

СРАВНИТЕЛЬНЫЙ ОБЗОР ИСПОЛЬЗОВАНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СЕРДЕЧНО-СОСУДИСТОГО РИСКА

Д.А. ШВЕЦ*, С.В. ПОВЕТКИН**

* *БУЗ Орловской области «Орловская областная клиническая больница»,
Бульвар Победы, д.10, г. Орёл, 302028, Россия, e-mail: denpost-card@mail.ru*

** *Курский государственный медицинский университет,
ул. К.Маркса, д.3, г. Курск, 305041 Россия, e-mail: clinfarm@kursknet.ru*

Аннотация. Постоянное увеличение объёма информации в кардиологии делает актуальным разработку новых методов анализа данных. Используя существующие подходы оценки риска, невозможно спрогнозировать около половины эпизодов острого коронарного синдрома. Машинное обучение больших данных может привести к улучшению результатов диагностики и лечения при более низких затратах. Индуктивный подход позволяет выявлять закономерности, возникающие в результате анализа данных и разрабатывать алгоритмы, способные учиться самостоятельно. Несмотря на то, что модели оценки сердечно-сосудистого риска, полученные на основе методов машинного обучения превосходят традиционные калькуляторы риска, до настоящего времени не было проведено ни одного крупномасштабного исследования с применением машинного обучения для доказательства прогностической роли в общей популяции с использованием рутинных клинических данных. Кроме того, нет чётких рекомендаций, какой из алгоритмов будет работать лучше в той или иной ситуации. Использование эмпирического подхода при выборе метода машинного обучения и принцип «чёрного ящика» затрудняют проведение крупных исследований и внедрение методов машинного обучения в клиническую практику. Данный обзор литературы знакомит исследователей с новыми направлениями, перспективами и трудностями при использовании методов машинного обучения для прогнозирования сердечно-сосудистого риска.

Ключевые слова: машинное обучение, сердечно-сосудистый риск.

COMPARATIVE REVIEW OF THE USE OF MACHINE LEARNING METHODS FOR PREDICTING CARDIOVASCULAR RISK

D.A. SHVETS*, S.V. POVETKIN**

* *Orel clinical regional hospital, Victory Boulevard, 10, Orel, 302028, Russia, e-mail: denpost-card@mail.ru*

** *Kursk State Medical University, K. Marx Str., 3, Kursk, 305041, Russia, e-mail: clinfarm@kursknet.ru*

Abstract. Constant augmentation in the volume of information in cardiology makes the development of new data analysis methods really timely and pressing. It is not possible to predict about half of the cases of Acute Coronary Syndrome while applying only already existing risk evaluation approaches. Machine learning of large data may result in the improvement of diagnostics results and treatment with lower expenditures. An inductive approach allows identifying regularities appearing as a result of data analysis and developing algorithms capable of self-education. Despite the fact that models of cardiovascular risks derived on the basis of machine learning methods surpass traditional risk calculators, so far not a single large scale research has been carried out which would employ machine learning with the purpose of confirmation of predictive role in a general population with the help of routine clinical findings. Besides, there are no well-defined recommendations which of the algorithms will work better in different situations. The use an empirical approach while choosing a machine learning method and black box principle complicate a wide-scale research and implementation machine learning method into clinical experience. The given background paper introduces researches to new directions, perspectives, and difficulties associated with machine learning methods for cardiovascular risk prediction.

Keywords: machine learning, cardiovascular risk.

В связи со старением населения, растущей урбанизацией, глобализацией, *сердечно-сосудистые заболевания* (ССЗ) остаются ведущей причиной инвалидности и преждевременной смерти во всем мире. Исследования в области сердечно-сосудистой медицины привели к терапевтическим достижениям и снижению темпов роста сердечно-сосудистой смертности за последние 50 лет. Несмотря на усилия по совершенствованию профилактики, диагностики и лечения, уровень смертности и заболеваемости ССЗ растёт из года в год. Исследования показали, что приблизительно 80% смертей, связанных с ССЗ, случаются в странах с низким и средним уровнем дохода, и ожидается, что к 2030 году их число возрастет до 23,3 миллионов [20, 23, 25].

В последнее время в кардиологии наблюдается некоторая стагнация в развитии новых методов лечения. Считается, что традиционная парадигма, основанная на крупномасштабных исследованиях конкретных патологий недостаточна для снижения бремени ССЗ [9]. Актуальна разработка новых форм сердечно-сосудистых исследований, способных экономически эффективно анализировать большие объемы информации. При этом, большие данные могут предложить альтернативу традиционному дедуктивному подходу. «Эра больших данных» в здравоохранении только начинается. Распространение электронной истории болезни, повышение доступности цифровых медицинских данных из таких источников, как приложения, биосенсоры и различные коммуникационные технологии способствовали увеличению доступности больших данных [23]. Эти постоянно растущие массивы данных позволяют применять аналитические методы, такие как машинное обучение (*machine learning (ML)*), для выполнения сложных итерационных операций и разработки алгоритмов прогнозирования. Использование инструментов анализа больших данных в кардиологии может привести к улучшению результатов лечения при более низких затратах [12, 17, 19]. В настоящее время количество наряду с качеством данных определяют точность полученного результата. Следовательно, анализируемые данные указывают исследователям на наиболее перспективное направление. Помогают в решении поставленных задач хранилища оперативных данных, облачные сервисы, которые периодически извлекают, реорганизуют данные для оперативного запроса и обеспечивают мгновенный доступ к информации [10, 20]. По мере того, как сбор информации становится проще и дешевле, можно ожидать увеличения количества исследований, основанных на объединении данных. Однако, данные часто собираются после выписки больного из стационара частично или полностью вручную из медицинской карты, что является ограничением для их сбора в рамках регистра [23]. Швеция и Великобритания одними из первых стали собирать данные по больным с *острым коронарным синдромом* (ОКС) в национальном масштабе [14, 20]. В ряде стран электронные медицинские записи в организациях здравоохранения хранятся на центральных серверах, что позволяет осуществлять разработку новых алгоритмов за пределами учреждения здравоохранения с использованием программного обеспечения для облачных вычислений [24]. Такие подходы имеют индуктивные элементы [9]. При этом, генерация гипотез не перекладывается на машины, а закономерности, возникающие в результате анализа данных, принимаются в качестве полезных подсказок. Дальнейший шаг – принятие клинических решений, при котором алгоритмы будут иметь возможность учиться самостоятельно [10, 17, 19].

