

Семантические модели для оценки влияния комплекса факторов на развитие заболеваний

В.В. Грибова, Д.Б. Окунь, Е.А. Шалфеева

Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия

Аннотация

Представлен анализ подходов и решений к прогнозу состояний, развития заболеваний и оценке рисков. Показано, что реализация программных сервисов на различных платформах затрудняет возможность их комплексного использования и выбор между имеющимися решениями. Это обусловило актуальность создания единой семантической модели заболеваний, интегрирующей различные методы и подходы и аккумулирующей знания о рисках и прогнозе в едином информационном пространстве. Предложена новая семантическая модель для учёта влияния комплекса факторов на развитие различных событий, угрожающих здоровью и жизни. Особенностью модели является независимость от конкретного заболевания либо группы заболеваний, что позволяет использовать её в различных разделах медицины. Данная модель апробирована на платформе *IACPaaS*, реализован программный решатель, позволяющий на основе базы знаний и анализа электронной медицинской карты пациента генерировать понятное для врачей объяснение. Применение новой модели для формирования знаний показано на примере оценки рисков и прогноза сердечно-сосудистых событий.

Ключевые слова: семантическая модель, оценка риска заболеваний, база знаний, медицинские интеллектуальные системы, система поддержки решений.

Цитирование: Грибова, В.В. Семантические модели для оценки влияния комплекса факторов на развитие заболеваний / В.В. Грибова, Д.Б. Окунь, Е.А. Шалфеева // *Онтология проектирования*. – 2021. – Т.11, №4(42). – С.464-477. – DOI: 10.18287/2223-9537-2021-11-4-464-477.

Введение

Среди основных задач, решаемых практикующим врачом, есть как диагностика и лечение заболеваний, так и сопутствующие им: мониторинг, оценка состояния, определение рисков заболеваний и жизнеопасных состояний, осложнений и смерти, перспектив развития, возникновения новых заболеваний, течения имеющихся заболеваний; прогноз качества жизни, раннее выявление заболеваний, оценка степени риска летальности и т.п.

Для поддержки решения таких задач применяют различные подходы и методы, начиная от простых расчётных формул (например, калькуляторы оценки риска сердечно-сосудистых заболеваний – ССЗ), до более сложных, основанных на методах математического моделирования (ММ), машинного обучения (МО) и искусственного интеллекта. Используемые методы имеют свои особенности, преимущества и недостатки. Наиболее распространёнными методами, связанными с прогнозом и оценкой рисков, состояний и т.п., являются методы, основанные на математическом моделировании и МО, зачастую они демонстрируют хорошие прогнозные результаты [1-5]. Методы на основе знаний (наборов качественных и количественных признаков, факторов, причинно-следственных и временных связей, часто с использованием нечёткой логики и основанные на моделях мягких вычислений), применяются реже, но также показывают достоверные результаты [6-8].

При решении сходных задач часто применяют разнообразные методы их решения, например, для оценки рисков ССЗ - различные шкалы (*SCORE*, *PROCAM*, *QRISK*, *FINRISK* и

др.), методы ММ и МО [1-4], реализованные на различных платформах. Это затрудняет выбор между имеющимися решениями и возможность их комплексного использования, поскольку большинство из них реализованы как отдельные сервисы, и если врач хочет использовать несколько сервисов, каждое решение требует ввода часто одной и той же информации. Как правило, результатом таких сервисов является вероятность события, состояния или заболевания, что недостаточно для принятия решений и врачу требуется объяснение либо схожая оценка вероятности при использовании других сервисов. Часть предложенных решений ориентирована на пациента и использует ограниченное количество параметров.

Указанные обстоятельства определяют актуальность создания единой семантической модели для оценки влияния комплекса факторов для прогноза и оценки рисков различных событий и заболеваний (далее – семантической модели), интегрирующей различные методы и подходы.

Целью данной работы является описание новой семантической модели и демонстрация её на примере оценки рисков и прогноза ССЗ.

1 Современное состояние сервисов

В разделе приведён обзор современного состояния сервисов в области оценки влияния комплекса факторов для прогноза и оценки рисков различных событий и заболеваний. Классификация программных средств данного назначения выполнена по следующим критериям.

1.1. Целевая аудитория: пациенты или врачи

Для пациентов в сервисах используются упрощённые наборы входной информации, позволяющей по доступным большинству пациентам признакам определить вероятность некоторого состояния или наступления некоторого события с генерацией полезных рекомендаций и предостережений [9-11].

1.2. Интеграция с электронной медицинской картой

Сервисы могут использовать информацию из истории болезни (ИБ), электронной медицинской картой (ЭМК) или быть реализованы как отдельно функционирующие сервисы. В первом случае сервисы (их немного) либо встроены в медицинские информационные системы, либо могут анализировать ЭМК, в том числе не полностью формализованные, имеющие компонент анализа текста (например, *Webiomed – WML.SymptomChecker, klinica.com*). Большинство сервисов реализованы как отдельные программы и от пользователя требуется онлайн заполнение формы для внесения результатов анализа, метрик (рост, вес, возраст, пол, частота сердечных сокращений - ЧСС), особенностей (курение, принимаемые препараты, наследственность и др.).

1.3. Количество анализируемых патологий

Большинство сервисов поддерживают анализ одной патологии (состояния или события); есть сервисы, поддерживающие анализ нескольких патологий. Например, сервис оценки кардиориска попутно может выявить хронические болезни почек (как фактора ССЗ) по результатам анализов [9-12].

1.4. Количество поддерживаемых методик оценки/прогноза заболеваний

Обычно каждая методика реализована отдельным сервисом; редко имеется возможность делать оценку некоторого важного риска по двум методикам (см. например рисунок 1) [12].

1.5. Метод оценки/прогноза

Можно выделить три метода для оценки комплекса факторов на развитие заболеваний и состояний, оценки персонафицированных рисков, наступления событий и т.п.

Первый метод, арифметический, представляется формулой расчёта наступления некоторого события или состояния. Типичными примерами являются [13, 14]: оценка риска смерти от ССЗ в ближайшие 10 лет по шкале *Score*, калькулятор *MELD* для оценки терминальных стадий заболеваний печени, калькулятор лимфомы Ходжкина, калькулятор оценки риска венозных тромбоэмболических осложнений и др. На вход калькуляторов поступает ряд измерений или индексов признаков, факторов (например, фактор курение –

присутствует или отсутствует, иногда указание количества лет). Результатом работы такого калькулятора является вероятность наступления одного или нескольких событий, выраженная либо в числовой шкале (например, вероятность наступления смертельного исхода от ССЗ в течение ближайших 10 лет), либо в категориальной (качественной) шкале (например, риск или польза - значениями низкий, умеренный, высокий и т.п.). Такой калькулятор может иметь условия использования. Например, калькулятор *Score* не используется, если присутствует ряд заболеваний или пороговых значений признаков: установленный диагноз ССЗ; сахарный диабет и др. Возможно также указание на некоторые признаки, факторы и заболевания, например, для *Score* – это ожирение, малоподвижный образ жизни и др.

