

**КАРПОВ О.Э.,**

член-корреспондент РАН, д.м.н., профессор, ФГБУ «Национальный медико-хирургический Центр имени Н.И. Пирогова» Минздрава России, г. Москва, Россия, e-mail: nmhc\_director@mail.ru

**ХРАМОВ А.Е.**

д.ф.-м.н., профессор, АНО ВО «Университет Иннополис», Казань, Россия, e-mail: hramovae@gmail.com

## ПРОГНОСТИЧЕСКАЯ МЕДИЦИНА

DOI: 1025881/18110193\_2021\_3\_20

**Аннотация.**

*В этой статье мы даем краткий обзор точной прогностической медицины, которая использует современные методы медицинской диагностики, подходы молекулярной медицины и информационной технологии, включая искусственный интеллект и машинное обучение, для анализа медицинских данных в целях улучшения диагностики, прогноза и лечения заболеваний. Развитие точной прогностической медицины меняет традиционный медицинский подход к болезням и пациентам: от реактивной парадигмы реакции на болезнь происходит переход к предсказанию и превентивному лечению хронических заболеваний. Текущие открытия в медицине позволяют углубить понимание причин болезней и разработать новые методы лечения. Одновременно, широкое внедрение высокотехнологичных методов, таких как новые методы медицинской визуализации (например, МРТ, МЭГ, БИКС и др.) и информационные технологии для анализа медицинских и биологических данных, позволяет исследователям и клиницистам иметь новую клиническую и статистическую информацию, значимую для прогноза заболевания конкретного пациента. Все это, а также успехи молекулярной медицины, приводят к значительному росту качества прогноза и лечения заболеваний в кардиологии, онкологии, неврологии и т.д.*

**Ключевые слова:** прогностические модели в медицине, медицинская визуализация, биомаркеры, информационные технологии в медицине.

**Для цитирования:** Карпов О.Э., Храмов А.Е. Прогностическая медицина. Врач и информационные технологии. 2021; 3: 20-37. doi: 1025881/18110193\_2021\_3\_20.

**KARPOV O.E.,**

Corresponding Member of the RAS, DSc, Professor, Pirogov National Medical and Surgical Center, Moscow, Russia, e-mail: nmhc\_director@mail.ru

**HRAMOV A.E.,**

DSc, Professor, Innopolis University, Kazan, Russia, e-mail: hramovae@gmail.com

**PREDICTIVE MEDICINE**

DOI: 1025881/18110193\_2021\_3\_20

**Abstract.**

*In this article, we provide a brief overview of precision and predictive medicine, which use modern medical diagnostics, molecular medicine approaches, and information technology. This includes artificial intelligence and machine learning to analyse medical data in order to improve the diagnosis, prognosis, and treatment of diseases. The development of effective predictive medicine is changing the traditional medical approach to diseases and patients from the reactive paradigm of response to diseases, to the prediction and preventive treatment of chronic diseases. Modern discoveries in medicine are enabling a deeper understanding of the causes of diseases and the development of new therapies. At the same time, the widespread introduction of high-tech methods, such as new methods of medical imaging (for example, MRI, MEG, NIRS, etc.) and information technologies for the analysis of medical and biological data now allows researchers and clinicians to have new clinical and statistical information for prognosis. All this, as well as the advances in molecular medicine, leads to a significant increase in the quality of prognosis and treatment of diseases in cardiology, oncology, neurology, etc.*

**Keywords:** prognostic model in medicine, medical visualization, biomarkers, information technologies in medicine

**How to cite:** Karpov O.E., Hramov A.E. Predictive Medicine. Medical doctor and information technology. 2021; 3: 20-37. (In Russ.). doi: 1025881/18110193\_2021\_3\_20.

## 1. ОСНОВНЫЕ ПОНЯТИЯ ТОЧНОЙ ПРОГНОСТИЧЕСКОЙ МЕДИЦИНЫ

Уже самые первые «врачеватели», медицинские трактаты которых дошли до нас, задумывались не только о лечении, но и прогностике заболеваний. Так, в папирусах из Рамессумы, датированных периодом 1850 года до н.э., представлены как рациональные, так и магические приемы врачевания, среди которых описываются методы прогнозирования течения беременности. Самым выдающимся врачом античности — Гиппократом (460–356 годы до н. э.) — было создано учение о прогностике, которое преследовало цель не только предсказать исход болезни, но определяло тактику лечения больного. Им были впервые описаны признаки болезни, таких как лицо умирающего больного: «нос острый, глаза впалые, виски вдавленные, уши холодные и стянутые, мочки ушей отверженные, кожа на лбу твердая, натянутая и сухая и цвет лица землистый, черный или бледный, или свинцовый и др.». Вместе с тем, исследователи в области медицины долгое время основное внимание уделяли лечению конкретных заболеваний, когда разрабатывались и совершенствовались терапевтические методы как ответ на наличие конкретных симптомов той или иной болезни.

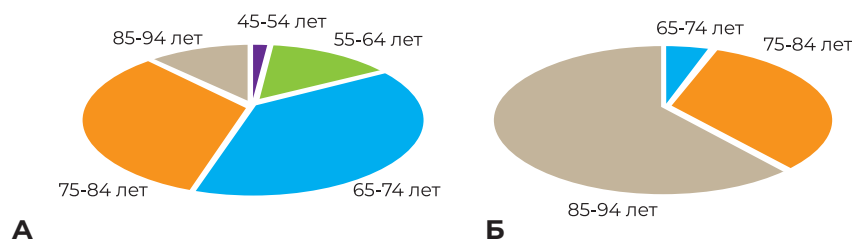
Современная медицина все чаще ставит вопросы не только лечения конкретных болезней, но и прогноза течения и терапии хронических заболеваний, возникающих из-за как излечимых, так и неизлечимых болезней [1]. Такая постановка вопроса требует разработку эффективных методов выявления принадлежности конкретного индивидуума к группам риска по различным хроническим заболеваниям в целях проведения адекватных профилактических и поддерживающих мероприятий для недопущения развития заболевания или его протекания в легкой форме. Особенно актуальным это становится в последнее время, когда значительный рост уровня жизни, правильная диета, психологическая установка на заботу о здоровье, профилактические мероприятия и успехи медицины в лечении текущих заболеваний привели к существенному росту продолжительности жизни. Как следствие, становится актуальным лечение таких болезней, которые еще полвека назад были

редкими, а теперь становятся все более распространенными среди пожилых пациентов.

В качестве примера можно упомянуть нейродегенеративные заболевания, риск которых резко увеличивается с возрастом пациента. В частности, от болезни Паркинсона страдает около 300 больных на 100 000 человек. Большинству людей ставят диагноз в возрасте старше 70 лет, только 15% случаев происходят среди людей моложе 50 лет. Последние оценки показывают, что число людей, живущих с болезнью Паркинсона, неуклонно увеличивается с возрастом [2]. Это иллюстрирует рис. 1(а), на котором показана возрастная структура пациентов с диагнозом болезнь Паркинсона [3]. Из представленных данных следует, что бремя болезни Паркинсона будет увеличиваться во многих странах по мере старения населения. Аналогичная ситуация имеет место для болезни Альцгеймера, как видно из рис. 1(б), на котором показана возрастная структура пожилых американских пациентов с возрастом более 65 лет, страдающих от болезни Альцгеймера [4]. Каждый десятый (10%) человек в США в возрасте 65 лет и старше страдает болезнью Альцгеймера. Процент людей с деменцией Альцгеймера увеличивается с возрастом: 3% людей в возрасте 65–74 лет, 17% людей в возрасте 75–84 лет и 32% людей в возрасте 85 лет и старше страдают болезнью Альцгеймера. У людей моложе 65 лет болезнь Альцгеймера также может развиваться, но она встречается гораздо реже, и ее точная частота неизвестна.

