

СЕДАШКИНА О.А.,

к.м.н., доцент, ФГБОУ ВО СамГМУ Минздрава России, г. Самара, Россия, e-mail: sedashkina@inbox.ru

КОЛСАНОВ А.В.,д.м.н., профессор, профессор РАН, ФГБОУ ВО СамГМУ Минздрава России, г. Самара, Россия,
e-mail: a.v.kolsanov@samsmu.ru

РАННЯЯ ДИАГНОСТИКА ХРОНИЧЕСКОЙ БОЛЕЗНИ ПОЧЕК У ДЕТЕЙ С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

DOI: 10.25881/18110193_2024_3_72

Аннотация. Актуальность. Диагностика ранней стадии хронической болезни почек (ХБП) является глобальной проблемой, поскольку чаще диагностируются поздние стадии заболевания. Разработка методов моделирования для принятия управленческих решений, направленных на повышение эффективности ранней диагностики ХБП, является важной научно-практической задачей, в решении которой большую поддержку может оказать использование алгоритмов машинного обучения (MLA).

Цель. Повышение точности диагностики ХБП с использованием данных анамнеза, клинико-инструментального, генетического обследования и MLA.

Материал и методы. Данные были получены из одностороннего ретроспективного когортного исследования (2011–2022 гг.) детей с ХБП 1-4 стадии в возрасте от 1 до 17 лет. В основную группу включены 128 детей с хроническими заболеваниями почек, в группу сравнения – 30 детей без патологии почек. Дети двух групп статистически значимо не различались по полу и возрасту. Для построения модели диагностики ХБП использованы данные анамнеза, клинико-инструментального и генетического обследования. Модель построена с применением MLA многофакторная логистическая регрессия (MLR). В модели использовано три переменных: СОЭ ($\beta = 0,392$; $p < 0,001$); эритроциты в моче ($\beta = 1,225$; $p < 0,001$); астеническое телосложение ($\beta = 5,792$; $p < 0,001$).

Результаты. Получена диагностическая модель, позволяющая на тестовой выборке выявлять ХБП с точностью 90,3% [80,6; 96,8]%, чувствительностью 92,0% [81,5; 100,0]%, специфичностью 83,3% [50,0; 100,0]%, ROC-AUC = 90,0% [77,2; 100,0]%. Полученная модель отличного качества ($> 90\%$), т.к. ROC-AUC составляет на тестовой выборке 0,90. Значение точки отсечения вероятности ХБП равно 0,25.

Выводы. Разработана и протестирована модель, которая с высокой точностью диагностирует на ранней стадии ХБП у детей.

Ключевые слова: хроническая болезнь почек у детей; машинное обучение; ранняя диагностика; боль в животе; вероятность; многофакторная логистическая регрессия.

Для цитирования: Седашкина О.А., Колсанов А.В. Ранняя диагностика хронической болезни почек у детей с помощью алгоритмов машинного обучения. Врач и информационные технологии. 2024; 3: 72-85. doi: 10.25881/18110193_2024_3_72.

SEDASHKINA O.A.,

PhD, Assoc. Prof., Samara State Medical University, Samara, Russia, e-mail: sedashkina@inbox.ru.

KOLSANOV A.V.,DSc, Prof., Prof. of the RAS, Samara State Medical University, Samara, Russia,
e-mail: a.v.kolsanov@samsmu.ru.

EARLY DIAGNOSIS OF CHRONIC KIDNEY DISEASE IN CHILDREN USING MACHINE LEARNING ALGORITHMS

DOI: 10.25881/18110193_2024_3_72

Abstract. *Background.* Diagnosis of early-stage chronic kidney disease (CKD) is a global challenge, as late-stage disease is more commonly diagnosed. The development of modeling methods for making management decisions aimed at improving the efficiency of early diagnosis of CKD is an important scientific and practical task, which can be greatly supported by the use of machine learning algorithms (MLA).

Aim. To improve the accuracy of diagnosis of CKD using data from history, clinical-instrumental, genetic examination and machine learning algorithms (MLA).

Methods. Data were obtained from a single-center retrospective catamnestic cohort study (2011–2022) of children with CKD stage 1–4 aged 1 to 17 years. The main group included 128 children with CKD, and the comparison group included 30 children without any kidney disease. Two groups were comparable by sex and age. The data of anamnesis, clinical-instrumental and genetic examination were used to build a model for CKD diagnosis. The model was built using the MLA multivariate logistic regression (MLR). Three variables were used in the model: erythrocyte sedimentation rate in blood ($\beta = 0,392$; $p < 0,001$); erythrocytes in urine ($\beta = 1,225$; $p < 0,001$); and asthenic physique ($\beta = 5,792$; $p < 0,001$).

Results. A diagnostics model was obtained allowing prediction of CKD on a test sample with accuracy of 90,3% [80,6; 96,8], sensitivity of 92,0% [81,5; 100,0], specificity of 83,3% [50,0; 100,0], ROC-AUC = 90,0% [77,2; 100,0]. The resulting model is of excellent quality (>90%) as the ROC-AUC is 0,90 on the test sample. The cut-off point value of the probability of CKD is 0,25.

Conclusions. We developed and tested the model that diagnoses early-stage CKD in children with high accuracy.

Keywords: chronic kidney disease in children; machine learning; early diagnosis; abdominal pain; probability; multivariate logistic regression.

For citation: Sedashkina O.A., Kolsanov A.V. Early diagnosis of chronic kidney disease in children using machine learning algorithms. Medical doctor and information technology. 2024; 3: 72-85. doi: 10.25881/18110193_2024_3_72.

ВВЕДЕНИЕ

Разработка методов моделирования для принятия управленческих решений, направленных на повышение эффективности ранней диагностики хронической болезни почек (ХБП), является важной научно-практической задачей, в решении которой большую поддержку может оказать использование методов машинного обучения [1].

Нефрологи всего мира сосредоточены на поиске решения, позволяющего отсрочить результат и уменьшить количество пациентов с почечной недостаточностью путем применения протоколов профилактики ХБП и мониторинга пациентов. По мнению Schena F.P. et al. (2022), врачи общей практики и нефрологи привлекаются на первой и на поздних стадиях заболевания, соответственно, поэтому ранняя диагностика ХБП является важным шагом в предотвращении прогрессирования поражения почек [2]. Авторы провели анализ публикаций об алгоритмах машинного обучения (MLA), которые могут предсказать ХБП. Выполнен обзор 55 статей о применении MLA при ХБП, в результате которого получены данные о том, что MLA используют огромное количество предикторов, комбинируя их нелинейными и интерактивными способами. Sanmarchi F. et al. (2023) стремились оценить, как искусственный интеллект (ИИ), включая MLA, используются для прогнозирования, диагностики и лечения ХБП [3].

Инструменты поддержки принятия решений на основе MLA были разработаны для многих аспектов лечения ХБП. Chen F. et al. (2023) утверждают о том, что алгоритмы, разработанные для ранней диагностики ХБП, могут помочь в планировании ранней нефропротективной терапии и отсрочить наступление заместительной почечной терапии, тем самым предлагая значимые клинические и экономические выгоды для пациентов и систем здравоохранения [4].