Основой сердечно-сосудистой эпидемиологии является прогнозирование риска негативных событий на относительно длительные периоды времени (5-10 лет). Тема прогноза имеет важное значение в кардиологии, закладывая основу для принятия клинических решений. Точная стратификация риска для пациента с ОКС остаётся сложной задачей, играющей важную роль в доказательной медицине [6, 14, 16]. Эпидемиологические исследования, организуемые для получения таких прогностических моделей, часто содержат сотни или тысячи переменных. Поэтому, методы *ML* эффективны в качестве средства выявления наилучших предикторов из большого количества точек фенотипических данных [3]. Коронарная болезнь сердца – многофакторное заболевание, требующее персонализированного подхода к лечению, что усложняет проведение эффективных лечебных и профилактических мероприятий. С помощью *ML* можно охватить сложность, присущую механизмам заболевания и выявить наиболее подходящие фенотипы для эффективного патогенетического лечения [9, 10]. Считается, что даже высокоэффективные методы лечения имеют неоднородный эффект на индивидуальном уровне. Исходя из этого, большие данные могут быть основой персонализированной медицины, когда алгоритмы *ML* предсказывают индивидуальный риск для пациента и более точно определяют точки приложения конкретных методов лечения. Многомерные данные, извлеченные такими технологиями, в сочетании с потенциалом для двустороннего взаимодействия в реальном времени между пациентами и врачами, обещают более детальную и динамичную индивидуализированную помощь [4, 6, 9, 25].

ML является новым разделом в области инновационной медицины, который все чаще используется в клинических исследованиях с целью улучшения прогностического моделирования и выяснения новых предикторов негативного исхода. Стимулирующим фактором для использования *ML* в кардиологии явилось появление беспрецедентно большого числа клинически значимых признаков и высокоспецифичных диагностических тестов [10]. Однако, несмотря на очевидные перспективы, использование *ML* в обработке больших данных вызывает много вопросов. На сегодняшний день не было проведено ни одного крупномасштабного исследования с применением *ML* для доказательства прогностической роли в общей популяции с использованием рутинных клинических данных [2, 8, 15, 17, 20, 24]. В последние годы появилось много данных, но не больше доказательств возможности клинического применения *ML*. Без таких доказательств потенциал больших данных и машинного обучения в сердечно-сосудистой медицине не может быть реализован [23].

Концепция больших данных была впервые выдвинута в 1997 году. Под этим термином понимают наборы данных, которые являются слишком большими или сложными для традиционной статистики. Не существует согласованного определения и признанного порога больших данных. По мнению некоторых авторов, для определения размера больших данных предложен $\log(n \times p)$ превосходящий или равный 7

(например, 100000 больных, имеющих 100 признаков). Помимо размера, проблема больших данных заключается в их сложности, основанной на гетерогенности, многомерности и том факте, что они являются динамическими (все предыдущие отдельные измерения динамически связаны). Наиболее популярное описание больших данных было предложено *D. Laney* в 2001 году и известно в академическом мире как “3Vs”: *volume, variety, velocity* (объем, разнообразие, скорость) [20]. *Искусственный интеллект* (ИИ) определяется как теория и разработка компьютерных программ и систем, способных выполнять сложные задачи, требующие человеческого уровня интеллекта [4, 23]. *ML* – это подмножество искусственного интеллекта, которое использует алгоритмы, автономно получающие знания путем извлечения шаблонов из данных [2]. *ML* относится к семейству алгоритмов, имеющих общую способность итеративно выяснять закономерности (обучаться), с целью оптимизации задач прогнозирования или классификации [4]. *ML* это союз между математикой и информатикой, который обусловлен уникальными вычислительными задачами построения статистических моделей из больших наборов данных [6]. Кроме того, *ML* – это программное решение на основании больших данных с возможностью делать прогнозы без предварительного программирования [20]. Искусственная *нейронная сеть* (НС) имитируют структуру нервной системы человека и являясь непараметрической методикой с хорошо известной способностью к обобщению. НС может предсказывать отдаленные исходы с заранее выбранными уровнями чувствительности и специфичности. Глубокое обучение – это тип машинного обучения, который включает в себя нейронные сети, предназначенные для моделирования высокоуровневых абстракций из множественных слоев обработки данных с часто чередующимися линейными и нелинейными преобразованиями [8].