До сих пор не существует достоверных тестов для ранней диагностики данных болезней, а также методик раннего выявления тех или иных биомаркеров для отнесения пациента к группе риска болезни Паркинсона и/или Альцгеймера. Основываясь на этом примере, можно сделать вывод, что перед медициной как научным направлением стоит важная и актуальная задача повышения качества прогнозирования факта того или иного заболевания, а также характера течения хронических болезней.

Современное развитие медицины связано с планомерным переходом от популяционной медицины, придерживающейся чисто биологической точки зрения, которую можно также рассматривать как традиционный подход анализа «одной клетки или одного белка» как источника проблемы со здоровьем, к прогностической



**Рисунок 1 — Возрастная структура пациентов с диагнозами распространенных в мире нейродегенеративных заболеваний: (а) болезни Паркинсона (на основе данных из [3]) и (б) болезни Альцгеймера (на основе данных из [4]).**

точной медицине, подразумевающей целостный подход к диагностике и лечению болезни и пациента, который определяется доступностью самых передовых технологий лечения и всестороннего знания о пациенте и заболевании в различных контекстах [5]. Наиболее последовательным сторонником такой точки зрения выступал профессор Лерой Худ, пионер в области системной биологии, который полагал, что будущая медицина станет более персонализированной, более точной с точки зрения прогноза течения заболевания и направленной на профилактику болезней [6,7]. Данная концепция в настоящее время оказывает существенное влияние на целый ряд направлений современной медицины, а также трансформирует различные подходы, связанные с модернизацией здравоохранения, в частности, с его цифровизацией [8–10]. Последнее позволяет учесть многие факторы, которые невозможно было анализировать ранее: образ жизни пациента, его генетику, разнообразные биомедицинские данные и т.д., которые теперь возможно накапливать в удобной форме с использованием баз данных и всесторонне анализировать с использованием методов искусственного интеллекта, оказывая врачу помощь в диагностике и выборе метода лечения пациента.

Медицина XXI века базируется на значительном прогрессе современной биологической и медицинской науки, которые формируют новые качества, обуславливающие возрастающее значение точной прогностической медицины.

1. **Профилактическое** свойство медицины предполагает способность превентивно, еще до появления симптомов, бороться с болезнями. Зная молекулярную картину пациента

и используя системный подход к анализу его состоянию, можно предвидеть заболевания, а также рецидивы или другие изменения состояния его здоровья.

2. **Персонализация** медицины была первоначально предложена как попытка использования достижений генетики с помощью применения наногеномики и нанопротеомики для адаптации медицинского вмешательства к конкретной молекулярной картине отдельного пациента [11; 12]. Однако, достаточно быстро данная концепция претерпела существенные изменения в сторону рассмотрения индивидуального пациента в целом, так что «адаптация медицинского вмешательства» означает учет индивидуальных способностей и потребностей пациента, его психофизиологического и когнитивного статуса [13; 14].
3. Современная медицина становится все более и более **прогнозирующей**, что означает, что она будет использовать информацию, полученную из последовательностей генома и молекулярных, клеточных и фенотипических измерений, чтобы обеспечить базовые значения, которые можно определить как здоровье и благополучие индивидуума, а затем использовать это «базовое» состояние для определения последующих переходов к заболеванию. Следует подчеркнуть, что прогностическая медицина обязательно основана на фактических биомедицинских данных [15; 16], что обычно характеризуется использованием различных ассистирующих врачу технологий — обработки больших данных с использованием метода искусственного интеллекта [17], математических моделей с

рядом неизвестных переменных для предсказания будущих событий пациента, связанных со здоровьем и т.д. [18]. Обычно такие решения, включающие модули интеллектуальной обработки данных и базы данных, объединяются под одной программной оболочкой системы поддержки принятия решений врачом (СППВР) [19].

Итак, необходимость разработки методов «прогнозирования здоровья», безусловно, является серьезной задачей, стоящей перед биомедицинскими исследованиями и клинической медициной. Для ее решения необходимо учитывать не просто диагностическую классификацию здоровья пациента, но также психологический и когнитивный профиль пациента, его образ жизни и ту среду, в которой он находится. В этом смысле прогностическая медицина ставит вопрос не только о медицинской диагностике пациента, но также оценке его состояния с использованием психометрических инструментов, которое включает когнитивные и психические аспекты.

Таким образом, основной целью прогностической медицины является в случае здоровых людей отнесение пациента в группу риска того или иного заболевания, а в случае больных — предсказание развития заболевания на его ранних и последующих стадиях. В рамках первого направления прогностическую медицину можно условно назвать «медициной здоровых людей», когда отнесение человека к группе риска позволяет ему изменить образ жизни и/или повысить частоту профилактических обследований, например, прохождение маммографии каждые 6 месяцев в случае, если пациент попадает в группу риска развития рака молочной железы. Такой подход требует, чтобы медицинские работники поддерживали долгосрочные отношения с пациентом, который должен научиться управлять симптомами, организовывать процесс лечения (например, регулярно принимать лекарства или проходить профилактическое обследование), поддерживать рекомендуемый образ жизни, а также справляться со стрессом и негативными эмоциями. Информационные технологии могут помочь врачу с мониторингом психологического состояния пациента, например, путем изучения, конечно с согласия пациента, его «цифрового следа» в социальных сетях

для выявления негативных или позитивных моментов его психического состояния.

Как уже отмечалось выше, прогностическая медицина тесно связана с персонализированной медициной, в основе которой изначально лежит известная связь между специфическими генетическими вариантами в генетическом коде и конкретным заболеванием. Например, женщины с мутацией в BRCA1 гене имеют 65% совокупный риск рака молочной железы [20; 21]. Другой пример — мутации в гене для фактора V предполагают повышение вероятности тромбоза глубоких вен [22]. Однако, в настоящее время, кроме генетического тестирования прогностическая медицина использует широкий спектр инструментов для прогнозирования здоровья и болезни, в том числе успехи IT-технологий, которые в настоящее время становятся одним из основных инструментов прогноза в различных, в том числе медицинских, сферах. Внедрение информационных технологий позволяет создать экосистему данных, которая объединяет биологическую информацию, клинические фенотипы и анамнез пациента благодаря визуализации, лабораторным тестам и медицинским картам, для многофакторного прогнозирования попадания пациента в ту или иную группу риска, идентифицировать и лечить заболевание, учитывая индивидуальные особенности человека, сокращая финансовые и временные усилия и улучшая качество жизни как здоровых людей, так и больных. Можно сказать, что предсказательная медицина изменяет парадигму традиционной медицины с реактивной на активную: не реакция на уже имеющееся заболевание, а предсказание и превентивное лечение хронических заболеваний. Как следствие, предсказательная медицина имеет значительный потенциал, чтобы значительно продлить продолжительность здорового состояния и, как следствие, общей продолжительности жизни путем снижения числа и остроты заболеваний в разных возрастных стратах общества.

## 2. БИОЛОГИЧЕСКИЕ МАРКЕРЫ ЗАБОЛЕВАНИЙ

В сфере здравоохранения точная прогностическая медицина является относительно новой областью. Особенности прогностической медицины можно определить как использование

лабораторных и генетических тестов для прогнозирования возникновения хронического заболевания у здорового индивидуума или ухудшения/улучшения состояния пациента для оценки развития заболевания, в частности, для выбора наиболее эффективной стратегии лечения [23; 24]. В данном контексте особое значение для отнесения индивидуума к группе риска, предсказания начала заболевания, прогноза и результата медицинской терапии приобретают биологические маркеры (*биомаркеры*).

Биомаркер указывает на медицинский (биологический) признак, который можно измерить объективно, точно и воспроизводимо и который можно использовать в качестве индикатора состояния всего организма. Впервые понятие биомаркера было введено в 2001 году Национальным институтом здоровья США. Оно формулировалось как объективная характеристика, которая фиксирует те процессы, которые происходят в клетке или организме в момент проведения исследования. Например, высокий уровень свинца в крови может указывать на необходимость проверки нервной системы и когнитивных расстройств, особенно у детей. Высокий уровень холестерина является распространенным биомаркером риска сердечных заболеваний. Всемирная организация здравоохранения определила биомаркер как «практически любое измерение, отражающее взаимодействие между биологической системой и потенциальной опасностью, которая может быть химической, физической или биологической. Измеренный ответ может быть функциональным и физиологическим, биохимическим на клеточном уровне или молекулярным взаимодействием» [25].

