

Харисова Зарина Ирековна кандидат технических наук, доцент кафедры
БМИ, Уфимский Уфимский университет науки и технологий
Россия, г. Уфа

Габдрафиков Аскар Ирикович магистрант Уфимский университет науки и
технологий
Россия, г. Уфа

ORCID 0009-0004-4710-3848

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАННЕЙ ДИАГНОСТИКИ ХРОНИЧЕСКОЙ БОЛЕЗНИ ПОЧЕК НА ОСНОВЕ ЛАБОРАТОРНЫХ ПОКАЗАТЕЛЕЙ

Аннотация: В статье рассматривается задача повышения точности ранней диагностики хронической болезни почек за счёт применения методов машинного обучения. Основное внимание уделено обоснованию использования алгоритмов классификации для анализа лабораторных показателей пациентов, а также исследованию влияния предварительной обработки данных на итоговое качество модели.

Ключевые слова: машинное обучение, хроническая болезнь почек, медицинская диагностика, классификация, лабораторные данные.

Annotation: This article examines the problem of improving the accuracy of early diagnosis of chronic kidney disease through the use of machine learning methods. The focus is on the rationale for using classification algorithms to analyze patient laboratory parameters, as well as the impact of data preprocessing on the final model performance.

Keywords: machine learning, chronic kidney disease, medical diagnostics, classification, laboratory data.

Хроническая болезнь почек относится к числу заболеваний, которые длительное время могут развиваться без выраженных симптомов, что существенно осложняет их своевременное выявление [10], [11]. На практике постановка диагноза основывается на анализе лабораторных показателей, отражающих состояние организма пациента. При этом врач вынужден учитывать сразу несколько параметров, между которыми могут существовать сложные взаимосвязи [4], [5].

Повышение уровня креатинина может свидетельствовать о снижении функции почек, однако само по себе это значение не всегда является достаточным для постановки диагноза. В сочетании с повышенным уровнем мочевины и пониженным гемоглобином вероятность наличия заболевания существенно возрастает, что подтверждается результатами современных исследований [2], [7]. Учет подобных комбинаций показателей в ручном режиме затруднен, особенно при большом объеме данных, что обуславливает необходимость автоматизации анализа [3], [12].

С увеличением объёма медицинских данных возрастает и сложность их интерпретации, что делает актуальным применение методов, способных выявлять скрытые зависимости между признаками [17], [18].

Рассматриваемая задача сводится к определению наличия или отсутствия хронической болезни почек на основе совокупности лабораторных показателей пациента. Такой подход соответствует классической задаче бинарной классификации, широко применяемой в медицинских исследованиях [1], [6]. Модель должна формировать вероятностную оценку наличия заболевания на основе входных данных.

Для иллюстрации можно рассмотреть пример: пациент с возрастом 55 лет, уровнем креатинина 2.1 мг/дл, мочевины 60 мг/дл и гемоглобина 9 г/дл относится к группе повышенного риска. Подобные зависимости эффективно выявляются алгоритмами машинного обучения, что подтверждается рядом работ [3], [8].

Выбор методов машинного обучения обусловлен их способностью учитывать как линейные, так и нелинейные зависимости между признаками, что делает их особенно эффективными при анализе медицинских данных [15], [16].

Используемый набор данных включает лабораторные показатели пациентов, представленные в виде числовых и категориальных признаков. Аналогичные структуры данных широко применяются в исследованиях, посвящённых диагностике хронической болезни почек [12], [19].

На практике данные выглядят следующим образом:

Пациент А — age: 48, bgr: 121, bu: 36, hemo: 15.4 → заболевание отсутствует.

Пациент В — age: 62, bgr: 180, bu: 72, hemo: 8.9 → заболевание присутствует.

Подобные примеры показывают, что различие между классами определяется не одним параметром, а их комбинацией [5], [14].

При этом данные содержат пропущенные значения. Например, у части пациентов может отсутствовать показатель гемоглобина или плотности мочи, что является типичной ситуацией в медицинских исследованиях.

