

Начала искусственного интеллекта для исследований гигиены труда **Principles of Artificial Intelligence for Occupational Health Research**

Сведения об авторе: Федорович Геннадий Викторович,
Технический директор, доктор физико-математических наук.

Место работы: ООО «НТМ-Защита»

About the author: Fedorovich Gennady Victorovich,
Technical Director, Doctor of Philosophy

Place of work: “NTM-Defence”, LLC

Контактная информация:

E-mail: fedorgv@gmail.com

Аннотация

Обсуждаются особенности организации исследований в области гигиены труда (ГТ) с использованием идеологии искусственного интеллекта (ИИ). Идеология ИИ перспективна для использования в качестве рациональной структуры исследований – оптимального паттерна, по которому следует выстраивать систему ГТ: фиксировать цели и задачи отдельных направлений и определить отношения между их результатами. Более сложные и функциональные программы исследований ГТ оптимизированы по производительности, адаптации интерфейса для разных пользователей, а также улучшают распознавание и предотвращение потенциальных угроз безопасности. Современные исследования ГТ приводят к образованию больших массивов медицинских данных (БМД). Для превращения БМД в знания, выводы и мотивы действия требуются специальные методы обучения ИИ на специально подобранных обучающих материалах. Логика ИИ предъявляет специфические требования к структуре БМД, к свидетельствам о силе и направлении причинно-следственных

зависимостей в ГТ. Одним из результатов анализа БМД является обнаружение полиморбидной структуры производственно-обусловленных заболеваний (ПОЗ). Анализ методами ИИ результатов нескольких натурных исследований ПОЗ выявил пары как несвязанных, так и скорелированных нозологий. В последнем случае анализ ситуации на основе логики Т. Байеса дает оценку причинно-следственных отношений между ними. Другая особенность ПОЗ – отклики на внешние производственные воздействия, выводящие организм из гомеостаза, в значительной степени однотипны для основных производственных факторов. Интегральная оценка методами ИИ состояния больного работника должна стать правилом в ГТ.

Abstract

The article discusses the specifics of organizing research in the field of occupational hygiene (OH) using the ideology of artificial intelligence (AI). The AI ideology is promising for use as a rational research structure - an optimal pattern for building an OH system: to record the goals and objectives of individual areas and determine the relationships between their results. More complex and functional OH research programs are optimized in terms of performance, interface adaptation for different users, and improve the recognition and prevention of potential safety threats. Modern OH research leads to the formation of large arrays of medical data (BMD). To transform BMD into knowledge, conclusions and motives for action, special methods of training AI on specially selected training materials are required. AI logic imposes specific requirements on the structure of BMD, on evidence of the strength and direction of cause-and-effect relationships in OH. One of the results of the BAM analysis is the detection of a polymorbid structure of occupational-related diseases (ORD). Analysis of the results of several field studies of the ORD using AI methods revealed pairs of both unrelated and correlated nosologies. In the latter case, the analysis of the situation based on T. Bayes' logic provides an assessment of the cause-and-effect relationships between them. Another feature of the ORD is that responses to external production effects that take the body out of homeostasis are largely uniform for

the main production factors. An integrated assessment of the condition of a sick worker using AI methods should become a rule in OH.

Ключевые слова: искусственный интеллект; гигиена труда; большие данные; управление данными; однотипность патологий; причинно-следственные связи.

Keywords: artificial intelligence; occupational hygiene; big data; data management; uniformity of pathologies; causal relationships.

УДК 004.855.331.45

Введение

Искусственный интеллект (далее – ИИ) – это универсальный термин для программ, которые выполняют сложные задачи, в общем случае требующие участия человека. Внедрение систем ИИ в медицине (также, как и в политике, бизнесе и т.д.) – уже происходит, но пока как тенденция дальнейшего развития. Технологии ИИ в корне меняют систему здравоохранения – за счет кардинальным образом переработанной медицинской диагностики, разработки новых лекарственных средств, а также в целом повышают качество услуг при одновременном снижении расходов для медицинских клиник [1]. В настоящее время формируется продвинутая идеология ИИ, она может быть использована для задания оптимального паттерна, по которому должна быть выстроена система исследований гигиены труда (далее - ГТ).

Ценность компьютерных комплексов ИИ в том, что их возможности не ограничиваются начальным вариантом программы, а могут постоянно развиваться. Способность к обучению и адаптации на основе новых данных, опыта и обратной связи позволяет ИИ с течением времени становиться более эффективным и точным. Иначе говоря, ИИ способен выходить за

границы первоначально установленных алгоритмов и адаптироваться к новым ситуациям без вмешательства оператора. Понимание контекста сложных задач отличает ИИ от традиционных программ, ограниченных исключительно заранее написанными скриптами и алгоритмами [2].

С этим же связана уникальность ситуации с изначальным выделением ИИ для разработки в качестве эмпирически данного объекта. Такая процедура методологически не состоятельна и не соответствует фактическому положению. Попытка дать этому объекту уже в исходном пункте абстрактные дефиниции предопределяет справочно-бюрократическое последующее описание, которое, как правило, выглядит более или менее упорядоченным набором привычных словесных клише. Как следствие – запутывание банальных проблем и превращение действительно сложных проблем в банальности. Реально ИИ должно определяться путем перечисления вполне конкретных понятий и законов. Лишь на этой основе могут быть использованы дефиниции, но только в качестве заключения – кратких итогов исследования.

Для выявления закономерностей и выработки рациональных суждений на основе образцовых наборов данных в ИИ используется машинное обучение (далее – МО). Само по себе МО – это наука о разработке алгоритмов и статистических обучающих моделей, предъявление которых компьютерным системам ИИ может быть использовано для обучения выполнению сложных задач без четких инструкций. Вместо них ИИ самостоятельно ищет закономерности и формирует выводы. МО следует рассматривать как подсистему ИИ, которая обучает компьютер различным приемам, позволяя генерировать модели, которые затем могут анализировать входные данные для получения результатов с необходимой точностью [3].

Обращает на себя внимание вполне общая для современной медицины тенденция – большой объем, разнообразие и высокая скорость накопления исходной медицинской информации. Это обусловлено [4]:

- увеличением количества, доступности и повышением качества медицинской диагностической аппаратуры;
- ростом скорости и плотности информационных обменов как между отдельными исследовательскими группами, так и между медиками теоретического и практического профиля.

Благодаря Интернету собираемые по всему миру медицинскими организациями данные объединяются, их суммарный объем превосходит возможности традиционных средств работы с такой информацией. Появился специальный термин «Большие медицинские данные» (далее – БМД). Предполагается, что работа с БМД требует развития новых методов их анализа [4], без чего медицина превращается из способа научного познания в средство накопления разрозненных фактов. Именно для ИИ характерна работа с большим количеством данных (со сбором, хранением и их передачей), с приданием им смысла, то есть с превращением данных в знания, выводы и мотивы действия. Человек эту работу в приемлемое время выполнить не в состоянии.

Поскольку ИИ охватывает широкий спектр исследований, на сегодня в МО разработано множество подходов. Можно указать на кластеризацию данных, байесовскую сеть, глубокое обучение и анализ дерева решений. Перспективна здесь методология глубокого обучения, проверенная в современных приложениях в различных областях [5]. Глубокое МО предполагает ИИ состоящим из нейронных сетей с несколькими скрытыми слоями. Методология обучения применяет нелинейные преобразования и модельные абстракции высокого уровня на больших базах данных. Последние достижения архитектуры глубокого обучения уже внесли значительный вклад в развитие ИИ во многих областях.

Несмотря на успех процедуры МО, некоторые проблемы должны быть серьезно рассмотрены для будущих применений. Одной из них является прямая зависимость эффективности ИИ от качества обучающих материалов (далее – ОМ). Кроме того, что объем, качество и сложность данных ОМ являются потенциальными резервами методики МО в ГТ [6], для инфраструктуры общественного здравоохранения вообще необходимы точные и надежные технологии диагностического тестирования [7].

В гигиенической эпидемиологии эмпирически подобранные образцовые наборы ОМ представляют собой хорошо известные соотношения Доза-Эффект (далее – ДЭ). Для вредных внешних факторов они кодируют причинно-следственные рассуждения и знания в предметной области (например, клинической и биологической) и при выполнении соответствующих требований обеспечивают добросовестное, точное и осмысленное приложение лучших результатов клинических исследований для выбора путей лечения конкретного пациента с патологией [8].

Последние годы ознаменовались развитием направления, сфокусированного на использовании результатов клинических исследований в качестве свидетельств в медицине (далее – СМ). В основном – для обоснования эффективности медицинских вмешательств [9]. При этом предполагается рациональность методов работы, осмысленность, точность, определенность и однозначность терминов и выводов. Утверждения должны допускать возможность их подтверждения или опровержения. В случае подтверждения справедливости гипотезы, следствия (приложения) из нее можно считать медицинскими фактами, основанными на свидетельствах и, тем самым – обоснованными СМ. Все это требует такого понимания исследовательских методов, которого до недавнего времени от врачей не требовалось.

