

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.9

## ТЕХНОЛОГИЯ BIG DATA: ПОТЕНЦИАЛ, ПРОБЛЕМЫ И ПРИМЕНЕНИЕ В МЕДИЦИНЕ И ЗДРАВООХРАНЕНИИ

МАГЕРРАМОВ З.Т., РАГИМОВА Н.А.,

АБДУЛЛАЕВ В.Г., МАГЕРРАМОВА Г.З.

Исследуются возможности применения технологии BIG DATA в медицине и в системе здравоохранения в целях совершенствования принятия решений в клинической практике. Проводится углубленный анализ состояния развития и особенностей применения технологии Big Data в сфере медицины и здравоохранения. Выявляются основные источники и характеристики медицинских данных. Показывается, что внедрение технологий Big Data в медицине может быть достигнуто при широком использовании представления медико-биологической информации в цифровом виде. Приводятся данные о том, что персонализированная медицина является одной из главных трендов в области сохранения здоровья, и интегрированная электронная медицинская карта занимает особое место в системе здравоохранения. Анализируется внедрение технологий Big Data в фармацевтике, биомедицине, психиатрии, эпидемиологии и в телемедицине. Подчеркивается, что для развития исследований геномики и микробиома человека особое место занимает технология Big Data. Формулируются перспективные направления применения технологии Big Data в сфере медицины и здравоохранения.

**Ключевые слова:** Big Data, медицинские данные, источник медицинских данных, цифровизация медицины, персонализированная медицина, электронная медицинская карта, фармацевтика, биоинформатика, биомедицина, геномика, микробиом, рак, сахарный диабет, психиатрия, эпидемиология, телемедицина.

**Keywords:** Big Data, medical data, a source of medical data, digitalization of medicine, the personalized medicine, the electronic medical record, pharmaceuticals, bioinformatics, biomedicine, genomics, a microbiome, cancer, diabetes, psychiatry, epidemiology, telemedicine.

### 1. Введение

Термин Big Data относится к наборам данных, размер которых превосходит возможности типичных баз данных по хранению, управлению и анализу информации. Сами по себе алгоритмы Big Data возникли при внедрении первых высокопроизводительных серверов (мэйнфреймов), обладающих достаточными ресурсами для оперативной обработки информации и пригодных для компьютерных вычислений и для дальнейшего анализа.

Термин Big Data впервые был озвучен в 2008 году на страницах спецвыпуска журнала Nature в статье главного редактора Клиффорда Линча. Этот номер издания был посвящен взрывному

росту глобальных объемов данных и их роли в науке.

С каждым годом увеличивается объем генерируемой человеком информации. И рост продолжается экспоненциально. Аналитики компании IBS оценили «весь мировой объем данных» такими величинами [1]: 2003 г. – 5 эксабайтов данных (1 ЭБ = 1 млрд гигабайтов); 2008 г. – 0,18 зеттабайта (1 ЗБ = 1024 эксабайта); 2015 г. – более 6,5 зеттабайтов; 2020 г. – 40-44 зеттабайта (прогноз); 2025 г. – этот объем вырастет еще в 10 раз.

Соответственно, управление структурированными и неструктурированными данными при помощи современных технологий – сфера, которая становится все более значимой [2].

По оценкам компании Frost & Sullivan [3] в 2022 году суммарный объем рынка технологий искусственного интеллекта (ИИ) увеличится до \$52,5 млрд или в 4 раза по сравнению с уровнем 2017 года (\$13,4 млрд). Ежегодный темп роста CAGR (Generally Accepted Risk Principles) в прогнозируемый период будет сохраняться на уровне 31%. Повсеместное внедрение технологий ИИ к 2030 году увеличит объем глобального рынка товаров и услуг на \$15,7 трлн.

Согласно отчету IDC [4], в связи с ростом объемов данных, генерируемых подключенными к интернету устройствами, датчиками и другими технологиями, доходы, связанные с большими данными, увеличатся со \$130 млрд в 2016 году до более чем \$203 млрд к 2020 году.

При этом крупнейшими сегментами рынка станут производственный сектор, финансы, здравоохранение, охрана окружающей среды и розничная торговля. Влияние различных отраслей на рынок аналитики больших данных показан на рисунке [5].

### 2. Особенности применения технологии Big Data в медицине и здравоохранении

Как видно из рисунка, рейтинг влияния здравоохранения на рынок аналитики больших данных достаточно высок. Исследования показывают, что более 30% всех данных, хранящихся на планете, составляют медицинские данные [6]. В будущем ожидается быстрое увеличение этой доли за счет создания новой и оцифровки уже имеющейся информации. Объем затрат на Big Data в области медицины к 2022 году достигнет 34,27 млрд долларов [7, 8]. Немаловажным является и тот факт, что вся эта информация уже не исключительно статистическая, а представляет собой массив неструктурированных и разнородных данных, что определяет различные инструменты и методы для их сбора, хранения и анализа. Согласно прогнозам Deloitte, к 2022 году объем расходов на мировом рынке здравоохранения достигнет \$10,059 трлн [9].



Мировым лидером по разработке и внедрению технологий Big Data в здравоохранении на сегодняшний день являются США. Главное основание для их развития – экономическая эффективность от их внедрения. По мнению аналитиков McKinsey Global Institute, использование технологий Big Data в здравоохранении США будет формировать финансовый поток объемом 300 млрд долл. в стоимостном выражении, из которых две трети – за счет снижения расходов системы здравоохранения США [10].

Стоит отметить, что сегодня Big Data – это не только информация, но и инструмент, который имеет неограниченные возможности и способствует получению новых решений, качественной трансформации процессов медицинской помощи и прогрессивному развитию системы здравоохранения в целом. Ученые исследуют возможности использования больших объемов данных для улучшения уровня доказательности и принятия клинических решений. Big Data в медицине и в здравоохранении позволяет совершенно по-новому подходить к вопросам оказания медицинской помощи как применительно к одному человеку, так и учитывая положение дел в масштабах города или даже страны. В настоящее время технологии Big Data используются для: совершенствования принятия решений в клинической практике; выявления пациентов группы риска, у которых возникли побочные реакции или установлена неэффективность лекарственных средств; повышения качества индивидуального обслуживания пациентов в целях получения лучших результатов; улучшения показателей доказанной эффективности лекарств с учетом данных, полученных во время реальной практики; улучшения планирования ресурсов здравоохранения.

Большой потенциал использования технологий Big Data в медицине связан с разработкой алгоритмов распознавания, дальнейшего анализа и интерпретации сигналов и изображений. Сигналы с носимых устройств, характеризующиеся большим объемом и скоростью поступления, особенно при непрерывном использовании в режиме реального времени, генерируемые множеством подключенных к пациенту датчиков, имеют большую сложность для обработки, хранения и анализа.

Благодаря достижениям в области Big Data возможно: реализовать в полной мере персонализированный подход к пациенту и его истории болезней; прогнозировать развитие болезни и риски при лечении; определять признаки болезни, паттернов ее протекания и терапии; осуществлять мониторинг состояния пациента с помощью различных устройств; оценивать эффективность лечения и выявления связей между причинами заболевания и его симптомами; более точно прогнозировать; рассчитывать необходимость в оборудовании и лекарственном обеспечении; управлять потоками пациентов и многое другое.

Цифровизация медицины – перспективное направление, которое не только унифицирует работу клиник или лабораторий, но и может спасти человеческие жизни.

В 2016 году рынок цифровой медицины оценивался в 179,6 млрд долларов США, по данным Transparency Market Research. По прогнозам аналитического агентства, GARP в этом сегменте составит 13,4%, вплоть до 2025 года, когда его объем превысит 536 млрд долларов. Более скромные прогнозы показывает статистическое агентство Statista: рынок цифровой медицины к 2020 году вырастет до 206 млрд [11].

Технологии Big Data в медицине используются в рамках следующих направлений: прогнозирование развития болезней; выявление генетических маркеров в онкологии; прогнозирование состояния здоровья младенцев; постановка диагнозов при помощи использования носимых устройств. В качестве средств поддержки Big Data можно выделить сервис Watson Health и систему QPID (Queryable Patient Interface Dossier). Watson Health представляет собой сервис от компании IBM в области медицинской аналитики на базе технологий Big Data. Watson Health позволяет принимать решения, основанные на большом количестве медицинских неструктурированных данных, среди которых можно отметить: клинические, научные, генетические и личные данные [12, 13]. Это особенно важно в связи с повсеместным распространением ГМО, пестицидов, химикатов и аллергенов, которые на уровне генетики негативно действуют на человеческий организм. Прогнозирование факторов риска в хирургии можно организовать на базе использования аналитической системы QPID, которая позволяет контролировать важную информацию о пациенте на протяжении курса лечения, прогнозировать хирургические риски с помощью информации с электронных медицинских карт. Система QPID выполняет автоматический поиск протоколов лечения, после чего выводит на экран результаты с рассчитанным красным, желтым или зеленым индикатором риска [14]. Данная технология позволяет найти точный и быстрый способ определения, какие пациенты характеризуются высоким риском развития сепсиса, что является критически важной задачей, и отслеживать побочные действия различных препаратов и вспышки эпидемий.

