

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

УДК 519.25:616.127

О СИСТЕМЕ ПРЕДИКТИВНЫХ КРИТЕРИЕВ НАРУШЕНИЯ СЕРДЕЧНОГО РИТМА У НОВОРОЖДЕННЫХ, ОСНОВАННОЙ НА АНАЛИЗЕ ДАННЫХ

Акишев Д.В., Тягушева Е.Н., Сыромясов А.О., Власова Т.И.,
Бадюкина Т.Е.

Национальный исследовательский Мордовский государственный университет,
г. Саранск

Поступила в редакцию 24.03.2025, после переработки 25.05.2025.

В статье рассматривается разработка алгоритмов машинного обучения для выявления нарушений сердечного ритма новорожденных. Исследованию подвергаются основные интервалы кардиографии. Источником данных служат истории болезней, которые содержат в себе показатели новорожденных детей в первый, третий и десятый дни жизни. Основное внимание акцентируется на статистическом анализе данных и построении регрессионных моделей, а также на сравнении результатов их работы. Применение разработанных моделей и расчет основных метрик показывает, что регрессионные алгоритмы в целом недостаточно хорошо объясняют изменения зависимой переменной, при этом модель MARS обладает большей предсказательной ценностью для сравнительно малых объемов выборки. Для улучшения диагностики и профилактики нарушений сердечного ритма требуются дальнейшая оптимизация, увеличение обучающей выборки и повышение разнообразия данных.

Ключевые слова: искусственный интеллект в педиатрии, прогнозирование нарушений сердечного ритма, машинное обучение, регрессионное моделирование, новорожденные.

Вестник ТвГУ. Серия: Прикладная математика. 2025. № 3. С. 78–92.
<https://doi.org/10.26456/vtppmk736>

Введение

Технологии искусственного интеллекта (ИИ), нейронные сети и машинное обучение в последнее время демонстрируют динамичное развитие, а круг их приложений постоянно расширяется; в частности, они часто применяются в медицине [1], где внедрение ИИ способствует более своевременному выявлению патологических процессов. Особое внимание использованию такого рода систем уделяется в педиатрической практике: повышенная уязвимость пациентов требует ускорения принятия врачебных решений и минимизации диагностических ошибок [2, 3].

© Акишев Д.В., Тягушева Е.Н., Сыромясов А.О., Власова Т.И., Бадюкина Т.Е., 2025

В последние годы существенные изменения, обусловленные активным развитием систем ИИ, произошли в исследованиях заболеваний сердечно-сосудистой системы. Большая часть работ в этой области сосредоточена на улучшении диагностических и прогностических возможностей с использованием набора данных, электрокардиограмм (ЭКГ), эхокардиограмм и других методов обследования [4]. Однако диагностика нарушений ритма сердца (НРС) у новорожденных и детей первого года жизни представляет особую сложность из-за отсутствия выраженной клинической картины и невозможности пациентов этого возраста описать свои жалобы. Это создает широкий пласт диагностических проблем для неонатологов и педиатров, а также требует разработки специализированных инструментов, включая обучение нейронных сетей.

Исследования, проведенные ранее «медицинской» частью коллектива авторов работы, позволили установить прогностическое значение динамического изменения ряда биохимических и электрокардиографических параметров в формировании НРС. Соответственно, главной целью настоящей статьи является разработка алгоритма раннего прогнозирования НРС с использованием модели машинного обучения на основании полученных данных об изменениях биомаркеров и электрофизиологических параметров у новорожденных детей.

1. Специфика исходных данных и формализация задачи

В рамках исследования был проведен ретроспективный анализ $n = 610$ историй болезни новорожденных детей (данные по республике Мордовия за несколько лет). Вектор x_i , описывающий i -ю историю болезни ($i = 1, 2, \dots, 610$), имел вид:

$$x_i = (GP_i, W_i, Apgar_i, M_i, ECO_i, CI_i, VGK_i, C_i^{(1)}, C_i^{(3)}, C_i^{(10)}).$$

Смысл указанных компонент таков [5]:

- $GP_i \in \{22, \dots, 41\}$ — период гестации в неделях (натуральное число);
- $W_i \in \mathbb{R}_+$ — масса ребенка при рождении;
- $Apgar_i = (A1_i, A5_i)$ — оценка новорожденного по шкале Апгар на первой и пятой минутах, соответственно; $A1_i$ и $A5_i$ — целые числа от 0 до 10;
- $M_i \in \{0, 1\}$ — метод родоразрешения: 0 — естественное, 1 — кесарево сечение;
- $ECO_i \in \{1, 2\}$ — способ зачатия: 1 — ЭКО, 2 — естественное;
- $CI_i, VGK_i \in \{1, 2, 3\}$ — степени тяжести церебральной ишемии и ВЖК.

