

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УДК 004.652; 004.853; 004.891

КУРОЧКИН

Александр Васильевич

**МОДЕЛЬ, МЕТОДЫ И АЛГОРИТМЫ ПОСТРОЕНИЯ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ ЭКСПЕРТНЫХ СИСТЕМ
УЧЕТА И АНАЛИЗА ДАННЫХ МЕДИЦИНСКИХ ОБСЛЕДОВАНИЙ**

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

по специальности

05.13.01 – системный анализ, управление и обработка
информации

Минск, 2022

Научная работа выполнена в **Белорусском государственном университете**.

НАУЧНЫЙ РУКОВОДИТЕЛЬ –

Садов Василий Сергеевич,
кандидат технических наук, доцент,
профессор кафедры интеллектуальных систем
Белорусского государственного университета.

ОФИЦИАЛЬНЫЕ ОППОНЕНТЫ:

Липницкий Станислав Феликсович,
доктор технических наук, доцент,
главный научный сотрудник отдела совместных
программ космических и информационных
технологий ГНУ «Объединенный институт
проблем информатики НАН Беларуси»;

Глухов Дмитрий Олегович,
кандидат технических наук, доцент,
доцент кафедры вычислительных систем и сетей
УО «Полоцкий государственный университет».

ОППОНИРУЮЩАЯ ОРГАНИЗАЦИЯ –

**УО «Белорусский государственный
университет информатики
и радиоэлектроники».**

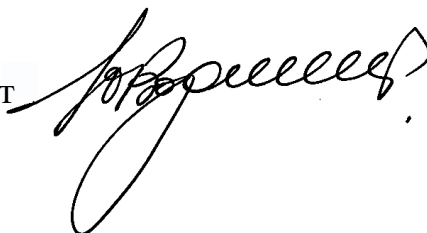
Защита состоится **25 февраля 2022 года** в **14.00** часов на заседании совета по защите диссертаций Д 02.01.14 при Белорусском государственном университете по адресу: *Минск, ул. Ленинградская 8 (корпус юридического факультета), ауд. 407*. Телефон ученого секретаря 209-57-09.

Почтовый адрес: пр-т Независимости 4, Минск, 220030.

С диссертацией можно ознакомиться в Фундаментальной библиотеке Белорусского государственного университета.

Автореферат разослан «18» января 2022 года.

Ученый секретарь совета
по защите диссертаций
кандидат физ.-мат. наук, доцент



Ю.И. Воротницкий

ВВЕДЕНИЕ

Широкое развитие и внедрение информационных технологий в разнообразные слои общества за последние годы дало существенный толчок к формированию принципиально новых подходов к работе с данными в различных областях. Значительную важность приобретают вопросы разработки интеллектуальных информационных систем, на основе которых могут быть реализованы комплексные системы поддержки принятия решений и управления знаниями. Стратегически важным направлением развития и внедрения решений на базе информационных технологий является медицина. В Республике Беларусь за последнее десятилетие уделялось и продолжает уделяться особое внимание вопросам расширения использования решений на базе информационных технологий в этой сфере, что подтверждается утверждением и реализацией многочисленных государственных программ различного масштаба.

В рамках данной работы исследуется синтез, интеграция и анализ различных видов хранилищ, экспертных систем и систем поддержки принятия решений в медицине. В работе рассматриваются основные аспекты разработки самих решающих модулей для различных видов задач в медицине, в том числе на основе строгой формализации экспертных знаний и на основе вывода зависимостей в имеющихся данных. Кроме этого, исследуются особенности использования разработанных решающих модулей непосредственно в работе медицинских специалистов с протоколами обследований и другими видами информации о пациенте. Также рассматривается задача интерпретации самого процесса принятия решений решающих модулей и оценки его корректности медицинскими специалистами.

В работе предложены модель, методы и алгоритмы для организации информационной поддержки и расширения возможностей по анализу данных в медицине. На их основе разработана технология построения специализированных медицинских экспертных систем, которая может использоваться для быстрого прототипирования и интеграции с другими медицинскими информационными системами.

С помощью предложенной технологии разработаны три специализированные информационные системы, относящиеся к различным областям медицины – пренатальной диагностике, офтальмологии и кардиологии. Предложенные методы и алгоритмы позволили не только организовать поддержку принятия решений в этих системах, но и осуществить анализ имеющихся данных и построить интерпретации процесса принятия решений, что позволило оценить корректность таких систем и выявить новые значимые результаты с точки зрения медицины.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с научными программами (проектами), темами

Тема диссертационной работы соответствует приоритетному направлению научных исследований Республики Беларусь на 2016–2020 годы «Информатика и космические исследования» (постановление Совета Министров Республики Беларусь от 12.03.2015 №190), научно-технической деятельности Республики Беларусь на 2016–2020 годы «Информационно-коммуникационные и авиакосмические технологии: биоэлектроника, биоинформатика и информационные технологии в медицине» (Указ Президента Республики Беларусь от 22.04.2015 №166) и государственной программе на 2016-2020 годы «Здоровье народа и демографическая безопасность Республики Беларусь» (постановление Совета Министров Республики Беларусь от 14.03.2016 № 200).

Диссертационная работа выполнялась в рамках:

- НИР №810/17 «Методы и алгоритмы комплексного анализа речевой и видео информации в коммуникационных процессах с целью поддержки принятия экспертных решений» на 2014–2018 годы,
- НИР №427/17 «Методы и алгоритмы оценки эмоционального состояния человека по динамике мимики его лица на видеопоследовательностях» государственной программы научных исследований «Информатика, космос и безопасность» на 2019-2020 годы,
- гранта БГУ «Трехмерная реконструкция результатов видеоэндоскопических медицинских обследований» на 2018 год,
- гранта БГУ «Экспертная медицинская система поддержки принятия решений на основе статистической обработки больших объемов данных» на 2019 год.

Цель и задачи исследования

Объектом исследования диссертационной работы являются медицинские информационные системы. Предметом исследования являются принципы построения медицинских экспертных систем, алгоритмы принятия решений и интерпретации полученных результатов.

Целью диссертационной работы является разработка унифицированной технологии построения специализированных медицинских систем для организации информационной поддержки диагностики и расширения возможностей по анализу данных в предметной области.

Для достижения цели, поставленной в диссертационной работе, необходимо решить следующие задачи:

- создать информационную модель обобщенной экспертной медицинской

- системы;
- разработать метамодель описания структуры данных и организации взаимодействия основных компонентов информационной системы;
 - разработать алгоритм генерации слоя доступа к данным и пользовательского интерфейса по метамодели;
 - разработать алгоритм агрегации данных, представленных гетерогенными источниками;
 - разработать алгоритмы поддержки принятия решений, учитывающие экспертные знания и исторические данные;
 - разработать методы анализа корректности алгоритмов поддержки принятия решений и объяснения их работы;
 - реализовать разработанные алгоритмы программно в рамках медицинских экспертных систем специального назначения;
 - исследовать корректность полученных результатов.

Научная новизна

1. Разработана информационная модель обобщенной экспертной медицинской системы, которая позволяет осуществить синтез основных структурных элементов специализированных медицинских информационных систем в конкретной области по описательному заданию в виде метамодели, а также интегрировать полученную систему с существующими источниками данных и другими системами.

2. Предложен метод семантической интерпретации процессов принятия решения в обученных моделях машинного обучения с учителем на основе локальных линейных аппроксимаций целевой функции модели, что позволяет экспертам оценивать корректность таких моделей и использовать полученные интерпретации для генерации новых экспертных знаний.