В шкалы обычно включаются хорошо изученные, легко определяемые показатели состояния здоровья пациента.

Второй метод, наиболее распространенный в настоящее время, использует ММ и МО. Его результатом является модель, основанная на логистической регрессии, нейросетевых алгоритмах и др. Типичными примерами являются модель прогнозирования развития ССЗ в течение 10 лет, прогнозирование смерти от ишемической болезни сердца (ИБС) и инсульта в течение 10 лет, прогнозирование осложнений у пациентов с сахарным диабетом 2 типа в течение 10 лет и др. [4, 5, 12]. Модели логистической регрессии иногда демонстрируют хорошую прогнозирующую ценность [3, 6]. Нейросетевые модели и системы определяют риски в виде вероятности [15-17]. Наиболее эффективно нейросетевые модели используются в области кардиологии [13, 17-20].

Для методов с обучением на множестве примеров и указанным заранее набором показателей пациента свойственна ограниченная применимость – требование к строго фиксирован-

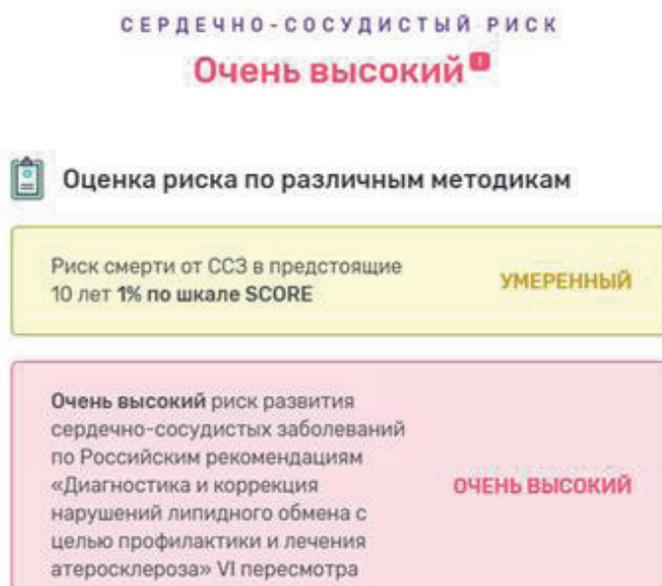


Рисунок 1 - Пример результата оценки риска ССЗ по различным методикам

ному набору параметров пациента [15]. Их число измеряется часто десятками, а отсутствие хотя бы одного в информации о пациенте может стать поводом отклонить метод.

В третьем методе используется подход на основе знаний, когда моделируется вероятность наличия какого-либо заболевания или состояния в настоящий момент времени. Типичными примерами являются многочисленные симптом-чекеры, системы диагностики заболеваний.

Иногда используется смешанный подход. Например, на первом этапе определяется состояние пациента на основе знаний, после чего выполняются вычисления по формуле.

К недостаткам сервисов, реализованных арифметическим способом или на основе методов ММ, относится отсутствие объяснений. Сервисы на основе знаний не всегда можно использовать для оценки или анализа будущих событий.

1.6. Генерация объяснения

Для выбора тактики лечения врачу необходимо объяснение ожидаемых исходов.

Для сервисов, реализованных на основе арифметических вычислений и методов МО, свойственно отсутствие обоснования решений – обнаруженных рисков, диагнозов и прогнозов (*Webiomed – WML.SymptomChecker, WebMD, Isabel*).

Для сервисов, использующих знания, объяснение может генерироваться. Например, у сервиса *Infermedica* [9] по каждому диагнозу из сформированного списка оно даётся в виде «Грипп. Умеренные доказательства – наличие симптомов: быстрое сердцебиение, головная боль, сухой кашель, ...). В некоторых случаях в качестве объяснения можно рассматривать описание типичных признаков, характеризующих состояние или заболевание. Объяснением могут являться вопросы на экране, на которые дан ответ «да» (см., например, *health.mail.ru*). Иногда формируется фрагментарный объясняющий ответ, например: «гнойный налет с миндалин не снимается. Можно предположить дифтерию» [21].

Системы поддержки принятия решения на основе онтологических знаний выдают детализированное объяснение [21-23]. Для его генерации используется информация о соответствии вариантов решения, получаемых из базы знаний (БЗ), информации об исходных данных задачи. Объяснение сводится к информации обо всех рассмотренных гипотезах при формировании решения, а также источниках знаний, по которым была создана БЗ. Понятность обеспечивается используемой терминологией и структурой (онтологией) объяснения. Объяснение, представленное в форме семантической сети, позволяет пользователю просмотреть любые его фрагменты с различной детализацией [22, 23].

1.7. Набор задач

Поддержка решения комплекса всех задач врача в полной мере нигде не реализована. На практике встречаются отдельные полезные сервисы, каждый из которых имеет свои входные данные, платформы, адреса и способы запуска. Врачу приходится инициировать каждый сервис отдельно и самостоятельно сравнивать результаты. Имеются сочетания двух или более задач для некоторой патологии: диагностика с лечением заболевания, диагностика с оценкой риска, прогноз заболевания с оценкой оценка рисков (смерти, осложнений) и др.

Есть ряд сервисов, которые врач вряд ли применит по отношению к пациенту: это вариант предполагаемого состояния в различные моменты времени (в течение 10 лет, через 30 лет...) со «сценарным прогнозированием», позволяющие задавать и менять влияющие условия его физиологических метрик, новых патологий или характеристик образа жизни [15].

1.8. Выводы

В настоящее время имеются различные прогнозные и оценочные модели, методы и платформы их реализации. Большинство решений представлено отдельными сервисами, реализующими одну прогнозную модель; есть примеры, объединяющие несколько моделей. Сервисы предназначены для врачей и пациентов (сервисы для пациентов используют ограниченный набор входных параметров). Сервисы (реализованные на основе знаний), как правило, генерируют объяснения/рекомендации.