Существует множество моделей, оценивающих риск ближайших и отдаленных осложнений при ОКС. Многие из них получены на основании небольших когортных исследований и не используются на практике, потому что врачи считают их неудобными и трудоемкими. Традиционные статистические подходы, такие как логистическая регрессия или модель рисков Кокса позволяют делать выводы об ассоциациях между ограниченным числом линейно связанных переменных без полной характеристики системы [4, 9, 19]. Появление новых данных в традиционных алгоритмах вызывает необходимость в их пересчете. При таком подходе устанавливаются многочисленные критерии исключения, вводятся заранее определенные переменные, с предполагаемой связью с исходом. Как правило не учитывается потенциальная прогностическая ценность взаимодействия между несколькими факторами риска и первичным результатом. Кроме того, традиционные модели построены на статических особенностях пациента и пренебрегают влиянием динамической информации о лечении. Вследствие таких ограничений существующие алгоритмы обычно демонстрируют скромную прогностическую эффективность [2, 8, 20]. Сердечно-сосудистых риск в настоящее время прогнозируют статистическими моделями, полученными на основании исследований 25-30-ти летней давности. Учитывая, что риск больных ОКС в данных исследованиях был выше, чем в настоящее время и, принимая во внимание возможное изменение роли отдельных факторов риска, диагностическая точность ранних статистических моделей снижается. Это особенно важно в связи с тем, что в последнее время произошли значительные изменения в эпидемиологии *инфаркта миокарда* (ИМ), характеризующиеся сдвигом в структуре заболеваемости, улучшением краткосрочных методов ведения и вторичной профилактики, снижением ранней смертности, переходом от инцидентов к рецидивам и от догоспитальной смертности к госпитализированным ИМ, а также увеличением бремени заболеваемости и смертности от некардиальных причин [4]. Используя существующие подходы оценки риска, невозможно спрогнозировать около половины эпизодов ОКС. Кроме того, традиционные оценки риска улучшили определение начальной терапевтической стратегии при ОКС, однако они не предназначены для учета последствий стационарной терапии или оценки риска более мягких исходов. Такой подход оказался менее точен в оценке риска отдельных пациентов. В тоже время, ряд больных получают ненужное профилактическое лечение на основании таких прогнозов. Указанные обстоятельства побуждают исследователей искать новые подходы к моделированию и предсказанию риска [5, 11, 14, 17, 18, 24].

Недостатком когортных исследований является чрезмерное упрощение модели для удобства использования в клинической практике. Поэтому, некоторые взаимосвязи не оцениваются, что потенциально приводит к менее удовлетворительным прогностическим характеристикам [4]. В целом, использование традиционных статистических методов может быть затруднительным или нецелесообразным с учетом сложного характера больших данных. Кроме того, включение основных (традиционных), но нелинейно связанных факторов риска в такие модели снижает их прогностическое значение. По этой причине всё чаще применяются новые статистические методы, такие как *ML* [3]. Если традиционная статистика оценивает и объясняет данные, то *ML* нацелено на достижение практического прогнозирования, на основе имеющихся данных. Алгоритмы *ML*, благодаря свойству обучаться, адаптируются к различным изменениям, каждый раз применяя новые данные для поиска решения поставленной задачи [3, 10]. С помощью методов *ML* можно идентифицировать больных после ОКС как с повышенным риском клинических событий, так и с высокой нагрузкой неконтролируемых факторов риска. Модели с *ML*, использующие

большие данные, обеспечивают долгосрочные прогнозы с точностью, сопоставимой или превосходящей хорошо проверенные оценки сердечно-сосудистого риска [15, 16].

Таблица 1

Сравнение прогностической точности оценки сердечно-сосудистого риска традиционными моделями и *ML*