3. Разработан алгоритм оптимизации процесса принятия решений на базе правил с использованием накопленных данных по пациентам, позволяющий одновременно учитывать экспертные знания и скрытые статистические зависимости этих данных при проектировании экспертных систем, что повышает их корректность и позволяет проверять выполнение существующих правил на произвольных наборах данных.

Положения, выносимые на защиту

1. Информационная модель обобщенной экспертной медицинской системы, позволяющая создать программную платформу для построения специализированных решений, отличающаяся организацией взаимодействия основных компонентов с помощью метамодели, которая позволяет

унифицировать процесс доступа к данным, организовать пользовательское взаимодействие с системой, а также осуществлять семантическую интерпретацию и оценку корректности принятых решений.

2. Методы интерпретации процесса принятия решений в обученных моделях машинного обучения с учителем на основе оценки значимости входных параметров и на основе локальных линейных аппроксимаций решающей функции модели, позволяющие устанавливать закономерности в процессе принятия решений, оценивать их корректность и использовать полученные результаты для генерации новых экспертных знаний.

3. Алгоритм оптимизации экспертных систем на базе правил с использованием накопленных данных по пациентам, основанный на формировании весов правил по скрытым статистическим зависимостям выборки, который позволяет корректировать процесс принятия решений с поступлением новых данных, и повышает F1-меру для бинарной классификации по данным исследований на 0,13 – 0,17 (95%-й доверительный интервал). Полученные результаты также могут использоваться для оценки корректности правил.

4. Информационно-аналитические экспертные системы учета и анализа данных медицинских обследований в различных областях медицины, основанные на предложенной информационной модели обобщенной экспертной медицинской системы, позволяющие выделить наиболее значимые показатели обследований, осуществить постановку диагноза с оценкой достоверности по показателю F1-меры 0,86 – 0,95, а также интерпретировать процесс принятия решения при помощи предложенных методов и алгоритмов, что расширяет диагностический и исследовательский потенциал при принятии решений в медицине.

Личный вклад соискателя ученой степени

Представленные в диссертационной работе результаты получены лично соискателем. Научный руководитель Садов В.С. участвовал в постановке задач и выборе методов исследования. Соавторы Прибушения О.В., Качан Т.В., Сенько К.В. оказывали консультацию по смежным вопросам медицинской области, внедрению в медицинскую практику. Научный руководитель и соавторы совместных работ участвовали в обсуждении полученных результатов.

Апробация диссертации и информация об использовании результатов

Основные результаты диссертационной работы докладывались на 12 международных научно-технических и научно-практических конференциях и семинарах: 3-я, 5-я Международная конференция «Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния»

(Минск, 28-29 апреля 2015 г., 16-17 мая 2019 г.); Международная научно-практическая конференция «Инновационные решения проблем экономики знаний Беларуси и Казахстана» (Минск, 13 октября 2016 г.); XV Белорусско-российская научно-техническая конференция «Технические средства защиты информации» (Минск, 6 июня 2017 г.); Международная научная конференция «Информационные технологии и системы» ИТС2017, ИТС2019, ИТС2020 (Минск, 25 октября 2017 г., 30 октября 2019 г., 18 ноября 2020 г.); International Conference «Open Semantic Technologies for Intelligent Systems» (Minsk, 21-23 February 2019, 19-22 February 2020, 16-18 September 2021); 62nd International Conference For Students Of Physics And Natural Sciences «Open Readings 2019» (Vilnius, Lithuania, 19-22 March 2019); Международная научно-практическая конференция «Компьютерные технологии и анализ данных (СТДА'2020)» (Минск, 23-24 апреля 2020 г.);

и на 1 республиканской научной конференции: «71-я Научная конференция студентов и аспирантов БГУ» (Минск, 15 – 24 мая 2014 г.).

Результаты диссертации внедрены в УЗ РНПЦ «Мать и дитя», УЗ 3-я городская клиническая больница г. Минска им. Е.В. Клумова, Белорусском государственном университете (имеется 5 актов о внедрении).

Опубликованность результатов диссертации

Основные результаты диссертации опубликованы в 23 научных работах, из которых 6 статей в научных изданиях в соответствии с п. 18 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь (общим объемом 3,5 авторского листа), 5 статей в других научных рецензируемых изданиях, 10 статей в сборниках материалов научных конференций, 2 тезисов докладов.

Структура и объем диссертации

Диссертация состоит из перечня сокращений и условных обозначений, введения, общей характеристики работы, четырех глав, заключения, библиографического списка и четырех приложений. Полный объем диссертации составляет 148 страниц, включая 26 рисунков на 14 страницах, 4 таблицы на 3 страницах, 4 приложения на 28 страницах. Библиографический список содержит 137 наименований, включая 23 публикации соискателя (на 12 страницах).

ОСНОВНАЯ ЧАСТЬ

В первой главе проведен анализ существующих подходов и принципов построения экспертных медицинских систем. Показано, что существующие подходы, алгоритмы и системы не обеспечивают решения многих важных и актуальных на сегодняшний день проблем:

- отсутствуют решения по обобщению структуры, упрощению разработки и

интеграции, а также организации связи с существующими источниками и хранилищами данных;

- не решены вопросы автоматизации процессов создания адаптированного под решаемые задачи пользовательского интерфейса;
- модели принятия решений на базе правил не предоставляют возможности подстройки при поступлении новых данных;
- модели на базе машинного обучения с учителем не предоставляют возможности учета экспертного опыта и, в общем виде, не поддаются интерпретации с точки зрения процесса принятия решений и значимости используемых признаков.

В результате проведенного анализа выделены следующие задачи:

- создать информационную модель обобщенной экспертной медицинской системы;
- разработать метамодель описания структуры данных и организации взаимодействия основных компонентов информационной системы;
- разработать алгоритм генерации слоя доступа к данным и пользовательского интерфейса по метамодели;
- разработать алгоритм агрегации данных, представленных гетерогенными источниками;
- разработать алгоритмы поддержки принятия решений, учитывающие экспертные знания и исторические данные;
- разработать методы анализа корректности алгоритмов поддержки принятия решений и объяснения их работы;
- реализовать разработанные алгоритмы программно в рамках медицинских экспертных систем специального назначения;
- исследовать корректность полученных результатов.

Во второй главе проводится анализ требований к функциональности интегрированных медицинских систем, на основании которого предлагается информационная модель обобщенной экспертной медицинской системы, а также рассматриваются аспекты работы с гетерогенными источниками информации. В основе предложенной модели лежит метамодель – описание формата данных, которыми оперирует система:

$$M = \{E, A, T, m_A, m_T, V_E, V_A, I_E, I_A\}, \quad (1)$$

где $E = \{E_i\}$ – множество типов сущностей E_i модели (например, информация о пациенте, протокол обследования и т.п.),

$E_i \in E$ – множество сущностей определенного i -го типа (например, множество протоколов лабораторного анализа крови)

A – множество атрибутов сущностей (например, острота зрения и т.п.),
 T – множество типов данных атрибутов,
 $T_j \in T$ – множество допустимых значений j -го типа атрибута,
 $m_A: (E \times A) \rightarrow \mathbb{B}$ – предикат принадлежности атрибута типу сущности,
 $m_T: (A \times T) \rightarrow \mathbb{B}$ – предикат принадлежности типа данных атрибуту,
 $V_E: \{E_i \rightarrow \mathbb{B}\}$ – множество валидационных правил сущностей,
 $V_A: \{(A \times T_j) \rightarrow \mathbb{B}\}$ – множество валидационных правил атрибутов,
 I_E – множество метаданных для сущностей,
 I_A – множество метаданных для атрибутов.

