

Участник:

ОБЩЕСТВО С ОГРАНИЧЕННОЙ  
ОТВЕТСТВЕННОСТЬЮ "ЛАБХАБ"  
(ООО «ЛАБХАБ»), резидент Сколково,  
Академии Инноваторов, фонда «МИК»

Паспорт проекта:

**«Программное обеспечение «ЛабЧекап» с технологией искусственного интеллекта  
для предиктивной диагностики ин витро и контроля качества исследований»**

Команда проекта:

Гимадиев Ринат Рашитович, руководитель проекта;

Щеголев Олег Борисович, инженер-программист;

Виноградов Михаил Андреевич, специалист по регуляторике и стандартизации.

## ОГЛАВЛЕНИЕ

1.	ВВЕДЕНИЕ	3
1.1.	ПРОБЛЕМАТИКА ИССЛЕДОВАНИЯ	3
1.2.	ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ	5
1.3.	ГИПОТЕЗА	5
1.4.	АКТУАЛЬНОСТЬ И НОВИЗНА	6
1.5.	ЦЕЛЬ И ЗАДАЧИ	6
1.6.	ОБЗОР ИСТОЧНИКОВ	7
1.7.	ИСПОЛЬЗОВАННЫЕ МЕТОДЫ	8
1.8.	СТРУКТУРА РАБОТЫ	8
2.	ОСНОВНАЯ ЧАСТЬ	10
2.1.	ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ И МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ	10
2.1.1.	КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ПЕРСОНАЛИЗИРОВАННОЙ ПРОФИЛАКТИКИ В СИСТЕМЕ ЛАБОРАТОРНОЙ ДИАГНОСТИКИ	10
2.1.2.	МЕТОДОЛОГИЯ ФОРМИРОВАНИЯ ИССЛЕДОВАТЕЛЬСКОЙ БАЗЫ ДАННЫХ И РАЗРАБОТКИ МОДЕЛЕЙ	10
2.2.	ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ И ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ	12
2.2.1.	ВАЛИДАЦИЯ ТОЧНОСТИ ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ	12
2.2.2.	АРХИТЕКТУРА И ГОТОВНОСТЬ К ВНЕДРЕНИЮ	14
2.2.3.	АНАЛИЗ ВКЛАДА ПРЕДИКТОРОВ И БИОЛОГИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ МОДЕЛЕЙ	14
2.3.	ВЫВОД О ЦЕННОСТИ, ПОЛЬЗЕ И ПЕРСПЕКТИВАХ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ	16
2.3.1.	НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ И КЛИНИЧЕСКАЯ ЦЕННОСТЬ	16
2.3.2.	ПРАКТИЧЕСКАЯ ПОЛЬЗА И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЙ ЭФФЕКТ	16
2.3.3.	ПЕРСПЕКТИВЫ РАЗВИТИЯ И МАСШТАБИРОВАНИЯ	18
3.	ЗАКЛЮЧЕНИЕ	19

## 1. ВВЕДЕНИЕ

### 1.1. ПРОБЛЕМАТИКА ИССЛЕДОВАНИЯ

**Проблема 1. Системные ограничения контроля качества лабораторных исследований.** Несмотря на растущий объём рынка лабораторной диагностики (мировой – \$86,5+ млрд, российский – около 105,7 млрд руб. [1,2]), сохраняется высокий уровень ошибок. До 2% проб крови бракуется из-за преаналитических ошибок [3,4]. На аналитическом этапе сохраняется зависимость от устаревших формул (например, Фридвальда, имеющей ограничения при триглицеридах (ТГ) >4,5 ммоль/л, которым пользуются 16% опрошенных специалистов КЛД [5]; формулы коррекции кальция также недостаточно точны [6]). На постаналитическом этапе добавляются риски из-за нестандартизированной интерпретации и несвоевременного оповещения критических значений [7]. Существующие системы контроля качества не способны автоматически выявлять эти клинически значимые отклонения, что снижает достоверность диагностики, ведёт к перерасходу реагентов и увеличению трудозатрат.

**Проблема 2. Недостаточная эффективность массового скрининга для ранней диагностики неинфекционных заболеваний.** Программы профилактических медицинских осмотров (ПМО) и диспансеризации определенных групп взрослого населения (ДОГВН) в РФ имеют ограниченный спектр исследований, что не позволяет оценить многофакторные риски у пациента [8,9]. Несмотря на массовый охват (109 млн человек в 2024 г. [10]), хронические неинфекционные заболевания (ХНИЗ) выявляются у >45% населения [11], нанося ущерб ~4% ВВП [12]. Наши данные пилотных тестирований показывают, что стандартный скрининг пропускает группы риска: 4,3% пациентов с повышенным холестерином липопротеинов низкой плотности (ХС-ЛПНП) при нормальном уровне общего холестерина (ОХС) и 2,4% с повышенным гликированным гемоглобином (HbA1c) при нормогликемии. Однако расширение перечня обследований экономически невыгодно для системы ОМС. Национальные проекты «Продолжительная и активная жизнь» и «Новые технологии сбережения здоровья» [13,14] ставят цель увеличить продолжительность жизни до 78 лет к 2030 году, что требует оптимизации первичного звена здравоохранения [15]. Таким образом, возможности скрининга ограничены устаревшими подходами и недостаточной цифровизацией.

### 1.2. ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ

**Объект исследования:** процесс лабораторной диагностики и профилактического скрининга в системе здравоохранения, включая методы контроля качества и алгоритмы интерпретации результатов.

**Предмет исследования:** разработка и валидация интеллектуальной программной платформы «ЛабЧекап», основанной на технологиях искусственного интеллекта (ИИ) и машинного обучения (МО) для сквозного контроля качества лабораторных данных и предиктивной аналитики рисков ХНИЗ на основе рутинных показателей диспансеризации.