Применение методов МО, ММ, инженерии знаний для решения сходных задач, делает актуальным их объединение в единый сервис, который бы позволил интегрировать методы решения задачи различными методами с генерацией объяснений полученных результатов. Интеграция таких методов с ИБ или ЭМК обеспечит возможность в фоновом режиме производить оценивание пациента при появлении новых данных о нём в ИБ или ЭМК. Учитывая, что задачи оценки рисков, прогноза состояния/заболевания пациентов актуальны для большого числа состояний и заболеваний [1-6, 14, 22, 24-26], такая модель не должна зависеть от конкретного заболевания, быть единой для различных состояний и заболеваний, что обеспечит её масштабируемость и позволит развивать функциональность системы.

2 Моделирование рисков развития заболеваний на основе инженерии знаний

Семантическое моделирование с помощью современных методов инженерии знаний позволяет представлять разные типы знаний о заболеваниях. Одни показывают зависимость состояний организма от сочетания наблюдаемых свойств и влияющих факторов (или вариантов развития патологий в зависимости от известных факторов), другие математически выражают зависимости показателей здоровья.

Для интеграции вариантов описания рисков заболеваний и альтернативных способов их определения предлагается следующая модель:

*{<Заболевание/патологическое состояние>_и,
<{риск, [имя/автор методики,] исходные данные, [условия применимости], способ определения}>_{ij}}*

Элементы модели:

Заболевание/патологическое состояние – имя заболевания или патологического состояния, для которого делается риск либо вычисляется прогноз.

Риск = <вид угрозы, вероятность угрозы>,

вид угрозы = (риск появления | риск перехода в тяжёлую стадию | прогноз продолжительности жизни | риск неблагоприятного исхода (смерти));

вероятность угрозы = (Очень высокий риск | Высокий риск | Умеренный риск | Низкий риск | Отсутствие риска).

Исходные данные – предикторы, используя которые вычисляется риск или угроза:

Исходные данные_и = {Наблюдение_{jk}}.

Каждый элемент исходных данных (наблюдение) может быть нетемпоральным и темпоральным. Для нетемпоральных исходных данных описание имеет следующий вид:

Наблюдение = (Фактор | Признак | Событие),

для темпоральных дополнительно описывается временная характеристика:

Наблюдение = <(Фактор | Признак | Событие) [, временная характеристика]>.

К факторам относятся метрики пациента, вредные привычки, имеющиеся заболевания и пр. Событием может быть лечебная процедура, операция, контакт с больным, с веществом и др. Признаки – это лабораторные или инструментальные исследования пациента, жалобы и другие показатели состояния организма пациента, наблюдаемые к моменту определения риска или прогноза.

При описании рисков заболеваний в исходных данных часто достаточно перечислить названия наблюдений, но надёжнее добавлять к ним ожидаемые диапазоны значений. Для

определения риска у конкретного пациента используют сведения о нём: набор исходных данных – названия наблюдений со значениями наблюдений.

Факторы описываются именем и диапазоном значений:

Фактор = <имя/название фактора, диапазон значений>.

Событие описывается своим именем, возможно, имеет уточняющие параметры (с именем и значением):

Событие = <название, {имя параметра, диапазон значений}>.

Значения бывают двух типов – количественные (числовые) с единицами измерения (ед.изм.) и качественные (строковые, категориальные).

Диапазон значений = ({значением категориального типа} | <числовой диапазон, ед.изм.>)

Признак может быть простым или составным. Простой признак определяется своим именем и значением (например, Изжога – имеется или нет), составной признак имеет имя, характеристику и значение (например, Боль в грудной клетке с Характеристиками Локализация, имеющая значение из диапазона: «за грудиной», «в подреберье» и т.п., Периодичность и т.д.).

Простой признак = название признака, диапазон значений.

Составной признак = название признака,

{название характеристики/элемента, диапазон значений}.

Временная характеристика – интервал времени или числовое значение, в течение которого фактор/событие/признак наблюдались или происходили. Например, <операция ЧКВ, длительность = 2 часа>.

Способ определения (величины) задаёт как вычисляется данный прогноз/риск по сведениям о конкретном пациенте:

способ определения = (декларативные знания для распознавания угрозы | вычисляемая величина).

Вычисляемая прогнозная шкала может опираться на формулу с вычислением, математическую модель, модель МО (стохастический градиентный бустинг| нейросеть или др.) или на другой способ получения вероятности риска, возможно явно показывающий вклад каждого предиктора в итоговую величину. Часто вычисленным значениям сопоставляются градации некоторой экспертной шкалы.

2.1 Модель вычисляемых прогнозных шкал

Если способ получения результата – модель МО или формула вычисления риска или прогноза, то в семантической модели задаются исходные данные, полученный результат, а в качестве способа его определения даётся ссылка на модель или формулу, по которой вычисляется результат. Выходные данные могут иметь различную структуру и способ описания. Часто результатом является градация по модели шкалы значений согласно предлагаемой методике.

Вычисляемая величина = (адрес или способ инициации вычислителя вероятности | формула вычисления | способ накопления баллов).

Вероятность угрозы = {градация (по шкале)k, нижняя границаk величины, верхняя границаk величины}.

Часто применение шкалы состоит в подсчёте количества баллов (нескольких признаков). При этом необходимо знать определения используемых в шкале признаков. Детальное изучение прогностической ценности каждого фактора, включённого в шкалу, может помочь в принятии оптимального клинического решения.

Для определения риска с использованием калькулятора оценки риска осложнений задаются веса (баллы) для отдельных наблюдений, участвующих в накоплении баллов:

способ накопления баллов = {баллы, {(соответствующее баллу) наблюдениеij}}.

Для установления вероятности угрозы *градациям по шкалам* ставятся *нижняя граница* величины и *верхняя граница* величины, которыми являются минимальный суммарный балл и максимальный суммарный балл.

С помощью математических методов, например, регрессионного анализа, производится выделение наиболее значимых для прогноза факторов и строится шкала, адаптированная для практики. Предполагается, что готов вычисляющий сервис (агент) и известны правила передачи ему параметров (наблюдений пациента).

Адрес или способ инициации вычислителя вероятности = ссылка на вычисляющего программного агента, последовательность параметров со значениями.

Формула для вычисления может быть задана явно. Определены структура и правила внесения формул для вычисления по набору исходных данных, наблюдаемых у пациента. Используется конструктор формулы через термины-переменные, константы и арифметические термины (сумма, произведение, частное, возведение в степень, логарифм и пр.). Модель связывает наблюдения пациента с параметрами запроса на вычисление по сконструированной формуле.

2.2 Модель заболеваний

Модель явно описанных знаний для распознавания имеет вид

декларативные знания для распознавания угрозы =
{условие на наблюдение jpr } $n=1, N$ [, минимальное количество $jr < N$]}>.

(Примечание. Некоторые знания не уточняют вероятность угрозы, а указывают её наличие («угроза j имеется»). «Угроза отсутствует» обычно специально не описывается. Примером «Угроза имеется» является («Риск сердечно-сосудистой смертности» + «имеется»).