На основании проекций из метамодели осуществляется формирование основных структурных составляющих информационной системы (рисунок 1) – подсистемы доступа к данным, подсистемы поддержки принятия решений и подсистемы пользовательского интерфейса [5, 6, 21].

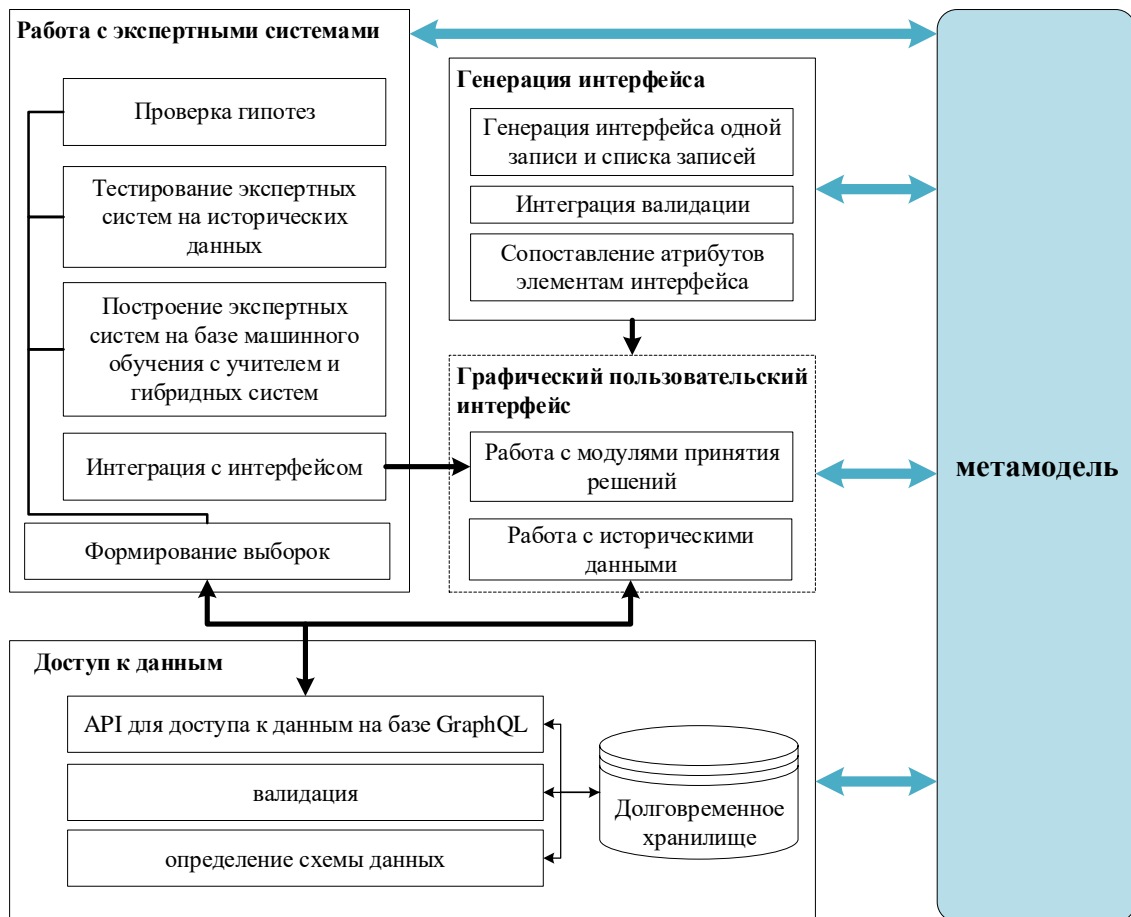


Рисунок 1. – Структура модели обобщенной экспертной медицинской системы

Подсистема доступа к данным использует метамодель (1) для формирования схемы данных, с помощью которой в дальнейшем осуществляется работа с информационным хранилищем системы. В работе разработаны проекции для реляционных и графовых СУБД.

Подсистема поддержки принятия решений использует информацию об атрибутах и типах из метамодели (1) для формирования выборок данных, на основании которых строятся, верифицируются и интегрируются различные модели поддержки принятия решений. Кроме того, эта подсистема отвечает за анализ значимости признаков и интерпретацию полученных результатов.

Пользовательский интерфейс реализует основные сценарии пользовательского взаимодействия с системой. Интерфейс определяется индивидуально для каждой информационной системы на основе анализа целей системы и, как правило, требует значительных затрат на разработку. Для решения этой задачи в работе предложен алгоритм процедурной генерации пользовательского интерфейса из описания, задаваемого метамоделью.

Для организации централизованного доступа к информации вводится схема данных S в виде мультиграфа, в котором определяется два основных типа вершин (узлов), соответствующих пациентам и областям обследований, а протоколы обследований задаются в виде рёбер между этими узлами:

$$S = \{V_p, V_a, E_p, K, W, p_{V_p}, p_{V_a}, p_{E_p}\}, \quad (2)$$

где V_p, V_a – множества вершин, соответствующих пациентам и областям исследований,

$E_p \subseteq V_p \times V_a$ – множество рёбер, соответствующих протоколу некоторого вида обследования $v_{a_i} \in V_a$ для некоторого пациента $v_{p_i} \in V_p$,

K – множество названий свойств,

W – множество значений свойств,

$p_{V_p}: V_p \rightarrow 2^{K \times W}, p_{V_a}: V_a \rightarrow 2^{K \times W}, p_{E_p}: E_p \rightarrow 2^{K \times W}$ – функции, ставящие в соответствие вершинам $v_{p_i} \in V_p, v_{a_j} \in V_a$ и рёбрам $e_{p_i} \in E_p$ некоторое множество пар ключ-значение $\{(k_i, w_i)\}, k_i \in K, w_i \in W$.

Предложенная схема реализована на основе графовой СУБД Neo4j. Ключевой особенностью предложенной схемы является низкая алгоритмическая сложность извлечения как информации о пациенте для формирования общей истории болезни, так и информации по конкретному виду обследования для анализа предметной области и формирования обучающих и валидационных выборок, что также подтверждено экспериментально [4, 9, 12, 20].

Для реализации доступа к данным, представленным гетерогенными источниками, предложен алгоритм агрегации с использованием формальных проекций, блок-схема которого представлена на рисунке 2. Каждая формальная проекция имеет вид $p: X \rightarrow \{o_i\}$, где X – множество всех возможных данных в информационной системе в исходном виде, $o_i: S \rightarrow S$ – описание операции по модификации графовой схемы (2). Сбор информации осуществляется парой

операций – проекция (map) для формирования операций и свёртка (reduce) для их выполнения, аналогично алгоритму MapReduce [17].