Замечены некоторые ограничения в интерпретации публикаций: материал для анализа был извлечен из опубликованных статей; базы данных, созданные на разных выборках, были объединены для последующего анализа; в базах данных были собраны переменные, которые не имели значения для диагностики и прогрессирования ХБП; не было представлено автором собственной информации о пациентах с заболеваниями почек, как и не собрана собственная база исходных данных; валидация результатов

на тестовой выборке не проводилась. В целом, сотрудничество между экспертами в области электроники, информатики и врачами было ограниченным [2, 4].

Несмотря на растущие перспективы MLA, остаются не решенными вопросы, препятствующие его клиническому внедрению. При проведении сравнения между инструментами ранней диагностики ХБП, основанными на MLA, в частности на многофакторной логистической регрессии (MLR) и других алгоритмах, была обнаружена схожая эффективность между моделями, разработанными с помощью MLA и традиционных методов логистической регрессии (LR) [4–7]. Кроме того, по мнению авторов, предыдущие исследования поставили под сомнение качество баз данных и методологию моделей прогнозирования и диагностики ХБП [8, 9].

Все эти обобщения послужили основанием для проведения настоящего исследования, в котором мы сосредоточились на создании диагностической модели ХБП с помощью MLA и постарались избежать недостатков, которые были выявлены в более ранних исследованиях.

Цель: повышение точности диагностики ХБП с использованием данных анамнеза, клинико-инструментального и генетического обследования и MLA.

МАТЕРИАЛЫ И МЕТОДЫ

Дизайн исследования: исходные данные были получены из одноцентрового ретроспективного когортного исследования (2011–2022 гг.) детей в возрасте 1–17 лет с ХБП 1–4 стадии. Для построения диагностической модели ХБП использованы данные анамнеза, клинико-инструментального и генетического обследования. В качестве целевой переменной явилось наличие или отсутствие ХБП. Статья написана в шаблоне TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis – Прозрачный отчет о многомерной модели прогнозирования для индивидуального прогноза или диагноза) [11].

УЧАСТНИКИ ИССЛЕДОВАНИЯ

Критерии отбора. В когортное исследование включены 158 детей. Основную группу составили 128 детей с неиммунной (разные формы пиелонефрита, тубулоинтерстициальный

нефрит, в том числе в исходе гемолитико-уремического синдрома, аномалии органов мочевой системы) и иммунной (первичный нефротический синдром, IgA-нефропатия, системная красная волчанка с волчаночным нефритом) патологией почек, группу контроля – 30 детей без патологии почек в возрасте от 1 до 17 лет. Сравнимые группы были репрезентативны.

Критерии не включения. Дети с поликистозной дисплазией почек, тубулопатией и рахитоподобными заболеваниями, капилляротоксическим нефритом, болезнью Альпорта, болезнью тонких мембран, постстрептококковым гломерулонефритом.

Критерии исключения. Дети, достигшие 5 стадии ХБП.

Условия проведения и продолжительность исследования. Выборку исследования формировали из числа детей, госпитализированных в отделение детской нефрологии Самарской областной клинической больницы имени В.Д. Середякина (г. Самара) с 2011 г. по 2022 г.

Процесс диагностики ХБП у детей реализован при использовании диагностической модели, которая была построена на обучающей выборке, а затем протестирована на тестовой выборке, составившей 20%. Обучающая выборка (N = 127) оказалась несбалансированной, т.е. пациентов с ХБП было больше (103 (81%)), чем пациентов без ХБП (24 (19%)). Для балансировки базы данных в части обучающей выборки в целях получения более стабильной модели был использован метод SMOTE (Synthetic Minority Over-sampling Technique). Получена выборка со 184 записями, в которой было 92 (50%) с ХБП и 92 (50%) – без ХБП.

Предикторы. Для разработки модели ранней диагностики ХБП использовали универсальные факторы риска (перенесённые критические состояния в перинатальном периоде с развитием острого повреждения почек (ОПП); пренатально установленный врожденный порок развития (ВПР) органов мочевой системы (ОМС); наличие генетической предрасположенности к заболеваниям почек, комплекса аллергических заболеваний в семье; носительство персистирующих инфекций, гемолитических штаммов стрептококка; социальная детерминанта; экологические условия) и специфические ранние факторы риска (ранние факторы риска прогрессирования

заболевания: особенности дебюта патологии почек и биомаркеры прогрессирования, аллергические заболевания, нарушения уродинамики при ВПР ОМС; показатели рутинного лабораторного исследования крови; показатели специального генетического анализа иммунных и неиммунных полиморфных маркеров генов ренин-ангиотензиновой системы, синтазы окиси азота-3, эндотелина-1, интерлейкина-4 и фактора некроза опухоли; результаты эхографического и доплерографического исследований).

Исходы/методы регистрации исходов. Результатом работы модели является раннее выявление вероятности прогрессирования патологии почек. К числу факторов прогрессирования нефропатии относили снижение скорости клубочковой фильтрации (СКФ) [12]. Диагноз ХБП выставлялся в соответствии с клиническими рекомендациями для ХБП Национального фонда почки США (2012) и Российскими национальными рекомендациями по ХБП (2021).

Индексный тест. В качестве индексного теста предлагается использовать формулу MLR, разработанную на основе MLA.

Обоснование выбора референсного теста. В качестве референсного теста выбрали определение расчетной СКФ по креатинину в рамках диагностических мероприятий. Чувствительность данного теста низкая и свидетельствует о далеко зашедшем хроническом заболевании почек. Других аналогов, которые максимально точно диагностируют ХБП, в настоящее время нет.

Статистическая обработка. Первичное формирование базы данных осуществлялось в электронных таблицах Excel (Microsoft Office 365). Количественные показатели оценивались на соответствие нормальному распределению, использовался критерий Шапиро-Уилка. Поскольку распределение большинства признаков не соответствовало нормальному, как центральная тенденция использовалась медиана, как показатели размаха – квартили (Me [Q1; Q3]). Для проверки гипотез относительно двух несвязанных выборок использовался U-критерий Манна-Уитни. Сравнение номинальных данных в группах проводилось при помощи критерия χ^2 Пирсона и точного критерия Фишера. В качестве количественной меры эффекта качественных признаков использовался показатель отношения

шансов (ОШ) с 95% доверительным интервалом. Для количественных признаков применялась логистическая регрессия, поскольку независимые переменные могли быть и категориальными, и количественными. Для определения связи между признаками использован коэффициент ранговой корреляции Спирмена (r) с 95% доверительным интервалом. Для оценки диагностической эффективности, точки разделения и операционных характеристик признаков применен ROC-анализ. Различия считались статистически значимыми при вероятности ошибки второго рода $p < 0,05$.

Обработка результатов проводилась средствами языка Питон (Python 3.11.). Построение диагностической модели было выполнено по следующей схеме. На первом этапе исходная выборка была разделена на обучающую и тестовую выборки в соотношении 80:20.

На втором этапе обучающая выборка была проверена на сбалансированность в зависимости от количества пациентов с исходом: 1 – есть ХБП, 0 – нет ХБП. Для балансировки базы данных в части обучающей выборки в целях получения более стабильной модели был использован метод SMOTE (Synthetic Minority Over-sampling Technique) [13, 14].