Стремительное развитие биоинформатики, а также возможности анализа технологий Big Data открывает принципиально новые методы диагностики и лечения, вызывает необходимость проведения дальнейших углубленных исследований. Применение Big Data в медицине позволит решить важные задачи из области биоинформатики, состоящие в создании и сопровождении баз данных и знаний, таких как специализированные базы белковых структур, нуклеотидных последовательностей генов, метаболических путей, клеточных ансамблей и других [8, 15]. Число и объём информации подобных баз данных непрерывно растёт, работа с такими огромными массивами информации требует принципиально новых подходов к обработке данных и соответствующего программного обеспечения [16].

Аналитика Big Data в медицине объединяет методы анализа из ряда научных областей, таких

как биоинформатика, медицинская визуализация, сенсорная и медицинская информатика, искусственный интеллект и других [17]. Технологии Big Data направлены на анализ все более сложных массивов медицинских данных, формирующихся из разнообразных по структуре, формату, достоверности источников информации. Big Data включает как структурированные данные, так и неструктурированные и полуструктурированные данные (документы XML). Около 80 % данных в медицине являются неструктурированными и представляют набор файлов, таблиц, рисунков, графиков и их описаний [18]. На сегодня достигнут значительный прогресс в использовании инструментальных технологий в сборе, хранении и анализе стоимости данных. Самой актуальной проблемой при оперировании Big Data стала разработка алгоритмов комплексного анализа и интерпретации данных в режиме реального времени. В совокупность подходов и технологий анализа Big Data входят средства массово-параллельной обработки неопределённо-структурированных данных, в частности NoSQL, алгоритмы MapReduce и средства проекта Hadoop, другие решения, обеспечивающие сходные по характеристикам возможности по обработке сверхбольших массивов данных, а также некоторые аппаратные средства.

Таким образом, Big Data в медицине – это непрерывно и быстро пополняемые электронные массивы качественно различающихся между собой медицинских и парамедицинских данных огромного объёма, которыми невозможно управлять посредством традиционных инструментов и методов программного и/или аппаратного обеспечения. Очевидно, что внедрение технологий Big Data позволит не только решать текущие задачи в медицине, но и даст возможность преодолеть недостижимые горизонты обработки медицинских данных в области здравоохранения. В то же время данное исследование нельзя считать законченным, необходимо продолжать работы по изучению технологий Big Data и алгоритмических решений обработки больших данных. Исследование особенностей и перспектив применения технологий Big Data в медицине и здравоохранении является актуальной научно-практической задачей.

### **3. Применение и потенциал Big Data в отдельных направлениях медицины и здравоохранения**

Применение Big Data в медицине и здравоохранении включает в себя интеграцию и анализ большого количества сложных разнородных показателей, таких как данные геномики, эпигеномики, транскриптомики, протеомики, метабомики, интерактомики, фармакогеномики, деза-

сомики, биомедицинские данные и данные электронных медицинских карт [8, 19, 20]. В 2015 году Apple и IBM также пришли к решению использования Big Data в сфере здравоохранения.

За последние несколько десятилетий медицинские идеи и технологии, считавшиеся ранее научной фантастикой, стремительно ворвались в фундаментальную и клиническую медицину, что позволяет ей совершить переход от стандартизированных методов лечения к индивидуальному подбору лекарственных средств с учетом генетических особенностей каждого конкретного пациента [21].

Кроме того, генерация идей на стыке медицины и информационных технологий меняет привычные подходы к мониторингу состояния здоровья и лечению. В данном контексте, рассматривая медицину как комплексную и специфическую сферу деятельности, которая характеризуется очень высоким уровнем разнородной лечебной и экономической информации, актуализируются две задачи – оптимизация клинической и экономической деятельности медицинского учреждения. Конечной целью этих задач является оптимизация деятельности системы здравоохранения в целом и любого медицинского учреждения в частности.

Не вызывает сомнения тот факт, что новейшие технологии и создание экспертных систем в медицине, способных на основе симптомов болезни и лабораторно-инструментальных данных предвидеть и достоверно прогнозировать возможные заболевания, могут значительно облегчить принятие решения специалистами в медицинской сфере. Такие технологии помогут выбрать врачу наиболее важные результаты в огромном количестве информации из истории болезни для назначения адекватного лечения.

Очевидно, что применение технологий Big Data для анализа все более сложных массивов медицинских данных открывает новые возможности в области здравоохранения. Основные задачи, стоящие перед разработчиками технологий Big Data в медицине, определяются, главным образом, особенностями циркулирующих в современном здравоохранении и биомедицине данных. Эти данные зачастую являются непреодолимыми для обработки с помощью традиционного программного обеспечения не только из-за их объема, но и из-за разнообразия типов данных и скорости, с которой они должны анализироваться.

### **3.1. Характеристики Big Data**

Известно, что Big Data как технология характеризуется несколькими «V», среди которых особое значения имеют следующие: объем (Volume), разнообразие (Variety), изменчивость

(Variability), скорость (Velocity), достоверность (Veracity) и ценность (Value) [22-24].

Имеющиеся в системе здравоохранения данные в полной мере отвечают этим характеристикам. Это и медицинские данные, описывающие все лечебно-профилактические мероприятия каждого пациента, и данные о ресурсах, которые участвуют в оказании этой помощи. Скорость появления новых данных в здравоохранении очень высока, однако их несвоевременная обработка при решении ряда задач очень быстро обесценивает полученную информацию. Кратко рассмотрим эти характеристики в контексте медицины.

**Объем (Volume) медицинских данных.** Медицинская информация представляется в виде больших массивов данных, требующих соответствующего места для хранения и инструментария для анализа. Непрерывное генерирование и накопление медицинской информации в ближайшие годы приведут к невероятному объему данных. В настоящее время медицинские данные включают в себя клинические данные, полученные врачами, личные медицинские записи пациентов (EHR), радиологические изображения (рентгеновские и маммологические снимки), МРТ (магнитно-резонансная томография), КТ (компьютерная томография), данные лабораторий и аптек, страховые претензии. Параллельно генерируется также огромное количество медицинской информации, не относящейся конкретно к пациенту. Это многочисленные медицинские публикации, научно-исследовательские отчеты, результаты исследований и разработок, социологические опросы и др. Новые типы медицинских данных, такие как многомасштабная 3D/4D визуализация, геномика, биометрические показания датчиков и т.п., также способствуют экспоненциальному росту медицинской информации.

**Разнообразие (Variety).** Медицинская информация является неоднородной по своей природе, она собирается из различных источников: флюорография, компьютерная томография, ультразвуковые исследования, лабораторные исследования, жалобы пациентов, анамнезы. В здравоохранении самой большой проблемой при работе с данными является их слабая структурированность в большинстве источников. Около 80% медицинских данных являются неструктурированными, причем наблюдается устойчивая тенденция к значительному росту скорости этой информации относительно таковой для структурированных данных. Неструктурированные – это данные, не имеющие заранее определенной структуры (медицинские изображения, видео, речь). Неструктурированные данные также могут быть представлены в виде текста, в котором да-

ты, числа и информация о наступлении каких-либо событий расположены в произвольном порядке и различных форматах. Это приводит к невозможности обработки данных информационными системами. Объединение большого количества данных из разных источников позволяет на основе их ретроспективного анализа выявлять закономерности, которые могут быть обнаружены исключительно при обработке больших массивов данных. Именно для такой информации и предназначены технологии больших данных.

**Изменчивость (Variability) медицинских данных.** Изменчивость информации: таковой является информация, непрерывно поступающая с биометрических датчиков или из интернета и имеющая важное значение для анализа, прогнозирования и принятия решений. Они может иметь пики и спады, сезонности, периодичность.

**Скорость (Velocity) производимых медицинских данных.** Постоянный поток новых данных накапливается беспрецедентными темпами, т.е. наблюдаемый рост объема и разнообразия данных непосредственно связан со скоростью, с которой они генерируются. Специфика проблемы в медицине заключается в том, что скорость пополнения данных также накладывает ограничения на обработку Big Data в медицине. Так, информация от приборов, следящих за пациентами отделения интенсивной терапии, поступает непрерывно, т.е. в режиме реального времени, и требуются ее немедленная обработка и анализ для своевременной выработки предварительного диагноза.

**Достоверность (Veracity) медицинских данных.** Этот признак отражает семантическую и синтаксическую определенность, качество, актуальность и надежность данных. Некоторые обеспечения достоверности данных специфичны именно для медицины, поскольку имеют дело с диагнозами, методами лечения, рецептами, процедурами. Так, результаты анализа Big Data должны быть безошибочными и надежными, поскольку от достоверности сделанного на основе анализа Big Data вывода может зависеть жизнь человека. С другой стороны, низкое качество медицинских данных, особенно неструктурированных, является одной из серьезных проблем: заполняемые часто с ошибками медицинские карточки, неверное истолкование и неточный цифровой ввод назначений врача в рецептах из-за плохого почерка относятся к разряду наиболее распространенных примеров [24]. Слишком высокая цена ошибки является причиной недоверия медицинской общественности к результатам анализа Big Data.