Величины $C_i^{(d)} \in \mathbb{R}^8$ суть векторы кардиографических параметров на d -й день наблюдений:

$$C_i^{(d)} = (HR, P, PQ, QRS, ST, lactate, QT, QTc),$$

причем частота сердечных сокращений (ЧСС) HR, интервал QT и его скорректированное значение QTc связаны между собой формулой Базетта [6]:

$$QTc = \frac{QT}{\sqrt{60/HR}}. \quad (1)$$

По этим данным требуется предсказать упомянутые выше интервал QT и его скорректированное значение QTc. Указанные параметры, измеряемые в первый, третий и десятый дни наблюдений, являются целевыми переменными задачи.

С точки зрения теории задача заключается в том, чтобы найти функцию $f : \mathbb{R}^m \rightarrow \mathbb{R}$, которая бы минимизировала ошибку между предсказанными и истинными значениями целевой переменной $Y \in \mathbb{R}$ (m — число параметров задачи). Если Y_i — наблюдаемые значения Y , соответствующие рядам данных X_i , то оптимальное решение f^* минимизирует среднее значение функции потерь L :

$$f^* = \arg \min_f \frac{1}{n} \sum_{i=1}^n L(f(X_i), Y_i). \quad (2)$$

Чтобы получить всестороннее представление об эффективности предсказаний, при определении (2) целесообразно использовать не одну, а несколько метрик качества. В настоящем исследовании в качестве функции потерь рассматривалась средняя абсолютная ошибка в процентах (MAPE):

$$L_{\text{MAPE}}(f(X_i), Y_i) = \left| \frac{Y_i - f(X_i)}{Y_i} \right| \cdot 100\%. \quad (3)$$

Помимо нее, ключевым показателем качества модели является коэффициент детерминации R^2 — доля общей дисперсии, объясняемая моделью:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - f(X_i))^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (4)$$

Здесь $\bar{Y}$ — среднее значение целевой переменной.

Также для количественного сравнения моделей с различной степенью сложности использовался информационный критерий Акаике (AIC), позволяющий оценить баланс между качеством аппроксимации и числом k параметров модели:

$$\text{AIC} = 2k - 2 \ln \tilde{L}; \quad (5)$$

здесь $\tilde{L}$ — максимальное значение функции правдоподобия. Форма этой функции выбирается в соответствии с предполагаемым распределением остатков (так, при нормальном распределении ошибок $\tilde{L}$ выражается через сумму их квадратов), что позволяет применять AIC в широком классе моделей. Более низкое значение AIC свидетельствует о лучшем соотношении качества и сложности модели [7].

2. Предварительный анализ данных

Перед созданием и обучением моделей был проведен комплексный анализ данных для выявления их статистических характеристик, структурных особенностей и потенциальных аномалий в них. Также были определены взаимосвязи между ключевыми переменными задачи.

2.1. Анализ распределений

На начальном этапе исследования применялась описательная статистика, включающая вычисление таких основных характеристик распределения каждого

признака, как среднее значение, стандартное отклонение, минимальные и максимальные значения, квартильные оценки и т.д. Далее для каждого прогнозируемого параметра были построены гистограммы с наложенной кривой плотности. Для выбора ширины т.н. бинов на гистограммах использовался метод Фридмана–Диакониса [8], поскольку он устойчив к выбросам и адаптируется к размеру выборки, что особенно важно при анализе медицинских данных.

Для статистической проверки соответствия распределения нормальному закону применялся тест Колмогорова–Смирнова с уровнем значимости 0.05. Такой выбор обусловлен тем, что данный критерий универсален и применим к выборкам объемом до нескольких тысяч наблюдений, не требует дискретизации данных и предварительной оценки параметров распределения. Кроме того, он обеспечивает чувствительность как к центральным отклонениям, так и к особенностям в хвостах распределения [9].

Проведенный анализ распределений целевых параметров на основе указанного теста и графиков показал, что временные срезы ЧСС и скорректированного интервала QTc демонстрируют признаки нормальности (p -value выше 0.05). В то же время для необработанного QT имеется статистически значимое отклонение от нормального распределения (p -value меньше 0.01). Сводная информация о результатах анализа распределений целевых величин указана в Таблице 1.

ТАБЛИЦА 1: Анализ распределений целевых величин

Параметр	Число бинов	p -value	Вывод
ЧСС (1-й день)	14	0.1386	нормальное
ЧСС (3-й день)	14	0.1118	нормальное
ЧСС (10-й день)	18	0.5300	нормальное
QT (1-й день)	18	0.0032	не нормальное
QT (3-й день)	24	0.0031	не нормальное
QT (10-й день)	21	0	не нормальное
QTc (1-й день)	24	0.4112	нормальное
QTc (3-й день)	21	0.3051	нормальное
QTc (10-й день)	18	0.1393	нормальное

Так, на Рис. 1 изображены распределения параметра HR на первый, третий и десятый дни наблюдения.