Автор, страна	Количество больных, источник данных: исследование, база данных, <i>EHR</i>	Методы <i>ML</i> , модель оценки риска	Факторы негативного исхода, длительность наблюдения, <i>AUC</i> метода
<i>Motwani M.</i> , США [10]	10300 <i>CONFIRM</i>	<i>LogitBoost</i> , <i>LR</i>	смертность 5 лет <i>LogitBoost</i> =0,79 <i>LR</i> =0,64
<i>Kakadiaris I.A.</i> , США [11]	6459 <i>MESA</i> и <i>FLEMENGHO</i>	<i>SVM</i>	<i>MACE</i> 13 лет <i>SVM</i> =0,94, калькулятор риска <i>ACC/AHA</i> =0,72.
<i>Pieszko K.</i> , Польша [16]	5053 <i>EHR</i>	<i>XGBoost</i> , <i>GRACE 2</i> (<i>LR</i>)	смертность 1 год <i>XGBoost</i> =0,72, калькулятор <i>GRACE</i> =0,71
<i>VanHouten J.P.</i> , США [21]	20078 <i>EHR</i>	<i>RF</i> , нейросеть, <i>RR</i> , <i>TIMI</i> , <i>GRACE</i> (<i>LR</i>)	смертность 30 дней <i>RF</i> =0,85, Нейросеть=0,82 <i>RR</i> =0,81, <i>TIMI</i> =0,74 <i>GRACE</i> =0,62.
<i>Wallert J.</i> , Швеция [22]	51943 <i>SWEDHEART/RIKS-HIA</i>	<i>LR</i> , <i>Boosted C5.0</i> (ансамбль решений), <i>RF</i> , <i>SVM</i>	смертность 2 года <i>LR</i> =0,83, <i>Boosted</i> =0,81 <i>RF</i> =0,842, <i>SVM</i> =0,845
<i>Weng S.F.</i> , Великобритания [24]	383592 <i>CPRD</i>	<i>LR</i> , <i>RF</i> , <i>XGBoost</i> , нейросеть	развитие <i>CC3</i> 10 лет <i>LR</i> =0,760, <i>RF</i> =0,745, <i>XGBoost</i> =0,761, нейросеть=0,764
<i>Ahmad T.</i> , Швеция [1]	44 886 <i>SwedeHF</i>	<i>k-NN</i> , модель СН в Сиэтле, <i>MAGGIC</i>	СН после <i>ОКС</i> 1 год <i>k-NN</i> =0,83 модель СН в Сиэтле=0,73 <i>MAGGIC</i> =0,74
<i>Al'Aref S.J.</i> , США [2]	479 804, <i>PCIRS</i>	<i>AdaBoost</i> , <i>XGBoost</i> , <i>RF</i> , <i>LR</i> .	Внутрибольничная смертность <i>AdaBoost</i> =0,923, <i>XGBoost</i> =0,906, <i>RF</i> =0,892 <i>LR</i> = 0,908
<i>Ambale-Venkatesh B.</i> , США [3]	6814 <i>MESA</i>	<i>RF</i> , регрессия Кокса	<i>MACE</i> 12 лет <i>RF</i> =0,86, регрессия Кокса=0,8
<i>Blom M.C.</i> , Швеция [5]	120940 Шведский регистр <i>ОКС</i> , <i>EHR</i>	<i>LR</i> , <i>SVM</i> , <i>k-NN</i> , <i>AdaBoost</i> , <i>RF</i> , <i>MLP</i> .	Смертность 30 дней <i>LR</i> =0,94, <i>SVM</i> =0,94, <i>k-NN</i> =0,92, <i>AdaBoost</i> =0,5, <i>RF</i> =0,93, <i>MLP</i> =0,87
<i>Duan H.</i> , Китай [7]	2930 <i>EHR</i>	<i>LR</i> , <i>Boosted-RMTM</i> , <i>mix model</i> , <i>CNN</i>	<i>MACE</i> во время госпитализации <i>LR</i> =0,64, <i>Boosted</i> =0,7, <i>mix model</i> =0,68, <i>CNN</i> =0,713
<i>Kwon J.M.</i> , Южная Корея [13]	25977 <i>KorMI</i>	<i>DL</i> , <i>RF</i> , <i>LR</i>	Смертность 6 месяцев <i>DL</i> =0,87, <i>RF</i> =0,85, <i>LR</i> =0,84

Примечание: *EHR* – электронная медицинская карта; большие сердечно-сосудистые события (*MACE*); площадь под *ROC*-кривой (*AUC*); *GRACE*, *TIMI* – модели на основе регистров *ОКС*; *сердечная недостаточность* (СН); *LogitBoost* – ансамблевый метод: сочетание логистической регрессии (*LR*) и *boosting*; *экстремальный градиентный boosting* (*XGBoost*); *random forest* (*RF*); *свёрточная нейронная сеть* (*CNN*); *многослойный перцептрон* (*MLP*); *глубокое обучение* (*DL*); *градиентный бустинг* (*Boosted-RMTM*), *смешанная модель* (*mix model*); *гребневая регрессия* (*RR*)

Кардиология является одной из областей медицины с наибольшей перспективой применения машинного обучения для извлечения закономерностей из больших данных. В последнее время в кардиологии всё чаще используются методы *ML* для прогнозирования выживаемости пациентов и для оценки эффективности различных лечебных стратегий [8, 11, 16]. Не существует универсального алгоритма для решения конкретной задачи. Наиболее популярные алгоритмы, используемые в машинном обучении: линейная и логистическая регрессия, линейный дискриминантный анализ, деревья принятия решений, наивный Байесовский классификатор, *k*-ближайших соседей (*k-nearest neighbors algorithm, k-NN*), метод опорных векторов (*support vector machine, SVM*), бэггинг (*bootstrap aggregating*) и *random forest, boosting* и *adaptive boosting (AdaBoost)*, нейросети [9, 10, 13, 23]. Существуют и сочетанные алгоритмы, использующие сильные стороны каждого: например, логистическая регрессия и нейросетевой анализ [15]. В некоторых исследованиях выявляются более высокие показатели точности методов *ML* (табл. 1) [5, 10, 13, 24].

Как видно из представленной таблицы, методы *ML* имеют преимущества над традиционными статистическими методами и моделями. В отдельных исследованиях, сравнивающих методы машинного обучения больших данных с традиционными статистическими методами, не показано значимого преимущества при оценке прогноза [10, 15, 18].

