

# Интеллектуальная система прогнозирования целесообразности применения компьютерной томографии

О. П. Шестерникова <sup>I</sup>, В. К. Финн <sup>II</sup>, К. А. Лесько <sup>III</sup>, Л. В. Винокурова <sup>III</sup>

<sup>I</sup> Центральный научно-исследовательский институт организации и информатизации здравоохранения МЗ РФ, Москва, Россия

<sup>II</sup> Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия

<sup>III</sup> Московский клинический научный центр им. А.С. Логинова, Москва,

**Аннотация.** В статье описываются принципы создания интеллектуальной системы, использующей ДСМ-метод автоматизированной поддержки исследования (ДСМ-метод АПИ), для прогнозирования целесообразности применения компьютерной томографии. Представлены процедуры одного из этапов применения ДСМ-метода АПИ – ДСМ-исследования, предназначенные для повышения надежности получаемых в системе эмпирических закономерностей. Приводятся полученные эмпирические закономерности и их экспертная оценка.

**Ключевые слова:** интеллектуальный анализ данных, ДСМ-метод, автоматизированная поддержка исследования, компьютерная томография, рак поджелудочной железы, хронический панкреатит, дифференциальная диагностика.

DOI 10.14357/20718594220201

## Введение

В [1, 2] описывается создание интеллектуальной системы (ИС), реализующей ДСМ-метод автоматизированной поддержки исследований (ДСМ-метод АПИ) для ранней дифференциальной диагностики заболеваний поджелудочной железы (ПЖ): рака поджелудочной железы (РПЖ) и хронического панкреатита (ХП). С помощью этой системы было проведено исследование, затрагивающее непосредственно дифференциальную диагностику, которое, в том числе, показало высокую связь результатов компьютерной томографии (КТ) и окончательного диагноза. Система сделала верный вывод в шести из девяти наблюдений проверочной группы. В остальных наблюдениях ошибочные выводы были вызваны тяжелыми

формами ХП (наличие ряда сопутствующих заболеваний, оказывающих негативное влияние на тяжесть течения патологического процесса в ПЖ) [2]. С помощью созданной системы и собранного массива данных можно решать другие близкие исследовательские задачи. Следующей такой задачей стало определение необходимости направления на исследование КТ исходя из клинических данных пациента и, таким образом, формирование скрининговых групп.

## 1. Постановка задачи

Одной из особенностей современной медицины является повсеместное использование методов лучевой диагностики, в первую очередь КТ. Данная методика занимает одну из ключевых позиций в диагностической процедуре при распо-

✉ Шестерникова Ольга Павловна. E-mail: oshesternikova@gmail.com

знании заболеваний поджелудочной железы, в первую очередь ХП и РПЖ. Значения чувствительности и специфичности КТ с внутривенным контрастированием в диагностике РПЖ достигают 82 и 65%, соответственно. Невысокие значения специфичности нивелируются высокой доступностью и пространственным разрешением КТ, позволяющим проводить одномоментную диагностику распространенности поражения и наличия сопутствующих изменений [3].

Во всем мире актуальность своевременной диагностики РПЖ несомненна и обусловлена высокой смертностью от этого заболевания. В структуре смертности от злокачественных новообразований в РПЖ занимает уже 6,7% в России в 2019 г. и 4,5% в мире в 2020 г. [4, 5]. Однако высокая эффективность и постоянно растущая доступность КТ ставит перед практикующими врачами важную задачу правильного определения показаний к проведению исследования.

Опираясь на известные и доказанные факторы, влияющие на возникновение и развитие ХП или РПЖ, система автоматизированной поддержки исследований с применением ДСМ-метода помогает выполнить эту задачу. Решение этой проблемы отразится в снижении ненужных исследований КТ пациентам со стабильным течением ХП и существенно сократит время ожидания КТ для пациентов с тяжелым течением ХП и подозрением на РПЖ. В итоге это приведет к возможному снижению смертности населения России от РПЖ.

В [6] представлена классификация методов анализа данных в медицине в соответствии с решаемыми задачами. Рассматриваемую задачу можно отнести к «обследованию» – постановке диагноза на основе совокупности симптомов. Основным методом решения такого рода задач с учителем указываются деревья решений [7]. В [8] методы построения деревьев решений также отмечаются как наиболее эффективные. В числе достоинств деревьев решений указываются простота интерпретации принимаемого решения с помощью порождаемых правил (так называемый «белый ящик»), а также вероятностный формат выходных данных, что интуитивно понятно для исследователя-медика.

Примерами исследований с деревьями решений могут служить работы: [9] – исследование связи между питанием и риском возникновения рака мочевого пузыря (использовался алгоритм C5.0); [10] – поиск новых биомаркеров

при развитии диабета 2-го типа из гестационного диабета и многие другие.

ДСМ-метод АПИ строит модель причинно-следственных отношений в предметной области. В отличие от деревьев решений в нем каждый пример из исходного массива соотносится не с одним классификационным правилом, а с несколькими гипотезами о причине. Также в ДСМ-методе итоговая классификация связана не с вероятностью, а с эффективной оценкой истинностного значения: +1 (для отнесения к группе с наличием эффекта), -1 (для отнесения к группе с отсутствием эффекта), 0 (для отнесения к фактическому противоречию),  $\tau$  (для ситуации неопределенности). Формирование скрининговых групп в нашем исследовании осуществлялось на основании различных возможных отношений причинности.

## 2. Методы, используемые в системе

Нулевым этапом применения ДСМ-метода АПИ можно считать подготовку базы фактов (БФ) из «сырых» данных: разработка языка представления фактов (их еще называют примерами), определение примеров с наличием эффекта (положительных) и его отсутствием (отрицательных), определение операции сходства.

