

Благосклонов Николай Алексеевич

**Диагностическая экспертная система по орфанной патологии на модели
наследственных лизосомных болезней накопления**

3.3.9. Медицинская информатика (медицинские науки)

Автореферат
диссертации на соискание ученой степени
кандидата медицинских наук

Москва

2025

Работа выполнена в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук»

Научный руководитель:

доктор медицинских наук, профессор

Кобринский Борис Аркадьевич

Официальные оппоненты:

доктор медицинских наук, доцент

Карась Сергей Иосифович

Научно-исследовательский институт кардиологии – филиал Федерального государственного бюджетного научного учреждения «Томский национальный исследовательский медицинский центр Российской академии наук», специалист отдела координации научной и образовательной деятельности

доктор медицинских наук

Владзимирский Антон Вячеславович

Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», заместитель директора по научной работе

Ведущая организация:

федеральное государственное автономное образовательное учреждение высшего образования Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет)

Защита диссертации состоится «_____» _____ 2025 года в _____ часов на заседании объединённого диссертационного совета 99.0.122.02 на базе ГНЦ РФ - ИМБП РАН и ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (Пироговский Университет) по адресу: 117513, г. Москва, ул. Островитянова, д. 1, стр. 6.

С диссертацией можно ознакомиться в научной библиотеке ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (Пироговский Университет) по адресу: 117513, г. Москва, ул. Островитянова, д. 1, стр. 5 и на сайте <https://rsmu.ru/>.

Автореферат разослан «_____» _____ 2025 года

Учёный секретарь диссертационного совета
доктор медицинских наук, доцент



Антонов Андрей Евгеньевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Наследственные болезни вызывают особенные трудности в диагностике на долабораторном этапе. Сложность первичной постановки диагноза объясняется тем, что большинство заболеваний генетической природы являются орфанными, то есть редко встречаемыми (менее 10 случаев на 100 000 населения). Одной из групп орфанных заболеваний являются лизосомные болезни накопления (ЛБН), которые в большинстве случаев проявляются в детском возрасте и характеризуются прогрессирующим течением. Низкая распространенность ЛБН приводит к тому, что даже у практикующих врачей-генетиков может отсутствовать личный опыт по распознаванию клинических форм (Шашель В.Н. и др., 2021) по фенотипическим проявлениям. Дополнительную сложность для дифференциальной диагностики ЛБН на долабораторном этапе создают (Горбунова В.Н. и др., 2021), с одной стороны, большая вариативность клинических проявлений, а с другой схожесть клинической картины. Подтверждение диагноза осуществляется посредством молекулярно-генетического тестирования, которое является дорогостоящим исследованием, в связи с чем возникает необходимость максимально возможного сужения диагностического ряда.

При своевременной и точной постановке диагноза появляется возможность с использованием новых патогенетических методов лечения предотвратить прогрессирование заболевания и повысить качество жизни пациента. Таким образом разработка и внедрение компьютерной системы поддержки принятия решений для дифференциальной диагностики ЛБН, ассистирующей врачу в поиске наиболее вероятных диагнозов по фенотипическим проявлениям, является актуальной задачей.

Степень разработанности темы исследования

Более 30 лет назад началось создание отечественных (Кобринский Б.А., 1991) и зарубежных (Gouvernet J. et al., 1985, Fryer A., 1991) компьютерных диагностических систем в области клинической генетики, продолжающееся в настоящее время (Allanson J.E. et al., 2009, Ronicke S. et al., 2019). Они позволили

повысить уровень распознавания генетически детерминированных болезней. Однако эти системы в недостаточной степени учитывали ранние проявления болезни и прогрессивность патологического процесса. Поэтому актуальным является поиск новых подходов к построению диагностических систем в предметной области наследственных болезней, учитывающих их прогрессирующую динамику. В том числе создание интеллектуальных систем, использующих модели представления экспертных знаний.

В настоящее время отсутствуют отечественные системы для поддержки постановки диагноза на долабораторном этапе у пациентов с подозрением на наследственные заболевания.

Целью диссертационной работы является повышение точности и своевременности идентификации генетических болезней на основе интеллектуальной (экспертной) системы поддержки принятия решений, формирующей узкий дифференциально-диагностический ряд на долабораторном этапе обследования пациентов (на примере ЛБН). Для достижения цели в диссертационной работе были поставлены следующие **задачи**:

1. Выделить диагностически значимые признаки ЛБН в литературных источниках и представить их в структурированном виде.
2. Разработать шкалы для экспертной оценки признаков (модальность, манифестация и выраженность) для четырех возрастных периодов.
3. Извлечь знания от врачей-экспертов о проявлениях ЛБН и представить их в формализованном виде.
4. Разработать модель интегрированной оценки клинической картины ЛБН в различных возрастных периодах.
5. Разработать классификационный алгоритм, обеспечивающий сравнение клинической картины пациента с эталонными описаниями болезни в различных возрастных периодах.
6. Разработать интеллектуальную (экспертную) систему для дифференциальной диагностики орфанных наследственных болезней и оценить её эффективность.

Научная новизна диссертационной работы состоит в разработке компьютерной системы, обеспечившей повышение эффективности диагностики редких наследственных заболеваний на основе методов искусственного интеллекта, и заключается в следующем:

1. Модифицирована модель триединства факторов уверенности представления экспертных знаний путём замены частоты встречаемости признака на модальность в дополнение к манифестации и выраженности.

2. Разработана интегрированная модель заболевания, включающая комплексные оценки признаков и учитывающая факторы уверенности экспертов к трём характеристикам признаков в разных возрастных периодах с целью отображения клинической картины эталонного варианта заболевания и конкретного случая.

3. Разработан алгоритм оценки степени сходства клинических проявлений нового случая с эталонными описаниями заболеваний в базе знаний на основе интегрированной модели заболевания, который обеспечивает ранжирование выдвигаемых в системе гипотез (возможных диагнозов) в ситуациях условной избыточности или неполноты данных о пациенте.

Теоретическая и практическая значимость работы

Теоретическая значимость заключается в создании научно-обоснованного представления знаний на основе характеристик атрибутов, модели заболевания и алгоритма, позволяющего осуществлять поддержку принятия клинических решений при диагностике наследственных ЛБН.