Предобработка данных

Подготовка данных направлена на устранение проблем, мешающих корректному обучению модели.

Если в записи пациента отсутствует значение, например, уровня гемоглобина, модель не сможет напрямую обработать такой объект. Поэтому пропущенные значения заменяются медианными значениями по выборке. Например, если средний уровень гемоглобина составляет 12 г/дл, то именно это значение подставляется вместо пропуска.

Категориальные признаки, такие как наличие альбумина (например, «normal» или «abnormal»), преобразуются в числовой формат, например 0 и 1, чтобы алгоритм мог их использовать [6], [13].

Нормализация данных позволяет устранить ситуацию, когда один признак имеет диапазон от 0 до 100, а другой — от 0 до 10, что может исказить процесс обучения.

Построение моделей

В рамках исследования были использованы несколько моделей, каждая из которых по-разному обрабатывает входные данные.

Логистическая регрессия строит зависимость между признаками и вероятностью заболевания. Например, при увеличении уровня креатинина вероятность заболевания может возрасти линейно [10].

Случайный лес работает иначе: он разбивает пространство признаков на области. Например, модель может «решить», что если креатинин выше 1.5 и гемоглобин ниже 10, то пациент относится к группе риска [3].

Градиентный бустинг идёт ещё дальше, последовательно уточняя такие правила. Например, сначала он может ошибочно классифицировать пациента с умеренными значениями, а затем корректирует модель, учитывая дополнительные признаки, такие как уровень глюкозы [8].

Оценка качества моделей

Для оценки качества моделей используются метрики, позволяющие понять, насколько корректно модель работает в разных ситуациях. Пример, из 100 пациентов модель определила 90 правильно. Это даёт точность 90%, однако важно понять, сколько больных пациентов было обнаружено. Если из 50 реально больных модель нашла только 35, это означает, что часть случаев была пропущена, что недопустимо в медицинской практике [9], [11]. Именно поэтому дополнительно используется метрика recall, отражающая полноту выявления заболевания.

Результаты и их анализ

Ансамблевые методы обеспечивают более высокое качество классификации. Например, модель градиентного бустинга смогла корректно классифицировать пациентов со сложными комбинациями показателей, где простые модели допускали ошибки.

Так, для пациента с умеренно повышенным креатинином, но сильно пониженным гемоглобином, простая модель могла дать неопределённый результат, тогда как более сложная модель правильно отнесла его к группе риска.

Библиографический список

1. Abdelhag M.E. Систематический обзор методов машинного обучения для диагностики и прогнозирования хронической болезни почек [электронный ресурс] // The Saudi Journal of Applied Sciences and Technology. URL: <https://journals.jazanu.edu.sa/ojs/index.php/SJAST/article/view/1151> (дата обращения: 29.04.2026).
2. Li J., Du X., Zhang R. и др. Модели машинного обучения для прогнозирования краткосрочного прогрессирования хронической болезни почек 4 стадии [электронный ресурс] // Scientific Reports. URL: <https://www.nature.com/articles/s41598-025-23037-4> (дата обращения: 29.04.2026).
3. Iftikhar H., Hashem A.F., Qureshi M. и др. Применение моделей машинного обучения для ранней диагностики хронической болезни почек

[электронный ресурс] // Diagnostics.
URL: <https://pubmed.ncbi.nlm.nih.gov/41153282/> (дата обращения: 29.04.2026).

4. Gogoi P., Valan J.A. Методы машинного обучения для диагностики и прогнозирования хронической болезни почек: современные подходы и проблемы [электронный ресурс] // International Urology and Nephrology.
URL: <https://pubmed.ncbi.nlm.nih.gov/39560857/> (дата обращения: 29.04.2026).

5. Nguyen D.P., Nguyen T.T., Linh V.T.T. и др. Методы машинного обучения при хронических заболеваниях почек: сравнительный анализ [электронный ресурс] // Bioinformatics and Biology Insights.
URL: <https://journals.sagepub.com/doi/10.1177/11779322251356563> (дата обращения: 29.04.2026).

