

РАЗРАБОТКА МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИИ НА ОСНОВЕ ПОКАЗАТЕЛЕЙ КЛИНИЧЕСКОГО АНАЛИЗА КРОВИ

С.А. Захарчук, Н.А. Миронов, А.Г. Плисюк, Я.А. Орлова

Медицинский научно-образовательный институт Московского государственного университета
им. М.В. Ломоносова, Москва, Россия

Цель: разработать и оценить модель машинного обучения для прогнозирования наличия артериальной гипертензии (АГ) по данным демографических показателей и параметров клинического анализа крови (КАК), а также определить вклад ключевых предикторов с использованием интерпретируемых методов. **Материалы и методы:** проанализирована база данных 5506 пациентов, последовательно обследованных в Университетской клинике Медицинского научно-образовательного института (МНОИ) МГУ им. М.В. Ломоносова в 2021–2022 гг. (2256 мужчин и 3250 женщин, со средним возрастом $57,23 \pm 17,60$ лет). АГ выявлена у 3697 (67%) обследованных, 1809 (33%) составили группу без АГ. В качестве переменных использовали пол, возраст, показатели КАК и производные воспалительные индексы (нейтрофильно-лимфоцитарное отношение (NLR), лимфоцитарно-моноцитарное отношение (LMR), тромбоцитарно-лимфоцитарное отношение (PLR)). Сравнивали разработанную стекинг-модель и модели Lasso и Ridge-регрессии, логистическую регрессию, случайный лес, градиентный бустинг. Качество оценивали по площади под ROC-кривой (AUC). **Результаты:** AUC составила 0,75–0,77 для отдельных моделей, тогда как стекинг-модель продемонстрировала AUC 0,99 на отложенной тестовой выборке. Разница AUC между стекинг-моделью и альтернативными подходами была статистически значимой ($p < 0,001$), средний прирост относительно логистической регрессии составил 22,9 % (абсолютная разница AUC). Ведущими маркерами стали возраст и эритроцитарные индексы RDW-SD, MCV, HGB и MCH. **Заключение:** по результатам внутренней валидации использование ансамблевых методов машинного обучения с включением показателей КАК может повысить дискриминационную способность моделей прогнозирования сердечно-сосудистых заболеваний, однако полученные оценки качества требуют подтверждения при внешней валидации на независимой когорте. Возраст и эритроцитарные индексы (RDW-SD, MCV, HGB, MCH) являются наиболее информативными предикторами АГ и могут рассматриваться как доступные маркерные кандидаты для дальнейших исследований и совершенствования стратификации риска.

Ключевые слова: артериальная гипертензия, машинное обучение, клинический анализ крови, стратификация сердечно-сосудистого риска.

Для цитирования: Захарчук С.А., Миронов Н.А., Плисюк А.Г., Орлова Я.А. Разработка модели машинного обучения для прогнозирования артериальной гипертензии на основе показателей клинического анализа крови. *Южно-Российский журнал терапевтической практики*. 2026;7(3):23-33. DOI: 10.21886/2712-8156-2026-7-3-23-33.

Контактное лицо: Софья Александровна Захарчук, szakharchuk5@gmail.com.

DEVELOPMENT OF A MACHINE-LEARNING MODEL FOR PREDICTING ARTERIAL HYPERTENSION BASED ON COMPLETE BLOOD COUNT PARAMETERS

S.A. Zakharchuk, N.A. Mironov, A.G. Plisyuk, Ya.A. Orlova

Medical Research and Educational Institute, Lomonosov Moscow State University, Moscow, Russia

Objective: development and evaluation of a machine learning model for predicting arterial hypertension using demographic characteristics and complete blood count parameters, and to assess the contribution of key predictors using interpretable methods. **Materials and methods:** we analyzed a dataset of 5,506 patients examined at the University Clinic of the Medical Research and Educational Institute of Lomonosov Moscow State University in 2021–2022 (2,256 men and 3,250 women, mean age 57.23 ± 17.60 years). Arterial hypertension was identified in 3,697 (67%) individuals, while 1,809 (33%) formed the non-hypertensive group. Variables included sex, age, complete blood count parameters, and derived inflammatory indices (NLR, LMR, PLR). The proposed stacking model was compared with Lasso and Ridge regression, logistic regression, random forest,

and gradient boosting. Model performance was assessed using the AUC. **Results:** AUC ranged from 0.75 to 0.77 for classic machine learning models, whereas the stacking model achieved an AUC of 0.99. The difference in AUC between the stacking model and alternative approaches was statistically significant ($p < 0.001$), with a mean improvement of 22.9% compared with logistic regression. The leading markers were age and the erythrocyte indices RDW-SD, MCV, HGB, and MCH. **Conclusion:** based on internal validation, ensemble machine learning methods incorporating complete blood count parameters may improve the discriminative performance of cardiovascular risk prediction models; however, these performance estimates require confirmation through external validation in an independent cohort. Age and erythrocyte indices (RDW-SD, MCV, HGB, MCH) were the most informative predictors of arterial hypertension and may serve as accessible marker candidates for further research and improved risk stratification.

Keywords: arterial hypertension, machine learning, complete blood count, cardiovascular risk stratification.

For citation: Zakharchuk S.A., Mironov N.A., Plisyuk A.G., Orlova Ya.A. Development of a machine-learning model for predicting arterial hypertension based on complete blood count parameters. *South Russian Journal of Therapeutic Practice*. 2026;7(3):23-33. DOI: 10.21886/2712-8156-2026-7-3-23-33.

Corresponding author: Sofya A. Zakharchuk, szakharchuk5@gmail.com.

Введение

Ежегодные профилактические осмотры в Российской Федерации позволяют выявить классические модифицируемые факторы риска сердечно-сосудистых заболеваний (ССЗ) и проводить их коррекцию. Однако эффективность профилактических вмешательств может быть повышена путём выявления лиц с высоким риском с помощью машинного обучения (МО) при анализе медицинских данных. Современные методы МО, включая алгоритмы глубокого обучения, открывают новые возможности для создания прогностических моделей, способных обнаруживать сложные нелинейные связи между множеством признаков и заболеванием [1–4]. На примерах различных областей медицины показано, что подобные модели могут достигать качества, сопоставимого с экспертной оценкой, в том числе при анализе сложных биомедицинских данных [4]. Применение МО в кардиологии уже позволило улучшить прогнозирование риска АГ за счёт учёта большого числа клинических и лабораторных предикторов [5]. При этом важно не только достичь высокой дискриминационной способности моделей, но и обеспечить интерпретируемость результатов для клиницистов, чтобы выявленные предикторы могли быть использованы как для разработки индивидуальной профилактической траектории пациента, так и для изучения механизмов развития и прогрессирования заболеваний. Для решения подобных задач широко используют логистическую регрессию и её регуляризованные варианты (Lasso, Ridge), позволяющие получить интерпретируемые коэффициенты влияния признаков [6, 7]. Новым подходом является использование ансамблевых алгоритмов, способных повысить точность за счёт объединения разных математических подходов.

