

Восточно-Казахстанский государственный технический университет
им Д.Серикбаева

УДК 004.893

На правах рукописи

БЕЛЬГИНОВА САУЛЕ АСКЕРБЕКОВНА

**Разработка информационной технологии поддержки принятия
решений при диагностировании заболеваний крови**

6D070300 – Информационные системы

Диссертация на соискание ученой степени
доктора философии (PhD)

Научные консультанты
д.ф.-м.н., профессор
Калимолдаев М.Н.,
PhD Увалиева И.М.

PhD Рустамов С.С.

Республика Казахстан
Усть-Каменогорск, 2019

СОДЕРЖАНИЕ

СОКРАЩЕНИЯ	4
ВВЕДЕНИЕ.....	5
1 АНАЛИЗ ПРОБЛЕМ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ПРИНЯТИЯ МЕДИЦИНСКИХ РЕШЕНИЙ ПРИ РЕАЛИЗАЦИИ Е- ЗДРАВООХРАНЕНИЯ.....	10
1.1 Системы поддержки принятия решений в медицине.....	10
1.2 Анализ реализации Концепции развития е-здравоохранения Республики Казахстан.....	14
1.3 Обзор существующих методов и моделей поддержки принятия медицинских решений	21
1.4 Существующие системы диагностирования заболеваний крови.....	26
1.5 Выводы по первому разделу	30
2 МЕТОДЫ И МОДЕЛИ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ	31
2.1 Постановка задачи и разработка функциональной модели диагностирования клинико-гематологических синдромов.....	31
2.2 Концептуальная модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови.....	36
2.3 Разработка комплекса методов вычислительной ИТ-инфраструктуры технологии диагностирования заболеваний крови.....	44
2.4 Выводы по второму разделу	61
3 АРХИТЕКТУРА И РЕАЛИЗАЦИЯ ПРОГРАММНОГО КОМПЛЕКСА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ	63
3.1 Общее описание и обоснование программного инструментария системы поддержки принятия решений.....	63
3.2 Проектирование архитектуры системы поддержки принятия решений	64
3.3 Модель классов системы поддержки принятия решений.....	69
3.4 Проектирование базы данных системы поддержки принятия решений.....	73
3.5 Выводы по третьему разделу	76
4 ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ	77
4.1 Описание экспериментального исследования	77
4.2 Реализация модели нечеткого логического вывода	82
4.3 Программная реализация вычислительного алгоритма морфологической классификации клинико-гематологических синдромов	92
4.4 Реализация алгоритма Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов	95

4.5 Анализ эффективности применения предложенной информационной технологии в задачах диагностирования заболеваний крови.....	97
4.6 Выводы по четвертому разделу	102
ЗАКЛЮЧЕНИЕ	103
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	104
ПРИЛОЖЕНИЕ А.....	113
ПРИЛОЖЕНИЕ Б	114
ПРИЛОЖЕНИЕ В.....	115
ПРИЛОЖЕНИЕ Г	116
ПРИЛОЖЕНИЕ Д.....	117
ПРИЛОЖЕНИЕ Е	118
ПРИЛОЖЕНИЕ Ж.....	119
ПРИЛОЖЕНИЕ И	120
ПРИЛОЖЕНИЕ К.....	124
ПРИЛОЖЕНИЕ Л.....	125

СОКРАЩЕНИЯ

Fer – ферритин - депо-белок железа.

FS - содержание ферритина в сыворотке крови.

HCT – показатель биохимического анализа крови, характеризующий гематокрит.

HGB – показатель биохимического анализа крови, характеризующий уровень гемоглобина.

MCH - показатель биохимического анализа крови, характеризующий средний содержания гемоглобина в эритроците.

MCHC - показатель биохимического анализа крови, характеризующий средней концентрации гемоглобина в эритроците.

MCV - показатель биохимического анализа крови, характеризующий средний объема эритроцита.

RBC - показатель биохимического анализа крови, характеризующий количества эритроцитов.

RTC - количество ретикулоцитов в крови.

SFe - сывороточное железо.

TFCS - общая железосвязывающая способность сыворотки крови.

АХЗ - анемия хронических заболеваний.

ЖДА - железодефицитная анемия, т.е. гематологический синдром, характеризующийся нарушением синтеза гемоглобина вследствие дефицита железа.

НПП - нечетко-продукционные правила.

ЛП – лингвистическая переменная.

КПВ - книга протоколов вывода.