На третьем этапе был применен корреляционный анализ, при котором был посчитан коэффициент ранговой корреляции Спирмена (r). Если пара переменных сильно ($r > 0,7$) коррелировала между собой, то для дальнейшего анализа оставлялась одна переменная (наиболее значимая), вторая переменная из дальнейшего анализа исключалась.

На четвертом этапе для отбора наиболее значимых переменных был применен метод RFE (Recursive Feature Elimination – рекурсивное исключение признаков) [15].

Для построения модели предсказания наличия/отсутствия исхода был применен MLA, позволяющий построить легко интерпретируемую модель: MLR.

Построенная модель была протестирована с использованием независимого тестового набора. Для оценки качества построенной модели были использованы следующие метрики качества на обучающей и тестовой выборках: точность, чувствительность, специфичность и ROC-AUC (Receiver Operating Characteristic

– рабочая характеристика приёмника; AUC-Area Under the Curve – площадь под кривой) [16]. Метрики были рассчитаны с 95% доверительными интервалами (ДИ). 95% доверительный интервал рассчитывался методом Бутстреп (Bootstrap) при выборке 1000 экземпляров. Также для максимизации чувствительности и специфичности было рассчитано пороговое значение для определения наличия/отсутствия изучаемого исхода.

Для оценки клинической полезности модели был проведен анализ кривой принятия решений (DCA - Decision curve analysis) [17] путем расчета чистых преимуществ модели по сравнению с решением «лечить всех» и «не лечить никого».

Этическая экспертиза. Протокол исследования согласован с этическим комитетом 12.04.2023 (выписка из протокола заседания Комитета по биоэтике при СамГМУ №263). Включение пациентов в исследование осуществляли после подписания законным представителем пациента или детьми старше 14 лет информированного добровольного согласия на проведение обследования ребенка и обработку персональных данных.

РЕЗУЛЬТАТЫ

Участники исследования.

В исследование включены 158 детей. Основную группу составили 128 детей в возрасте от 1 до 17 лет, 88 девочек (68,8%) и 40 мальчиков (31,2%) с установленным нозологическим диагнозом поражения почек на различных стадиях ХБП. Группу контроля составили 30 детей в возрасте от 1 до 17 лет, 18 девочек (60,0%) и 12 мальчиков (40,0%) без патологии почек. Медиана возраста в основной группе составила 6 [5,0; 8,0] лет, медиана возраста в контрольной группе составила 5 [5,0; 6,75] лет. В выделенных группах дети статистически значимо не различались по полу ($p = 0,931$) и возрасту ($p = 0,148$).

Результаты диагностического исследования:

При отборе признаков использовали метрику качества – площадь под ROC-кривой – AUC, чем выше была AUC при добавлении признака в модель, тем больше оценивали значимость.

У детей с ХБП наблюдался мочевого синдром (гематурия) в отличие от детей группы контроля, у которых такого синдрома не наблюдали (2,0 [0,0; 4,0] против 0, $p < 0,001$). Далее мы получили данные о том, может ли полученный признак иметь самостоятельное значение в диагностике

ХБП. Наличие эритроцитов в моче обеспечивает диагностику ХБП без учета взаимодействия с другими симптомами с достаточно высоким качеством (ROC-AUC = 0,852), при умеренной точности 65,2% [58,9; 71,5], низкой чувствительности 57,0% [50,0; 64,5] и высокой специфичности признака – 100,0% [100,0; 100,0] в точке отсечения 1 эритроцит в поле зрения. Это дает право полагать, что при обнаружении более 1 эритроцита в поле зрения в моче в течение определенного времени, согласно определению ХБП, признак «гематурия» предсказывает формирование хронического

процесса в почках. При отсутствии или при наличии гематурии менее 1 эритроцита в поле зрения можно предположить о низкой вероятности развития хронического процесса в почках.

На рисунке 1 показано, что группа контроля (ХБП-) и основная группа (ХБП+) статистически значимо различаются по переменной «Эритроциты_моча>1» ($p<0,001$). В группе контроля получены данные об отсутствии или наличии 1 эритроцита в моче в поле зрения (эритроцитов в моче ≤ 1) у всех 30 детей (100,0%). В основной группе отсутствие или наличие 1 эритроцита