Настоящее исследование проведено для разработки подходов к повышению эффективно-

сти прогнозирования ССЗ на модели АГ, которая остаётся одной из наиболее распространённых и социально значимых неинфекционных патологий во всём мире [8–11]. Сходные тенденции отмечаются и в отношении других сердечно-сосудистых состояний, включая хроническую сердечную недостаточность [12] и осложнения системного атеросклероза, профилактика которых опирается на широкое использование лабораторных биомаркеров [13, 14]. В Российской Федерации распространённость АГ среди взрослого населения превышает 40% [15], что подтверждается данными современных клинических рекомендаций [16] и результатами все-российских скрининговых программ [17]. В качестве потенциальных маркеров АГ изучались показатели развернутого КАК, а также комплексные индексы, отражающие уровень системного воспаления [18–20], что может быть практично в условиях профилактических осмотров благодаря доступности таких показателей. Накопленные данные свидетельствуют о существенной роли иммуновоспалительных механизмов в патогенезе АГ [21, 22] и подтверждают информативность параметров клинического анализа крови у пациентов с гипертензией различной этиологии [23].

Цель исследования — разработка ансамблевой стекинг-модели для прогнозирования АГ, обладающей лучшей дискриминационной способностью в сравнении с классическими регрессионными моделями и базовыми моделями МО, а также анализ значимости ключевых признаков лучшей по качеству модели.

Материалы и методы

В ретроспективном исследовании использована база данных 5506 пациентов, последовательно проходивших обследование в Университетской клинике МНОИ МГУ им. М.В. Ломоносова в 2021–2022 гг. Исследование одобрено

Таблица / Table 1

Структура анализируемой базы данных
Structure of the analyzed dataset

Компонент базы	Данные
Демографические данные	Пол, возраст
Диагноз	АГ
Лабораторные данные	
Клинический анализ крови	Лейкоциты (WBC) Эритроциты (RBC) Гемоглобин (HGB) Гематокрит (HCT) Средний объём эритроцита (MCV) Среднее содержание гемоглобина (MCH) Средняя концентрация гемоглобина (MCHC) Тромбоциты (PLT) Тромбоцит (PCT) Ширина распределения эритроцитов (RDW-CV) Ширина распределения эритроцитов (RDW-SD) Средний объём тромбоцита (MPV) Большие тромбоциты (P-LCR) Нейтрофилы # % (Neu) Лимфоциты # % (Lym) Моноциты # % (Mon) Эозинофилы # % (EO) Базофилы # % (Bas) Нормобласты # % (NRBC) IG #, IG % Расчётные воспалительные индексы NLR, LMR, PLR

локальным этическим комитетом (ЛЭК) МНОИ МГУ им. М.В. Ломоносова (Выписка №09/20 от 14.09.2020), все участники исследования подписали форму информированного согласия. В исследование включали пациентов старше 18 лет при наличии результатов клинического анализа крови и доступности медицинской документации, достаточной для верификации диагноза АГ. Критериями не включения стали недостаточность данных, беременность и период лактации, острое инфекционное заболевание на момент обследования, установленный диагноз (психическое заболевание), алкоголизма и/или наркомании, приём кортикостероидов и/или регулярная терапия нестероидными противовоспалительными препаратами (НПВС) в предшествующие 3 месяца, активная противоопухолевая терапия в течение 6 месяцев до скрининга.

Скрининг проведён для 6983 пациентов. Пропуски в количественных признаках заменяли медианными значениями, рассчитанными на обучающей выборке. Выбросы, выходящие за пределы клинически допустимых диапазонов и расценённые как артефакты измерения, исключали, а физиологически правдоподобные экстремальные значения сохраняли в анализе.

В итоговую выборку включены данные обследований 2256 мужчин (41%) и 3250 женщин

(59%), а средний возраст участников составил $57,23 \pm 17,60$ лет. Диагноз «АГ» был установлен у 3697 пациентов (67%) в соответствии с действующими клиническими рекомендациями при стойком повышении систолического АД ≥ 140 мм рт.ст. и/или диастолического АД ≥ 90 мм рт.ст., подтверждённом на двух и более визитах, а 1809 обследованных (33%) не имели данного заболевания.

Компоненты базы данных включали демографические параметры (пол, возраст), данные КАК с воспалительными индексами (NLR, LMR, PLR) и информацию о наличии АГ. Структура базы данных представлена в таблице 1.

Характеристика сравниваемых групп представлена в таблице 2. Пациенты с АГ были значимо старше, имели более высокие значения HGB, MCH, RDW-SD и воспалительных индексов NLR и PLR, а также более низкие значения MCV ($p < 0,05$ для всех сравнений).

Были обучены регрессионные модели с регуляризацией (Lasso, Ridge), классические модели МО (логистическая регрессия, случайный лес, градиентный бустинг), а также создана ансамблевая стекинг-модель, где в качестве мета-модели использована логистическая регрессия, а базовые модели включали Random Forest, Gradient Boosting, XGBoost, CatBoost, LightGBM.

Таблица / Table 2

Характеристика пациентов сравниваемых групп
Baseline characteristics of the study groups