Основные виды машинного обучения: классическое (контролируемое и неконтролируемое), обучение с подкреплением, ансамблевые методы и нейросети. Контролируемое обучение осуществляется на основе помеченных данных и обычно используется для оценки риска. Типичными задачами, решаемыми при обучении под наблюдением (контролем), являются классификация и регрессия. Лучший пример контролируемого обучения в кардиологии - *Framingham Risk Score* при ишемической болезни сердца. Неконтролируемое обучение выводит структурные отношения и зависимости из данных без каких-либо предварительных знаний или ограничений [5, 12, 20]. В данную группу входят кластеризация, поиск правил и уменьшение размерности (обобщение). Эти модели используются главным образом для получения новых сведений о структуре данных, что позволяет развивать прецизионное фенотипирование и дифференцированный подход к лечению различных подтипов сердечно-сосудистых заболеваний. В обучении с подкреплением модель учится принимать решения, получая вознаграждение за свои действия [9, 10, 25]. В настоящее время существует большое разнообразие алгоритмов машинного обучения и моделей данных. При этом нет чётких рекомендаций, какой из алгоритмов будет работать лучше в той или иной ситуации. Эмпирический подход состоит в том, чтобы попробовать несколько алгоритмов, что может занять много времени. Специалисты по машинному обучению имеют набор инструментов для извлечения функций и предварительной обработки, а также подмножество контролируемых и неконтролируемых алгоритмов обучения, которые они привыкли использовать [6]. Нейросетевые модели считаются «чёрными ящиками» и, как правило, дают мало информации о том, что и как они изучают [10, 13]. *CNN* представляют собой эволюцию традиционных нейронных сетей и в настоящее время являются самыми популярными моделями. *CNN* сильно не зависят от местоположения и масштаба объекта и требуют меньше вычислительной мощности. Модель, которая хорошо обучена на одной большой исходной задаче, может быть использована для быстрого успеха для решения другой. *Рекуррентные нейронные сети (RNN)* могут быть использованы для изучения и прогнозирования последовательных данных, таких как электрокардиограмма, непрерывные измерения артериального давления и кривые перфузии. Вместо типичных сетей прямой передачи они не обрабатывают данные на определенном временном интервале, а отслеживают предшествующие данные и изменения данных с течением времени [4, 6].

Несмотря на оптимизм, нужно осознавать недостатки и потенциальные угрозы, которые присущи таким мощным инструментам, как искусственный интеллект. Существенным фактором, затрудняющим внедрение машинного обучения в медицине, является потребность в больших данных, которые не всегда доступны. До настоящего времени подавляющее большинство медицинских карт предназначены не для исследований, а для отчётов страховым компаниям [4]. Кроме того, при глобализации регистров нужно учесть существующие различия в рекомендациях [17, 18]. Качество моделей прогнозирования зависит от набора данных. При низком качестве исходных данных существует возможность распространения дезинформации, что может привести к неверной интерпретации данных и ненужному вмешательству. Принцип «чёрного ящика» может привести к неточности прогноза, величина которого будет увеличиваться при дальнейшем смещении данных. Кроме того, отсутствуют доказательства возможности интеграции полученных данных в клиническую практику [10, 13, 18]. Эти непредвиденные последствия являются общими для любой новой технологии и должны быть тщательно рассмотрены перед внедрением моделей в больших масштабах. Нужны большие сравнительные исследования, которые позволят доказать преимущества технологии *ML* [17]. Существуют различные проблемы, тормозящие процесс внедрения больших данных в кардиологии. Среди них: чрезмерное количество пропусков и проблема дисбаланса данных [23]. Одной из проблем больших медицинских данных является сложность интеграции результатов анализа в ежедневную кардиологическую практику. Интерпретация результатов анализа может быть искажена субъективными предположениями или манипуляциями аналитиков. Качество данных может

является сомнительным, что не позволяет сделать однозначные выводы [20]. Техника глубокого обучения не может быть запрограммирована на получение новых знаний о физиологии сердца, а скорее зависит от повседневных особенностей, которые могут не иметь интуитивного смысла за пределами наборов данных. Несмотря на это, существует оптимизм в использовании сложных статистических методов анализа в раскрытии причинно-следственных выводов [10]. В настоящее время искусственные сети могут дать лишь узкие представления о патофизиологических процессах и далеки от решения абстрактных задач, которые под силу только человеческому мозгу. Таким образом, существуют противоречивые данные о возможности клинического применения ИИ. Считается, что большинство современных моделей *DL* не предназначены для прогнозирования, а также не имеют возможности прогнозировать предполагаемое время наступления неблагоприятного события [4]. Прежде чем пытаться предсказывать развитие заболевания, нужно иметь максимальное количество патофизиологических данных. Однако, учесть все предикторы невозможно. Трудно предположить, что с помощью неконтролируемого обучения с его ограниченным набором аберрантных путей можно раскрыть патофизиологический смысл заболевания. Вследствие этого концепция прецизионной медицины вряд ли будет полезна [6, 10, 18, 25]. К сожалению, анализ больших массивов данных все еще зачастую не оптимален из-за недостаточного знания исследователем имеющихся статистических и методических инструментов [20]. Несмотря на улучшение технологии, нельзя быть окончательно уверенным в её надёжности. *ML* можно использовать как дополнительную помощь для оптимизации в принятии решения [14, 25]. Неизвестна практическая и медицинская ценность информации, полученной на основе данной технологии [25]. Существует сложность с объяснением механизмов и подходов при оценке ССР [13]. Большая проблема в настоящее время – защита конфиденциальности данных. Как правило цель исследования увязывается с индивидуальным согласием на момент сбора данных, что затрудняет их повторное использование. Тем не менее, многие ценные данные могут быть «обезличены» и использованы повторно [10, 14, 20, 23]. Сложности существуют при использовании современных мобильных устройств для наблюдения за пациентами после ОКС. Непонятны механизмы финансовой и юридической ответственности производителей таких устройств [6].