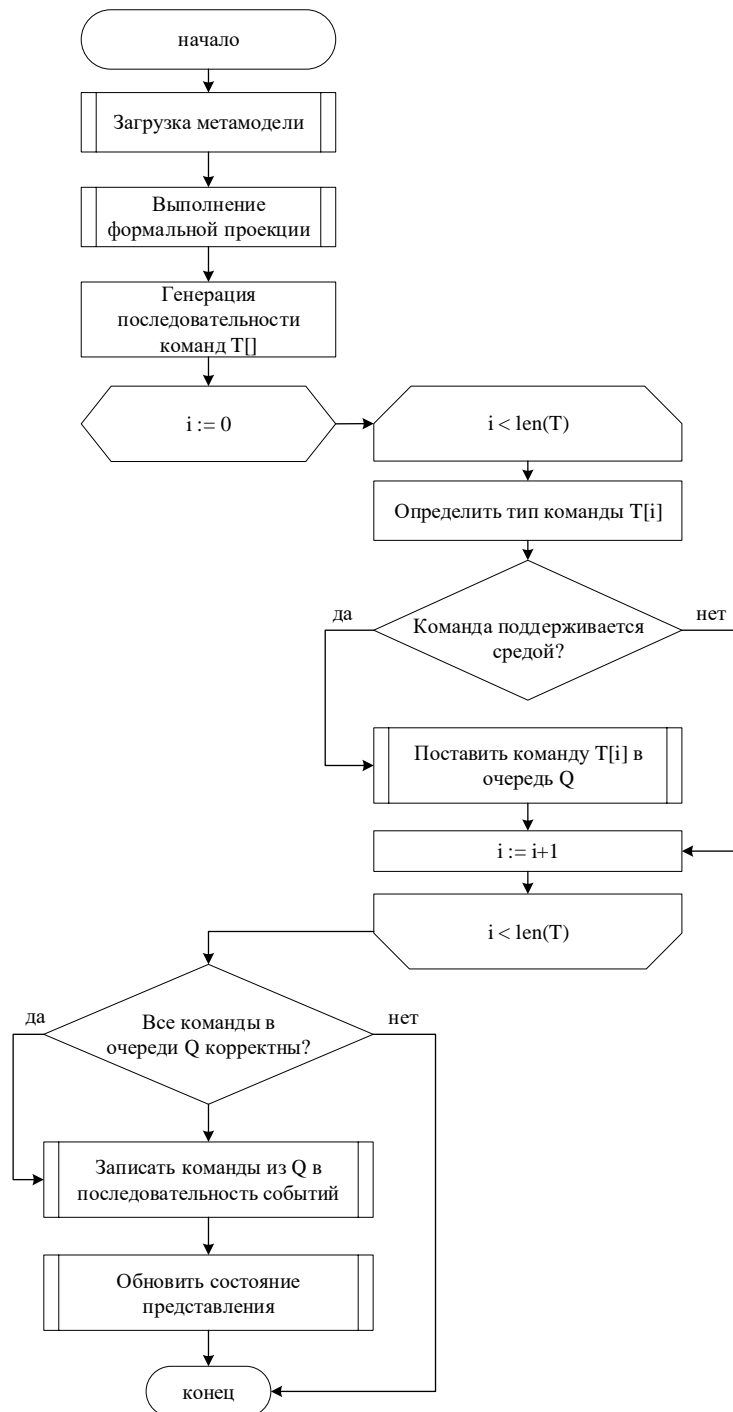


Рисунок 2. – Алгоритм извлечения информации из отдельного источника для обновления централизованного состояния с использованием формальных проекций

В третьей главе рассматриваются аспекты разработки и верификации решающих модулей медицинских экспертных систем.

Основным недостатком известных моделей поддержки принятия решений на основе правил (в частности, моделей на базе нечеткого вывода) является

отсутствие привязки процесса принятия решений к историческим данным. Для решения этой проблемы предложен алгоритм оптимизации параметров системы нечеткого вывода с использованием исторических данных на основании весовых коэффициентов правил, блок-схема которого представлена на рисунке 3.

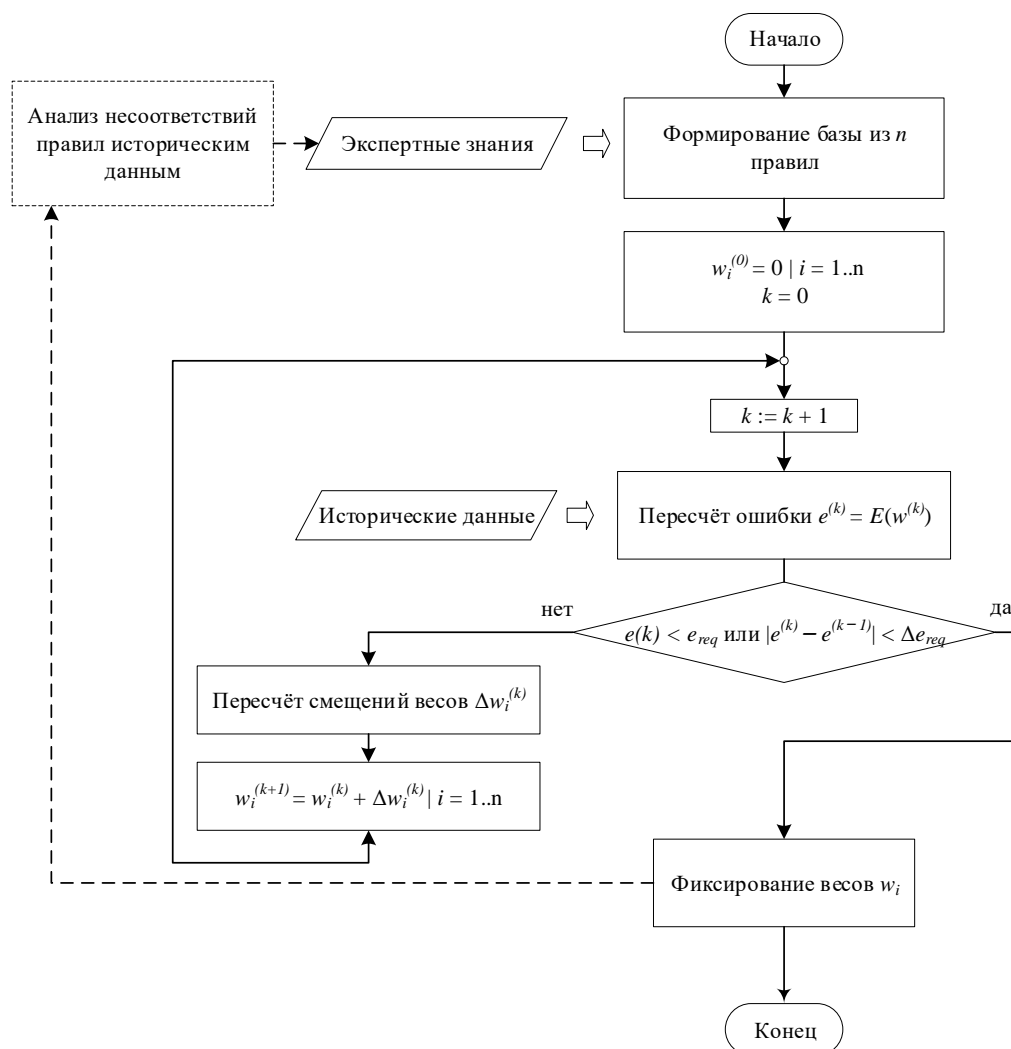


Рисунок 3. – Блок-схема процесса оптимизации системы нечеткого вывода с использованием исторических данных на основании весовых коэффициентов правил

По предложенному алгоритму экспертная система формируется как система нечеткого вывода, строится набор правил, но для каждого из правил вводится весовой коэффициент $w_i \in [0; 1]$. Вектор весов правил $\vec{w} = \{w_i\}$ выступает в качестве набора регулируемых параметров системы.

