

Вареха Л.А.¹, Кокорин В.А.^{1,5}, Святов И.С.², Тяжельников А.А.^{2,5}, Дорожкова В.С.²,
Гимадиев Р.Р.^{1,3,4,5}, Стуклов Н.И.¹, Кочетов А.Г.⁵, Гордиенко К.В.⁴, Губина Е.В.³, Вареха Н.В.¹,
Чаусова С.В.⁵, Щеголев О.Б.³, Силкина Т.А.⁶



<https://elibrary.ru/hhmfeb>

РАСЧЕТ КОНЦЕНТРАЦИИ ХОЛЕСТЕРИНА ЛИПОПРОТЕИНОВ НИЗКОЙ ПЛОТНОСТИ НА ОСНОВЕ ДАННЫХ ЛИПИДНОГО ПРОФИЛЯ С ПРИМЕНЕНИЕМ МАШИННОГО ОБУЧЕНИЯ

¹ФГАОУ ВО «Российский университет дружбы народов имени Патрика Лумумбы», 117198, Москва, Россия;

² ГБУЗ «Городская поликлиника №220» Департамента здравоохранения г. Москвы», 123022, Москва, Россия;

³ ООО «ЛАБХАБ», 119002, Москва, Россия;

⁴ ГНЦ РФ – ИМБП РАН, 123007, Москва, Россия;

⁵ ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава РФ (Пироговский Университет), 117513, Москва, Россия;

⁶ ООО «Лаборатория Гемотест», 123112, Москва, Россия

В условиях реализации стратегии РФ по развитию цифровых технологий в медицине актуально изучение возможностей искусственного интеллекта (ИИ) и машинного обучения (МО) в клинической лабораторной диагностике.

Цель исследования – создать модель МО для расчета концентрации холестерина липопротеинов низкой плотности на основе показателей липидного профиля и оценить ее точность относительно традиционных прямого и расчётных методов.

Материал и методы. Обучение и тестирование модели проведено на выборке из 631 267 пациентов с использованием обезличенного набора данных, включающего демографические характеристики и результаты показателей липидограммы: концентрации общего холестерина, холестерина липопротеинов высокой плотности, триглицеридов и холестерина липопротеинов низкой плотности.

Результаты. Модель регрессии показала высокую точность в прогнозе уровня холестерина липопротеинов низкой плотности с $R^2 = 0,952$, превосходя традиционные формулы для пациентов при уровнях триглицеридов более 4,5 ммоль/л и холестерина липопротеинов низкой плотности более 3,0 ммоль/л: средняя абсолютная ошибка между предсказанным и известным значением холестерина липопротеинов низкой плотности составила 0,476 и 0,168 против 1,173, 0,562, 0,796 и 0,187, 0,173, 0,200 ммоль/л, рассчитанным по формулам Фридвальда, Мартина и Сэмпсона, соответственно.

Заключение. Разработанная модель, обученная и протестированная на крупной российской выборке пациентов, по точности расчета концентрации холестерина липопротеинов низкой плотности превосходит традиционные формулы при высоких уровнях триглицеридов и холестерина липопротеинов низкой плотности, и не уступает при иных диапазонах, что подтверждает надежность регрессионной модели для клинического использования в условиях отсутствия прямого измерения концентрации холестерина липопротеинов низкой плотности.

Ключевые слова: машинное обучение; искусственный интеллект; холестерин липопротеинов низкой плотности; уравнение Фридвальда; уравнение Мартина; уравнение Сэмпсона

Для цитирования: Вареха Л.А., Кокорин В.А., Святов И.С. Тяжельников А.А., Дорожкова В.С., Гимадиев Р.Р., Стуклов Н.И., Кочетов А.Г., Гордиенко К.В., Губина Е.В., Вареха Н.В., Чаусова С.В., Щеголев О.Б., Силкина Т.А. Расчет концентрации холестерина липопротеинов низкой плотности на основе данных липидного профиля с применением машинного обучения. *Клиническая лабораторная диагностика.* 2026; 71 (2): 133-138

DOI: <https://doi.org/10.51620/0869-2084-2026-71-2-133-138>

EDN: HHMFEB

Для корреспонденции: Вареха Любовь Андреевна, аспирант каф. госпитальной терапии с курсами эндокринологии, гематологии и КЛД; ассистент кафедры гистологии, цитологии и эмбриологии МИ РУДН им. Патрика Лумумбы; e-mail: eykhenvald-la@rudn.ru

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Финансирование. Исследование не имело спонсорской поддержки.