Биомаркеры широко используются для разработки лекарств и оценки клинических результатов. Нынешний подход к клиническим испытаниям, основанный на методах доказательной медицины, предполагает в идеале универсальность того или иного метода лечения для всех испытуемых (англ. “one-size-fits-all”, т.е. эффект лечения одинаков для всей выборки). Однако, будущее медицины видится в том, чтобы обеспечить наиболее оптимальное лечение в нужное время для конкретного пациента, определяя различные подгруппы в зависимости от определенных биомаркеров, которые отвечают на оптимальную терапию [26].

Роль биомаркеров для прогнозирования возникновения заболевания у здоровых индивидуумов важна, поскольку единственным объективным критерием наличия ранних изменений в организме в случае отсутствия клинических проявлений может быть только объективно измеряемые характеристики организма. Подобные биомаркеры могут быть очень разнообразными и включать в себя измерения как в пассивном состоянии организма, например, общий анализ крови, так и в активном в виде ответа организма на некоторые воздействия. В ряде случаев последний подход оказывается более информативным, так как позволяет оценить ответ той или иной функциональной системы организма человека на некоторое специфическое воздействие. В качестве примера можно привести функциональные пробы с фотостимуляцией, гипервентиляцией, коразолом, депривацией сна и т.д. при ЭЭГ исследованиях для провоцирования эпилептических разрядов, что позволяет сократить длительность рутинных наблюдений и измерений.

### **3. ПРОГНОСТИЧЕСКАЯ МЕДИЦИНА И ИНФОРМАЦИОННАЯ РЕВОЛЮЦИЯ В ОБРАБОТКЕ БОЛЬШИХ ДАННЫХ**

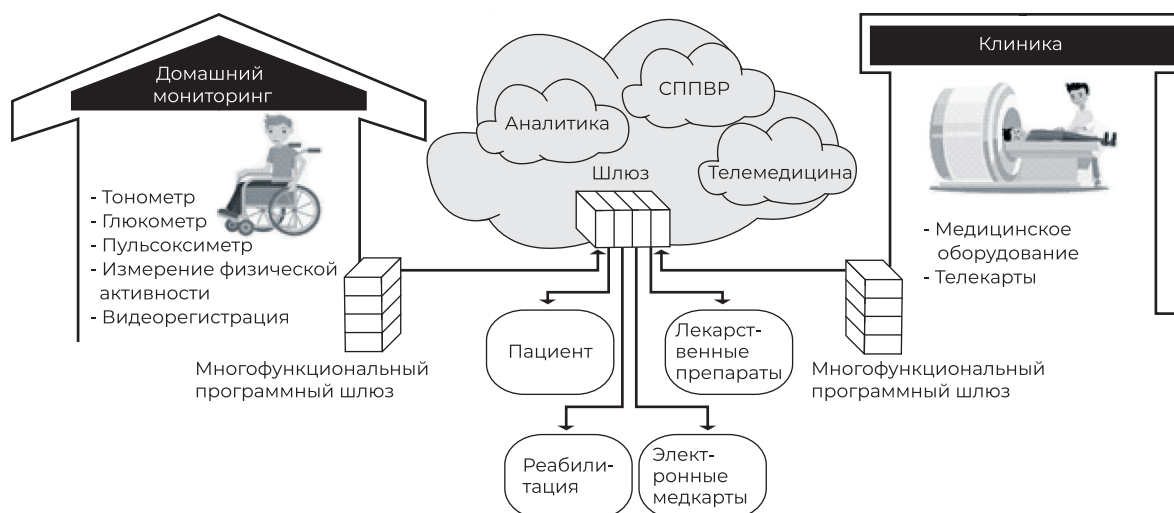
Итак, прогностическая медицина является важным направлением современной медицины, в основе которой лежит новая концепция и новые приложения для здравоохранения [1]. Если персонализированная медицина учитывает геном пациента, а также его фенотип, психофизиологические показатели и социальный контекст, прогностическая медицина является подходом, который в значительной степени основан на больших данных, аналитике и информации. Фактически, прогностическая медицина формирует для своей эффективной реализации обширную «экосистему», которая включает пациентов, клиницистов, исследователей, технологии, генетику и анализ больших данных [27]. Для реализации методов прогностической медицины крайне важно определять биомаркеры заболеваний, используя методы молекулярных исследований (геномные, протеомные, эпигенетические и др. исследования), функциональных проб и др. либо отдельно, либо в сочетании с информацией об окружающей среде и образе жизни пациента [28] с целью формирования

прогностических данных для диагностических и терапевтических стратегий, основанных на индивидуальных особенностях и потребностях пациента [29].

Концепция прогностической медицины практически невозможна без медицинских клиник и исследовательских институтов, которые собирают и накапливают медицинскую информацию в виде биомедицинских изображений (рентгеновские снимки, МРТ, ПЭТ и т.д.), сигналов (ЭКГ, ЭЭГ, МЭГ, BOLD-сигналы и т.д.) и данных (результаты биохимических анализов, тестов, геномных исследований). Однако, обработка накопленных больших данных и поиск в них закономерностей невозможна без современных компьютерных аналитических методов, среди которых в последнее время особую роль и значение приобрели технологии искусственного интеллекта и машинного обучения [30]. Эти информационные технологии, которые вызвали в последние несколько лет «революцию больших данных» в различных отраслях экономики и социальной жизни, позволяют обработать огромные объемы различных видов медицинской информации [31–34]. Отметим, что методы машинного обучения в отличие от традиционных (например, статистических) методов анализа имеют ряд преимуществ. Во-первых, в этом случае особенности рассматриваемого явления анализируются всесторонним образом, существует возможность выявления как линейных, так и нелинейных закономерностей. Во-вторых, методы машинного обучения, такие как различные типы искусственных нейронных сетей, классификаторов, деревьев принятия решений и т.д. [35], предпочтительнее для всестороннего нелинейного анализа данных по сравнению с традиционными, например, статистическими методами. В-третьих, анализ больших данных позволяет не только отвечать на ранее поставленные вопросы (т.е. подтверждать уже высказанные гипотезы), но и сформулировать новые гипотезы и/или устанавливать новые закономерности [36].

Следует также отметить, что прогресс в накоплении больших объемов медицинских данных был бы невозможен без широкого внедрения различных подходов к цифровой трансформации здравоохранения. В настоящее время мониторинг физиологических параметров пациента с использованием технологий интернета вещей

становится стандартным подходом в здравоохранении, способствуя индивидуальному медицинскому обслуживанию для любого человека [37]. Все чаще для обработки такого повседневного мониторинга применяются методы машинного обучения, которые наиболее часто используются для распознавания активности и анализа движений тела пациентов, например, для оценки правильности реабилитационных процедур [38]. У медицинского интернета вещей (ИВ) большие перспективы в современном цифровом здравоохранении, и в настоящее время наблюдается устойчивый рост применений ИВ в медицинской отрасли. В настоящий момент создается широкий спектр приложений и устройств ИВ, специально разработанных для нужд и настроек здравоохранения, таких как датчики и приложения для удаленного мониторинга, консультаций и дистанционного оказания медицинских услуг. Более продвинутые варианты использования ИВ могут включать в себя облачные технологии, создавая единую среду, с помощью которой осуществляется мониторинг критически важных медицинских устройств и оборудования с возможностью получения предупреждений, когда они требуют обслуживания или замены. Все устройства и решения ИВ в сфере медицины можно разделить на два типа: (1) предназначенные для больниц и специалистов, которые в них работают; (2) ориентированные на пациента и предназначенные для использования в домашних условиях. Такие «умные» устройства (например, умные гаджеты, датчики, измерители сердечного ритма), подключенные к облачным хранилищам данных, собирают и обрабатывают биомедицинскую информацию — общие показатели здоровья, аллергические реакции, результаты анализов, уровень физической активности пациента. Устройства ИВ помогают медицинскому персоналу оперативно формировать персональную статистику пациентов, их медицинские карты. При этом врачи могут применять информацию из подключенных систем, чтобы эффективнее лечить пациентов, оказывать им своевременный и соответствующий уход, а также предупреждать обострения хронических заболеваний. Схема использования технологий ИВ, взаимодействующих с облачным хранилищем данных и интеллектуальной системой управления, представлена на рис. 2.