Ядром ДСМ-метода АПИ является ДСМ-рассуждение, которое представляет собой синтез процедур индукции (порождаются гипотезы о причинах), аналогии (порождаются гипотезы о предсказаниях) и абдукции (принятие полученных гипотез посредством объяснимости исходного массива). Для порождения гипотез о причине используется понятие стратегии ( $Str$ ), определяемой парой предикатов сходства для положительных и отрицательных примеров. В системе в качестве предикатов сходства использовались 4 метода [12]: простой метод сходства ( $a$ ), метод сходства с запретом на контрпример ( $ab$ ), метод различия ( $ad_0$ ) и метод различия с запретом на контрпример ( $abd_0$ ), которые в итоге формируют 16 возможных стратегий. Формальное определение предикатов сходства приведено в Приложении 1. Имя стратегии имеет вид  $Str_{x,y}$ , где  $x, y \in \{a, ab, ad_0, abd_0\}$ . Множество стратегий образуют изоморфные решетки по логической силе для каждого из четырех правил индуктивного вывода [12]. Стратегия характеризует тип причинности в предметной области. Пример решетки для

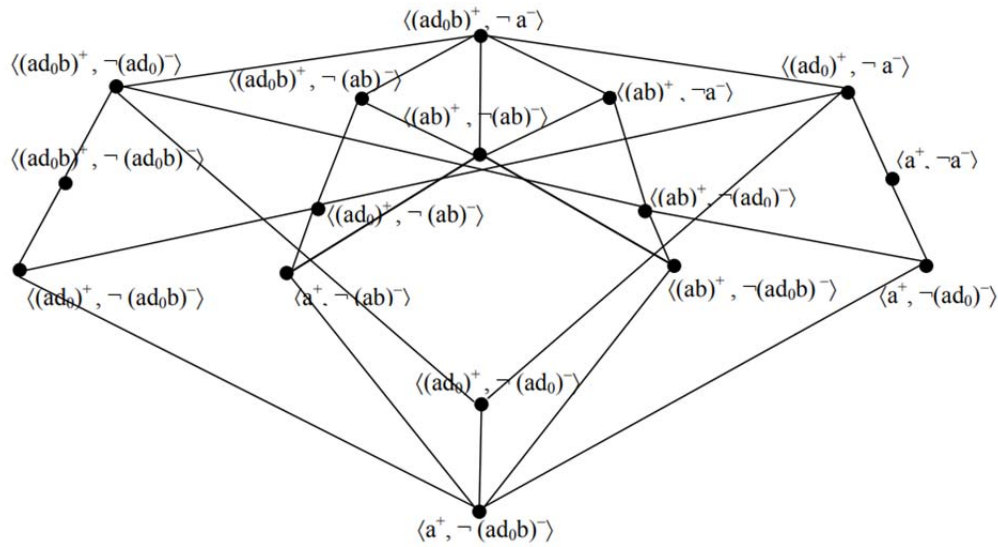


Рис. 1. Диаграмма стратегий для правила индуктивного вывода положительных гипотез

правила индуктивного вывода положительных гипотез в виде диаграммы приведен на Рис. 1.

Применение ДСМ-рассуждения к расширяющимся базам фактов называется ДСМ-исследованием. Главным результатом ДСМ-исследования является нахождение эмпирических закономерностей. Алгоритм для нахождения эмпирических закономерностей в системе основывался на определении эмпирической закономерности, данном в [11], и учитывал сохранения оценок истинностных значений гипотез во всех возможных перестановках расширений массива для минимизации случайности расширения, а также монотонное неубывание показателя объясняемости полученными гипотезами исходного массива. Тип закономерности (определяется согласно дереву Т-классификаций в [11] и обозначается латинскими буквами от *a* до *n*) характеризует устойчивость гипотезы при расширении данных. Типам закономерностей соответствует отношение частичного порядка с наибольшим и наименьшим элементами. На Рис. 2 приведена диаграмма этого частичного порядка.

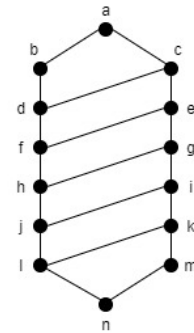


Рис. 2. Диаграмма типов закономерностей

Таким образом, в системе эмпирические закономерности получают оценку вида  $\langle z, Str \rangle$ , где *Str* – стратегия, *z* – тип закономерности. Этой оценке соответствует прямое произведение частичных порядков для стратегий и типов закономерностей. Оценка может быть усилена параметром *k* ( $\langle z, Str, k \rangle$ ), обозначающим количество родителей гипотезы, т.е. объектов исходного массива, порождающих эту гипотезу.

Такая оценка соответствует шкале качества гипотез [11, таблица 5]. Другим результатом ДСМ-исследования является разбиение множества стратегий относительно отношений эквивалентности. Отношение эквивалентности стратегий является средством моделирования предметной области. Оно может задаваться несколькими способами. В [11] приведены два способа определения эквивалентности: слабая эквивалентность (Df.21–3, совпадают правильные предсказания, в нашем исследовании обозначается  $E_1$ ) и сильная эквивалентность (Df.20–3, совпадают правильные предсказания и ошибки, в нашем исследовании обозначается  $E_{13}$ ). Также в системе была добавлена эквивалентность стратегий по эмпирическим закономерностям (обозначается  $E_2$ ). Эквивалентность стратегий является частью модели предметной области. Представляется полезным рассмотреть

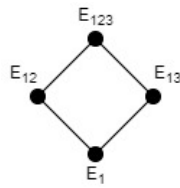


Рис. 3. Решетка отношений эквивалентности стратегий

отношения эквивалентности, образованные конъюнкцией условий:  $E_{12} = E_1 \& E_2$  и  $E_{123} = E_{13} \& E_2$ . Определяемые отношения эквивалентности образуют решетку, представленную на Рис. 3.

### 3. Проведение исследования

На момент проведения исследования массив данных по пациентам содержал 279 историй болезни. Для возможности использования средств нахождения эмпирических закономерностей массив был разделен на 4 расширения, близких по количеству входящих в них примеров. Дополнительно использовались 20 примеров для предсказания, не входящих в исходные массивы. Положительными примерами были пациенты, которым рекомендовано прохождение КТ (179 из 279), отрицательными — те пациенты, для которых прохождение КТ не обязательно (80 из 279). Этот дополнительный признак задавался экспертом. На основании созданной расширяющейся последовательности баз фактов были получены эмпирические закономерности, соответствующие скрининговым группам.

В качестве признаков были выбраны возраст (пациенты разбивались по декадам), наличие в анамнезе алкогольной и табачной зависимостей, индекс массы тела, длительность заболевания, боль (степень силы), наличие в анамнезе сахарного диабета (СД). Все признаки относились к типу «выбор одного из возможных качественных значений» [13] и имели соответствующее сходство, определяемое как совпадение выбранного значения.