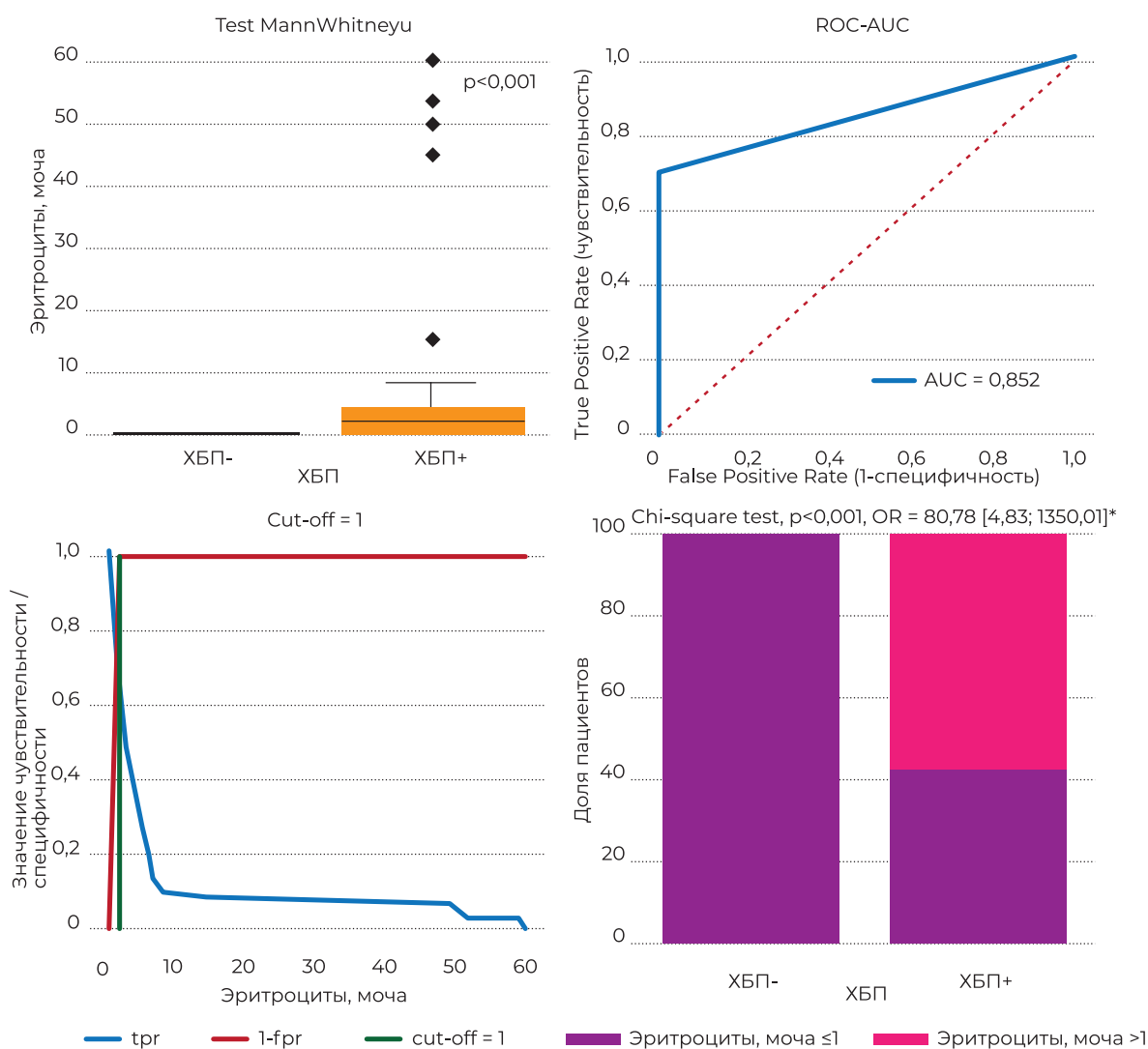


Рисунок 1 — Метрики оценки диагностической значимости признака «гематурия» (эритроциты в моче). «ХБП-» — группа контроля, «ХБП+» — основная группа, «эритроциты_моча» — гематурия.

в поле зрения отмечено у 55 (42,97%), а более 1 эритроцита в поле зрения – у 73 (57,03%) детей. Отношение шансов = 80,78 [4,83; 1350,01], $p < 0,001$. Иными словами, вероятность встретить пациентов с «Эритроциты_моча>1» среди пациентов с ХБП в 80,78 [4,83; 1350,01] раз выше, чем в группе пациентов без ХБП.

Следующий признак – астеническое телосложение, который замечен у 76,6% детей с ХБП по сравнению 10,0% детей группы сравнения ($p < 0,001$; ОШ = 29,4). Однофакторный регрессионный анализ показал, что астеническое телосложение ($B = 3,381$) является значимым фактором риска ХБП.

При оценке вероятности самостоятельного значения признака в диагностике ХБП получили результат, представленный на рисунке 2.

Если использовать показатель «астеническое телосложение» как самостоятельный (единственный) предиктор ХБП, то он обеспечивает диагностику ХБП с достаточно высоким качеством ($ROC-AUC = 0,833$) при умеренной точности 79,1% [73,4; 84,2], средней чувствительности 76,6% [70,1; 82,8] и высокой специфичности признака – 90,0% [80,0; 97,1]. Это дает право полагать, что при обнаружении астенического телосложения высока вероятность наличия ХБП.

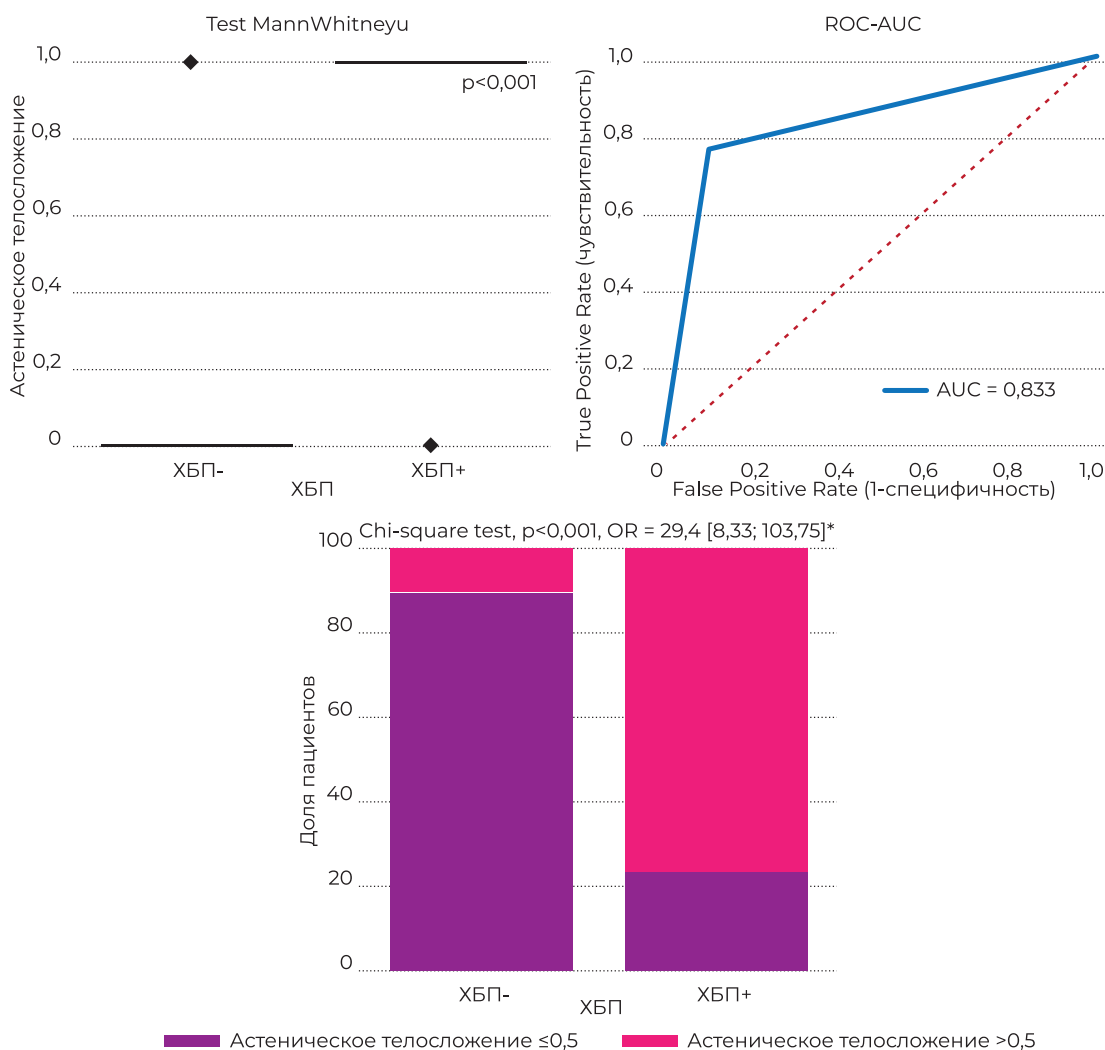


Рисунок 2 — Метрики оценки диагностической значимости признака «астеническое телосложение».

А при его отсутствии можно предположить низкую вероятность развития хронического процесса в почках (рисунок 2).

Третьим признаком, вошедшим в диагностическую модель ХБП у детей, является ускорение СОЭ (6,0 [3,0; 9,0] против 4,0 [2,0; 6,0], $p = 0,002$). Однофакторный регрессионный анализ показал признак «ускорение СОЭ» ($B = 0,219$), как значимый фактор риска ХБП.

Далее мы получили данные о том, может ли полученный признак иметь самостоятельное значение в диагностике ХБП. Скорость оседания эритроцитов показала хорошую

диагностическую значимость, площадь под ROC-кривой составила 0,686 при низкой точности (57%), низкой чувствительности (53,9%) и большей специфичности (70%) в точке отсечения 5 мм/час (рисунок 3).