Результаты, изложенные в диссертации, получили практическую реализацию в виде компьютерной интеллектуальной системы ГенДиЭС (Генетическая Диагностическая Экспертная Система) для дифференциальной диагностики на долабораторном этапе группы орфанных наследственных болезней – 30 клинических форм ЛБН.

Изложенный в диссертации подход может использоваться для задач разработки экспертных систем по дифференциальной диагностике заболеваний различного генеза, для которых затруднительно или невозможно применять

вычислительные подходы для установления статистически значимых причинно-следственных связей между фенотипическими проявлениями и клинически-подтверждёнными диагнозами.

Методология и методы исследования

В работе использовались методы инженерии знаний при структурировании информации из литературных источников и работе с врачами-экспертами, элементы нечёткой логики при разработке ранговых и интервальных шкал. База знаний формировалась на основе матричной модели представления знаний.

При проведении клинической апробации отсутствовало прямое взаимодействие как с пациентами и/или их законными представителями, так и с их персональными данными. Валидация и верификация проводились на основе анонимизированных выписок из электронных медицинских карт, подготовленных сотрудниками медицинских организаций.

Положения, выносимые на защиту:

1. Модель заболевания, содержащая описывающие нозологическую форму признаки и их экспертные оценки, обеспечивает сопоставление нового случая с эталонными описаниями заболеваний.
2. Алгоритм принятия решений, учитывающий избыточность или неполноту данных у пациента, обеспечивает ранжирование выдвигаемых в системе гипотез.
3. «Текстологическая карта», как форма для структурированного представления знаний, извлечённых из литературных источников, обеспечивает единообразное описание клинической картины ЛБН в разные возрастные периоды.
4. Интеллектуальная (экспертная) система поддержки принятия решений для дифференциальной диагностики 30 клинических форм ЛБН у детей обеспечивает эффективную поддержку врача на долабораторном этапе.

Степень достоверности и апробация результатов

Достоверность результатов подтверждена проверкой разработанной системы ГенДиЭС на верифицированных случаях из литературных источников и в процессе валидации и верификации на данных из электронных медицинских карт, полученных из 4 медицинских организаций разных регионов Российской

Федерации. Публикации по результатам этапов исследования были осуществлены в рецензируемых отечественных научных журналах и трудах национальных и международных конференций как по искусственному интеллекту, так и по медицине (педиатрии и генетике). Результаты исследования являются воспроизводимыми, что подтверждается сведениями о верифицированных диагнозах пациентов, полученных из медицинских организаций разного уровня.

Апробация диссертационной работы проведена 23.06.2025 на заседании семинара Отделения 7 ФИЦ ИУ РАН (протокол №6).

Основные материалы диссертации и результаты исследования докладывались на 16 российских и зарубежных научно-практических медицинских и информационно-технологических конференциях.

Внедрение результатов исследования

Результаты диссертационной работы используются в клинической практике врачами-генетиками Медико-генетический центр (МГЦ) ГБУЗ МО МОНИКИ им. М.Ф. Владимирского (г. Москва), Институт педиатрии Университетской клиники (ИП УК) ФГБОУ ВО «ПИМУ» Минздрава России (г. Нижний Новгород).

Результаты диссертационного исследования внедрены в учебный процесс на кафедре медицинской кибернетики и информатики им. С.А. Гаспаряна Института биомедицины (МБФ) ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (Пироговский Университет) и на кафедре медицинской генетики ФГБОУ ДПО РМАНПО Минздрава России.

Публикации по теме диссертации

Результаты диссертационного исследования опубликованы в 16 печатных работах, из которых 2 научные статьи в рецензируемых изданиях из перечня ВАК по специальности 3.3.9 Медицинская информатика (медицинские науки), 1 индексирована в единой библиографической и реферативной базе данных рецензируемой научной литературы Scopus, 3 – в RSCI, 1 – в РИНЦ, 6 научных статей в материалах конференций, 1 тезисы, получено 1 свидетельство на регистрацию базы данных и 1 свидетельство на регистрацию программы для ЭВМ.

Структура и объём диссертации

Диссертация состоит из введения, трёх глав, заключения, списка сокращений и условных обозначений, списка литературы и 5 приложений. Диссертация изложена на 158 страницах машинописного текста, из них 101 страницу составляет основной текст работы, включая 22 рисунка и 16 таблиц. Список литературы включает 81 наименование (31 отечественный и 50 зарубежных источников) на 8 страницах.

СОДЕРЖАНИЕ РАБОТЫ

В первой главе представлен аналитический обзор литературы. В первом разделе дана характеристика ЛБН. Рассмотрены сложности диагностики и важность своевременной постановки диагноза для назначения специального лечения. Во втором разделе представлен обзор систем поддержки принятия решений при наследственных заболеваниях.

Вторая глава содержит описание материалов и методов. Экспертами был определён перечень из 30 клинических форм ЛБН, входящих в дифференциально-диагностический ряд для разработки интеллектуальной системы – 15 мукополисахаридозов (МПС), 7 ганглиозидозов (ГЗ) и 8 муколипидозов (МЛП). На этапе формирования базы знаний проводился анализ 45 медицинских литературных источников по диагностике ЛБН на русском и английском языках и онлайн баз данных OMIM, GARD, HPO и OrphaNet для уточнения характеристик встречаемости признаков при клинических формах ЛБН.

Для комплексной проверки разработанной системы использовались ретроспективные данные 142 пациентов с верифицированными лабораторными или молекулярно-генетическими методами диагнозами ЛБН: 50 клинических случаев из литературы для тестирования и 92 деперсонифицированные выписки из электронных медицинских карт (ЭМК) пациентов для валидации и верификации. Включались пациенты обоих полов, с рождения до 18 лет с наличием одной из 30 исследуемых в рамках данной работы клинических форм, направленные в специализированное учреждение. Не включались пациенты с иными диагнозами основного заболевания (с другим генетическим дефектом). Валидация

осуществлялась на выборке анонимизированных выписок из 54 ЭМК, которые были получены из ОСП НИКИ педиатрии и детской хирургии им. академика Ю. Е. Вельтищева ФГАОУ ВО РНИМУ им. Н. И. Пирогова Минздрава России, г. Москва – 45 МПС и 9 МЛП. Верификация проводилась на деперсонифицированных данных 38 пациентов, полученных из: МГЦ ГБУЗ МО МОНИКИ им. М.Ф. Владимирского, г. Москва – 15 МПС; ГАУЗ СО «КДЦ «ОЗМР», г. Екатеринбург – 16 МПС, 1 МЛП и 3 ГЗ; ИП УК ФГБОУ ВО «ПИМУ» Минздрава России, г. Нижний Новгород – 2 МПС и 1 ГЗ.