Условие на наблюдение = название наблюдения,

(значение категориального\перечисляемого типа | <число, ед.изм. | Тренд >

| {название характеристики, (значение категориального типа | <число, ед.изм.> | Тренд)})

[, условие на время измерения].

В некоторых случаях знания об угрозе охватывают временные отношения между наблюдаемыми состояниями в периодах развития разной длительности и эффекты лечебных воздействий:

<угроза j имеется – {{Фактор i ,}} {(Период i , Интервал длительности i),

{Условие на наблюдение jk } $k=1, N$ [, минимальное количество $j < N$]}.

На основе представленной модели на платформе *IACPaaS* [27-29] разработана онтология (см. рисунок 2).

3 Разработка БЗ

Семантические модели, реализованные на платформе *IACPaaS*, позволяют формировать разные типы знаний в любых областях медицины и интегрировать их в процессе поддержки решения важных задач прогноза, оценки состояния и рисков неблагоприятных процессов.

В качестве примера приведено описание БЗ по оценке рисков болезней системы кровообращения (см. рисунок 3). Наиболее востребованы БЗ оценки риска развития ССЗ (см. рисунок 4). Если способом вычисления результата является формула, то в БЗ создаётся модель вычисления, либо задаётся ссылка на агент, производящий вычисления (см. рисунок 5).

Заключение

Определение рисков заболеваний, прогноз возможных осложнений, смерти, исхода лечения являются неотъемлемыми составляющими практикующего врача. Для помощи ему в

повседневной деятельности предложен набор моделей и методов, основанных как на достаточно простых вычислениях (калькуляторах), так и на более сложных методах ММ, МО и искусственного интеллекта.

По ряду заболеваний (например, ССЗ) существует большое количество моделей, что затрудняет их выбор и использование, поскольку большинство решений представлено отдельными сервисами.

Для объединения различных подходов к решению практических задач, аккумулирования знаний в едином информационном пространстве предложена семантическая модель для оценки влияния комплекса факторов на развитие заболеваний и оценки персонафицированных рисков. Особенностью модели является независимость от конкретного заболевания либо группы заболеваний, что позволяет использовать её как для конкретного заболевания, так и для произвольной группы заболеваний. Данная модель реализована на платформе *IACPaas*. На платформе реализован решатель, интегрированный с ИБ и ЭМК, позволяющий на основе БЗ и анализа всех разделов медицинских документов (ИБ, ЭМК) генерировать объяснение.

Возможность объединения на основе онтологии моделей прогноза, оценки рисков и др. по различным заболеваниям позволяет комплексно решать важные задачи практической медицины. При этом генерируется объяснение в терминах, понятных врачу с указанием источника, по которому применённые знания формировались.

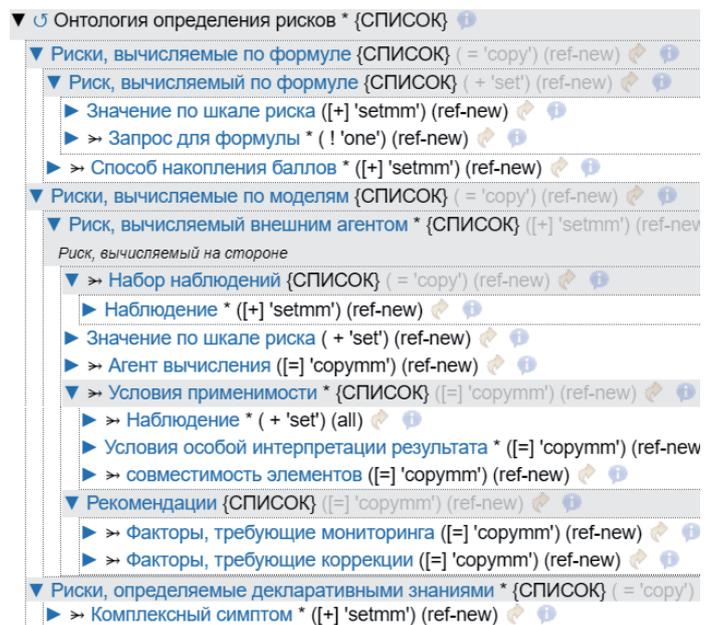


Рисунок 2 - Онтология знаний о связи рисков заболеваний с их показателями [30]

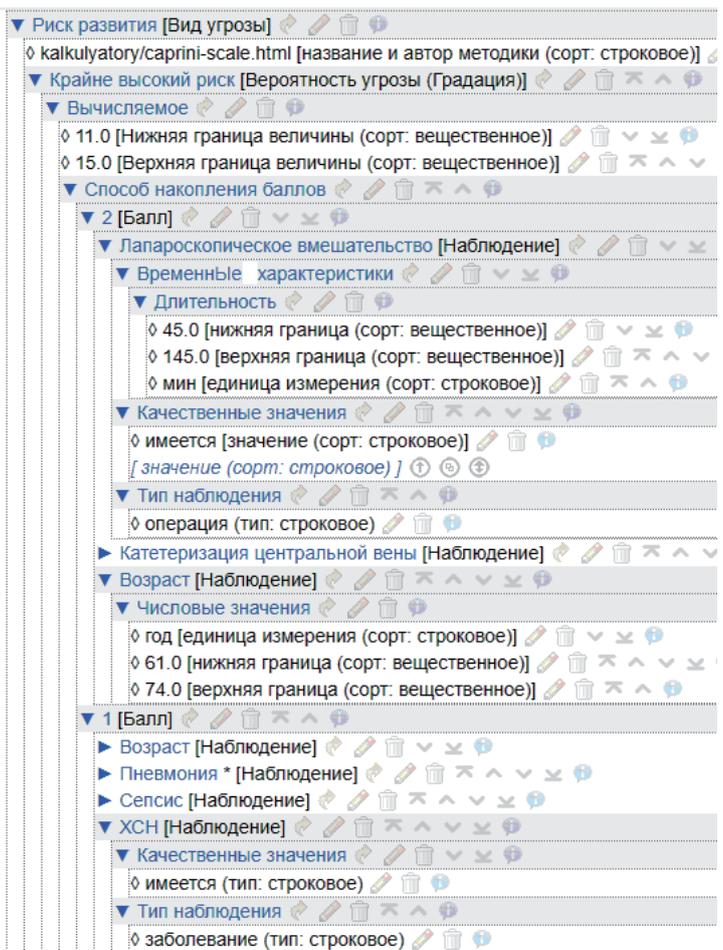


Рисунок 3 - Скриншот фрагмента БЗ на платформе *IACPaas* по оценке рисков венозного тромбоза