Поступила 25.09.2025

Принята к печати 31.12.2025

Опубликовано 01.02.2026

Varekha L.A.¹, Kokorin V.A.^{1,5}, Sviatov I.S.², Tyazhelnikov A.A.^{2,5}, Dorozhkova V.S.², Gimadiev R.R.^{1,3,4,5}, Stuklov N.I.¹, Kochetov A.G.⁵, Gordienko K.V.⁴, Gubina E.V.³, Varekha N.V.¹, Chausova S.V.⁵, Shchegolev O.B.³, Silkina T.A.⁶

CALCULATION OF LOW-DENSITY LIPOPROTEIN CHOLESTEROL CONCENTRATION FROM LIPID PROFILE DATA USING MACHINE LEARNING

¹Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University), 117198, Moscow, Russia;

²City Outpatient Clinic No. 220, 123022, Moscow, Russia;

³LABHUB LLC, 119002, Moscow, Russia;

⁴Institute of Biomedical Problems of the Russian Academy of Sciences, 123007 Moscow, Russia

⁵N.I. Pirogov Russian National Research Medical University, Ministry of Health of Russia, 117513, Moscow, Russia;

⁶Gemotest Laboratory LLC, 123112, Moscow, Russia

In the context of implementing the Russian Federation's strategy for the development of digital technologies in medicine, studying the possibilities of artificial intelligence and machine learning in clinical laboratory diagnostics is highly relevant.

Objective. To develop a machine learning model for calculating low-density lipoprotein cholesterol concentration based on lipid profile parameters and evaluate its accuracy compared to traditional direct and calculated methods.

Material and methods. Model training and testing were performed on a sample of 631,267 patients using an anonymized dataset that included demographic characteristics and lipid profile results: concentrations of total cholesterol, high-density lipoprotein cholesterol, triglycerides, and low-density lipoprotein cholesterol.

Results. The regression model demonstrated high accuracy in predicting low-density lipoprotein cholesterol levels with $R^2 = 0.952$, outperforming traditional formulas for patients with triglyceride levels above 4.5 mmol/L and low-density lipoprotein cholesterol levels above 3.0 mmol/L: the mean absolute error between predicted and known low-density lipoprotein cholesterol values was 0.476 and 0.168 versus 1.173, 0.562, 0.796 and 0.187, 0.173, 0.200 mmol/L calculated using the Friedewald, Martin, and Sampson formulas, respectively.

Conclusion. The developed model, trained and tested on a large Russian patient cohort, outperforms traditional formulas at high triglyceride and low-density lipoprotein cholesterol levels, and performs comparably at other ranges in terms of accuracy for calculating low-density lipoprotein cholesterol concentration, confirming the reliability of the regression model for clinical use when direct measurement of low-density lipoprotein cholesterol concentration is unavailable.

Key words: machine learning; artificial intelligence; low-density lipoprotein cholesterol; Friedewald equation; Martin equation; Sampson equation

For citation: Varekha L.A., Kokorin V.A., Sviatov I.S., Tyazhelnikov A.A., Dorozhkova V.S., Gimadiev R.R., Stuklov N.I., Kochetov A.G., Gordienko K.V., Gubina E.V., Varekha N.V., Chausova S.V., Shchegolev O.B., Silkina T.A. Calculation of low-density lipoprotein cholesterol concentration from lipid profile data using machine learning. *Klinicheskaya Laboratornaya Diagnostika (Russian Clinical Laboratory Diagnostics)*. 2026; 71 (2): 133-138 (in Russ.).

DOI: <https://doi.org/10.51620/0869-2084-2026-71-2-133-138>

EDN: HHMFEB

For correspondence: Varekha Lyubov Andreevna, postgraduate student of the Department of Hospital Therapy with courses in endocrinology, hematology and KLD; assistant of the Department of Histology, Cytology and Embryology of the Patrice Lumumba RUDN Medical Institute; e-mail: eykhenvvald-la@rudn.ru

Information about authors:

Varekha L.A.,	https://orcid.org/0000-0002-8709-3056 ;
Kokorin V.A.,	https://orcid.org/0000-0001-8614-6542 ;
Sviatov I.S.,	https://orcid.org/0009-0006-3948-3697 ;
Tyazhelnikov A.A.,	https://orcid.org/0000-0002-2191-0623 ;
Dorozhkova V.S.,	https://orcid.org/0009-0001-5386-2646 ;
Gimadiev R.R.,	https://orcid.org/0000-0002-9567-3317 ;
Stuklov N.I.,	https://orcid.org/0000-0002-4546-1578 ;
Kochetov A.G.,	https://orcid.org/0000-0003-3632-291X ;
Gordienko K.V.,	https://orcid.org/0000-0002-5381-1013 ;
Gubina E.V.,	https://orcid.org/0009-0004-7507-8205 ;
Varekha N.V.,	https://orcid.org/0009-0004-0969-6144 ;
Chausova S.V.,	https://orcid.org/0000-0001-5799-5908 ;
Shchegolev O.B.,	https://orcid.org/0000-0002-3493-1415 ;
Silkina T.A.,	https://orcid.org/0009-0007-2869-9510 .

Conflict of interests. The authors declare absence of conflict of interests.

Acknowledgment. The study had no sponsor support.