ППРДЗК - поддержка принятия решений при диагностировании заболеваний крови.

ВВЕДЕНИЕ

Актуальность работы. В настоящее время одним из направлений развития информационных технологий является разработка и применение информационных медицинских систем, их интеграция в единое информационное пространство. Для решения этих задач требуется развитие систем поддержки принятия решений, способствующим повышению эффективности лечебно-диагностических бизнес-процессов, которые должны найти свое место в существующих интегрированных системах. Одновременно не лишне вспомнить, что именно с математического моделирования физиологических процессов и диагностических вычислительных систем начиналась медицинская кибернетика, и они уже тогда продемонстрировали свою эффективность и целесообразность использования в практическом здравоохранении [1].

Одним из важнейших стратегических направлений развития системы здравоохранения Республики Казахстан является создание и развитие технологической инфраструктуры единого информационного пространства здравоохранения. К сегодняшнему дню в казахстанском здравоохранении наблюдается переход к автоматизации обработки медицинской информации и ведения документооборота [2]. Это предполагает повышение скорости обработки, тем самым улучшение качества обслуживания пациентов, облегчение работы врачебного и медицинского персонала.

В целях реализации электронного здравоохранения в Казахстане предполагается разработка и внедрение стандартов, которые дают возможность реализации сервис ориентированной архитектуры. Она в свою очередь обеспечит полную интероперабельность между информационными системами, участвующими в поддержке процессов здравоохранения [3]. Т.е. данная архитектура позволяет системе взаимодействовать и функционировать с другими продуктами или системами без каких-либо ограничений доступа и реализации.

Сегодня в стране активно внедряется информационно-технологическая платформа, так называемая единая информационная система здравоохранения, основной целью которой является создание информационной структуры здравоохранения Республики Казахстан, соответствующей уровню экономического, социального, технического, технологического развития общества и обеспечивающей рациональное использование ресурсов здравоохранения при более качественном предоставлении медицинских услуг населению. Создание единой информационной системы здравоохранения предполагает выполнение ряда задач, основанных на разработке и внедрении единых стандартов обмена медицинскими данными, применении единой системы идентификации объектов учета и субъектов информационного взаимодействия в здравоохранении, а также централизованного ведения и открытого доступа к базе данных единых классификаторов, справочников и нормативов, в том числе к базе данных стандартов медицинской помощи,

протоколов ведения больных, государственным реестрам лекарственных средств и изделий медицинского назначения [4].

Диссертационная работа посвящена решению актуальной задачи совершенствования системы электронного здравоохранения, которая является одной из направлений государственной программы «Цифровой Казахстан» [5]. Так как здоровье нации является одним из приоритетных направлений государственной политики Республики Казахстан, на сегодняшний момент очень важно развивать электронное здравоохранение, внедрять информационно-коммуникационные технологии в систему здравоохранения, позволяющие вывести на новый уровень качество оказания медицинских услуг, что и определяют актуальность исследования.

Объектом исследования является система показателей клинико-гематологических синдромов.

Предмет исследования является информационная технология поддержки принятия решений при диагностировании клинико-гематологических синдромов.

Цель работы - разработать информационную технологию поддержки принятия решений при диагностировании заболеваний крови, позволяющая повысить эффективность бизнес-процессов мониторинга клинико-гематологических синдромов.

Задачи исследования. Для достижения поставленной цели диссертационного исследования требуется решение следующих задач:

- изучить существующие системы поддержки принятия решений в медицине;
- провести аналитический обзор существующих методов и моделей поддержки принятия медицинских решений и практических ИТ-решений диагностирования клинико-гематологических синдромов;
- разработать концептуальную модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови;
- разработать вычислительную ИТ-инфраструктуру технологии диагностирования заболеваний крови, объединяющую интеллектуальные методы обработки показателей клинико-гематологических синдромов;
- разработать архитектурный проект программного комплекса поддержки принятия решений при диагностировании заболеваний крови;
- описать результаты реализации модели нечеткого логического вывода и вычислительного алгоритма морфологической классификации клинико-гематологических синдромов;
- проанализировать эффективность применения предложенной информационной технологии в задачах диагностирования заболеваний крови.

Научная новизна диссертационного исследования заключается в том, что впервые для повышения эффективности процессов диагностирования заболеваний крови предложена модель ИТ-инфраструктуры клинико-гематологических синдромов, интегрирующая информационную, вычислительную и коммуникационную инфраструктуру.