Считается, что исследования больших данных и медицинская информатика помогут преобразовать медицинскую практику. Наличие больших данных и новых методов анализа могут привести к разработке программ, оценивающих состояние больного с точностью врача. Применение прогностических моделей на основе рутинных клинико-инструментальных и лабораторных данных позволит проводить индивидуальную терапию пациента и прогнозировать ответ на терапию [4]. Большие данные обещают изменить многие решения в области здравоохранения, основываясь на более мощной прогностической аналитике. Однако, в настоящее время отсутствуют убедительные доказательства этой концепции [18]. Существуют оптимистичные перспективы использования ИИ в сердечно-сосудистой медицине. Результаты показали, что оценка риска, основанная на ИИ, превосходит традиционное клиническое суждение и улучшает способность предсказывать будущие события, что является основным направлением в эпоху прецизионной медицины [25]. Однако, необходимо ещё изучить целесообразность применения машинного обучения в клинической практике [9, 24]. Дополнительные опции электронных устройств в виде исследования физиологических потенциалов, способствуют накоплению больших наборов данных. Такой подход создаст основу для непрерывного динамического мониторинга состояния сердечно-сосудистой системы, что является шагом к персонализированной медицине [14]. В ближайшие десятилетия ряд хорошо разработанных исследований с применением больших данных откроет новую парадигму в медицине. Компьютерное моделирование позволит моделировать персонализированное сердце путем интеграции нескольких диагностических данных и обеспечит платформу для виртуальной оценки и оптимизации терапии. Такие синтетические данные, созданные путем анализа существующих данных с использованием методов машинного обучения, могут быть эффективной альтернативой для подготовки обучающих данных [10].

Заключение. Машинное обучение больших данных в настоящее время всесторонне изучается. Однако, назвать искусственным интеллектом современный уровень машинного обучения нельзя. Машина служит помощником врачу, экономя время, выполняя рутинную работу, однако, в настоящее время нет основания считать, что машина может принимать решения самостоятельно. Эмпатия, проявление и восприятие эмоций недостижимо для современного уровня машинного обучения. Увеличение объёма данных и наличие высоких вычислительных скоростей ещё недостаточно для создания разумных алгоритмов, способных предсказывать событие. Очевидно, что всегда имеет место неточность прогноза, связанная с невозможностью сбора всей информации о процессе или явлении. Непрерывное мониторирование биосигналов (электрокардиограмма, артериальное давление, эхокардиография и др.) предполагает появление большого количества информации в режиме реального времени. Можно предположить, что многие диагностические методы исследования перейдут в цифровой формат. Врачебный контроль при сборе и оценке информации будет снижаться, ограничиваясь проведением дифференциального диагноза и выбором наиболее подходящего метода исследования, в том числе оптимального алгоритма анализа

данных. Причём, данные алгоритмы могут иметь для лечебного процесса такое же важное значение, как и другие методы исследования.

Предполагается, что объединение различных регистров ОКС позволит создать и обучить большой объём данных. Хранилище обученных данных должно быть доступно для любого лечебного учреждения. Двустороннее взаимодействие с такими данными позволит сопоставлять новый случай ОКС с группой идентичных данных с целью получения прогностических характеристик. Так как диагностические возможности каждого медицинского центра различные, то общий набор признаков (факторов) по каждому новому случаю ОКС будет различаться. В силу такого несовпадения формата данных и в зависимости от самих данных (количественные, порядковые, повторные измерения), могут потребоваться различные алгоритмы, которые наиболее эффективно прогнозируют исходы и определяют лечебную тактику. Таким образом, при единых данных могут быть созданы различные алгоритмы, отличающиеся прогностической мощностью и областью применения.

Убедительных доказательств преимуществ машинного обучения над другими методами статистического анализа не выявлено. Учитывая потенциал всестороннего индуктивного анализа больших данных, можно ожидать появления больших многоцентровых сравнительных исследований с использованием различных алгоритмов машинного обучения.

Литература

1. Ahmad T., Lund L.H., Rao P. Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to Therapy in a Large Cohort of Heart Failure Patients // *J Am Heart Assoc.* 2018. Vol. 7(8). P. e008081. DOI: 10.1161/JAHA.117.008081
2. Al'Aref S.J., Singh G., van Rosendaal A.R.. Determinants of In-Hospital Mortality After Percutaneous Coronary Intervention: A Machine Learning Approach // *Journal of the American Heart Association.* 2019. Vol. 8(5). P. 1-9. DOI: <https://doi.org/10.1161/JAHA.118.011160>
3. Ambale-Venkatesh B., Yang X., Wu C.O. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis // *Circ Res.* 2017. Vol. 121(9). P. 1092–1101. DOI: 10.1161/CIRCRESAHA.117.311312
4. Benjamins J.W., van Leeuwen K., Hofstra L. A primer in artificial intelligence in cardiovascular medicine // *Neth Heart J.* 2019. Vol. 27(9). P. 392–402. DOI: 10.1007/s12471-019-1286-6
5. Blom M.C., Ashfaq A., Sant'Anna A. Training machine learning models to predict 30-day mortality in patients discharged from the emergency department: a retrospective, population-based registry study // *BMJ Open.* 2019. Vol. 9(8). P. e028015. DOI: 10.1136/bmjopen-2018-028015
6. Deo R.C. Machine Learning in Medicine // *Circulation.* 2015. Vol. 132(17). P. 1920–1930. DOI: doi.org/10.1161/CIRCULATIONAHA.115.001593
7. Duan H., Sun Z., Dong W. Utilizing dynamic treatment information for MACE prediction of acute coronary syndrome // *BMC Med Inform Decis Mak.* 2019. Vol. 19(5). P. 1–11. DOI: 10.1186/s12911-018-0730-7
8. Henglin M., Stein G., Hushcha P.V. Machine Learning Approaches in Cardiovascular Imaging // *Circulation: Cardiovascular Imaging.* 2017. Vol. 10(10). P. e005614. DOI: doi.org/10.1161/CIRCIMAGING.117.005614
9. Johnson K.W., Shameer K., Glicksberg B.S. Enabling Precision Cardiology Through Multiscale Biology and Systems Medicine // *JACC: Basic to Translational Science.* 2017. Vol. 2(3). P. 311–237. DOI: doi.org/10.1016/j.jacbts.2016.11.010
10. Kagiya N., Shrestha S., Farjo P.D. Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease // *Journal of the American Heart Association.* 2019. Vol. 27(8). P. e012788. DOI: doi.org/10.1161/JAHA.119.012788
11. Kakadiaris I.A., Vrigkas M., Yen A.A. Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA // *Journal of the American Heart Association.* 2018. Vol. 7(22). P. e00947620. DOI: doi.org/10.1161/JAHA.118.009476
12. Krittanawong C., Zhang H., Wang Z. Review Topic of the Week Artificial Intelligence in Precision Cardiovascular Medicine // *Journal of the American College of Cardiology.* 2017. Vol. 69(21). P. 2657–2664. DOI: doi.org/10.1016/j.jacc.2017.03.571
13. Kwon J.M., Jeon K.H., Kim H.M. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction // *PLoS One.* 2019. Vol. 14(10). P. e0224502. DOI: [10.1371/journal.pone.0224502](https://doi.org/10.1371/journal.pone.0224502)
14. Mayer-Schönberger V. Big Data for cardiology: novel discovery? // *European Heart Journal.* 2016. Vol. 37(12). P. 996–1001. DOI: doi.org/10.1093/eurheartj/ehv648
15. Myers P.D., Scirica B.M., Stult C.M. Machine Learning Improves Risk Stratification After Acute Coronary Syndrome // *Scientific Reports.* 2017. Vol. 7(1). P. 1–12. DOI: doi.org/10.1038/s41598-017-12951-x