Received 25.09.2025

Accepted 31.12.2025

Published 01.02.2026

ВВЕДЕНИЕ

Холестерин липопротеинов низкой плотности (ХС-ЛПНП) является ключевым индикатором оценки сердечно-сосудистых рисков [1, 2]. Согласно имеющимся данным, около 60 % населения России страдают гиперхолестеринемией, характеризующейся повышенными значениями ХС-ЛПНП более 3,0 ммоль/л [3]. Несмотря на важность точного измерения этого параметра, финансовые ограничения осложняют широкое применение прямого тестирования, способствуя использованию расчетных методов определения концентрации ХС-ЛПНП.

В 1972 году W.T. Friedewald и соавторы [4], анализируя данные 448 пациентов, разработали формулу:

$$\text{ЛПНП} \left(\frac{\text{мг}}{\text{дл}} \right) = \text{ОХС} - \text{ЛПВП} - \frac{\text{ТГ}}{5},$$

где: ОХС - общий холестерин, ЛПВП – холестерин

липопротеинов высокой плотности, ТГ – триглицерины, ($\text{TG} \div 5$) – липопротеины очень низкой плотности (ЛПОНП).

Спустя более 50 лет эта формула остается широко используемой в медицинской практике и упоминается в качестве расчетного показателя в клинических рекомендациях. По данным нашего опроса в 2024 году в профильном канале Telegram, 16 % из 525 специалистов медицинских лабораторий используют уравнение Фридвальда [5]. Известно, что точность оценки существенно снижается при таких состояниях, как сахарный диабет, патология почек и печени, $\text{TG} \geq 4,5$ ммоль/л (400 мг/дл) или ХС-ЛПНП $< 1,8$ ммоль/л (70 мг/дл) [6]. Недостатки формулы побуждали и побуждают к ее совершенствованию и разработке новых подходов для определения расчетного показателя ХС-ЛПНП. В 2013 году S. Martin и соавторы [7] предложили новое урав-

нение на основе данных 1 350 908 липидных профилей:

$$\text{ЛПНП (мг/дл)} = \text{ОХС} - \text{ЛПВП} - \frac{\text{ТГ}}{\text{фактор}},$$

где: фактор представляет стратоспецифичное, зависящее от уровня триглицеридов и общего холестерина, медианное соотношение ТГ и ЛПОНП. Уравнение Мартина, разработанное с использованием традиционного линейного регрессионного анализа и демонстрирующее более высокую точность по сравнению с формулой Фридвальда, имеет погрешности, особенно при более низких уровнях ХС-ЛПНП [8].

В 2020 году M. Sampson и соавторы [9], используя линейную регрессию и данные 8656 пациентов, разработали уравнение:

$$\text{ЛПНП (мг/дл)} = \frac{\text{ОХС}}{0,948} - \frac{\text{ЛПВП}}{0,971} - \left(\frac{\text{ТГ}}{8,56} + \frac{\text{ТГ} \times \text{ЛПВП}}{2140} + \frac{\text{ТГ}^2}{16100} \right) - 9,44,$$

где: неЛПВП – фракции холестерина за вычетом ХС-ЛПВП в размерности мг/дл. Как и уравнение Мартина, оно превосходит формулу Фридвальда по точности, особенно при гипертриглицеридемии. В отличие от метода Мартина, формула Сэмпсона основана на эталонном методе бета-количественного анализа с применением ультрацентрифугирования с последующим химическим определением фракций холестерина.

В последние годы исследователи активно применяют методы искусственного интеллекта (ИИ), в частности машинного обучения (МО), для повышения точности оценки ХС-ЛПНП в стандартной липидной панели: ОХС, ТГ, ХС-ЛПВП, ХС-ЛПНП. Существует ряд зарубежных разработок в этой сфере [10–15], представления в научной литературе аналогичных отечественных исследований пока не проводилось.

ЦЕЛЬ ИССЛЕДОВАНИЯ – создать модель МО для расчета концентрации ХС-ЛПНП на основе показателей липидного профиля и оценить ее точность относительно традиционных прямого и расчётных методов.

МАТЕРИАЛ И МЕТОДЫ

Ретроспективное исследование выполнено в соответствии с этическими принципами Хельсинской декларации Всемирной медицинской ассоциации (Форталеза, 2013). Исследование основано на анализе деидентифицированного набора лабораторных данных 631 267 пациентов за 2022–2025 годы, полученного из лабораторных информационных систем (ЛИС) нескольких сетевых лабораторий в соответствии с научно-практическими договорами. Анонимизация данных исключает возможность идентификации отдельных пациентов. В соответствии со статьей 32 Хельсинской декларации и учитывая ретроспективный характер исследования, получение индивидуального информированного согласия пациентов не требовалось. Собрана информация о концентрациях ОХС, ХС-ЛПВП, ХС-ЛПНП, ТГ, о поле и возрасте пациентов. Критерием включения в исследование являлось наличие всех вышеуказанных данных пациента, критерием невключения – возраст менее 18 лет.