На основании совокупности исторических данных $\{(\vec{x}_i, \vec{y}_i)\}$ подбирается такой набор весов, который минимизирует ошибку системы нечеткого вывода $J(\vec{w})$, определяемую в данной работе по методу наименьших квадратов:

$$J(\vec{w}) = \sum_{i=1}^n (M(\vec{x}_i, \vec{w}) - \vec{y}_i)^2, \quad (3)$$

$$\vec{w}_{opt} = \operatorname{argmin}(J(\vec{w})) \Leftrightarrow J(\vec{w}_{opt}) = \min_{\vec{w} \in \mathbb{R}^n} J(\vec{w}). \quad (4)$$

В результате подбора весов формируется гибридная система нечеткого вывода со взвешенными правилами. На основании полученных данных можно также сделать вывод о корректности тех или иных правил для дальнейшей их доработки с экспертом и генерации новых экспертных знаний [2, 7, 16].

Основным недостатком моделей на основе алгоритмов машинного обучения с учителем является сложность интерпретации полученных решений. Для решения этой проблемы в работе предлагаются методы анализа значимости признаков, а также построения интерпретаций с использованием локальных линейных аппроксимаций.

Предложенные методы анализа значимости признаков основаны на последовательном обучении нейросетевой модели с одной и той же архитектурой на нескольких случайных разбиениях исходного набора данных с фиксированием весов. Для выборки $\{(\vec{x}_i, y_i)\}$ из n элементов в m -мерном признаковом пространстве $\vec{x}_i \in \mathbb{R}^m$ для задачи бинарной классификации $y_i \in \{0; 1\}$ и набора исключенных признаков $E \subset \overline{1, m}$ формируется некоторое количество q случайных разбиений для обучающей выборки $T \subset \{(\vec{x}_i, \vec{y}_i)\}$ и валидационной выборки $V \subset \{(\vec{x}_i, \vec{y}_i)\}$. При обучении веса связи нейронов, соответствующих исключенным признакам, со всеми нейронами первого скрытого слоя, принимаются равными нулю. После получения обученной модели для данного набора исключенных признаков $C_i^{(E)}: \mathbb{R}^m \rightarrow \{0; 1\}$ на каждом из q разбиений подсчитывается F_1 -мера. В качестве степени влияния k -го признака оценивается суммарное снижение F_1 -меры по сравнению с исходным классификатором ω_k . Для относительной оценки значимости признаков используется нормированное на максимум значение $\omega_k^* = \omega_k / \max_{i \in \overline{1, n}} \omega_i$.

Анализ значимости признаков использовался в исследованиях по выявлению рассеянного склероза и прогнозированию исхода системной тромболитической терапии. Медицинская интерпретация полученных результатов приводится в статьях [9] и [11].

Для формирования интерпретаций по обученной модели в условиях нелинейной функциональной зависимости между входными и выходными признаками наиболее простым для восприятия и интерпретации является представление, приближенное к линейному разделению. На основании этого предположения в работе предлагается осуществлять построение семантических интерпретаций с использованием локальных линейных аппроксимаций. Для обучающей выборки T из k элементов в задаче бинарной классификации по n признакам и классификатора $C(\vec{x}): \mathbb{R}^n \rightarrow \mathbb{B}$ введем меру корректности на

обучающей выборке T , например, F_1 -меру. Для нелинейного параметрического бинарного классификатора $C_f(\vec{x}, \Theta): \mathbb{R}^n \rightarrow \mathbb{B}$ с набором параметров Θ , которые подобраны в результате обучения, из входных векторов обучающей выборки T , выбирается вектор, ближайший к \vec{x}^* . После нахождения ближайшей точки и минимального расстояния d_{min} осуществляется случайная генерация набора точек $\{\vec{x}^{(s)}\}$ фиксированной длины m , которые удалены от \vec{x}^* не больше, чем на d_{min} . Для каждой из m полученных точек $\{\vec{x}^{(s)}\}$ определяется соответствующее значение $y_i^{(s)}$, вычисленное обученным классификатором. Набор точек $\{\vec{x}^{(s)}\}$ и соответствующих им выходных значений классификатора $\{y_i^{(s)}\}$ характеризуют поведение классификатора C_f в окрестности точки \vec{x}^* . Для построения локальной линейной аппроксимации осуществляется построение линейного классификатора $C^*(\vec{x}, \vec{\theta}): \mathbb{R}^n \rightarrow \mathbb{B}$ с использованием этих точек в качестве обучающей выборки. Полученный линейный классификатор задаёт гиперплоскость $p^*(\vec{x}) = \theta_0 + \vec{\theta} \cdot \vec{x}$, которая разделяет исходное n -мерное пространство решений и задается при помощи $(n + 1)$ -мерного вектора параметров классификатора $\vec{\theta}$. По локальному линейному классификатору C^* и соответствующей ему разделяющей гиперплоскости $p^*(\vec{x})$ может быть рассмотрена точка пересечения этой гиперплоскости с осью каждого из n исходных признаков. Пусть угол между единичным вектором оси x_n и вектором нормали к плоскости $p^*(\vec{x})$ составляет φ_{n_0} , а точка пересечения гиперплоскости $p^*(\vec{x})$ с осью x_n , если она существует, находится в $x_n = x_{n_0}$. Тогда в качестве значения веса влияния n -го признака на результат классификации может использоваться значение $w_n^* = \cos \varphi_{n_0}$. Для исключения влияния размерности - значения для каждой из осей могут нормироваться. Семантически полученные значения весов w_i^* и точек пересечения x_{n_0} могут быть интерпретированы в виде логических высказываний следующего вида: «Для точки \vec{x}^* обученный классификатор C_f принял решение $y_f^* = C_f(\vec{x}^*, \Theta)$, причем с весом w_i^* на это решение повлиял тот факт, что i -й признак принял значение x_i^* , большее (меньшее), чем x_{i_0} ». Полученные значения весов w_i^* могут быть отсортированы по убыванию для определения признаков, имеющих наибольшее влияние на выходной результат. [10, 18].

В четвертой главе приводится описание программной реализации метамоделей и обобщенного фреймворка, на базе которого предложенные методы и алгоритмы использованы при создании трех узкоспециализированных медицинских экспертных систем.

Экспертная программа для определения хориальности и прогнозирования течения беременности при многоплодии служит для оценки хориальности многоплодных беременностей на раннем этапе. В качестве источника данных в системе использовалась информация об исходах многоплодных беременностей

в табличном виде. Решающий модуль реализован на базе системы нечеткого вывода по алгоритму Такаги-Сугено-Канга. Сгенерированный интерфейс пользователя оформлен в виде приложения рабочего стола под управлением Electron. При оптимизации относительно части выборки с использованием алгоритма, предложенного в четвертой главе, установлена некорректность одного из правил, а также повышена F1-меру классификации с 0.731 до 0.956 [7]. Разработанная программа зарегистрирована в НЦИС (рег. св. № 986 от 23.11.2017), а также внедрена в медицинскую практику в ГУ РНПЦ «Мать и дитя» (акт о внедрении от 30.10.2020).