Морфологический, синтаксический и семантический методы анализа текстов, как комплексный текстологический подход, использовались при анализе литературных источников с описаниями ЛБН.

В извлечении знаний от двух экспертов, в качестве которых выступали врач-специалисты в области педиатрии и медицинской генетики, с учёными степенями докторов наук, учёными званиями профессор, имеющие непосредственный личный многолетний опыт работы с пациентами с ЛБН, применялся коммуникативный метод.

Присвоение факторов уверенности (меры доверия экспертов) характеристикам признаков осуществлялось с использованием модифицированной стэнфордской модели (Shortliffe E.H., Buchanan B.G., 1984), представленной тремя экспертными оценками (Кобринский Б.А., 2018).

Разработаны порядковые лингвистические шкалы для ранжирования описаний качественных признаков и интервальные шкалы для семантического представления числовых значений.

В проектировании базы знаний применялся матричный метод.

Для программной реализации компьютерной системы использовались: высокоуровневый объектно-ориентированный язык программирования Java, фреймворки Spring Boot и Bootstrap. В качестве СУБД для базы знаний использован PostgreSQL 16.

Метрика точности работы системы рассчитывалась как отношение суммы истинно положительных и истинно отрицательных результатов дифференциальной диагностики к общему количеству случаев.

Для расчета показателей 95% доверительных интервалов применен метод Клоппера-Пирсона с использованием интернет-ресурса MedCalc (доступен по https://www.medcalc.org/calc/diagnostic_test.php).

Третья глава посвящена описанию результатов. Разработана экспертная система ГенДиЭС поддержки принятия решений по дифференциальной диагностике орфанных наследственных заболеваний (на модели ЛБН).

Схематично основные архитектурные компоненты и направления потоков данных представлены на рисунке 1.

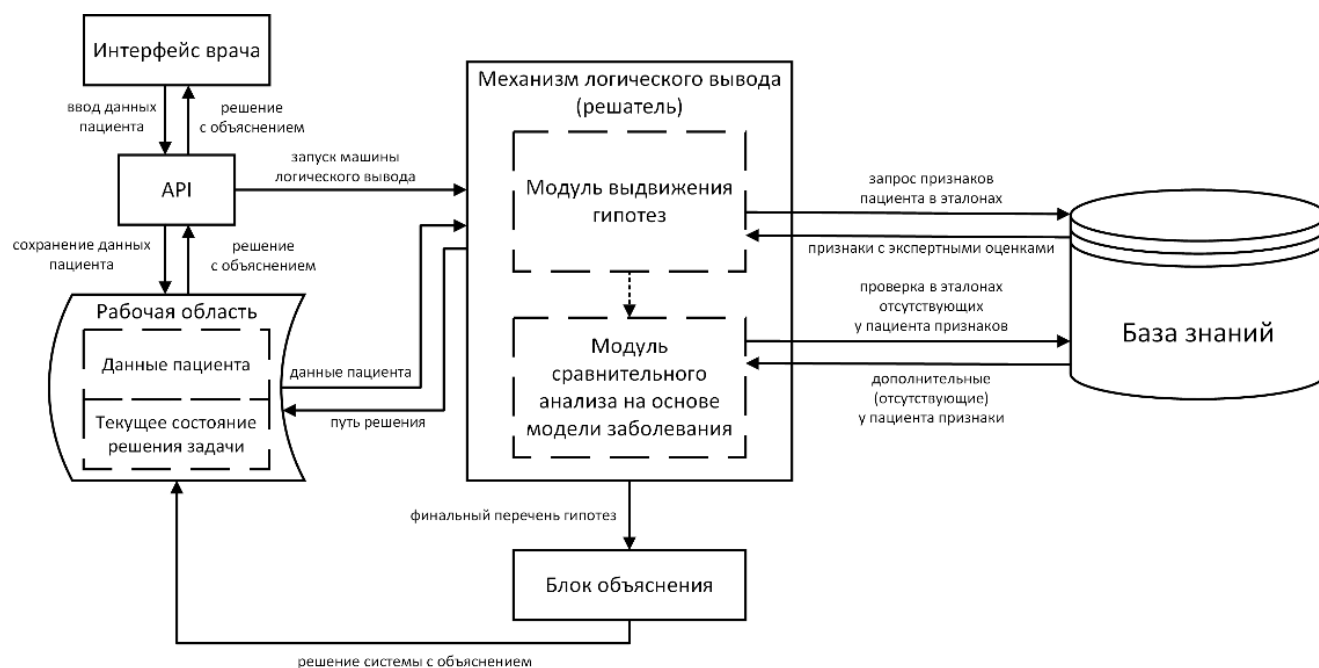


Рисунок 1 – Структурно-функциональная схема системы ГенДиЭС

Система ГенДиЭС состоит из ряда взаимодействующих модулей. Поступление в систему данных нового пациента запускает работу программного модуля API, который выполняет функцию управляющего контроллера, инициируя работу механизма логического вывода (интеллектуального решателя). Введённые врачом данные передаются в рабочую область системы для временного хранения. Из рабочей области данные пациента поступают в решатель, который сопоставляет поступившую информацию с эталонными описаниями базы знаний, в том числе

осуществляет поиск признаков, отсутствовавших у пациента, но указанных в эталонах, для подсказки врачу при дополнительном осмотре больного. Промежуточные решения системы, включая отвергнутые гипотезы, сохраняются в рабочей памяти для возможности отслеживания пути принятия решения. Финальный ранжированный перечень гипотез с признаками передаётся в блок объяснений, где осуществляется интерпретация решения по заданным шаблонам в понятный для врача-пользователя вид. Решение системы с объяснением сохраняется в рабочей области и передаётся на интерфейс врача.