### **1.3. ГИПОТЕЗА**

Мы предполагаем, что интеграция интеллектуальной программной надстройки, реализующей сквозной ИИ-контроль качества и способной с высокой точностью прогнозировать концентрации дорогостоящих биомаркеров второго уровня / стратифицировать риски на основе данных первичного скрининга, позволит:

1. Существенно повысить достоверность и экономическую эффективность лабораторных исследований. Платформа усиливает внутренний контроль, применяя улучшенные расчетные методики для связанных параметров и увеличивая точность в ситуациях, где классические формулы дают погрешность. Например, ХС-ЛПНП традиционно вычисляют по формуле Фридвальда при отсутствии гипертриглицеридемии (~30% взрослого населения), тогда как модели МО позволяют обойти эти ограничения с точностью прогноза, сопоставимой с лабораторной (не превышает допустимый порог общей погрешности для прямых измерений ХС-ЛПНП, установленный протоколами контроля качества на уровне 12% (CHOLESTEROL REFERENCE METHOD LABORATORY NETWORK, 2024).

2. Трансформировать массовый скрининг в системе ОМС в персонализированный инструмент ранней доклинической диагностики ХНИЗ без пропорционального роста затрат. Платформа позволяет с высокой точностью рассчитывать концентрации важных, но экономически затратных показателей второго уровня (таких как ХС-ЛПНП, HbA1c, ферритин и др.) и автоматически выявлять факторы риска ХНИЗ (по отклонению от пороговых значений анализов «золотого стандарта»). Сформировать технологическую основу для предиктивной и профилактической медицины в рамках национальных проектов, преодолевая барьер недоверия к «черному ящику» ИИ через клиническую валидацию и интерпретируемость результатов.

### **1.4. АКТУАЛЬНОСТЬ И НОВИЗНА**

Актуальность проекта обусловлена его полным соответствием ключевым государственным инициативам в области здравоохранения и цифровой трансформации. Работа напрямую способствует достижению национальных целей, установленных Указом Президента РФ № 309 от 07.05.2024, в рамках национальных проектов «Новые технологии сохранения здоровья» и «Продолжительная и активная жизнь», направленных на увеличение продолжительности и качества жизни. Проект соответствует Указу Президента

Российской Федерации от 28.02.2024 г. № 145 «О Стратегии научно-технологического развития Российской Федерации» и реализует приоритеты научно-технологических направлений, утвержденные Указом Президента РФ № 529 от 18.06.2024, включая развитие превентивной и персонализированной медицины, применение биомедицинских технологий долголетия и сквозных технологий искусственного интеллекта, а также соответствует направлению перехода к персонализированной, предиктивной и профилактической медицине. Разработка также отвечает задачам «Стратегии развития здравоохранения до 2030 года» (Указ № 896 от 08.12.2025) и вносит вклад в цифровую трансформацию отрасли, создавая интеллектуальную систему предиктивной аналитики и поддержки врачебных решений, что закреплено в Распоряжении Правительства РФ № 959-р от 17.04.2024. Таким образом, проект «ЛабЧекап» является технологическим ответом на актуальные вызовы системы здравоохранения, обеспечивая инструмент для повышения эффективности, доступности и качества медицинской помощи в рамках обозначенных государственных приоритетов.

#### **Научная новизна:**

- Создана уникальная обезличенная база данных пар «рутинный скрининг – золотой стандарт» объемом >1,5 млн пациентов.
- Впервые разработана и валидирована методология построения предиктивных моделей МО для лабораторной диагностики на основе рутинного скрининга.
- Доказана возможность с высокой точностью прогнозировать концентрации дорогостоящих биомаркеров второго уровня на основе минимального набора рутинных показателей и стратифицировать риски ХНИЗ по отклонению от пороговых значений анализов «золотого стандарта».
- Научная новизна и значимость проекта подтверждены подготовкой 5 диссертационных исследований и 3 дипломных работ по его материалам, запланированных к защите в 2026-2027 годах (специальности: внутренние болезни, КЛД, медицинская информатика).

#### **Технологическая новизна:**

- Комплексная платформа, объединяющая модули ИИ-контроля качества лабораторного процесса и предиктивной аналитики для профилактики ХНИЗ.
- Готовое API-решение с микросервисной архитектурой, независимое от производителя оборудования и тест-систем, интегрируемое в действующие информационные системы.
- Реализован ансамбль моделей МО (LightGBM, CatBoost), обеспечивающий точность, сопоставимую с лабораторными методами, при времени отклика <1 секунды.

#### **Конкурентные преимущества:**

Наш проект занимает специализированную нишу, фокусируясь исключительно на лабораторных данных для контроля качества, рутинной диспансеризации для определения уровня некоторых показателей и рисков, что позволяет достичь высокой точности при минимальном наборе входных параметров. Отличие от аналогов:

- Традиционная лабораторная диагностика (прямое измерение показателей). Преимущество «ЛабЧекап»: экономичность (себестоимость прогноза <1 рубля против 200-1000 руб. (например, ХС-ЛПНП, HbA1c) и масштабируемость (не требует дополнительного взятия крови - 200-300 рублей), скорость выполнения прогноза превосходит лабораторное исследование (<1 сек. против 10-30 минут, не считая время на логистику).
- Webiomed: ориентирован на разнородные медицинские данные (АД, ЧСС, рост, вес, ОХС, ГПН, ОАК, ЭКГ, внутриглазное давление, флюорография, цитологические исследования шейки матки, маммографии), тогда как «ЛабЧекап» работает с минимальным набором рутинных показателей, что обеспечивает более быстрое внедрение и низкую стоимость.
- СберМедИИ: фокусируется на поддержке диагноза, наш проект решает задачу на более раннем этапе – массовый предиктивный скрининг.

## 1.5. ЦЕЛЬ И ЗАДАЧИ

**Цель:** создать и внедрить в клиническую практику интеллектуальную платформу «ЛабЧекап» для повышения качества лабораторной диагностики и персонализации массового скрининга путем прогнозирования рисков ХНИЗ на основе рутинных данных и лабораторных параметров первого этапа диспансеризации.