**Ценность (Value) накопленных данных.** Этот показатель, представляющий интерес для различных заинтересованных сторон и лиц, которые принимают решение, характеризует Big Data с позиций их полезности и привнесения определенной ценности для медицинского учреждения и системы здравоохранения в целом (например, усовершенствование бизнес-процессов, определение бизнес-стратегий, оптимизация расходов) [23].

Кроме этих характеристик, важными являются также [25]:

**Сложность.** Медицинские документы являются сложными по своей природе для структурирования и математического описания.

**Исключительная важность.** Медицинская информация является исключительно важной, поскольку она служит для диагностики и лечения заболеваний.

**Чувствительность.** Медицинская информация является чувствительной к мнению и оценкам врача, медицинским осмотрам и клиническим анализам. Большинство диагнозов и анализов содержит некоторый процент ошибок [2].

**Правовые, этические и социальные аспекты.** Медицинская информация — это в основном данные пациентов. В связи с этим возникают вопросы обеспечения конфиденциальности персональных медицинских данных в соответствии с действующим законодательством.

### 3.2. Источники медицинских данных

Существует три основных источника Big Data в медицине:

**первый:** электронные медицинские карты, данные о страховании и счетах, данные о лекарственных средствах и неотложной помощи, данные о результатах проведения клинических исследований и испытаний, данные с мониторов и записывающих устройств;

**второй:** данные геномики, данные протеомики, данные метаболомики;

**третий:** информация о пациентах, получаемая со смартфонов и из социальных сетей.

Источники медицинских данных включают в себя [25]:

- клинические данные для поддержки принятия решений различной специализации (диагностическая, прогностическая, с элементами искусственного интеллекта, управления, уход за больными) в виде стандартизированных данных из электронных историй болезни;
- данные генотипа;
- зарегистрированные данные с датчиков мониторинга и записывающих устройств;
- генерируемые экспертами конкретные показатели, письменные заметки и медицинские рецепты;

- звукозаписи и визуальные образы;
- данные специализированных исследований;
- данные о лекарственных препаратах;
- данные неотложной помощи;
- административно-паспортные данные;
- данные о страховании и медицинском страховании;
- социальные публикации в СМИ, в том числе Twitter-каналы, блоги, обновления статуса на Facebook и других платформ и веб-страниц;
- данные об опыте и результатах использования методов нетрадиционной медицины и непрофессиональных инициатив в области здравоохранения и медицины;
- нормативные и законодательные документы из области социальной медицины, общественно-го здравоохранения, рынка здравоохранения, политики и культуры;
- данные медицинской науки;
- социальные медиаданные;
- данные об окружающей среде (экологические данные).

К источникам Big Data относятся также другие биометрические данные – отсканированные отпечатки пальцев, почерк, сетчатка глаза, рентгеновские и другие медицинские изображения, значения жизненно важных показателей (артериальное давление, пульс).

В будущем около двух третей всех связанных с медициной данных будут поступать из созданных пациентом источников [8].

### 3.3. Персонализированная медицина

Один из главных трендов последних лет в области сохранения здоровья – персонализированная медицина, в задачи которой входит обеспечение пациента теми лекарствами и лечебными технологиями, которые подходят именно ему, а не какому-то «среднестатистическому» больному. Например, в онкологии персональный подход обеспечивается использованием диагностических тестов для выявления конкретных мутаций в опухолевых клетках и подбором и даже разработкой соответствующих противоопухолевых препаратов.

Персонализированная медицина – платформа, которая использует информацию о генах человека, белках и окружающей среде для предотвращения, своевременной диагностики и лечения заболеваний. Другими словами, ее именуют медициной 4P, или P4 predictive, или предсказательной, preventive, или предупредительной, personalize, или персонализированной, и participant, или партисипативной. Основные ее принципы более детально можно представить следующим образом [26, 27]:

- предиктивная медицина – предсказание особенностей состояния здоровья (возможные в будущем заболевания, реагирование на лекарства и факторы внешней среды) конкретного человека до появления первых симптомов;
  - превентивная медицина – проведение профилактических мероприятий в отношении возможных, предсказанных заболеваний до возникновения первых проявлений;
  - персонализированная медицина – выбор лечебных воздействий с учетом индивидуальных (генетических) особенностей конкретного человека;
  - партисипативная медицина – активное участие человека в профилактике возможных заболеваний и их лечении; в определенной степени это связано с прохождением диспансерных осмотров, с использованием возможностей m-Health и со здоровым образом жизни, при учете выявленных индивидуальных факторов риска.
- К основным преимуществам медицины P4 авторы [27] относят:
- обнаружение болезни на более ранней стадии, когда ее лечение эффективнее и дешевле;
  - разделение пациентов на сходные группы для выбора оптимальной терапии;
  - уменьшение побочных реакций на лекарства путем более эффективной ранней оценки индивидуальных ответных реакций;
  - улучшение отбора новых биохимических показателей, позволяющих контролировать действие лекарственных препаратов;
  - снижение времени, стоимости и частоты отказов при клинических испытаниях новых методов лечения.

Персонализированная медицина может:

1. Окончательно перенести акцент с лечения на раннюю профилактику, что будет осуществимо в случае научно обоснованного предвидения заболевания для каждого человека.
2. Обеспечить выбор терапии при исключении в этом процессе проб и ошибок, приводящих к задержке в начале правильного лечения.
3. Способствовать повышению качества жизни пациентов, что является одной из конечных целей лечения.
4. Обеспечить рентабельность лечения.

### 3.4. Интегрированная электронная медицинская карта

Особое место в системе здравоохранения занимает интегрированная электронная медицинская карта (ИЭМК), позволяющая перейти на единый формат структурированных электронных медицинских документов в медицинских организациях и являющаяся совокупностью электронных персональных медицинских записей, относящих-

ся к одному человеку, собираемых и используемых несколькими медицинскими организациями. В будущем история всех медицинских действий с самого рождения человека будет храниться в электронной базе. Инициатива ввести EHR (Electronic Health Records) для каждого человека исходила от президента США Джорджа Буша-младшего еще в 2004 году [28]. Электронная медицинская карта – это система, которая собирает информацию с ряда источников. Она заполняется с рождения и пополняется в каждой медицинской организации, которая осуществляет наблюдение, обследование, лечение и реабилитацию пациента (включая оказание экстренной помощи) в течение всей его жизни. Массивы персональных данных электронных медицинских карт, интегрируемых за все периоды жизни людей, должны создаваться в первую очередь в целях предоставления врачам любых медицинских организаций необходимой им информации в любое время и в любом месте. Здесь находится сведения о диагнозах пациента, принимаемых им лекарствах, текущих проблемах со здоровьем, пройденных процедурах, результатах тестов и клинических замечаниях. Данный тип медкарты способен рассылать уведомления пациентам о необходимости прохождения нового обследования. В EHR также предусмотрен контроль за выполнением пациентом врачебных предписаний. Это яркий пример прогнозной аналитики в здравоохранении. Эксперты Deloitte Center прогнозируют, что в ближайшие годы благодаря гаджетам пациенты будут знать о своем здоровье практически все и смогут лично участвовать в выборе оптимального лечения.

ИЭМК пациентов с подключенными к Интернет носимыми smart-устройствами, контролирующими жизненно важные показатели здоровья, позволяет врачам наблюдать за больным в режиме реального времени.

Благодаря мобильным технологиям и интернету вещей пациенты могут собирать данные о своем здоровье и делиться ими со своим врачом (например, приложения [HealthKit](#) от Apple или [Health Vault](#) от Microsoft).

### 3.5. Big Data в фармацевтике

Создание, испытание и промышленное производство лекарств очень дорого стоит. Разработка нового препарата стоит в среднем \$500 млн, в эту сумму входит также обработка большого количества данных: о побочных эффектах, фармакологических свойствах [29]. Компании, которые разработали за 10 лет более шести новых лекарств, тратили на разработку и исследования в среднем \$5,8 млрд [30].

При этом цель далеко не всегда оправдывает средства. Ошибки в расчетах, затягивание иссле-

дований и другие неприятности временами заставляют фармацевтические компании терпеть серьезные убытки. Технология Big Data способна существенно облегчить их работу и позволяет [31]:

- ускорить процесс клинических испытаний (компании смогут искать в базах данных наиболее подходящих для участия в медицинских экспериментах добровольцев в режиме реального времени);
- наладить эффективную коммуникацию (данные можно не только собирать, но обмениваться ими, получая таким образом новые сведения и поддерживая связи с другими фирмами или независимыми учеными);
- улучшить продажи фармацевтической продукции (например, анализ Big Data позволит компаниям вычлнять среди своих потребителей тех, кто больше других нуждается в их разработках);
- прогнозировать эффективность лекарств (с помощью предсказательного моделирования фармацевты смогут подбирать препараты для пациента на основе его генетики, истории болезни и образа жизни).