Медицина, основанная на свидетельствах – это перевод определения Evidence Based Medicine [9]. Часто «Evidence» переводится как «Доказательство» и все вместе – как «Доказательная Медицина». Это не

вполне точно. Дело в том, что доказательство некоторой гипотезы – это утверждение из той же области, что и сама гипотеза. Можно говорить о математическом доказательстве в математике, или о физическом в физике и т.д., но не в химии, например. Для разных областей правильнее употреблять термин «Свидетельство». Именно этим правилом будем руководствоваться далее, при обосновании статистическими (вероятностными) данными эпидемиологических утверждений (например - о риске).

Не следует воспринимать последующее как законченную кодификацию нормативно-правовой базы использования концепций ИИ в исследованиях по ГТ. Это, скорее, изложение тех принципов и идей, какими стоило бы руководствоваться при создании такой базы. В соответствии с этим статья содержит мало новой конкретики. Детализация – в публикациях, ссылки на которые приводятся в тексте. Фактически изложенное ниже – это развернутое обсуждение вопросов, поставленных в статье [10]. Основное внимание уделяется рационализации структуры исследований, фиксации целей и задач отдельных направлений и определению связи между результатами.

Исследования наиболее значимых эпистемологических аспектов системы ГТ (структуры и границы этой области знаний) сводятся к решению следующей последовательности проблем:

- подбор выборок ОМ – образцовых наборов пар «вход-выход» для обучения системы ИИ (в медицине это соотношения ДЭ);
- при этом – использование методов СМ для обеспечения должного качества выборки ОМ;
- для этого – структурирование (табулирование) клинических данных БМД, выявление закономерностей (свидетельств), обычно скрытых под внешней случайностью;

- отбор свидетельств силы мультиморбидных ассоциаций в парах ОМ и указаний на направление причинно-следственных связей в каждой паре переменных для разбиения их на входные и выходные;

- отобранные свидетельства используются для формирования СМ заключений (выводов) относительно обусловленности отдельных заболеваний вредными производственными факторами (далее – ВПФ);

- специальные разделы ГТ посвящены общим вопросам медицины труда, профессиональной патологии и неспецифическим синдромам, сопровождающим воздействие ВПФ;

- здесь обнаруживается однотипность откликов на воздействия ВПФ, выводящие организм из гомеостаза (во всяком случае – на начальных стадиях); это позволяет объединить развитие производственно-обусловленных заболеваний (далее – ПОЗ) в один общий коморбидный синдром профессиональных заболеваний.

Начала ИИ, предлагаемые для анализа и обобщения клинических данных, достаточно просты, но новые требования необходимо ясно сформулировать и не отступать от них. При этом, как всякая наука, ГТ должна основываться не на интуитивных догадках и отдельных умозаключениях, а на рациональных результатах опытов и экспериментов, изучаемых с помощью математических приемов [11]. Сами по себе результаты должны использовать стандартизованную терминологию и сопрягаться друг с другом, образуя единую структуру (онтологическую модель) исследований ИИ в сфере ГТ. Именно с этого (с краткого обзора существующих практик и перспективных стандартов для терминов и концепций ИИ) начинается последующее изложение.

1. Терминология (стандарты)

Начала ИИ могут выбираться по-разному, в зависимости от темы исследований ГТ. Но, во всех вариантах содержится нечто устойчивое –

система постулатов, определяющих основные понятия, т.е. система определений. Именно они должны обеспечить высокий качественный уровень разработки, управления и использования контента ГТ. Это онтологические стандарты соглашений о смысле языковых выражений. Здесь недопустима небрежность уже в изначальных терминах – все должно быть ясно и однозначно определено, и неизменно от одного исследования к другому. Нормализация терминологии позволяет избежать путаницы и недопонимания.

Ниже приводится (с краткими комментариями) перечень основных терминов, используемых в ИИ исследованиях ГТ.

1.1. Большие медицинские данные (БМД)

Содержание термина БМД дано выше. Он отмечает большие массивы данных, специфика работы с которыми характерна для ИИ. Исходные БМД – это выраженные в разной форме «сырые» факты, которые должны быть обработаны, организованы, поставлены в контекст. Клинические регистры – база этого процесса использования реальной клинической практики. Они позволяют собирать данные о применении медицинских технологий и проводить мониторинг результатов терапии. Такие данные нужны, в частности, для обучения ИИ (см. ниже).

Клинические регистры выделены Управлением по контролю за продуктами питания и лекарственными средствами США (U.S. Food and Drug Administration; US FDA) как источник данных для получения реальных клинических свидетельств, отличных от традиционных [12]. Речь идёт об автоматизированных системах регистров, позволяющих проводить как эпидемиологические исследования, так и мониторинг индивидуального клинического состояния пациентов на всех этапах лечебно-диагностического процесса, анализировать накопленные данные и использовать их для совершенствования диагностики и лечения заболеваний, оценивать эффективность и безопасность различных методов

и средств лечения, качество жизни пациентов и качество специализированной высокотехнологичной медицинской помощи, прогнозировать объёмы её финансирования. Таким образом, регистры пациентов являются электронной базой данных для принятия врачебных решений [13].

В медицинской литературе популярны систематические обзоры, с включенными оценками показателей здоровья. Общие требования к источникам данных приведены в документах Всемирной Организации Здравоохранения. Работы, содержащие медицинские данные, описывают результаты обработки медицинских карт как минимум нескольких сот пациентов. Однако, есть работы, содержащие информацию о тысячах [14], десятках и сотнях тысяч, до нескольких миллионов пациентов [15].

Основное, что характерно для БМД в медицине – их статистическая природа. Если при фиксированных с заданной точностью начальных условиях и параметрах явления процесс демонстрирует одну и ту же временную реализацию, то естественно его динамическое описание. Однако, в медицине при каждом новом исследовании «в одних и тех же условиях» реализация процесса обычно отличается от предыдущих. Неоднозначность естественно связать с влиянием многочисленных неконтролируемых факторов, которые всегда имеют место в реальном мире. При этом целесообразно отказаться от детерминированного описания и использовать аппарат теории вероятностей и случайных процессов [16 – 18]. Такие данные (именно они обозначаются термином БМД) требуют (из-за своего большого объема) специальных методов обработки (логических, алгоритмических или статистических), больших вычислительных ресурсов и специального программного обеспечения.

Первое, что здесь следует учитывать – множественность взаимодействующих процессов. Так как в общем случае отдельная причина болезни лишь частично ответственна за результат, а последний лишь частично зависит от отдельной причины, то множественность причин

обуславливает случайность результатов и для анализа ситуации необходим статистический подход. Статистически значимые результаты образуются комплексом индивидуальных данных об отдельных пациентах.

С одной стороны, это обстоятельство «размывает» результаты, делая их вероятностными, внося в них неопределенность. Однако, с другой стороны, дает мощный исследовательский инструмент. Тут важна близость понятий «вероятность» и «риск». Первое (априорная и апостериорная вероятность) используется в рамках статистического подхода, второе (риск безусловный и условный) принято в эпидемиологии [19]. Молчаливо подразумевается, что следует инверсировать результаты статистического анализа ситуации: выносить эпидемиологические суждения об уровне внутренней связи событий по статистическим показателям (относительной вероятности) их проявления в выбранных группах.

1.2. Эпидемиологические риски и вероятности

В большинстве нормативных документов предполагается (как само собой разумеющееся), что эпидемиологические риски и вероятности – одно и то же. Так, в [20] определяется «Риск - вероятность причинения вреда жизни или здоровью граждан, ... с учетом тяжести этого вреда». В качестве оценки профессионального риска берется статистическая вероятность ущерба для здоровья работников (с учетом его тяжести) от действия вредных и опасных факторов рабочей среды и трудовой нагрузки (см. [20], там же библиография вопроса).

Близкое определение риска приведено в стандарте [21] «Риск: сочетание (произведение) вероятности (или частоты) нанесения ущерба и тяжести этого ущерба». Предлагается рассчитывать риск, как математическое ожидание дискретной случайной величины – ущерба здоровью и жизни работника.

Следует иметь в виду, что идентичность понятий – всего лишь предположение, которое не всегда срабатывает. Например, в

эпидемиологии встречается несколько рисков (принята полукачественная градация рисков – от пренебрежимо малого до экстремально высокого). Это верно и для вероятностей, но их градация определяется по другим основаниям. Появляется проблема соответствия одного другому. Для сколько-нибудь обоснованного решения необходимо отдельное обсуждение.

Другой пример: на вероятность влияет время развития заболевания – чем оно больше, тем выше вероятность найти больного среди всех пациентов. Для переноса этого результата в эпидемиологию требуется дополнительное соглашение.