### **Задачи:**

1. Создать модуль сквозного интеллектуального контроля качества для автоматического выявления лабораторных ошибок и повышения достоверности расчетных показателей.
2. Разработать и валидировать модели МО для прогнозирования концентраций лабораторных маркеров второго уровня с высокой точностью и стратификации рисков ХНИЗ по отклонению от пороговых значений тестов «золотого стандарта».
3. Реализовать промышленное API-решение, обеспечивающее интеграцию платформы с медицинскими информационными системами и работу в режиме реального времени.
4. Провести серию пилотных тестирований в медицинских учреждениях для подтверждения клинической эффективности, экономической целесообразности и готовности к масштабированию.
5. Обеспечить регуляторное оформление решения как медицинского изделия класса 3.

## 1.6. ОБЗОР ИСТОЧНИКОВ

1.	Clinical Diagnostics Industry Size and Share Forecast Outlook (2025 to 2035) // Future Market Insights URL: <a href="https://www.futuremarketinsights.com/reports/clinical-diagnostics-market">https://www.futuremarketinsights.com/reports/clinical-diagnostics-market</a>
----	--

2.	Тренды рынка медицинских изделий в России 2025 // АналитикаМед URL: <a href="https://analitikamed.ru/articles/22/">https://analitikamed.ru/articles/22/</a> .
3.	Getawa S, Aynalem M, Melku M, Adane T. Blood specimen rejection rate in clinical laboratory: A systematic review and meta-analysis. <i>Pract Lab Med.</i> 2022;33:e00303. Published 2022 Dec 19. doi:10.1016/j.plabm.2022.e00303
4.	John GK, Favaloro EJ, Austin S, Islam MZ, Santhakumar AB. From errors to excellence: the pre-analytical journey to improved quality in diagnostics. A scoping review. <i>Clin Chem Lab Med.</i> 2025;63(7):1243-1259. Published 2025 Jan 28. doi:10.1515/cclm-2024-1277
5.	URL: <a href="https://t.me/Lab_Hub/1961">https://t.me/Lab_Hub/1961</a> . (дата обращения: 18.06.2025)
6.	Hung W. L. et al. A new predictive equation for estimating serum ionized calcium levels in patients on chronic hemodialysis // <i>Medical Science Monitor: International Medical Journal of Experimental and Clinical Research.</i> – 2023. – Т. 29. – С. e941321-1. doi.org/10.12659/MSM.941321
7.	Plebani M. Harmonizing the post-analytical phase: focus on the laboratory report. <i>Clin Chem Lab Med.</i> 2024;62(6):1053-1062. Published 2024 Jan 8. doi:10.1515/cclm-2023-1402
8.	МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ПРИКАЗ от 27 апреля 2021 г. N 404н ОБ УТВЕРЖДЕНИИ ПОРЯДКА ПРОВЕДЕНИЯ ПРОФИЛАКТИЧЕСКОГО МЕДИЦИНСКОГО ОСМОТРА И ДИСПАНСЕРИЗАЦИИ ОПРЕДЕЛЕННЫХ ГРУПП ВЗРОСЛОГО НАСЕЛЕНИЯ [Электронный ресурс] – URL: <a href="https://normativ.kontur.ru/document?moduleId=1&amp;documentId=477456">https://normativ.kontur.ru/document?moduleId=1&amp;documentId=477456</a>
9.	<a href="https://preventmed.space/contest">https://preventmed.space/contest</a> (см. кейс N4 и 5 для конкурса, озвученные Российским обществом профилактики неинфекционных заболеваний и Главным внештатным специалистом по терапии и общей врачебной практике Драпкиной О.М.)
10.	ТАСС: новости о новых регионах // ТАСС. – URL: <a href="https://tass.ru/obschestvo/23506323">https://tass.ru/obschestvo/23506323</a>
11.	Мурашко сообщил о наличии хронических неинфекционных заболеваний у 45% россиян // Интерфакс. – URL: <a href="https://www.interfax.ru/russia/1015238">https://www.interfax.ru/russia/1015238</a>
12.	В Минздраве оценили ущерб экономике от хронических заболеваний населения // РБК. – URL: <a href="https://www.rbc.ru/economics/05/06/2024/666069359a79475e59400016">https://www.rbc.ru/economics/05/06/2024/666069359a79475e59400016</a>
13.	Национальный проект «Продолжительная и активная жизнь» // Правительство России : официальный сайт. –URL: <a href="http://government.ru/rugovclassifier/section/2641/">http://government.ru/rugovclassifier/section/2641/</a>
14.	Национальный проект «Новые технологии сбережения здоровья» // Правительство России : официальный сайт. –URL: <a href="http://government.ru/rugovclassifier/section/2641/">http://government.ru/rugovclassifier/section/2641/</a>
15.	В фокусе внимания — хронические заболевания // Коммерсантъ. – 07.06.2024. – URL: <a href="https://www.kommersant.ru/doc/6745632">https://www.kommersant.ru/doc/6745632</a>

## 1.7. ИСПОЛЬЗОВАННЫЕ МЕТОДЫ

В работе применялся комплекс междисциплинарных методов:

- Методы Data Science и МО: ансамбль методов на основе линейной регрессии и градиентного бустинга на решающих деревьях для алгоритмов регрессии и классификации, методы предобработки и валидации данных.
- Клинико-лабораторные методы: ретроспективный анализ парных лабораторных данных “предикторы” - “целевая переменная”, формирование клинически релевантных дата-сетов.
- Методы разработки программного обеспечения: микросервисная архитектура, REST API, контейнеризация (Docker).
- Статистические методы с помощью пакета `scipy.stats`, оценка метрик качества моделей: коэффициент детерминации ( $R^2$ ), средняя абсолютная ошибка (MAE), средняя абсолютная

ошибка в процентах (MAPE), площадь под кривой ROC-анализа (AUC-ROC), чувствительность, специфичность.