По рисунку видно, что во всех трех случаях данные преимущественно находятся в диапазоне от 100 до 180, но в первый и третий дни также есть экстремальные значения, достигающие 80 и 220 ударов в минуту. Пик ядерной кривой находится около 140, что указывает на моды распределений.

Изучение гистограмм скорректированных интервалов QTc позволяет сделать вывод, что в целом все три распределения сосредоточены в диапазоне от 350 до 450 мс, причем явные выбросы отсутствуют. Гистограммы третьего и десятого дней имеют слабую правостороннюю асимметрию (Рис. 2).

Установленное в Таблице 1 отклонение распределения QT от нормального может указывать на наличие выбросов или иные его особенности, требующие дополнительной трансформации данных или выбора робастных методов моделирования.

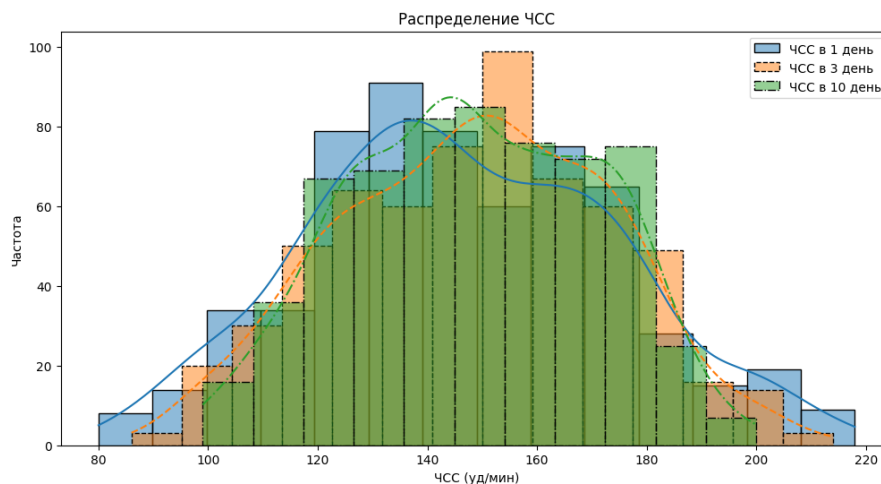


Рис. 1: Распределение HR

2.2. Корреляционный анализ

Для того, чтобы понять степень взаимосвязи между электрокардиографическими признаками и отобрать наиболее информативные входные данные для регрессионных моделей, была рассчитана матрица корреляции Спирмена. Выбор был сделан в пользу этого метода, т.к. он не требует нормального распределения данных (см. результаты предыдущего подраздела) и чувствителен к монотонным, но не обязательно линейным зависимостям между переменными. Коэффициент корреляции Спирмена основан на ранговой шкале [9], что позволяет выявлять взаимосвязи даже при наличии выбросов или данных нелинейной природы и делает его подходящим инструментом для анализа медицинских данных.

Для начала была построена матрица корреляции между целевыми переменными и параметрами, не зависящими от дня наблюдений, такими, как период гестации, оценка по шкале Апгар и т.д. (Рис. 3).

Анализируя данную таблицу, можно сделать вывод, что между целевым параметрами, не зависящими от дня наблюдений, присутствует высокий уровень корреляции. Это вполне объяснимо: каждый из них в своем аспекте описывает состояние новорожденного в одни и те же начальные моменты его жизни.

Слабая корреляция присутствует между ЧСС и QT в первый день, с одной стороны, и сроком гестации, массой, шкалой Апгар на первой и пятой минутах и диагнозом ВЖК, с другой.

Значение QTc сильно коррелирует с QT и ЧСС в соответствующие дни, что объясняется формулой (1), но почти не коррелирует с другими параметрами. Кроме того, степень связи между целевыми переменными в разные дни невысока и падает со временем. Это является отражением постепенной нормализации состояния пациентов при должном наблюдении и уходе.

Далее были построены матрицы корреляции между исходными (P , PQ , QRS , уровень лактата) и целевыми (HR , QT , QTc) параметрами каждого из дней по отдельности. Как выяснилось, значимая корреляция между целевыми и входными