Показатель	АГ (n=3697)	Без АГ (n=1809)	p
Мужской пол, n (%)	1571 (42,5)	685 (37,9)	0,001
Возраст, лет	63,5±13,8	44,4±17,5	<0,001
WBC, ×10 ⁹ /л	6,8 [5,7; 8,1]	6,5 [5,5; 7,8]	0,002
RBC, ×10 ¹² /л	4,72±0,48	4,65±0,45	0,013
HGB, г/л	141,2±14,6	137,8±13,9	<0,001
HCT, %	41,9±3,8	41,0±3,6	<0,001
MCV, фл	89,1±4,9	90,3±4,4	<0,001
MCH, пг	30,1±1,9	29,7±1,8	<0,001
MCHC, г/л	337,4±10,2	336,9±9,8	0,214
PLT, ×10 ⁹ /л	245 [208; 287]	251 [214; 292]	0,038
PCT, %	0,26 [0,22; 0,30]	0,26 [0,23; 0,30]	0,412
RDW-CV, %	13,4 [12,9; 14,1]	13,1 [12,7; 13,7]	<0,001
RDW-SD, фл	43,8 [41,6; 46,3]	42,5 [40,7; 44,6]	<0,001
MPV, фл	10,4±0,9	10,3±0,9	0,157
P-LCR, %	28,6±7,3	28,1±7,1	0,089
Нейтрофилы, ×10 ⁹ /л	4,1 [3,3; 5,1]	3,8 [3,1; 4,8]	<0,001
Лимфоциты, ×10 ⁹ /л	2,0 [1,6; 2,4]	2,1 [1,7; 2,5]	0,024
Моноциты, ×10 ⁹ /л	0,52 [0,43; 0,63]	0,50 [0,41; 0,60]	0,008
Эозинофилы, ×10 ⁹ /л	0,15 [0,09; 0,23]	0,14 [0,09; 0,22]	0,341
Базофилы, ×10 ⁹ /л	0,04 [0,03; 0,06]	0,04 [0,03; 0,06]	0,672
NLR	2,05 [1,58; 2,71]	1,84 [1,45; 2,42]	<0,001
LMR	3,82 [3,01; 4,84]	4,10 [3,25; 5,12]	<0,001
PLR	124 [98; 156]	119 [95; 149]	0,003

В качестве целевой метрики выбрана AUC, поскольку способность правильно различать предсказанные классы является ключевой характеристикой многих современных медицинских тест-систем. Учитывая дисбаланс классов (67% пациентов с АГ), дополнительно рассчитывали площадь под precision-recall-кривой (PR-AUC), а также чувствительность, специфичность и F1-меру при пороге классификации, определённом по индексу Йодена на обучающей выборке. Калибровку моделей оценивали с помощью калибровочных кривых и показателя Брайера.

Данные предварительно разделяли на обучающую и тестовую выборки в соотношении 80:20 до выполнения операций, способных использовать распределение признаков или информацию об исходе. На обучающей выборке проводили настройку гиперпараметров с использованием стратифицированной k-fold кросс-валидации. Для формирования признаков второго уровня стекинг-модели использовали только out-of-fold-предсказания базовых алгоритмов. После получения предсказаний для всей обучающей выборки на них обучали мета-модель. Для тестовой выборки базовые модели переобучали на всей обучающей выборке, после чего их предсказания передавали в мета-модель.

Качество модели оценивалось на отложенной тестовой выборке. Тестовая выборка не использовалась для подбора гиперпараметров, отбора признаков, промежуточного обучения и оценки промежуточных результатов.

Использовали предсказания базовых моделей в качестве входа мета-модели, что предотвращает утечку данных. Методом bootstrap рассчитывались распределения AUC для каждой модели, на основе которых определялись среднее значение, стандартная ошибка и 95%-ный доверительный интервал. Для оценки различий между моделями рассчитаны разница средних AUC и t-критерий для независимых переменных. Эффективной считалась модель, у которой прирост AUC составил 5% и p-уровень значимости <0,05. Выполнено сравнение моделей при помощи графиков плотности распределения и диаграмм размаха с доверительными интервалами.

Для эффективных моделей построены и охарактеризованы графики зависимости пяти наиболее значимых факторов и целевой переменной.

Результаты

В ходе проведённого комплексного анализа эффективности различных методов МО и ре-

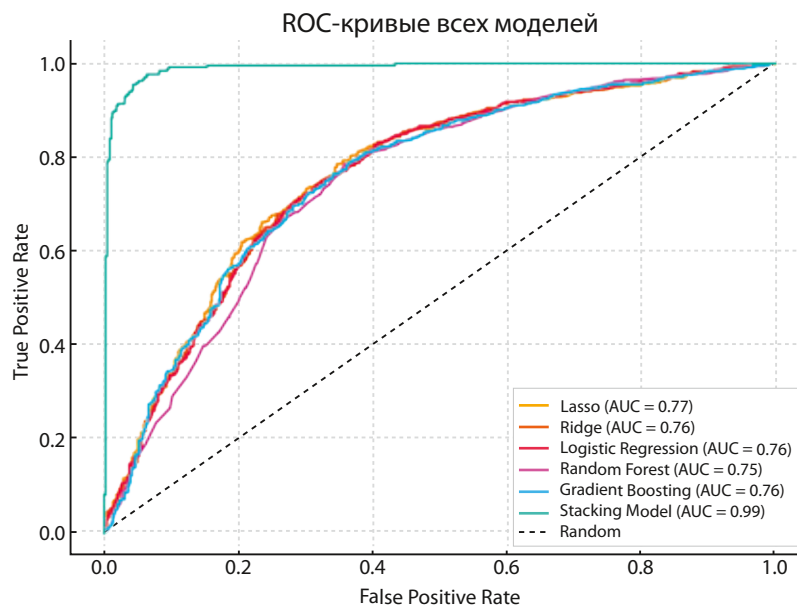


Рисунок 1. ROC-кривые всех моделей
Figure 1. ROC curves of all models

Таблица / Table 3

Метрики качества моделей на тестовой выборке
Performance metrics of the models on the test set

Модель	AUC (95% ДИ)	Чувствительность	Специфичность	F1	PR-AUC	Brier
Lasso-регрессия	0,77 (0,74–0,80)	0,72	0,69	0,78	0,86	0,192
Ridge-регрессия	0,76 (0,73–0,79)	0,71	0,68	0,77	0,85	0,195
Логистическая регрессия	0,76 (0,73–0,79)	0,71	0,68	0,77	0,85	0,196
Случайный лес	0,75 (0,72–0,78)	0,70	0,67	0,76	0,84	0,201
Градиентный бустинг	0,76 (0,73–0,79)	0,71	0,69	0,77	0,85	0,194
Стекинг-модель	0,99 (0,98–0,99)	0,95	0,94	0,96	0,99	0,045

грессионного моделирования были сопоставлены показатели качества классификации АГ. Для каждой из моделей были построены ROC-кривые и рассчитаны значения AUC. Данные были разделены на обучающую и тестовую выборки. Подбор гиперпараметров и обучение стекинг-модели выполнялись только на обучающей части с использованием кросс-валидации. В результате Lasso-регрессия показала AUC 0,77, Ridge и классическая логистическая регрессия по 0,76, случайный лес — 0,75, градиентный бустинг — 0,76, тогда как стекинг-ансамбль продемонстрировал AUC 0,99 (рис. 1). При интерпретации столь высокого значения AUC учитывалась необходимость исключения возможной утечки данных и корректной процедуры валидации стекинг-модели. Дополнительные метрики качества моделей на тестовой выборке представлены в таблице 3. Стекинг-модель продемонстрировала чувствительность 0,95 и специ-

фичность 0,94 при пороге, определённом по индексу Йодена, а также наилучшую калибровку. PR-AUC стекинг-модели составила 0,99 против 0,84–0,86 у остальных моделей.

