проводилось	Кросс-валидация
6	ХСНсФВ	XGBoost показал наибольший AUC 0,98 в обучающей и 0,896 в тестовой выборке	Не проводилось	Только тренировка и тест
7	ХСНсФВ, ХСНнФВ	AUC — 0,863	Не проводилось	Не проводилась
8	ХСНсФВ	C-statistic через 1 год — 0,73	Показала более высокую дискриминацию, чем шкала риска MAGGIC (статистика C через 1 год: 0,60)	Внешняя валидация
9	ХСНсФВ	Регрессионные модели показали для 30-дневного периода AUC 0,83. Для результатов за 1 год наиболее эффективной была модель Histogram-based Gradient Boosting Classifier с AUC 0,78	Не проводилось	Кросс-валидация
10	ХСНсФВ	Значения AUC составили для каждой модели: LightGBM — 0,88, RF — 0,87, XGBoost — 0,86 и Multilayer Perceptron — 0,84. Модель LightGBM достигла AUC 0,87 при проверке на внешней базе данных	Не проводилось	Внешняя валидация
11	ХСНнФВ	Наилучшие результаты показала GLM с AUC 0,745	Не проводилось	Только тренировка
12	ХСНсФВ	AUC — 0,770	Модель превзошла шкалу оценки риска повторной госпитализации Йельского университета за 30 дней, для которой AUC — 0,6	Только тренировка
13	ХСНсФВ	AUC — 0,755	Модель превзошла шкалу оценки риска повторной госпитализации Йельского университета за 30 дней, для которой AUC — 0,6	Только тренировка
14	ХСНсФВ	AUC — 0,845	Не проводилось	Внешняя валидация
15	ХСНнФВ	Модель дерева решений показала AUC — 0,91. Модель RF показала AUC — 0,94	Кластеры значительно улучшили прогностические показатели по сравнению с оценкой по шкале MAGGIC для всех основных исследуемых конечных точек	Внешняя валидация
16	ХСНсФВ	C-statistic при прогнозировании первичной комбинированной конечной точки составил 0,62. При добавлении к наиболее подходящей модели 10 определенных оптимальных предикторов показатели улучшились C-statistic до 0,73	Не проводилось	Внешняя валидация
17	ХСНсФВ, ХСНнФВ	C-statistic — 0,7	Не проводилось	Не проводилась
18	ХСНнФВ	Модели глубокого обучения без буферного окна показали гораздо более высокую эффективность прогноза госпитализации (AUC 0,938-0,977), чем модели с 7-дневным буферным окном (AUC 0,492-0,595). Лучшие показатели AUC для моделей 30-дневных и 90-дневных повторных госпитализаций по поводу ХСН с использованием глубокого обучения составили 0,597 и 0,614, соответственно	Не проводилось	Кросс-валидация

Таблица 2. Продолжение

Порядковый номер исследования	Фенотип ХСН	Метрика эффективности: AUC-ROC, C-statistic	Сравнение с традиционной шкалой	Тип валидации
19	ХСНунФВ	Наилучший показатель выявлен для XGBoost AUC — 0,916. За ней последовала RF модель с AUC 0,900	Для прогнозирования смертности пациентов с сердечной недостаточностью со сниженной фракцией выброса использовались система оценки Seattle Heart Failure Model, калькулятор BCN Bio-HF и шкала MAGGIC, все значения AUC или c-statistic которых были ниже 0,8. Шкала BIOSTAT-HF показала аналогичные результаты: c-statistic составили 0,74 (ХСНсФВ) и 0,72 (ХСНнФВ)	Кросс-валидация
20	ХСНсФВ	AUC — 0,778, 0,782 и 0,772 через 1, 3 и 5 лет после выписки, соответственно	Не проводилось	Кросс-валидация

Сокращения: ХСН — хроническая сердечная недостаточность, ХСНсФВ — хроническая сердечная недостаточность с сохраненной фракцией выброса, ХСНнФВ — хроническая сердечная недостаточность со сниженной фракцией выброса, ХСНунФВ — хроническая сердечная недостаточность с умеренно сниженной фракцией выброса, AUC — площадь под ROC-кривой, RF — модель случайного леса, XGBoost — модель градиентного бустинга.