Лабораторные показатели определялись в сыворотке крови на доступных в Российской Федерации автоматических анализаторах: Roche (Германия), Abbott (США), Beckman Coulter (США), Mindray (КНР). В рамках данного ретроспективного исследования целенаправленная гармонизация результатов измерения ХС-ЛПНП между

разными аналитическими системами не проводилась. Все лаборатории-поставщики данных являются участниками Федеральной системы внешней оценки качества (ФСВОК), что обеспечивает соответствие методов установленным критериям приемлемости в рамках рутинной клинической практики. Крупный объем выборки ($n = 631\,267$) способствует статистическому нивелированию случайной погрешности измерений.

Выбросы идентифицированы как значения, превышающие 99-й процентиль или находящиеся ниже 1-го процентиля, и далее анализировали на предмет биологической интерпретируемости. Значения, соответствующие известным патологическим состояниям (например, семейной гиперхолестеринемии или тяжелой гипертриглицеридемии), сохранены для дальнейшего анализа. Значения, которые могли быть артефактами методики (например, отрицательный уровень ХС-ЛПНП) или являлись физиологически невозможными, исключены из анализа.

Прогнозирование уровня ХС-ЛПНП проведено с учётом значений ОХС, ТГ, ХС-ЛПВП, пола, возраста на основе разработанной модели, комбинирующей методы линейной регрессии и градиентного бустинга на решающих деревьях с использованием платформы LightAutoML [16]. Измеренный колориметрическим фотометрическим методом, уровень ХС-ЛПНП использован в качестве значений целевой переменной при обучении модели. Точность модели дополнительно оценена при гипертриглицеридемии, при которой используемые в настоящее время расчётные формулы являются неточными. С этой целью проведена стратификация пациентов по уровню ТГ $< 4,5$ и $\geq 4,5$ ммоль/л. Стратификация пациентов проведена в зависимости от уровня ХС-ЛПНП в соответствии с классификацией риска сердечно-сосудистых событий согласно клиническим рекомендациям по ведению пациентов с нарушениями липидного обмена по 5 категориям: $< 1,4$; $1,4\text{--}1,8$; $1,8\text{--}2,6$; $2,6\text{--}3,0$ и $\geq 3,0$ ммоль/л [3].

Набор данных случайным образом разделен на обучающую (80 %) и тестовую (20 %) выборки. Оптимизация гиперпараметров выполнена с помощью 5-кратной кросс-валидации, включавшей подбор следующих параметров: скорость обучения, глубина деревьев, минимальное число примеров на листе, параметры регуляризации [17]. Задача оптимизации – максимизировать коэффициент детерминации R^2 на кросс-валидации. После выбора оптимальной модели ее точность оценивалась на тестовом наборе. Проведено сравнение полученных значений с результатами, полученными по формулам Фридвальда, Мартина и Сэмпсона на том же тестовом наборе.

Программное обеспечение с ИИ для расчета концентрации ХС-ЛПНП соответствует требованиям нормативных документов, регулирующих использование ИИ в медицине, включая «Кодекс этики применения искусственного интеллекта в сфере охраны здоровья» (версия 2.1, утв. 14.02.2025), ГОСТ Р 59921.1-2022 «Системы искусственного интеллекта в клинической медицине», постановление Правительства РФ № 1906 от 24.11.2020 «О внесении изменений в Правила государственной регистрации медицинских изделий», что обеспечивает соблюдение принципов безопасности, клинической обоснованности, прозрачности алгорит-

мов и правовых норм внедрения цифровых медицинских технологий.

Статистический анализ включал предварительную оценку нормальности распределения с помощью теста Шапиро-Уилка и Д'Агостино-Пирсона [18]. Поскольку данные не соответствовали нормальному распределению, применялись непараметрические методы статистического анализа: медиана с интерквартильным размахом для описательной статистики, корреляционный анализ Спирмена. Качество разработанной модели оценивали путем сравнения абсолютных отклонений ее результатов от истинных значений ХС-ЛПНП с отклонениями, полученными по известным формулам. Статистическая значимость различий между этими группами отклонений определена с помощью теста Данна с поправкой Бенъямини-Криттерса-Йекутелли. Критический уровень значимости установлен на уровне $\alpha = 0,05$ при двустороннем тестировании гипотез. Статистический анализ выполнен с использованием программного обеспечения GraphPad Prism 8.0.

РЕЗУЛЬТАТЫ

Характеристики набора данных до и после деления на группы по уровням ТГ и ХС-ЛПНП представлены в табл. 1–3.

В исследуемой выборке гиперхолестеринемия ($\text{OXC} \geq 5,0 \text{ ммоль/л}$) выявлена у 64,4 % обследуемых. Повышенный уровень ХС-ЛПНП ($\geq 3,0 \text{ ммоль/л}$) зафиксирован в 63,5 % случаев. Гипертриглицеридемия ($\text{TG} \geq 1,7 \text{ ммоль/л}$) обнаружена у 30 % обследуемых. Низкий уровень ХС-ЛПВП ($< 1,0 \text{ ммоль/л}$ у мужчин и $< 1,2 \text{ ммоль/л}$ у женщин) отмечен в 19,4 % случаев. У 28,0 % пациентов наблюдалось значительно повышенное содержание ОХС ($> 6,2 \text{ ммоль/л}$), для 22,1 % – высокая концентрация ХС-ЛПНП ($> 4,2 \text{ ммоль/л}$).