Извлечение знаний о фенотипических проявлениях заболеваний проходило в два этапа. На первом этапе осуществлялся текстологический анализ специальных литературных источников для выделения признаков (симптомов), характеризующих клинические проявления ЛБН в различные возрастные периоды. Разработана форма структурированного представления знаний из литературы – «текстологическая карта», при заполнении которой были использованы специально созданные шкалы. Возрастная шкала включала 4 градации: до 1-го года жизни ребенка, 1–3 года, 4–6 лет, 7 лет и старше. Шкала присутствия (наличие / отсутствие) признаков была представлена в форме “+” и “-”. Ранговая шкала уровней проявления признаков включала 5 лингвистических нечётких оценок выраженности: очень сильно, сильно, умеренно, слабо, очень слабо. Лингвистическая шкала частоты встречаемости признаков имела следующую структуру: очень часто (>80%), часто (60–80%), сравнительно часто (30–60%), редко (15–30%), очень редко (<15%).

По итогам анализа литературных источников для каждой нозологической формы были сформированы текстологические карты. В качестве примера в таблице 1 представлен фрагмент текстологической карты GM1 ГЗ.

Таблица 1 – Фрагмент текстологической карты GM-1 ГЗ

Признак	Возрастная группа	Присутствие	Уровень проявления	Встречаемость	Источник (автор)	Примечания
1	2	3	4	5	6	7
утолщение суставов запястья	2	+	-	-	Jones	Только наличие

Продолжение таблицы 1

1	2	3	4	5	6	7
«вишнёвая косточка» на сетчатке	1	+	-	средне	Harrison, Berman, Campdela creu, Serkov	Без выраженности
	2	+	-	средне		
	3	+	-	редко		

На втором этапе эксперты уточняли и корректировали знания о клинических проявлениях ЛБН, используя текстологические карты. Были выделены 35 наиболее релевантных фенотипических признака. Триединство факторов уверенности экспертов в отношении выраженности признаков, сроков манифестации и частоты встречаемости (Кобринский Б.А., 2018) было модифицировано – частота встречаемости заменена модальностью с целью отражения патогномоничности. Для каждого j -го признака в каждом из возрастных интервалов i -го заболевания экспертами были выставлены: коэффициент модальности (важности для диагностики) и факторы уверенности манифестации и выраженности.

Модальность (mod_{ij}), как оценка диагностической значимости (Грибова В.В. и др., 2019), определяет уровень релевантности признаков на модифицированной стэнфордской шкале $\{-1\} \cup [0;1]$, которая характеризуется следующими градациями: -1 – не может наблюдаться при данной клинической форме; 0 – отсутствует по возрасту или случайный; $(0;0,3]$ – очень редкий; $(0,3;0,6]$ – второстепенный; $(0,6;0,8]$ –необходимый; $(0,8;1]$ – главный.

Манифестация (man_{ij}) оценивалась фактором уверенности экспертов в том, что признак проявляется в данном возрастном периоде. Для этого была использована модифицированная стэнфордская шкала. Если авторами использовалась непрерывная шкала $[0;1]$ (Shortliffe E.H., Buchanan B.G., 1984), то в настоящем исследовании она определяется как $\{-1\} \cup [0;1]$, где: -1 характеризует невозможность проявления данного признака в конкретной возрастной группе по физиологическим причинам; значение в интервале $[0;1]$ характеризует меру доверия экспертов к тому, что признак манифестирует в данном возрастном

периоде, при этом 0 характеризует норму, либо отсутствие симптома в силу особенностей течения заболевания.

Выраженность (sev_{ij}) характеризует уверенность экспертов в уровне проявления данного признака в конкретной возрастной группе. Шкала для оценки выраженности признака была определена в интервале $[0;1]$, где 0 соответствует ситуации, когда признак отсутствует у пациента, а 1 – максимальной выраженности признака.

Экспертами были выставлены 12 600 оценок для 35 признаков 30 клинических форм ЛБН по 4 возрастным периодам. Оценки признаков выставлялись экспертами с учётом отличающихся по тяжести клинических форм, характеризующихся нечёткостью симптомов. Эти оценки были формализованы в виде матрицы «болезни – признаки» (таблица 2), где строки – это симптомы (j), а столбцы – экспертные оценки в 4 возрастных периодах по каждой из клинических форм заболеваний (i).

Таблица 2 – Вид матрицы «болезни – признаки»

Название признака	Название i -го заболевания											
	до 1 года			1-3 года			4-6 лет			7 лет и старше		
	mod	man	sev	mod	man	sev	mod	man	sev	mod	man	sev
Признак 1	mod_{i1}	man_{i1}	sev_{i1}	mod_{i1}	man_{i1}	sev_{i1}	mod_{i1}	man_{i1}	sev_{i1}	mod_{i1}	man_{i1}	sev_{i1}
Признак j	mod_{ij}	man_{ij}	sev_{ij}	mod_{ij}	man_{ij}	sev_{ij}	mod_{ij}	man_{ij}	sev_{ij}	mod_{ij}	man_{ij}	sev_{ij}

Матричный формат (Skillen J.D., 2017) позволяет отображать в формализованном виде комбинации симптомов для рассматриваемого заболевания.

Схема базы данных для базы знаний системы ГенДиЭС представлена на рисунке 2.