Особую актуальность технологии Big Data приобрели в процессе анализа молекулярных данных, получаемых путем изучения структуры молекул ДНК (геномика и эпигеномика), молекул РНК (транскриптомика), белковых молекул (протеомика), взаимодействия клеточных метаболитов (метабономика) и других типов подобных данных, где скорость генерации информации в разы превышает скорость ее обработки [21].

Наличие данных о ДНК пациента позволяет более эффективно назначать ему лечение, что значительно снижает затраты на фармацевтическое обеспечение и терапию. Согласно исследованиям ученых UniSA, фармакогенетическое (PGx) тестирование позволит сэкономить порядка \$2,4 млрд ежегодно, и это только в Австралии [9].

Сейчас фармпроизводители стремятся получить доступ к медицинским данным пациентов. Например, фармхолдинг [Roche](#) приобрел стартап [Flatiron Health](#), специализирующийся на сборе клинических данных онкологических пациентов, за \$2 млрд. Исследования в области диабета проводят компании AstraZeneca и Sanofi, а Pfizer и Bristol-Myers Squibb занялись исследованиями в сфере профилактики инсульта – и все это на основе сбора и использования реальных данных о заболеваниях [9].

Компании, специализирующиеся на аналитике больших данных, для открытия новых лекарственных препаратов используют алгоритмы анализа в облаке. Так, компания Numerate создает платформу, которая включает в себя програм-

мы, специально предназначенные для лечения сердечно-сосудистых и нейродегенеративных заболеваний. В совокупности с целенаправленным и индивидуальным подходом к процессу лечения пациенты будут принимать только эффективные лекарства, которые позволят им быстрее пойти на поправку. Это означает, что они больше не будут тратить средства на препараты, которые не оказывали или не окажут должного эффекта.

Большие данные также применяются в прогнозировании побочных эффектов для конкретных соединений и компонентов еще до начала клинического испытания. В настоящее время существует метод, который предсказывает общую токсичность препарата. Ранее в ходе испытаний на людях интоксикация могла быть обнаружена слишком поздно. Используя метод Проктора, который включает анализ 48 различных характеристик лекарственных препаратов, компании могут сэкономить время, деньги и сохранить жизни [29].

Теперь фармацевтические компании могут использовать Big Data для совместной работы со страховыми компаниями, фирмами по управлению базами данных и учеными, работающими в различных учреждениях. Обмениваясь информацией со страховыми компаниями и провайдерами/поставщиками внутри своей сети, фармацевтическая компания может расширить свою базу данных для обеспечения будущих клинических испытаний и предсказательного моделирования.

### 3.6. Big Data в биомедицине

Число пациентов с тяжело излечимыми или неподдающимися лечению диагнозами, такими как сахарный диабет и рак, постоянно растет. Сегодня во всем мире от сахарного диабета страдают 400 млн человек [32], и их число постоянно растет, а в 2017 году заболели раком 24,5 млн человек (16,8 млн без учета немеланомного рака кожи) и 9,6 млн человек скончались от онкологических заболеваний. За последние 10 лет заболеваемость раком в мире выросла на 33% [33]. По прогнозам американского National Cancer Institute, всего через 10-15 лет у 23,6 млн американцев ежегодно будут диагностировать рак. По мере роста населения и распространения рака затраты на терапию пациентов также увеличиваются: национальные расходы на лечение рака в Соединенных Штатах только в 2017 году составили 147,3 млрд долларов. В России полный курс каждого ракового пациента обходится в несколько миллионов рублей, по данным "Коммерсантъ". При этом далеко не всегда медицина оказывается сильнее болезни: статистика показывает, что за 2017 год рак забрал жизни 289 тысяч россиян [11].

**Геномика человека.** Геномика занимается расшифровкой и определением последовательности нуклеотидов в ДНК (дезоксирибонуклеиновая кислота). Методы полного геномного секвенирования генерируют такой большой объем данных, содержащих информацию об отдельных участках генома, что проблемой становится не только их обработка, но и запись на информационный носитель и передача копий данных в другую лабораторию. Все гены 20000 данных, полученных при изучении одной-единственной опухоли (на самом деле больше), вместо 20-30 (и это если брать по максимуму) параметров, определяются классическими методами. Это очень большая проблема, но это ещё не всё. В последние годы стремительно развивается технология «по клеточного» анализа, позволяющего исследовать уровень активности генов не в образце, а в каждой конкретной клетке. С помощью этого метода с одного образца ткани можно собрать уже не 20000 данных, а  $20000 \cdot N$ , где  $N$  – количество клеток, причем не обязательно раковых. А речь идет ещё об одном пациенте и одном-единственном образце опухоли. А таких пациентов в мире (потенциально) миллионы и миллионы, и опухоль у каждого может быть не одна. Так рождаются огромные таблицы, бесконечные ряды данных [34].

Эти данные поступают в разных форматах, которые традиционно обрабатываются разными программами. Разработанный в Гарвардской Медицинской Школе «Персональный геномный проект» (PGP), использующий большие данные, поставил перед собой цель секвенирования и публикации готовых геномов и медицинских документов 100000 добровольцев, чтобы направить исследования на персональную геномику и персональную медицину [35]. В рамках проекта планируется выполнить расшифровку персональных геномов множества добровольцев, и их генетические данные вместе с записями из медицинских карт будут доступны для дальнейшего анализа [36].

Методы компьютерного анализа в сравнительной геномике технологии Big Data применяются в биоинформатике. Проект GenePool занимается секвенированием и выявлением биомаркеров и систематизирует весь огромный поток полученных данных [37, 38]. Применение технологии Big Data, расшифровывающей этот массив данных, и выявляет закономерности в геномном мире. С помощью технологий Big Data ученые проанализировали геномы многих тысяч человек – больных и здоровых. Основываясь на анализе огромного количества информации генетического кода, система подбирает лечение. Так, болезнь Альцгеймера зависит от генов, участвующих в



развитии рассеянного склероза и паркинсонизма [36, 39].

Исследования ассоциации генома (Genomwide association) представляют собой первые крупные попытки связать большие геномные данные с фенотипом и сделать медицину более точной путем обнаружения вариантов, которые влияют на заболевание и метаболизм лекарственных средств. Изучение ассоциации генома – подход, включающий в себя быстрое сканирование маркеров по всем наборам ДНК или геномов многих людей для поиска генетических вариаций, связанных с конкретным заболеванием. После выявления новых генетических ассоциаций исследователи могут использовать эту информацию для разработки лучших стратегий выявления, лечения и профилактики заболевания. Эти информации особенно полезны при поиске генетических вариаций, которые способствуют сложным заболеваниям, таким как астма, рак, диабет, сердечные и психические. Например, архив изображений рака, связанный с Атласом генома рака, включает в себя 31 сайт и более 3 млн изображений, загруженных за первый год работы. Стандартные подходы основаны на статистических моделях, таких как логистическая регрессия и линейные смешанные модели [26], с акцентом на исправление потенциальных смещений и смещающих факторов и приведение шкалы данных.

Сотрудники Национального медицинского исследовательского центра имени В.А. Алмазова разработали математическую модель, которая поможет быстро и эффективно находить мутации в индивидуальных геномах при анализе большого количества образцов. Ее можно применять как в медицине, чтобы выявлять генные мутации, вызывающие предрасположенность к различным заболеваниям, так и в других областях, например, в сельском хозяйстве для селекционной работы. Ученые протестировали разработанный метод на синтетическом наборе данных проекта «1000 геномов» и успешно определили 97% мутаций, присутствующих менее чем у 10% носителей [40]. Расшифровка генома человека заняла 10 лет. Теперь это можно сделать за 1 неделю.

В Германии уже сегодня благодаря технологиям больших данных онкологические заболевания либо предрасположенность к ним выявляются по анализу крови пациентов и доноров. В результате своевременной диагностики существенно снижаются затраты государства и самих людей, а также невероятно повышается эффективность лечения. Ведь один из самых главных врагов пациента, запустившего болезнь, – время. Обратимся к упомянутой ранее онкологии. Диагностика и подбор нужной схемы лечения может

забрать драгоценные минуты, которые так важны в оперативном реагировании при обнаружении злокачественных образований [41].

**Микробиом человека.** Другим направлением биомедицины, развитие которого невозможно без применения подходов и технологий Big Data, является исследование микробиома. В США проект по исследованию микробиома человека «Human Microbiome Project» был запущен одновременно с известным проектом по исследованию генома человека Human Genome Project. В ходе его реализации в рамках Национальных институтов здоровья США создан специальный центр Data Analysis and Coordination Center. Реализуется совместный китайско-европейский проект MetaHit, где ведутся активные исследования в этом направлении [10].