1.3.Машинное обучение

МО – это наука о разработке алгоритмов и статистического обучающего материала, предъявление которого компьютерным системам ИИ используется для обучения выполнению сложных задач без четких предварительных инструкций. Вместо них ИИ самостоятельно ищет закономерности и формирует выводы [6]. Задаваемые МО алгоритмы позволяют генерировать методы анализа БМД [7]. Одна из возможностей здесь – это модель генеративного машинного обучения ИИ (далее – ГМО).

ГМО – это модель ИИ нейронной сети, предварительно обученная с помощью МО для выявления шаблонов и обработки больших объемов статистических входных и выходных данных. Основы ГМО – это обучение по прецедентам. Методы решения соответствующих задач созданы за последние 10 – 20 лет. Начала ИИ исследований ГТ можно использовать для описания исходных идей эвристики и математических основ их формализации. ГМО дает возможность ИИ вырабатывать общие утверждения, объединенные в единое целое и позволяющие по определенным правилам выводиться новые утверждения относительно ГТ.

Перспективна здесь методология глубокого обучения, проверенная в современных приложениях в различных областях [5]. Глубокое МО предполагает структуру ИИ как нейронной сети из нескольких слоев. Внедрение архитектуры глубокого обучения уже внесло значительный вклад в развитие ИИ во многих областях.

Эти обстоятельства следует учитывать уже на стадии подготовки исходных данных, которые дают возможность применения новых методик ИИ. Хранение и накопление данных без выявления взаимосвязей между ними при помощи современных алгоритмов МО является порочной практикой, не приносящей реальной пользы.

1.4. Обучающие материалы

Так как ИИ – это программа, обучающаяся по набору ОМ, то качество работы ИИ (а следовательно – ГТ) непосредственно определяется качеством материала в ОМ. Точные и надежные технологии диагностического тестирования необходимы для ГТ и шире - для инфраструктуры общественного здравоохранения. Формально задача обучения может быть определена так: есть множество входных данных $\{x_1 \dots x_n\} \in X$, по которым необходимо получить некоторое множество ответов $\{y_1 \dots y_m\} \in Y$. При этом предполагается, что существует зависимость: $Y = F(X)$. Обучение сводится к поиску решающей функции F , приближающей Y на всем множестве X , по предъявленному набору тестовых данных «вход-выход», желательно исчерпывающего возможные рабочие ситуации. ОМ должны моделировать реальную работу ИИ – содержать входные наборы данных и соответствующие им выходные результаты анализа [6], [7].

Сложность работы в сфере ИИ обуславливается неточностью, противоречивостью, разнородностью данных, которые при этом обычно имеют большие объемы. Для их обработки требуются специальные программные средства – инструменты преобразования «сырых» данных в

информацию, в знания. Алгоритмы обработки данных в программах ИИ должны иметь возможность обучаться по прецедентам. Процессы получения БМД, организации их поиска и хранения, анализа, изменения и удаления объединяются в рамках ИИ под названием «управление данными». Специфика подбора ОМ для работы в области ГТ описана ниже (см. разд.2).

1.5. Доза-Эффект

Аналог ОМ для медицины труда общей практики – соотношения ДЭ [8]. Возможность оценки взаимоотношения между дозой и результатом предоставляют упорядоченные категории. Грубые бинарные переменные (например, да или нет; высокое или низкое) часто неадекватны, поскольку уверенность в воздействии может быть низкой и объединять большое количество различных уровней воздействия внутри потенциально не гомогенных групп риска. Оценка по непрерывной шкале воздействия в значимый период времени придает исследованию значительно большую чувствительность.

В гигиене труда ДЭ — это функция, показывающая реакцию организма работника на стресс-фактор в зависимости от его величины (концентрация токсичного вещества или загрязнителя, температура, интенсивность облучения, и т. д.). Под «эффектом» исследователь может иметь в виду физиологический или биохимический процесс, или даже уровень смертности; следовательно, единицами измерения могут быть количество особей (в случае смертности), упорядоченные описательные категории (например, степень патологических повреждений), либо уровни физических или химических сдвигов в организме. Обычно в клинических исследованиях изучаются эффекты на разных организационных уровнях – клеточном, тканевом, организменном, популяционном.

Основы такого подхода изложены в книге [22]. Определение «дозы» представляет самостоятельную (не медицинскую) проблему, а «эффект»

предлагается описывать биометрическими функциями, отражающими скорость развития заболеваний. В рамках такого подхода к ГТ доза определяет вариации параметров биометрических функций в зависимости от уровня и времени воздействия ВПФ.

1.5.1. Дозы ВПФ

Оценка дозы по результатам разовых измерений ВПФ нетривиальна. Из-за комплексного характера работ по ГТ и их экономического значения, выбор системы контроля за условиями труда необходимо основывать на базовых особенностях воздействия ВПФ, которые позволяют определить общие подходы к методике выполнения дозовых измерений [23].

Измерения дозы многих ВПФ в условиях реального производства представляются трудно выполнимыми из-за постоянно меняющейся производственной обстановки. Однократные измерения дают только мгновенный снимок ситуации, а для постоянного индивидуального мониторинга требуется слишком большое количество сложного оборудования и обслуживающего его высококвалифицированного персонала. В качестве альтернативы предлагается измерять не нормируемую величину, а нечто, что должно быть пересчитано в нее. Немного, однако, ясности относительно методов пересчета одного в другое.

Представления о проблемах с методикой определения доз воздействия, можно получить, рассматривая ситуацию, когда ВПФ включаются в случайные моменты времени и действуют со случайной интенсивностью, а работники случайно перемещаются между различными производственными участками [23]. В этом, достаточно общем примере, в качестве ВПФ выбрано ионизирующее излучение, однако, с точностью до терминологии, концепция методики выполнения дозовых измерений, остается общей и для других ВПФ и для бытовых условий.

Воздействие ВПФ на работников можно рассматривать как последовательность испытаний Бернулли с определенной вероятностью

экспозиции. Распределение вероятности задается биномиальным законом. Помимо средних значений дозы вероятностный подход позволяет оценить неопределенности различных типов. Достоверность результатов проверяется путем сравнения реальных отклонений результатов измерений ВПФ от среднего значения со среднеквадратическим отклонением. Если первое меньше второго, результатам измерений можно доверять.

1.5.2. Эффекты ВПФ

При обсуждении источников БМД в разд. 1.1 указывалась роль медицинских регистров, которые выделены US FDA в качестве источника данных и для получения свидетельств из реальной клинической практики. Автоматизированные системы регистров позволяют проводить как эпидемиологические исследования, так и мониторинг индивидуального клинического состояния пациентов на всех этапах лечебно-диагностического процесса, анализировать накопленные данные и использовать их для совершенствования диагностики и лечения заболеваний, оценивать эффективность и безопасность различных методов и средств лечения. В качестве составной части обучающих выборок должны использоваться результаты реальной клинической практики, моделирующие реальную работу ИИ.

Концепция ДЭ небызупречна. Невозможно общее определение «дозы»: анализ эффектов акустического шума, например, принципиально отличается от анализа пылевого воздействия [11]. Для некоторых видов ВПФ (например – электромагнитных полей) механизм воздействия неясен, и доза вводится по аналогии с другими факторами, в расчете на то, что это понятие будет уточнено по мере углубления понимания процесса воздействия на организм.

1.6.Свидетельства в медицине

Последние несколько десятков лет ознаменовались развитием в медицине СМ – направления, сфокусированного на использования клинических исследований для эмпирической демонстрации эффективности медицинских вмешательств [24]. В рамках СМ анализ “по аналогии” позволяет делать лишь более-менее содержательные предположения. Все чаще и чаще врачам предлагается углубленный подход для того, чтобы принимать ИИ решения и анализировать, что стоит за теми или иными клиническими рекомендациями. При этом предполагается рациональность методов работы, осмысленность, точность, определенность и однозначность терминов и выводов. Утверждения должны допускать возможность их подтверждения или опровержения. Все это непосредственно относится к развитию ГТ, т.к. это подраздел медицины, основанной на свидетельствах, предполагающий поиск, сравнение, обобщение и широкое распространение полученных доказательств для использования в интересах больных [24].

Предложенная в начале 1990-х гг. концепция СМ подразумевает критическую оценку медицинской информации с целью установления её надежности и достоверности. Для этого СМ должна опираться на рациональные (количественные) клинические данные, прошедшие проверку в ходе крупных рандомизированных испытаний. В случае подтверждения справедливости гипотезы, следствия (приложения) из нее можно считать медицинскими фактами, основанными на свидетельствах и, тем самым – относящимися к СМ.

1.7. Мультиморбидность профпатологий

Одно из открытий, сделанных благодаря работе с БМД методами ИИ, это обнаружение явления мультиморбидности в ГТ [25]. Мультиморбидность («коморбидность», «полиморбидность» и др. – несколько нозологических форм у одного больного одновременно) довольно распространена, так что полиморбидный больной сегодня — это обычное явление в ежедневной

медицинской практике. Редкостью является выявление у пациента только одного заболевания. Соответственно, при тотальной распространенности полиморбидной (интегральной) патологии должна стать правилом интегральная оценка состояния больного. Вопросы патогенеза сочетанных заболеваний, их эффективное и долгосрочное лечение являются одной из самых актуальных проблем медицины и здравоохранения, с которыми сталкиваются пациенты, ученые, работники здравоохранения и общество в целом [26].