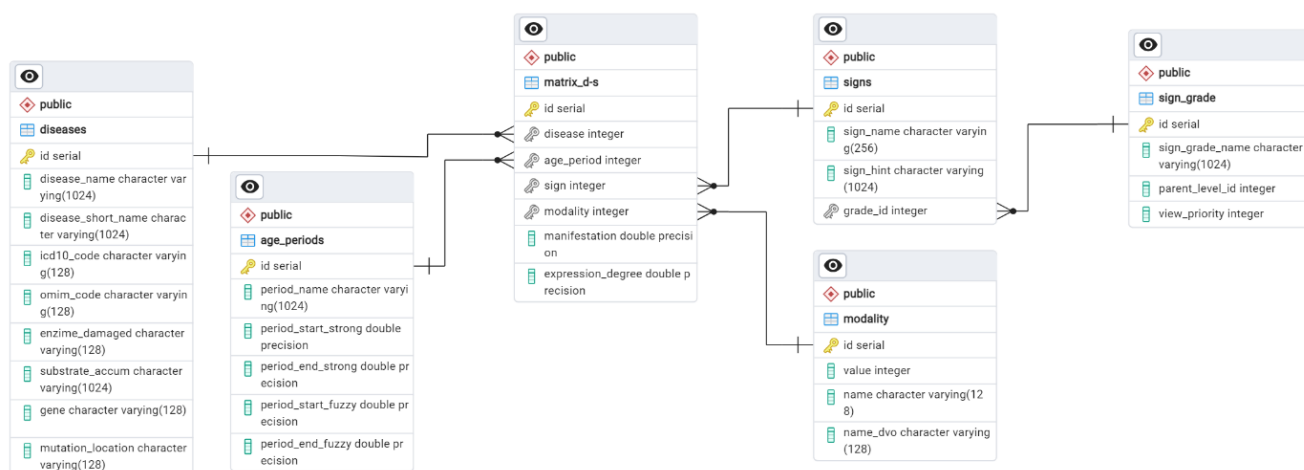


Рисунок 2 – Схема базы данных для базы знаний

Таблица *matrix_d-s* является ключевой таблицей и содержит формализованное представление экспертных знаний в виде матрицы «болезни – признаки», которое включает экспертные оценки для признаков по четырём возрастным периодам для диагностируемых клинических форм. Фрагмент матрицы представлен в таблице 3.

Таблица 3 – Фрагмент таблицы матрица «болезни – признаки» для заболевания с идентификатором «1» (МПС I типа Н) в четвёртой возрастной группе (7 лет и старше)

id	Заболевание	Возрастная группа	Признак	Модалность	Манифестация	Выраженность
	disease	age_period	sign	modality	manifestation	expression_degree
106	1	4	1	0.8	1	1
107	1	4	2	1	0.9	0.9
108	1	4	3	0.4	1	0.8

Как видно из таблицы 3, вид матрицы «болезни – признаки» в базе знаний был преобразован.

Для оценки сходства новых случаев с правилами базы знаний была разработана модель заболевания. Для получения комплексной оценки j -го признака i -го заболевания был введён мультипликативный критерий P_{ij} экспертных оценок, с учётом их равноценности, в определённом возрастном периоде:

$$P_{ij} = (mod_{ij} \cdot man_{ij} \cdot sev_{ij}) \quad (1)$$

Интегрированная аддитивная числовая модель I_i (для каждой возрастной группы) – суммарный показатель, характеризующий клиническую картину i -го заболевания с учетом полученных ранее комплексных оценок P_{ij} :

$$I_i = \sum_{j=1}^n P_{ij}, \quad (2)$$

где n – количество признаков заболевания.

Модель заболевания использовалась в двух вариантах: $\tilde{I}_i$ – общая числовая оценка заболевания, которая включает все описанные экспертами признаки и является эталоном; $\hat{I}_i$ – частная числовая оценка заболевания, рассчитываемая как сумма признаков пациента – конкретного клинического случая.

Дифференциальная диагностика (выдвижение и отбор гипотез, указывающих на возможный диагноз) осуществляется в решателе системы. Для этого используется алгоритм принятия решений, в основу которого положен поэтапный перебор гипотез в пространстве рассматриваемых состояний (рисунок 3).

Решатель включает 2 модуля: выдвижения гипотез о диагнозах на основе сопоставительного анализа по наличию признаков и ранжирования гипотез с использованием модели заболевания.

На I этапе работает модуль выдвижения гипотез, который включает 4 шага:

Шаг I.0. Поступление из рабочей области информации о пациенте.

Шаг I.1. Определение возрастной группы пациента, и сопоставление его признаков с 30 эталонными экспертными описаниями болезней.

Шаг I.2. Формирование трёх групп признаков для пациента, в соответствии с их ролью для гипотезы i -го заболевания в базе знаний:

- Признаки «за» гипотезу – если $mod_{ij} > 0$.
- Признаки «против» гипотезы – если $mod_{ij} = -1$.
- Признаки «вне эталона» – если $mod_{ij} = 0$ (условно избыточные признаки).

Если при сопоставлении данных пациента с эталоном есть хотя бы один признак «против», либо нет ни одного признака «за», то гипотеза отвергается.