**Рисунок 2 — Схема использования технологий интернета вещей в прогностической медицине.**

Следует также отметить, что огромную роль при накоплении биомедицинских данных и дальнейшем их исследовании играют открытые репозитории и базы данных с соответствующей обезличенной биомедицинской информацией. Подобные медицинские информационные базы данных позволяют исследователям из различных областей применять и апробировать свои методы на реальных данных. Среди таковых можно назвать международный портал электрофизиологии эпилепсии (<http://ieeg.org>), который является облачной открытой платформой для исследователей эпилепсии. Эта информационная платформа была первоначально разработана для совместного использования больших наборов инвазивных внутречерепных записей, но позже была доработана для включения доклинических данных и дополнительных данных других модальностей (то есть изображений, снимков МРТ, документов и т.д.). Портал содержит более 1200 наборов данных неинвазивной и инвазивной ЭЭГ, полученных как на животных моделях эпилепсии, так и на пациентах. Этим порталом пользуется 500 исследователей из более чем 30 стран [39]. Он обеспечивает основу для разработки инструментов для хранения, обработки и анализа больших данных в виде экспериментальных материалов по эпилепсии. Ключевые исследователи в области эпилепсии

делятся своими данными на этой платформе и имеют возможность устанавливать научные коллаборации и использовать данные своих коллег. Другой пример общедоступной базы данных по эпилепсии — это европейский проект EPILEPSIAE (<http://www.epilepsiae.eu/>), который содержит долговременные записи ЭЭГ от 275 пациентов с эпилепсией [40]. Следует приветствовать и активно входящую в норму научной жизни рекомендацию ведущих биологических и медицинских журналов на одновременную публикацию статьи вместе с экспериментальными данными, на основе которых были выполнены исследования. Это служит повышению воспроизводимости публикуемых результатов в области биомедицины и физиологии и пополняет глобальную базу данных открытых материалов.

Таким образом, развитие прогностической медицины невозможно без информационных технологий, которые являются инструментом использования биомедицинских больших данных для решения фундаментальных научных вопросов, создания систем поддержки принятия врачебных решений и прогноза течения заболеваний. С одной стороны, современные IT-технологии, которые позволяют эффективно управлять большими данными, лежат в основе систем обучения, предоставляют индивидуальный доступ к данным, поддерживают трансляционные



исследования и т.д. [41]. С другой стороны, клиническая информатика имеет решающее значение для прогностической медицины, предоставляя клиницистам инструменты, способные дать информацию об индивидууме, подверженном риску, начале заболевания и способах его наиболее эффективного лечения с учетом анамнеза пациента, его образа жизни, групп риска и т.д. Ярким примером важности информационных технологий в медицине является активный переход на использование медицинских карт. В Соединенных Штатах использование электронных медицинских карт согласно различным источникам [42–44] 42 выросло с 12% до 90–98% среди врачей с 2007 по 2018 год. Одной из крупнейших электронных баз данных в США является централизованная электронная медицинская карта (Centricity Electronic Medical Record, CEMR), которая включает обезличенные клинические данные на уровне пациентов из независимых практик врачей, академических медицинских центров, больниц и крупных интегрированных сетей доставки [45]. При среднем сроке наблюдения 4,5 года исследовательская база данных CEMR охватывает более 35 000 поставщиков медицинских услуг из всех штатов США, из которых примерно 70% являются поставщиками первичной медицинской помощи. Лонгитюдные электронные медицинские записи доступны для более чем 45 миллионов человек с 1995 по сентябрь 2018 года с исчерпывающей информацией на уровне пациентов о демографических характеристиках, антропометрических показателях, событиях болезни, лекарствах, а также клинических и лабораторных показателях [46]. База данных CEMR широко используется в академических исследованиях [47–49]. Очевидно, что использование подобных технологий будет способствовать активному внедрению элементов технологий прогностической медицины в рутинную клиническую практику.

#### **4. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В ПРОГНОСТИЧЕСКОЙ МЕДИЦИНЕ**

В настоящее время методы интеллектуального анализа больших медицинских данных обычно применяются в клинических условиях для анализа ретроспективных данных, что дает исследователям возможность использовать

большие объемы данных, регулярно собираемых во время рутинной медицинской деятельности. Более того, в настоящее время клиницисты могут использовать методы интеллектуального анализа биомедицинских данных, чтобы учитывать их при постановке диагноза и определении оптимального метода лечения. Успехи молекулярной медицины способствуют переходу от популяционной медицины к персонализированной, в частности, путем выявления генетических или геномных сигнатур заболеваний. В настоящее время задача заключается в использовании интеллектуального анализа данных для создания моделей, способных учитывать динамический временной характер клинической помощи, а также использовать разнообразную информацию, собираемую в реальном времени у пациента. Центральной задачей здесь является разработка систем поддержки принятия решений, в которых были бы реализованы возможности использования биомедицинских данных различной модальности и интеллектуальной системы для анализа этих данных.

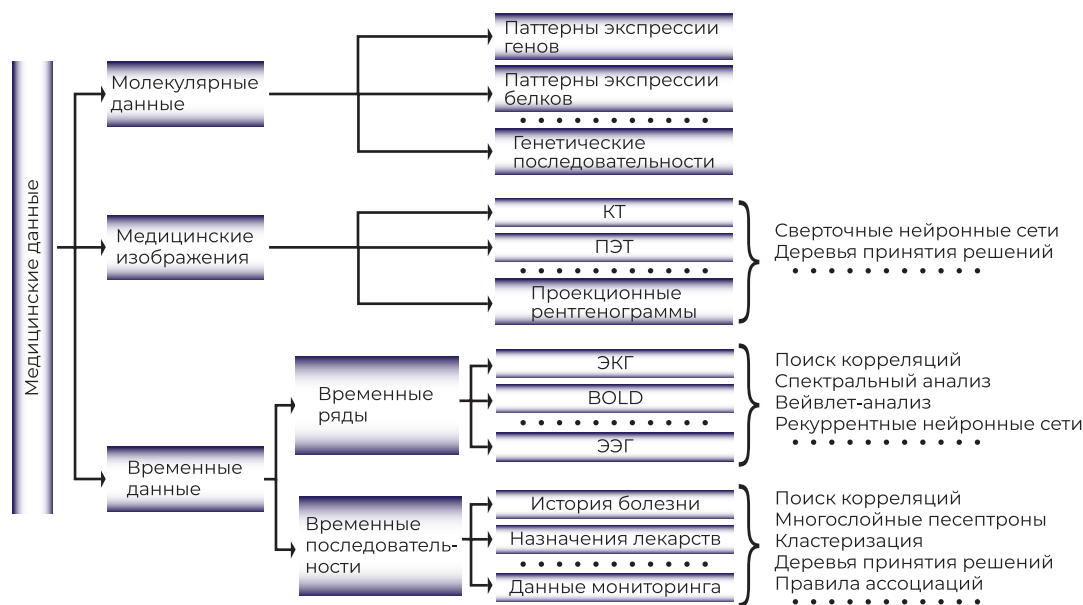
Следует отметить, что различные интеллектуальные системы прогноза в медицине могут варьироваться от простой стратификации пациентов на основе известных факторов риска, таких как возраст или образ жизни, до прогноза того, какое воздействие лечение или лекарственное средство может оказать на данного пациента. В клиническом контексте прогнозы могут поддерживать диагностические, терапевтические или контрольные задачи. *Диагноз* связан с классификацией пациентов по классам или подклассам заболеваний на основе различных данных пациента и охватывает широкий спектр клинических случаев, например, сортировку в отделениях неотложной помощи больницы в зависимости от тяжести их состояния. *Терапевтический прогноз* связан с выбором наиболее подходящего лечения для пациента. Этот вид прогноза очень распространен в клинических условиях, как для планирования лекарственной терапии, так и для прогноза хирургических вмешательств. В качестве примера можно привести назначение лечения пациентам с различными злокачественными опухолями, которые могут потребовать разных терапевтических стратегий лечения. Наконец, *прогнозы в клиническом мониторинге* имеют важное значение в нескольких