Для определенности, далее будем общее состояние мультиморбидности подразделять на два основных: поли- и коморбидность. Предполагаем, что они обозначают различные состояния, с разной иерархией нозологий. В то время как полиморбидность должна рассматриваться как сосуществование двух или более заболеваний у человека, при этом ни одно из них не выделяется в качестве основного, коморбидность определяется как наличие или возникновение синдрома - отдельных дополнительных нозологических единиц у пациента с индексным (основным) заболеванием (далее – ОЗ) [26].

В настоящее время основным из имеющихся у больного нескольких заболеваний принято считать ту нозологическую форму, которая сама по себе или вследствие своих осложнений обуславливает наибольшую угрозу трудоспособности и жизни больного и требует неотложного лечения. Важно подчеркнуть, что это определение ОЗ аргументировано не физиологическими показаниями. С этой точки зрения выбор ОЗ в значительной степени произволен и определяется обстоятельствами, сторонними по отношению к медицине.

Здесь не столько терминологическая, сколько сущностная проблема. Не следует понимать мультиморбидность как сумму (результат сложения) того или иного количества болезней и автоматического утяжеления состояния больного; за ним, вероятно, стоят те закономерности формирования патологии у человека и та сущность болезни, которые еще предстоит изучить и понять. Знание причины болезни определяет наиболее

эффективный способ ее предотвращения и лечения. *Diagnosis bona – curatio bona*. Понятно, что основные теоретические разработки и прикладные усилия в фундаментальной медицине направлены на установление причин и механизмов развития болезней. Для реализации этого направления необходимо количественное описание нозологической структуры (паттерна) мультиморбидности.

Мультиморбидность видоизменяет классическую клинику течения заболеваний, утяжеляет состояние больных, удлиняет диагностический процесс и усложняет лечение. В частности, у мультиморбидных заболеваний другая эпидемиологическая основа и последовательность проявления отдельных нозологий. Это привносит особенности мультиморбидности в профпатологию, делает специфичными ее предикторы, варианты клинического течения, расшифровку сложных механизмов патогенеза, выбор рационального лечения, индивидуальных подходов к ведению и реабилитации пациентов [27]. Требуются специальные исследования патогенеза отдельных групп заболеваний: мультиморбидность как целое ПОЗ больше суммы отдельных нозологий. Это физиологическое образование более сложное и более высокого уровня биологической организации.

2. Управление данными

ИИ – мощное орудие исследования, но требует предварительной организации материала – объединение и структурирование результатов БМД клинических исследований. Они должны собираться под определенные задачи, единообразно, с ясно определенной целью. Процессы поиска и получения БМД, организации их хранения и анализа, изменения и удаления, объединяются в рамках ИИ под названием «Управление Данными» (далее – УД).

Как уже отмечалось выше, ИИ для исследований ГТ – это обучающаяся программа, предназначенная для работы с БМД. Методика работы с большими объемами данных для пересчета входных данных в выходные вырабатывается в процессе МО, сводящегося к предъявлению программе ОМ - обучающих (образцовых) наборов входных данных и соответствующих им выходных результатов. Качество ИИ практически полностью определяется качеством ОМ, поэтому УД должно обеспечить надежность и достоверность медицинской информации.

Логика ИИ требует задания комплексов мультиморбидных нозологий минимальным числом рационально определенных характеристик. Основная проблема здесь – оценить насколько явно скрытые вероятности, определяющие структуру мультиморбидной патологии, отражаются в реальных статистических результатах. Вне ИИ проблема УД не только не решается, но и не ставится. Отсутствие единого методологического подхода к оценке свидетельств явления мультиморбидности препятствует проведению обобщающего фундаментального исследования его свойств и закономерностей.

Различные аспекты мультиморбидности могут быть оценены ИИ с разных точек зрения: величины, тяжести, характера и бремени основного и сопутствующих заболеваний. Самый распространенный подход – редукция результатов измерений сосуществующих заболеваний в одну суммарную переменную, которая подобно характеристике состояния одного заболевания может подвергаться любому анализу [28]. При таком подходе должны объединяться количество болезней, доля потерянных лет жизни, риск смерти и весовые коэффициенты нетрудоспособности и тяжести сопутствующих заболеваний в одну суммарную переменную. Выбор подходящего способа редукции не должен быть произвольно-сочиненным.

В современной статистике разработаны весьма эффективные методы исследования БМД. Самый простой способ получения единственного значения на основе ряда результатов измерений — усреднение, или

вычисление среднего значения. Способ не лучший и не единственный. Следует иметь в виду, что свойства систем многих объектов описываются в терминах специфических переменных, характерных для средних по ансамблю: статистическим весом, энтропией, температурой и т.п. Эти переменные не очевидны, они не измеряются непосредственно и могут быть только вычислены по результатам наблюдения за системой. Тем не менее, именно эти переменные дают адекватное описание системы.

Ниже (разд. 3 и 4) приведены два характерных примера подготовки и аналитического описания таких специфических переменных БМД в качестве свидетельств.

3. Подготовка таблиц БМД для обучения ИИ

Машинное обучение ИИ – это дисциплина, находящаяся на пересечении теории вероятностей, математической статистики, численных методов оптимизации, а также дискретного анализа. Решением задач извлечения знаний из данных БМД занимается еще только формирующаяся область «Интеллектуальный анализ данных». Результаты фильтрации пар зависимостей ДЭ, предназначенные в качестве составной части выборок для использования в обучающих материалах ИИ, должны отражать силу и направление связей: что в паре считать входом, а что выходом.

Одним из основных приемов статистического анализа БМД является группировка объектов. Обычно (в эпидемиологии – в частности) статистические результаты натурного исследования группируются в таблицы, которые являются наиболее универсальным средством изучения причинно-следственных отношений [30]. Обычно это факторная таблица, в статистике – метод представления перекрестной классификации категориальных данных, предназначенный для исследования статистической связи между ними. При этом категории одного показателя указаны горизонтально (в строках), а категории другого показателя – вертикально (в столбцах). Так организуется сводка единичных факторов,

образующая совокупность данных наблюдения. Статистически значимое количество единичных данных об отдельных объектах наблюдения образует комплекс статистических результатов. На этом этапе проявляются общие черты и закономерности исследуемых явлений.

Выше (см. п. 1.5. Доза-эффект) при обсуждении проблемы создания обучающей выборки, в частности – обоснования вычленения взаимосвязанных пар из БМД и деления их на входные и выходные множества, указывалось на необходимость предъявления свидетельств причинно-следственных соотношений в этих парах. Сила ассоциаций и направления связей должны учитываться при включении связанных пар в выборки ОМ для обучения ИИ. Такую возможность предоставляет обращение к вероятностной логике Т. Байеса для оценки свидетельств [29].

Основная идея подхода Т. Байеса – оценка изменения вероятности события в зависимости от того, в какой группе она определяется. Первоначальная (априорная) оценка вероятности, сделанная для большой (неограниченной) группы меняется на более специфическую (апостериорную) вероятность для исследуемой (ограниченной) группы. Следующий шаг – инверсия результата: вынесение суждения об уровне внутренней связи событий по относительной вероятности их проявления в выбранных группах.

3.1. Индексация ячеек статистических таблиц

Для определенности высказываний будем полагать, что проблема состоит в оценке силы и направления причинно-следственной зависимости между экспонирующим фактором А (ВПФ или ОЗ для коморбидности) и заболеванием В. Будем отмечать численности n_{ik} индексом i , относящимся к А и индексом k , относящимся к В. Эти числа заполняют 2×2 таблицу:

Таблица 1. Статистика заболеваемости. Обозначения.

$A(i) \setminus B(k)$	$k=0$	$k=1$	
-----------------------	-------	-------	--

$i=0$	n_{00}	n_{01}	n_{0*}
$i=1$	n_{10}	n_{11}	n_{1*}
	n_{*0}	n_{*1}	n_{**}

Смысл отдельных чисел:

n_{00} – количество не экспонированных здоровых (нет ни А, ни В);

n_{10} – экспонированы А, не больны (здоровы) В;

n_{01} - не экспонированы А, больны В;

n_{11} - экспонированы А и больны В.

Вдобавок к указанным числам вводятся дополнительные (маргинальные), отмеченные индексом *, переменные. Это результаты суммирования n_{ik} :

n_{1*} - экспонированы А (неважно - больны ли они В);

n_{*1} - число больных В (неважно - экспонированы ли они А);

n_{**} - общее число пациентов.

Возможен обратный пересчет маргинальных переменных в числа заполнения.

3.1.1. Таблицы сопряженности

Числа $\{n_{00} \dots n_{11}\}$ образуют таблицу сопряженности (далее – ТС). В статистике это метод представления перекрестной классификации категориальных данных, предназначенный для исследования статистической связи между ними. При этом категории одного показателя указаны горизонтально (в строках), а категории другого показателя указаны вертикально (в столбцах).