Были правильно предсказаны все 20 примеров, заданных для предсказания, из которых полученные эмпирические закономерности предсказывают 17 из 20 (85%) примеров. К закономерностям были добавлены гипотезы не являющиеся эмпирическими закономерностями, но имеющими предсказательную силу. Та-

ким образом были предсказаны оставшиеся 3 примера. Минимальные<sup>1</sup> из полученных гипотез были предоставлены экспертам для оценки по шкале от 1 до 3, где 1 означает необоснованность с медицинской точки зрения, а 3 — максимальную обоснованность. После этого все предсказанные примеры были добавлены к расширениям, а экспертами была предложена новая группа для предсказания, состоящая из 10 примеров, также не содержащихся в исходных массивах. К ним была применена аналогичная ДСМ-эвристика исследования. В результате были получены эмпирические закономерности, предсказывающие 9 из 10 примеров (90%). Гипотезы, имеющие предсказательную силу, но не являющиеся закономерностями, в этом случае не были найдены. Один пример остался непредсказанным (Рис. 4 и 5).

### 4. Результаты

Положительные закономерности в первом эксперименте получены стратегиями  $Str_{ab,ab}$ ,  $Str_{ab,abd0}$ ,  $Str_{a,ab}$ ,  $Str_{a,abd0}$ , остальные не порождают закономерности (Рис. 6).

Закономерности были типизированы как  $a$  (наибольший элемент),  $b$  (отличается от  $a$  тем, что не выполняется условие монотонного роста объясняемости при расширении массива) и  $n$  (наименьший элемент). Оценки, полученные закономерностями, соответствуют поддиаграмме прямого произведения частичных порядков стратегий и типов закономерностей (Рис. 7).

Отрицательные закономерности получены единственной стратегией  $Str_{abd0,a}$ , являющейся минимальной по логической силе для правила индуктивного вывода отрицательных гипотез, и все имели тип  $a$ .

Во втором эксперименте эмпирические закономерности были получены теми же стратегиями, что и в первом эксперименте, однако их типы изменились: вместо  $b$  появился тип  $a$ , для которого характерен монотонный рост объясняемости. Поддиаграмма для положительных закономерностей во втором эксперименте приведена на Рис. 8.

<sup>1</sup>Под минимальными имеются ввиду гипотезы, такие, что никакая другая полученная гипотеза не состоит из подмножества признаков этих гипотез. Формальное описание минимальных гипотез, например, в [14].

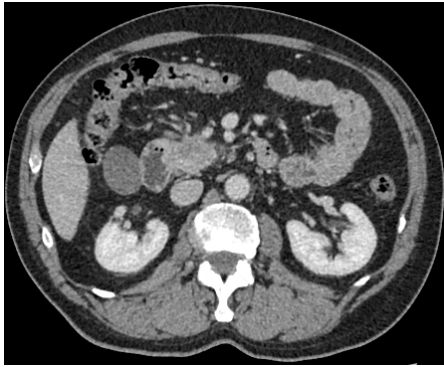


Рис. 4. Пациент К., 59 лет.  
КТ с внутривенным контрастированием в аксиальной плоскости. Морфологически подтвержденный РПЖ (Т3N0M0). Картина новообразования в головке поджелудочной железы. Высокая клиническая обоснованность необходимости в проведении КТ



Рис. 5. Пациентка Д., 60 лет.  
Хронический панкреатит без признаков осложнений заболевания. Картина диффузных изменений. Высокая клиническая обоснованность отсутствия необходимости в проведении КТ

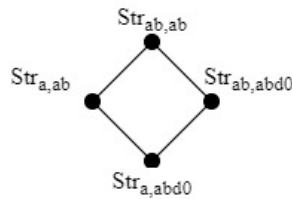


Рис. 6. Поддиаграмма стратегий, порождающих закономерности, для правила индуктивного вывода положительных гипотез

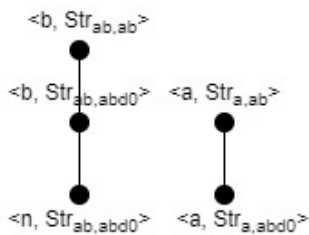


Рис. 7. Поддиаграмма прямого произведения частичных порядков стратегий и типов закономерностей для положительных эмпирических закономерностей (первая группа на предсказание)

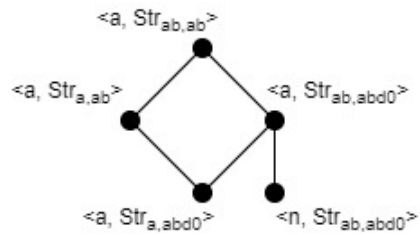


Рис. 8. Поддиаграмма произведения решеток оценок положительных эмпирических закономерностей для второй группы на предсказание

Все минимальные гипотезы о причине понижения оценок, соответствующие эмпирическим закономерностям, с оценками эксперта перечислены в Приложении 2. Стоит отметить, что в целом понижение оценок согласуется с уменьшением силы стратегий. Самая сильная по логической силе стратегия ( $Str_{ab,ab}$ ) получила все максимальные оценки, самая ми-

нимальная по силе ( $Str_{a,abd0}$ ) — ни одной максимальной в случае первого эксперимента, и только одну максимальную в случае второго.

Для последнего расширения в обоих экспериментах были посчитаны правильные предсказания (обозначаются  $l_0$ ) и ошибки, классифицированные в три типа:  $a$  — грубые ошибки (примеры с наличием эффекта предсказаны как

примеры с отсутствием эффекта и наоборот), *b* – негрубые ошибки (примеры предсказаны как «фактическое противоречие») и *c* – отказ от предсказания системой. Таблицы с данными представлены в Приложении 3.