контекстах, например, в отделениях интенсивной терапии, в которых прогноз постоянно обновляется на основе данных мониторинга. В последнем случае особую роль приобретает анализ в реальном времени (on-line режим) данных и временных рядов состояния пациента. Например, данные непрерывного мониторинга кардиоваскулярной системы становятся основным материалом для прогноза потребности в экстренной помощи при сердечно-сосудистых заболеваниях [50; 51]. В последние годы «классические» задачи прогнозирования, связанные с клинической деятельностью, получили новые решения, основанные на данных из молекулярной медицины и различных методов диагностики состояния пациента. Благодаря этим данным теперь возможно создавать модели на основе очень большого набора прогностически ценных физиологических или биологических переменных — биомаркеров, которые выступают как индикаторы состояния пациента. Одновременно, число таких переменных, характеризующих пациента, становится слишком много и врачу нужна система поддержки и помощи принятия решения, чтобы корректно учесть все факторы.

Необходимо отметить, что можно говорить о различных стратегиях выделения биомаркеров

в зависимости от модальности анализируемых биомедицинских данных (Рис. 3). Среди них особое внимание нужно уделить трем типам данных — молекулярным данным [52; 53], изображениям [54] и временным данным [55].

*Молекулярная диагностика* играет значительную роль в персонализированной медицине, выделяя с использованием инструментом молекулярной биологии биомаркеры в геноме и протеоме, а также анализируя как клетки человеческого организма выражают свои гены в виде белков. Обнаружение тех или иных генетических или геномных сигнатур используется для диагностики и мониторинга заболевания, выявления риска и принятия решения о том, какие методы лечения лучше всего подходят для отдельных пациентов. Молекулярная диагностика использует такие методы, как масс-спектрометрия и генные чипы, для определения паттернов экспрессии генов и белков. В качестве исходных биоматериалов наиболее часто используются образцы крови или тканей пациента, с которыми проводятся биологические анализы *in vitro*, такие как ПЦР-ИФА или флуоресцентная гибридизация *in situ*. В результате таких анализов выявляются молекулы (часто присутствующие в образцах в низких концентрациях), которые



**Рисунок 3 — Модальности медицинских данных, используемых прогностической медициной.**

являются маркером заболевания или риска в образце, взятом у пациента.

В настоящее время наиболее перспективными *инструментами визуализации* в медицине являются методы микроскопии, ультразвуковых исследований, рентгенографии, компьютерной томографии (КТ), магнитно-резонансной томографии (МРТ), позитронно-эмиссионной томографии (ПЭТ) и т.д. Достижения в области биомедицинской инженерии постоянно улучшают пространственное и временное разрешение таких изображений.

Весьма перспективным подходом к анализу биомедицинских изображений является применение методов искусственного интеллекта, который представляет собой попытку смоделировать некоторые функции человеческого сознания. Наиболее востребованным методом искусственного интеллекта в контексте анализа данных и прогнозирования является машинное обучение, которое можно определить как формирование элементов интеллекта без предварительного программирования за счет того или иного обучения моделей. Например, в настоящее время машинное обучение начинает находить все более широкое применение для анализа рентгеновских изображений [54; 56; 57] в целях компьютерной диагностики, например, для обнаружения тех или иных особенностей, которые были бы полезны при хирургическом вмешательстве. В радиологии использование методов искусственного интеллекта позволяет проводить автоматическую классификацию снимков и локализовать область патологии. Одновременно методы машинного обучения могут быть использованы для облегчения хирургического вмешательства для предварительной сегментации изображений [56]. В пандемии коронавирусной инфекции большую роль для первичной диагностики начинают играть системы интеллектуального анализа КТ/рентгеновских снимков легких пациентов с подозрением на COVID-19 [58].

Под методами машинного обучения понимаются различные математические модели, которые могут обучаться выделять определенные характерные свойства данных, классифицировать данные и прогнозировать дальнейшую динамику. Для решения подобных задач наиболее универсальным инструментом являются

искусственные нейронные сети (ИНС) [59], которые представляют собой набор входных нейронов, на которые подаются анализируемые данные, скрытые слои с искусственными нейронами, которые активируются через соединения между нейронами слоев, а также выходные нейроны, результаты активации которых и дают решение задачи классификации и/или прогноза. Такие ИНС при большом числе нейронов в скрытых слоях представляют собой сложные высокопроизводительные вычислительные системы, которые после соответствующего обучения, заключающегося в настройке связей и параметров самих нейронов, обеспечивают эффективное решение задач классификации и прогноза.

Наиболее распространенным методом обучения для классификации медицинских изображений является обучение с учителем. Оно подразумевает, что у нас есть обучающая выборка изображений, которая уже размечена опытным специалистом-клиницистом, и на которой происходит настройка вычислительной интеллектуальной системы для получения близкой к 100% точности классификации. В качестве таких данных для обучения могут выступать собранные ранее данные большого числа пациентов и размещенные в открытых репозиториях, о которых мы говорили в предыдущем разделе. Далее ожидается, что все новые изображения будут также классифицироваться обученной ИНС с высокой степенью точности. Вместе с тем, на этом пути может возникнуть проблема переобучения, которая требует специальных подходов к ее решению [60]. Обычно для классификации рентгеновских снимков и изображений КТ или МРТ наиболее часто применяется такая разновидность ИНС как сверточная нейронная сеть (СНН) [59].

Менее часто в медицине применяются методы обучения ИНС без учителя в силу неопределенности конечного результата такой обработки данных. Вместе с тем для работы с медицинскими данными широко применяются другие методы машинного обучения для кластеризации, то есть для разделения данных на группы, например, метод k-средних или иерархические методы для выделения особенностей в группе испытуемых [61].

Следует отметить значительную сложность применения методов искусственного интеллекта в медицинских диагностических приложениях,

которые связаны с тем, что ИНС представляет собой черный ящик без четкого понимания тех правил, которые они вырабатывают при обучении и на которых основаны их высокая эффективность классификации и прогноза. Это затрудняет сертификацию и использование подобных интеллектуальных систем. Поэтому в последнее время для интерпретации и классификации данных все чаще применяются деревья решений, поскольку они характеризуются простыми и прозрачными правилами принятия решений. Деревья решений — это способ представления правил в иерархической последовательной логической структуре, который позволяет соотнести набор данных на входе дерева с одним или несколькими выходными узлами. Под правилом здесь понимается логическая конструкция, представленная в виде «если [логическое условие]..., то [действие]...». Такая модификация деревьев решений, как случайный лес (англ. random forest), находит широкое применение при анализе данных молекулярной медицины [62; 63].

Другая интересная и важная модальность данных, используемых прогностической медициной — это временные данные [55]. Следует отметить, что реальные клинические сценарии часто характеризуются разнообразными стратегиями сбора данных, которые приводят к необходимости работы с различными типами временных данных. С одной стороны, имеются временные ряды внутрибольничного мониторинга. К ним относятся, например, физиологические сигналы, отслеживаемые с помощью электронных устройств или специализированных датчиков (например, непрерывных мониторов уровня глюкозы в крови) у постели больного, или длительные записи ЭЭГ при мониторинге эпилепсии. С другой стороны, собираются данные периодических осмотров или лабораторных анализов. Это, например, случай больных раком, которые проходят периодические обследования, во время которых оцениваются определенные биомаркеры. Во втором случае обычно формируется набор данных, содержащий короткие временные ряды (до 10–50 временных точек), которые собираются нерегулярно из-за неравномерного графика посещений или проведенных измерений. Это приводит к созданию выборки отсчетов, которая варьируется как по периоду наблюдения конкретного пациента, так и от пациента к пациенту.