Экспертная система учета и ранней диагностики дегенеративных оптиконейропатий на основе нейросетевого анализа позволяет обрабатывать данные об офтальмологическом обследовании пациентов, а также формировать прогноз о наличии глаукомной оптиконейропатии и дегенеративной оптиконейропатии. В качестве источника данных использовались табличные данные и информация протоколов диагностического оборудования в формате PDF. Итоговый набор данных для обучения содержал информацию о 408 глазах с диагностированным рассеянным склерозом, 380 глазах с диагностированной глаукомой, и 326 глазах контрольной группы без нарушений. Решающие модули основаны на многослойных перцептронах прямого распространения. Нейросетевой классификатор рассеянного склероза характеризуется следующими статистическими показателями на независимой выборке: чувствительность – 89,5%, специфичность – 85,7%, F1-мера – 0,876, площадь под ROC-кривой (AUC) – 0,822. Нейросетевой классификатор глаукомной оптиконейропатии имеет следующие характеристики: чувствительность – 91,0%, специфичность – 91,0%, F1-мера – 0,910, площадь под ROC-кривой (AUC) – 0,871. Полученные характеристики свидетельствуют о более точной классификации по сравнению с существующими подходами [8]. Анализ значимости признаков при помощи методов, предложенных в четвертой главе, указал на более высокую чувствительность показателей сканирующей лазерной поляриметрии и позволил выявить вторичность поражения зрительного нерва и сетчатки, что является значимым результатом с точки зрения медицины [9].

Программный комплекс «Экспертная система учета и ранней диагностики дегенеративных оптиконейропатий на основе нейросетевого анализа» зарегистрирован в НЦИС (рег. св. №1257 от 05.02.2020), а также внедрен в УЗ «3-я городская клиническая больница Е.В. Клумова».

Система прогнозирования исхода системной тромболитической терапии (сТЛТ) у пациентов с ишемическим инсультом предназначена для определения стратегии лечения при диагностике симптомов раннего ишемического инсульта. Особенностью разрабатываемой системы является использование веб-

интерфейса, адаптированного для работы на мобильных устройствах (смартфонах и планшетах), т. к. формирование экспертной оценки по первичным показателям может потребоваться непосредственно на выезде бригады скорой медицинской помощи [11]. На момент подготовки текста работы система доступна по адресу <https://www.bsmu.by/page/59/7408/>. Метод оценки значимости признаков, описанный в четвертой главе, позволил выявить факторы, обуславливающие благоприятный функциональный исход терапии как с точки зрения безопасности, так и с позиций маршрутизации пациентов [11]. Решающий модуль представлен нейронной сетью прямого распространения. Исходный набор данных включал 174 положительных экземпляров и 82 отрицательных, соответствующих 1 основной группе и 1 контрольной группе, а также 408 положительных экземпляров и 126 отрицательных, соответствующих 2 основной группе и 2 контрольной группе. При внешней валидационной проверке были получены следующие результаты: чувствительность 88,5%, специфичность 98,2%, F1-мера 0.945, площадь под ROC-кривой 0,93 [11]. Разработанное приложение и полученные результаты использованы при подготовке инструкции по применению Министерства здравоохранения Республики Беларусь «Метод прогнозирования исходов системной тромболитической терапии у пациентов с инфарктом головного мозга с помощью нейросетевой модели» (№034-0521 от 21.05.2021).

В приложении приведены численные данные, используемые для приведенных в работе таблиц и графиков, свидетельства о регистрации компьютерных программ, удостоверение на рационализаторское предложение и акты внедрения, подтверждающие апробацию и тестирование разработанных в диссертационной работе решений.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Разработана модель обобщенной экспертной медицинской системы с возможностями обработки медицинских данных, генерации пользовательского интерфейса и интеграции систем поддержки принятия решений на основе описательного задания с помощью метамоделей, позволяющая осуществлять прототипирование приложений и проверку методических гипотез, что упрощает внедрение таких систем в медицинскую практику и повышает качество оказываемых медицинских услуг [1, 5, 6, 13, 21].

2. Разработан алгоритм агрегации и трансформации слабоструктурированных и денормализованных данных на основании формальных проекций, что позволяет унифицировать процесс обработки

информации из различных видов хранилищ и осуществить извлечение данных в общую схему. Предлагаемый механизм может применяться в качестве промежуточного слоя доступа к данным для облегчения формирования запросов, не заменяя при этом уже существующие системы и используемые в них хранилища, что позволяет упростить и ускорить развертывание системы [4, 17, 20, 22].

3. Предложена графовая схема данных для организации и обработки полученной агрегированной из разных источников информации, что позволяет обеспечить поддержку сложных запросов к нескольким разнородным источникам информации и снижает время выполнения таких запросов на 10-15% по сравнению с традиционными хранилищами. Предложенный подход позволяет оптимизировать выборку информации по совокупности обследований конкретного пациента для просмотра общей истории болезни, формирования клинической картины и упрощения постановки комплексного диагноза, а также выборку по совокупности протоколов обследований конкретного вида обследования, что позволяет анализировать статистические зависимости, а также разрабатывать и тестировать медицинские экспертные системы и системы поддержки принятия решений, опирающиеся на исторические данные [4, 12, 20].

4. Предложены методы оценки влияния входных параметров и семантической интерпретации процессов принятия решения в обученных моделях машинного обучения с учителем, что позволяет оценивать корректность таких моделей с точки зрения экспертного опыта, а также использовать полученные интерпретации для генерации новых экспертных знаний. Методы апробированы на медицинских данных и позволили выявить новые зависимости в соответствующих областях, которые были дополнительно исследованы и подтверждены медицинскими специалистами [3, 9–11, 14, 18].

5. Разработан алгоритм подстройки экспертных систем на базе правил с использованием исторических данных, основанный на подборе весов правил по скрытым статистическим зависимостям выборки, что позволяет модифицировать процесс принятия решений с поступлением новых данных и повышает F1-меру для бинарной классификации по исследованным данным на 0,13 – 0,17. Разработанный алгоритм может использоваться не только для оптимизации процесса принятия решения с использованием исторических данных, но и для переоценки экспертных знаний, поиска аномалий в исторических данных, а также установления новых эмпирических зависимостей в предметных областях [2, 7, 8, 15, 16, 23].

6. На базе предложенной модели обобщенной экспертной медицинской системы, разработанных алгоритмов и методов созданы три программные комплекса для организации поддержки принятия решений в различных областях

медицины. Проведено внедрение и тестирование предложенных решений [5–7, 9, 11, 13, 15, 19].

Рекомендации по практическому использованию результатов

1. Созданная компьютерная программа «Экспертная программа для определения хориальности и прогнозирования течения беременности при многоплодии» зарегистрирована в Национальном центре интеллектуальной собственности (№ 986 от 23.11.2017 г.), может использоваться в медико-генетических центрах и позволяет осуществлять оценку плацентации при многоплодной беременности.

2. Разработанная компьютерная программа «Экспертная программа для определения хориальности и прогнозирования течения беременности при многоплодии» используется в клинической практике в ГУ РНПЦ «Мать и дитя» с базой данных многоплодных беременностей, на основании чего оформлены акт внедрения от 30.10.2020 и рационализаторское предложение под наименованием «Оценка плацентации при многоплодной беременности с использованием современных экспертных компьютерных программ» №113 от 02.03.2018.

3. Разработанная компьютерная программа «Экспертная система учета и ранней диагностики дегенеративных оптиконеуропатий» зарегистрирована в Национальном центре интеллектуальной собственности (№ 1257 от 06.02.2020 г.) и позволяет повысить эффективность диагностики глаукомной оптиконеуропатии у пациентов с первичной открытоугольной глаукомой и повысить эффективность выявления дегенеративной оптиконеуропатии у пациентов с рассеянным склерозом на основе использования искусственных нейронных сетей.