Научные положения и результаты, выносимые на защиту. В диссертационной работе предложена информационная технология поддержки принятия решений при диагностировании заболеваний крови, основанная на ИТ-инфраструктуре диагностирования клинико-гематологических синдромов, интегрирующая вычислительный алгоритм морфологической классификации, метод оценки информативности диагностических показателей и гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов. В диссертации разработана архитектура и реализация модулей программного комплекса поддержки принятия решений при диагностировании заболеваний крови.

Основные научные положения, выносимые на защиту:

- концептуальная модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови;
- модель интегрированной ИТ-инфраструктуры клинико-гематологических синдромов, объединяющая информационную, вычислительную и коммуникационную инфраструктуру;
- архитектура программного комплекса диагностирования клинико-гематологических синдромов, основанная на интеграции трехуровневой, объектно-ориентированной и компонентной архитектурных шаблонов.

Практическая ценность результатов диссертационного исследования заключается в применимости предложенных в диссертации концептуальной модели информационной технологии поддержки принятия решений и ИТ-инфраструктуры диагностирования заболеваний крови, а также архитектурного проекта системы, для решения задач информационно-аналитической поддержки принятия управленческих решений при сопровождении бизнес-процессов мониторинга клинико-гематологических синдромов.

Методы исследования. В работе используется аппарат теории управления системами, теории статистического анализа, теории искусственного интеллекта, теории экономико-математических методов и моделей, теории проектирования информационных систем.

Реализация результатов работы. Основные научные и практические результаты диссертационной работы, представленные в виде программного модуля диагностирования клинико-гематологических синдромов приняты для внедрения на базе ГКП на ПХВ «Городской ревматологический центр г. Алматы» (Приложение А) и программный модуль «Оценка информативности диагностических показателей клинико-гематологических синдромов» внедрен в АО «Казахский институт онкологии и радиологии» г. Алматы (Приложение Б). Получено свидетельство о государственной регистрации прав на объект авторского права под названием «Программный модуль диагностирования клинико-гематологических синдромов (программа для ЭВМ)» за №4737 от «1» августа 2019 года (Приложение В).

Апробация работы. Основные результаты диссертационной работы доложены и обсуждены на следующих международных конференциях: «Computer science and Applied Mathematics» (Алматы, 2016) II International

scientific conference «Computer science and Applied Mathematics»» (Алматы, 2017) [6]; 4th International Conference on Computer and Technology Applications (Стамбул, 2018) [7]; 4th International Conference on Engineering and MIS2018 (Стамбул, 2018) [8]; «Communications in Computer and Information Science» (Усть-Каменогорск, 2018) [9]; «Application of Information and Communication Technologies - AICT2018» (Алматы, 2018) [10]; 5th International Conference on Engineering and MIS 2019 (Астана, 2019) [11]; 12th IEEE International Conference «Application of Information and Communication Technologies - AICT2019 (Баку, 2019) [12].

Публикации. По теме диссертационной работы опубликовано 26 научных работ, в том числе 1 статья в журнале имеющих в базе данных Scopus показатель процентиля по CiteScore равный 33, еще 7 статей были опубликованы в изданиях, индексируемые в базах данных Scopus и Web of Science, 8 работ в изданиях рекомендуемых Комитетом по контролю в сфере образования и науки МОН РК, 16 работ в сборниках международных конференций, в том числе 2 статьи в изданиях дальнего зарубежья.

Публикация «Decision support system for diagnosing anemia» и «Informational and analytical system to diagnose anemia», опубликованных в научных изданиях Scopus (и Web of Science) были процитированы 4 и 3 раз соответственно. Индекс Хирша автора по базе данных Scopus равен 2, по Web of Science – 1 (Приложение Г).

Структура и объем диссертации. Диссертация состоит из введения, 4 разделов, заключения, списка использованных источников из 113 наименований, изложенных на 126 страницах компьютерного текста, включает 57 рисунков, 26 таблиц и 10 приложений.

В первом разделе диссертации представлены результаты анализа текущего положения, предпосылки и основные направления развития электронного здравоохранения в Казахстане. Обоснована актуальность проектирования систем поддержки принятия решений для диагностирования различных заболеваний, повышающие эффективность деятельности медицинских работников путем внедрения различных методов анализа данных и медицинских данных. Проведен обзор существующих методов и моделей поддержки принятия медицинских решений, а также изучены архитектурные модели и модели управление информационными системами здравоохранения.