Как следствие, такие особенности данных не позволяют для клинических целей использовать весь спектр традиционных методов, развитых для анализа временных рядов, например, спектральный или вейвлет-анализ.

Наряду с временными рядами биомедицинских сигналов и данных важным источником информации о прогнозе состояния пациента могут стать «административные» данные [55; 64]. Административные записи содержат данные, которые также отображают историю болезни пациента: рецепты на лекарства, госпитализации, амбулаторные посещения, дневные стационары и т.д. Такие данные обычно собираются для административных целей и, как правило, не содержат клинических значений (например, известно, что пациент поступил в больницу с определенным диагнозом и что были проведены некоторые лабораторные анализы, но конкретные результаты тестов из административных данных неизвестны). Базы данных, содержащие административную информацию, зарегистрированную, например, органами здравоохранения или страховыми компаниями, до настоящего времени изучались весьма мало в силу своей плохой формализации и из-за огромного количества сложных разнородных данных, хранящихся в них. Вместе с тем, в настоящее время в связи с развитием технологий искусственного интеллекта, который все более эффективно позволяет формализовать данные, записанные на естественном языке, распознавать рукописные тексты и т.д., открываются новые перспективы вовлечь эти данные в контексте исследований прогностической медицины.

Конечной целью исследования временных данных применительно к клинической области является детектирование соответствующих временных структур (паттернов), то есть некоторую последовательность событий, которые клинически значимы для прогноза конкретного состояния пациента. Есть принципиальная разница между двумя типами временных данных — временными рядами и административными данными. Первые представляют собой значения некоторой переменной через равные промежутки времени (например, отсчеты сигнала ЭКГ или МЭГ), а вторые — это последовательность событий без четкого постоянного временного интервала между ними.

Биомедицинские временные ряды обычно требуют предварительной обработки для дальнейшего использования в задачах прогноза. Часто необходимо выделить основные свойства и особенности таких сигналов, характеризующих важное для нас клиническое явление. Например, применение нелинейного анализа на основе выделения главных компонент совместно с ИНН позволяет детектировать ишемию на сигналах ЭКГ [65]. При анализе физиологических рядов, таких как ЭКГ, ЭЭГ, МЭГ, широко применяется вейвлет-анализ, который имеет большие прогностические возможности в целом ряде биомедицинских задач [66–68].

Другая ситуация определяет последовательность временных событий, которая может быть определена как список событий, где каждое событие связано с одним и тем же лицом. В клинических приложениях последовательность событий может представлять собой, например, историю болезни пациента с точки зрения госпитализации, назначения лекарств и лабораторных анализов. В последнем случае это может быть, например, последовательность временных интервалов, в которых уровень глюкозы в крови выходил за пределы нормы. Таким образом, временные последовательности событий могут происходить как из административных данных, так и из правильно обработанных данных временных рядов. Извлечение значимых паттернов из временных последовательностей событий является очень важной областью исследования временных данных для формирования прогноза в медицине [69]. Такие алгоритмы могут значительно расширить набор инструментов поддержки принятия решений врачом для прогнозирования будущих событий. Модели поиска закономерностей в последовательностях базируются на техниках выявления ассоциаций [70]. Как частный случай этих методов, были разработаны подходы для выявления правил временной ассоциации, в которых набор одно-временных событий (паттернов) предшествует другому интересующему нас событию [71].

В настоящее время такие техники выделения временных паттернов в основном применяются к клиническим временным последовательностям. Например, был предложен алгоритм для анализа связанных временных интервалов, который применялся к набору данных о пациентах

с диабетом, включающим данные об уровне глюкозы в крови, холестерина, гемоглобина и применяемых лекарств [72]. В работе [73] предложен метод поиска временных шаблонов в данных на основе динамического выделения временных ассоциаций на разных иерархических уровнях. Метод был применен к набору данных о ВИЧ-пациентах для изучения временных взаимосвязей между новыми мутациями, вызванными используемой терапией. Примеры таких методов можно продолжать, они в основном, как говорилось выше, ограничиваются анализом клинических данных. Исследования прогностических возможностей анализа административных баз данных, содержащих историю болезни пациентов только начинается. За рубежом методы временного анализа были применены в области выявления побочных реакций на лекарства, используя базы данных рецептов и клинических данных [74–76].

В работе [77] рассмотрен метод для извлечения неочевидных паттернов из административных данных о состоянии здоровья с учетом последовательности событий, связанных с назначением лекарств. В работе [78] был предложен метод обнаружения паттернов в больших наборах данных записей пациентов, который основан на статистическом и графическом подходе для представления связи между назначениями лекарств и клиническими событиями, зарегистрированными во временной истории пациентов. В работах [77; 78] методы выделения паттернов во временных последовательностях были применены для оценки взаимосвязей между назначениями лекарств и вариациями в клинических состояниях пациентов с диабетом. В работе [79] та же методология применяется для оценки затрат, связанных с фармакологическим лечением диабета.

## **5. ПРОГНОСТИЧЕСКИЕ МОДЕЛИ И СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ**

В данной статье мы рассмотрели цели и задачи прогностической медицины, останавливаясь на различных подходах, которые могут помочь в решении ее основной задачи — прогноза возникновения хронических заболеваний у здоровых индивидуумов или прогноз течения заболевания у диагностированных пациентов.

Прогностическая медицина базируется на двух достижениях современной науки. Во-первых, все более точные и надежные методы клинической диагностики, которые основаны как на успехах молекулярной медицины (как мы уже отмечали, это привело к развитию персонализированной медицины), так и на совершенствовании различных методов визуализации — МРТ, ПЭТ, нейровизуализации и т.д. Во-вторых, это развитие информационных технологий в сфере интеллектуального анализа данных. Сочетание обоих этих подходов приводит к формированию новой парадигмы медицины — переходу от реактивной концепции в медицине, направленной на реакцию на заболевание, к активной, основанной на надежном прогнозе течения заболевания.

Создание и широкий доступ к большим коллекциям клинических данных, которые часто находятся в открытом доступе в обезличенном виде, позволяет проводить тщательный ретроспективный анализ, который может дать учреждениям здравоохранения беспрецедентную возможность лучше понять природу и особенности проходящих клинических процессов. Кроме того, наличие крупномасштабных молекулярных данных может дать понимание индивидуальных особенностей каждого отдельного пациента и предложить изменения в стратегиях принятия решений, основанных на персонализации. Поэтому область интеллектуального анализа данных, которая основана на развитии различных систем искусственного интеллекта, может помочь медицине в ее трансформации от популяционной через персонализированную к прогностической медицине.

Для достижения данной цели решающее значение имеет использование методов, способных

в автоматическом режиме в реальном времени работать с клиническими изображениями, различными типами временной информацией, а также разработка новых инструментов анализа данных, способных интегрировать данные и знания в согласованную структуру. Последнее возможно реализовать путем разработки систем поддержки принятия решений врачом (СППВР), в которых включались бы интеллектуальные модули с заложенными в них прогностическими моделями, разработка становится актуальной и важной задачей. В первую очередь при создании СППВР необходимо предусмотреть, что различные источники данных должны быть должным образом интегрированы, при этом интеллектуальный анализ данных, как получаемых в реальном времени, так и собранных в ретроспективе, являются центральным этапом, на котором выделяются наиболее важные с точки зрения прогноза биомаркеры. Для оценки модели необходим независимый набор данных для оценки эффективности прогнозирования. Рабочая прогностическая модель должна работать в рамках клинического контекста и трансформироваться при накоплении новых данных [55].