Методом bootstrap рассчитаны распределения и сформированы гистограммы распределений обученных моделей (рис. 2).

Статистический парный анализ AUC между стекинг-моделью и каждой из других подтвердил значимый прирост AUC по сравнению со всеми моделями, как продемонстрировано в таблице 4. Средний прирост составил 22,9 % AUC (абсолютная разница) при $p < 0,001$.

Данные результаты демонстрируют, что объединение прогнозов нескольких базовых моделей в стекинг-ансамбль позволяет значительно повысить способность модели различать классы. Потенциальное преимущество стекинг-подхода объясняется тем, что разные модели фокусируются на различных признаках и ошибках. В

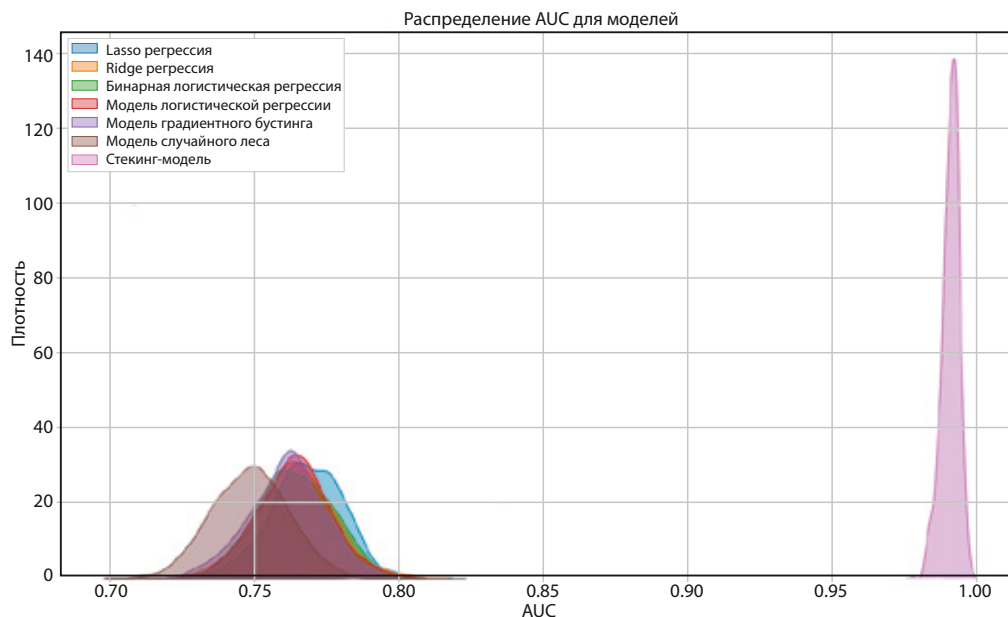


Рисунок 2. Гистограмма распределения AUC для моделей
Figure 2. Histogram of AUC distribution across models

Таблица / Table 4

Эффективность моделей для предсказания АГ
Performance of models for AH prediction

Модель 1	Модель 2	Δ AUC	p-уровень значимости
Lasso регрессия	Стекинг модель	-0,2223	<0,001
Ridge регрессия	Стекинг модель	-0,2265	<0,001
Бинарная логистическая регрессия	Стекинг модель	-0,2267	<0,001
Модель логистической регрессии	Стекинг модель	-0,227	<0,001
Модель градиентного бустинга	Стекинг модель	-0,2296	<0,001
Модель случайного леса	Стекинг модель	-0,242	<0,001

результате комбинации слабых и средних классификаторов удаётся нивелировать их индивидуальные ограничения и совместно извлекать информацию из всех доступных переменных.

Особое внимание было уделено анализу значимости признаков, наиболее сильно влияющих на предсказание АГ в итоговой стекинг-модели, как продемонстрировано в таблице 5.

На первом месте по показателю перестановочной важности оказался возраст пациента. Это соответствует хорошо известным эпидемиологическим данным: с увеличением возраста повышается риск развития гипертензии в связи с возраст-ассоциированными изменениями сосудистой стенки и снижением эластичности артерий. Анализ зависимости предсказанной вероятности АГ от возраста показал, что в молодом и среднем возрасте (примерно до 50 лет) предсказанная вероятность остаётся низкой и даже демонстрирует умеренное снижение, после чего начиная примерно с 55–60 лет наблюдает-

ся устойчивый рост (рис. 3). В интервале 60–75 лет вероятность увеличивается постепенно (до 0,10–0,12), а в возрасте около 80 лет отмечается резкий скачок предсказанной вероятности до 0,20 с достижением максимума порядка 0,22–0,23 в диапазоне 80–85 лет.

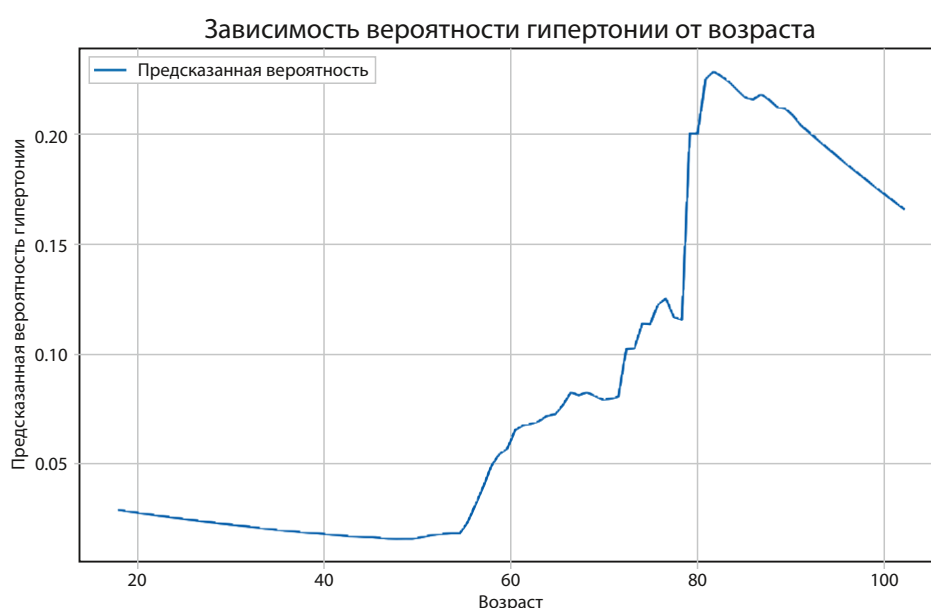
Среди гематологических маркеров вторую по значимости роль играет ширина распределения эритроцитов (RDW-SD). Этот параметр отражает степень анизоцитоза — вариабельность размеров эритроцитов. Повышенные значения RDW-SD в ряде исследований связывают с хроническим воспалением, окислительным стрессом и ухудшением функции эндотелия сосудов, что может коррелировать с риском АГ. Наш анализ подтвердил, что при росте RDW-SD предсказанная вероятность наличия АГ также возрастает.