Классификация предсказаний используется в формировании разбиения множества стратегий на классы эквивалентности по отношениям  $E_1, E_2, E_{13}, E_{123}$ . Отношения эквивалентности, учитывающие эмпирические закономерности ( $E_2$  и  $E_{123}$ ), рассматривались отдельно для положительных и отрицательных закономерностей. Результаты для обоих экспериментов представлены в Табл. 1 и 2. Каждая строка в таблице соответствует разбиению множества стратегий на классы эквивалентности для одного из отношений. Если в класс эквивалентности входит несколько стратегий, то ячейки соответствующей строки для них объединены, и для класса приведены поддиаграммы диаграмм стратегий для правил вывода, которые включают стратегии, формирующие класс. Также

стоит отметить, что отношение может быть определено не на всем множестве исходных стратегий (ячейки, соответствующие стратегиям, для которых отношение не определено, выделены серым цветом). Большое число полученных классов эквивалентности содержат только одну стратегию, что может свидетельствовать о том, что хорошо определен тип причинности.

### 5. Интерпретация результата

Важность ранней диагностики РПЖ и осложнений ХП несомненна и диктует необходимость рационального использования возможностей высокотехнологичных методов визуализации, в первую очередь, КТ. Данный метод позволяет быстро и достаточно точно проводить раннюю диагностику РПЖ и осложнений ХП.

Широкий спектр показаний к проведению КТ при заболеваниях поджелудочной железы и большое количество таких пациентов создает

Табл. 1. Классы эквивалентности для отношений  $E_1, E_2, E_{13}, E_{123}$ , эксперимент 1

	Str <sub>ab,ab</sub>	Str <sub>a,abd0</sub>	Str <sub>ab,abd0</sub>	Str <sub>a,ab</sub>	Str <sub>ad0,ad0</sub>	Str <sub>abd0,ad0</sub>	Str <sub>ad0,a</sub>	Str <sub>ad0b,a</sub>	Str <sub>ab,abd0</sub>	Str <sub>ad0,ab</sub>	Str <sub>ad0b,ab</sub>	Str <sub>ab,a</sub>	Str <sub>ad0,ad0b</sub>	Str <sub>a,ad0</sub>	Str <sub>ad0b,ad0b</sub>	Str <sub>a,a</sub>
$E_1$	Кл <sup>2</sup> 1	Кл2	Кл3	Кл4					Кл6	Кл7	Кл8	Кл9				
$E_2^+$	Кл1	Кл2	Кл3	Кл4												
$E_2^-$								Кл1								
$E_{13}$	Кл1	Кл2	Кл3	Кл4	Кл5	Кл6	Кл7	Кл8	Кл9	Кл10	Кл11	Кл12	Кл13	Кл14	Кл15	Кл16
$E_{123}^+$	Кл1	Кл2	Кл3	Кл4												
$E_{123}^-$								Кл1								

Табл. 2. Классы эквивалентности для отношений  $E_1, E_2, E_{13}, E_{123}$ , эксперимент 2

	Str <sub>ab,ab</sub>	Str <sub>a,abd0</sub>	Str <sub>ab,abd0</sub>	Str <sub>a,ab</sub>	Str <sub>ad0,ad0</sub>	Str <sub>abd0,ad0</sub>	Str <sub>ad0,a</sub>	Str <sub>ad0b,a</sub>	Str <sub>ab,abd0</sub>	Str <sub>ad0,ab</sub>	Str <sub>ad0b,ab</sub>	Str <sub>ab,a</sub>	Str <sub>a,ad0</sub>	Str <sub>ad0,ad0b</sub>	Str <sub>ad0b,ad0b</sub>	Str <sub>a,a</sub>
$E_1$	Кл1	Кл2	Кл3	Кл4				Кл6	Кл7			Кл9	Кл10			
$E_2^+$	Кл1	Кл2	Кл3	Кл4												
$E_2^-$								Кл1								
$E_{13}$	Кл1	Кл2	Кл3	Кл4	Кл5		Кл6	Кл7	Кл8	Кл9	Кл10		Кл11		Кл12	
$E_{123}^+$	Кл1	Кл2	Кл3	Кл4												
$E_{123}^-$								Кл1								

<sup>2</sup> Кл — здесь и далее в таблицах сокращение от класса эквивалентности.

<sup>3</sup> Приведена только поддиаграмма для правил вывода отрицательных гипотез, так как стратегии, входящие в класс эквивалентности, правильно предсказывают только отрицательные примеры

<sup>4</sup> Приведена только поддиаграмма для правил вывода положительных гипотез, так как стратегии, входящие в класс эквивалентности, правильно предсказывают только положительные примеры

<sup>5</sup> Приведена только поддиаграмма для правил вывода отрицательных гипотез, так как стратегии, входящие в класс эквивалентности, правильно предсказывают только отрицательные примеры

сложности для врачей-клиницистов – терапевтов, гастроэнтерологов и хирургов – при определении необходимости проведения КТ. Кроме того, также возникают условия для увеличения сложности для врачей-рентгенологов и службы лучевой диагностики в целом за счет повышения трудовой нагрузки и снижения доступности обследования для пациентов.

Включенные в БФ признаки входят в набор фактов, которые врач выясняет у пациента при сборе анамнеза и осмотре. Такой подход соответствует принципам клинического мышления врача и полностью оправдан для оценки необходимости проведения КТ. Использование анамнестических фактов, являющихся доказанными факторами риска развития РПЖ, предикторами тяжелого течения или признаками скрытого течения указанных заболеваний позволяет расценить результаты данного исследования логичными и достоверными [15]. Отобранные факты для работы ИС могут быть выяснены у пациента на доврачебном этапе оказания медицинской помощи.

Последующее использование полученных в представленном исследовании результатов ускорит анализ данных о состоянии здоровья пациента и его анамнезе, что не только упростит работу врача, но и снизит вероятность дефектов в работе специалиста, связанных с информационной перегрузкой. ИС не ставит себе целью заменить клиническое мышление врача, но будучи внедренной в клиническую практику ИС, окажет значимую консультативную помощь врачу в обработке массива информации. Более того, рациональный подход к назначению КТ,

который можно осуществить с помощью представленного варианта ИС, снизит избыточную нагрузку на службу лучевой диагностики и повысит доступность метода. В конечном счете такой подход улучшит возможности ранней диагностики РПЖ и приведет к снижению смертности от злокачественных новообразований.

Таким образом, применение ИС открывает широкие возможности для врачей благодаря интегральному подходу к ускоренному анализу каждого пациента и прецизионному определению необходимости проведения КТ пациентам с заболеваниями поджелудочной железы.