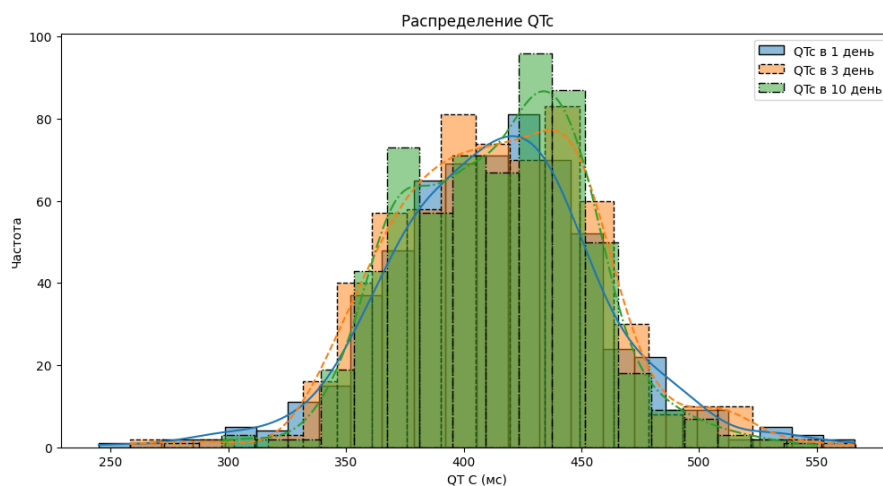


Рис. 2: Распределение QTc

параметрами отсутствует, из чего следует, что эти параметры слабо или вообще не объясняют дисперсию целевых переменных и их использование в прогнозе не является целесообразным.

Помимо количественных в анализе учитывались также и категориальные переменные: метод родоразрешения, наличие ЭКО и т.д. Для оценки зависимости между ними и количественными признаками использовался коэффициент корреляционного отношения η [10]. Результаты поиска указанных зависимостей отображены на Рис. 4.

Анализируя представленную матрицу, можно сделать вывод, что на состояние пациентов влияет наличие ЭКО: между этим показателем, с одной стороны, и ЧСС в первый и третий дни имеется некоторая корреляция. Данный факт впоследствии может быть использован в клинической практике.

3. Построение и тестирование регрессионных моделей

3.1. Выбор моделей

Поскольку явных зависимостей между различными параметрами задачи обнаружить не удалось, классические подходы к описанию искомой функции f (начиная с линейной регрессии) оказываются нерезультативными. С другой стороны, малый объем выборки ($n = 610$) делает невозможным эффективное обучение и использование нейронных сетей.

В связи с указанными обстоятельствами для поиска зависимостей одних электрофизиологических параметров от других использовались следующие модели, описанные, например, в [7]:

- GAM, позволяющая выявлять нелинейные, но интерпретируемые зависимости между признаками. В данном исследовании при использовании GAM применялись кубические B-сплайны, а число узлов было равно 5.

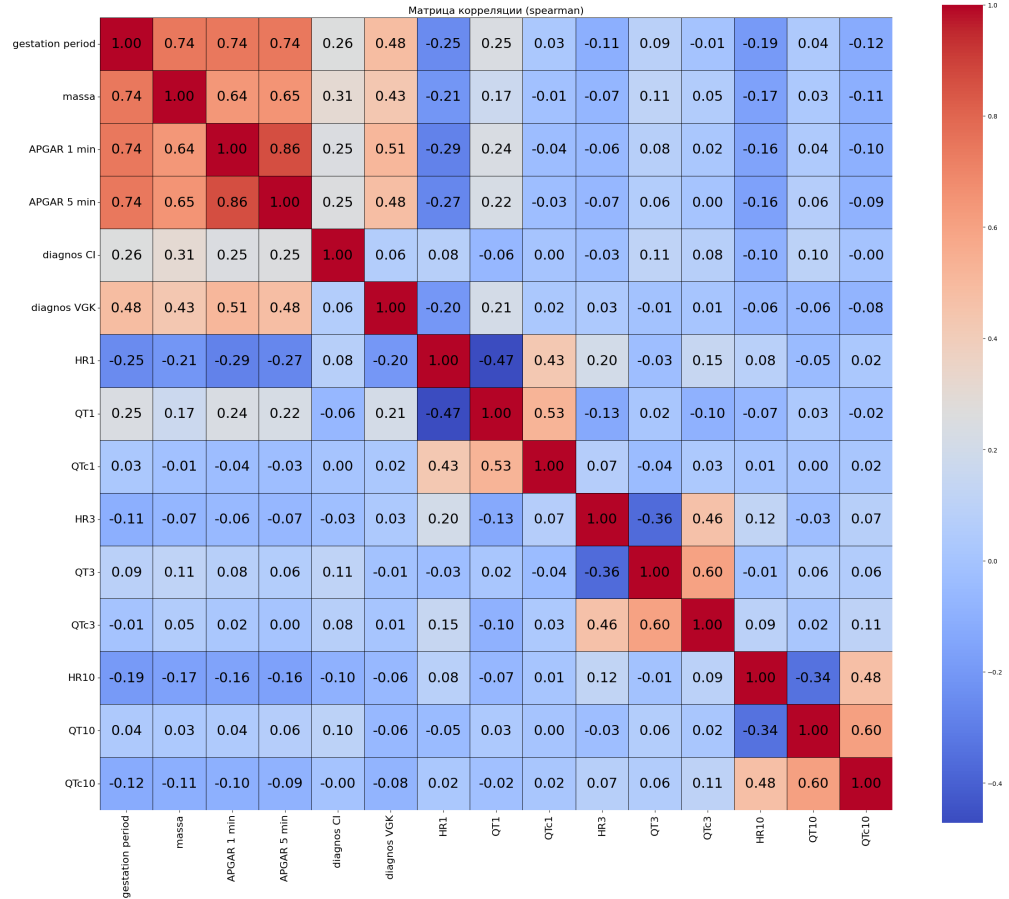
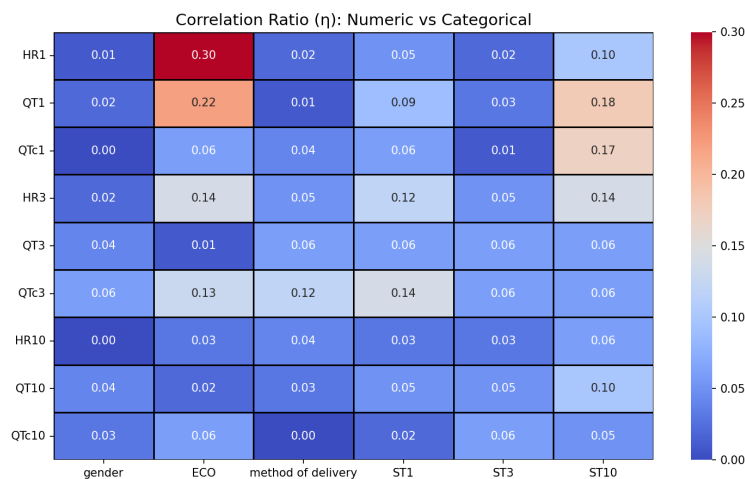