Третьим по важности маркером выступил средний объём эритроцитов (MCV). В полученных данных снижение уровня MCV ассоциировалось с наличием АГ, что может отражать вклад состояний,

Таблица / Table 5

Коэффициенты признаков стекинг-модели для предсказания АГ
Permutation importance of features in the stacking model for AH prediction

Признак	Коэффициент перестановочной важности
Возраст	0,230
Ширина распределения эритроцитов (RDW-SD)	0,012
Средний объём эритроцита (MCV)	0,011
Гемоглобин (HGB)	0,009
Среднее содержание гемоглобина (MCH)	0,005

**Рисунок 3.** Зависимость вероятности АГ от возраста**Figure 3.** Predicted probability of AH as a function of age

сопровождающихся микроцитозом, в профиль риска. Предполагаемые механизмы могут включать изменение реологических свойств крови и повышение вязкости, что потенциально влияет на периферическое сопротивление сосудов.

На четвёртом месте по важности расположена концентрация гемоглобина (HGB). Увеличение уровня гемоглобина может повышать вязкость крови и периферическое сопротивление сосудов, что потенциально связано с ростом артериального давления (АД).

Пятым по значимости признаком является среднее содержание гемоглобина в эритроците (MCH). Рост уровня MCH отражает увеличение массы гемоглобина в каждой клетке и может быть связан с изменениями реологии крови в сочетании с другими показателями эритроцитарного звена.

Зависимость предсказанной вероятности АГ от перечисленных эритроцитарных индексов представлена на рисунке 4.

Обсуждение

Полученные данные подтверждают ключевую роль возраста как фактора риска развития АГ. Этот вывод согласуется с подавляющим большинством эпидемиологических исследований [24]. Аналогично в 11-летней проспективной модели риска (HUNT, Норвегия) возраст вошёл в число важнейших предикторов будущего развития АГ наряду с исходным уровнем АД и индексом массы тела (ИМТ) [25]. Выявленная в нашем исследовании зависимость риска АГ от возраста полностью соответствует существующим научным данным и подчёркивает необходимость учёта возраста при стратификации риска.

Особый интерес представляют гематологические показатели красной крови. Наши результаты демонстрируют, что показатели эритроцитарного звена (включая RDW-SD, MCV, HGB и MCH) внесли вклад в прогнозирование наличия АГ. Ряд работ указывает, что повышенный уро-

Зависимость вероятности гипертензии от гематологических маркеров

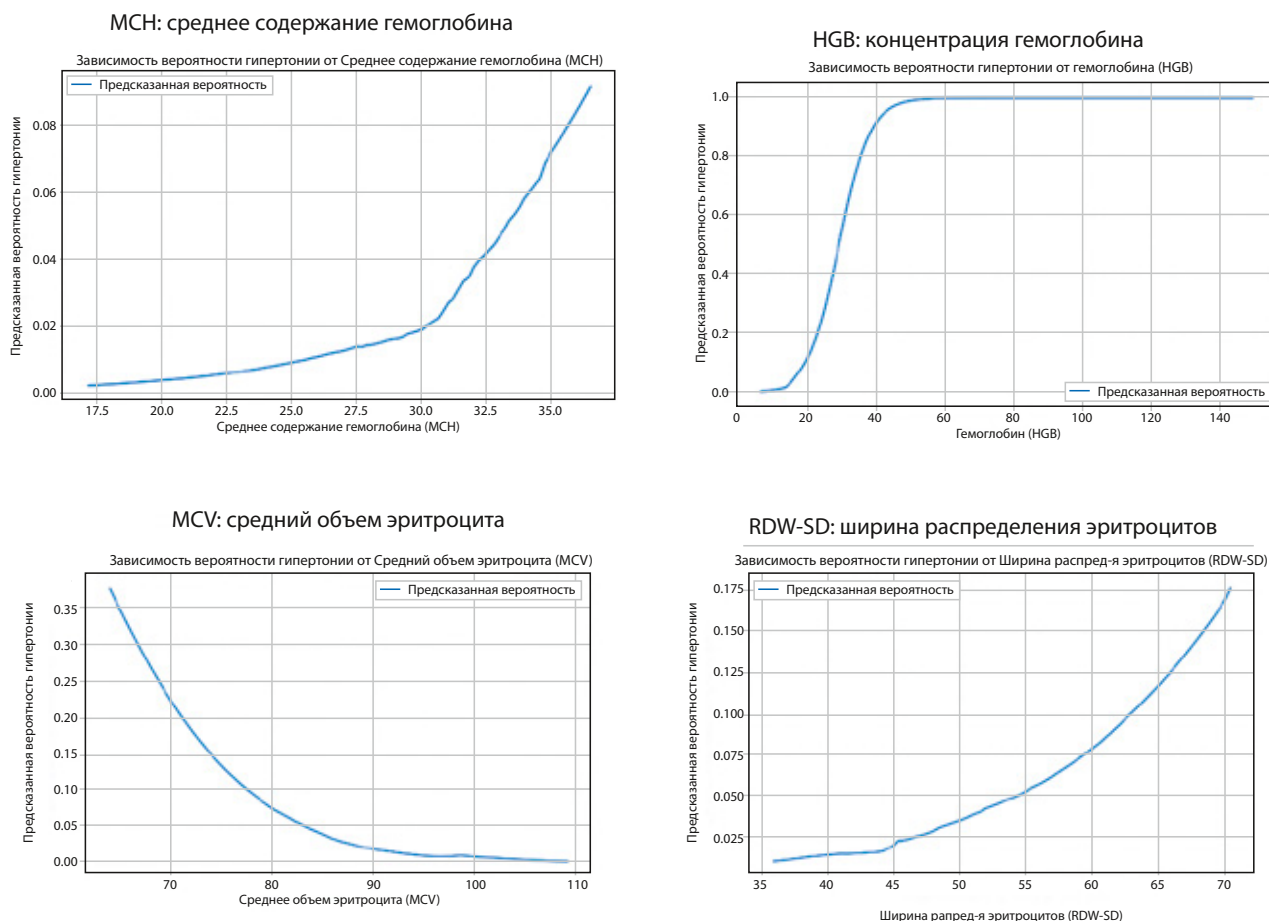


Рисунок 4. Зависимость вероятности АГ от параметров КАК
Figure 4. Predicted probability of AH as a function of CBC parameters