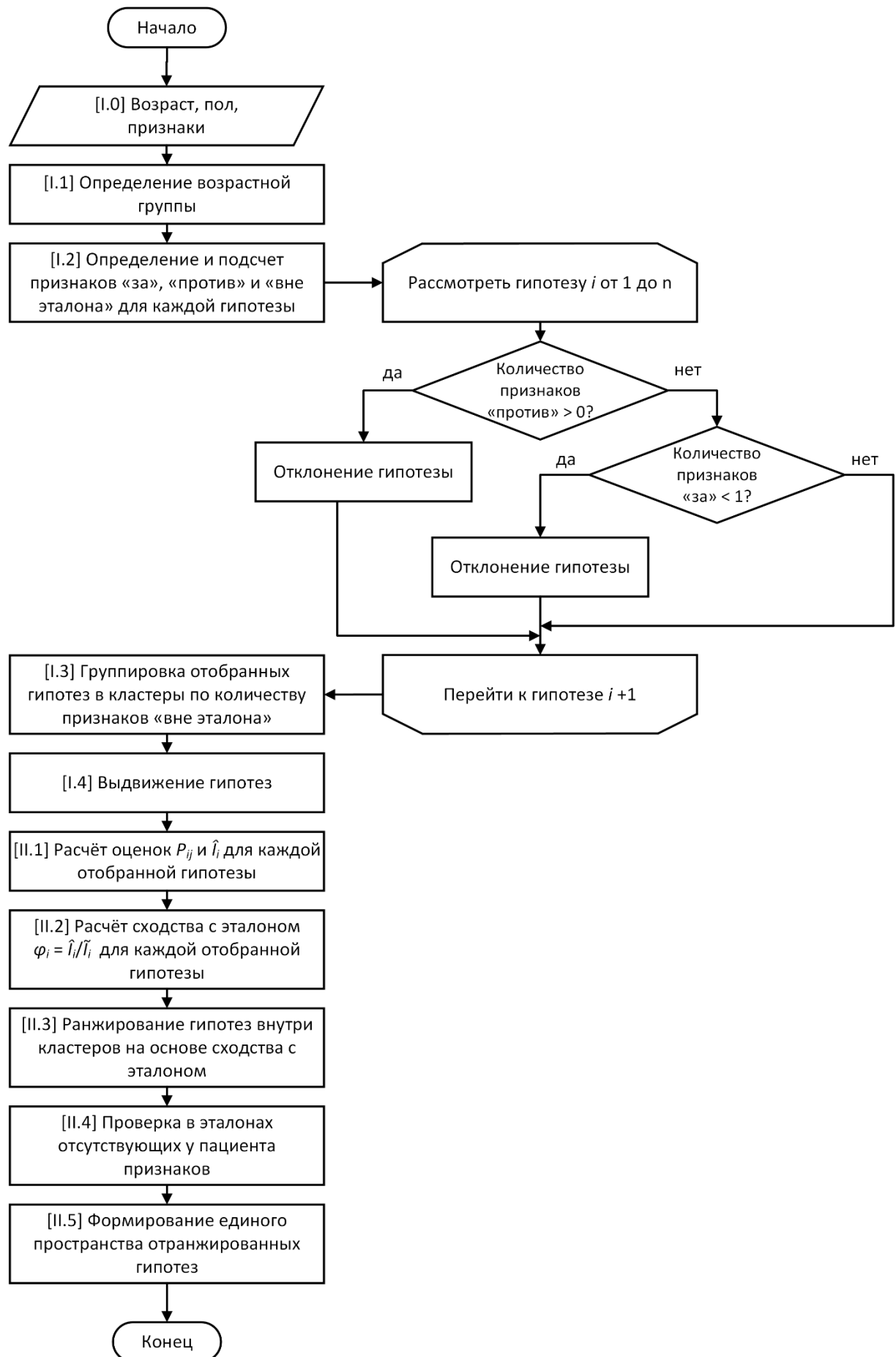


Рисунок 3 – Алгоритм принятия решений в системе ГенДиЭС

Шаг I.3. Объединение выдвинутых гипотез в «кластеры» на основании количества признаков «вне эталона». Вычисляется количество признаков, для которых значения $mod_{ij} = 0$ и гипотеза помещается в соответствующий «кластер» под порядковым номером, равным количеству mod_{ij} .

Упорядочивание «кластеров» осуществляется по принципу – чем меньше признаков «вне эталона», тем более значим «кластер».

Шаг I.4. «Кластеры» гипотез передаются на II этап работы решателя.

На II этапе осуществляется пошаговая работа модуля ранжирования гипотез:

Шаг II.1. Расчёт оценок P_{ij} и $\hat{I}_i$ для каждой отобранной гипотезы.

Шаг II.2. Расчёт степени сходства φ_i нового случая с эталонными вариантами на основе признаков «за» по формуле:

$$\varphi_i = \hat{I}_i / \tilde{I}_i. \quad (3)$$

Оценка сходства по данной формуле позволяет судить насколько интегральная оценка нового случая близка к эталонной оценке заболевания.

Шаг II.3. Ранжирование гипотез внутри каждого «кластера» по уменьшению степени сходства с эталоном φ_i .

Шаг II.4. Запрос к базе знаний о признаках, которые содержатся в эталонных описаниях базы знаний, но не были отмечены врачом у пациента. Присоединение к гипотезам из базы знаний справочной информации о заболеваниях.

Шаг II.5. Формирование единого ранжированного списка гипотез в порядке рангового места «кластеров».

При ранжировании выдвигаемых гипотез алгоритм обрабатывает данные как в ситуации избыточно отмеченных врачом у больного признаков, так и при неполных данных. Финальный ранжированный перечень гипотез (дифференциально-диагностический ряд) передаётся в блок объяснения.

Блок объяснения формирует объяснение для врача на ограниченном естественном языке по заданному шаблону. В качестве объяснения система предоставляет для каждой гипотезы три дополняющие друг друга группы признаков: (1) послужившие основанием для обоснования, (2) присутствовавшие у

пациента, но отсутствующие в эталонном описании, (3) отсутствовавшие у пациента, но являющиеся частью эталонного варианта. В качестве справочной информации осуществляется перевод отобранных полей из таблицы *diseases* в понятный для врача вид, включающий для каждого заболевания: код МКБ-10, код OMIM, дефицитный фермент, накапливаемый субстрат, мутантный ген, локализация гена. При этом врачу-пользователю выводится (по умолчанию) ограниченный, как принято, список из первых 5 гипотез-диагнозов (или весь список по желанию).

При проверке работы системы точность на тестовой выборке (50 больных) составила 0,84 (95% ДИ [0,71;0,93]), на валидационной (54 больных) – 0,87 (95% ДИ [0,75;0,95]), на верификационной (38 больных) – 0,90 (95% ДИ [0,75;0,97]). По результатам анализа ошибок системы ГенДиЭС установлено, что в 18 из 19 неверно распознанных больных, верный диагноз присутствовал в расширенном ряду гипотез, выдаваемых системой. Нахождение верных диагнозов на более низких позициях ранжированного перечня предполагаемых системой диагнозов было обусловлено, в частности, нетипичными проявлениями заболеваний.

Приведённые выше результаты свидетельствуют о воспроизводимости результата точности системы на данных, полученных из различных источников. Показано, что набор правил, включающий три экспертные оценки каждого признака, позволяет с достаточно высокой точностью осуществлять диагностику клинических форм сходных заболеваний. По данным валидации и верификации можно сделать вывод о возможности применения разработанной системы в медицинских организациях.

Заключение

В рамках настоящей работы была разработана и валидирована экспертная система поддержки принятия врачебных решений ГенДиЭС при дифференциальной диагностике редких наследственных заболеваний на долабораторном этапе обследования пациента (на примере ЛБН).