## Заключение

Результатом исследования стал интеллектуальный анализ данных, т.е. применение ИС-ДСМ (ИС, реализующей ДСМ-метод АПИ) для получения обоснованных результатов, учитывающих типы причинных отношений [16]. Окончательный результат представлен в виде внутренних оценок системы, образованных парой стратегия-закономерность ( $\langle z, Str \rangle$ ), что говорит о спецификации характера причины и конструктивной обоснованности гипотез. Оценка уточняется за счет добавления количества примеров, порождающих гипотезу ( $k$ ). Внутренняя оценка хорошо согласуется с оценкой эксперта. Два описанных в статье эксперимента свидетельствуют как об адекватности применяемых методов, так и об удачной параметризации, предложенной экспертами. Оба эксперимента реализовали высокоэффективное применение ДСМ-метода.

## Приложение 1. Определения предикатов сходства

$M_{a,n}^{\sigma}(V, W)$  ( $\sigma = +, -, 0, \tau$ ) являются минимальными предикатами, реализующими индуктивные выводы. Они могут быть дополнены более информативными усилениями.

Запрет на контрпример:

$$(b)^+ \forall X \forall Y ((V \square X) \& (W \square Y)) \rightarrow (J_{(1,n)}(X \Rightarrow_1 Y) \vee J_{(\tau,n)}(X \Rightarrow_1 Y))$$

$$(b)^- \forall X \forall Y (((V \square X) \& (W \square Y)) \rightarrow (J_{(-1,n)}(X \Rightarrow_1 Y) \vee J_{(\tau,n)}(X \Rightarrow_1 Y)))$$

$$M_{ab,n}^+ (V, W) \Leftrightarrow M_{a,n}^+ (V, W) \& (b)^+, M_{ab,n}^- (V, W) \Leftrightarrow M_{a,n}^- (V, W) \& (b)^-$$

Метод различия:

$$(d_0)^+ \forall X \forall Y \forall Z \forall U ((J_{(1,n)}(X \Rightarrow_1 Y) \& (W \square Y) \& (V \square X) \& ((X - V) \square Z) \& ((X - V) \neq \emptyset) \& \neg(V \square Z)) \rightarrow \neg J_{(1,n)}(Z \Rightarrow_1 U) \vee \neg(W \square U))$$

$$(d_0)^- \forall X \forall Y \forall Z \forall U ((J_{(-1,n)}(X \Rightarrow_1 Y) \& (W \square Y) \& (V \square X) \& ((X - V) \square Z) \& ((X - V) \neq \emptyset) \& \neg(V \square Z)) \rightarrow \neg J_{(-1,n)}(Z \Rightarrow_1 U) \vee \neg(W \square U))$$

6 Здесь « $\leftrightarrow$ » обозначает операцию булевой разности

$$M_{ad0,n}^+(V,W) \Leftrightarrow M_{a,n}^+(V,W) \& (d_0)^+, M_{ad0,n}^-(V,W) \Leftrightarrow M_{a,n}^-(V,W) \& (d_0)^-$$

Метод различия с запретом на контрпримеры:

$$M_{ad0b,n}^+(V,W) \Leftrightarrow M_{a,n}^+(V,W) \& (b)^+ \& (d_0)^+, M_{ad0b,n}^-(V,W) \Leftrightarrow M_{a,n}^-(V,W) \& (b)^- \& (d_0)^-$$

## Приложение 2. Перечень полученных закономерностей

Признаки и сокращения:

- Возраст разбит по декадам: 3 — 30-39, 4 — 40-49, 5 — 50-59, 6 — 60-69, 7 — 70-79, 8 — 80-89;
- Алкоголь (алкогольная зависимость) имеет возможные значения Да/Нет;
- Табакокурение (табачная зависимость) имеет возможные значения Да/Нет;
- Индекс массы тела (ИМТ) имеет значения: Ниже — ниже нормы, Норм — норма, Выше — выше нормы;
- Длительность заболевания сравнивалась по ключевым точкам: 1 год, 2 года, 3 года, 5 лет;
- Боль (выраженность боли) имеет возможные значения Да (сильная боль)/Нет (несильная боль);
- Сахарный диабет (СД) имеет возможные значения Да/Нет.

$k$  — количество «родителей» гипотезы (примеров, которые ее порождают).

Если соответствующая закономерность порождается несколькими стратегиями – выбиралась наибольшая (так как стратегии, которыми получены закономерности в наших экспериментах, оказались сравнимы). Таким образом, каждой эмпирической закономерности была приписана одна оценка.

### Эксперимент 1. Положительные закономерности

$\langle b, Str_{ab,ab} \rangle$

#	Возраст (декада)	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6 (Сильная)	СД	$k$	Оценка эксперта
2105				Норм	< 3 лет	Да		17	3
2168		Нет			< 3 лет	Да		16	3
132			Нет		< 2 лет	Да		13	3
205					< 1 года	Да		12	3
2114	5				< 3 лет	Да		6	3
3145		Нет	Нет		От 1 до 2 лет			6	3
3321	6				< 2 лет	Да		6	3
2178	5	Нет		Норм		Да		5	3
2538	5	Нет				Да	Нет	4	3
2548	5		Да		< 3 лет		Нет	4	3
3028	5				От 1 до 3 лет		Нет	4	3
3314	6	Да			< 5 лет		Нет	4	3
3315		Да			< 1 года		Нет	4	3
9		Да	Нет		< 5 лет	Да	Нет	3	3
1520	5			Норм		Да	Да	3	3

$\langle b, Str_{ab,abd0} \rangle$

#	Возраст (декада)	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6 (Сильная)	СД	$k$	Оценка эксперта
1537	5	Нет				Да		6	3

$\langle a, Str_{a,ab} \rangle$

#	Возраст (декада)	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6 (Сильная)	СД	$k$	Оценка эксперта
33				Норм		Да		23	3
689				Норм	< 2 лет			19	2
207		Нет			< 2 лет			18	3
1089			Нет			Да		18	2
1280		Нет			< 5 лет	Да		17	2