вень гемоглобина и гематокрита может способствовать гипертензии за счёт увеличения вязкости крови и периферического сопротивления сосудов [26]. Одновременно с этим обсуждается роль окислительного стресса и активных форм кислорода как универсальных механизмов сосудистого ремоделирования и эндотелиальной дисфункции [27]. Например, в крупной работе (MASHAD, 9808 участников) у лиц с АГ наблюдались более высокие уровни HGB и MCH, тогда как уровни MCV и RDW-SD у них были несколько ниже, чем в контрольной группе [28]. Авторы связывают это с гемоконцентрацией и особенностями эритропоэза при АГ, отмечая при этом, что именно уровень HCT оказался независимым предиктором гипертензии в многофакторной модели [28]. Повышенная вариабельность размеров эритроцитов считается маркером хронического воспаления и дисфункции костного мозга, что может сопутствовать кардиометаболическим нарушениям [29].

В крупном популяционном анализе данных NHANES за 1999–2018 гг. также подтверждена ассоциация повышенных значений RDW с распространённостью АГ среди взрослых [30], а в недавнем исследовании показано, что комбинации эритроцитарных характеристик оказывают совместное влияние на риск развития гипертензии [31]. В российском исследовании среди молодых мужчин с АГ первой степени у 37% пациентов выявлено повышение уровня RDW-SD. У таких пациентов фиксировали более высокое АД, больший ИМТ, нарушения липидного профиля и эндотелиальную дисфункцию по сравнению с лицами с нормальным уровнем RDW-SD [32]. Авторы делают вывод, что повышенный уровень RDW-SD может служить ранним маркером неблагоприятного течения гипертензии у молодых пациентов. Таким образом, наши результаты о значимости RDW-SD подтверждаются рядом исследований,

хотя окончательный механизм этой связи остаётся предметом обсуждения.

Дополнительный интерес представляет совместный учёт нескольких эритроцитарных индексов, поскольку их комбинации могут повышать прогностическую информативность по сравнению с использованием одного показателя. В частности, показано, что совместный учёт RDW-SD и MCV улучшает прогнозирование осложнений гипертонии. Добавление MCV усиливает связь RDW-SD с поражением органов-мишеней [26]. Недавно Yu S. и соавт. предложили интегральный показатель — произведение RDW-SD и MCV. Было установлено, что его высокие значения ассоциированы с гипертрофией миокарда левого желудочка у пациентов с АГ [33]. Данный факт согласуется с тем, что в нашей модели MCV и MCH вошли в число значимых предикторов, и может указывать на вклад изменений морфологии эритроцитов в профиль сердечно-сосудистого риска. При этом выявленные ассоциации следует интерпретировать как маркерные, а не причинно-следственные, поскольку эритроцитарные индексы могут отражать сопутствующие метаболические и воспалительные процессы, влияющие на ремоделирование сердца и сосудов.

В совокупности выявленные нами ассоциации (возраст, RDW-SD, MCV, HGB, MCH) во многом совпадают с данными литературы. Наше исследование дополняет доказательства в пользу того, что эритроцитарные индексы являются доступными маркерами, способными улучшить прогнозирование риска АГ, наряду с традиционными факторами.

Ограничения исследования

К ограничениям настоящей работы следует отнести её ретроспективный характер, монокентрический дизайн и отсутствие внешней валидации разработанной модели на независимой когорте. Полученное значение $AUC=0,99$ значительно превышает результаты базовых моделей (0,75–0,77), что требует тщательной проверки на возможную утечку данных через out-of-fold-предсказания мета-модели и подтверждения в проспективных исследованиях с независимыми выборками для исключения возможного переобучения и оценки переносимости модели на иные клинические популяции, в том числе с учётом возможных популяционных и этнических различий в течении АГ [34].

Заключение

Применение современных методов МО способно повысить точность прогнозирования ССЗ, особенно если в анализ вовлечено большое число новых предикторов. Мы продемонстрировали значимость включения параметров КАК и разработали стекинг-модель, показавшую более высокую дискриминационную способность, по сравнению с классическими подходами, на внутренней тестовой выборке. Учитывая величину полученного значения $AUC(0,99)$ и отсутствие внешней валидации, это преимущество следует рассматривать как предварительное до подтверждения на независимой когорте. Сравнение с литературными данными показывает, что полученная в нашей работе оценка $AUC(0,99)$ формально находится на уровне верхней границы опубликованных значений (0,77–0,92) [5], однако прямое сопоставление ограничено различиями в дизайне, популяции и схемах валидации, а окончательное суждение о сравнительной эффективности возможно только после внешней проверки на независимой когорте. В то же время, как и отмечалось другими авторами, прирост качества относительно простой логистической регрессии в нашем случае составил 22,9%, что соотносится с ожидаемым диапазоном улучшений при использовании ансамблевых методов. Для оценки вклада отдельных предикторов в прогнозирование использовались методы интерпретации на основе перестановочной важности признаков, которые подтвердили ведущую роль возраста и высокую значимость RDW-SD, а также вклад MCV, HGB и MCH. Полученные результаты согласуются с современной концепцией персонализированной медицины [35] и могут способствовать совершенствованию существующих подходов к стратификации сердечно-сосудистого риска [36, 37]. Таким образом, полученные данные вписываются в контекст существующих научных работ, а выявленные нами предикторы (возраст, RDW-SD, MCV, HGB, MCH) могут рассматриваться как перспективные маркерные кандидаты для дальнейших исследований механизмов, ассоциированных с развитием АГ.