В результате анализа Big Data становится возможным выявить такие неожиданные взаимосвязи или закономерности, которые человек не в состоянии обнаружить. Так, технологии Big Data позволяют одновременно обрабатывать базы историй болезней пациентов, геномных данных и отчетов медицинских исследований и в итоге выдавать полезную информацию для принятия наилучшего решения относительно лечения конкретного пациента. Особую актуальность технологии Big Data приобрели в процессе анализа молекулярных данных, получаемых путем изучения структуры молекул ДНК (геномика и эпигеномика), молекул РНК (транскриптомика), белковых молекул (протеомика), взаимодействия клеточных метаболитов (метаболомика) и других типов подобных данных, где скорость генерации информации в разы превышает скорость ее обработки. Это, в свою очередь, обозначило одно из наиболее перспективных направлений использования технологий Big Data в медицине, так как решение проблемы эффективного анализа данных предоставляет значительное преимущество в ранней диагностике онкологических заболеваний.

Наиболее остро необходимость новых программно-технических средств, опирающихся на методы анализа больших объемов данных, наблюдается в биоинформатике и биомедицине.

Традиционные алгоритмы анализа данных не справляются с поставленными перед ними задачами.

### 3.7. Big Data в психиатрии

Первые исследования с использованием технологий больших данных показали хорошие результаты и в области психиатрии. Одним из пионеров применения Big Data в психиатрии стала американская ученая индийского происхождения Мунмун Де Чаудхури (Munmun De Choudhury) [42]. Вместе с коллегами она использовала про-

гностическую модель для анализа неструктурированных открытых данных из социальных сетей для определения распространения депрессии в США.

В программу статистического анализа с алгоритмом машинного обучения загрузили сообщения из Twitter, Facebook и Reddit с геометками. Посты отбирались по признакам присутствия в них слов, указывающих на депрессивное состояние пользователей.

Данные, полученные в 50 штатах, сравнили с показателями Центров по контролю и профилактике заболеваний США (федеральное агентство при министерстве здравоохранения). Результаты исследования практически совпали с официальной статистикой.

Другой пример успешного применения анализа больших данных в психиатрии – проект World Well Being, или «Мировое здоровье». В рамках одного из экспериментов исследовались 148 миллионов твитов, содержащих негативно-окрашенные слова и выражения. Данные анализа сопоставили с официальной статистикой смертности от сердечно-сосудистых заболеваний. Корреляция оказалась даже выше, чем при сравнении цифр смертности с другими статистическими данными – социально-экономического положения регионов, демографических показателей и состояния здравоохранения. Перспективу использования цифровых технологий в психиатрии признали во Всемирной организации здравоохранения.

### **3.8. Big Data в эпидемиологии**

Большие данные, помимо уже известных и распространенных задач, можно использовать в том числе для борьбы с заболеваниями и отслеживания роста эпидемий. Применение технологии Big Data в эпидемиологии позволяет строить как географические и социальные модели здоровья населения, так и предиктивные модели развития эпидемических вспышек.

Так, еще за девять дней до того, как вспышка инфекционных заболеваний, таких как вирус Эбола, была официально объявлена эпидемией, группа исследователей и ученых из Бостона при помощи больших данных смогла обнаружить распространение лихорадки в Гвинее [43].

Картину движения эпидемии смертельного вируса по Западной Африке составил стартап HealthMap, работающий на базе алгоритма, который учитывает упоминания в социальных медиа, сводки местных новостей и другие данные, доступные в Сети.

Системы больших данных могут оказаться полезными в первую очередь не для обнаружения уже проявившихся вспышек тех или иных заболеваний, а для предсказания потенциально воз-

можных эпидемий такого рода благодаря анализу доступной информации. В этом случае практически те же технологии, которые помогают маркетологам демонстрировать потребителям таргетированную рекламу или предлагать музыку и видео для просмотра, могут быть использованы для борьбы против инфекционных заболеваний, таких как Эбола.

А вот в случае с пандемией коронавируса Covid-19 этого не произошло.

11 марта 2020 года ВОЗ объявила о пандемии нового коронавируса Covid-19, который в декабре 2019 был впервые обнаружен в китайском мегаполисе Ухань. С тех пор вирус стремительно распространяется по всей планете, вызывая острые респираторные заболевания.

По данным ВОЗ на 14 апреля 2020 года, эпидемия коронавирусной инфекции затронула уже 213 стран и территорий [44], количество заразившихся коронавирусом в мире превысило 2 млн человек. По данным Университета имени Джона Хопкинса, оно составило 2 019 320 человек, умерли 119 483 человека [45]. В результате пандемии почти во всех странах объявили агрессивный карантин, в разных странах на разные сроки, вследствие чего был нанесен колоссальный экономический ущерб.

И в случае пандемии Data Science демонстрировал своё превосходство, так как помогал бороться с вирусом с момента его возникновения. В частности, Big Data система машинного обучения для мониторинга здоровья BlueDot зафиксировала вспышку пневмонии, вызванной коронавирусом еще 31 декабря 2019 года, т.е. раньше специалистов ВОЗ [46]. Также искусственный интеллект (ИИ) BlueDot агрегирует и косвенные данные: посты в блогах и на форумах, а также брони авиабилетов по всему миру. Благодаря этому система верно определила не только место возникновения эпидемии, но и пути ее дальнейшего распространения: от китайского Уханя в Бангкок, Сеул, Тайбэй и Токио. Machine Learning и другие методы Data Science используются для разработки вакцины и лекарств от коронавируса. Например, ИИ-система AlphaFold исследовательского подразделения DeepMind корпорации Google предсказала возможную структуру некоторых белков COVID-19, используя технологию матричного моделирования. Точное знание компонентного и структурного состава коронавируса позволит синтезировать вещества для лечения заболеваний, вызванных этой инфекцией. Пока еще полученные исследования не проверены экспериментально, но они помогут ученым понять, как функционирует вирус, чтобы создать гипотезы по эффективной терапии.

Подавление вспышки COVID-19 в Китае заключалось в агрессивном карантине, жестком обеспечении изоляции и инвазивном наблюдении с помощью передовых инструментов Big Data и искусственного интеллекта. Другой показательный пример использования Artificial Intelligence в борьбе с Covid-19 – это автоматическая диагностика заболевших с помощью алгоритмов Machine Learning для распознавания данных компьютерной томографии. В Китае также был создан подвижный робот для бесконтактного измерения температуры тела. Благодаря интеграции этого устройства с национальной системой распознавания лиц и наличию технологии 5G, оно способно выявить отсутствие медицинской маски и сделать личное предупреждение. Этот робот способен одновременно отслеживать целых 32 объекта во время патрулирования улиц, что весьма востребовано в условиях всеобщего карантина в Китае [47].

Технологии распознавания лиц с помощью искусственного интеллекта и другие инструменты Big Data использовались для отслеживания контактов и управления приоритетными группами населения. С применением метода анализа больших данных в Китае во время распространения пандемии выслеживали каждого пассажира не только в городских, но и в междугородних автобусах, поездах, самолетах и в любых иных типах транспорта [46].

### **3.9. Big Data в телемедицине**

Еще один пример того, как работает Big Data в здравоохранении – развитие телемедицины. Она присутствует на рынке медицинских услуг более 40 лет, но только сегодня, с появлением смартфонов, беспроводных переносимых устройств и видеоконференций, телемедицина смогла полностью раскрыться. Под телемедициной следует понимать предоставления клинических услуг на расстоянии (например, мониторинг состояния пациента и консультации) и взаимодействие медицинских работников между собой с помощью телекоммуникационных технологий. При том телемедицина допускает не только общение тет-а-тет с квалифицированным специалистом в дистанционном режиме, но и самодиагностирование через посещение сервисов вроде webmd.com. Этот вид услуг предоставляет Healthtap [48].

С развитием телемедицины человек экономит время и силы, потому что может пообщаться со специалистом онлайн. Это актуально для жителей мегаполисов, у которых часто не хватает времени следить за своим здоровьем и сидеть в очередях. Но телемедицинские технологии – это еще и выход для людей, живущих в сельской местности, потому что, как правило, высококвалифицированные врачи работают в городах. Эта

технология также направлена на создание приложений, где нейросети на основе симптомов будут ставить диагнозы и назначать лечение.

Ещё одно из главных преимуществ телемедицины – возможность обеспечить уход физически немощным пациентам, а также тем, кто уязвим к инфекциям. Кроме того, дистанционная работа с пациентами повышает и качество медицинской помощи: лечащие врачи имеют постоянный доступ к точным данным о состоянии здоровья пациента и могут оперативно дать экспертные рекомендации при его внезапном ухудшении.

Поэтому спрос на дистанционное здравоохранение растет невероятными темпами: по данным Cleveland Clinic Innovation, 90% медицинских организаций в настоящее время строят или готовятся внедрять программу телездоровоохранения.