в прогнозировании: в исследовании Soltani F, et al. [27] выявили кластер с HR 2,99 (95% доверительный интервал (CI): 2,41–3,7, $p < 0,001$) по сравнению с референсной группой, в работе Segar MW, et al. [26] максимальный HR составил 2,46 (95% CI: 1,66–3,67, $p < 0,001$), а Monzo L, et al. [14] продемонстрировали кластер с HR 2,36 (95% CI: 2,0–2,78, $p < 0,001$). Таким образом, модели кластеризации успешно идентифицировали фенотипы со статистически значимым повышением риска неблагоприятных исходов. Примечательно, что свою эффективность в работе с кластерами показала модель RF, AUC которой составил 0,94 [14]. В исследовании [25] с использованием RF было выявлено 3 клинически значимых кластера пациентов с ХСНсФВ. Анализ кривых выживаемости Каплана-Мейера продемонстрировал статистически значимые различия в вероятности госпитализации между выделенными кластерами (лог-ранг тест 0,0021, $p < 0,05$). Визуальный анализ кривых показал, что пациенты в феногруппе 1 имели наихудший прогноз неблагоприятных исходов, в то время как феногруппа 3 ассоциировалась с наиболее благоприятным прогнозом. Несмотря на отсутствие расчета HR, исследование [25] подтверждает прогностическую ценность выявленных фенотипов.

Методологическое качество работ и риск смещения

Подавляющее большинство исследований относятся к группе с высоким риском смещения (14 из 20, 70%). Наиболее частой причиной, приведшей к такому соотношению, являлось отсутствие внешней валидации ($n=14$, 100%), и только в исследовании [18] дополнительным фактором смещения оказалось оставление пропущенных значений во время анализа предикторов в местах их пропусков. Лишь 6 исследований были оценены как имеющие низкий риск смещения, и все они характеризовались наличием внешней валидации [8, 14, 17, 21, 25, 26].

Сравнение эффективности моделей ML между группами с ХСНнФВ и ХСНсФВ, ХСНунФВ

Сравнительный анализ эффективности моделей ML между группой пациентов с ХСНнФВ ($n=7$) и объединенной группой других фенотипов ($n=13$) не

выявил статистически значимых различий в значениях AUC (критерий Манна-Уитни, $p=0,48$). Медиана AUC в обеих группах составила 0,812.

Обсуждение

Настоящий систематический обзор, включивший 20 исследований, продемонстрировал, что, несмотря на вариабельную, но в целом умеренно-высокую заявленную эффективность (AUC) моделей ML для прогнозирования исходов при ХСН, статистически значимых различий между фенотипами ХСНнФВ и ХСНсФВ/ХСНунФВ выявлено не было ($p=0,48$). Критической проблемой оказалось методологическое качество: 70% исследований имеют высокий риск систематической ошибки по PROBAST, преимущественно из-за отсутствия внешней валидации.

Отсутствие значимых различий между фенотипами ХСН, несмотря на их патофизиологическую гетерогенность, позволяет предположить, что модели ML выявляют универсальные предикторы неблагоприятных исходов (такие как сопутствующие заболевания, биохимические маркеры, принимаемые лекарственные препараты), общие для всех типов ХСН, а не специфические патофизиологические механизмы [13, 28]. Альтернативно, этот результат может отражать методологическую незрелость области и ограниченную мощность анализа вследствие малого количества исследований.

Анализ сравнительной эффективности моделей ML и традиционных шкал стратификации риска выявил противоречивую картину. В тех исследованиях, где проводилось прямое сравнение [13, 14, 21–23], модели ML демонстрировали статистически значимое превосходство над клиническими шкалами (MAGGIC, SHFM и др.), что согласуется с данными других авторов [11, 29]. Однако лишь 30% включенных в обзор работ содержали такое сравнение, что существенно ограничивает возможность формулирования окончательных выводов о клинической ценности ML-подхода.

Как отмечается в литературе, интерпретируемость моделей ML долгое время оставалась серьезным пре-

пятствием для их клинического внедрения, что породило термин "черный ящик" [22]. Однако наша оценка по PROBAST показала, что более фундаментальной проблемой является отсутствие должной процедуры внешней валидации. Хотя методы объяснимого ИИ (например, SHAP-анализ) могли бы повысить доверие клиницистов, в рассмотренных исследованиях они применялись эпизодически. Это ограничивает не только интерпретацию результатов, но и возможность выявления новых биомедицинских закономерностей. Оценка моделей обязательно должна включать по крайней мере один независимый набор данных внешнего тестирования для понимания их воспроизводимости. Более того, поскольку модели ML обычно разрабатываются на основе ретроспективных данных, важно оценивать модель в перспективных условиях, поскольку результаты лабораторных тестов могут меняться с течением времени [30]. Таким образом, выявленная нами проблема воспроизводимости, выражающаяся в распространенном отсутствии внешней валидации, является основной причиной, ограничивающей переход от многообещающих разработок к клинической практике. Этот методологический недостаток создает так называемую "иллюзию точности", когда модели демонстрируют выдающиеся результаты на внутренних данных, но их реальная эффективность в новых условиях остается недоказанной. Проблема носит системный характер для медицинского ML: как показывают исследования в других областях медицины, внешняя валидация проводится лишь в 20-30% работ [11, 31-33], но данное ограничение еще не осознано в полной мере в кардиологическом сообществе.