Статистически значимая корреляция ($r > 0,3$; $p < 0,05$) с целевым признаком ХС-ЛПНП наблюдалась только для ОХС ($r = 0,94$; рис. 1). Некоторые признаки оказались слабо коррелирующими друг с другом (TG и OXC $r = 0,31/0,37$; TG и LPVP – $r = -0,39/-0,31$; OXC и LPVP – $r = 0,31/0,34$).

Анализ объясненной дисперсии (доля изменчивости ХС-ЛПНП, объясняемой моделью) дополнил корреляционные матрицы – наиболее надежными предикторами для разработанной модели явились ОХС, ХС-

Таблица 1
Распределение возрастных характеристик и липидного профиля по половому признаку

Показатель	Женщины (<i>n</i> = 409 779)	Мужчины (<i>n</i> = 221 488)
Возраст, годы	59 [46; 69]	54 [41; 65]
ОХС, ммоль/л	5,5 [4,64; 6,39]	5,1 [4,26; 6,04]
ХС-ЛПНП, ммоль/л	3,3 [2,61; 4,09]	3,2 [2,49; 3,97]
ХС-ЛПВП, ммоль/л	1,5 [1,28; 1,78]	1,2 [1,04; 1,44]
ТГ, ммоль/л	1,2 [0,87; 1,71]	1,3 [0,91; 1,80]

Примечание. *n* – здесь и далее число пациентов, удовлетворяющих условию (в данном случае пол: женщины или мужчины).

Таблица 2
Распределение возрастных характеристик и липидного профиля по половому признаку и уровням ТГ

Уровень ТГ, ммоль/л	Пол	<i>n</i>	Возраст, годы	ОХС, ммоль/л	ХС-ЛПВП, ммоль/л	ХС-ЛПНП, ммоль/л	ТГ, ммоль/л
< 4,5	Ж	405841	59 [46; 69]	5,47 [4,63; 6,37]	1,52 [1,29; 1,79]	3,32 [2,60; 4,09]	1,20 [0,86; 1,69]
	М	214920	54 [41; 65]	5,10 [4,23; 5,99]	1,23 [1,05; 1,44]	3,22 [2,48; 3,97]	1,32 [0,93; 1,91]
$\geq 4,5$	Ж	3938	58 [50; 65]	6,82 [5,86; 7,94]	1,18 [1,00; 1,40]	3,74 [2,93; 4,57]	5,42 [4,85; 6,60]
	М	6568	49 [41; 57]	6,42 [5,50; 7,44]	1,05 [0,90; 1,24]	3,53 [2,80; 4,31]	5,69 [4,95; 7,17]

Таблица 3
Распределение возрастных характеристик и липидного профиля по половому признаку и уровням ХС-ЛПНП

Уровень ХС-ЛПНП, ммоль/л	Пол	<i>n</i>	Возраст, годы	ОХС, ммоль/л	ХС-ЛПВП, ммоль/л	ХС-ЛПНП, ммоль/л	ТГ, ммоль/л
до 1,4	М	6567	65 [55; 73]	2,84 [2,56; 3,2]	1,06 [0,87; 1,3]	1,2 [1,03; 1,32]	1,11 [0,77; 1,77]
	Ж	6963	69 [57; 76]	3,15 [2,82; 3,5]	1,33 [1,07; 1,64]	1,22 [1,05; 1,32]	1,12 [0,79; 1,67]
от 1,4 до 1,8	М	12203	64 [51; 72]	3,27 [3,02; 3,57]	1,12 [0,93; 1,34]	1,64 [1,53; 1,72]	1,05 [0,77; 1,54]
	Ж	16371	66 [48; 75]	3,59 [3,3; 3,89]	1,41 [1,17; 1,7]	1,65 [1,54; 1,73]	1,04 [0,74; 1,46]
от 1,8 до 2,6	М	44693	58 [41; 68]	3,98 [3,68; 4,3]	1,17 [0,99; 1,4]	2,26 [2,05; 2,44]	1,11 [0,8; 1,64]
	Ж	78835	59 [39; 71]	4,26 [3,95; 4,59]	1,48 [1,24; 1,76]	2,28 [2,07; 2,45]	1,02 [0,74; 1,45]
от 2,6 до 3,0	М	29828	54 [40; 65]	4,61 [4,37; 4,9]	1,2 [1,01; 1,43]	2,81 [2,71; 2,91]	1,22 [0,87; 1,8]
	Ж	55781	55 [40; 68]	4,86 [4,6; 5,15]	1,51 [1,27; 1,79]	2,81 [2,71; 2,91]	1,06 [0,77; 1,5]
от 3,0	М	128197	52 [41; 62]	5,86 [5,3; 6,56]	1,26 [1,08; 1,46]	3,84 [3,41; 4,4]	1,51 [1,07; 2,18]
	Ж	251829	59 (48; 67)	6,15 (5,57; 6,88)	1,54 (1,31; 1,8)	3,9 (3,43; 4,5)	1,32 (0,96; 1,85)

ЛПВП, ТГ с коэффициентом детерминации $R^2 = 0,951$. Возраст и пол суммарно дали прирост коэффициента детерминации менее 1 %.