Финансирование. Исследование выполнено в рамках государственного задания Московского государственного университета им. М.В. Ломоносова.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА/REFERENCES

1. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*. 2016;375(13):1216-1219. DOI: 10.1056/NEJMp1606181
2. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med*. 2017;376(26):2507-2509. DOI: 10.1056/NEJMp1702071
3. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246. DOI: 10.1093/bib/bbx044
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. Erratum in: *Nature*. 2017;546(7660):686. DOI: 10.1038/nature21056
5. Zhao H, Zhang X, Xu Y, Gao L, Ma Z, Sun Y, et al. Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. *Front Public Health*. 2021;9:619429. DOI: 10.3389/fpubh.2021.619429
6. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol*. 2001;54(10):979-985. DOI: 10.1016/s0895-4356(01)00372-9
7. Boateng EY, Abaye DA. A Review of the Logistic Regression Model with Emphasis on Medical Research. *J Data Anal Inf Process*. 2019;7(4):190-207. DOI: 10.4236/jdaip.2019.74012
8. Amini M, Zayeri F, Salehi M. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *BMC Public Health*. 2021;21(1):401. DOI: 10.1186/s12889-021-10429-0
9. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation*. 2019;139(10):e56-e528. Erratum in: *Circulation*. 2020;141(2):e33. DOI: 10.1161/CIR.0000000000000659.
10. Ma LY, Chen WW, Gao RL, Liu LS, Zhu ML, Wang YJ, et al. China cardiovascular diseases report 2018: an updated summary. *J Geriatr Cardiol*. 2020;17(1):1-8. DOI: 10.11909/j.issn.1671-5411.2020.01.001
11. Coronado F, Melvin SC, Bell RA, Zhao G. Global Responses to Prevent, Manage, and Control Cardiovascular Diseases. *Prev Chronic Dis*. 2022;19:E84. DOI: 10.5888/pcd19.220347
12. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2021;42(36):3599-3726. Erratum in: *Eur Heart J*. 2021;42(48):4901. DOI: 10.1093/eurheartj/ehab368.
13. Morrow DA, Cannon CP, Jesse RL, Newby LK, Ravkilde J, Storrow AB, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes. *Circulation*. 2007;115(13):e356-75. DOI: 10.1161/CIRCULATIONAHA.107.182882
14. Dobesh PP, Finks SW, Trujillo TC. Dual Antiplatelet Therapy for Long-term Secondary Prevention of Atherosclerotic Cardiovascular Events. *Clin Ther*. 2020;42(10):2084-2097. DOI: 10.1016/j.clinthera.2020.08.003
15. Ерина А.М., Ротарь О.П., Солнцев В.Н., Шальнова С.А., Деев А.Д., Баранова Е.И., и др. Эпидемиология артериальной гипертензии в Российской Федерации – важность выбора критериев диагностики. *Кардиология*. 2019;59(6):5-11. Erina A.M., Rotar O.P., Solntsev V.N., Shalnova S.A., Deev A.D., Baranova E.I., et al. Epidemiology of Arterial Hypertension in Russian Federation – Importance of Choice of Criteria of Diagnosis. *Kardiologiya*. 2019;59(6):5-11. (In Russ.) DOI: 10.18087/cardio.2019.6.2595
16. Кобалава Ж.Д., Конради А.О., Недогода С.В., Шляхто Е.В., Арутюнов Г.П., Баранова Е.И., и др. Артериальная гипертензия у взрослых. Клинические рекомендации 2024. *Российский кардиологический журнал*. 2024;29(9):6117. Kobalava Zh.D., Konradi A.O., Nedogoda S.V., Shlyakhto E.V., Arutyunov G.P., Baranova E.I., et al. 2024 Clinical practice guidelines for Hypertension in adults. *Russian Journal of Cardiology*. 2024;29(9):6117. (In Russ.) DOI: 10.15829/1560-4071-2024-6117
17. Ротарь О.П., Ильянова И.Н., Бояринова М.А., Могучая Е.В., Толкунова К.М., Дьячков В.А., и др. Результаты Все-российского скрининга артериальной гипертензии 2023. *Российский кардиологический журнал*. 2024;29(5):5931. Rotar O.P., Ilyanova I.N., Boyarinova M.A., Moguchaya E.V., Tolkunova K.M., Dyachkov V.A., et al. 2023 All-Russian screening for hypertension: results. *Russian Journal of Cardiology*. 2024;29(5):5931. (In Russ.) DOI: 10.15829/1560-4071-2024-5931
18. Layton AT. AI, Machine Learning, and ChatGPT in Hypertension. *Hypertension*. 2024;81(4):709-716. DOI: 10.1161/HYPERTENSIONAHA.124.19468
19. Islam SMS, Talukder A, Awal MA, Siddiqui MMU, Ahamed MM, Ahamed B, et al. Machine Learning Approaches for Predicting Hypertension and Its Associated Factors Using Population-Level Data From Three South Asian Countries. *Front Cardiovasc Med*. 2022;9:839379. DOI: 10.3389/fcvm.2022.839379
20. Khamissi FZ, Sun L, Johnson P, Shah S, Benjamin IJ. Machine Learning and Artificial Intelligence for Research on Hypertension. *Am J Hypertens*. 2025;38(11):867-871. DOI: 10.1093/ajh/hpaf051
21. Guzik TJ, Nosalaki R, Maffia P, Drummond GR. Immune and inflammatory mechanisms in hypertension. *Nat Rev Cardiol*. 2024;21(6):396-416. DOI: 10.1038/s41569-023-00964-1
22. Мордовин В.Ф., Зюбанова И.В., Манукян М.А., Доржиева И.К., Вторушина А.А., Хунхинова С.А., и др. Роль иммунно-воспалительных механизмов в патогенезе артериальной гипертензии. *Сибирский журнал клинической и экспериментальной медицины*. 2023;38(1):21-27. Mordovin V.F., Zyubanova I.V., Manukyan M.A., Dorzhieva I.K., Vtorushina A.A., Khunkhinova S.A., et al. The role of immune-inflammatory mechanisms in the pathogenesis of hypertension. *Siberian Journal of Clinical and Experimental Medicine*. 2023;38(1):21-27. (In Russ.) DOI: 10.29001/2073-8552-2023-38-1-21-27
23. Поселюгина О.Б., Коричкина Л.Н., Стеблецова Н.И., Бородина В.Н., Данилина К.С., Маслов А.Н., и др. Показатели клинического анализа крови у больных с эссенциальной и ренопаренхиматозной артериальной гипертензией. *Трудный пациент*. 2021;19(2):14-17. Poselyugina O.B., Korichkina L.N., Stebletsova N.I., Borodina V.N., Danilina K.S., Maslov A.N., et al. Clinical Blood Test Indicators in Patients with Essential and Renal Parenchymal Arterial Hypertension. *Trudnyj Pacient = Difficult Patient*. 2021;19(2):14-17. (In Russ.) DOI: 10.224412/2074-1005-2021-2-14-17