4. Разработанная компьютерная программа «Экспертная система учета и ранней диагностики дегенеративных оптиконеуропатий» используется в лечебной практике городского офтальмологического консультативно-диагностического центра УЗ «3-я городская клинической больницы имени Е. В. Клумова».

5. Разработанная компьютерная программа «Система прогнозирования исхода системной тромболитической терапии» позволяет определить стратегии лечения при диагностике симптомов раннего ишемического инсульта у пациентов с ишемическим инсультом.

6. Разработанные компьютерная программа «Экспертная система учета и ранней диагностики дегенеративных оптиконеуропатий» и алгоритм агрегации, трансформации и индексирования слабоструктурированных и денормализованных данных могут использоваться в образовательном процессе и внедрены на факультете радиофизики и компьютерных технологий Белорусского государственного университета.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ УЧЕНОЙ СТЕПЕНИ***Статьи в научных изданиях в соответствии с п. 18 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь и рецензируемых зарубежных изданиях***

1. Курочкин, А. В. Использование моделей динамической топологии одноранговой сети для эмуляции внутрисетевого взаимодействия / А. В. Курочкин, И. А. Адуцкевич // *Электроника инфо.* – 2015. – № 4. – С. 50–53.

2. Курочкин, А. В. Оптимизация процесса принятия решений в медицинских экспертных системах на базе нечеткой логики с использованием исторических данных / А. В. Курочкин, В. С. Садов, О. М. Демиденко // *Проблемы физики, математики и техники.* – 2019. – № 1 (38). – С. 78–84.

3. Курочкин, А. В. Методы анализа и семантической интерпретации процессов принятия решения в классификационных нейросетевых моделях машинного обучения с учителем / А. В. Курочкин, В. С. Садов // *Вестн. Полоц. гос. ун-та. Сер. С, Фундам. науки.* – 2019. – № 12. – С. 57–61.

4. Курочкин, А. В. Агрегация и индексирование данных нескольких источников на основе графовой модели в базах данных медицинских экспертных систем / А. В. Курочкин, В. С. Садов // *Информатика.* – 2020. – Т. 17, № 3. – С. 25–35.

5. Kurachkin, A. Integration medical data management and decision-making systems with common metamodel / A. Kurachkin // *International Journal of Open Information Technologies.* – 2020. – Vol. 8, № 12. – P. 49–53.

6. Kurochkin, A. Integrated medical information and decision-support system development based on shared metamodel definition / A. Kurochkin, V. Sadau, A. Halavatyi // *Open semantic technologies for intelligent systems : research papers coll. = Открытые семантические технологии проектирования интеллектуальных систем : сб. науч. тр. / Belarusian State Univ. of Informatics a. Radioelectronics.* – Minsk, 2021. – Iss. 5. – P. 237–340.

Статьи в других научных рецензируемых изданиях

7. Прибушения, О. В. Оценка плацентации при многоплодной беременности с использованием современных экспертных компьютерных программ / О. В. Прибушения, А. В. Курочкин // *Современные перинатальные медицинские технологии в решении проблем демографической безопасности : сб. науч. тр. / Мин. здрав. Респ. Бел. ; ГУ «Респ. НПЦ «Мать и дитя» ; редкол.: К. У. Вильчук [и др.].* – Минск, 2017. – Вып. 10. – С. 106–111.

8. Kurochkin, A. Fuzzy logic inference ruleset augmentation with sample data in medical decision-making systems / A. Kurochkin, V. Sadau // *Open Semantic Technologies for Intelligent Systems : research papers coll. / Belarusian State university of Informatics and Radioelectronics.* – Minsk, 2019. – Iss. 3. – P. 329–332.

9. Роль искусственных нейронных сетей в выявлении ранней гибели ганглионарных клеток сетчатки у пациентов с дегенеративными оптиконеуропатиями / Т. В. Качан, А. В. Курочкин, Е. А. Головатая, Л. Н. Марченко, А. С. Федулов, А. А. Далидович, О. В. Скрыпник, Т. А. Муштина // *Офтальмология. Вост. Европа.* – 2019. – Т. 9, № 4. – С. 446–457.

10. Kurachkin, A. Decision-making process analysis and semantic explanation extraction from trained supervised machine learning models in medical expert systems / A. Kurochkin, V. Sadau, T. Kachan // *Open Semantic Technologies for Intelligent Systems : research papers coll. / Belarusian State University of Informatics and Radioelectronics.* – Minsk, 2020. – Iss. 4. – P. 283–286.

11. Прогнозирование исхода тромболитической терапии у пациентов с ишемическим инсультом на основе применения нейросетевого анализа / К. В. Сенько, А. С. Федулов, А. В. Курочкин, Е. А. Головатая // *Неврология и нейрохирургия. Вост. Европа.* – 2020. – Т. 10, № 3. – С. 353–366.

Статьи в сборниках материалов научных конференций

12. Курочкин, А. В. Исследование методов моделирования динамики топологии ячеистых одноранговых сетей / А. В. Курочкин, И. А. Адуцкевич // *Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния : материалы Третьей Междунар. науч.-практ. конф., Минск, 28–29 апр. 2015 г. / Ин-т приклад. физ. проблем Белорус. гос. ун-та ; редкол.: В. И. Попечиц (гл. ред.) [и др.].* – Минск, 2015. – С. 181–183.

13. Доросинец, Р. О. Микросервисный подход к архитектуре веб-приложений / Р. О. Доросинец, А. В. Курочкин // *Инновационные решения проблем экономики знаний Беларуси и Казахстана : сб. материалов науч.-практ. конф., Минск, 13 окт. 2016 г. / Белорус. нац. техн. ун-т.* – Минск, 2016. – С. 257–258.

14. Михайловский, И. А. Предварительная обработка информации в задаче семантического анализа текстов на естественном языке / И. А. Михайловский, А. В. Курочкин // *Инновационные решения проблем экономики знаний Беларуси и Казахстана : сб. материалов науч.-практ. конф., Минск, 13 окт. 2016 г. / Белорус. нац. техн. ун-т.* – Минск, 2016. – С. 282–283.

15. Курочкин, А. В. Экспертная медицинская система по определению хориальности на основе системы нечеткой логики / А. В. Курочкин, О. В. Прибушения, В. С. Садов // *Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф., Минск, 25 окт. 2017 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.].* – Минск, 2017. – С. 92–93.

16. Курочкин, А. В. Оптимизация параметров экспертной системы оценки хориальности многоплодных беременностей на основе системы нечеткой логики по историческим данным / А. В. Курочкин, В. С. Садов // *Актуальные вопросы современной науки : сб. ст. по материалам XIV Междунар. науч.-практ.*

конф., Томск, 26 сент. 2018 г. : в 2 ч. / НИЦ Вестник науки. – Уфа, 2018. – Ч. 1. – С. 13–17.

17. Чернявский, А. Ф. Обработка и индексирование денормализованных и слабоструктурированных данных / А. Ф. Чернявский, А. В. Курочкин, В. С. Садов // Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния : материалы пятой Междунар. науч.-практ. конф., Минск, 16–17 мая 2019 г. / Ин-т приклад. физ. проблем Белорус. гос. ун-та ; редкол.: В. И. Попечиц [и др.]. – Минск, 2019. – С. 164–165.

18. Курочкин, А. В. Анализ процесса принятия решения в модели нейросетевой классификации методом локальных линейных аппроксимаций / А. В. Курочкин, В. С. Садов // Информационные технологии и системы 2019 (ИТС 2019) : материалы междунар. науч. конф., Минск, 30 окт. 2019 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2019. – С. 228–229.

