

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

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ПОЧЕК

Аннотация: В статье рассматривается применение методов машинного обучения для решения задачи диагностики заболеваний почек на основе лабораторных данных. Актуальность исследования обусловлена высокой распространённостью хронической болезни почек и необходимостью её раннего выявления.

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

Annotation: This article examines the application of machine learning methods to the problem of diagnosing kidney diseases based on laboratory data. The relevance of the study is due to the high prevalence of chronic kidney disease and the need for its early detection.

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

Хроническая болезнь почек (ХБП) является одним из наиболее распространённых заболеваний, характеризующихся длительным бессимптомным течением [1]. На ранних стадиях заболевание часто не диагностируется, что приводит к ухудшению состояния пациента и усложняет лечение [3].

Современные методы диагностики требуют анализа большого количества лабораторных показателей и высокой квалификации медицинского персонала [4]. В связи с этим возрастает интерес к применению методов машинного обучения, способных автоматизировать процесс диагностики и повысить её точность [5].

В задачах медицинской диагностики широко применяются методы машинного обучения, включая логистическую регрессию, метод опорных

векторов, деревья решений и ансамблевые методы [6]. Эти алгоритмы позволяют эффективно решать задачи классификации и прогнозирования [7].

Метод случайного леса является одним из наиболее популярных алгоритмов благодаря высокой точности и устойчивости к переобучению [4]. Он основан на построении ансамбля решающих деревьев и использовании механизма голосования [8]. Также активно применяются нейронные сети, способные выявлять сложные нелинейные зависимости в данных [9].

В работе использован набор данных с платформы Kaggle [10], содержащий клинично-лабораторные показатели пациентов. К числу признаков относятся возраст (age), уровень глюкозы в крови (bgr), содержание мочевины (bu), уровень гемоглобина (hemo), креатинин (sc) и другие параметры [11]. Предварительная обработка включала заполнение пропущенных значений, нормализацию данных и кодирование категориальных признаков [12].

Анализ распределения признака hemo показал, что у пациентов с заболеванием почек наблюдается пониженный уровень гемоглобина по сравнению со здоровыми пациентами. Аналогично, показатель bu (мочевина) имеет тенденцию к увеличению при наличии патологии.

Данные были разделены на обучающую и тестовую выборки в соотношении 80/20 [13]. В качестве основного алгоритма использован метод случайного леса [8].

Пример работы модели, для пациента со следующими параметрами: age = 48, bgr = 180, bu = 60, hemo = 9.5 модель классифицирует состояние как наличие заболевания почек с высокой вероятностью.

Для пациента: age = 32, bgr = 110, bu = 25, hemo = 14.2 модель относит его к классу здоровых.

Обучение модели проводилось с использованием Python и библиотеки Scikit-learn [14].

В результате была получена модель с точностью около 0,96 [15].

Если вероятность, выдаваемая моделью, превышает 0,5, пациент классифицируется как имеющий заболевание. Например, при вероятности 0,87 модель уверенно относит пациента к группе риска.

Анализ важности признаков показал, что наибольшее влияние оказывают hemo, bu и bgr [11].

Полученные результаты подтверждают эффективность применения методов машинного обучения для диагностики заболеваний почек [5]. Практические примеры показывают, что модель способна выявлять закономерности, соответствующие медицинской практике.

Однако следует учитывать зависимость результатов от качества данных и необходимость клинической валидации [17].

Библиографический список

1. Левин А. и др. Глобальное состояние здоровья почек [электронный ресурс] / URL: [https://www.thelancet.com/article/S0140-6736\(17\)30788-2/fulltext](https://www.thelancet.com/article/S0140-6736(17)30788-2/fulltext) (дата обращения: 25.04.2026).
2. Обермейер З., Эмануэль Э. Прогнозирование будущего — большие данные в медицине [электронный ресурс] / URL: <https://www.nejm.org/doi/full/10.1056/NEJMp1606181> (дата обращения: 25.04.2026).
3. Вебстер А.С. и др. Хроническая болезнь почек [электронный ресурс] / URL: [https://www.thelancet.com/article/S0140-6736\(16\)32064-5/fulltext](https://www.thelancet.com/article/S0140-6736(16)32064-5/fulltext) (дата обращения: 25.04.2026).
4. Брейман Л. Случайные леса [электронный ресурс] / URL: <https://link.springer.com/article/10.1023/A:1010933404324> (дата обращения: 25.04.2026).
5. Прогнозирование хронической болезни почек на основе алгоритмов машинного обучения [электронный ресурс] / URL: <https://ijitce.org/index.php/ijitce/article/view/1189> (дата обращения: 25.04.2026).
6. Коциантис С. Контролируемое машинное обучение: обзор [электронный ресурс] / URL: <https://www.informatica.si/index.php/informatica/article/view/148> (дата обращения: 25.04.2026).
7. Митчелл Т. Машинное обучение [электронный ресурс] / URL: <http://www.cs.cmu.edu/~tom/mlbook.html> (дата обращения: 25.04.2026).

8. Хастии Т., Тибширани Р., Фридман Дж. Элементы статистического обучения [электронный ресурс] / URL: <https://hastie.su.domains/ElemStatLearn/> (дата обращения: 25.04.2026).
9. Лиав А., Винер М. Классификация и регрессия с использованием randomForest [электронный ресурс] / URL: https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf (дата обращения: 25.04.2026).
10. Гудфеллоу И., Бенджио Й., Курвиль А. Глубокое обучение [электронный ресурс] / URL: <https://www.deeplearningbook.org> (дата обращения: 25.04.2026).
11. Набор данных по заболеваниям почек (Kaggle) [электронный ресурс] / URL: <https://www.kaggle.com/datasets/mansoordaku/ckdisease> (дата обращения: 25.04.2026).
12. Дуа Д., Графф К. Репозиторий машинного обучения UCI [электронный ресурс] / URL: <https://archive.ics.uci.edu> (дата обращения: 25.04.2026).
13. Хан Дж., Камбер М. Интеллектуальный анализ данных: концепции и методы [электронный ресурс] / URL: <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques> (дата обращения: 25.04.2026).
14. Жерон А. Практическое машинное обучение с использованием Scikit-Learn [электронный ресурс] / URL: <https://github.com/ageron/handson-ml> (дата обращения: 25.04.2026).
15. Педрегоса Ф. и др. Scikit-learn: машинное обучение на Python [электронный ресурс] / URL: <https://jmlr.org/papers/v12/pedregosa11a.html> (дата обращения: 25.04.2026).
16. Чикко Д. Десять практических советов по машинному обучению в вычислительной биологии [электронный ресурс] / URL: <https://biodatamining.biomedcentral.com/articles/10.1186/s13040-017-0155-3> (дата обращения: 25.04.2026).

17. Эстева А. и др. Руководство по глубокому обучению в здравоохранении [электронный ресурс] / URL: <https://www.nature.com/articles/s41591-018-0316-z> (дата обращения: 25.04.2026).

18. Топол Э. Глубокая медицина [электронный ресурс] / URL: <https://erictopol.com/deep-medicine> (дата обращения: 25.04.2026).

19. Раджкомар А. и др. Машинное обучение в медицине [электронный ресурс] / URL: <https://www.nejm.org/doi/full/10.1056/NEJMra1814259> (дата обращения: 25.04.2026).

20. Чжан З. Машинное обучение в анализе медицинских данных [электронный ресурс] / URL: <https://atm.amegroups.org/article/view/11025> (дата обращения: 25.04.2026).