В основной группе обнаружена $COЭ \leq 5$ мм/ч у 59 (46,09%) детей, а $COЭ > 5$ мм/ч у 69 (53,91%) детей. В группе контроля уровень $COЭ \leq 5$ мм/ч обнаружен у 21 (70,0%) детей, а уровень $COЭ > 5$ мм/час у 9 (30,0%) детей. Отношение шансов = 2,73 [1,16; 6,42], $p = 0,018$.

Можно сделать вывод, что гематурия, астеническое телосложение и ускорение СОЭ являются

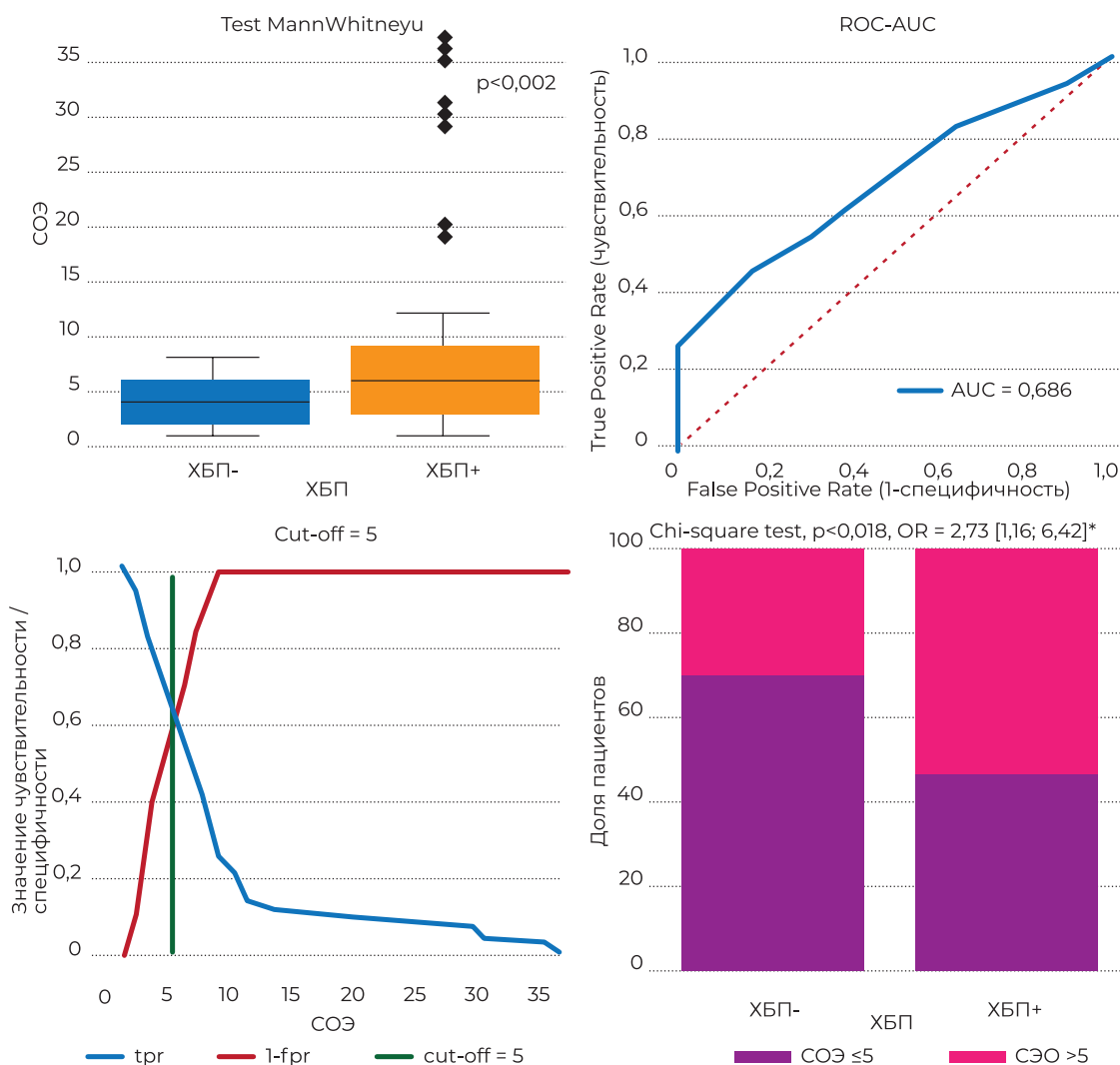


Рисунок 3 — Метрики оценки диагностической значимости признака «ускорение СОЭ». «ХБП-» — группа контроля, «ХБП+» — основная группа, «СОЭ» — скорость оседания эритроцитов.

Таблица 1 — Метрики оценки качества диагностической силы переменных как самостоятельный (единственный) предиктор ХБП

Мера оценки	Гематурия	Астеническое телосложение	Ускорения СОЭ
ROC-AUC	85,2% [81,8; 88,4]	83,3% [77,3; 88,6]	68,6% [60,8; 76,0]
Точность	65,2% [58,9; 71,5]	79,1% [73,4; 84,2]	57,0% [50,6; 63,3]
Чувствительность (1)	57,0% [50,0; 64,5]	76,6% [70,1; 82,8]	53,9% [46,9; 61,2]
Специфичность (0)	100,0% [100,0; 100,0]	90,0% [80,0; 97,1]	70,0% [55,9; 83,3]

Таблица 2 — Сравнение обучающей и тестовой выборок

Предикторы	Все пациенты (n = 158)	Тестовая выборка (n = 31)	Обучающая выборка (n = 127)	p
ХБП				
0 – нет	30 (19,0%)	6 (19,0%)	24 (18,9%)	p = 1,000
1 – да	128 (81,0%)	25 (81,0%)	103 (81,1%)	
Пол				
0 – мужской	106 (67,1%)	21 (68,0%)	85 (66,9%)	p = 0,931
1 – женский	52 (32,9%)	10 (32,0%)	42 (33,1%)	
Возраст, лет	6,0 [5,0; 8,0]	7,0 [5,0; 8,0]	6,0 [4,0; 8,0]	p = 0,195
Астеническое телосложение у ребенка				
0 – нет	57 (36,1%)	11 (35,0%)	46 (36,2%)	p = 0,939
1 – да	101 (63,9%)	20 (65,0%)	81 (63,8%)	
СОЭ	5,0 [3,0; 8,0]	4,0 [3,0; 6,0]	5,0 [3,0; 7,0]	p = 0,293
Гематурия (потеря эритроцитов с мочой)	4,52 [4,19; 4,75]	4,55 [4,28; 4,86]	4,52 [4,19; 4,75]	p = 0,237

важными факторами диагностики ХБП. Однако, если посмотреть на предсказательную силу этих переменных, видно, что чувствительность составляет 53,9–76,6%, специфичность 70–100%. При этом точность ни одного из факторов не превышает 80% (Таблица 1).

Для повышения точности диагностики ХБП было принято решение о построении модели для практического применения на новых пациентах, которые будут обращаться за медицинской помощью в будущем.

Разработка диагностической модели.