Рис. 3: Корреляция между целевыми переменными и параметрами, не зависящими от дня наблюдений

- MARS, автоматически строящая кусочно-линейные аппроксимации с отбором важных взаимодействий. Степень взаимодействия была выбрана равной 1 (аддитивная модель), во время прямого прохода происходило жадное добавление базисных функций, а при регуляризации и обратном проходе выполнялось автоматическое отсечение по GCV-критерию.
- SVM, демонстрирующая устойчивую работу на сложных распределениях. При реализации SVM в качестве ядра была выбрана радиально-базисная функция (RBF) с параметром регуляризации $C = 1$.

В качестве альтернативы применялась библиотека FLAML [11], позволяющая автоматически подбирать лучшие алгоритмы и их параметры по заданной метрике качества; в нашем случае в роли такой метрики выступила R^2 .

Согласно предварительным расчетам, вне зависимости от дня наблюдений наибольшей предсказуемостью отличались HR и QTc. При их прогнозировании учи-

Рис. 4: Матрица корреляционного отношения η

тывались лишь те входные параметры, которые при корреляционном анализе продемонстрировали значимую связь с искомой характеристикой (модуль коэффициента корреляции не менее 0.1). Так, при моделировании QTc на третий день наблюдений учитывались период гестации, степень тяжести церебральной ишемии, значения ЧСС, ST, P и PQ в первый день, а также ЧСС и ST на третий день.

Общие результаты сравнения моделей при прогнозировании значений этих переменных по критериям MAPE, R^2 и AIC (3)–(5) приведены в Таблице 2. В ней указаны модели, оказавшиеся лучшими по тому или иному критерию; если две модели показывают близкие значения метрики качества, они перечислены через запятую, при этом «выигрывающая» модель названа первой.

Таблица 2: Результаты сравнения моделей по критериям MAPE, R^2 и AIC

Целевой параметр	MAPE	R^2	AIC
ЧСС (1-й день)	GAM, MARS	MARS	MARS
QTc (1-й день)	FLAML, MARS	FLAML, MARS	MARS
ЧСС (3-й день)	MARS	MARS	MARS
QTc(3-й день)	GAM	MARS	MARS
ЧСС (10-й день)	GAM	SVM	SVM
QTc(10-й день)	GAM	MARS	MARS

Следует отметить, что модель GAM, показывающая лучшие результаты по критерию MAPE для параметра QTc на третий и десятый дни наблюдения, дает отрицательное значение критерия R^2 . Это означает, что она плохо объясняет дисперсию целевой переменной, что делает ее применение сомнительным. При прогнозировании ЧСС на десятый день все модели, кроме SVM, дают отрицательный коэффициент детерминации.

Однако аномальное поведение R^2 именно при прогнозировании ЧСС не столь критично, поскольку наиболее информативным с медицинской точки зрения является значение QTc — синтетического параметра, объединяющего в себе как частоту сердечных сокращений, так и величину интервала QT. Поэтому, основываясь на данных Таблицы 2, можно сделать вывод, что наибольшей предсказательной способностью для описанной в исследовании выборки обладает модель MARS.

К примеру, в Таблице 3 более подробно приведены значения критериев качества при прогнозировании величины QTc на третий день наблюдений.

ТАБЛИЦА 3: Критерии качества при прогнозировании QTc на третий день

Модель	MAPE	R^2	AIC
FLAML (ExtraTree)	7.11	0.205	898.58
GAM	7.07	−0.412	996.01
SVM	8.05	0.091	907.36
MARS	7.14	0.232	889.67

Рис. 5 иллюстрирует приближение реальных данных QTc третьего дня (первые 30 наблюдений) различными моделями.