6. Ahmed I., Chowdhury T.E. и др. Анализ эффективности алгоритмов машинного обучения для прогнозирования хронической болезни почек [электронный ресурс] // arXiv.
URL: <https://arxiv.org/abs/2510.09493> (дата обращения: 29.04.2026).

7. Ma J., Wang J., Lu L. и др. Разработка динамической модели прогнозирования почечной недостаточности на основе глубокого обучения [электронный ресурс] // arXiv.
URL: <https://arxiv.org/abs/2501.16388> (дата обращения: 29.04.2026).

8. Haque M.E., Islam S.M.J. и др. Повышение точности диагностики хронической болезни почек с использованием CatBoost и объяснимого ИИ [электронный ресурс] // arXiv.
URL: <https://arxiv.org/abs/2504.04262> (дата обращения: 29.04.2026).

9. Ayub A.B., Niha N.S. Глубокая нейронная сеть для раннего прогнозирования хронической болезни почек с использованием объяснимого ИИ [электронный ресурс] // arXiv.
URL: <https://arxiv.org/abs/2512.09244> (дата обращения: 29.04.2026).

10. Islam M.A., Majumder M.Z.H., Hussein M.A. Прогнозирование хронической болезни почек с использованием алгоритмов машинного обучения [электронный ресурс] // Journal of Pathology Informatics. URL: <https://pubmed.ncbi.nlm.nih.gov/36714452/> (дата обращения: 29.04.2026).

11. Takavatakarn K., Oh W., Cheng E. и др. Модели машинного обучения для прогнозирования терминальной стадии почечной недостаточности [электронный ресурс] // BMC Nephrology. URL: <https://pubmed.ncbi.nlm.nih.gov/38114923/> (дата обращения: 29.04.2026).

12. Debal D.A., Sitote T.M. Прогнозирование хронической болезни почек с использованием методов машинного обучения [электронный ресурс] // Journal of Big Data. URL: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00657-5> (дата обращения: 29.04.2026).

13. Shamrat F.J.M., Ghosh P., Sadek M.H. Реализация алгоритмов машинного обучения для прогнозирования заболеваний почек [электронный ресурс] // Материалы международной конференции по вычислительным технологиям. URL: <https://ieeexplore.ieee.org/document/9038804> (дата обращения: 29.04.2026).

14. Charleonnann A., Fufaung T., Niyomwong T. Прогностическая аналитика хронической болезни почек с использованием методов машинного обучения [электронный ресурс] // Материалы международной конференции по информационным технологиям. URL: <https://ieeexplore.ieee.org/document/9155414> (дата обращения: 29.04.2026).

15. Esteva A., Robicquet A., Ramsundar B. и др. Руководство по применению глубокого обучения в здравоохранении [электронный ресурс] // Nature Medicine. T.27 URL: <https://doi.org/10.1038/s41591-021-01315-5> (дата обращения: 29.04.2026).

16. Topol E.J. Высокоточная медицина: синтез искусственного интеллекта и клинической практики [электронный ресурс] // Nature Medicine.

Т. 25 .URL: <https://doi.org/10.1038/s41591-018-0300-7> (дата обращения: 29.04.2026).

17. Beam A.L., Kohane I.S. Большие данные и машинное обучение в здравоохранении [электронный ресурс] // JAMA. Т. 319, № 13. URL: <https://doi.org/10.1001/jama.2017.18391> (дата обращения: 29.04.2026).

18. Rajkomar A., Dean J., Kohane I. Машинное обучение в медицине [электронный ресурс] // New England Journal of Medicine. Т. 380. URL: <https://doi.org/10.1056/NEJMra1814259> (дата обращения: 29.04.2026).

19. Dua D., Graff С. Репозиторий машинного обучения UCI: набор данных по хронической болезни почек. [электронный ресурс] URL: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease (дата обращения: 29.04.2026).