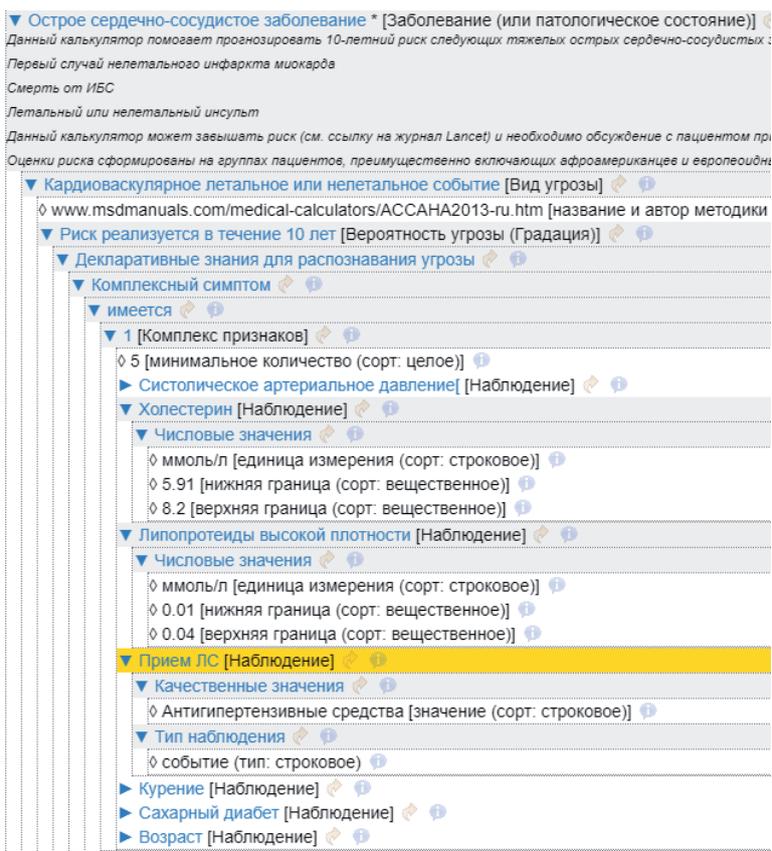


Рисунок 4 - Скриншот фрагмента БЗ на платформе IACPaaS по оценке рисков ССЗ

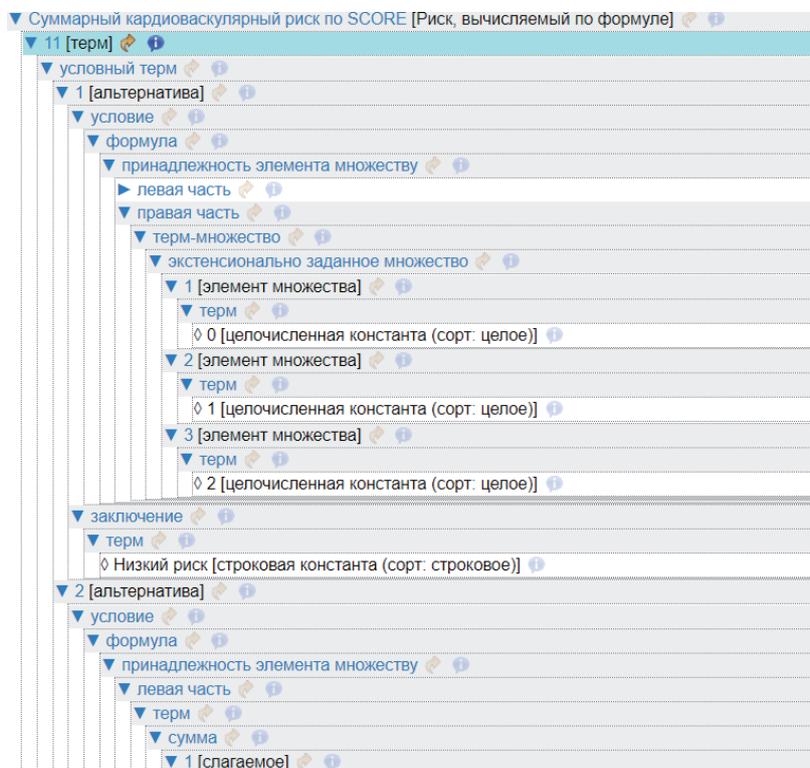


Рисунок 5 - Скриншот фрагмента БЗ на платформе IACPaaS по вычислению кардиоваскулярного риска по формуле

Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 18-29-03131 и 19-29-01077.