Легализация телемедицины и бум телекоммуникаций стимулировали развитие отрасли: глобальный рынок решений для дистанционного взаимодействия с пациентами будет расти со среднегодовой скоростью 16,2% в следующие 4 года и к концу периода достигнет 18,2 млрд долларов [11].

Консолидация непрерывно генерируемых в процессе отслеживания состояния пациента медицинской информации с данными EHR и одновременный анализ огромного объема разнородных данных в масштабе реального времени предоставляют профессионалу комплексную связную картину общего состояния здоровья пациента. Это, в свою очередь, будет способствовать поддержке принятия обоснованных решений по дистанционному диагностированию больного в режиме реального времени. Согласно прогнозам [24], возможность выработки в режиме реального времени аналитики меняющихся данных большого объема по всем направлениям медицины может произвести революцию в этой сфере.

### **4. Перспективные направления применения технологий Big Data в медицине и здравоохранении**

Хотя потенциал анализа Big Data является многообещающим и перспективным направлением, существует ряд проблем, связанных с применением Big Data в медицине. Аналитики и практики видят причину этого в недостаточной адаптации имеющихся на рынке платформ Big Data к специфическим особенностям медицинской отрасли, требованиям к конфиденциальности и безопасности медицинских данных, стандартизации врачебных решений. Четких доказательств практической пользы (в том числе и экономической эффективности) использования Big Data мало. Существует ряд методологических вопросов, таких как низкое качество, нестабильность и

неструктурированность, синтаксические и семантические проблемы данных, проверка их достоверности, мониторинг безопасности в режиме реального времени, происхождение данных, обилие лишней информации, валидация, аналитические, правовые и этические вопросы, продолжающиеся дискуссии о возможности использования Big Data в соответствии с критериям доказательной медицины. Часто в результате анализа Big Data выявляются связи между событиями, которые на самом деле не могли никак повлиять друг на друга, при этом число ложных корреляций увеличивается с количеством анализируемых данных [49, 50]. Необходимо приложить дальнейшие усилия для улучшения качества и унификации электронных медицинских карт пациентов. Существует значительный риск искажения информации (в частности, проект Google Flu Trends) [8]. Традиционных механизмов безопасности, таких как брандмауэры и антивирусное программное обеспечение, устанавливаемых на компьютерах, недостаточно для эффективной защиты Big Data.

Динамика развития технологий и сервисов Big Data и прогнозы аналитиков усиливают уверенность в том, что в ближайшем будущем будет наблюдаться значительный рост внедрения Big Data-решений в медицину. В этом контексте стоит отметить перспективные направления применения технологий анализа больших объемов информации в сфере информатизации медицины и практического здравоохранения:

1. Прежде всего, это интеграция геномной информации с электронными историями болезни, которая приобретает особую актуальность для развития персонифицированной медицины, предполагающей эффективно организованный доступ к любой совокупности медицинских записей и первичных результатов исследований пациента. Персонализированная медицина – новое направление, способное значительно улучшить медицинское обслуживание детей и взрослых и учитывать при назначении терапии индивидуальную чувствительность пациента к лекарствам.

2. Предполагается, что в будущем геномика, являющаяся хранилищем биообразцов, наряду с медицинскими изображениями и мониторингом на основе датчиков физиологических систем, а также электронные медицинские карты будут представлять самые большие объемы данных. При этом исследователю, который будет обрабатывать эти данные, необходимо уметь извлекать из него конкретные участки, делать необходимые запросы, применять к их результатам программные инструменты биоинформатики. В таких запросах могут участвовать десятки тысяч

файлов и десятки инструментов, передающих друг другу результаты своей работы.

3. Перспективным направлением является совмещение информации о генотипе и фенотипе. Как показывает практика, это направление очень выгодно с исследовательской точки зрения.

4. Цифровизация медицины — ещё одно перспективное направление, которое не только унифицирует работу клиник или лабораторий, но и может спасти человеческие жизни. Обработка и хранение Big Data позволяют делать постановку диагнозов точнее, проверять медицинские данные без нагрузки на врачей, а также интегрировать результаты исследований, выполненных на разных устройствах в общую систему.

5. Современные компьютерные технологии оказывают существенную практическую помощь в проведении эпидемиологического надзора, повышая его оперативность, качество эпидемиологической диагностики, обеспечивая своевременность принятия адекватных управленческих решений и проведение профилактических и противоэпидемических мер. Современный этап развития эпидемиологического надзора ставит перед специалистами новые серьезные задачи, ориентированные на потенциал охраны общественного здоровья, готовность к чрезвычайным ситуациям, эпиднадзора и реагирования, решение которых является залогом успешной борьбы и профилактики инфекционных болезней. Для решения проблем современной эпидемиологии необходимо усовершенствование хранения и конфиденциальности данных в рамках новых видов эпидемиологических исследований.

6. Изменение способов взаимодействия врачей и пациентов, доступа последних к медицинским услугам, результатам лечения и медицинской информации, а также подготовка кадров для работы с большими медицинскими данными и предоставление медицинских услуг в режиме реального времени

Не подлежит сомнению, что технологии Big Data могут оказать существенное воздействие на медицину и здравоохранение, причем как в плане повышения точности диагностики, эффективности лечения и своевременного предупреждения болезней, так и в контексте улучшения работы самих медицинских учреждений. Ожидается, что с помощью применения современных аналитических систем эффективность лечения будет повышена благодаря обработке всей доступной информации. В практику врачей войдет широкое использование систем поддержки принятия решений и экспертных систем нового поколения, позволяющих предоставить врачам всеобъемлющий доступ к опыту коллег посредством анализа электронных медицинских паспортов паци-

ентов как в локальном варианте, так и в масштабе страны. Это, в свою очередь, позволит минимизировать субъективный человеческий фактор при принятии врачебных решений о стратегии лечения пациента [24].

#### **Выводы**

В сфере цифровой медицины, в том числе в клиниках и исследовательских центрах мирового уровня, до сих пор много пробелов и уязвимых зон. Но практика показывает: технологизация даже тех процессов, которые на первый взгляд ей не поддаются, а также выстраивание целостной системы сбора и обмена информацией уже сегодня помогают нащупать точки роста, которые определяют, каким будет здравоохранение в обозримом будущем. Обработка Big Data в медицине представляет собой перспективный процесс исследования и анализа большого количества сложных гетерогенных данных различной природы: биомедицинских, электронных медицинских карт, фармацевтических, правовых, страховых, публикаций в социальных сетях и других. Внедрение технологий Big Data приведет к разработке оптимальных способов обработки, анализа и извлечения полезных знаний из этих больших объемов медицинских данных. Вследствие этого Big Data сможет улучшить качество обслуживания пациентов, диагностировать различную патологию на ранней стадии, прогнозировать и предотвращать развитие заболеваний, возникновение эпидемий, предоставлять наиболее эффективные персонализированные методы лечения, контролировать качество работы медицинских учреждений.

#### **Литература:**

1. <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>
2. <https://habr.com/ru/company/mailru/blog/449370/>
3. <https://ww2.frost.com/news/press-releases/>
4. *Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC.*
5. [http://www.tadviser.ru/index.php/Статья:Большие\\_данные\\_\(Big\\_Data\)\\_мировой\\_рынок](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)_мировой_рынок)
6. *Manchini M.* Exploiting Big Data for improving healthcare services // *Journal of e-Learning and Knowledge Society.* 2014. Vol.10, n.2. P.23–33.
7. *Hackenberger B.K.* Data by data, Big Data. *Croat. Med. J.* 2019. Vol. 60, № 3. P. 290-292.
8. *Петров В.В., Минцер О.П., Крючин А.А., Крючина Е.А.* Перспективы и проблемы использования технологий big data в медицине. *Медицина информатика та інженерія.* 2019. № 3. С. 19-30.
9. <https://www.forbes.ru/obshchestvo/371437-doktor-google-glavnye-trendy-medicinskogo-rynka>
10. *Цветкова Л.А., Черченко О.В.* Внедрение технологий big data в здравоохранение: оценка технологических и коммерческих перспектив // *Экономика науки.* 2016. Том 2, № 2. С. 138-150.