Модель ранней диагностики ХБП была построена на обучающей выборке, а затем модель была протестирована на тестовой выборке, которая составила 20%. Итого общая выборка в 158 пациентов была случайным образом разделена на обучающую выборку в 127 (80%) пациентов и тестовую выборку в 31 (20%) пациент. Обучающая

и тестовая выборка статистически значимо не различались по исходу (наличие ХБП) и исследуемым предикторам, что указывает на рациональное разделение набора данных (Таблица 2).

Обучающая выборка (N = 127) оказалась несбалансированной, т.е. пациентов с ХБП было больше 103 (81%), чем пациентов без ХБП – 24 (19%). Для балансировки базы данных в части обучающей выборки в целях получения более стабильной модели ранней диагностики, был использован метод SMOTE. В результате была получена выборка с 206 записями, в которой было 103 (50%) пациента с ХБП и 103 (50%) – без ХБП.

Для построения модели MLR использованы три переменных (Таблица 3): астеническое телосложение у ребенка, СОЭ, потеря эритроцитов с мочой. Модель имеет следующий вид:

Вероятность диагностики ХБП = $1/(1 + e^{-(x)})$, где $x = +10,649 + 5,792 \cdot (\text{астеническое телосложение}) + 0,392 \cdot (\text{СОЭ}) + 1,225 \cdot (\text{эритроциты в моче})$.

Таблица 3 — Характеристика переменных модели MLR

Предикторы	coef (B)	Exp (B)	p
Астеническое телосложение у ребенка	5,792	327,67 [59,24; 1811,76]	p<0,001*
СОЭ	0,392	1,48 [1,188; 1,84]	p<0,001*
Потеря эритроцитов с мочой	1,225	0,04 [0,01; 0,23]	p<0,001*
constanta	10,649		p = 0,004*

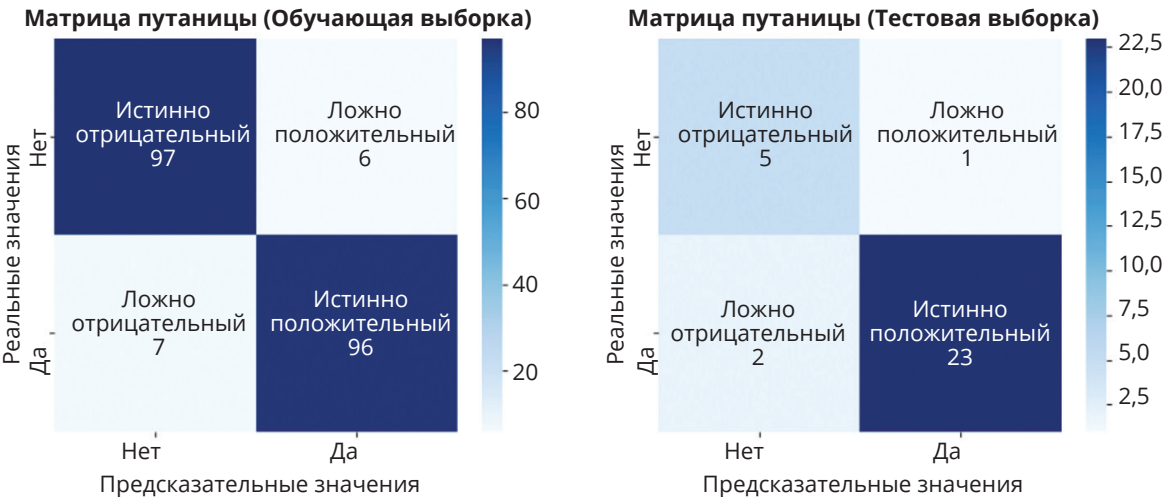


Рисунок 4 — Матрица путаницы для обучающей и тестовой выборок.

Таблица 4 — Метрики оценки качества модели ранней диагностики ХБП у детей

Мера оценки	MLR	
	Обучение	Тестирование
ROC-AUC	95,9% [93,2; 98,1]	90,0% [77,2; 100,0]
Точность	93,7% [90,8; 96,1]	90,3% [80,6; 96,8]
Чувствительность	93,2% [88,7; 97,1]	92,0% [81,5; 100,0]
Специфичность	94,2% [90,4; 98,0]	83,3% [50,0; 100,0]
R ²	74,8% [63,1; 84,5]	

Для оценки информативности построенной модели были рассчитаны матрицы путаницы отдельно на обучающей и тестовой выборках (Рисунок 4). На основе матриц путаницы были рассчитаны метрики качества построенной модели (Таблица 4). На обучающей выборке из 206 записей модель MLR диагностирует с точностью 93,7% [90,8; 96,1], чувствительностью 93,2% [88,7; 97,1], специфичностью 94,2% [90,4; 98,0], ROC-AUC = 95,9% [93,2; 98,1], описывает 74,8% [63,1; 84,5] дисперсии (Рисунок 5).

На тестовой выборке из 31 пациента модель показала высокую точность 90,3% [80,6; 96,8], высокую чувствительность 92,0% [81,5; 100,0], и чуть ниже специфичность 83,3% [50,0; 100,0], ROC-AUC = 90,0% [77,2; 100,0]. Полученная модель отличного качества (>90%), т.к. площадь под ROC-кривой составляет на тестовой выборке 0,90. Значение точки отсечения (cut-off) вероятности ХБП равно 0,25 (Рисунок 5). Это означает, что при получении значения<0,25 вероятность развития ХБП отсутствует, а при значении≥0,25 можно предположить высокий риск развития ХБП.

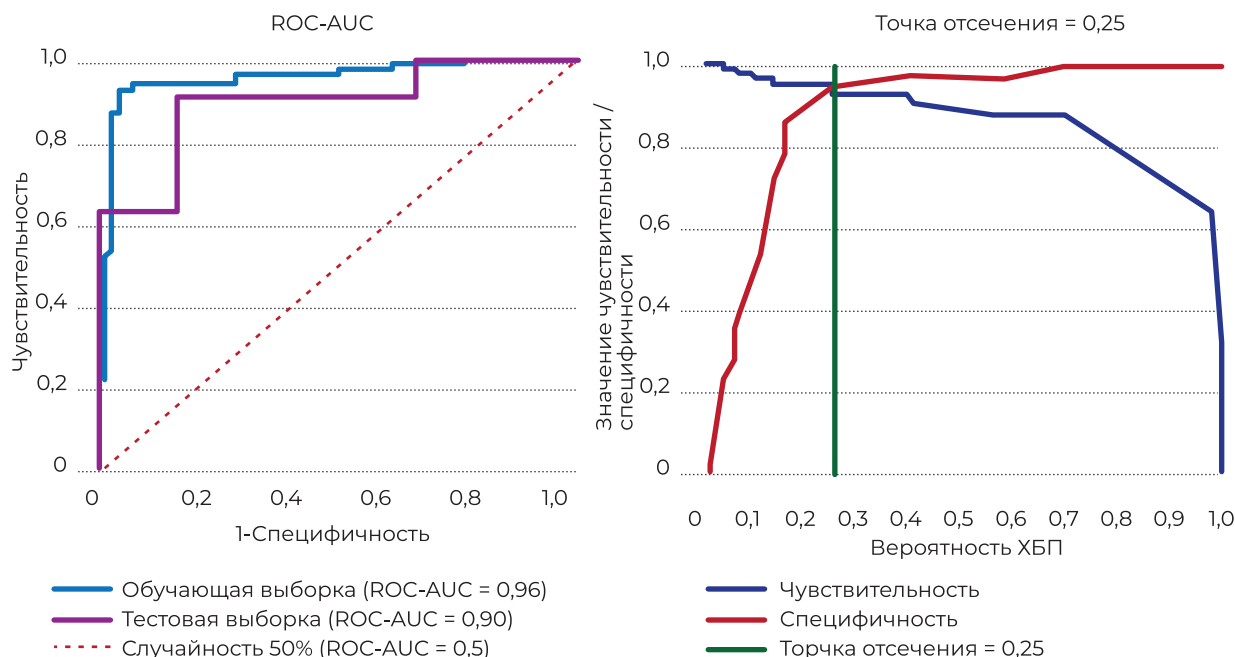


Рисунок 5 — ROC-кривые и Cut-off (точка отсечения) для диагностической модели в обучающей и тестовой выборках.