Список источников

- [1] **Шахгельдян, К.И.** Оценка предиктивного потенциала дооперационных факторов риска фибрилляции предсердий у больных ишемической болезнью сердца после коронарного шунтирования / К.И. Шахгельдян, В.Ю. Рублев, Б.И. Гельцер, Б.О. Щеглов, В.Г. Ширококов, М.К. Духтаева, К.В. Чернышева // Сибирский журнал клинической и экспериментальной медицины. 2020. Т.35. № 4. С.128-136. DOI:10.29001/2073-8552-2020-35-4-128-136.
- [2] **Бродская, Т.А.** Оценка предиктивного потенциала факторов сердечно-сосудистого риска и их ассоциаций с жесткостью артерий у лиц европейской и корейской этнических групп / Т.А. Бродская, В.А. Невзорова, К.И. Шахгельдян, Б.И. Гельцер, Д.А. Вражнов, Ю.В. Кистенев // Российский кардиологический журнал. 2021; 26(5): 4230. DOI:10.15829/1560-4071-2021-4230.
- [3] **Гельцер, Б.И.** Методы машинного обучения в прогнозировании летальных исходов в стационаре у больных ишемической болезнью сердца после коронарного шунтирования / Б.И. Гельцер, К.И. Шахгельдян, В.Ю. Рублев, В.Н. Котельников, А.Б. Кригер, В.Г. Ширококов // Кардиология. 2020, 10(60): с.38-46. DOI:10.18087/cardio.2020.10.n1170.
- [4] **Мун, С.А.** Регрессионный анализ в медико-биологических исследованиях / С.А. Мун, А.Н. Глушов, Т.А. Штернис, С.А. Ларин, С.А. Максимов // Кемерово: Кемеровская государственная медицинская академия, 2012. 119 с.
- [5] **Гельцер, Б.И.** Результаты моделирования должных величин силы дыхательных мышц на основе метода искусственных нейронных сетей / Б.И. Гельцер, В.Н. Котельников, К.И. Шахгельдян, И.Г. Курпатов // Российский физиологический журнал им. И.М.Сеченова. 2018; 104 (9): 1065–1074. DOI:10.7868/S0869813918090058.
- [6] **Акопян, В.С.** Комплексная модель оценки риска развития и прогрессирования возрастной макулярной дегенерации / В.С. Акопян, Н.С. Семенова, Е.А. Новикова, А.Г. Тоницкий, И.Н. Нечаев // Офтальмология, 2012, Т. 9, № 3. DOI:10.18008/1816-5095-2012-3-52-56.
- [7] **Iniesta, R.** Combining clinical variables to optimize prediction of antidepressant treatment outcomes / R. Iniesta, K. Malki, W. Maier, M. Rietschel, O. Mors, J. Hauser, et al. // *J Psychiatr Res.* (2016) 78: 94–102. DOI:10.1016/j.jpsychires.2016.03.016.
- [8] **Medeiros, F.A.** Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models / F.A. Medeiros, M.T. Leite, L.M. Zangwill, R.N. Weinreb // *Investig Ophthalmol Vis Sci.* 2011; 52: 5794-5803.
- [9] Symptomate: Diagnostic technology is powered by Infermedica. – <https://symptomate.com/diagnosis/#0-66>.
- [10] WebMD: Logo for WebMD Symptom Checker WITH BODY MAP. – <https://symptoms.webmd.com/default.htm>.
- [11] Isabel: The Symptom Checker doctors use and trust. <https://symptomchecker.isabelhealthcare.com/>.
- [12] Webiomed. Демонстрационное приложение для оценки сердечно-сосудистого риска: Оценка сердечно-сосудистого риска. <https://webiomed.ai/dhra/>.
- [13] **Kwon, J.M.** Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction / J.M. Kwon, K.H. Jeon, H.M. Kim, et al. *PLoS One.* 2019;14(10):e0224502. DOI:10.1371/journal.pone.0224502.
- [14] Калькулятор для оценки риска ампутации конечности и пользы от реваскуляризации: Экспертные решения для предотвращения ампутаций нижних конечностей. – <https://kink.ru/specialistam/kalkulyator/>.
- [15] **Ясницкий, Л.Н.** Нейроэкспертная система диагностики, прогнозирования и управления рисками сердечно-сосудистых заболеваний / Л.Н. Ясницкий, Ф.М. Черепанов // Прикладная математика и вопросы управления. 2018. № 3. С.107-126. DOI: 10.15593/2499-9873/2018.3.08.
- [16] **Выучейская, М.В.** и др. Нейросетевые технологии в диагностике заболеваний (обзор) / М.В. Выучейская и др. // Журн. мед.-биол. исследований. 2018; 3 (6): 284–94. DOI: 10.17238/2542-1298.2018.6.3.284.
- [17] **Manyam, R.B.** Deep Learning Approach for Predicting 30 Day Readmissions after Coronary Artery Bypass Graft Surgery / R.B. Manyam, Y Zhang, W Keeling, et al.. *NeurIPS (ML4H)*. 2018; 222 ArXiv:abs/1812.00596.
- [18] **Chekroud, A.M.** Cross-trial prediction of treatment outcome in depression: a machine learning approach / A.M. Chekroud, R.J. Zotti, Z. Shehzad, R. Gueorguieva, M.K. Johnson, M.H. Trivedi, et al.. *Lancet Psychiatry* (2016) 3: 243–50. DOI: 10.1016/S2215-0366(15)00471-X.

- [19] **Kim, Y.J.** Deep learning-based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining / Y.J. Kim, M. Saqlian, J.Y. Lee. *Personal and Ubiquitous Computing*. 2019; 23(4):485-94. DOI: 10.1007/s00779-019-01248-7.
- [20] **Hill, N.R.** Predicting atrial fibrillation in primary care using machine learning / N.R. Hill, D. Ayoubkhani, P. McEwan, et al. *PLoS One*. 2019; 14(11): e0224582. DOI:10.1371/journal.pone.0224582.
- [21] Карта симптомов: @Здоровье. <https://health.mail.ru/symptoms/>.
- [22] **Грибова, В.В.** Облачный сервис поддержки принятия решений в кардиологии на основе формализованных знаний / В.В. Грибова, М.В. Петряева, Е.А. Шалфеева // *Сибирский журнал клинической и экспериментальной медицины*. 2020. №4. С.32-38. DOI: 10.29001/2073-8552-2020-35-4-32-38.
- [23] **Грибова, В.В.** Облачная среда для поддержки клинической медицины и образования / В.В. Грибова, Ф.М. Москаленко, Д.Б. Окунь, М.В. Петряева // *Врач и информационные технологии*. 2016. №1. С.60-66.
- [24] **Kolek, M.J.** Genetic and clinical risk prediction model for postoperative atrial fibrillation / M.J. Kolek, J.D. Muehlschlegel, W.S. Bush, et al. *Circ Arrhythm Electrophysiol*. 2015; 8(1): 25-31. - DOI: 10.1161/CIRCEP.114.002300.
- [25] **Inohara, T.** Risk stratification model for in-hospital death in patients undergoing percutaneous coronary intervention: a nationwide retrospective cohort study in Japan / T. Inohara, S. Kohsaka, K. Yamaji, et al. *BMJ Open*. 2019;9(5):e026683. DOI: 10.1136/bmjopen-2018-026683.
- [26] **Shahian, M.D.** STS 2018 adult cardiac risk models part 1 / M.D. Shahian, P. Jeffrey, B. Vinay, et al. *Ann Thorac Surg*. 2018; 105: 1411-8. DOI: 10.1016/j.athoracsur.2018.03.002.
- [27] **Грибова, В.** The methods and the IACPaaS Platform tools for semantic representation of knowledge and development of declarative components for intelligent systems / V. Gribova, A. Kleshev, Ph. Moskalenko, V. Timchenko, L. Fedorischev, E. Shalfeeva // *Open Semantic Technologies for Intelligent Systems*. 2019. 3. P.21-24.
- [28] **Грибова, В.В.** Облачная платформа IACPaaS для разработки оболочек интеллектуальных сервисов: состояние и перспективы развития / В.В. Грибова, А.С. Клешев, Ф.М. Москаленко, В.А. Тимченко, Л.А. Федорищев, Е.А. Шалфеева // *Программные продукты и системы*. 2018. Т. 31. № 3. С.527-536. DOI: 10.15827/0236-235X.031.3.527-536.
- [29] **Грибова, В.** The IACPaaS Cloud Platform: Features and Perspectives / V. Gribova, A. Kleshev, P. Moskalenko, V. Timchenko, L. Fedorischev, E. Shalfeeva. // *IEEE. Xplore*. 2017. P. 80-84. DOI: 10.1109/RPC.2017.8168073.
- [30] IACPcloud: Онтология определения рисков. - <https://owncloud.dvo.ru/s/3YnamqX98eiQo4p>.