Коэффициент детерминации R^2 регрессионной модели составил 0,952, что указывает на успешное объяснение 95,2 % вариации концентрации ХС-ЛПНП выбранными предикторами. Средняя абсолютная ошибка (МАЕ), составившая 0,156 ммоль/л, отражает среднее откло-

нение предсказанных значений от фактических во всем диапазоне концентрации ХС-ЛПНП, включая высокие значения. Средняя относительная ошибка (MAPE) составила 5,2 %, что свидетельствует о незначительных отклонениях предсказаний (в среднем на 5,2 % от фактических значений). Корень из среднеквадратичной ошибки (RMSE) составил 0,240 ммоль/л, что подтверждает высокую точность предсказаний с учетом коррекции на большие отклонения.

Метрики высокой точности расчета ХС-ЛПНП с помощью разработанной модели по сравнению с традиционными методами при различных диапазонах ТГ и ХС-ЛПНП по средней абсолютной ошибке и корню из среднеквадратичной ошибки, продемонстрированы в табл. 4 и на рис. 2.

ОБСУЖДЕНИЕ

Важность разработки метода МО для прогнозирования концентрации ЛПНП обусловлена высокой клинической значимостью данного биомаркера как ключевого фактора сердечно-сосудистого риска на фоне ограниченной доступности точных методов его определения в рутинной лабораторной практике. Как ранее отмечено, традиционные расчетные формулы имеют существенные ограничения при гипертриглицеридемии, прямые методы определения ХС-ЛПНП остаются дорогостоящими и технически сложными. Результаты представленного исследования по применению алгоритмов МО демонстрируют высокую точность прогнозирования концентрации ХС-ЛПНП на основе доступных лабораторных параметров, что способствует улучшению качества диагностики дислипидемии, оптимизации стратификации сердечно-сосудистого риска и снижению экономических затрат на лабораторные исследования при сохранении высокой диагностической точности.

Представлена новая модель МО для прогнозирования концентрации холестерина липопротеинов низкой плотности, основанная на базовых демографических и лабораторных параметрах (пол, возраст, общий холестерин, холестерин липопротеинов высокой плотности и триглицериды). Модель обучена и валидирована на репрезентативной когорте из 631 267 пациентов, обеспечивающей высокую статистическую мощность исследования. Анализ распределения дислипидемии в изученной популяции выявил ее высокую распространенность (64%), соответствующую данным крупного популяционного исследования ЭССЭ-РФ [3].

Сравнительный анализ продемонстрировал пре- восходство разработанной модели над традиционными расчетными уравнениями по показателю сред-

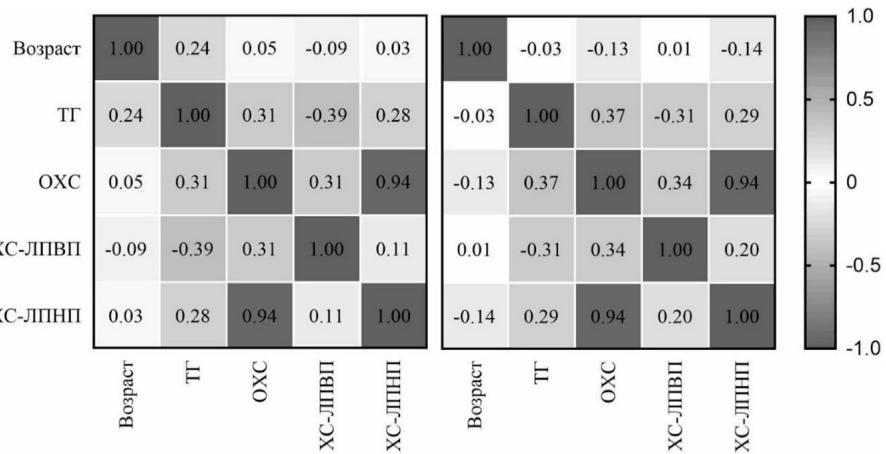


Рис. 1. Тепловая карта корреляционных связей (критерий Спирмена) между показателями липидограммы для пациентов женского (слева) и мужского пола (справа).

Таблица 4
Метрика качества модели расчета концентрации ХС-ЛПНП в зависимости от диапазона уровней ТГ и ХС-ЛПНП

Группа, ммоль/л	МАЕ, ммоль/л			
	Уравнение Фридвальда	Уравнение Мартина	Уравнение Сэмпсона	Новая модель (разработанная)
ТГ < 4,5	0,143*	0,155*	0,163*	0,150
ТГ ≥ 4,5	1,173*	0,562*	0,796*	0,476
ХС-ЛПНП < 1,4	0,114*	0,217*	0,160 [#]	0,227
ХС-ЛПНП 1,4–1,8	0,092*	0,150*	0,121*	0,145
ХС-ЛПНП 1,8–2,6	0,117*	0,139*	0,127*	0,131
ХС-ЛПНП 2,6–3,0	0,131	0,139*	0,139*	0,131
ХС-ЛПНП ≥ 3,0	0,187*	0,173*	0,200*	0,168

Примечание. * – $p < 0,01$; # – $p < 0,05$ при сравнении известных уравнений с разработанной моделью.