Для преодоления выявленных ограничений необходимо активно внедрять стандарты отчетности, такие как TRIPOD-ML, и специально разработанные списки приводящих к конечной точке первичных событий, включающих частоту госпитализаций с ХСН и смертность, связанную с ХСН. Соблюдение данных стандартов значительно повысит прозрачность, качество и, в конечном итоге, доверие к прогностическим моделям [34]. Другой проблемой является отсутствие единообразия в определении фенотипов ХСН в различных исследованиях. Если ранее ХСНсФВ определялся как ФВ ЛЖ >40% или 45%, то современные рекомендации предполагают использование порога ≥50%. Эта терминологическая неоднородность создает дополнительные сложности при сравнительном анализе и интерпретации результатов.

Практические рекомендации

Несмотря на перспективные показатели эффективности, ни одна из рассмотренных моделей ML не обладает достаточной доказательной базой для немедленного внедрения в рутинную клиническую прак-

тику в Российской Федерации. Клинiciстам стоит проявлять скептический подход к заявлениям о высокой эффективности моделей, не прошедших внешнюю валидацию.

Для научного сообщества результаты нашего обзора указывают на необходимость смены исследовательской парадигмы: приоритетом должна стать не максимальная точность на внутренних данных, а демонстрация надежности и воспроизводимости моделей. Ключевыми шагами являются соблюдение стандартов отчетности (TRIPOD-ML), обязательное планирование внешней валидации и переход к проспективным исследованиям, оценивающим реальное влияние моделей на процесс принятия клинических решений и исходы у пациентов.

Сильные стороны и ограничения

Сильными сторонами нашего исследования являются: использование валидированного инструмента PROBAST для системной оценки методологического качества, применение четких критериев PICO, сравнительный анализ эффективности моделей между разными фенотипами ХСН, включение исследований, использующих как контролируемое, так и неконтролируемое обучение.

Основными ограничениями являются: ограниченное количество исследований, что могло повлиять на мощность статистического анализа, значительная гетерогенность исследований в отношении критериев отбора пациентов и используемых предикторов, ставшая причиной исключения проведения метаанализа.

Заключение

Проведенный систематический обзор демонстрирует, что модели ML показывают умеренно-высокую эффективность в прогнозировании исходов у пациентов с ХСН, в ряде случаев превосходя традиционные шкалы. Статистически значимых различий в эффективности между моделями для ХСНсФВ/ХСНунФВ и ХСНнФВ выявлено не было. Ключевой проблемой области является отсутствие внешней валидации в большинстве исследований (70%), что подчеркивает необходимость демонстрации эффективности моделей на независимых когортах пациентов как обязательного условия их внедрения в клиническую практику.

Отношения и деятельность: все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

Декларация ИИ. Технологии генеративного ИИ использовались исключительно как вспомогательный инструмент на финальных стадиях работы над текстом рукописи.