Полученные значения качества модели на двух выборках (обучающая выборка: ROC-AUC = 95,9% [93,2; 98,1], $p < 0,001$ и тестовая выборка: ROC-AUC = 90,0% [77,2; 100,0], $p < 0,001$), соответствуют отличной дискриминационной способности модели, а отсутствие значимых различий параметра ROC-AUC в когортах обучения и тестирования свидетельствует об ее устойчивости при проверке.

Чтобы оценить клиническую полезность модели ранней диагностики ХБП у детей, был использован анализ кривой принятия решения. Анализ проводился путем расчета чистых преимуществ модели предсказания ХБП по сравнению с порогами риска в комбинированных наборах обучения и тестирования. Анализ кривой решения определил, что диапазон пороговых вероятностей составляет от 0 до 0,96, в которых модель имеет большую ценность, чем «лечить всех» или «не лечить никого» (Рисунок 6).

ОБСУЖДЕНИЕ

Резюме основного результата исследования. Построена модель, предсказывающая на обучающей выборке с высокой точностью, чувствительностью и специфичностью ХБП у детей.

Полученная регрессионная модель отличного качества (>90%).

Преимущество нашего исследования. Переменные были получены из собственного ретроспективного когортного исследования, база данных формировалась в режиме реального времени, использованы количественные и качественные переменные клинико-генетического и инструментального обследования пациентов, при построении модели ранней диагностики ХБП применен метод MLR (RFE – RandomForest), проведена валидация результатов. Поиск показал отсутствие аналогов диагностических моделей ХБП у детей.

Обсуждение основного результата исследования. Основным вкладом нашего исследования было установление клинических признаков или симптомов, которые могут быть полезны для выявления детей с повышенным риском развития ХБП. Факторы, которые были выбраны в качестве предикторов диагноза ХБП: астеническое телосложение у ребенка, потеря эритроцитов с мочой, СОЭ.

Маркеры, выявленные в нашем исследовании, не являются новыми открытиями, но они совместно могут быть актуальны и служить показателями

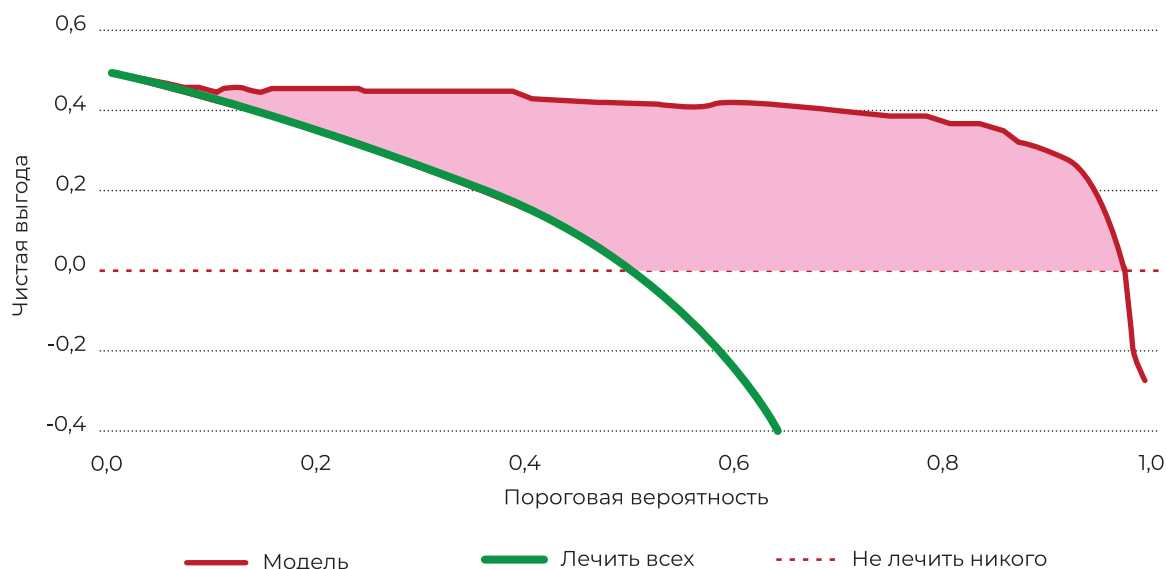


Рисунок 6 — Кривая принятия решений, отображающая клиническую выгоду от предпочтений врача.

ХБП. Наша модель, основанная на этих маркерах, может помочь выявлять ХБП у детей на ранней стадии, что указывает на необходимость более комплексных обследований пациента.

Доказано, что ХБП может развиваться на фоне заболевания почек, протекающего с синдромом гематурии, но на практике гематурию не рассматривают как фактор прогрессирования ХБП, а маломанифестные формы заболевания, протекающие с гематурией, часто не находят должной настороженности врача.

Дополнительным заметным признаком, наблюдаемым в нашей модели, было замедление роста (астеническое телосложение). Известно, что задержке роста у детей с ХБП способствует множество различных причин [12]. Наши результаты показывают, что этот симптом может возникнуть рано, и его следует рассматривать как тревожный признак для диагностики заболевания.

Связь ускорения СОЭ и ХБП неоднозначна. СОЭ является традиционным маркером воспаления, который используется для оценки активности воспаления при различных заболеваниях иммунного [13] и инфекционного генеза, однако до сих пор недостаточно знаний о значимости этого системного воспалительного маркера при диспротеинемии, способствующей увеличению агрегации эритроцитов,

при повышенном уровне холестерина, анемии, алкалозе, азотемии, сдвиге pH крови и наличии других нарушений гомеостаза, которые развиваются при хронической патологии почек.

Предложенный нами способ ранней диагностики ХБП учитывает линейное влияние предикторов на исход. Диагностическая модель, созданная в ходе текущего исследования, показала очень хорошую точность. Эти результаты позволяют предположить, что модель может быть полезным инструментом для более раннего выявления случаев ХБП в общеклинических педиатрических условиях, которые в дальнейшем должны быть подтверждены при детальном изучении врачом нефрологом.

Существуют разные способы определения внутренней валидации модели. Наиболее известными методами, используемыми для оценки качества модели, являются бутстрап, перекрестная проверка (cross validation) и проверка на тестовой выборке. В этом исследовании мы использовали бутстрап, который, как утверждают Steyerberg et al., является наиболее эффективным методом проверки качества построенной модели [14].