## **2. ОСНОВНАЯ ЧАСТЬ**

### **2.1. ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ И МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ**

#### **2.1.1. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ПЕРСОНАЛИЗИРОВАННОЙ ДИАГНОСТИКИ В СИСТЕМЕ ЛАБОРАТОРНОЙ МЕДИЦИНЫ**

Проект «ЛабЧекап» основан на междисциплинарном подходе. В основе диагностики лежат строгие взаимосвязи между лабораторными показателями, которые существующие формулы (например, Фридвальда, Пэйна) описывают упрощенно. Наш подход использует МО для точного моделирования этих сложных зависимостей. МО реализует принцип обучения на уникальной базе данных пар «рутинные показатели – эталонный результат» и позволяет выявлять скрытые закономерности. Строгая валидация гарантирует, что модели не просто запоминают данные, а способны к обобщению. С позиции системной медицины платформа становится инструментом персонализированной медицины (предиктивность, профилактика, персонализация), автоматизируя цикл контроля качества и решая ключевое противоречие скрининга между эффективностью и затратами. ИИ позволяет «виртуально» расширить диагностическую панель без увеличения расходов. Таким образом, гипотеза проекта заключается в том, что выявление и формализация сложных клинико-лабораторных взаимосвязей методами ИИ позволяет создать «цифрового двойника» точного анализа, работающего в масштабах всей системы здравоохранения.

#### **2.1.2. МЕТОДОЛОГИЯ ФОРМИРОВАНИЯ ИССЛЕДОВАТЕЛЬСКОЙ БАЗЫ ДАННЫХ И РАЗРАБОТКИ МОДЕЛЕЙ**

Ключевым фактором, обеспечивающим высокую точность решения, стала уникальная методология формирования дата-сета для обучения. Для каждого прогнозируемого маркера собиралась парная выборка, где в один временной промежуток (в рамках одного визита пациента) были выполнены:

– Входные данные (предикторы): стандартный набор первого этапа диспансеризации (пол, возраст, ОАК, СОЭ, ОХС, глюкоза плазмы натощак (ГПН) или другие комбинации рутинных тестов.

– Целевая переменная (золотой стандарт): референсное измерение искомого маркера прямым лабораторным методом («золотой стандарт»), например, колориметрический фотометрический метод для ХС-ЛПНП или ВЭЖХ для HbA1c, иммунохимический метод для определения ферритина.

**Критерии формирования дата-сета:**

- Клиническая релевантность. Использовались только парные данные, что исключало влияние временных изменений в состоянии пациента.
- Аппаратная независимость. Данные были собраны с аналитических тест-систем ведущих мировых и российских производителей (Roche, Abbott, Siemens, Mindray, Вектор-Бест и др.), что обеспечило устойчивость моделей к вариациям методик.
- Репрезентативность. Выборка включала данные из нескольких сетевых лабораторий, представляющих разные регионы, полученных от лабораторий-партнеров (ООО "Лаборатория Гемотест", ООО "Диалаб Плюс", ГБУЗ «МНПЦЛИ ДЗМ», другие). Общий объем обработанных и использованных для обучения обезличенных лабораторных отчетов превысил 1,5 млн пациентов. При проведении ретроспективного исследования, а также при подготовке научных публикаций на его основе, руководствовались положениями Хельсинкской декларации (в редакции 2013 г.). В соответствии со статьей 32 данного документа получение индивидуального информированного согласия пациентов для ретроспективного анализа обезличенных медицинских данных не является обязательным. Все данные, использованные для разработки и обучения моделей ИИ, были предварительно де-идентифицированы. Работа с такими наборами данных регламентирована договорами о передаче информации, заключенными с партнерскими сетевыми лабораториями в полном соответствии с требованиями Федерального закона № 152-ФЗ «О персональных данных».

**Алгоритмический стек и валидация.** Для задач регрессии (прогноз концентрации) основной алгоритм – линейная регрессия и градиентный бустинг на решающих деревьях. Для задач классификации (оценка риска) также использовался ансамбль методов на основе LightGBM и CatBoost. Для каждой модели применялся строгий протокол валидации: стратифицированное разделение на обучающую (80%) и тестовую (20%) выборки, кросс-валидация для настройки гиперпараметров. Все итоговые метрики представлены исключительно по изолированной тестовой выборке, которую модель никогда «не видела».

## **2.2. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ И РЕЗУЛЬТАТЫ**

### **2.2.1. ВАЛИДАЦИЯ ТОЧНОСТИ ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ**

На независимых тестовых выборках были получены следующие ключевые метрики (20% от исходного набора данных), подтверждающие высокую точность моделей: контроля качества (табл. 1), регрессионных и классификационных моделей первого этапа диспансеризации (табл. 2,3), классификационных прогностических моделей на основе рутинных лабораторных исследований (табл. 4).

Таблица 1. Метрики регрессионных моделей модуля ИИ-контроля качества

Модель	Метрики точности	Набор данных, n
Прогноз уровня ХС-ЛПНП по данным липидограммы (пол, возраст, ОХС, ТГ, ХС-ЛПВП)	$R^2=0,95$ в прогнозе ХС-ЛПНП, превосходит традиционные формулы при уровнях ТГ>4,5 ммоль/л: МАЕ между предсказанным и известным значением ХС-ЛПНП составила 0,476 против 1,173, 0,562, 0,796 ммоль/л, рассчитанным по формулам Фридвальда, Мартина и Сэмпсона, соотв.	631 267
Прогноз уровня ионизированного кальция ( $iCa^{2+}$ ) по данным показателей, входящих в формулу Пэйна (пол, возраст, альбумин, кальций общий)	$R^2=0,64$ в прогнозе $iCa^{2+}$ . Модель превосходит традиционные формулы расчета $Ca^{2+}$ во всех клинически значимых диапазонах: МАЕ между рассчитанным значением и измеренным составила: 0,034 (на всём диапазоне), 0,043 (гипокальциемия), 0,031 (нормокальциемия), 0,054 (гиперкальциемия) ммоль/л; для формулы 1 ( $Ca^{2+}$ = общий кальций $\times$ 0,5): 0,042, 0,059, 0,038, 0,071 ммоль/л; для формулы 2 ( $Ca^{2+}$ = (общий кальций + 0,02 $\times$ (40-альбумин)) $\times$ 0,5): 0,045, 0,049, 0,044, 0,057 ммоль/л.	24 780
Прогноз уровня истинного уровня $K^+$ по общему белку, кальцию, креатинину, мочеvine, ГПН, $Na^+$ , $Cl^-$ , ОАК и индексу гемоллиза	$R^2=0,42$ , медианная абсолютная ошибка между значениями прогнозируемого и измеренного уровня калия составила 0,34 ммоль/л.	758