Изучение прогностических моделей, которые можно применять в клинической практике, требует всесторонней верификации (Рис. 4). Прежде всего, прогностическая модель должна быть обоснована статистически. После первичной оценки ее качества, выполненной методами перекрестной проверки, оценка эффективности должна обязательно проводиться на достаточно большом независимом тестовом наборе. Более того, процесс обучения должен быть максимально прозрачным и воспроизводимым. Для медицины является критическим контекст безопасности, и все этапы моделирования,



Рисунок 4 — Шаги построения клинической прогностической модели.

включая выбор параметров проектирования, должны быть четкими и обоснованными. Наконец, выбор прогностической модели также связан с ее возможностью развертывания в клиническом контексте: модели, которые легко объяснить и которые подтверждаются статистическими данными, с большей вероятностью будут приняты в клинической практике, чем модели черного ящика. Это ограничивает применение ряда технологий машинного обучения, которые непрозрачны, в частности, многие модели ИНС фактически представляют «черный ящик», который невозможно описать кроме как полным описанием модели в виде задания всех параметров ИНС и их сложно воспроизвести. Интеллектуальные системы СППВР могут быть существенно упрощены путем продолжения исследований, направленных на уменьшение

объема информации и выделение конкретных наиболее значимых биомаркеров, полезных для диагностики, прогноза и лечения того или иного заболевания.

В настоящее время делаются попытки разработки систем прогноза течения заболеваний для отдельных болезней. Систем, оценивающих совокупное состояние пациента, в настоящее время практически не существует. Остается также нерешенным важный и открытый вопрос горизонта прогноза состояния пациента. Его увеличение, без сомнения, очень важно при лечении ряда болезней, особенно в их острой фазе, что позволило бы назначать более эффективное и щадящее лечение, а также надежно выделять критические состояния пациента. Все эти задачи являются одними из самых важных для развития точной прогностической медицины.

## ЛИТЕРАТУРА/REFERENCES

1. Hodson R. Precision medicine. *Nature*. 2016; 537(7619): 49.
2. Hirsch L, Jette N, Frolkis A, Steeves T, Pringsheim T. The incidence of Parkinson's disease: a systematic review and meta-analysis. *Neuroepidemiology*. 2016; 46(4): 292-300.
3. Driver JA, Logroscino G, Gaziano JM, Kurth T. Incidence and remaining lifetime risk of Parkinson disease in advanced age. *Neurology*. 2009; 72(5): 432-438.
4. Alzheimer's Association Report. 2020 Alzheimer's disease facts and figures. *Alzheimer's Dement*. 2020; 16: 391-460.
5. Pravettoni G, Triberti S. A "P5" Approach to Healthcare and Health Technology. In *P5 eHealth: An Agenda for the Health Technologies of the Future*. 2020; Springer, Cham: 3-17.
6. Hood L. Systems medicine, transformational technologies and the emergence of predictive, personalized, preventive and participatory (P4) medicine. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. ACM. 2009.
7. Hood L. Systems biology and p4 medicine: past, present, and future. *Rambam Maimonides medical journal*. 2013; 4(2).
8. Cesario A, Auffray C, Russo P, Hood L. P4 medicine needs P4 education. *Current Pharmaceutical Design*. 2014; 20(38): 6071-6072.
9. Pulciani S, Di Lonardo A, Fagnani C, Taruscio D. P4 medicine versus Hippocrates. *Annali dell'Istituto superiore di sanita*. 2017; 53(3): 185-191.
10. Леонов С.А. Интеграция здравоохранения, образования и информационно-коммуникационных технологий в рамках цифровизации отечественной медицины // *Актуальные проблемы экономики и управления*. — 2018. — №3. — С.35. [Leonov S.A. Integraciya zdravoohraneniya, obrazovaniya i informacionno-kommunikacionnyh tekhnologij v ramkah cifrovizacii otechestvennoj mediciny. *Aktual'nye problemy ekonomiki i upravleniya*. 2018; 3: 35. (In Russ).]
11. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*. 1998; 95(25): 14863-14868.
12. Nicolini C, Bragazzi N, Pechkova E. Nanoproteomics enabling personalized nanomedicine. *Advanced drug delivery reviews*. 2012; 64(13): 1522-1531.
13. Cutica I, Mc Vie G, Pravettoni G. Personalised medicine: the cognitive side of patients. *European Journal of Internal Medicine*. 2014; 25(8): 685-688.
14. Vickers NJ. Animal Communication: When I'm Calling You, Will You Answer Too? *Current Biology*. 2017; 27(14): R713-R715.

15. Domenighetti G, Grilli R, Liberati A. Promoting consumers' demand for evidence-based medicine. *International Journal of Technology Assessment in Health Care*. 1998; 14(1): 97-105.
16. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *New biotechnology*. 2012; 29(6): 613-624.
17. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2011; 1(5): 416-430.
18. Barbolosi D, Ciccolini J, Lacarelle B, Barlési F, André N. Computational oncology — mathematical modelling of drug regimens for precision medicine. *Nature reviews Clinical oncology*. 2016; 13(4): 242.
19. Грибова В.В., Петряева М.В., Окунь Д.Б., Шалфеева Е.А. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. — 2018. — №8(1). — С.27. [Gribova VV, Petryaeva MV, Okun' DB, Shalfeeva EA. Ontologiya medicinskoj diagnostiki dlya intellektual'nyh sistem podderzhki prinyatiya reshenij. *Ontologiya proektirovaniya*. 2018; 8(1): 27. (In Russ).]
20. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Breast Cancer Linkage Consortium. American journal of human genetics*. 1995; 56(1): 265.
21. Carter RF. BRCA1, BRCA2 and breast cancer: a concise clinical review. *Clinical and Investigative Medicine*. 2001; 24(3): 147.
22. Kujovich JL. Factor v Leiden thrombophilia. *Genetics in Medicine*. 2011; 13(1): 1-16.
23. Torkamani A, Andersen KG, Steinhubl SR, Topol EJ. High-definition medicine. *Cell*. 2017; 170(5): 828-843.
24. Tuena C, Semonella M, Fernández-Álvarez J, Colombo D, Cipresso P. Predictive Precision Medicine: Towards the Computational Challenge. In *P5 eHealth: An Agenda for the Health Technologies of the Future*. 2020. Springer, Cham: 71-86.
25. Strimbu K, Tavel JA. What are biomarkers? *Current Opinion in HIV and AIDS*. 2010; 5(6): 463.
26. Chen JJ, Lu TP, Chen YC, Lin WJ. Predictive biomarkers for treatment selection: statistical considerations. *Biomarkers in medicine*. 2015; 9(11): 1121-1135.
27. Ginsburg GS, Phillips KA. Precision medicine: from science to value. *Health Affairs*. 2018; 37(5): 694-701.
28. Wang E, Cho WC, Wong SC, Liu S. Disease biomarkers for precision medicine: challenges and future opportunities. *Genomics, proteomics & bioinformatics*. 2017; 15(2): 57.
29. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. *N Engl J Med*. 2012; 366(6): 489-491.
30. Mitchell RS, Michalski JG, Carbonell TM. *An artificial intelligence approach*. 2013. Springer, Berlin.
31. Giger ML. Machine learning in medical imaging. *Journal of the American College of Radiology*. 2018; 15(3): 512-520.
32. Cipresso P, Matic A, Giakoumis D, Ostrovsky Y. *Advances in computational psychometrics. Computational and mathematical methods in medicine*, 2015.
33. Leff DR, Yang GZ. Big data for precision medicine. *Engineering*. 2015; 1(3): 277-279.
34. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *The New England journal of medicine*. 2016; 375(13): 1216.
35. Alpaydin E. *Introduction to machine learning*. MIT press, 2020.
36. Mayer Schönberger V, Ingelsson E. Big Data and medicine: a big deal? *Journal of internal medicine*. 2018; 283(5): 418-429.
37. Jacob Rodrigues M, Postolache O, Cercas F. Physiological and Behavior Monitoring Systems for Smart Healthcare Environments: A Review. *Sensors*. 2020; 20(8): 2186.
38. Arac A. Machine Learning for 3D Kinematic Analysis of Movements in Neurorehabilitation. *Current Neurology and Neuroscience Reports*. 2020; 20(8): 1-6.
39. Wagenaar JB, Worrell GA, Ives Z, Dümpelmann M, Litt B, Schulze-Bonhage A. Collaborating and sharing data in epilepsy research. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*. 2015; 32(3): 235.
40. Klatt J, Feldwisch Drentrup H, Ihle M, Navarro V, Neufang M, Teixeira C, Le Van Quyen M. The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients. *Epilepsia*. 2012; 53(9): 1669-1676.