Построенная нами модель для раннего выявления ХБП у детей, не является первой. Так

Nogueira et al. в 2019 году построили модель ранней диагностики ХБП детей из Бразилии ($N = 752$), основанной на следующих факторах риска: рецидивирующая инфекция мочевыводящих путей, сглаживание кривой роста, полиурия, никтурия, аномальная струя мочи, гипертония в анамнезе, пенистая моча, отеки и госпитализация по причине мочеполовых инфекций в неонатальном периоде [15]. Полученная ими модель имела достаточно высокий ROC-AUC = 0,927 [0,910; 0,946]. Однако этот инструмент в настоящее время не валидирован на российской популяции.

На сегодняшний день наилучшей практикой в отечественной детской нефрологии является диагностика болезни почек, развившейся на фоне ВПР ОМС, поскольку с 2001 года внедрен и отлично работает мониторинг ВПР в антенатальном и раннем постнатальном периодах. Между тем, по нашим данным, выявляются ВПР ОМС у 31,6% детей в антенатальном и у 12,5% постнатальном периодах. У остальной доли пациентов ВПР ОМС выявлены поздно: у 16,6% детей – случайно, у 39,1% детей – при обращении в детскую поликлинику по поводу болей в животе или присоединения мочевого синдрома. Итого, чувствительность существующих методов ранней диагностики ХБП у детей составляет 44,1%.

Поэтому пришло время улучшить диагностику заболевания с помощью новых технологий. ИИ и машинное обучение имеют значительный потенциал в современном здравоохранении. Модели, разработанные для прогнозирования и диагностики ХБП, потенциально способны обеспечить раннее распознавание заболевания, что может привести к своевременной диагностике и профилактике заболеваний почек, помочь в раннем планировании ухода и выделить ресурсы для наиболее значимых клинических и экономических мероприятий.

Поскольку количество алгоритмов растет, основное внимание должно быть направлено на устранение препятствий на пути клинической реализации. Целью написания статьи явилась оценка метода разработки модели и сравнение производительности MLA с традиционными диагностическими методами. Это будет информировать врачей и технических специалистов о текущем развитии MLA в прогнозировании и

ранней диагностике ХБП, а также будет способствовать разработке и стандартизации будущих моделей. Алгоритмы ИИ могут быть полезным инструментом, помогающим принимать клинические решения и более эффективно распределять напряженные ресурсы.

Ограничения исследования. Основным ограничением является одноцентровой характер исследования. Модель ранней диагностики ХБП у детей была проверена только на основе внутренних данных, а проверка на внешних данных не проводилась. Для его проверки необходимы дальнейшие многоцентровые исследования с более крупными выборками.

ЗАКЛЮЧЕНИЕ

Получена диагностическая модель отличного качества (>90%). Интеллектуальная система поддержки врачебных решений является мощным инструментом для оптимизации ранней диагностики ХБП на основе предложенной модели и заключается в использовании ее в клинической практике. Внедрение системы поддержки принятия врачебных решений в первичное педиатрическое звено здравоохранения значительно повысит эффективность и качество раннего выявления пациентов с ХБП с целью их отбора в группы для направления к нефрологу с последующим назначением раннего нефропротективного лечения. Чем раньше пациенту назначена индивидуальная схема нефропротективной терапии, тем лучшие показатели выживаемости и функции почек демонстрируются.

Машинное обучение является многообещающим инструментом для раннего выявления ХБП у детей. Тем не менее, необходима дальнейшая работа над интерпретируемостью, обобщаемостью и справедливостью моделей, чтобы обеспечить безопасное применение таких технологий в повседневной клинической практике.

Значение для практики. Построенная модель помогает диагностировать ХБП на уровне первичного звена здравоохранения. Ранняя диагностика и начало нефропротективного лечения могут привести к лучшим результатам лечения и потенциально продлить жизнь.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов (The authors declare no conflict of interest).

ЛИТЕРАТУРА/REFERENCES

1. Бодрин К.А., Красноперова А.А. Использование технологий машинного обучения в медицине // Теория и практика современной науки. – 2018. – №10(40). – С.52-56. [Bodrin KA, Krasnoperova AA. Machine learning technologies in medicine. Theory and practice of modern science. 2018; 10(40): 52-56. (In Russ.)]
2. Schena FP, Magistroni R, Narducci F, et al. Artificial intelligence in glomerular diseases. *Pediatr Nephrol.* 2022; 37(11): 2533-2545. doi: 10.1007/s00467-021-05419-8.
3. Sanmarchi F, Fanconi C, Golinelli D, et al. Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review. *J Nephrol.* 2023; 36(4): 1101-1117. doi: 10.1007/s40620-023-01573-4.
4. Chen F, Kantagowit P, Nopsopon T, et al. Prediction and diagnosis of chronic kidney disease development and progression using machine-learning: Protocol for a systematic review and meta-analysis of reporting standards and model performance. *PLoS One.* 2023; 18(2): e0278729. doi: 10.1371/journal.pone.0278729.
5. Scikit-learn. Decision tree, regressor. [cited 2023 Oct 10]. Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html#sklearn.tree.DecisionTreeRegressor>.
6. Scikit-learn. Common errors in the interpretation of linear model coefficients. [cited 2023 Oct 10]. Available from: https://scikit-learn.org/stable/auto_examples/inspection/plot_linear_model_coefficient_interpretation.html#sphx-glr-auto-examples-inspection-plot-linear-model-coefficient-interpretation-py.
7. Scikit-learn. Robust scaling. [cited 2023 Oct 10]. Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>.
8. Scikit-learn. Lasso regression. [cited 2023 Oct 10]. Available at: scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html. Scikit-learn. Cross-validation. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_validate.html.
9. Носова Г.С., Абдуллин А.Х. Машинное обучение на основе непараметрического и нелинейного алгоритма Random Forest (RF) // Инновации. Наука. Образование. – 2021. – №35. – С.33-39. [Nosova GS, Abdullin AH. Machine learning based on nonparametric and nonlinear Random Forest (RF) algorithm. *Innovation. The science. Education.* 2021; 35: 33-39. (In Russ.)]
10. Scikit-learn. Random forest, regressor. [cited 2023 Oct 10] Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>.
11. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2020; 395(10225): 709-733. doi: 10.1016/S0140-6736(20)30045-3.
12. Rees L, Mak RH. Nutrition and growth in children with chronic kidney disease. *Nat Rev Nephrol.* 2011; 7(11): 615-23. doi: 10.1038/nrneph.2011.137.
13. Dima A, Opris D, Jurcut C, et al. Is there still a place for erythrocyte sedimentation rate and C-reactive protein in systemic lupus erythematosus? *Lupus.* 2016; 25(11): 1173-1179. doi: 10.1177/0961203316651742.
14. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001; 54(8): 774-81.
15. Nogueira PCK, Konstantyner T, Carvalho MFC, et al. Development of a risk score for earlier diagnosis of chronic kidney disease in children. *PLoS One.* 2019; 14(4): e0215100. doi: 10.1371/journal.pone.0215100.