ТС 2x2 является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения. Каждая ячейка таблицы указывает число или долю лиц, имеющих характеристики из соответствующего ряда и столбца. В эпидемиологии она используется в клинических испытаниях, например,

для сравнения бинарных исходов в двух группах [30]. Одна из проблем традиционной эпидемиологии заключается в наблюдении и количественной оценке ассоциаций. В этой области предложено несколько десятков индексов, указывающих на силу вероятностной связи событий [32].

ТС часто используются для проверки гипотезы о наличии связи между двумя признаками с использованием точного теста Фишера-Пирсона χ^2 , или коэффициента (критерия) согласия Yule's Q. Критерии Q и χ^2 – это чисто статистические характеристики ситуации [33].

Например, при оценке связности используется коэффициент сходимости (в статистике - Yule's Q), он является мерой связи между двумя бинарными переменными:

$$Y = [(n_{00} \times n_{11})^{1/2} - (n_{10} \times n_{01})^{1/2}] / [(n_{00} \times n_{11})^{1/2} + (n_{10} \times n_{01})^{1/2}] ;$$

$$Q = 2*Y/[1+Y^2] .$$

Для оценки достоверности статистических гипотез часто используется р-значение статистической значимости. Величина р может быть оценена критерием Пирсона χ^2 [34]. Последний определяется по формуле

$$\chi^2 = \sum_{ik} \{(n_{ik} - m_{ik})^2 / m_{ik}\}$$

где n_{ik} – реальные числа заполнения ячеек ТС, а m_{ik} – ожидаемые числа заполнения в предположении о независимости событий В от фактора А. В последнем случае в ячейках ТС должны находиться следующие числа:

Таблица 2. Ожидаемые числа заполнения 2x2 ТС

m_{ik}	=0	=1
i=0	$n_{*0} \times n_{*0} / n^{**}$	$n_{*0} \times n_{*1} / n^{**}$
i=1	$n_{*0} \times n_{*1} / n^{**}$	$n_{*1} \times n_{*1} / n^{**}$

Чтобы различать значимые и незначимые результаты, обычно используют уровень $p = 0,05$. Если задать его таким, то критическое $\chi^2 = 3,84$. Расчеты конкретных данных натурных измерений показывают, что такой величине критерия Пирсона соответствует критический Yule's $Q = 0,3$. Задавая другой уровень значимости, приходим к другой оценке критической величины критерия χ^2 и другой оценке критического Yule's Q .

3.1.2. Таблицы распространенности

После несложных преобразований данные ТС позволяют оценивать распространённость заболеваний — медико-статистический показатель, исчисляемый количеством заболеваний на популяцию определенной численности (обычно 1000 чел.). При работе с БМД в области ГТ обычно приводят данные о моментной распространенности полиморбидных нозологий. Это вероятность обнаружения той или иной нозологии: отношения сумм n_{1*} и n_{*1} к полному числу пациентов n^{**} . Не столь наглядно ТС свидетельствует о направлении и уровне причинно-следственных зависимостей в изучаемой системе ВПФ – ПОЗ. Для того, чтобы оценить это свидетельство, необходимы методики анализа сложных (вероятностных) ситуаций.

3.2. Логика Т. Байеса в ИИ. Свидетельства ПОЗ.

Если между ВПФ и ПОЗ существует однозначное соответствие, т.е. если ПОЗ может возникнуть только в результате действия ВПФ, то число n_{01} должно быть равно нулю: не может быть ПОЗ, если на работника не действует ВПФ. Как правило, соответствие всего лишь вероятно, т.е. ПОЗ может быть вызвано ВПФ, но с ненулевой вероятностью может развиваться и самостоятельно. В этих случаях следует ожидать, что величина n_{01} будет относительно мала, но не строго равна нулю. Еще точнее – следует говорить о малой вероятности одной ситуации относительно других. Когда в анализ

включаются относительные вероятности, для оценки свидетельств следует использовать вероятностную логику Т. Байеса [34].

Для использования логики Т. Байеса при анализе причинно-следственных зависимостей в ГТ, данные ячеек ТС пересчитываются в условные вероятности обнаружения ПОЗ (нозологическое следствие) при наличии определенных ВПФ (факторы-причины). Можно указать на работу [33], в которой диагностировались причины ПОЗ. Описан алгоритм, подразумевающий специальную технологию сбора, анализа и представления данных эпидемиологических исследований. Анализ причинно-следственных отношений указывает на ВПФ как на причину ПОЗ [35], [36]. Показано, что оценки силы вероятностной связи ПОЗ с ВПФ, основанные на статистике Фишера χ^2 , приводят к тем же качественным выводам, что и анализ апостериорной вероятности по Т. Байесу, однако последний позволяет наглядно проследить логику возникновения ситуации, дает направление и количественную меру воздействия ВПФ, то есть позволяет достоверно диагностировать заболевание в качестве ПОЗ.

3.2.1. Апостериорные вероятности. Выводы для ГТ.

Формальное определение случайной величины в теории вероятностей сводится к оценке в пространстве элементарных событий меры подмножества исследуемых событий. В это определение явно входит представление о пространстве элементарных событий. Если последнее определять по-разному, получим различную вероятность одного и того же подмножества исследуемых событий.

При статистическом анализе это – безусловная и условная вероятности. В традиционной эпидемиологии такой подход сводится к сопоставлению апостериорного и априорного рисков. Безусловная вероятность (априорный риск) обнаружить больного $P(\text{ПОЗ}=1)$ – это доля больных среди всех обследуемых работников:

$$P(\text{ПОЗ}=1) = n_{*1} / n^{**} .$$

Более информативна величина вероятности заболевания среди экспонированных работников (подвергавшихся воздействию ВПФ). Это условная вероятность, которую в эпидемиологии определяют через апостериорный риск:

$$P(\text{ПОЗ}=1 \mid \text{ВПФ} = 1) = n_{11} / n_{1*}.$$

Частное от деления $P(\text{ПОЗ} = 1 \mid \text{ВПФ} = 1)$ на $P(\text{ПОЗ}=1)$ показывает уровень влияния ВПФ на вероятность заболевания ПОЗ. В эпидемиологии это относительный риск заболевания под действием ВПФ:

$$RP_{11} = (n_{11} * n_{**}) / (n_{*1} * n_{1*})$$

Если величина RP_{11} близка к единице, это означает, что воздействие ВПФ незначительно: в подгруппе экспонированных риск заболеть ПОЗ такой же, как и во всей обследуемой группе. Значимое превышение RP_{11} над единицей означает сильную связь ПОЗ с ВПФ: уровень заболеваемости среди экспонированных ВПФ существенно больше, чем в среднем по обследуемой группе. Следовательно, фактор RP_{11} определяет коэффициент доверия Байеса к гипотезе о положительной связи ПОЗ с ВПФ.

Аналогично интерпретируются другие относительные риски R_{ik} [34]. Примеры использования ТС для анализа конкретной ситуации с бронхолёгочной патологией у рабочих, занятых добычей, транспортировкой и обогащением апатит-нефелиновых руд в климатических условиях Крайнего Севера даны в работах [35], [36]. Другие примеры анализа ТС для оценки эффектов экспонирования ВПФ работников, занятых в сварочном производстве, демонстрирующие адекватность подхода ИИ при исследованиях ГТ на трех крупных предприятиях машиностроения в г. Санкт - Петербурге: ОАО «Ижорские заводы», ФГУП «Адмиралтейские верфи» и ОАО судостроительный завод «Северная верфь» даны в [37].

Результаты таких оценок эффектов экспонирования ВПФ указывают на надежность свидетельств, основанных на вероятностной логике Т. Байеса.

4. Статистические ансамбли

Другой пример аналитического описания БМД опирается на предложенную Д. Гиббсом в 1902 году концепцию Статистических Ансамблей (далее – СА). СА – это совокупность большого (возможно – бесконечного) числа экземпляров системы, находящихся в одинаковых макроскопических состояниях; при этом микроскопические состояния у них могут различаться. Иными словами, СА – это набор макроскопически одинаковых (с точностью до пренебрежимо малых флуктуаций) систем, анализируемых для статистического описания одной из них. При ИИ анализе ГТ предполагается, что ухудшение здоровья со стажем работы выражается в накоплении функциональных изменений в состоянии работника. В этом случае СА перестает быть описанием и становится результатом измерения, а оно, как известно, подвержено случайности. Распределение ошибок измерений подчиняется некому универсальному правилу, которое называют законом случайного распределения. Это основополагающая для статистической теории идея.

4.1. Распределение нозологий

Степень выраженности накапливаемых функциональных изменений описывается неким показателем k со счетным числом состояний $\{k_j\}$. О самом показателе не предполагается ничего, кроме того, что он существует и его рост со временем (возрастом) отражает ухудшение состояния здоровья за счет накопления неблагоприятных функциональных изменений в организме. Из любого j -того состояния возможны переходы с вероятностью p либо в состояние $j+1$, либо в состояние $j-1$ с вероятностью q , либо можно остаться в состоянии j с вероятностью $1-p-q$. Для отражения тенденции ухудшения состояния с возрастом следует предполагать, что $p > q$.