#	Возраст (декада)	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6 (Сильная)	СД	k	Оценка эксперта
1832			Нет		< 2 лет			16	3
1267		Нет				Да	Нет	15	2
1837					< 1 года			14	3
795	5				< 5 лет			9	2
794		Да		Норм	< 5 лет			8	2
1178	6				< 5 лет	Да		8	3
1428	5						Нет	8	2
2120					> 2 лет	Да	Нет	8	1
3618	6				< 3 лет		Нет	8	2
1834	6		Нет		< 5 лет		Нет	7	3
1644		Да	Нет				Нет	6	2
2552		Нет	Да		< 3 лет			6	3
3609	6			Норм	< 3 лет			6	3
69			Нет	Выше			Нет	5	2
681	5		Да	Норм				5	3
3411	4		Да		> 1 года		Нет	5	1
2276		Нет		Выше			Нет	4	1
3314	6	Да			< 5 лет		Нет	4	2
3424	4	Да			> 1 года			4	1

 $\langle n, \text{Str}_{ab,abd0} \rangle$ 

#	Возраст (декада)	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6 (Сильная)	СД	k	Оценка эксперта
23073	6		Да		От 2 до 5 лет		Нет	3	2

 $\langle a, \text{Str}_{a,abd0} \rangle$ 

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта
694				Норм	< 5 лет			23	1
1265		Нет				Да		21	2
257		Нет			< 5 лет			20	2
4			Нет				Нет	19	1
802			Нет		< 5 лет			19	2
2744		Нет	Нет	Норм				11	1
1843	6				< 5 лет		Нет	10	2
2103					> 2 лет	Да		10	1
76	5					Да		8	2
54	5		Да					7	2
1274	4				> 1 года		Нет	6	1
2154					От 2 до 5 лет		Нет	5	1
478		Да		Выше	> 3 лет			4	2
3424	4	Да			> 1 года			4	2
1724				Выше	> 3 лет		Нет	3	2
1286	4		Нет		> 1 года			2	2

**Эксперимент 1. Отрицательные закономерности** $\langle a, \text{Str}_{abd0,a} \rangle$ 

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта
69	3				> 3 лет			7	2
353			Нет		> 3 лет			7	2

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта
269					> 1 года	Нет	Да	6	3
546		Нет			> 5 лет	Нет		6	3
308	5					Нет		3	3
399	4			Норм	> 5 лет		Да	3	2
267	6			Выше	> 1 года	Нет		2	3
480	6		Да		От 1 до 2 лет	Нет		2	3
672	4	Нет	Да	Норм	> 2 лет			2	2

### Эксперимент 1. Отрицательные гипотезы, не являющиеся закономерностями

Str<sub>ab,ad0</sub>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта
271					> 1 года			69	3

Str<sub>ab,a</sub>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта
62						Нет		70	2
224							Нет	51	1
4097				Норм				47	1
4100			Да					45	1
364					< 5 лет			33	1
4108		Да						33	1
1249				Выше				25	1
1879				Ниже				8	1

### Эксперимент 2. Положительные закономерности

<a, Str<sub>ab,ab</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
3315		Да			< 1 года		Нет	4	3	Да

<a, Str<sub>a,ab</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
884		Нет		Норм				20	2	Нет
1103			Нет	Норм				15	2	Нет
2520		Нет	Да					13	2	Нет
2101		Нет			> 2 лет			12	2	Нет
1835				Норм	< 1 года			11	3	Нет
4010		Нет					Да	9	2	Нет
1857	6			Норм	< 5 лет			8	3	Нет
2728				Норм	< 2 лет		Да	8	2	Нет
1830		Да			< 1 года			6	3	Нет

<a, Str<sub>ab,abd0</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
3145		Нет	Нет		От 1 до 2 лет			6	3	Да
3315		Да			< 1 года		Нет	4	3	Да

<n, Str<sub>ab,abd0</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
23055	5		Нет	Ниже	> 1 года			3	2	Нет

<a, Str<sub>a,abd0</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
37		Нет			> 1 года			22	2	Нет
75	5							14	1	Нет
121			Нет		> 1 года			14	1	Нет
494							Да	14	2	Нет
2520		Нет	Да					13	2	Нет
1855	6				< 5 лет			12	3	Нет
486				Выше				11	1	Нет
998		Да	Нет					7	1	Нет
601	3	Да		Норм				5	1	Нет

**Эксперимент 2. Отрицательные закономерности**<a, Str<sub>abd0,a</sub>>

#	Возраст	Алкоголь	Табакокурение	ИМТ	Длительность	Боль > 6	СД	k	Оценка эксперта	Была в Эксперименте 1
473				Норм			Нет	7	3	Нет
474			Да				Нет	8	3	Нет
479			Да		< 5 лет	Нет		6	2	Нет

**Приложение 3. Перечень правильных предсказаний и ошибок по стратегиям для последнего расширения**

Для эксперимента 1

	l0		a		b		c	
	+	-	+	-	+	-	+	-
Str <sub>ab,ab</sub>	8	3		3	2	1	1	2
Str <sub>a,abd0</sub>	11			9				
Str <sub>ab,bd0</sub>	10			4			1	5
Str <sub>a,ab</sub>	9			5	2	4		
Str <sub>ad0,ad0</sub>		9	5		3		3	
Str <sub>abd0,ad0</sub>		9	8				3	
Str <sub>ad0,a</sub>		9	8		3			
Str <sub>abd0,a</sub>		9	11					
Str <sub>ab,ad0</sub>	3	5	1		7	4		
Str <sub>ad0,ab</sub>	2	4	1		1		7	5
Str <sub>ab,a</sub>		5	1		10	4		
Str <sub>abd0,ab</sub>		4	2				9	5
Str <sub>ad0,abd0</sub>	3						8	9
Str <sub>a,ad0</sub>	3				8	9		
Str <sub>abd0,abd0</sub>							11	9
Str <sub>a,a</sub>					11	9		

Для эксперимента 2

	l0		a		b		c	
	+	-	+	-	+	-	+	-
Str <sub>a,abd0</sub>	7			3				
Str <sub>ab,abd0</sub>	5						2	3
Str <sub>ab,ab</sub>	4	1	1		1		1	2
Str <sub>a,ab</sub>	5			2	2	1		
Str <sub>ab,ad0</sub>	2	2			3	2		1
Str <sub>ab,a</sub>		3	2		5			