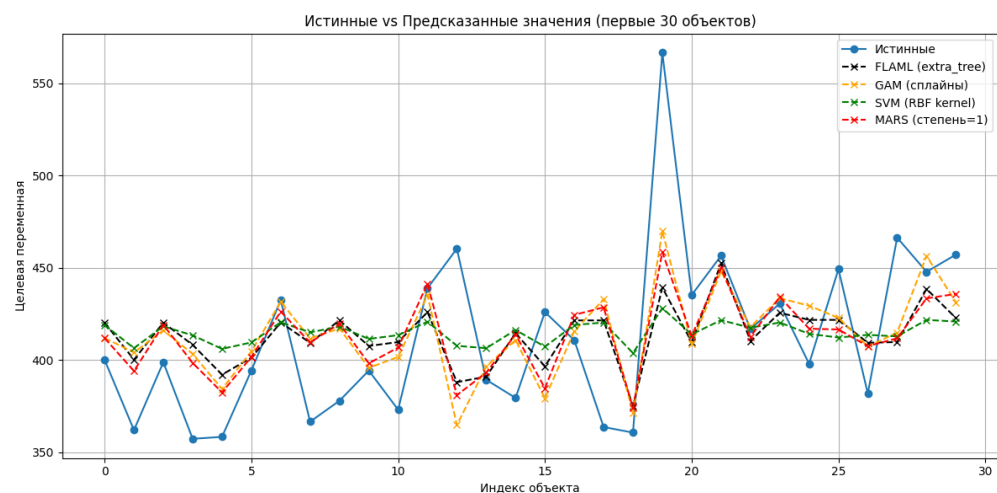


Рис. 5: Сравнение предсказанных QTc (третий день жизни) с наблюдаемыми

Как видно, используемые модели в целом верно отслеживают качественные особенности поведения прогнозируемого параметра («больше — меньше»), но не всегда корректно предсказывают его количественные скачки. Невысокие коэффициенты детерминации, отраженные в Таблице 3, говорят о том, что качество прогноза оставляет желать лучшего. Для его улучшения объем выборки должен быть кратно увеличен; возможно, этого удастся добиться, используя данные не только по республике Мордовия, но и по другим регионам России.

3.2. Проверка моделей

Как было установлено выше, модель MARS с некоторыми оговорками может быть использована для предсказания значений QTc. Для практической проверки этого утверждения она была протестирована на двух клинических случаях.

Клинический случай № 1.

Девочка А., масса тела 2790 г, возраст 1 сутки жизни, диагноз: церебральная ишемия II степени, синдром гипервозбудимости. Согласно данным ЭКГ: ЧСС — 167 уд/мин, продолжительность зубца Р — 0.06 с, интервала PQ — 0.1 с, длительность комплекса QRS — 0.06 с, сегмент ST находится на изолинии 3.

Прогнозируемое значение для интервала QTc на третий день жизни составило 400 мс. При клиническом наблюдении на третьи сутки зарегистрировано фактическое значение QTc = 417 мс.

Клинический случай № 2.

Мальчик Б., масса тела 2910 г, в возрасте 1 сутки жизни, с диагнозом: церебральная ишемия I степени, синдром гипервозбудимости. По данным ЭКГ: ЧСС — 137 уд/мин, продолжительность зубца Р — 0.07 с, интервала PQ — 0.09 с, длительность комплекса QRS — 0.04 с, сегмент ST находится на изолинии 3.

Ожидаемое значение интервала QTc на третьи сутки жизни оказалось равно 423 мс. При анализе ЭКГ в стационаре для того же периода отмечено значение QTc = 363 мс.

Итак, по результатам апробации на двух клинических случаях предиктивная модель интервала QTc у новорожденных детей с церебральной ишемией продемонстрировала удовлетворительные результаты.

Заключение

Таким образом, в рамках данной работы получен алгоритм раннего прогнозирования сердечных патологий у новорожденных детей на базе оптимального подбора модели машинного обучения. Проведенный сравнительный анализ показал, что при малых объемах данных модель MARS способна предсказывать значения QTc у данной категории пациентов с удовлетворительной точностью. В дальнейшем этот результат может быть использован для прогнозирования попадания новорожденного в группу риска. Кроме того, в ходе предварительного корреляционного анализа было установлено умеренное влияние факта ЭКО на ЧСС новорожденного.

Возможность использования модели для первичной оценки ЭКГ-параметров у новорожденных детей подтверждена рассмотрением конкретных клинических случаев.

Проведенный анализ текущих метрик указал на необходимость дальнейшей доработки и совершенствования алгоритма для повышения его эффективности в

диагностике, лечении и профилактике НРС с учётом принципов персонализированной медицины. В частности, повышению точности алгоритма будет способствовать расширение исследуемой выборки детей неонатального периода.