Во втором разделе диссертации описана разработка концептуальной модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови, основанная на модуле оперативной памяти, базы знаний, модуль интерфейса пользователя, процедуры логического вывода, а также модуле наполнения и проверки базы знаний. Описывается разработка ИТ-инфраструктура диагностирования клиничко-гематологических синдромов, которая включает вычислительный алгоритм морфологической классификации, метод оценки информативности диагностических показателей и гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клиничко-гематологических синдромов.

В третьем разделе диссертации представлена разработка архитектуры программного комплекса диагностирования клинико-гематологических синдромов, основанная на интеграции трехуровневой, объектно-ориентированной и компонентной архитектурных шаблонов. Представлено описание и обоснование программного инструментария системы поддержки принятия решений. Разработана и описана модель классов системы, а также была спроектирована база данных диагностирования клинико-гематологических синдромов для поддержки принятия решений. Были выявлены основные требования, которые должны предъявляться к реализуемой системе. Проведен анализ и сделан выбор инструментальных средств для разработки системы диагностирования клинико-гематологических синдромов.

В четвертом разделе диссертации описаны результаты экспериментального исследования предложенной информационной технологии поддержки принятия решений, а также была произведена оценка эффективности применения предложенной информационной технологии в задачах диагностирования заболеваний крови. Представлена программная реализация вычисленного алгоритма диагностирования клинико-гематологических синдромов, основанная на алгоритме диагностики анемий и позволяющая классифицировать анемии по морфологическим признакам и дать количественную характеристику выраженности патологического процесса.

1 АНАЛИЗ ПРОБЛЕМ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ПРИНЯТИЯ МЕДИЦИНСКИХ РЕШЕНИЙ ПРИ РЕАЛИЗАЦИИ Е-ЗДРАВООХРАНЕНИЯ

1.1 Системы поддержки принятия решений в медицине

1.1.1 Понятие медицинской информации

Основным понятием медицинских систем является понятие медицинской информации. Однозначного определения понятия информации не существует, так как оно является одним из базовых общенаучных терминов. Поэтому для каждой науки этот термин должен определяться специфически, путем подачи различных примеров касательно той области, в котором он применим.

В современной медицинской литературе, так же как и в других научных направлениях, существует множество определений понятия «информация». Если за основу толкования термина «информация» брать медицинское направление, то словосочетание «медицинская информация» в широком смысле можно рассматривать как любую информацию, которая относится к медицинской отрасли. А в узком смысле – это информация, относящаяся непосредственно к человеку как к пациенту, то есть информация о его здоровье, особенностях организма, перенесенных заболеваниях и т.д. [13]. Наиболее расширенное толкование данному термину дает другой автор: «медицинская информация - это совокупность данных о пациентах и заболеваниях, образующаяся при их взаимодействии с адекватными им методами и снимающая неопределенность и неполноту предварительных знаний» [14].

Любая информация, в том числе и медицинская, имеет свойства, созвучные с информационными свойствами в целом, но отражающие специфику, присущую только ей.

Медицинская информация, по характеру объективности разделяется на объективную и субъективную. Информация, полученная с помощью различных медицинских оборудований, представляет собой объективную информацию. В качестве такой информации можно рассматривать данные магнитно-резонансной томографии, рентгенографии, электрокардиограммы, электрокардиографии, сфигмоманометров, офтальмоскопов, отоскопов, лабораторные данные биохимического анализа крови, и т.п. К субъективным данным относятся данные, которые получают, не прибегая к каким-либо специальным электронным средствам, то есть информация, полученная путем визуального осмотра доктором пациента, пальпации, перкуссии или аускультации, а также данные анамнеза. В связи с тем, что сбор такого рода информации осуществляется врачом, влияние человеческого фактора ведет к росту субъективного характера данных.

Объективность медицинской информации зависит от того, в какой степени эффективно сочетаются диагностические и терапевтические методы представления данных, которые в совокупности определяют «золотой стандарт» в медицине.

Если говорить о достоверности медицинских данных, прежде всего, учитывается качество полученной медицинской информации, например, сведения, полученные в результате использования диагностических медицинских оборудований или лабораторными анализами. Присутствие разных шумов в биологических сигналах, выход итогов лабораторных анализов за пределы нормальных значений, присутствие нестандартных ошибок в анализах — это признаки недостоверности медицинских данных. Обычно, качество клинического исследования обусловлено степенью случайных и систематических погрешностей их результатов.