Сведения об авторах



Грибова Валерия Викторовна, 1965 г. рождения. Окончила Ленинградский политехнический институт (1989), д.т.н. (2007). Зав. лабораторией интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН, заместитель директора по научной работе, вице-президент Российской ассоциации искусственного интеллекта. В списке научных трудов более 250 работ в области искусственного интеллекта, проблемно-ориентированных систем, основанных на знаниях, специализированных программных моделей и систем. AuthorID (РИНЦ): 7400; Author ID (Scopus): 7801667631; ORCID: 0000-0001-9393-351X; Researcher ID (WoS): Q-4250- 2016. gribova@iacp.dvo.ru.

Окунь Дмитрий Борисович, 1973 г. рождения. Окончил Владивостокский государственный медицинский университет (1996), к.м.н. (2000). Научный сотрудник лаборатории Интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН. AuthorID (РИНЦ): 642886; Author ID (Scopus): 57204598165; ORCID: 0000-0002-6300-846X. Researcher ID (WoS): AAV-1824-2020. okdm@dvo.ru.



Шалфеева Елена Артефьевна, 1967 г. рождения. Окончила Дальневосточный государственный университет (1989), к.т.н. (2000). Старший научный сотрудник лаборатории Интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН, доцент. В списке научных трудов более 80 работ. AuthorID (РИНЦ): 11682; Author ID (Scopus): 6508163590 ORCID: 0000-0001-5536-2875. Researcher ID (WoS): L-9623-2015. shalfe@iacp.dvo.ru.

Поступила в редакцию 23.11.2021, после рецензирования 2.12.2021. Принята к публикации 13.12.2021.

Semantic models for assessing the influence of a combination of factors on the development of diseases

V.V. Gribova, D.B. Okun, E.A. Shalfeeva

Abstract

The analysis of approaches and solutions to the problem of risk assessment and prognosis of conditions and development of diseases is presented. It is shown that the implementation of software services on various platforms complicates the possibility of their comprehensive use and the choice between the available solutions. This has risen the urgency of creating a unified semantic model of diseases that integrates various methods and approaches to solving this problem and accumulates knowledge about risks and prognosis in a unified information space. A new semantic model is proposed to take into account influence of a combination of factors on development of various events that threaten health and life. The feature of the model is its independence from a specific disease or a group of diseases, which allows it to be used in various branches of medicine. This model has been tested on the IACPaaS platform. A software solver has been implemented that allows generating a clear explanation based on the knowledge base and analysis of the patient's electronic medical record. The application of the new model for the formation of knowledge is shown on the example of risk assessment and prognosis of cardiovascular events.

Key words: *semantic model, disease risk assessment, knowledge base, medical intelligent systems, decision support system.*

Citation: *Gribova VV., Okun DB., Shalfeeva EA. Semantic models for assessing the influence of a combination of factors on the development of diseases [In Russian]. *Ontology of designing*. 2021; 11(4): 464-477. DOI: 10.18287/2223-9537-2021-11-4-464-477.*

Acknowledgment: This work was carried out with partial financial support from the Russian Foundation for Basic Research, projects 18-29-03131 and 19-29-01077.

List of figures

Figure 1 - An example of the risk assessing result of cardiovascular diseases using various methods

Figure 2 - Ontology of knowledge about the relationship of disease risks with their indicators [30]

Figure 3 - Screenshot of a knowledge base fragment on the IACPaaS platform for assessing the risks of venous thrombosis

Figure 4 - Screenshot of a knowledge base fragment on the IACPaaS platform for risk assessment of cardiovascular diseases

Figure 5 - Screenshot of a knowledge base fragment on the IACPaaS platform for calculating cardiovascular risk using the formula

References

- [1] **Shakhgeldyan KI, Rublev VYu, Geltser BI, Shcheglov BO, Shirobokov VG, Dukhtaeva MK, Chernysheva KV.** Predictive potential assessment of preoperative risk factors for atrial fibrillation in patients with coronary artery disease after coronary artery bypass grafting [In Russian]. *The Siberian Journal of Clinical and Experimental Medicine*. 2020; 35(4):128-136. DOI: 10.29001/2073-8552-2020-35-4-128-136.
- [2] **Brodskaya TA, Nevzorova VA, Shakhgeldyan KI, Geltser BI, Vrazhnov DA, Kistenev YuV.** Predictive potential assessment of cardiovascular risk factors and their associations with arterial stiffness in people of European and Korean ethnic groups [In Russian]. *Russian Journal of Cardiology*. 2021; 26(5):4230. DOI: 10.15829/1560-4071-2021-4230.
- [3] **Geltser BI, Shahgeldyan KJ, Rublev VYu, Kotelnikov VN, Krieger AB, Shirobokov VG.** Machine Learning Methods for Prediction of Hospital Mortality in Patients with Coronary Heart Disease after Coronary Artery Bypass Grafting [In Russian]. *Kardiologiya*. 2020;60(10):38-46. DOI: 10.18087/cardio.2020.10.n1170.
- [4] **Mun SA, Glushov AN, Shternis TA, Larin SA, Maksimov SA.** Regression analysis in biomedical research [In Russian]. Kemerovo: KemGMA. 2012. 119 p.