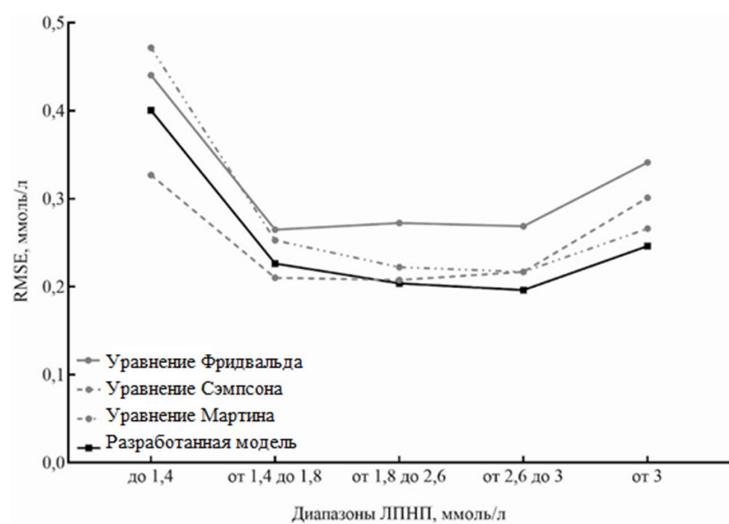


Рис. 2. RMSE сравниваемых уравнений и модели МО; по оси абсцисс представлены результаты прямого измерения ХС-ЛПНП. RMSE – корень из среднеквадратичной ошибки.

ней абсолютной ошибки в клинически значимых диапазонах концентраций. При уровне триглицеридов $\geq 4,5$ ммоль/л и холестерина ЛПНП $\geq 3,0$ ммоль/л модель МО показала MAE 0,476 и 0,168 ммоль/л соответственно, в то время как уравнения Фридвальда, Мартина и Сэмпсона демонстрировали существенно большие погрешности (1,173, 0,562, 0,796

и 0,187, 0,173, 0,200 ммоль/л, соответственно). Полученные результаты согласуются с данными современных зарубежных исследований [10–15].

Ограничения формулы Фридвалльда обусловлены использованием фиксированного коэффициента для расчета холестерина ЛПОНП на основе концентрации триглицеридов. При гипертриглицеридемии $\geq 4,5$ ммоль/л соотношение триглицеридов к холестерину в ЛПОНП превышает стандартный коэффициент вследствие присутствия хиломикронов и их ремнантов, при водящих к систематическому занижению концентрации холестерина ЛПНП. Данное ограничение приобретает особую клиническую значимость в контексте установленной ассоциации повышенного уровня триглицеридов с риском сердечно-сосудистых заболеваний.

Предложенная модель МО представляет адаптивное решение для определения холестерина ЛПНП с возможностью дополнительного обучения на локальных выборках для повышения точности в специфических популяциях. Универсальность архитектуры модели позволяет ее применение для прогнозирования любого параметра липидного профиля при неполном наборе исходных данных. Техническая реализация модели обеспечивает её представление в виде веб-сервиса с программным интерфейсом (API), что значительно облегчает интеграцию в ЛИС. Это расширяет возможности ее практического применения для стратификации сердечно-сосудистого риска и реализации программ внутрилабораторного контроля качества исследований липидного обмена.

ЗАКЛЮЧЕНИЕ

Разработанная модель МО для прогнозирования концентрации холестерина ЛПНП демонстрирует значительное превосходство над традиционными расчетными методами в клинически значимых диапазонах концентраций, особенно при гипертриглицеридемии и повышенных уровнях ХС-ЛПНП. Валидация на крупной репрезентативной когорте подтверждает надежность и воспроизводимость полученных результатов.

Практическая значимость исследования заключается в создании универсального инструмента для уточненного определения ХС-ЛПНП на основе рутинных лабораторных параметров, что особенно актуально в условиях ограниченной доступности прямых методов измерения. Адаптивность модели к локальным популяционным особенностям и возможность интеграции в ЛИС обеспечивают широкие перспективы ее внедрения в клиническую практику.

Полученные результаты вносят вклад в развитие методов ИИ в клинической лабораторной диагностике и открывают новые возможности для оптимизации стратификации сердечно-сосудистого риска. Дальнейшие исследования должны быть направлены на валидацию модели в различных этнических группах и клинических популяциях для подтверждения ее универсальной применимости.

ращения: 18.06.2025).