11. <https://www.comnews.ru/digital-economy/content/116081/2018-11-28/zhit-dolshe-iot-chat-boty-i-big-data-pomogut-tyazhelobolnym-pacientam>
12. <https://www.ibm.com/watson/health/>
13. *Зелинский С.С., Удуд Е.А., Новрузова В.И.* BIG DATA в медицине. Направления использования стимулирование инновационного развития общества в стратегическом периоде // *Сборник статей Международной научно-практической конференции.* Уфа. 2018. Издательство: ООО "Аэтерна".
14. <https://healthitanalytics.com/news/qpid-epic-ehr-combine-for-big-data-analytics-at-partners-health>
15. *Ioannidis JPA, Houry MJ.* Evidence-based medicine and big genomic data. *Hum Mol Genet.* – 2018 May 1;27(R1):R2-R7.
16. *Grüning B., Lampa S., Vaudel M., Blankenberg D.* Software engineering for scientific big data analysis *Gigascience.* 2019. May 1;8(5). pii: giz054.
17. *Abidi S. S. R.* Intelligent health data analytics: A convergence of artificial intelligence and big data. *Health Manage Forum.* 2019. Vol. 32. № 4. P. 178-182.
18. *Luo J., Wu M., Gopukumar D., Zhao Y.* Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights.* 2016. № 8. P. 1-10.
19. *Karen Y. He, Dongliang Ge, Max M.* Big Data Analytics for Genomic Medicine. *Int. J. Mol. Sci.* 2017. Vol. 18. № 2. P. E412.
20. *Cobb A. N., Benjamin A. J., Huang E. S., Kuo P. C.* Big data: More than big data sets. *Surgery.* 2018. Vol. 164, № 4. P. 640-642.
21. *Карнаухов Н. С., Ильяхин П. Г.* Возможности технологий «Big Data» в медицине. *Врач и информационные технологии.* 2019, № 1. С. 59-63.
22. *Магеррамов З.Т., Абдуллаев В.Г., Магеррамова А.З.* Big Data: проблемы, методы анализа, алгоритмы // *Радиоэлектроника и Информатика.* № 3(78), 2017. С. 42-51. <http://openarchive.nure.ua/handle/document/5653>
23. *Herman R, Williams P.* Big Data in healthcare: what is it used for? <http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1021&context=acis>
24. *Мамедова М.Г.* Big Data в электронной медицине: возможности, вызовы и перспективы. *İnformasiya texnologiyaları problemləri.* 2016. №2. С. 9–29.
25. *Маркина Н.В., Касюк С.Т., Шамаева Т.Н.* Анализ данных в медицинских информационных системах с использованием технологии Data Mining. *Информатика, вычислительная техника и управление.* Серия «Естественные и технические науки». 2019. №6. С. 111–116.
26. *Кобринский Б.А.* Персонализированная медицина: геном, электронное здравоохранение и интеллектуальные системы. Часть 1. Геномика и мониторинг клинических данных. *Рос вестн перинатол и педиатр* 2017. 62:(5): 16–20. DOI: 10.21508/1027-4065-2017-62-5-16-20.
27. *Hood L., Galas D.* P4 Medicine: Personalized, Predictive, Preventive, Participatory: A Change of View that Changes Everything: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. 2008. <http://cra.org/ccc/resources/ccc-ledwhitepapers>
28. <http://ru.datasides.com/big-data-medicine/>
29. <https://gmpnews.ru/2017/06/ispolzovanie-big-data-v-farmaceuticheskoy-industrii/>

30. <https://www.forbes.ru/biznes/355471-vremya-innovaciy-kak-obuzdat-ceny-na-lekarstva>
31. <https://iotconf.ru/ru/article/kak-big-data-pomoget-farmatsevtike-69841>
32. <https://www.if24.ru/diabet-pod-kontrol-berut-tehnologii/>
33. <http://www.oncology.ru/news/2019/10/14/>
34. Кондратова М.С. Кривое зеркало жизни: Главные мифы о раке, и что современная наука думает о них. 2019. 310 с.
35. [https://ru.wikipedia.org/wiki/Персональная\\_геномика](https://ru.wikipedia.org/wiki/Персональная_геномика)
36. Мурадова Г.И. Большие данные в системе здравоохранения // *Informasiya tehnologiyalari problemləri*. 2016. №2. С. 98–106.
37. How GenePool Works. [www.stationxinc.com/how-genepool-works](http://www.stationxinc.com/how-genepool-works)
38. Научно-популярный портал о генетике <http://mygenome.su/articles/112/>
39. Bettens K., Slegers K., Broeckhoven C., Genetic insights in Alzheimer's disease/ *The lancet neurology*, 2013, vol.12, no.1, pp.92–104.
40. [https://www.gazeta.ru/science/news/2019/03/12/n\\_12738421.shtml](https://www.gazeta.ru/science/news/2019/03/12/n_12738421.shtml)
41. [https://www.cnews.ru/articles/dmitrij\\_shepelyavyj\\_bolshie\\_dannye\\_na](https://www.cnews.ru/articles/dmitrij_shepelyavyj_bolshie_dannye_na)
42. <https://aboutdata.ru/2017/05/29/big-data-in-psychiatry/>
43. [https://www.cnews.ru/news/top/bolshie\\_dannye\\_mogut\\_pomoch\\_v\\_borbe](https://www.cnews.ru/news/top/bolshie_dannye_mogut_pomoch_v_borbe)
44. <https://ren.tv/news/v-mire/685999-voz-koronavirus-zatronul-213-stran-i-territorii>
45. <https://www.vedomosti.ru/society/news/2020/04/14/827957-dva-milliona>
46. <https://review.uz.ru/post/big-data-i-machine-learning-protiv-covid-19>
47. <https://nv.ua/techno/innovations/koronavirus-v-kitae-sozdali-robotov-dlya-borby-s-epidemiej-50072122.html>
48. Ижунин М.А. Big Data в здравоохранении. // Молодой ученый. Международный научный журнал. 2019. № 50 (288). С. 8-10.
49. Fahr P., Buchanan J., Wordsworth S. A Review of the Challenges of Using Biomedical Big Data for Economic Evaluations of Precision Medicine. *Appl. Health Econ Health Policy*. 2019. 3 p.
50. Xu L., Wang S., Zhan S. Randomized controlled trial based on big data. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2019. Vol. 40, № 6. P. 702-706.
- Transliterated bibliography:**
- <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>
  - <https://habr.com/ru/company/mailru/blog/449370/>
  - <https://ww2.frost.com/news/press-releases/>
  - Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC.
  - [http://www.tadviser.ru/index.php/Статья:Большие\\_данные\\_\(Big\\_Data\)\\_мировой\\_рынок](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)_мировой_рынок)
  - Manchini M. Exploiting Big Data for improving healthcare services // *Journal of e-Learning and Knowledge Society*. 2014. Vol.10, n.2. P.23–33.
  - Hackenberger B.K. Data by data, Big Data. *Croat. Med. J.* 2019. Vol. 60, № 3. P. 290-292.
  - Petrov V.V., Mincer O.P., Krjuchin A.A., Krjuchina E.A. Perspektivy i problemy ispol'zovaniya tehnologij big data v medicine. *Medichna informatika ta inzhenerija*. 2019. № 3. S. 19-30.
  - <https://www.forbes.ru/obshchestvo/371437-doktor-google-glavnye-trendy-medicinskogo-rynka>
  - Cvetkova L.A., Cherchenko O.V. Vnedrenie tehnologij big data v zdravoohranenie: ocenka tehnologicheskikh i kommercheskikh perspektiv // *Jekonomika nauki*. 2016. Tom 2, № 2. С. 138-150.
  - <https://www.comnews.ru/digital-economy/content/116081/2018-11-28/zhit-dolshe-iot-chat-boty-i-big-data-pomogut-tyazhelobolnym-pacientam>
  - <https://www.ibm.com/watson/health/>
  - Zelinskij S.S., Udud E.A., Novruzova V.I. BIG DATA v medicine. Napravlenija ispol'zovaniya stimulirovanie innovacionnogo razvitija obshhestva v strategicheskome periode // *Sbornik statej Mezhdunarodnoj nauchno-prakticheskoy konferencii*. Ufa. 2018 Izdatel'stvo: OOO "Ajeterna".
  - <https://healthitanalytics.com/news/qpid-epic-ehr-combine-for-big-data-analytics-at-partners-health>
  - Ioannidis JPA, Houry MJ. Evidence-based medicine and big genomic data. *Hum Mol Genet.* – 2018 May 1;27(R1):R2-R7.
  - Grüning B., Lampa S., Vaudel M., Blankenberg D. Software engineering for scientific big data analysis *Gigascience*. 2019. May 1;8(5). pii: giz054.
  - Abidi S. S. R. Intelligent health data analytics: A convergence of artificial intelligence and big data. *Healthc Manage Forum*. 2019. Vol. 32. № 4. P. 178-182.
  - Luo J., Wu M., Gopukumar D., Zhao Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights*. 2016. № 8. P. 1-10.
  - Karen Y. He, Dongliang Ge, Max M. Big Data Analytics for Genomic Medicine. *Int. J. Mol. Sci.* 2017. Vol. 18. № 2. P. E412.
  - Cobb A. N., Benjamin A. J., Huang E. S., Kuo P. C. Big data: More than big data sets. *Surgery*. 2018. Vol. 164, № 4. P. 640-642.
  - Karnauhov N.S., Il'juhin R.G. Vozmozhnosti tehnologij «Big Data» v medicine. *Vrach i informacionnye tehnologii*. 2019, № 1. С. 59-63.
  - Magerramov Z.T., Abdullaev V.G., Magerramova A.Z. Big Data: problemy, metody analiza, algoritmy // *Radiotelektronika i Informatika*. № 3(78), 2017. С. 42-51. <http://openarchive.nure.ua/handle/document/5653>
  - Herman R, Williams P. Big Data in healthcare: what is it used for? <http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1021&context=aegis>
  - Mamedova M.G. Big Data v jelektronnoj medicine: vozmozhnosti, vyzovy i perspektivy. *Informasiya tehnologiyalari problemləri*. 2016. №2. С. 9–29.
  - Markina N.V., Kasjuk S.T., Shamaeva T.N. Analiz dannyh v medicinskih informacionnyh sistemah s ispol'zovaniem tehnologij Data Mining. *Informatika, vychislitel'naja tehnika i upravlenie*. Serija «Estestvennye i tehnicheckie nauki». 2019. №6. S. 111–116.
  - Kobrinskij B.A. Personalizirovannaja medicina: genom, jelektronnoe zdravoohranenie i intellektual'nye sistemy. Chast' 1. *Genomika i monitoring klinicheskikh dannyh*. *Ros vestn perinatol i pediatri* 2017. 62:(5): 16–20. DOI: 10.21508/1027-4065-2017-62-5-16-20.
  - Hood L., Galas D. P4 Medicine: Personalized, Predictive, Preventive, Participatory: A Change of View that Changes Everything: A white paper prepared for the