Чтобы координата k_n выросла от нулевого начального значения до заданного k , нужно сделать n_0 «нулевых» шагов, n_1 положительных и n_2 отрицательных, причем должно быть

$$n_1 - n_2 = k, \quad n_0 = n - (n_1 + n_2).$$

Комбинаторным решением задачи определения вероятности результата $k_n = k$ является обобщенный биномиальный закон

$$P(k_n = k) = \sum \frac{n!}{n_0! n_1! n_2!} p^{n_1} q^{n_2} (1 - p - q)^{n_0}$$

Вероятность определяется суммой большого числа небольших слагаемых, для нее из центральной предельной теоремы следует нормальное распределения вероятности со средним значением $\langle k \rangle = n \cdot \nu$ (здесь $\nu = p - q$) и дисперсией $D = p + q - (p - q)^2$

$$P(k) = \frac{1}{\sqrt{2\pi D n}} \exp\left[-\frac{(k - \langle k \rangle)^2}{2 D n}\right]$$

Непосредственно измеряемое среднеквадратическое отклонение σ связано с дисперсией соотношением $n \cdot D = \sigma^2 = \langle (k - \langle k \rangle)^2 \rangle$.

Распределение $P(k)$ получено при вполне общих предположениях о динамике развития заболевания. Его адекватность для описания реальной ситуации можно увидеть после сопоставления теоретических результатов с данными наблюдений. Такие данные, например, приведены в работе [40], вместе с распределением $P(k)$ они показаны на графике Рис.1. Вдоль оси абсцисс – индекс полиморбидности (количество нозологий у одного больного). Значения индекса 1: нет заболеваний; 2: 1-2; 3: 3-6; 4: 6-9; 5: более 9 нозологий.

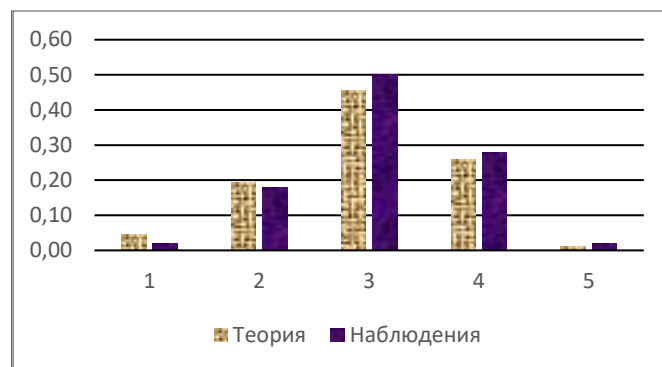


Рис.1. Распределение индекса полиморбидности (теория и наблюдения).

Видно удовлетворительное совпадение теоретических и натурных результатов. Это можно рассматривать как свидетельство адекватности концепции СА.

4.2. Свидетельства в ГТ. Метод линейной регрессии распределения.

Особенностью распределения $P(k)$, характеризующей СА работников как объект ГТ, является возможность линеаризации зависимости логарифма вероятности от квадрата разности наблюдаемого числа k нозологий и средним значением $\langle k \rangle$: $x = (k - \langle k \rangle)^2$. Для использования распределения $P(k)$ в качестве свидетельства предлагается метод линейной регрессии распределения (далее - МЛРР). Он позволяет оценить реальные статистические параметры заболеваемости.

Логарифм плотности распределения, т.е. величина

$$F = \ln[P(k)] = -\ln[\sqrt{2\pi}\sigma] - \left(\frac{x}{2\sigma^2}\right)$$

должна быть линейной функцией переменной x . Именно это считаем линейной регрессией распределения. Обе половины (относительно медианы) распределения $P(k)$ должны лежать на одной и той же прямой. Более того, оба коэффициента линейной регрессии для этих половин должны приводить к одному и тому же значению σ . Таким образом, результаты МЛРР дают вполне надежные свидетельства адекватности концепции СА в области ГТ.

4.2.1. Возрастная полиморбидность.

В работе [40] приведены результаты анализа натурных исследований индекса полиморбидности для заболеваемости неинфекционной патологией 133 лиц пожилого и старческого возраста. Существуют веские основания

относить к такой патологии возрастную полиморбидность, которая представляет собой вялотекущее хроническое системное воспаление, обусловленное старением иммунной системы. Результаты натурных исследований – распределение индекса полиморбидности – приведены в виде гистограммы. Результат перехода к новой функции распределения от нового аргумента x показан на Рис.2, график (1). Принято $\langle k \rangle = 4,075$. Здесь же дана формула линейной регрессии. Коэффициенты регрессии

$$\ln[\sqrt{2\pi}\sigma] = 1,193; \quad \frac{1}{(2\sigma^2)} = 0,279$$

Из первого уравнения получим $\sigma = 1,315$, а из второго $\sigma = 1,34$ (Среднее = 1,33). Близость двух оценок величины σ свидетельствует об адекватности статистической картины полиморбидности и о правильном выборе параметра $\langle k \rangle$.

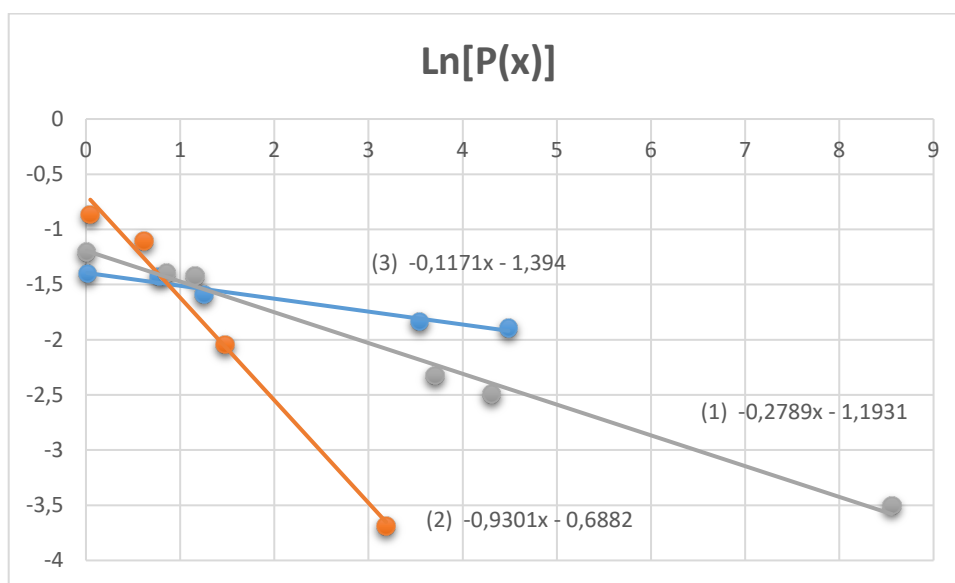


Рисунок 2. Результаты МЛРР для индекса полиморбидности

Определение медианного значения индекса полиморбидности $\langle k \rangle$ и стандартного отклонения σ позволяет последовательно вычислить реальные величины скорости роста $v = \langle k \rangle / n$ числа нозологий и дисперсии $D = \sigma^2 / n$

Получено $v = 0,509$ и $D = 0,220$. Можно пойти далее и оценить вероятность (на принятом временном шаге) заболеть $p = (D + v + v^2) / 2 = 0,49$ и вероятность остаться здоровым $1 - p - q = 0,52$. Вероятность выздороветь q пренебрежимо мала (в пределах погрешности измерений).

Дисперсия непосредственно связана с выборочным стандартным отклонением – количественной мерой разброса результатов измерений. Стандартное отклонение показывает, насколько данные по выборке близки к среднему — или, в практическом смысле, какова погрешность измерения. Если отклонение невелико, все результаты группируются вокруг среднего. Именно объяснение разброса данных (относительно среднего) случайной ошибкой дает ключ к пониманию адекватности ИИ анализа ГТ.

4.2.2. ПОЗ шахтеров

На Рис.2 приведен еще один результат МЛРР (график (2)), для распределения индекса полиморбидности, полученного в условиях, существенно отличающихся от [40]. Этот результат получен в [41] при наблюдении за шахтерами с полиморбидным ПОЗ – патологией костно-мышечной и периферической нервной системы. Здесь появляется очевидный дополнительный фактор – ВПФ. В возникновении ПОЗ каждый вид ВПФ (пыль, шум, физическая перегрузка, переохлаждение, аллергены и др.) с его интенсивностью и длительностью воздействия является ведущим полиэтиологического заболевания. При этом на организм работающего во вредных и опасных условиях труда, как правило, влияют одновременно несколько ВПФ. Каждый вид ВПФ дает свою характерную нозологию, что может приводить к формированию сочетанной патологии как ответной реакции организма на множественное воздействие.