Литература/References

- Larina VN, Kokorin VA, Larin VG, et al. Decompensated heart failure: a reconceptualization in the light of updated consensus statement of the European Society of Cardiology. *Russian Journal of Cardiology*. 2023;28(12):5581. (In Russ.) Ларина В.Н., Кокорин В.А., Ларин В.Г. и др. Декомпенсация хронической сердечной недостаточности: новый взгляд на проблему в свете обновленного консенсуса экспертов Европейского общества кардиологов. *Российский кардиологический журнал*. 2023;28(12):5581. doi:10.15829/1560-4071-2023-5581.
- Drapkina OM, Boytsov SA, Omelyanovskiy VV, et al. Socio-economic impact of heart failure in Russia. *Russian Journal of Cardiology*. 2021;26(6):4490. (In Russ.) Драпкина О.М., Бойцов С.А., Омеляновский В.В. и др. Социально-экономический ущерб, обусловленный хронической сердечной недостаточностью, в Российской Федерации. *Российский кардиологический журнал*. 2021;26(6):4490. doi:10.15829/1560-4071-2021-4490.
- Izraiq M, AlBalbissi K, Alawaisheh R, et al. Comparative Analysis of Heart Failure with Preserved Vs Reduced Ejection Fraction: Patient Characteristics, Outcomes, Mortality Prediction, and Machine Learning Model Development in the JoHF R. *Int J Gen Med*. 2024;17:3083-91. doi:10.2147/IJGM.S465388.
- Galyavich AS, Tereshchenko SN, Uskash TM, et al. 2024 Clinical practice guidelines for Chronic heart failure. *Russian Journal of Cardiology*. 2024;29(11):6162. (In Russ.) Галевич А.С., Терещенко С.Н., Ускаш Т.М. и др. Хроническая сердечная недостаточность. Клинические рекомендации 2024. *Российский кардиологический журнал*. 2024;29(11):6162. doi:10.15829/1560-4071-2024-6162.
- Bozkurt B, Coats AJS, Tsutsui H, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure: Endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association. *Eur J Heart Fail*. 2021;23(3):352-80. doi:10.1002/ehf2.1215.
- Stolfo D, Fabris E, Lund LH, et al. From mid-range to mildly reduced ejection fraction heart failure: A call to treat. *Eur J Intern Med*. 2022;103:29-35. doi:10.1016/j.ejim.2022.05.030.
- Mardanov BU, Kokozheva MA, Mamedov MN. Biochemical profile and prognosis of patients with heart failure and diabetes. *Russian Journal of Cardiology*. 2020;25(10):3841. (In Russ.) Марданов Б.У., Кокочева М.А., Мамедов М.Н. Особенности биохимического профиля и прогноз больных хронической сердечной недостаточностью и сахарным диабетом. *Российский кардиологический журнал*. 2020;25(10):3841. doi:10.15829/1560-4071-2020-3841.
- Chang CY, Chen CC, Tsai ML, et al. Predicting Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction by Using Machine Learning. *JACC Asia*. 2024;4(12):956-68. doi:10.1016/j.jacasi.2024.09.003.
- Skoularigkis S, Kourek C, Xanthopoulos A, et al. Prognostic Models in Heart Failure: Hope or Hype? *J Pers Med*. 2025;15(8):345. doi:10.3390/jpm15080345.
- Tong R, Zhu Z, Ling J. Comparison of linear and non-linear machine learning models for time-dependent readmission or mortality prediction among hospitalized heart failure patients. *Heliyon*. 2023;9(5):e16068. doi:10.1016/j.heliyon.2023.e16068.
- Shin S, Austin PC, Ross HJ, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. *ESC Heart Fail*. 2021;8(1):106-15. doi:10.1002/ehf2.13073.
- Tohyama T, Ide T, Ikeda M, et al. Machine learning-based model for predicting 1 year mortality of hospitalized patients with heart failure. *ESC Heart Fail*. 2021;8(5):4077-85. doi:10.1002/ehf2.13556.
- Park J, Zhong X, Babaie Sarjaloo F, et al. Tailored risk assessment of 90-day acute heart failure readmission or all-cause death to heart failure with preserved versus reduced ejection fraction. *Clin Cardiol*. 2022;45(4):370-8. doi:10.1002/clc.23780.
- Monzo L, Bresso E, Dickstein K, et al. Machine learning approach to identify phenotypes in patients with ischaemic heart failure with reduced ejection fraction. *Eur J Heart Fail*. 2025;27(12):3378-91. doi:10.1002/ehf2.3547.
- Ma F, Hu Y, Han P, et al. Machine Learning-Driven Prediction of One-Year Readmission in HFpEF Patients: The Key Role of Inflammation. *Clin Interv Aging*. 2025;20:1071-84. doi:10.2147/CIA.S528442.
- Wang Z, Chen X, Tan X, et al. Using Deep Learning to Identify High-Risk Patients with Heart Failure with Reduced Ejection Fraction. *J Health Econ Outcomes Res*. 2021;8(2):6-13. doi:10.36469/jheor.2021.25753.