	<i>l0</i>		<i>a</i>		<i>b</i>		<i>c</i>	
	+	-	+	-	+	-	+	-
Str <sub>ad0,a</sub> , Str <sub>abd0,a</sub>		3	7					
Str <sub>a,ad0</sub>	2			1	5			2
Str <sub>ad0,ad0</sub> , Str <sub>abd0,ad0</sub>		2	5				2	1
Str <sub>ad0,ab</sub> , Str <sub>abd0,ab</sub>		1	2				5	2
Str <sub>ad0,abd0</sub> , Str <sub>abd0,abd0</sub>							7	3
Str <sub>a,a</sub>					7	3		

## Литература

- Шестерникова О.П., Финн В.К., Винокурова Л.В., Лесько К.А., Варванина Г.Г., Тюляева Е.Ю.. Интеллектуальная система для диагностики заболеваний поджелудочной железы // Научно-техническая информация. Серия 2. 2019. № 10. С. 41–48.
- Винокурова Л.В., Финн В.К., Лесько К.А., Шестерникова О.П., Варванина Г.Г. Первые результаты изучения возможностей интеллектуальной системы автоматизированной поддержки исследований в дифференциальной диагностике рака поджелудочной железы// Материалы 47-й научной сессии ЦНИИ гастроэнтерологии «Гастроэнтерология в меняющейся реальности». 4–5 марта 2021 г., Москва. Терапевтический архив. 2021; 93 (2, Приложение): 77.
- Zhang J., Zuo C.J, Jia N.Y, et al. Cross-modality PET/CT and contrast-enhanced CT imaging for pancreatic cancer. *World J Gastroenterol.* 2015;21(10). P.2988-2996.
- Состояние онкологической помощи населению России в 2019 году. Под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. М.: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России. 2020. 239 с.
- Sung H., Ferlay J., Siegel R.L., Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021 Feb 4.
- Баранов А. А., Намазова-Баранова Л. С., Смирнов И. В., Девяткин Д. А., Шелманов А. О., Вишнева Е. А., Антонова Е. В., Смирнов В. И., Латышев А. В. Методы и средства комплексного интеллектуального анализа медицинских данных // Труды ИСА РАН. 2015. Том 65. Вып. 2. С.81–93.
- Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание. Пер. с англ. М.: Издательский дом «Вильямс», 2003. 864 с. (Luger, George F. Artificial intelligence : structures and strategies for complex problem solving / George F. Luger.- 4th ed. - Addison-Wesley, 2002).
- Linden A., Yarnold P. R. (2016) Using data mining techniques to characterize participation in observational studies. *J Eval Clin Pract*, 22: 839– 847. doi: 10.1111/jep.12515. URL: <https://onlinelibrary.wiley.com/doi/10.1111/jep.12515>
- Yu E., Wesselius A., Sinhart C., Wolk A., Stern M., Jiang X., Zeegers M. (2020). A data mining approach to investigate food groups related to incidence of bladder cancer in the BLadder cancer Epidemiology and Nutritional Determinants International Study. *British Journal of Nutrition*, 124(6), 611-619.
- Khan S.R., Mohan H., Liu Y., Batchuluun B., Gohil H., Al Rijjal D., Manialawy Y., Cox B.J., Gunderson E.P., Wheeler M.B. The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. *Diabetologia.* 2019 Apr;62(4):687-703.
- Финн В.К., Шестерникова О.П. Эвристика обнаружения эмпирических закономерностей посредством ДСМ-рассуждений // Научно-техническая информация. Серия 2. 2018. № 9. С. 7–42.
- Финн В.К. Дистрибутивные решетки индуктивных процедур//Научно-техническая информация. Серия 2. 2014. № 11. С. 1–30.
- Панкратова Е.С, Виноградов Д.В. Формальное описание настройки интеллектуальных ДСМ-систем // Научно-техническая информация. Серия.2. 2011. № 9. С.1–5.
- Гусакова С.М., Михеенкова М.А. Интеллектуальный анализ данных как инструмент формирования структуры социума // Научно-техническая информация. Серия 2. 2016. № 8. С. 9–18.
- Klein A.P. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat Rev Gastroenterol Hepatol.* 2021;18(7):493-502// Midha S, Chawla S, Garg PK. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. *Cancer Lett.* 2016;381(1):269-277.
- Финн В.К. Точная эпистемология и искусственный интеллект //Научно-техническая информация. Серия. 2. 2020. №6. С.1–36.

**Шестерникова Ольга Павловна.** Инженер-программист отдела разработки программного обеспечения. ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» МЗ РФ. Области исследований: интеллектуальные системы в медицине, анализ данных. E-mail: oshesternikova@gmail.com

**Финн Виктор Константинович.** Доктор технических наук, профессор, Заслуженный деятель науки РФ. Главный научный сотрудник. Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Научный руководитель Отделения интеллектуальных систем в гуманитарной сфере. Российский государственный гуманитарный университет. Области исследований: искусственный интеллект, математическая логика, интеллектуальные системы, машинное обучение, теория познания, логика и методология науки. E-mail: ira.finn@gmail.com

**Лесько Константин Александрович.** Кандидат медицинских наук. Врач рентгенолог. Московский клинический научный центр им. А.С. Логинова. Области исследований: рентгенология, радиология, панкреатология и онкология. E-mail: k.lesko@mknc.ru

**Винокурова Людмила Васильевна.** Доктор медицинских наук. Ведущий научный сотрудник отделения патологии поджелудочной железы и желчных путей. Московский клинический научный центр им. А.С. Логинова. Области исследований: терапия, гастроэнтерология, панкреатология. E-mail: vinokurova52@mail.ru

## Intelligent System for Predicting the Feasibility of Using Computed Tomography

O. P. Shesternikova<sup>1</sup>, V. K. Finn<sup>2</sup>, K. A. Lesko<sup>3</sup>, L. V. Vinokurova<sup>3</sup>

<sup>1</sup> Russian Research Institute of Health, Moscow, Russia

<sup>2</sup> Federal Research Center "Computer Science and Control", the Russian Academy of Sciences, Moscow, Russia

<sup>3</sup> Moscow Clinical Scientific Center named after A.S. Loginov, Moscow, Russia

**Abstract.** The article describes principles of creating an intelligent system using JSM-method of automated research support (JSM-method ARS) to predict the necessity for computed tomography application. The procedures of JSM-research (one of the JSM-method ARS stages) designed to increase the reliability of the regularities obtained in the system are described. The obtained regularities and their expert ratings are given.