Несмотря на, возможно, отличную природу полиморбидных ПОЗ, здесь также можно использовать методы математической статистики. В частности, описывать ситуацию с ПОЗ распределением индексов

полиморбидности $P(k)$. Это обстоятельство позволяет использовать МЛРР для оценки параметров распределений индексов полиморбидности ПОЗ.

Согласно работе [41], из 1591 больных с ПОЗ и сопутствующими полипатиями, мононозология отмечалась лишь в 5% и бинозология - в 13% случаев. Наибольшей была доля лиц с тремя сопутствующими заболеваниями (42%), на втором месте находились лица с четырьмя сопутствующими заболеваниями (33%). Доля лиц с числом сопутствующих заболеваний свыше 5 составила 2,5%.

Графическая иллюстрация этих результатов для функции распределения $P(k)$ в зависимости от аргумента x приведена на Рис. 2, график (2). К этому графику дана формула линейной регрессии. Коэффициенты регрессии

$$a = \frac{1}{(2\sigma^2)} = 0,93; b = \text{Ln}[\sqrt{2\pi}\sigma] = 0,69$$

Из первого уравнения получим $\sigma = 1/(2*a)^{1/2} = 0,733$, а из второго $\sigma = \exp(b)/(2\pi)^{1/2} = 0,794$ (Среднее = 0,764). Соответственно, при $n \approx 6$; $v \approx 0,643$ и $D = 0,117$. Вероятность (на принятом временном шаге) заболеть $p = 0,59$ и вероятность остаться здоровым $1 - p - q = 0,47$. Вероятность выздороветь q здесь также пренебрежимо мала (в пределах погрешности измерений) как и в целом по популяции.

Поучительно сопоставить эти результаты с полученными выше (см. разд. 4.2.1) для общесоматической возрастной полиморбидности [40]. Вероятность заболевания была оценена как $p = 0,49$, что заметно меньше чем $p = 0,59$ для ПОЗ. Рост вероятности p для ВПФ, по-видимому, связан с большей интенсивностью воздействия вредных производственных факторов на организм работника, чем возрастное хроническое системное воспаление.

Для ПОЗ последний результат тривиален, но он рационален т.е. предложенный метод дает количественную характеристику эффекта, критерии обоснованности можно рационально сформулировать и оценить. Фактически здесь нет места для произвольных характеристик явления или

их субъективных оценок, сводящих на нет результаты многих довольно трудоемких исследований. В этом смысле результаты канонизируют СМ подход ИИ при исследованиях ГТ. Обратная задача – оценка уровня ВПФ по риску ПОЗ – описана в работах [38], [39].

4.2.3. Психосоматические заболевания

Большинство доступных свидетельств ассоциации хронических заболеваний могло быть собрано в исследованиях отдельных ОЗ, однако сопутствующие заболевания обычно исключались. Существует растущая потребность в свидетельствах, которые можно использовать для выявления ОЗ при коморбидных заболеваниях пациентов. Возможным решением может быть поиск конкретных комбинаций заболеваний с высокой распространенностью и участие в рандомизированных клинических испытаниях пациентов с комбинациями заболеваний. В работе [42] приведены результаты таких испытаний с ассоциациями заболеваний от умеренных до сильных (Yules $Q > 0,50$). Подобный описанному выше анализ структуры и распространенности коморбидной соматической патологии у госпитализированных пациентов с пограничными психическими расстройствами показывает ведущую роль психических или социальных проблем.

5.Отклики. Медицинские данные. Общий коморбидный синдром.

Отклики на внешние воздействия, выводящие организм из гомеостаза, в значительной степени однотипны для основных ВПФ (запыленность воздуха, акустический шум, микроклимат и пр.). Впервые внимание на роль однотипности откликов обратил Г. Селье, предложив теорию общего адаптационного синдрома [43]. Она представляет чёткое биологическое объяснение наблюдений того, как организм реагирует и приспосабливается к стрессу. В своих исследованиях Г. Селье заметил, что для самых разных внешних стрессоров вполне схожи гормональные

ответные реакции, активизирующиеся для восстановления и сохранения внутреннего баланса в организме. Эти процессы являются главным содержанием общего адаптационного синдрома.

Нечто похожее следует ожидать и для патофизиологических процессов, приводящих к развитию и прогрессированию ПОЗ [44]. Важнейшую роль в формировании коморбидности здесь играют универсальные патофизиологические механизмы и развивающиеся общепатологические процессы (хроническое воспаление, дистрофия, нарушения крово- и лимфообращения и др.). Реально, изучение коморбидных нозологий ограничивается фенотипическим уровнем и характеризуется недостатком знания их структурно-генетических основ. Тем не менее, уже и фенотипический уровень исследований имеет большое практическое значение, так как принципиально важно при обнаружении определенных болезней диагностировать характерную для них структуру коморбидности.

Однотипность реакций организма определяется едиными для них факторами риска, такими как сердечно-сосудистые и церебро-васкулярные заболевания. Среди типичных симптомов отмечается артериальная гипертензия, гипергликемия с атеросклерозом и дислипидемией, ожирение (особенно абдоминальный тип), метаболический синдром, хроническая болезнь почек.

Атеросклероз проявляется уплотнением сосудистой стенки и образованием атеросклеротических бляшек. Это динамичный процесс, для которого характерно развитие изменений в стенке артерии. Из клинических проявлений – поражение артерий эластического типа, таких как аорта, подвздошные сосуды, а также крупные и средние артерии мышечного типа (коронарные, сонные, внутримозговые, артерии нижних конечностей). Основным осложнением атеросклероза коронарных артерий является ишемия, которая проявляется стенокардией, инфарктом и кардиосклерозом, ведущим к прогрессирующей сердечной недостаточности. Поражение магистральных артерий головного мозга проявляется симптомами его

хронической ишемии с последующим развитием атеросклеротической энцефалопатии и инсульта. Атеросклероз артерий нижних конечностей сопровождается клинической картиной перемежающейся хромоты. При отсутствии соответствующего лечения заболевание прогрессирует и может закончиться развитием гангрены нижних конечностей [45].

Многие из представленных составляющих факторов риска, обычно расцениваются как самостоятельные заболевания, приводящие к развитию или ухудшению прогноза сосуществующих с ними болезней у человека. Клиническая картина взаимодействия нескольких патологий у одного больного хорошо известна врачам, поскольку редко в настоящее время (особенно среди пациентов среднего и пожилого возраста), можно встретить только одну патологию [45].

Заключение (In Summa)

Использование ИИ предоставляет потенциальную возможность создавать сложные и функционально эффективные программы исследований ГТ. Как отмечено во Введении, для дальнейшего развития этой темы требуется перечисление основных концепций работы ИИ в области ГТ. Это могут быть краткие выводы из изложенного выше.

ИИ – это программа, способная к обучению и адаптации к новым сценариям работы без вмешательства оператора. Обучение происходит при предъявлении программе образцовых наборов пар «вход-выход», по возможности полно исчерпывающих реальные ситуации. Должное качество обучающего материала обеспечивается статистическим анализом полиморбидных ассоциаций в образцовых парах и направления причинно-следственных связей в каждой паре для разбиения их на входные и выходные. Так отобранные свидетельства используются для формирования эпидемиологических заключений (выводов) относительно обусловленности отдельных ПОЗ теми или иными ВПФ.

В современных исследованиях ГТ образуются БМД за счет увеличения количества, доступности и повышения качества медицинской диагностической аппаратуры, а также – за счет развития медицинских регистров для организации и хранения данных реальной клинической практики. БМД требуются для формирования (обучения) ИИ и его работы. При этом идеология ИИ предъявляет специфические требования к структуре БМД. Например, в связи с обнаружением явления полиморбидности в структуре ПОЗ, должна стать правилом интегральная оценка состояния больного.

Методы ИИ позволяют выявлять вполне общие закономерности (свидетельства) развития мультиморбидных нозологий обычно скрытые под внешней случайностью. Так обнаруживается однотипность откликов на воздействия ВПФ, выводящие организм из гомеостаза; это позволяет объединить развитие ПОЗ в один общий коморбидный синдром.

Продвинутую идеологию ИИ можно использовать для задания оптимального паттерна, по которому должны быть организованы исследования в области ГТ. Это позволяет перейти от интуитивных догадок и умозаключений к рациональной структуре исследований: фиксировать цели и задачи отдельных направлений и определить отношения между их результатами. Они могут быть использованы для оптимизации производительности, адаптации интерфейса для разных пользователей, а также для улучшения распознавания и предотвращения потенциальных угроз безопасности.

Библиография

1. Шевский В.И., Шейман И.М.; Проблемы формирования интегрированной системы здравоохранения; Вопросы государственного и муниципального управления. 2013. № 3. с. 24 – 48.