Список литературы

- [1] Chi S., Wang Y., Zhang Y., Zhu W., Li J. Graph Neural Network Based Multi-Label Hierarchical Classification for Disease Predictions in General Practice // *Studies in Health Technology and Informatics*. Vol. 310. 2024. Pp. 725–729. <https://doi.org/10.3233/SHTI231060>
- [2] Aylward B.S., Abbas H., Taraman S., Salomon C., Gal-Szabo D., Kraft C., Ehwerhemuepha L., Chang A., Wall D.P. An introduction to artificial intelligence in developmental and behavioral pediatrics // *Journal of Developmental and Behavioral Pediatrics*. 2023. Vol. 44, № 2. Pp. e126–e134. <https://doi.org/10.1097/DBP.0000000000001149>
- [3] Berghea E.C., Ionescu M.D., Gheorghiu R.M., Tincu I.F., Cobilinschi C.O., Craiu M., Bălgrădean M., Berghea F. Integrating Artificial Intelligence in Pediatric Healthcare: Parental Perceptions and Ethical Implications // *Children*. 2024. Vol. 11, № 2. ID 240. <https://doi.org/10.3390/children11020240>
- [4] Makimoto H., Takahide K. Adopting artificial intelligence in cardiovascular medicine: A scoping review // *Hypertension Research*. 2024. Vol. 47, № 3. Pp. 685–699. <https://doi.org/10.1038/s41440-023-01469-7>
- [5] Obermeyer Z., Ezekiel J.E. Predicting the future — big data, machine learning, and clinical medicine // *The New England Journal of Medicine*. 2016. Vol. 375, № 13. Pp. 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- [6] Bazett H.C. An analysis of the time-relations of electrocardiograms // *Heart*. 1920. № 7. Pp. 353–370.
- [7] Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Springer.
- [8] Freedman D., Diaconis P. On the Histogram as a Density Estimator: L_2 Theory // *Z. Wahrscheinlichkeitstheorie verw. Gebiete*. 1981. Vol. 57. Pp. 453–476.
- [9] Лагутин М.Б. *Наглядная математическая статистика*. М.: БИНОМ. Лаборатория знаний, 2009. 472 с.
- [10] Гланц С. *Медико-биологическая статистика*. М.: Практика, 1998. 459 с.
- [11] AutoML – Regression [Electronic resource] // AutoML-Regression. URL: <https://microsoft.github.io/FLAML/docs/Examples/>.

Образец цитирования

Акишев Д.В., Тягушева Е.Н., Сыромясов А.О., Власова Т.И., Бадюкина Т.Е. О системе предиктивных критериев нарушения сердечного ритма у новорожденных,

основанной на анализе данных // Вестник ТвГУ. Серия: Прикладная математика. 2025. № 3. С. 78–92. <https://doi.org/10.26456/vtpmk736>

Сведения об авторах

1. Акишев Даниил Витальевич

магистрант факультета математики и информационных технологий Национального исследовательского Мордовского государственного университета им. Н.П. Огарёва.

Россия, 430005, г. Саранск, ул. Большевикская, д. 68/1, Мордовский государственный университет. E-mail: rlegar@mail.ru

2. Тягушева Евгения Николаевна

студент Медицинского института Национального исследовательского Мордовского государственного университета им. Н.П. Огарёва.

Россия, 430005, г. Саранск, ул. Большевикская, д. 68/1, Мордовский государственный университет. E-mail: evgenia.tyagsuheva@yandex.ru

3. Сыромясов Алексей Олегович

доцент кафедры прикладной математики Национального исследовательского Мордовского государственного университета им. Н.П. Огарёва.

Россия, 430005, г. Саранск, ул. Большевикская, д. 68/1, Мордовский государственный университет. E-mail: syal1@yandex.ru

4. Власова Татьяна Ивановна

зав. кафедрой нормальной и патологической физиологии Национального исследовательского Мордовского государственного университета им. Н.П. Огарёва.

Россия, 430005, г. Саранск, ул. Большевикская, д. 68/1, Мордовский государственный университет. E-mail: vasova-t.i@mrsu.ru

5. Бадокина Татьяна Евгеньевна

доцент кафедры анализа данных и искусственного интеллекта Национального исследовательского Мордовского государственного университета им. Н.П. Огарёва.

Россия, 430005, г. Саранск, ул. Большевикская, д. 68/1, Мордовский государственный университет. E-mail: badokinate@yandex.ru

ON THE SYSTEM OF PREDICTIVE CRITERIA FOR CARDIAC ARRHYTHMIAS IN NEWBORNS BASED ON DATA ANALYSIS

Akischev D.V., Tiagusheva E.N., Syromiasov A.O., Vlasova T.I.,
Badokina T.E.

Ogarev Mordovia State University, Saransk

Received 24.03.2025, revised 25.05.2025.