Таблица 2. Метрики точности регрессионных моделей прогноза концентраций тестов на основе данных первого этапа диспансеризации

Модель	Метрики точности	Набор данных, n
Прогноз уровня ХС-ЛПНП (ЛипидоСкрининг)	Тестовая выборка: $R^2=0,87$ , МАЕ=0,27 ммоль/л, МАРЕ=9%. Валидационная выборка n=10 000 пациентов ГБУЗ «МНПЦЛИ ДЗМ» - пилотное тестирование фонда «МИК»: $R^2=0,90$ , МАЕ=0,28 ммоль/л, МАРЕ=10%; в сети клиник «Будь Здоров»: $R^2=0,86$ , МАЕ=0,28 ммоль/л.	279 325
Прогноз уровня ХС-неЛВП	$R^2=0,93$ , МАЕ=0,25 ммоль/л, МАРЕ = 7%	279 325
Прогноз индекса TyG	$R^2=0,60$ , МАЕ=0,16	279 325
Прогноз уровня HbA1c (ДиаСкрининг)	Тестовая выборка: $R^2=0,77$ , МАЕ=0,39%, МАРЕ=6%. Валидационная выборка n=101 пациентов ГБУЗ «ГП №220 ДЗМ» - пилотное тестирование при содействии Московского центра инновационных технологий в здравоохранении: $R^2=0,57$ , МАЕ=0,62%. Валидационная выборка n=180 пациентов ГБУЗ «КДП №121 ДЗМ» - пилотное тестирование: $R^2=0,55$ , МАЕ = 0,65%.	54 952
Прогноз уровня ферритина	МАЕ=15,3 мкг/л	133 568
Прогноз уровня железа	МАЕ=3,5 мкмоль/л	133 568
Прогноз уровня фруктозамина	Тестовая выборка: $R^2=0,61$ , МАЕ=29,2 мкмоль/л, МАРЕ=13%. При валидации модели на когорте пациентов с анемией (n=72) получены сопоставимые показатели точности по сравнению с общей выборкой: $R^2=0,64$ , МАЕ=23,7 мкмоль/л, МАРЕ=10%.	735

Таблица 3. Метрики точности классификационных моделей стратификации риска ХНИЗ на основе данных первого этапа диспансеризации

Модель	Метрики точности	Набор данных, n
--------	------------------	-----------------

↑ ХС-ЛПНП ≥3,0 ммоль/л	AUC ROC = 0,97, точность = 0,90, полнота = 0,90 и точность положительного прогноза = 0,93. Валидационная выборка n=10 000 пациентов ГБУЗ «МНПЦЛИ ДЗМ» - пилот фонда «МИК»: AUC ROC = 0,97, точность = 0,91, чувствительность 92%, специфичность 89%. AUC ROC в прогнозе ХС-ЛПНП ≥3,0 ммоль/л у лиц с нормальным уровнем ОХС (<5 ммоль/л) составил 0,91 на представленной выборке (n=432, 4,3%)	279 325
↑ HbA1c ≥6,5%	AUC ROC=0,96, точность=0,93, полнота=0,89 и точность положительного прогноза=0,87. AUC ROC в прогнозе HbA1c ≥6,5 % у лиц с нормальным уровнем ГПН (<6,1 ммоль/л) составил 0,83 (ранняя диагностика, n=1 314) на тестовой выборке. Валидационная выборка n=101 пациентов ГБУЗ «ГП №220 ДЗМ» - пилотное тестирование при содействии Московского центра инновационных технологий в здравоохранении: AUC ROC=0,82.	54 952
↑ индекса HOMA-IR >2,7	AUC-ROC=0,86	50 574
↓ ферритина <15,0 мкг/л	AUC-ROC=0,96	133 568
↓ сывороточного железа <10,7 мкмоль/л)	AUC-ROC=0,96	133 568
↑ NTproBNP >125 пг/мл	AUC-ROC=0,88	1 667
↑ BNP >35 пг/мл	AUC-ROC=0,75	780
↑ гомоцистеина >15 мкмоль/л	AUC-ROC=0,76	20 145
↑ мочевой кислоты >360 мкмоль/л (УрикоСкрининг)	AUC-ROC=0,77	84 335
↓ СКФ <60 мл/мин/1,73 м <sup>2</sup>	AUC-ROC=0,92	170 759
Гипотиреоз (ТТГ >4,94 мЕд/л и Т4 св. <9,0 пмоль/л)	AUC-ROC=0,72, чувствительность 72%, специфичность 65%	105 625
Гипертиреоз (ТТГ <0,35 мЕд/л и Т4 св. >19,05 пмоль/л)	AUC-ROC=0,78, чувствительность 74%, специфичность 67%	105 625
↓ общего тестостерона <12 нмоль/л	AUC ROC=0,65	13 799
Снижение АМГ <1,2 нг/мл	AUC ROC=0,78	1464
↑ общего ПСА >4,0 нг/мл	AUC ROC=0,75	3 157
↑ АФП >10 нг/мл	AUC ROC=0,84	2 008
↑ СА-125 ≥30 МЕ/мл	AUC ROC=0,75	7 640
Обнаружение в кале гемоглобина (Hb)	AUC ROC=0,68	2 745
Обнаружение в кале Hb и/или гаптоглобина	AUC ROC=0,70	231
Обнаружение в кале Hb и/или трансферрина	AUC ROC=0,75	424
↑ СОЭ >20 мм/ч для М, >30 мм/ч для Ж	AUC ROC=0,84	62 148
↑ СРБ >5 мг/л	AUC ROC=0,81	62 148
↑ АЛТ >26,0 и >19,0 Ед/л у М и Ж	AUC ROC=0,70	199 898
↑ АСТ >40,0 и >32,0 Ед/л у М и Ж	AUC ROC=0,70	199 898
↑ креатинина >104,0 и >90,0 мкмоль/л у М и Ж	AUC ROC=0,78	170 581
↑ мочевины >7,1 и >8,2 ммоль/л у лиц 18-60 и >60 лет, соотв.	AUC ROC=0,80	170 581
↑ цистатина С >1,2 мг/л	AUC ROC=0,82	394
Прогноз “+” anti-HCV	AUC ROC=0,79, чувствительность 11%, специфичность 99%	66 487