- Hu Y, Ma F, Hu M, et al. Development and validation of a machine learning model to predict the risk of readmission within one year in HFpEF patients: Short title: Prediction of HFpEF readmission. *Int J Med Inform*. 2025;194:105703. doi:10.1016/j.ijmedinf.2024.105703.
- Ru B, Tan X, Liu Y, et al. Comparison of Machine Learning Algorithms for Predicting Hospital Readmissions and Worsening Heart Failure Events in Patients With Heart Failure With Reduced Ejection Fraction: Modeling Study. *JMIR Form Res*. 2023;7:e41775. doi:10.2196/41775.
- Glif I, Abdelhedi R, Ouarda W, et al. Oxidative stress markers-driven prognostic model to predict post-discharge mortality in heart failure with reduced ejection fraction. *Front Cardiovasc Med*. 2022;9:1017673. doi:10.3389/fcvm.2022.1017673.
- Zheng B, Liang T, Mei J, et al. Prediction of 90 day readmission in heart failure with preserved ejection fraction by interpretable machine learning. *ESC Heart Fail*. 2024;11(6):4267-76. doi:10.1002/ehf2.15033.
- McDowell K, Kondo T, Talebi A, et al. Prognostic Models for Mortality and Morbidity in Heart Failure With Preserved Ejection Fraction. *JAMA Cardiol*. 2024;9(5):457-65. doi:10.1001/jamacardio.2024.0284.
- Tian P, Liang L, Zhao X, et al. Machine Learning for Mortality Prediction in Patients With Heart Failure With Mildly Reduced Ejection Fraction. *J Am Heart Assoc*. 2023;12(12):e029124. doi:10.1161/JAHA.122.029124.
- Angraal S, Mortazavi BJ, Gupta A, et al. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. *JACC Heart Fail*. 2020;8(1):12-21. doi:10.1016/j.jchf.2019.06.013.
- Dragomiretskaya NA, Tolmacheva AV, Vetluzhskaya MV, et al. Multifactor model to predict adverse outcomes in patients with chronic heart failure with preserved and moderately reduced left ventricular ejection fraction. *Practical medicine*. 2024;22(3):48-56. (In Russ.) Драгомирецкая Н.А., Толмачева А.В., Ветлужская М.В. и др. Многофакторная модель прогнозирования неблагоприятных исходов у пациентов с хронической сердечной недостаточностью с сохраненной и умеренно сниженной фракцией выброса левого желудочка. *Практическая медицина*. 2024;22(3):48-56. doi:10.32000/2072-1757-2024-3-48-56.
- Kyodo A, Kanaoka K, Keshi A, et al. Heart failure with preserved ejection fraction phenogroup classification using machine learning. *ESC Heart Fail*. 2023;10(3):2019-30. doi:10.1002/ehf2.14368.
- Segar MW, Patel KV, Ayers C, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. 2020;22(1):148-58. doi:10.1002/ehf2.1621.
- Soltani F, Jenkins DA, Kaura A, et al. Phenogrouping heart failure with preserved or mildly reduced ejection fraction using electronic health record data. *BMC Cardiovasc Disord*. 2024;24(1):343. doi:10.1186/s12872-024-03987-9.
- Jawadi Z, He R, Srivastava PK, et al. Predicting in-hospital mortality among patients admitted with a diagnosis of heart failure: a machine learning approach. *ESC Heart Fail*. 2024;11(5):2490-8. doi:10.1002/ehf2.14796.
- Hajishah H, Kazemi D, Safaee E, et al. Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis. *BMC Cardiovasc Disord*. 2025;25(1):264. doi:10.1186/s12872-025-04700-0.
- Yang HS, Rhoads DD, Sepulveda J, et al. Building the Model. *Arch Pathol Lab Med*. 2023;147(7):826-36. doi:10.5858/arpa.2021-0635-RA.
- Smith LA, Oakden-Rayner L, Bird A, et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Digit Health*. 2023;5(12):e872-e881. doi:10.1016/S2589-7500(23)00177-2.
- Zhou Y, Ge YT, Shi XL, et al. Machine learning predictive models for acute pancreatitis: A systematic review. *Int J Med Inform*. 2022;157:104641. doi:10.1016/j.ijmedinf.2021.104641.
- Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput Methods Programs Biomed*. 2021;208:106288. doi:10.1016/j.cmpb.2021.106288.
- Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73. doi:10.7326/M14-0698.

Адреса организаций авторов: ФГБОУ ВО Воронежский государственный медицинский университет им. Н.Н. Бурденко Минздрава России, Студенческая ул., д. 10, Воронеж, 394036, Россия; ФГАОУ ВО Белгородский государственный национальный исследовательский университет, ул. Победы, д. 85, Белгород, 308007, Россия.

Addresses of the authors' institutions: Burdenko Voronezh State Medical University, Studencheskaya str., 10, Voronezh, 394036, Russia; Belgorod State National Research University, Pobedy str., 85, Belgorod, 308007, Russia.