24. Бойцов С.А., Драпкина О.М., Шляхто Е.В., Конради А.О., Баланова Ю.А., Жернакова Ю.В., и др. Исследование ЭССЕ-РФ (Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской Федерации). Десять лет спустя. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):3007. Boytsov S.A., Drapkina O.M., Shlyakhto E.V., Konradi A.O., Balanova Yu.A., Zhernakova Yu.V., et al. Epidemiology of Cardiovascular Diseases and their Risk Factors in Regions of Russian Federation (ESSE-RF) study. Ten years later. *Cardiovascular Therapy and Prevention*. 2021;20(5):3007. (In Russ.) DOI: 10.15829/1728-8800-2021-3007
25. Schjerven FE, Ingeström EML, Steinsland I, Lindseth F. Development of risk models of incident hypertension using machine learning on the HUNT study data. *Sci Rep*. 2024;14(1):5609. DOI: 10.1038/s41598-024-56170-7
26. Zhan YL, Zou B, Kang T, Xiong LB, Zou J, Wei YF. Multiplicative interaction between mean corpuscular volume and red cell distribution width with target organ damage in hypertensive patients. *J Clin Lab Anal*. 2017;31(5):e22082. DOI: 10.1002/jcla.22082
27. Forrester SJ, Kikuchi DS, Hernandez MS, Xu Q, Griendling KK. Reactive Oxygen Species in Metabolic and Inflammatory Signaling. *Circ Res*. 2018;122(6):877-902. DOI: 10.1161/CIRCRESAHA.117.311401
28. Emamian M, Hasanian SM, Tayefi M, Bijari M, Movahedian Far F, Shafiee M, et al. Association of hematocrit with blood pressure and hypertension. *J Clin Lab Anal*. 2017;31(6):e22124. DOI: 10.1002/jcla.22124
29. Danese E, Lippi G, Montagnana M. Red blood cell distribution width and cardiovascular diseases. *J Thorac Dis*. 2015;7(10):E402-11. DOI: 10.3978/j.issn.2072-1439.2015.10.04
30. Chen Y, Hou X, Zhong J, Liu K. Association between red cell distribution width and hypertension: Results from NHANES 1999-2018. *PLoS One*. 2024;19(5):e0303279. DOI: 10.1371/journal.pone.0303279
31. He Z, Chen Z, Wang Y, Qin H, Wu W, Fu P, et al. Individual and joint effects of red blood cell traits on hypertension: a longitudinal analysis. *Eur J Prev Cardiol*. 2026;33(7):1140-1150. DOI: 10.1093/eurjpc/zwaf093
32. Миронов Н.А., Приезжев А.В., Свешникова А.Н., Луговцов А.Е., Каранадзе Н.А., Дячук Л.И., и др. Связь изменений микрореологии крови, системы гемостаза и функционального статуса пациентов с хронической сердечной недостаточностью: обоснование и протокол исследования. *Кардиологический вестник*. 2024;19(1):79-83. Mironov N.A., Priezzhev A.V., Sveshnikova A.N., Lugovtsov A.E., Karanadze N.A., Dyachuk L.I., et al. Correlation between blood rheology, hemostasis and functional status in patients with chronic heart failure: rationale and study protocol. *Russian Cardiology Bulletin*. 2024;19(1):79-83. (In Russ.) DOI: 10.17116/Cardiobulletin20241901179
33. Yu S, Xiong L, Wei D, Zhu H, Cai X, Shao L, et al. Prediction of the left ventricular mass index in hypertensive patients using the product of red cell distribution width and mean corpuscular volume. *Medicine (Baltimore)*. 2024;103(14):e37685. DOI: 10.1097/MD.00000000000037685
34. Williams SK, Ravenell J, Seyedali S, Nayef S, Ogedegbe G. Hypertension Treatment in Blacks: Discussion of the U.S. Clinical Practice Guidelines. *Prog Cardiovasc Dis*. 2016;59(3):282-288. DOI: 10.1016/j.pcad.2016.09.004
35. Sheikhy A, Fallahzadeh A, Aghaei Meybodi HR, Hasanzad M, Tajdini M, Hosseini K. Personalized medicine in cardiovascular disease: review of literature. *J Diabetes Metab Disord*. 2021;20(2):1793-1805. DOI: 10.1007/s40200-021-00840-0
36. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140(11):e563-e595. Erratum in: *Circulation*. 2019;140(11):e647-e648. Erratum in: *Circulation*. 2020;141(4):e59. Erratum in: *Circulation*. 2020;141(16):e773. DOI: 10.1161/CIR.0000000000000677
37. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2021;42(34):3227-3337. Erratum in: *Eur Heart J*. 2022;43(42):4468. DOI: 10.1093/eurheartj/ehab484

Информация об авторах

Захарчук Софья Александровна, студент, факультет фундаментальной медицины, Медицинский научно-образовательный институт Московского государственного университета им. М.В. Ломоносова, Москва, Россия, ORCID: <https://orcid.org/0009-0008-8220-351X>, szakharchuk5@gmail.com.

Миронов Никита Александрович, к. м. н., ассистент кафедры терапии, старший научный сотрудник Центра регенеративной медицины, Медицинский научно-образовательный институт Московского государственного университета им. М.В. Ломоносова, Москва, Россия, ORCID: <https://orcid.org/0000-0001-6729-4371>, nikimir29@mail.ru.

Плисюк Алина Геннадьевна, к. м. н., доцент кафедры терапии, Медицинский научно-образовательный институт Московского государственного университета им. М.В. Ломоносова, Москва, Россия, ORCID: <https://orcid.org/0000-0003-2015-4712>, apl.cardio@yandex.ru.

Орлова Яна Артуровна, д. м. н., проф., заведующий отделом возраст-ассоциированных заболеваний, Медицинский научно-образовательный институт Московского государственного университета им. М.В. Ломоносова, Москва, Россия, ORCID: <https://orcid.org/0000-0002-8160-5612>, 5163002@bk.ru.

Information about the authors

Sofya A. Zakharchuk, student, Faculty of Medicine, Medical Research and Educational Institute, Lomonosov Moscow State University, Moscow, Russia, ORCID: <https://orcid.org/0009-0008-8220-351X>, szakharchuk5@gmail.com.

Nikita A. Mironov, Cand. Sci. (Med.), Assistant Professor, Department of Therapy; Senior Researcher, Center for Regenerative Medicine, Medical Research and Educational Institute, Lomonosov Moscow State University, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-6729-4371>, nikimir29@mail.ru.

Alina G. Plisyuk, Cand. Sci. (Med.), Associate Professor, Department of Therapy, Medical Research and Educational Institute, Lomonosov Moscow State University, Moscow, Russia, ORCID: <https://orcid.org/0000-0003-2015-4712>, apl.cardio@yandex.ru.

Yana A. Orlova, Dr. Sci. (Med.), Professor, Head of the Department of Age-Associated Diseases, Medical Research and Educational Institute, Lomonosov Moscow State University, Moscow, Russia, ORCID: <https://orcid.org/0000-0002-8160-5612>, 5163002@bk.ru.

Получено / Received: 13.07.2026

Принято к печати / Accepted: 21.08.2026