Триединство модифицированных экспертных факторов уверенности позволило достаточно тонко описывать различия сходных дифференцируемых

заболеваний на основе количественной оценки нечётко проявляющихся признаков в клинической картине наследственных ЛБН. Структурированные экспертные знания были представлены в форме матрицы «болезни – признаки», где нозологические формы прогрессирующих заболеваний представлены по четырём возрастным периодам, в каждом из которых указаны экспертные оценки модальности, манифестации и выраженности признаков.

В системе реализована интегральная модель заболевания, являющаяся суммой комплексных оценок признаков в пределах возрастной группы, которая, в свою очередь, является произведением трёх экспертных оценок – модальности, манифестации и выраженности. Общая интегральная модель, рассчитанная на основе всех признаков из базы знаний, является эталонной оценкой заболевания. Частная интегральная модель, рассчитанная только по указанным врачом признакам пациента, представляет собой оценку конкретного случая.

Для выдвижения, подтверждения и ранжирования диагностических гипотез был разработан алгоритм принятия решений, который включает ряд шагов, позволяющих эффективно формировать ограниченный ряд диагностических гипотез. В алгоритме принятия решений системы предусмотрена обработка в условиях не только неполных данных пациентов, но и в случаях так называемой условной избыточности, то есть ситуации, когда врач отмечает признаки, которые отсутствуют в эталонных описаниях.

Критерием успешной работы системы было определено вхождение верифицированного диагноза в ранжированный ряд из 5 гипотез. В аналогичных интеллектуальных системах для поддержки принятия решений в клинической генетике выдаётся от 3 до 10 диагнозов или полный список. В ходе проверки точности системы на разных выборках (тестовая, валидационная, верификационная) была продемонстрирована воспроизводимость результатов, что свидетельствует о возможности применения системы в МГЦ (консультациях) и специализированных стационарах.

Выводы

1. На основе анализа отечественных и зарубежных литературных источников были выделены диагностически значимые признаки для 30 клинических форм ЛБН, представленные в структурированном виде в форме «текстологической карты», в которой указывались: наличие или отсутствие признака, возраст проявления, уровень проявления, частота встречаемости признака и ссылка на источник.

2. Разработаны три экспертные характеристики признаков для четырёх возрастных периодах (с рождения до 1 года, от 1 года до 3 лет, от 4 до 6 лет, 7 лет и старше) и соответствующие им шкалы: для коэффициента модальности $\{-1\} \cup [0;1]$, для манифестации $\{-1\} \cup [0;1]$, для выраженности $[0;1]$.

3. Знания уточнены и структурированы с участием врачей-экспертов в области клинической генетики с использованием разработанных шкал для трёх экспертных оценок признаков. Получено 12 600 экспертных оценок для 35 признаков, отобранных специалистами для диагностики 30 клинических форм ЛБН в рамках выделенных 4 возрастных групп. Признаки с экспертными оценками представлены в виде матрицы «болезни – признаки», на основе которой сформирована база знаний.

4. Построена модель интегрированной оценки клинической картины ЛБН, которая представляет собой сумму комплексных оценок признаков в рамках возрастных периодов. Комплексная оценка признака является как произведением экспертных оценок. Общая интегрированная оценка – это эталонная оценка заболевания – сумма всех комплексных оценок признаков, характеризующих нозологическую форму. Частная интегрированная оценка – сумма комплексных оценок признаков, которые были отмечены врачом у пациента.

5. Разработан алгоритм принятия решений по дифференциальной диагностике, обеспечивающий сравнение клинической картины пациента с эталонными описаниями нозологических форм в различных возрастных периодах в базе знаний. Алгоритм функционирует на основе поэтапного перебора гипотез с последовательным двухэтапным ранжированием по диагностической роли

признаков пациента для рассматриваемого диагноза и по степени сходства с эталонным описанием. Алгоритм позволяет обрабатывать случаи как с неполными данными пациентов, так и с условно избыточными (отмеченными врачом, но отсутствующими в эталонных описаниях). На основе алгоритма был реализован модуль логического вывода (решатель) экспертной системы.

6. Создана экспертная система ГенДиЭС для дифференциальной диагностики наследственных лизосомных болезней на долабораторном этапе. Решением системы является перечень из 5 наиболее вероятных диагнозов, каждый из которых сопровождается объяснением, содержащим признаки: (а) послужившие основанием для обоснования выдвинутой гипотезы, (б) присутствовавшие у пациента, но отсутствующие в эталонном описании нозологической формы, (в) отсутствовавшие у пациента, но являющиеся частью эталонного варианта, с включением дополнительной информации (код МКБ-10, код OMIM, дефицитный фермент, накапливаемый субстрат, мутантный ген, локализация гена). Система реализована в виде *web*-приложения на клиент-серверной архитектуре. Опытная эксплуатация (валидация) системы проводилась на реальных деперсонифицированных данных 54 пациентов (выписках из ЭМК) одной медицинской организации – точность составила 0,87 (95% ДИ [0,75;0,95]). Верификация системы проводилась на 38 случаях из реальной клинической практики (деперсонифицированных выписках из ЭМК) из трёх медицинских организаций различных регионов Российской Федерации – точность составила 0,90 (95% ДИ [0,75;0,97]).

Практические рекомендации

Экспертная система ГенДиЭС может найти применение в МГЦ (центрах) и в отделениях наследственных заболеваний педиатрических медицинских организаций для ассистирования врачам-генетикам в проведении дифференциальной диагностики на долабораторном этапе при подозрении на заболевание из группы ЛБН.

Экспертная система ГенДиЭС может использоваться медицинскими организациями как внешнее *web*-приложение, размещённое в сети интернет, через

разработанный интерфейс системы, либо как сторонний сервис на основе доступа через API. Система ГенДиЭС может быть установлена также на сервере внутри контура медицинской организации с аналогичными вариантами обращения к поддержке принятия решений.

Экспертная система ГенДиЭС рекомендуется к использованию в образовательном процессе для обучающихся по программам специалитета, как пример практической реализации экспертных систем, и в ординатуре или для повышения квалификации как программное средство для решения ситуационных задач в клинической генетике с использованием систем искусственного интеллекта.

Разработанные модель заболевания и алгоритм принятия решений могут быть использованы в интеллектуальных системах для решения задач, требующих классификации объектов.