- [5] **Geltcer BI, Kotelnikov VN, Shahgeldyan KI, Kurpatov IG.** Results of modeling the proper values of the strength of the respiratory muscles based on the method of artificial neural networks [In Russian]. *Russian physiological journal. I.M.Sechenov.* 2018; 104 (9): 1065–1074. DOI: 10.7868/S0869813918090058.
- [6] **Akopyan VS, Semenova NS, Novikova EA, Tonivitskiy AG, Nechayev IN.** The complex model of risk and progression of AMD estimation [In Russian]. *Ophthalmology in Russia.* 2012; 9(3):52-56. DOI: 10.18008/1816-5095-2012-3-52-56.
- [7] **Iniesta R, Malki K, Maier W, Rietschel M, Mors O, Hauser J, et al.** Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J. Psychiatr Res.* 2016; 78: 94–102. DOI: 10.1016/j.jpsychires.2016.03.016.
- [8] **Medeiros FA, Leite MT, Zangwill LM, Weinreb RN.** Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Investig Ophthalmol Vis Sci.* 2011; 52: 5794-5803.
- [9] Symptomat: Diagnostic technology is powered by Infermedica. <https://symptomate.com/diagnosis/#0-66>.
- [10] WebMD: Logo for WebMD Symptom Checker WITH BODY MAP. <https://symptoms.webmd.com/default.htm>.
- [11] Isabel: The Symptom Checker doctors use and trust. <https://symptomchecker.isabelhealthcare.com/>.
- [12] Webiomed. Demonstration App for Cardiovascular Risk Assessment: Cardiovascular risk assessment. <https://webiomed.ai/dhra/>.
- [13] **Kwon JM, Jeon KH, Kim HM, et al.** Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *PLoS One.* 2019; 14(10): e0224502. DOI: 10.1371/journal.pone.0224502.
- [14] Calculator for assessing the risk of limb amputation and the benefits of revascularization: Expert solutions for the prevention of lower limb amputation [In Russian]. <https://kink.ru/specialistam/kalkulyator/>.
- [15] **Yasnitskiy LN, Cherepanov FM.** Neuroexpert system for diagnostics, prediction and risk management of cardiovascular diseases [In Russian]. *Applied Mathematics and Management Issues.* 2018. № 3. S. 107-126. (In Russ.). DOI: 10.15593/2499-9873/2018.3.08.
- [16] **Vyucheyskaya MV, Kraynova IN, Gribanov AV.** Neural Network Technologies in Medical Diagnosis (Review) [In Russian]. *Journal of Medical and Biological Research,* 2018; 6(3): 284-294. DOI: 10.17238/issn2542-1298.2018.6.3.284.
- [17] **Manyam RB, Zhang Y, Keeling W, et al.** Deep Learning Approach for Predicting 30 Day Readmissions after Coronary Artery Bypass Graft Surgery. *NeurIPS (ML4H).* 2018; 222 ArXiv:abs/1812.00596.
- [18] **Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al.** Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* (2016) 3:243–50. DOI: 10.1016/S2215-0366(15)00471-X.
- [19] **Kim YJ, Saqlian M, Lee JY.** Deep learning-based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining. *Personal and Ubiquitous Computing.* 2019; 23(4): 485-94. DOI: 10.1007/s00779-019-01248-7.
- [20] **Hill NR, Ayoubkhani D, McEwan P, et al.** Predicting atrial fibrillation in primary care using machine learning. *PLoS One.* 2019;14(11):e0224582. DOI: 10.1371/journal.pone.0224582.
- [21] Symptom Map: @Zdorov'e. <https://health.mail.ru/symptoms/>.
- [22] **Gribova VV, Petryaeva MV, Shalfeeva EA.** Cloud decision support service in cardiology based on formalized knowledge [In Russian]. *The Siberian Journal of Clinical and Experimental Medicine.* 2020; 35(4): 32-38. DOI: 10.29001/2073-8552-2020-35-4-32-38.
- [23] **Gribova VV, Moskalenko FM, Okun DB, Petryaeva MV.** Cloud environment to support clinical medicine and education [In Russian]. *Physician and Information Technology.* 2016; 1: 60-66.
- [24] **Kolek MJ, Muehlschlegel JD, Bush WS, et al.** Genetic and clinical risk prediction model for postoperative atrial fibrillation. *Circ Arrhythm Electrophysiol.* 2015; 8(1): 25-31. DOI: 10.1161/CIRCEP.114.002300.
- [25] **Inohara T, Kohsaka S, Yamaji K, et al.** Risk stratification model for in-hospital death in patients undergoing percutaneous coronary intervention: a nationwide retrospective cohort study in Japan. *BMJ Open.* 2019; 9(5): e026683. DOI: 10.1136/bmjopen-2018-026683.
- [26] **Shahian MD, Jeffrey P, Vinay B, et al.** STS 2018 ADULT CARDIAC RISK MODELS PART 1. *Ann Thorac Surg.* 2018; 105: 1411-8. DOI: 10.1016/j.athoracsur.2018.03.002.
- [27] **Gribova V, Kleshev A, Moskalenko Ph, Timchenko V, Fedorischev L, Shalfeeva E.** The methods and the IAC-PaaS Platform tools for semantic representation of knowledge and development of declarative components for intelligent systems. *Open Semantic Technologies for Intelligent Systems.* 2019; 3: 21-24.
- [28] **Gribova VV, Kleshchev AS, Moskalenko FM, Timchenko VA, Fedorishchev LA, Shalfeeva EA.** IACPaaS cloud platform for the development of intelligent services shells: state and development prospects [In Russian]. *Software products and systems.* 2018; 31(3): 527-536. DOI: 10.15827/0236-235X.031.3.527-536.
- [29] **Gribova V, Kleshev A, Moskalenko P, Timchenko V, Fedorischev L, Shalfeeva E.** The IACPaaS Cloud Platform: Features and Perspectives. *IEEE. Xplore.* 2017. P. 80-84. DOI: 10.1109/RPC.2017.8168073.

[30] IACPcloud: Ontology of risk assessment. <https://owncloud.dvo.ru/s/3YnamqX98eiQo4p>.

About the authors

Valeriya Victorovna Gribova (b.1965) graduated from the Leningrad Polytechnic University in 1989, Professor's degree (2007). She is a Head of the Intelligent systems lab in the Institute for Automation & Control Processes of the FEB RAS, Research Deputy Director, an expert of Analytic Center in Government of Russian Federation, a Vice-President of Russian Association of Artificial Intelligence. She is a co-author of more than 250 publications in the fields of AI, informatics, program models, technologies and systems. AuthorID (RSCI): 7400; Author ID (Scopus): 7801667631; ORCID: 0000-0001-9393-351X; Researcher ID (WoS): Q-4250-2016. gribova@iacp.dvo.ru.

Dmitry Borisovich Okun (b.1973) graduated from the Vladivostok State Medical University in 1996, PhD in Medical sciences (2000). Research associate of the Intelligent systems lab at the Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences. AuthorID (RSCI): 642886; Author ID (Scopus): 57204598165; ORCID: 0000-0002-6300-846X. Researcher ID (WoS): AAV-1824-2020. okdm@dvo.ru

Elena Arefjevna Shalfeeva (b.1967) graduated from the Far Eastern State University (Vladivostok) in 1989, PhD (2000). She is a Senior Researcher at the Intelligent systems lab in the Institute for Automation & Control Processes of the FEB RAS, an assistant professor. She is a co-author of more than 80 publications in the fields of Program models and systems and AI. AuthorID (RSCI): 11682; Author ID (Scopus): 6508163590 ORCID: 0000-0001-5536-2875. Researcher ID (WoS): L-9623-2015. shalff@iacp.dvo.ru.

Received November 23, 2021. Revised December 02, 2021. Accepted December 13, 2021.