Таблица 4. Метрики точности классификационных моделей стратификации риска ХНИЗ на основе данных рутинных анализов

Модель	Метрики точности	Набор данных, n
Прогноз выраженного фиброза печени на основе данных FIB-4 (АСТ, АЛТ, тромбоциты и возраст)	AUC ROC=0,82 превзошла диагностическую точность метода FIB-4 (AUC ROC=0,74) для верификации выраженного фиброза (F3-F4) по данным жесткости печени по FibroScan.	272
Прогноз клинически значимой мутации в гене UGT1A1 на основе результатов фракций билирубина, гемоглобина, АЛТ, АСТ, пола и возраста	Доля правильно выявленных пациентов с синдромом Жильбера (recall) среди всех, у кого действительно присутствует мутация 7ТА/7ТА составила 0,87.	1 003

### 2.2.2. АРХИТЕКТУРА И ГОТОВНОСТЬ К ВНЕДРЕНИЮ

Решение реализовано как промышленное API-решение:

- Технологический стек: Python, Flask, Scikit-learn, LightGBM/CatBoost, Docker, MongoDB, S3-совместимое хранилище.
- Архитектура: микросервисная, обеспечивающая отказоустойчивость и легкое обновление моделей.
- Интеграция: стандартизированные REST API для подключения к ЛИС.
- Производительность: время обработки одного запроса (от данных до результата) – менее 1 секунды, что позволяет работать в реальном времени.

### 2.2.3. АНАЛИЗ ВКЛАДА ПРЕДИКТОРОВ И БИОЛОГИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ МОДЕЛЕЙ

Для подтверждения научной обоснованности моделей был проведен анализ вклада предикторов (feature importance). Результаты демонстрируют, что ключевыми предикторами для прогноза являются клинически значимые и логичные параметры.

- Прогноз мочевой кислоты. ГПН и возраст – главные предикторы, что отражает связь гиперурикемии с метаболическим синдромом и возрастными изменениями. Лейкоциты (WBC) и СОЭ указывают на воспалительный компонент.
- Прогноз HbA1c. ГПН ожидаемо на первом месте. Возраст, WBC и эритроцитарные индексы (MCH, RDW, MCHC) показывают влияние хронической гипергликемии на кроветворение и клеточный метаболизм.
- Прогноз гомоцистеина. Возраст и MCV – ключевые предикторы, что соответствует известной связи гипергомоцистеинемии с возрастными изменениями и макроцитозом. ГПН и WBC отражают метаболические и воспалительные корреляции.
- Прогноз ХС-ЛПНП, не-ЛВП. ОХС – доминирующий предиктор. Пол, возраст, ГПН подтверждают влияние демографии и метаболизма на липидный профиль.

- Прогноз индекса TyG (ТГ-ГПН). ГПН и ОХС - ведущие факторы, что точно отражает суть индекса как маркера инсулинорезистентности и дислипидемии.
- Прогноз скрытой крови в кале. СОЭ и ОХС - неочевидные, но значимые предикторы, возможно указывающие на системное воспаление. Показатели лейкоцитарной формулы отражают иммунный ответ.
- Прогноз инсулина. ГПН - главный предиктор, что соответствует физиологии. WBC, эритроциты, СОЭ подчеркивают связь гиперинсулинемии с воспалением и оксидативным стрессом.
- Прогноз онкомаркеров (СА-125, АФП, ПСА). RDW (показатель анизоцитоза эритроцитов) и возраст стабильно входят в топ-предикторов, что может указывать на системные изменения при онкологических процессах. Для ПСА возраст - ожидаемый главный фактор.

Таким образом, модели выявляют не только очевидные прямые связи (например, ГПН в прогнозе HbA1c), но и сложные непрямые взаимосвязи (эритроцитарные индексы, биохимические маркеры), что демонстрирует способность ИИ обнаруживать глубинные патофизиологические паттерны. Это подтверждает, что решение работает как инструмент для системного анализа здоровья, а не просто как калькулятор отдельных параметров.

## **2.3. ВЫВОД О ЦЕННОСТИ, ПОЛЬЗЕ И ПЕРСПЕКТИВАХ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ**

### **2.3.1. НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ И КЛИНИЧЕСКАЯ ЦЕННОСТЬ**

1. Доказана осуществимость новой клинической методологии. Впервые предложена и валидирована методология построения предиктивных моделей для лабораторной диагностики на основе данных первого этапа диспансеризации. Экспериментально доказано, что на основе минимального набора рутинных данных возможно с лабораторной точностью прогнозировать концентрации ключевых биомаркеров и оценивать индивидуальные риски ХНИЗ. Клиническая ценность заключается в создании инструмента для качественного расширения скрининга без дополнительных затрат и инвазивных процедур для пациента, что позволяет выявлять ранее упускаемые группы риска (например, пациентов с нормальным ОХС, но повышенным ХС-ЛПНП).
2. Создан уникальный исследовательский актив. Сформирована одна из крупнейших в РФ специализированных баз данных «скрининг - эталон» объемом >1,5 млн записей. Этот набор данных представляет долгосрочную научную ценность для разработки и валидации новых диагностических алгоритмов в области цифровой медицины.
3. Получен патентный задел. Зарегистрировано 8 РИД (свидетельства на ПО и БД), поданы заявки на патенты и товарные знаки, что обеспечивает правовую защиту разработки.