- Computing Community Consortium committee of the Computing Research Association. 2008. <http://cra.org/ccc/resources/ccc-ledwhitepapers>
28. <http://ru.datasides.com/big-data-medicine/>
29. <https://gmpnews.ru/2017/06/ispolzovanie-big-data-v-farmaceuticheskoy-industrii/>
30. <https://www.forbes.ru/biznes/355471-vremya-innovacij-kak-obuzdat-ceny-na-lekarstva>
31. <https://iotconf.ru/ru/article/kak-big-data-pomoget-farmatsevtike-69841>
32. <https://www.if24.ru/diabet-pod-kontrol-berut-tehnologii/>
33. <http://www.oncology.ru/news/2019/10/14/>
34. Kondratova M.S. Krivoje zerkalo zhizni: Glavnye mify o rake, i chto sovremennaja nauka dumaet o nih. 2019. 310 s.
35. [https://ru.wikipedia.org/wiki/Персональная\\_геномика](https://ru.wikipedia.org/wiki/Персональная_геномика)
36. Muradova G.I. Bol'shie dannye v sisteme zdavoohraneniya // Informasiya texnologiyalari problemleri. 2016. №2. S. 98–106.
37. How GenePool Works. [www.stationxinc.com/how-gene-pool-works](http://www.stationxinc.com/how-gene-pool-works)
38. Научно-популярный портал о генетике <http://mygenome.su/articles/112/>
39. Bettens K., Slegers K., Broeckhoven C., Genetic insights in Alzheimer's disease/ The lancet neurology, 2013, vol.12, no.1, pp.92–104.
40. [https://www.gazeta.ru/science/news/2019/03/12/n\\_12738421.shtml](https://www.gazeta.ru/science/news/2019/03/12/n_12738421.shtml)
41. [https://www.cnews.ru/articles/dmitrij\\_shepelyavyj\\_bolshie\\_dannye\\_na](https://www.cnews.ru/articles/dmitrij_shepelyavyj_bolshie_dannye_na)
42. <https://aboutdata.ru/2017/05/29/big-data-in-psychiatry/>
43. [https://www.cnews.ru/news/top/bolshie\\_dannye\\_mogut\\_pomoch\\_v\\_borbe](https://www.cnews.ru/news/top/bolshie_dannye_mogut_pomoch_v_borbe)
44. <https://ren.tv/news/v-mire/685999-voz-koronavirus-zatronul-213-stran-i-territorii>
45. <https://www.vedomosti.ru/society/news/2020/04/14/827957-dva-milliona>
46. <https://review.uz/ru/post/big-data-i-machine-learning-protiv-covid-19>
47. <https://nv.ua/techno/innovations/koronavirus-v-kitae-sozdali-robotov-dlya-borby-s-epidemiej-50072122.html>
48. Izhunin M.A. Big Data v zdavoohranenii. // Molodoj uchenyj. Mezhdunarodnyj nauchnyj zhurnal. 2019. № 50 (288). S. 8-10.
49. Fahr P., Buchanan J., Wordsworth S. A Review of the Challenges of Using Biomedical Big Data for Economic Evaluations of Precision Medicine. Appl. Health Econ Health Policy. 2019. 3 p.
50. Xu L., Wang S., Zhan S. Randomized controlled trial based on big data. Zhonghua Liu Xing Bing Xue Za Zhi. 2019. Vol. 40, № 6. P. 702-706.

Поступила в редколлегию 20.03.2020

**Рецензент:** д-р техн. наук, проф. Омаров М.А.

**Магеррамов Закир Тулуевич**, канд. техн. наук, доцент кафедры «Информационная технология и программирования» Азербайджанского технического университета. Научные интересы: численные методы, моделирование и оптимальное управление, информационные технологии, объектно-ориентированное программирование. Увлечения: научные книги, художественная литература (классика), мир животных. Адрес: Азербайджан, AZ1114, Баку, ул. И. Джумшудова, 1/7, кв. 110, тел.

(99412)5689951, (050)3212595, e-mail:

[zakirmaharramov@rambler.ru](mailto:zakirmaharramov@rambler.ru).

**Рагимова Назиля Алиевна**, канд. техн. наук, зав. кафедрой «Компьютерная инженерия» Азербайджанского государственного университета нефти и промышленности. Научные интересы: Cloud technology, Data Mining, экспертные системы, big data. Увлечения: научные книги, художественная литература, спорт. Адрес: Азербайджан, Баку, (050)3136431, e-mail: [ragimova.n.a@gmail.com](mailto:ragimova.n.a@gmail.com).

**Абдуллаев Вугар Гаджимахмудович**, канд. техн. наук, доцент кафедры «Компьютерная инженерия» Азербайджанского государственного университета нефти и промышленности, Институт Кибернетики НАНА. Научные интересы: киберфизические системы, большие данные, информационные технологии. Увлечения: электронная коммерция, B2B, B2C проекты, научные книги, спорт. Адрес: Азербайджан, AZ1129, Баку, ул. М. Гади, 53, кв. 81, тел. (99412)5712428, (050)3325483, e-mail: [abdulvugar@mail.com](mailto:abdulvugar@mail.com).

**Магеррамова Гюльнар Закировна**, стоматолог-эндодонтолог, аспирантка кафедры стоматологии и челюстно-лицевой хирургии Института Усовершенствования врачей имени А.Алиева (Азербайджан). Научные интересы: применение цифровых технологий в медицине, big data. Увлечения: изучение иностранных языков, латино-американские танцы. Адрес: Азербайджан, Баку, ул. Ильгар Джумшудова 1/7, кв.110, e-mail: [maharramovagulnar@gmail.com](mailto:maharramovagulnar@gmail.com).

**Maharramov Zakir Tuluevich**, PhD, Associate Professor of the Department of Information Technology and Programming, Azerbaijan Technical University. Scientific interests: numerical methods, modeling and optimal control, information technology, object-oriented programming. Hobbies: scientific books, fiction (classics), the world of animals. Address: Azerbaijan, AZ1114, Baku, st. I. Dzhumshudova, 1/7, apt. 110, tel. (99412) 5689951, (050) 3212595, e-mail: [zakirmaharramov@rambler.ru](mailto:zakirmaharramov@rambler.ru).

**Ragimova Nazilya Alievna**, PhD, head of Computer Engineering Department, Azerbaijan State University of Oil and Industry. Scientific interests: Cloud technology, Data Mining, expert systems, big data. Hobbies: scientific books, fiction, sports. Address: Azerbaijan, Baku, (050) 3136431, e-mail: [ragimova.n.a@gmail.com](mailto:ragimova.n.a@gmail.com).

**Abdullaev Vugar Gadzhimakhmudovich**, PhD, Associate Professor of Computer Engineering Department, Azerbaijan State University of Oil and Industry, Institute of Cybernetics of ANAS. Scientific interests: cyber-physical systems, big data, information technology. Hobbies: e-commerce, B2B, B2C projects, scientific books, sports. Address: Azerbaijan, AZ1129, Baku, st. M. Gadi, 53, apt. 81, tel. (99412) 5712428, (050) 3325483, e-mail: [abdulvugar@mail.com](mailto:abdulvugar@mail.com).

**Maharramova Gulnar Zakirovna**, stomatologist-endodontologist, post-graduate student of the Department of Dentistry and Maxillofacial Surgery of the Institute for Advanced Training of Doctors named after A. Aliyev (Azerbaijan). Scientific interests: application of digital technologies in medicine, big data. Hobbies: studying foreign languages, Latin American dances. Address: Azerbaijan, Baku, st. Ilgar Dzhumshudova 1/7, apartment 110, e-mail: [maharramovagulnar@gmail.com](mailto:maharramovagulnar@gmail.com)