16. Pieszko K., Hiczkiewicz J., Budzianowski P. Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers // *Dis Markers*. 2019. ID 9056402. P. 1–9. DOI: 10.1155/2019/9056402
17. Rodriguez F., Scheinker D., Harrington R.A. Promise and Perils of Big Data and Artificial Intelligence in Clinical Medicine and Biomedical Research // *Circulation Research*. 2018. Vol. 123(12). P. 1282–1284. DOI: doi.org/10.1161/CIRCRESAHA.118.314119
18. Shah S.H., Arnett D., Houser S.R., Ginsburg G.S., MacRae C., Mital S., Loscalzo J., Hall J.L. Opportunities for the Cardiovascular Community in the Precision Medicine Initiative // *Circulation*. 2016. Vol. 133(2). P. 226–231. DOI: doi.org/10.1161/CIRCULATIONAHA.115.019475
19. Shah R.U., Rumsfeld J.S. Big Data in Cardiology // *European Heart Journal*. 2017. Vol. 38(24). P. 1865–1867. DOI: doi.org/10.1093/eurheartj/ehx284
20. Silverio A., Cavallo P., De Rosa R. Big Health Data and Cardiovascular Diseases: A Challenge for Research, an Opportunity for Clinical Care // *Front Med (Lausanne)*. 2019. Vol. 25(6). P. 1–10. DOI: 10.3389/fmed.2019.00036
21. VanHouten J.P., Starmer J.M., Lorenzi N.M. Machine learning for risk prediction of acute coronary syndrome // *AMIA Annu Symp Proc*. 2014. Vol. 2014. P. 1940–1949.
22. Wallert J., Tomasoni M., Madison G. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data // *BMC Med Inform Decis Mak*. 2017. Vol. 17(1). P. 1–11. DOI: 10.1186/s12911-017-0500-Y
23. Weintraub W.S., Fahed A.C., Rumsfeld J.S. Translational Medicine in the Era of Big Data and Machine Learning // *Circulation Research*. 2018. Vol. 123(11). P. 1202–1204. DOI: doi.org/10.1161/CIRCRESAHA.118.313944
24. Weng S.F., Reys J., Kai J. Can machine-learning improve cardiovascular risk prediction using routine clinical data? // *PLoS One*. 2017. Vol. 12(4). P. e0174944. DOI: 10.1371/journal.pone.0174944
25. Yan Y., Zhang J.W., Zang G.Y. The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine? // *J Geriatr Cardiol*. 2019. Vol. 16(8). P. 585–591. DOI: 10.11909/j.issn.1671-5411.2019.08.010