4. Опубликовано >35 публикаций по теме проекта “ЛабЧекап” (см. приложение).

### 2.3.2. ПРАКТИЧЕСКАЯ ПОЛЬЗА И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЙ ЭФФЕКТ

Для системы ОМС технология «ЛабЧекап» позволяет кардинально снизить затраты на массовый профилактический скрининг, обеспечивая при этом высокую клиническую точность. Экономический эффект достигается за счет замены дорогостоящих лабораторных тестов на высокоточный ИИ-прогноз, выполняемый на основе уже собранных в ходе диспансеризации (табл. 5,6) рутинных данных (табл. 7,8), а также для ранней диагностики ХНИЗ. Бизнес-модель: гос. контракты, B2B SaaS-подписки.

Таблица 5. Экономическая эффективность прогноза уровня ХС-ЛПНП.

Параметр	Значение	Комментарий
Взрослое население РФ (N)	113 млн чел.	Целевая группа для скрининга
Распространенность дислипидемии	59,2%	Согласно данным ЭССЕ-РФ
Целевая группа для скрининга ХС-ЛПНП	66,9 млн чел.	$N \times 59,2\% = 113 \text{ млн} \times 0,592$
Затраты по традиционной схеме (прямое измерение):		
Стоимость теста ОМС (ХС-ЛПНП)	54,14 руб.	Тариф ОМС
Стоимость обработки пробы	24,72 руб.	Включает забор крови и логистику
Итого стоимость 1 анализа	78,86 руб.	
Годовые затраты на скрининг	5,28 млрд руб.	$66,9 \text{ млн чел.} \times 78,86 \text{ руб.}$
Затраты с использованием «ЛабЧекап»:		
Стоимость одного прогноза	0,5 руб.	Расчет ИИ-моделью
Годовые затраты на скрининг	33,5 млн руб.	$66,9 \text{ млн чел.} \times 0,5 \text{ руб.}$
Годовая экономия для системы ОМС	$\approx 5,24 \text{ млрд руб.}$	$5,28 \text{ млрд руб.} - 33,5 \text{ млн руб.}$
Снижение затрат	$\sim 99\%$	$(5,24 / 5,28 \text{ млрд руб.}) \times 100\%$

Таблица 6. Экономическая эффективность прогноза уровня HbA1c.

Параметр	Значение	Комментарий
Пациенты с СД в РФ (оценочно)	5,93 млн чел.	Взрослое население, нуждающееся в регулярном контроле HbA1c
Затраты по традиционной схеме:		
Стоимость теста ОМС (HbA1c) + обработка крови	269,51 руб.	Тариф ОМС
Годовые затраты на скрининг	$\sim 1,60 \text{ млрд руб.}$	$5,93 \text{ млн чел.} \times 269,51 \text{ руб.}$
Затраты с использованием «ЛабЧекап»:		
Стоимость одного прогноза	0,5 руб.	Расчет ИИ-моделью
Годовые затраты на скрининг	$\sim 3,0 \text{ млн руб.}$	$5,93 \text{ млн чел.} \times 0,5 \text{ руб.}$
Годовая экономия для системы ОМС	$\approx 1,60 \text{ млрд руб.}$	$1,60 \text{ млрд руб.} - 3,0 \text{ млн руб.}$
Снижение затрат	$\sim 99\%$	$(1,59 / 1,6 \text{ млрд руб.}) \times 100\%$

Таблица 7. Сравнение традиционного подхода и решения “ЛабЧекап” в скрининге липидного профиля на примере потока 10 000 пациентов в месяц.

Параметр	Значение	Комментарий
Параметр	Традиционная схема (Формула Фридвальда + прямое измерение)	Схема с использованием «ЛабЧекап» (ML-прогноз)
Охват пациентов	Раздельный подход: 70% (7 000 чел.) - расчет по формуле Фридвальда (при ТГ $\leq 4,5$ ммоль/л). 30% (3 000 чел.) - прямое измерение (при ТГ $>4,5$ ммоль/л).	Единый подход для 100% пациентов: 10 000 чел. - точный прогноз ML-моделью (без ограничений по ТГ).
Клиническая точность	Неоднородная:	Единая и высокая для всех:

	Потенциальная погрешность у 70% пациентов. Высокая точность только у 30% пациентов (прямое измерение).	Точность прогноза сопоставима с «золотым стандартом» для 100% пациентов ( $R^2 > 0,87$ ).
Затраты на 1 анализ/прогноз	0 руб. – для расчета по формуле. ~40 руб. – себестоимость реагентов.	0,5 руб. – стоимость одного прогноза ИИ-моделью.
Месячные затраты	120 000 руб. (3 000 анализов × 40 руб.)	5 000 руб. (10 000 прогнозов × 0,5 руб.)
Годовые затраты	~1 440 000 руб. (120 000 руб./мес. × 12)	~60 000 руб. (5 000 руб./мес. × 12)
Годовая экономия	—	~1 380 000 руб. (1 440 000 - 60 000 руб.)
Снижение затрат	—	~96% (1 380 000 / 1 440 000 × 100%)
Ключевой эффект	Двойной стандарт качества: часть пациентов получает результат с погрешностью.	Гарантированное высокое качество диагностики для всех пациентов при радикальном снижении издержек.

Таблица 8. Экономический эффект от внедрения «ЛабЧекап» для прогноза уровня ионизированного кальция на примере поломки электролитного анализатора.