The paper examines machine-learning algorithms for detecting cardiac arrhythmias of newborns. Main cardiography intervals are analyzed. Medical histories containing indicators of newborn children for the first, third and the tenth days of life act as the source of data. The authors pay main attention to statistical analysis of this data and to construction of regression models; the results of such analysis are examined as well. Application of constructed models together with calculation of the main metrics shows that explanatory quality of regression algorithms in general is not enough good. Besides that, MARS has more predictive power than other models if the volume of sample group is small enough. To improve diagnostics and to make prevention of cardiac arrhythmias more effective it is necessary to provide further optimization, to enlarge training sample and to increase data diversity.

Keywords: artificial intelligence in pediatrics, prediction of cardiac arrhythmias, machine learning, regression modeling, newborns.

Citation

Akischev D.V., Tiagusheva E.N., Syromiasov A.O., Vlasova T.I., Badokina T.E., “On the system of predictive criteria for cardiac arrhythmias in newborns based on data analysis”, *Vestnik Tvgu. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2025, № 3, 78–92 (in Russian). <https://doi.org/10.26456/vtpmk736>

References

- [1] Chi S., Wang Y., Zhang Y., Zhu W., Li J., “Graph Neural Network Based Multi-Label Hierarchical Classification for Disease Predictions in General Practice”, *Studies in Health Technology and Informatics*. V.310, 2024, 725–729, <https://doi.org/10.3233/SHTI231060>.
- [2] Aylward B.S., Abbas H., Taraman S., Salomon C., Gal-Szabo D., Kraft C., Ehwerhemuepha L., Chang A., Wall D.P., “An introduction to artificial intelligence in developmental and behavioral pediatrics”, *Journal of Developmental and Behavioral Pediatrics*, 44:2 (2023), e126–e134, <https://doi.org/10.1097/DBP.0000000000001149>.

- [3] Berghea E.C., Ionescu M.D., Gheorghiu R.M., Tincu I.F., Cobilinschi C.O., Craiu M., Bălgrădean M., Berghea F., “Integrating Artificial Intelligence in Pediatric Healthcare: Parental Perceptions and Ethical Implications”, *Children*, **11**:2 (2024), 240, <https://doi.org/10.3390/children11020240>.
- [4] Makimoto H., Takahide K., “Adopting artificial intelligence in cardiovascular medicine: A scoping review”, *Hypertension Research*, **47**:3 (2024), 685–699, <https://doi.org/10.1038/s41440-023-01469-7>.
- [5] Obermeyer Z., Ezekiel J.E., “Predicting the future — big data, machine learning, and clinical medicine”, *The New England Journal of Medicine*, **375**:13 (2016), 1216–1219, <https://doi.org/10.1056/NEJMp1606181>.
- [6] Bazett H.C., “An analysis of the time-relations of electrocardiograms”, *Heart*, 1920, № 7, 353–370.
- [7] Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, Springer.
- [8] Freedman D., Diaconis P., “On the Histogram as a Density Estimator: L_2 Theory”, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **57** (1981), 453–476.
- [9] Lagutin M.B., *Naglyadnaya matematicheskaya statistika [Visual mathematical statistics]*, BINOM. Laboratoriya znaniy, Moscow, 2009 (in Russian), 472 pp.
- [10] Glancz S., *Mediko-biologicheskaya statistika [Medical and biological statistics]*, Praktika, Moscow, 1998 (in Russian), 459 pp.
- [11] *AutoML – Regression*, AutoML-Regression, <https://microsoft.github.io/FLAML/docs/Examples/>.

Author Info

1. Akishev Daniil Vital’evich

Master Student at Mathematics and Information Technologies Faculty, Ogarev Mordovia State University.

Russia, 430005, Saransk, Bolshevikskaya str., 68/1, Ogarev Mordovia State University. E-mail: rlegar@mail.ru

2. Tiagusheva Evgenia Nikolaevna

Student of Medicine Institute, Ogarev Mordovia State University.

Russia, 430005, Saransk, Bolshevikskaya str., 68/1, Ogarev Mordovia State University. E-mail: evgenia.tyagsuheva@yandex.ru

3. Syromiasov Aleksei Olegovich

Associate Professor at Applied Mathematics Department, Ogarev Mordovia State University.

Russia, 430005, Saransk, Bolshevikskaya str., 68/1, Ogarev Mordovia State University. E-mail: syal1@yandex.ru

4. Vlasova Tatiana Ivanovna

Head of Normal and Pathological Physiology Department, Ogarev Mordovia State University.

Russia, 430005, Saransk, Bolshevikskaya str., 68/1, Ogarev Mordovia State University. E-mail: vlasova-t.i@mrsu.ru

5. Badokina Tatiana Evgenyevna

Associate Professor at Data Analysis and Artificial Intelligence Department, Ogarev Mordovia State University.

Russia, 430005, Saransk, Bolshevikskaya str., 68/1, Ogarev Mordovia State University. E-mail: badokinate@yandex.ru