41. Frey LJ, Bernstam EV, Denny JC. Precision medicine informatics. *Journal of the American Medical Informatics Association*. 2016; 23(4): 668-670.
42. Hsiao CJ, Hing E, Ashman J. Trends in Electronic Health Record System Use Among Office-based Physicians, United States. 2014. 2007-2012 (No. 75). US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
43. Kruse CS, Stein A, Thomas H, Kaur H. The use of electronic health records to support population health: a systematic review of the literature. *Journal of medical systems*. 2018; 42(11): 1-16.
44. Hecht J. The future of electronic health records. *Nature*. 2019; 573(7775): S114-S114.
45. GE Healthcare Systems. 2011. Centricity Electronic Medical Record: Experience that counts. URL:<http://www3.gehealthcare.com/media/Downloads/us/Product/Product-Categories/Healthcare%20IT/Electronic%20Medical%20Records/ITP01981010ENUScentricityemrbrochure.pdf?Parent=%7B3FB3AC2B-38EE-4838-B06E-4472AFC090F0%7D>.
46. Montvida O, Dibato JE, Paul S. Evaluating the representativeness of US centricity electronic medical records with reports from the centers for disease control and prevention: comparative study on office visits and cardiometabolic conditions. *JMIR Medical Informatics*. 2020; 8(6): e17174.
47. Brixner D, Said Q, Kirkness C, Oberg B, Ben Joseph R, Oderda G. Assessment of cardiometabolic risk factors in a national primary care electronic health record database. *Value in health*. 2007; 10: S29-S36.
48. Crawford AG, Cote C, Couto, et al. Comparison of GE Centricity Electronic Medical Record database and National Ambulatory Medical Care Survey findings on the prevalence of major conditions in the United States. *Population health management*. 2010; 13(3): 139-150.
49. Montvida O, Shaw J, Atherton JJ, Stringer F, Paul SK. Long-term trends in antidiabetes drug usage in the US: real-world evidence in patients newly diagnosed with type 2 diabetes. *Diabetes care*. 2018; 41(1): 69-78.
50. Lundin P, Eriksson SV, Strandberg LE, Rehnqvist N. Prognostic information from on-line vectorcardiography in acute myocardial infarction. *The American journal of cardiology*. 1994; 74(11): 1103-1108.
51. Yang L, Xia C, Mu Y, Guan L, Wang C, Tang Q, Shih MC. Prognostic value of real time myocardial contrast echocardiography after percutaneous coronary intervention. *Echocardiography*. 2016; 33(3): 421-430.
52. Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a Kim case study. *Nature Reviews Cancer*. 2016; 16(8): 525.
53. Hoggatt J. Personalized Medicine — Trends in Molecular Diagnostics. *Molecular diagnosis & therapy*. 2011; 15(1): 53-55.
54. Kim J, Hong J, Park H. Prospects of deep learning for medical imaging. *Precision and Future Medicine*. 2018; 2(2): 37-52.
55. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2011; 1(5): 416-430.
56. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y. Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*. 2017; 2(4): 230-243.
57. Ker J, Lipo W, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access*. 2018; 6: 9375-9389.
58. Hu S, Gao Y, Niu Z, Jiang Y, Li L, Xiao X, Ye H. Weakly supervised deep learning for covid-19 infection detection and classification from ct images. *IEEE Access*. 2020; 8: 118869-118883.
59. Rojas R. *Neural networks: a systematic introduction*. Springer Science & Business Media. 2013.
60. Dietterich T. Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*. 1995; 27(3): 326-327.

61. Chholak P, Niso G, Maksimenko VA, Kurkin SA, Frolov NS, Pitsik EN, Hramov AE, Pisarchik AN. Visual and kinesthetic modes affect motor imagery classification in untrained subjects. *Scientific reports*. 2019; 9(1): 1-12.
62. Barrett JH, Cairns D A. Application of the random forest classification method to peaks detected from mass spectrometric proteomic profiles of cancer patients and controls. *Statistical Applications in Genetics and Molecular Biology*. 2008; 7(2).
63. Fan S, Kind T, Cajka T, Hazen SL, Tang WW, Kaddurah-Daouk R, Fiehn O. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. *Analytical chemistry*. 2019; 91(5): 3590-3596.
64. Adlassnig KP, Combi C, Das AK, Keravnou ET, Pozzi G. Temporal representation and reasoning in medicine: Research directions and challenges. *Artificial intelligence in medicine*. 2006; 38(2): 101-113.
65. Sternickel K. Automatic pattern recognition in ECG time series. *Computer methods and programs in biomedicine*. 2002; 68(2): 109-115.
66. Aldroubi A. *Wavelets in medicine and biology*. Routledge. 2017.
67. Hramov AE, Koronovskii AA, Makarov VA, Pavlov AN, Sitnikova E. *Wavelets in neuroscience*. Springer Berlin Heidelberg. 2015.
68. Maksimenko VA, van Heukelum S, Makarov VV, Kelderhuis J, Lüttjohann A, Koronovskii AA, Hramov AE, van Luijckelaar G. Absence seizure control by a brain computer interface. *Scientific Reports*. 2017; 7(1): 1-8.
69. Fournier-Viger P, Lin JCW, Kiran RU, Koh YS, Thomas R. A survey of sequential pattern mining. *Data Science and Pattern Recognition*. 2017; 1(1): 54-77.
70. Sethi A, Mahajan P. Association Rule Mining: A Review. *International Journal of Computer Science*. 2012; 1(9).
71. Shukla DP, Patel SB, Sen AK. A literature review in health informatics using data mining techniques. *International Journal of Software and Hardware Research in Engineering*. 2014; 2(2): 123-129.
72. Moskovitch R, Shahar Y. Medical temporal-knowledge discovery via temporal abstraction. In *AMIA annual symposium proceedings*. American Medical Informatics Association. 2009: 452).
73. Tu SW, Tennakoon L, O'Connor M, Shankar R, Das A. Using an integrated ontology and information model for querying and reasoning about phenotypes: the case of autism. In *AMIA Annual Symposium Proceedings*. American Medical Informatics Association. 2008: 727.
74. Jin H, Chen J, He H, Williams GJ, Kelman C, O'Keefe CM. Mining unexpected temporal associations: applications in detecting adverse drug reactions. *IEEE Transactions on Information Technology in Biomedicine*. 2008; 12(4): 488-500.
75. Chazard E, Preda C, Merlin B, Ficheur G, Beuscart R. Data-mining-based detection of adverse drug events. *Stud Health Technol Inform*. 2009; 150: 552-556.
76. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Mining and Knowledge Discovery*. 2010; 20(3): 361-387.
77. Combi C, Franceschet M, Peron A. Representing and reasoning about temporal granularities. *Journal of Logic and Computation*. 2004; 14(1): 51-77.
78. Combi C, Pinciroli F, Pozzi G. Managing different time granularities of clinical information by an interval-based temporal data model. *Methods of information in medicine*. 1995; 34(05): 458-474.
79. Concaro S, Sacchi L, Cerra C, Stefanelli M, Fratino P, Bellazzi R. Temporal data mining for the assessment of the costs related to diabetes mellitus pharmacological treatment. In *AMIA Annual Symposium Proceedings*. American Medical Informatics Association. 2009: 119.