Параметр	Аутсорсинг (традиционный путь)	Решение с «ЛабЧекап»	Эффект
Метод	Прямое измерение в сторонней лаборатории	Прогноз по Ca общ. + альбумин (ML-модель)	—
Стоимость 1 анализа	300 руб.	50,5 руб. (50 руб. + 0,5 руб.)	Снижение на 249,5 руб. (-83%)
Месячный объем (200 шт.)	60 000 руб.	10 100 руб.	Экономия: 49 900 руб./мес
Годовые затраты	720 000 руб.	121 200 руб.	Экономия: 598 800 руб./год
Точность	Высокая («золотой стандарт»)	Сопоставима с лабораторной ( $R^2=0,64$ , MAE=0,034 ммоль/л), превосходит формулы коррекции	Качество результата сохраняется

- Повышение эффективности: раннее выявление дополнительных 1-3% пациентов с скрытыми рисками (как показали пилоты) позволит вовремя начать профилактику, снизив будущие затраты на лечение осложнений ХНИЗ.
- Разгрузка врачей: автоматическая выдача готовой интерпретации с оценкой риска сокращает время на анализ результатов и принятие решений.

## 2. Для медицинских лабораторий:

- Снижение операционных затрат: ИИ-контроль качества сокращает брак проб, расход реагентов и время оборота теста.
- Повышение клинической ценности: врачи получают более достоверные данные и расширенную аналитику без кратного увеличения стоимости исследования.

3. Для пациента: получение более полной и точной информации о рисках здоровья на ранней, доклинической стадии, что позволяет вовремя принять превентивные меры и избежать развития тяжелых заболеваний.

### 2.3.3. ПЕРСПЕКТИВЫ РАЗВИТИЯ И МАСШТАБИРОВАНИЯ

**Стратегическая цель:** государственная регистрация разработанного нами ПО с ИИ «ЛабЧекап» как медицинского изделия 3 класса риска и обеспечить первые коммерческие внедрения в РФ до конца 2027 года.

**Задачи:**

1. Регуляторно-юридическая подготовка: (1) Внедрение СМК соответствующей требованиям ГОСТ ISO 13485-2017 до июля 2026, подготовка пакета документов для регистрации (Технические Условия, Эксплуатационная Документация, Документация Жизненного Цикла ПО согласно ГОСТ 62304-2022, Нормативная Документация, Верификация и Валидация ПО) к маю 2026; (2) Технические испытания в аккредитованном испытательном центре и Клинические испытания на базе сертифицированной лаборатории (RU-1027709005642) ООО «Лаборатория Гемотест» до мая 2027 или в лаборатории «СМД»; (3) Подача документов в Росздравнадзор к июню 2027.

2. Технологическое развитие: (1) Расширение базы данных до 5 млн записей и разработка алгоритмов адаптивного дообучения (2026); (2) Ретроспективное пилотное тестирование в  $\geq 5$  региональных учреждениях (2026); (3) Интеграция с ЛИС/МИС (ЕГИСЗ, ЕМИАС) в  $\geq 5$  учреждениях (2026–27); (4) Разработка  $\geq 50$  новых прогностических моделей (2026–27).

3. Научно-методическая работа: (1) Публикация  $\geq 15$  статей и  $\geq 50$  тезисов (2026–27); (2) Регистрация  $\geq 10$  РИД, в т.ч. в ЕАЭС (2026–27); (3) Разработка проекта клинических рекомендаций (2026); (4) Образовательные мероприятия и взаимодействие с ГВС и Минздравом (2026); (5) Участие в конференции в качестве экспонента ИТМ ИИ (февраль 2026).

4. Коммерческая подготовка: (1) Запуск B2B-модулей контроля качества до лета 2026; (2) Проведение коммерческих проспективных пилотов; (3) Заключение  $\geq 10$  соглашений о намерениях с сетевыми лабораториями до конца 2026; (4) Запуск MVP B2C-сервиса в III квартале 2026; (5) Подготовка пакета для госзакупок к I кварталу 2028.

### **3. ЗАКЛЮЧЕНИЕ**

Настоящий проект представляет собой законченный цикл научно-исследовательских и опытно-конструкторских работ, в результате которого создана и апробирована интеллектуальная платформа «ЛабЧекап». Основным научно-техническим результатом работы является доказательство эффективности применения ансамблей моделей машинного обучения для решения двух взаимосвязанных задач: сквозного интеллектуального контроля качества на всех этапах лабораторного процесса и предиктивной аналитики рисков хронических неинфекционных заболеваний на основе ограниченного набора рутинных скрининговых данных. Валидация разработанных алгоритмов на независимых тестовых выборках, включая пилотные испытания в клинических учреждениях, продемонстрировала метрики точности ( $R^2$  до 0,95; AUC-ROC до 0,97; MAPE 6–10%), статистически значимо превосходящие традиционные расчетные

методики и сопоставимые с референсными лабораторными методами, что подтверждает достижение заявленной цели.

Сформированная в процессе исследований уникальная база данных «рутинный скрининг – золотой стандарт» объемом свыше 1,5 млн обезличенных записей пациентов представляет собой уникальный научный задел и исследовательскую основу для дальнейшего развития методов цифровой медицины. Практическая значимость полученных результатов заключается в создании готового к промышленной интеграции API-сервиса, который обеспечивает не только экономический эффект за счет виртуального расширения диагностических возможностей при пропорциональном снижении затрат, но и качественный сдвиг в организации профилактического звена. Платформа позволяет трансформировать массовый скрининг в персонализированный инструмент превентивной медицины, способствуя ранней доклинической диагностике, что полностью соответствует стратегическим ориентирам национальных проектов в сфере здравоохранения и цифровизации.

Таким образом, проект завершился созданием научно-обоснованного и технологически реализованного решения, прошедшего стадию клинической апробации и обладающего высоким потенциалом для масштабирования после получения регистрационного удостоверения. Полученные результаты закладывают основу для формирования нового стандарта в области интеллектуальной обработки лабораторных данных, направленного на повышение доступности, эффективности и качества медицинской помощи на национальном уровне.