**Keywords:** data mining, intelligent data analysis, JSM-method, automated research support, computed tomography, pancreatic cancer, chronic pancreatitis, differential diagnosis.

DOI 10.14357/20718594220201

## References

1. Shesternikova O.P., Finn V.K., Vinokurova L.V., Lesko K.A., Varvanina G.G., Tyulyaeva E.Yu. 2019. *Интеллектуальная система для диагностики заболеваний поджелудочной железы [Intelligent system for diagnostics of pancreatic diseases]. Научно-техническая информация. Серия 2 [Scientific and technical information. Series 2.]* 10: 41-48.
2. Vinokurova L.V., Finn V.K., Les'ko K.A., Shesternikova O.P., Varvanina G.G. 4-5.03.2021. *Первые результаты изучения возможностей интеллектуальной системы автоматизированной поддержки исследований в дифференциальной диагностике рака поджелудочной железы [The first results of studying the capabilities of an intelligent automated research support system in the differential diagnosis of pancreatic cancer]. Материалы 47-й научной сессии ЦНИИ гастроэнтерологии «Гастроэнтерология в меняющемся мире» [47th scientific session of the Central Research Institute of Gastroenterology "Gastroenterology in a changing reality"]. *Терапевтический архив.* 2021; 93 (2, Приложение) [Therapeutic Archive. 2021; 93 (2, Appendix)]: 77.*
3. Zhang J, Zuo CJ, Jia NY, et al. Cross-modality PET/CT and contrast-enhanced CT imaging for pancreatic cancer. *World J Gastroenterol.* 2015;21(10):2988-2996.
4. *Состояние онкологической помощи населению России в 2019 году [The state of oncological care to the population of Russia in 2019.].* Под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. 2020. Moscow: National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, 239 p.
5. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021 Feb 4.
6. Baranov A. A., Namazova-Baranova L. S., Smirnov I. V., Devyatkin D. A., Shelmanov A. O., Vishneva E. A., Antonova E. V., Smirnov V. I., Latyshev A. V. 2015. *Методы и средства комплексного интеллектуального анализа медицинских данных [Methods and means of complex intellectual analysis of medical data].* *Труды ИСА РАН [Proceedings of ISA RAS].* 65.2: 81-93
7. Luger, George F. *Artificial intelligence: structures and strategies for complex problem solving / George F. Luger.* - 4th ed. - Addison-Wesley, 2002
8. Linden, A., and Yarnold, P. R. (2016) Using data mining techniques to characterize participation in observational studies. *J Eval Clin Pract*, 22: 839-847.
9. Yu, E., Wesselius, A., Sinhart, C., Wolk, A., Stern, M., Jiang, X., Zeegers, M. (2020). A data mining approach to investigate food groups related to incidence of bladder cancer in the BLadder cancer Epidemiology and Nutritional Determinants International Study. *British Journal of Nutrition*, 124(6), 611-619.
10. Khan SR, Mohan H, Liu Y, Batchuluun B, Gohil H, Al Rijjal D, Manialawy Y, Cox BJ, Gunderson EP, Wheeler MB. The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. *Diabetologia.* 2019 Apr; 62(4):687-703.
11. Finn V.K., Shesternikova O.P. 2018. *Эвристика обнаружения эмпирических закономерностей посредством DSM-рассуждений [Heuristics of detecting*

- empirical regularities using JSM-reasoning]. Nauchno-tehnicheskaya informaciya. Seriya 2 [Scientific and technical information. Series 2]. 9: 7-42.
12. Finn V.K. 2014. Distributive Lattices of Inductive JSM Procedures. Automatic Documentation and Mathematical Linguistics. Vol. 48. 6: 265-295
  13. Pankratova E.S, Vinogradov D.V. 2011. Formalnoe opisaniye nastrojki intellektualnyh DSM-sistem [Formal description of the configuration of intelligent JSM-systems]. Nauchno-tehnicheskaya informaciya. Seriya 2 [Scientific and technical information. Series 2]. 9: 1-5.
  14. Gusakova S.M., Miheenkova M.A. 2016. Intellektual'nyj analiz dannyh kak instrument formirovaniya struktury sociuma [Data mining as a tool for shaping the structure of society]. Nauchno-tehnicheskaya informaciya. Seriya 2 [Scientific and technical information. Series 2]. 8: 9-18.
  15. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021; 18(7):493-502. doi:10.1038/s41575-021-00457-x // Midha S, Chawla S, Garg PK. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. Cancer Lett. 2016; 381(1):269-277.
  16. Finn V.K. 2020. Tochnaya epistemologiya i iskusstvennyj intellekt [Precise epistemology and artificial intelligence]. Nauchno-tehnicheskaya informaciya. Seriya 2 [Scientific and technical information. Series 2]. 6:1-36

**Shesternikova Olga P.** IT engineer, Russian Research Institute of Health (RIH). Research areas: medicine intelligent systems, data mining. E-mail: oshesternikova@gmail.com

**Finn Viktor K.** Doctor of Technical Sciences, Professor, Honored Scientist of the Russian Federation. Chief Researcher, Federal Research Center "Computer Science and Control", the Russian Academy of Sciences (FRC CSC RAS), Research Head of the Department of Intelligent Systems in Humanities, Russian State University for Humanities (RSUH). Research areas: artificial intelligence, mathematical logic, intelligent systems, machine learning, theory of cognition, logic and methodology of science. E-mail: ira.finn@gmail.com

**Lesko Konstantin A.** Candidate of Medical Sciences. Radiologist, Moscow Clinical Scientific Center named after A. S. Loginov (MCSC). Research areas: radiology, pancreatology and oncology. E-mail: k.lesko@mknc.ru

**Vinokurova Lyudmila V.** Doctor of Medical Sciences. Leading researcher, Moscow Clinical Scientific Center named after A. S. Loginov (MCSC). Research areas: therapy, gastroenterology, pancreatology. E-mail: vinokurova52@mail.ru