Список работ, опубликованных по теме диссертации

1. Kobrinskii, B. A. Knowledge Engineering in Construction of Expert Systems on Hereditary Diseases [Text] / B. A. Kobrinskii, N. S. Demikova, N. A. Blagosklonov // Artificial Intelligence. 16th Russian Conference, RCAI 2018, Moscow, Russia, September 24-27, 2018, Proceedings. – 2018. – P. 35-45.
2. Благодосклонов, Н. А. Разработка основанной на знаниях системы диагностики орфанных заболеваний [Текст] / Н. А. Благодосклонов, Б. А. Кобринский // Врач и информационные технологии. – 2019. – № 4. – С. 72-78.
3. Кобринский, Б. А. Система для компьютерной диагностики наследственных заболеваний [Текст] / Б. А. Кобринский, Н. А. Благодосклонов, Н. С. Демикова // Медицинская генетика. – 2020. – Т. 19, № 8(217). – С. 9-11.
4. Blagosklonov, N. A. Model of integral evaluation of expert knowledge for the diagnosis of lysosomal storage diseases [Text] / N. A. Blagosklonov, B. A. Kobrinskii // CEUR Workshop Proceedings. Vol.2648. Proceedings of the Russian Advances in Artificial Intelligence 2020. Moscow, Russia, October 10-16, 2020. – 2020. – P. 250-264.

5. Благосклонов, Н. А. Дифференциальная диагностика наследственных болезней обмена с использованием системы на основе экспертных знаний [Текст] / Н. А. Благосклонов, Б. А. Кобринский // Сибирский журнал клинической и экспериментальной медицины. – 2020. – Т. 35, № 4. – С. 71-78.
6. Благосклонов, Н. А. Экспертная система для диагностики наследственных заболеваний [Текст] / Н. А. Благосклонов, Б. А. Кобринский // Вестник новых медицинских технологий. – 2021. – № 4. – С. 98-102.
7. Blagosklonov, N. A. Implementation of fuzziness of the hereditary diseases clinical picture in an expert diagnostic system [Text] / N. A. Blagosklonov // CEUR Workshop Proceedings. 2021. Vol.2965. Russian Advances in Fuzzy Systems and Soft Computing: Selected Contributions to the 10th International Conference “Integrated Models and Soft Computing in Artificial Intelligence” (IMSC-2021). Kolomna, Russian Federation, May 17–20, 2021. – 2021. – P. 331-335.
8. Компьютерная система для дифференциальной диагностики лизосомных болезней накопления на основе методов искусственного интеллекта [Текст] / Б. А. Кобринский, Н. А. Благосклонов, Н. С. Демикова [и др.] // Бюллетень сибирской медицины. – 2022. – Т. 21, № 2. – С. 67-73.
9. Expert System for the Diagnosis of Orphan Diseases [Text] / B. A. Kobrinskii, N. A. Blagosklonov, V. V. Gribova, E. A. Shalfeeva // Lecture Notes in Networks and Systems: Proceedings of the Sixth International Scientific Conference “Intelligent Information Technologies for Industry” (IITI'22) / eds. S. Kovalev [et al.]. – 2023. – P. 251-260.
10. Благосклонов, Н. А. Интеллектуальная система поддержки дифференциальной диагностики орфанных заболеваний на долабораторном этапе [Текст] / Н. А. Благосклонов, Е. А. Николаева // Российский вестник перинатологии и педиатрии. – 2023. – Т. 68, № 4. – С. 254-255.
11. Blagosklonov, N. A. Architecture of an Expert System to Support Diagnostic Decisions for Hereditary Diseases [Text] / N. A. Blagosklonov, B. A. Kobrinskii // Studies in Computational Intelligence / eds. A. V. Samsonovich, T. Liu. – Cham : Springer Nature Switzerland, 2024. – Vol. 1130 LNCS. – P. 113-120.

12. Благосклонов, Н. А. Принятие диагностических решений экспертной системой в условиях неполноты или избыточности данных [Текст] / Н. А. Благосклонов, Б. А. Кобринский // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов XII Международной научно-практической конференции (ИММВ-2024, Коломна, 14-17 мая 2024 г.). В 2-х т. Т.1. – Смоленск : Универсум, 2024. – С. 55-63.

13. Кобринский, Б. А. Система искусственного интеллекта для диагностики редких заболеваний: принципы построения и результаты клинической апробации [Текст] / Б. А. Кобринский, Н. А. Благосклонов // Сибирский журнал клинической и экспериментальной медицины. – 2025. – Т. 40, № 2. – С. 218-225.

14. Благосклонов, Н. А. Интеллектуальная система для диагностики лизосомных болезней накопления: анализ результатов верификации [Текст] / Н. А. Благосклонов, Ю. Ю. Коталевская // Менеджер здравоохранения. – 2025. – № 9. – С. 68-76.

15. Свидетельство о гос. регистрации базы данных 2023622835 Российская Федерация. Электронная база данных для базы знаний программы «ГенДиЭС». Версия 1.0 [Текст] // Н. А. Благосклонов, Б. А. Кобринский; правообладатель: Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) – № 2023622582; заявл. 08.08.2023; опубл. 17.08.2023.

16. Свидетельство о гос. регистрации программы для ЭВМ 2024689938 Российская Федерация. Экспертная система «ГенДиЭС» для дифференциальной диагностики наследственных лизосомных болезней на долабораторном этапе [Текст] // Н. А. Благосклонов, Б. А. Кобринский, А. А. Николаев, Н. С. Демикова; правообладатель: Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) – № 2024689716; заявл. 05.12.2024; опубл. 11.12.2024.

СПИСОК СОКРАЩЕНИЙ

ГЗ	–	Ганглиозидоз
ДИ	–	Доверительный интервал
ИП УК	–	Институт педиатрии Университетской клиники
ЛБН	–	Лизосомные болезни накопления
МГЦ	–	Медико-генетический центр
МЛП	–	Муколипидоз
МПС	–	Мукополисахаридоз
СУБД	–	Система управления базами данных
ЭВМ	–	Электронная вычислительная машина
ЭМК	–	Электронная медицинская карта
API	–	Application programming interface