References

1. Ahmad T, Lund LH, Rao P. Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to Therapy in a Large Cohort of Heart Failure Patients. *J Am Heart Assoc*. 2018;7(8):e008081. DOI: 10.1161/JAHA.117.008081
2. Al'Aref SJ, Singh G, van Rosendael AR. Determinants of In-Hospital Mortality After Percutaneous Coronary Intervention: A Machine Learning Approach. *Journal of the American Heart Association*. 2019;8(5):1-9. DOI: https://doi.org/10.1161/JAHA.118.011160
3. Ambale-Venkatesh B, Yang X, Wu CO. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res*. 2017;121(9):1092-101. DOI: 10.1161/CIRCRESAHA.117.311312
4. Benjamins JW, van Leeuwen K, Hofstra L. A primer in artificial intelligence in cardiovascular medicine. *Neth Heart J*. 2019;27(9):392-402. DOI: 10.1007/s12471-019-1286-6
5. Blom MC, Ashfaq A, Sant'Anna A. Training machine learning models to predict 30-day mortality in patients discharged from the emergency department: a retrospective, population-based registry study. *BMJ Open*. 2019;9(8):e028015. DOI: 10.1136/bmjopen-2018-028015
6. Deo RC. Machine Learning in Medicine. *Circulation*. 2015;132(17):1920-30. DOI: doi.org/10.1161/CIRCULATIONAHA.115.001593
7. Duan H, Sun Z, Dong W. Utilizing dynamic treatment information for MACE prediction of acute coronary syndrome. *BMC Med Inform Decis Mak*. 2019;19(5):1-11. DOI: 10.1186/s12911-018-0730-7
8. Henglin M, Stein G, Hushcha PV. Machine Learning Approaches in Cardiovascular Imaging. *Circulation: Cardiovascular Imaging*. 2017;10(10):e005614. DOI: doi.org/10.1161/CIRCIMAGING.117.005614
9. Johnson KW, Shameer K, Glicksberg BS. Enabling Precision Cardiology Through Multiscale Biology and Systems Medicine. *JACC: Basic to Translational Science*. 2017;2(3):311-37. DOI: doi.org/10.1016/j.jacbts.2016.11.010
10. Kagiya N, Shrestha S, Farjo PD. Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. *Journal of the American Heart Association*. 2019;27(8):e012788 doi.org/10.1161/JAHA.119.012788
11. Kakadiaris IA, Vrigkas M, Yen AA. Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA. *Journal of the American Heart Association*. 2018;7(22):e00947620. DOI: doi.org/10.1161/JAHA.118.009476

12. Krittanawong C, Zhang H, Wang Z. Review Topic of the Week Artificial Intelligence in Precision Cardiovascular Medicine. *Journal of the American College of Cardiology*. 2017;69(21):2657-64. DOI: doi.org/10.1016/j.jacc.2017.03.571
13. Kwon JM, Jeon KH, Kim HM. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *PLoS One*. 2019;14(10):e0224502. DOI: 10.1371/journal.pone.0224502
14. Mayer-Schönberger V. Big Data for cardiology: novel discovery? *European Heart Journal*. 2016;37(12):996-1001. DOI: doi.org/10.1093/eurheartj/ehv648
15. Myers PD, Scirica BM, Stult CM. Machine Learning Improves Risk Stratification After Acute Coronary Syndrome. *Scientific Reports*. 2017;7(1):1-12. DOI: doi.org/10.1038/s41598-017-12951-x
16. Pieszko K, Hiczkiewicz J, Budzianowski P. Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers. *Dis Markers*. 2019;1-9. DOI: 10.1155/2019/9056402
17. Rodriguez F, Scheinker D, Harrington RA. Promise and Perils of Big Data and Artificial Intelligence in Clinical Medicine and Biomedical Research. *Circulation Research*. 2018;123(12):1282-4. DOI: doi.org/10.1161/CIRCRESAHA.118.314119
18. Shah SH, Arnett D, Houser SR, Ginsburg GS, MacRae C, Mital S, Loscalzo J, Hall JL. Opportunities for the Cardiovascular Community in the Precision Medicine Initiative. *Circulation*. 2016;133(2):226-31. DOI: doi.org/10.1161/CIRCULATIONAHA.115.019475
19. Shah RU, Rumsfeld JS. Big Data in Cardiology. *European Heart Journal*. 2017;38(24):1865-7. DOI: doi.org/10.1093/eurheartj/ehx284
20. Silverio A, Cavallo P, De Rosa R. Big Health Data and Cardiovascular Diseases: A Challenge for Research, an Opportunity for Clinical Care. *Front Med (Lausanne)*. 2019;25(6):1-10. DOI: 10.3389/fmed.2019.00036
21. VanHouten JP, Starmer JM, Lorenzi NM. Machine learning for risk prediction of acute coronary syndrome. *AMIA Annu Symp Proc*. 2014;2014:1940-9.
22. Wallert J, Tomasoni M, Madison G. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. *BMC Med Inform Decis Mak*. 2017;17(1):1-11. DOI: 10.1186/s12911-017-0500-Y
23. Weintraub WS, Fahed AC, Rumsfeld JS. Translational Medicine in the Era of Big Data and Machine Learning. *Circulation Research*. 2018;123(11):1202-4. DOI: doi.org/10.1161/CIRCRESAHA.118.313944
24. Weng SF, Reys J, Kai J. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12(4):e0174944. DOI: 10.1371/journal.pone.0174944
25. Yan Y, Zhang JW, Zang GY. The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine? *J Geriatr Cardiol*. 2019;16(8):585-91. DOI: 10.11909/j.issn.1671-5411.2019.08.010

Библиографическая ссылка:

Швец Д.А., Поветкин С.В. Сравнительный обзор использования методов машинного обучения для прогнозирования сердечно-сосудистого риска // Вестник новых медицинских технологий. Электронное периодическое издание. 2020. №5. Публикация 1-11. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2020-5/1-11.pdf> (дата обращения: 12.10.2020). DOI: 10.24411/2075-4094-2020-16711*

Bibliographic reference:

Shvets DA, Povetkin SV. Sravnitel'nyj obzor ispol'zovaniya metodov mashinnogo obuchenija dlja prognozirovaniya serdechno-sosudistogo riska [Comparative review of the use of machine learning methods for predicting cardiovascular risk]. *Journal of New Medical Technologies, e-edition*. 2020 [cited 2020 Oct 12];5 [about 9 p.]. Russian. Available from: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2020-5/1-11.pdf>. DOI: 10.24411/2075-4094-2020-16711

* номера страниц смотреть после выхода полной версии журнала: URL: <http://medtsu.tula.ru/VNMT/Bulletin/E2020-5/e2020-5.pdf>