2. Osondu O. A First Course in Artificial Intelligence. Bentham Science Publishers Ltd., 2021. ISBN: 978-1-68108-854-9 ;
DOI: [10.2174/97816810885321210101](https://doi.org/10.2174/97816810885321210101)
3. Safdar S, Zafar S, Zafar N, Khan N F (2017) Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Artificial Intelligence Review: p. 1-17
4. Borges do Nascimento I.J., Marcolino M.S., Abdulazeem H.M., et al.; Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies. Journal of medical internet research 2021; 23(4): e 27275
DOI: 10.2196/27275 PMID: 33847586 PMCID: 8080139
5. Liu W. Deep learning hashing for mobile visual search. EURASIP Journal on Image and Video Processing 17, (2017). DOI: <https://doi.org/10.1186/s13640-017-0167-4>
6. Кафтаников И.Л., Парасич А.В. Проблемы формирования обучающей выборки в задачах машинного обучения // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2016. Т. 16, № 3. С. 15–24 ;
DOI: 10.14529/ctcr160302; УДК 004.855.5
7. Парасич А.В., Парасич В.А., Парасич И.В. ; Формирование обучающей выборки в задачах машинного обучения // Информационно-управляющие системы, 2021 (№ 4), с. 61.
DOI: <https://doi.org/10.31799/1684-8853-2021-4-61-70>
8. Mhuler B.; Modeling of Dose-Response Relationships; Environ Health Perspect, 1981 (V. 42) P. 23—27.
9. Guyatt G.H., Sackett D.L., Cook D.J. Users' guides to the medical literature. Evidence-Based Medicine Working Group. JAMA. 1993 Dec 1; 270 (21): 2598–601.
DOI: 10.1001/jama.270.21.2598. PMID: 8230645.

10. Федорович Г.В. Материалы обучения искусственного интеллекта охране труда (к вебинару КИОУТ). БиОТ, № 4, 2023 г., С.42 – 46, УДК 004.855.331.45

DOI: 10.54904/52952_2023_4_42

11. Федорович Г. В. Рациональная эпидемиология профессиональных заболеваний (Модели и методы). Saarbrücken, Deutschland, Palmarium Academic Publishing, 2014: P. 343. ISBN-13: 978-3-639-82722-4.

DOI: <http://elibrary.ru/item.asp?id=23256439>

12. Марцевич С.Ю., Лукина Ю.В., Кутишенко Н.П., и др. Краткий обзор методических рекомендаций "Медицинские регистры. Роль в доказательной медицине. Рекомендации по созданию". *Кардиоваскулярная терапия и профилактика*. 2023;22(6):3615.

DOI: <https://doi.org/10.15829/1728-8800-2023-3615>.

13. Счастливцев И.В., Навасардян А.Р., Лобастов К.В. Регистры, их место в иерархии исследований и значение для клинициста на примере регистра RIETE. *Флебология*. 2022;16(3): стр. 227-237.

DOI: <https://doi.org/10.17116/flebo202216031227>

14. Tran J, Norton R, Conrad N, Rahimian F, Canoy D, Nazarzadeh M, et al. (2018) Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: A population-based cohort study. *PLoS Med* 15(3): e1002513. DOI: 10.1371/journal.pmed.1002513.

15. Conrad N, Judge A, Tran J, et. al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet*. 2018 Feb 10;391(10120):572-580.

DOI: 10.1016/S0140-6736(17)32520-5

16. Ломовской Р. А. Статистический анализ данных: методы и приложения. Актуальные исследования. 2023. №15 (145). с. 29-34.

URL: <https://apni.ru/article/6013-statisticheskij-analiz-dannikh-metodi-i-prilo>

17. Урбах В. Ю. Математическая статистика для биологов и медиков. М. Изд-во АН СССР, 1963, 323 с.
18. Сепетлиев Д. А. Статистические методы в научных медицинских исследованиях. М., «Медицина», 1968, 419 с.
19. Федорович Г.В., Эпидемиологический анализ характеристик профессионального риска. БиОТ № 3, 2012 г., с. 49-53
20. Руководство Р 2.2.1766-03 по оценке профессионального риска для здоровья работников. Организационно-методические основы, принципы и критерии оценки. Минздрав России, Москва 2004
21. ГОСТ Р 12.0.010–2009. Национальный стандарт Российской Федерации. Системы управления охраной труда. Определение опасностей и оценка рисков. Москва 2009.
22. Федорович Г.В., Зависимость «доза-эффект» в гигиене труда. (Риск-ориентированный подход), Saarbrücken, DeuOMchland, Palmarium Academic Publishing, ISBN-13: 978-620-2-38060-7. 2017:201.
[URL: http://www.palmarium-publishing.ru/;](http://www.palmarium-publishing.ru/)
[https://www.morebooks.shop/store/gb/book/;](https://www.morebooks.shop/store/gb/book/)
23. Федорович Г.В. О методике исследования доз вредных производственных воздействий// Безопасность труда в промышленности. 2022. № 6. С. 60–67.
 DOI: 10.24000/0409-2961-2022-6-60-67
24. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. Evidence-Based Medicine Working Group. JAMA. 1993; 270(21): 2598-2601. PMID: 8230645.
 DOI: 10.1001/jama.270.21.2598.
25. Тарловская Е.И. Коморбидность и полиморбидность – современная трактовка и насущные задачи, стоящие перед терапевтическим сообществом // Кардиология; 2018 ; № 58 (S9); стр. 29–38.
 DOI: 10.18087/cardio.2562.

26. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med*. 2012 Mar-Apr;10(2):142-51. PMID: 22412006; DOI: 10.1370/afm.1337.
27. Белялов Ф.И. Лечение внутренних болезней в условиях коморбидности: монография. – Иркутск: РИОИГИУВа, 2012 ; 285 с.
28. Федорович Г.В. Таблицы сопряженности в эпидемиологии профессиональных заболеваний, *Фундаментальная медицина*; 2014 (06.10) URL; <http://www.medlinks.ru/article.php?sid=60880>
29. Хей Дж., Введение в методы байесовского статистического вывода. — М.: Финансы и статистика, 1987. 336 с.
30. Аптон Г., Анализ таблиц сопряженности. М.: Финансы и статистика, 1982. 143 с.
31. Youden, W.J. Index for rating diagnostic tes.; *Cancer*, (1950) № 3, pp. 32–35.
32. Yule G. On the Methods of Measuring Association Between Two Attributes, *Journal of the Royal Statistical Society*, (1912), Vol. 75, №. 6, pp. 579–652.
33. Joffe, M., Gambhir, M., Chadeau-Hyam, M. *et al.* Causal diagrams in systems epidemiology; *Emergig Themes in Epidemiology*, (2012) v. 9, №1. DOI: [10.1186/1742-7622-9-1](https://doi.org/10.1186/1742-7622-9-1).
34. Федорович Г. В., Теорема Т. Байеса в профэпидемиологии. Безопасность и охрана труда. 2023; № 2; стр. 5–9; УДК 331.45. DOI: 10.54904/52952_2023_2_5_9
35. Сюрин С.А. Вдох или выдох; *Безопасность и охрана труда*, № 2, 2013 г., стр. 67 – 72.
36. Сюрин С.А. Особенности формирования профессиональной патологии у работников различного передела никеля в условиях Крайнего Севера. *БиОТ*, № 1, 2012 г., с.50 – 51.

37. Кусраева З.С. Современный подход к оценке профессионального риска при выполнении электродуговой сварки и резки металлов // Доклад на XI съезде гигиенистов. М., апрель 2012 г.
38. Федорович Г.В., Классификация условий труда по эпидемиологическим данным // Безопасность и охрана труда, № 4, 2011 г., с.49 – 52
39. Федорович Г.В., Эпидемиологические методы диагностики профессиональных заболеваний. БиОТ № 3 2018 г. с. 20-28
40. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med.* 2012 Mar-Apr;10(2):142-51. DOI: 10.1370/afm.1337.
41. Шпагина Л.Н. Полипатии у шахтеров с патологией периферической нервной системы и костно-мышечной системы: формирование и особенности структуры. Фундаментальные исследования. – 2012. – № 8 (часть 1) – С. 190-192
42. Краснов А. А., Синенченко А. Г., Козлова С. Н. Психосоматическая коморбидность как фактор риска развития неотложных состояний у пациентов с пограничными психическими расстройствами. *Мед. вест. Северного Кавказа.* 2022;17(1):17-20.
DOI – <https://doi.org/10.14300/mnnc.2022.17005>
43. Селье Г. Очерки об адаптационном синдроме. — М.: Медгиз, 1960. — 255 с.
44. Наумова Л.А., Осипова О.Н. // Коморбидность: механизмы патогенеза, клиническое значение // Современные проблемы науки и образования; 2016 г. № 5;
URL: <https://science-education.ru/ru/article/view>
45. Виноградов В.В. Гормоны, адаптация и системные реакции организма. Минск: Наука и техника. — 1989. — 223 с.
46. Chrysant S.G. A new paradigm in the treatment of the cardiovascular disease continuum: focus on prevention; *Hippokratia*; 2011, v. 15, № 1; P. 7-11