REFERENCES

1. Lewington S., Whitlock G., Clarke R., Sherliker P., Emberson J. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet.* 2007; 370(9602): 1829-39. DOI: 10.1016/S0140-6736(07)61778-4.
2. Mach F., Baigent C., Catapano A.L., Koskinas K.C., Casula M., Badimon L. et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur. Heart J.* 2020; 21; 41(44): 4255. DOI: 10.1093/euroheartj/ehz826.
3. Ministry of Health of the Russian Federation. Lipid metabolism disorders. 2023. [Электронный ресурс]. URL: https://cr.minsdrav.gov.ru/preview-cr/752_1. (дата обращения: 03.06.2025). (in Russian)
4. Friedewald W.T., Levy R.I., Fredrickson D.S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 1972; 18(6):499-502.
5. [Electronic resource]. URL: https://t.me/Lab_Hub/1961. (date of access: 18.06.2025). (in Russian)
6. Arnett D.K., Blumenthal R.S., Albert M.A., Buroker A.B., Goldberger Z.D., Hahn E. J. et al. 2019 ACC/AHA Guideline on the primary prevention of cardiovascular disease: A report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Circulation.* 2020; 141(16): e774. DOI: 10.1161/CIR.0000000000000771.
7. Martin S.S., Blaha M.J., Elshazly M.B., Toth P.P., Kwiterovich P.O., Blumenthal R.S. et al. Comparison of a novel method vs the Friedewald equation for estimating low-density lipoprotein cholesterol levels from the standard lipid profile. *JAMA.* 2013; 310(19): 2061-8. DOI: 10.1001/jama.2013.280532.
8. Quispe R., Hendrani A., Elshazly M.B. Michos E.D., McEvoy J.W., Blaha M.J. et al. Accuracy of low-density lipoprotein cholesterol estimation at very low levels. *BMC Med.* 2017; 15(1): 83. DOI: 10.1186/s12916-017-0852-2.
9. Sampson M., Ling C., Sun Q., Harb R., Ashmaig M., Warnick R. et al. New equation for calculation of low-density lipoprotein cholesterol in patients with normolipidemia and/or hypertriglyceridemia. *JAMA Cardiol.* 2020; 5(5): 613. DOI: 10.1001/jamacardio.2020.0526.
10. Sezer S., Oter A., Ersoz B., Topcuoglu C., Bulbul H.İ., Sagiroglu S. et al. Explainable artificial intelligence for LDL cholesterol prediction and classification. *Clin. Biochem.* 2024; 130: 110791. DOI: 10.1016/j.clinbiochem.2024.110791.
11. Çubukçu H.C., Topcu D.İ. Estimation of low-density lipoprotein cholesterol concentration using machine learning. *Lab. Med.* 2022; 53(2): 161-71. DOI: 10.1093/labmed/lmab065.
12. Oh G.C., Ko T., Kim J.H., Lee M.H., Choi S.W., Bae Y.S. et al. Estimation of low-density lipoprotein cholesterol levels using machine learning. *Int. J. Cardiol.* 2022; 352: 144-9. DOI: 10.1016/j.ijcard.2022.01.029.
13. Kwon Y.J., Lee H., Baik S.J., Chang H.J., Lee J.W. Comparison of a machine learning method and various equations for estimating low-density lipoprotein cholesterol in Korean populations. *Front. Cardiovasc. Med.* 2022; 9: 824574. DOI: 10.3389/fcvm.2022.824574.
14. Barakett-Hamade V., Ghayad J.P., Mchanta G., Sleilaty G. Is machine learning-derived low-density lipoprotein cholesterol estimation more reliable than standard closed form equations? Insights from a laboratory database by comparison with a direct homogeneous assay. *Clin. Chim. Acta.* 2021; 519: 220-6. DOI: 10.1016/j.cca.2021.05.008.
15. Chen L., Rong C., Ma P., Li Y., Deng X., Hua M. A new equation for estimating low-density lipoprotein cholesterol concentration based on machine learning. *Medicine (Baltimore).* 2024; 103(15): e37766. DOI: 10.1097/MD.00000000000037766.
16. Vakhrushev A., Ryzhkov A., Savchenko M., Simakov D., Damdinov R., Tuzhilin A. LightAutoML: AutoML solution for a large financial services ecosystem. *arXiv.* 2021; 2109.01528. DOI: 10.48550/arXiv.2109.01528.
17. Fushiki T. Estimation of prediction error by using K-fold crossvalidation. *Statistics and Computing.* 2011; 21: 137-46. DOI: 10.1007/s11222-009-9153-8.
18. Das K.R., Imon A. A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics.* 2016; 5(1): 5-12. DOI: 10.11648/j.ajtas.20160501.12.



ЛИТЕРАТУРА (ПП. 1-2, 4, 6-18 СМ. REFERENCES)

3. Министерство здравоохранения Российской Федерации. Нарушения липидного обмена. 2023. [Электронный ресурс]. URL: https://cr.minsdrav.gov.ru/preview-cr/752_1. (дата обращения: 03.06.2025).
5. [Электронный ресурс]. URL: https://t.me/Lab_Hub/1961. (дата об-