19. Ляпко, А. М. Библиотека для создания экспертных систем на базе алгоритмов нечеткого вывода на языке TypeScript / А.М. Ляпко, А. В. Курочкин // Информационные технологии и системы 2019 (ИТС 2019) : материалы междунар. науч. конф., Минск, 30 окт. 2019 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2019. – С. 118–119.

20. Курочкин, А. В. Использование графовой схемы для организации хранения и обработки данных медицинских систем / А. В. Курочкин, В. С. Садов // Компьютерные технологии и анализ данных (СТДА'2020) : материалы междунар. науч.-практ. конф., Минск, 23-24 апр. 2020 / Белорус. гос. ун-т ; И-т ИТ и бизнес-администрирования ; ИПУП «Иссофт Солюшенз» ; редкол.: В. В. Скакун [и др.]. – Минск : БГУ, 2020. – С. 248–252.

21. Курочкин, А. В. Обобщенная модель медицинских экспертных систем поддержки принятия решений / А. В. Курочкин // Информационные технологии и системы (ИТС 2020) : материалы междунар. конф., Минск, 18 ноября 2020 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 84–85.

Тезисы докладов

22. Курочкин, А. В. Защита информации и информационная безопасность в медицинских системах / А. В. Курочкин, Е. А. Головатая // Технические средства защиты информации : тез. докл. XV Белорус.-рос. науч.-техн. конф., Минск, 6 июня 2017 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Т. В. Борботько [и др.]. – Минск, 2017. – С. 52.

23. Kurachkin, A. Fuzzy inference system for automated image-based karyotyping / A. Kurachkin, V. Sadau // 62nd international conference for students of physics and natural sciences «Open readings 2019» : book of abstr., Vilnius, Lithuania, March 19-22, 2019 / Vilnius Univ. ; ed.: E. Skliutas [et al.]. – Vilnius, 2019. – P. 470.

Курочкин Александр Васильевич

Модель, методы и алгоритмы построения информационно-аналитических экспертных систем учета и анализа данных медицинских обследований

Ключевые слова: экспертные медицинские системы, обработка данных, семантическая интерпретация, машинное обучение.

Целью диссертационной работы является формирование обобщенного подхода к разработке, статистической верификации, семантической интерпретации и внедрению информационно-аналитических экспертных медицинских систем для повышения эффективности оказания медицинской помощи.

Методы исследования: машинное обучение, нейронные сети, нечеткая логика, методы численной оптимизации, методы статистического анализа, а также методы и средства разработки программного обеспечения.

Получены новые результаты: модель экспертной медицинской системы поддержки принятия решений; алгоритм агрегации, трансформации и индексирования слабоструктурированных и денормализованных данных на основании формальных проекций и графовой схемы данных, позволяющий унифицировать процесс обработки информации из различных видов хранилищ; метод оценки влияния входных параметров и семантической интерпретации процессов принятия решения в обученных моделях машинного обучения с учителем на основе локальных линейных аппроксимаций, позволяющий оценивать семантическую корректность таких моделей и использовать полученные интерпретации для генерации новых экспертных знаний; алгоритм оптимизации экспертных систем на базе правил с использованием исторических данных, основанный на подборе весов правил по скрытым статистическим зависимостям выборки.

Результаты апробировались и используются УЗ «3-я городская клинической больницы имени Е.В. Клумова»; ГУ Республиканский научно-практический центр «Мать и дитя»; кафедрой интеллектуальных систем Белорусского государственного университета.

Областью применения разработанных моделей и алгоритмов являются системы поддержки принятия решений медицинской диагностики.

Курачкін Аляксандр Васільевіч**Мадэль, метады і алгарытмы пабудовы інфармацыйна-аналітычных экспертных сістэм уліку і аналізу дадзеных медыцынскіх абследаванняў**

Ключавыя словы: экспертныя медыцынскія сістэмы, апрацоўка дадзеных, семантычная інтэрпрэтацыя, машыннае навучанне.

Мэтай дысертацыйнай працы з'яўляецца фарміраванне абагульненага падыходу да распрацоўкі, статыстычнай верыфікацыі, семантычнай інтэрпрэтацыі і ўкараненню інфармацыйна-аналітычных экспертных медыцынскіх сістэм для павышэння эфектыўнасці аказання медыцынскай дапамогі.

Метады даследавання: машыннае навучанне, нейронныя сеці, нячоткая логіка, метады лічбавай аптымізацыі, метады статыстычнага аналізу, а таксама метады і сродкі распрацоўкі праграмнага забеспячэння.

Атрыманы новыя вынікі: мадэль экспертная медыцынскай сістэмы падтрымкі прыняцця рашэнняў; алгарытм агрэгавання, трансфармацыі і індэксавання слабоструктуріраваных і денормалізаваных дадзеных на падставе фармальных праекцый і графавая схема дадзеных, які дазваляе ўніфікаваць працэс апрацоўкі інфармацыі з розных відаў сховішчаў; метады ацэнкі ўплыву ўваходных параметраў і семантычнай інтэрпрэтацыі працэсаў прыняцця рашэння ў навучаных мадэлях машыннага навучання з настаўнікам на аснове лакальных лінейных апраксімацыі, які дазваляе ацэньваць семантычную карэктнасць такіх мадэляў і выкарыстоўваць атрыманыя інтэрпрэтацыі для генерацыі новых экспертных ведаў; алгарытм аптымізацыі экспертных сістэм на базе правілаў з выкарыстаннем гістарычных дадзеных, заснаваны на падборы вагаў правіл па схаваным статыстычным залежнасцяў выбаркі.

Вынікі апрабаваны і выкарыстоўваюцца УЗ «3-я гарадская клінічная бальніца імя А.В. Клумава»; ГУ Рэспубліканскі навукова-практычны цэнтр «Маці і дзіця»; кафедрай інтэлектуальных сістэм Беларускага дзяржаўнага ўніверсітэта.

Вобласцю ужывання распрацаваных мадэляў і алгарытмаў з'яўляюцца сістэмы падтрымкі прыняцця рашэнняў медыцынскай дыягностыкі.

SUMMARY**Kurachkin Aliaksandr Vasilevich****Model, methods and algorithms for constructing information and analytical expert systems for storing and analyzing medical research data**

Keywords: expert medical systems, data processing, semantic interpretation, machine learning.

The aim of the thesis is to form a generalized approach to the development, statistical verification, semantic interpretation and implementation of information and analytical expert medical systems to improve the efficiency of medical care.

Research methods: machine learning, neural networks, fuzzy logic, numerical optimization methods, statistical analysis methods, as well as software development methods and tools.

New results have been obtained: a model of an expert medical decision support system; an algorithm for aggregation, transformation and indexing of loosely structured and denormalized data based on formal projections, and a graph data schema, which allows unifying the process of processing information from various types of storages; a method for assessing the influence of input parameters and semantic interpretation of decision-making processes in trained supervised machine learning models based on local linear approximations, which makes it possible to verify the semantic correctness of such models and use the resulting interpretations to generate new expert knowledge; an algorithm for optimizing rule-based expert systems using historical data, based on the rule weight optimization from hidden statistical dependences in the historical data sample.

The results were tested and used by the 3rd city clinical hospital named after E.V. Klumova; Republican scientific and practical center «Mother and child»; Department of Intelligent Systems of the Belarusian State University.

Application fields of the developed models and algorithms are decision support systems for medical diagnostics.