Когда речь идет о доступности медицинских данных, то они определяются уровнем возможности получения медицинскими работниками или пациентами, какой-либо информации медицинского характера. Такая информация обладает ограничениями в правах доступа, касательно степени секретности всевозможных персональных сведений пациентов.

Согласно Статье 27 «Принципы информатизации в области здравоохранения» Кодекса Республики Казахстан «О здоровье народа и системе здравоохранения» одним из основных принципов является обеспечение сохранности и конфиденциальности электронных информационных ресурсов, содержащих персональные данные физических лиц (пациентов), и доступа пациента к своим персональным данным. (Статья 28, пункты 5, 6) [15].

Актуальность медицинской информации подразумевает соответствия данных к текущему времени или к текущему состоянию здоровья пациента. Только актуальные данные могут обеспечить однозначную интерпретацию заболевания. Для этого необходимо постоянное обновление медицинских данных.

1.1.2 Характеристики медицинских систем поддержки принятия решений

На сегодняшний день одним из актуальных задач является разработка и применение информационных медицинских систем и их интеграция в единое информационное пространство. Существуют различные виды системы поддержки принятия решений медицинского назначения, относительно их применения, которые представлена на рисунке 1.1.

Медицинские СППР в основном направлены для решения следующих задач:

- дифференциальное диагностирование;
- прогнозирование хода заболевания и определение клинической картины;
- наблюдение и регистрация текущего состояния пациента;
- оценка степени выраженности синдрома болезни в режиме «реального времени».

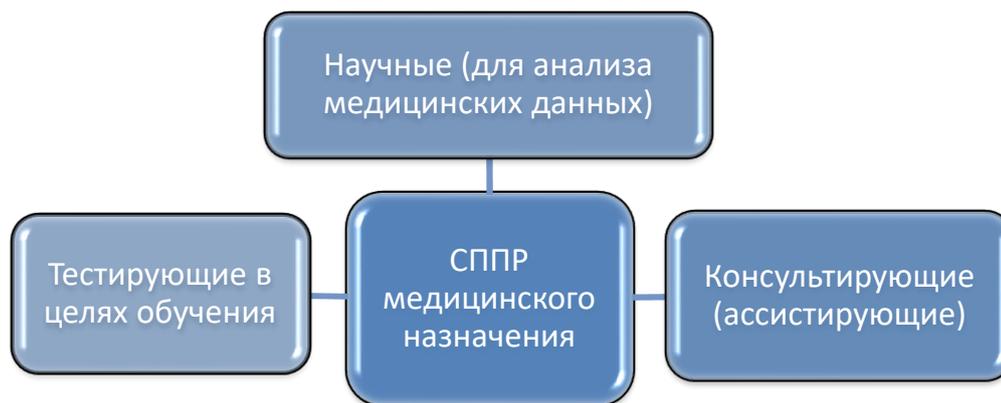


Рисунок 1.1 - Виды медицинских систем поддержки принятия решений

Современные системы поддержки принятия решений зачастую основываются на знаниях экспертов и носят интеллектуальный характер. Поэтому в процессе их формирования необходимо точно определить способ представления медицинских данных для дальнейшего их использования (Рисунок 1.2).

Способы при постановке задачи и представлении данных для построения СППР может характеризоваться следующим образом:

- применение шаблонов заболеваний, которые представляют собой логические выражения, состоящие из теоретически допустимых клинических состояний, определяющих конкретное заболевание;
- формирование наборов симптомов заболевания, которые характеризуются интервалом принадлежности количественных или качественных значений признаков, определяющих болезнь;
- формирование исключаящего набора, который указывает на слабую вероятность определенного заболевания, который также может характеризоваться как набор взаимоисключающих симптомов;
- определение неоднозначных данных, которые характеризуются трудностью однозначной интерпретации клинических проявлений, и как следствие, предполагают применение методов нечеткой логики;
- формирование причинно-следственных связей, которые проявляются в виде определения сведений о патологии, ставшей причиной указанного заболевания либо, напротив, выступает в качестве его следствия;
- формирование ассоциативной связи, дающая возможность учитывать, каким состоянием было вызвано определенное заболевание, или какое заболевание может вызвать такое состояние, какие критические состояния могут проявляться одновременно;
- определение вероятности возникновения заболевания, посредством оценки степени уверенности среди рассматриваемых гипотез.

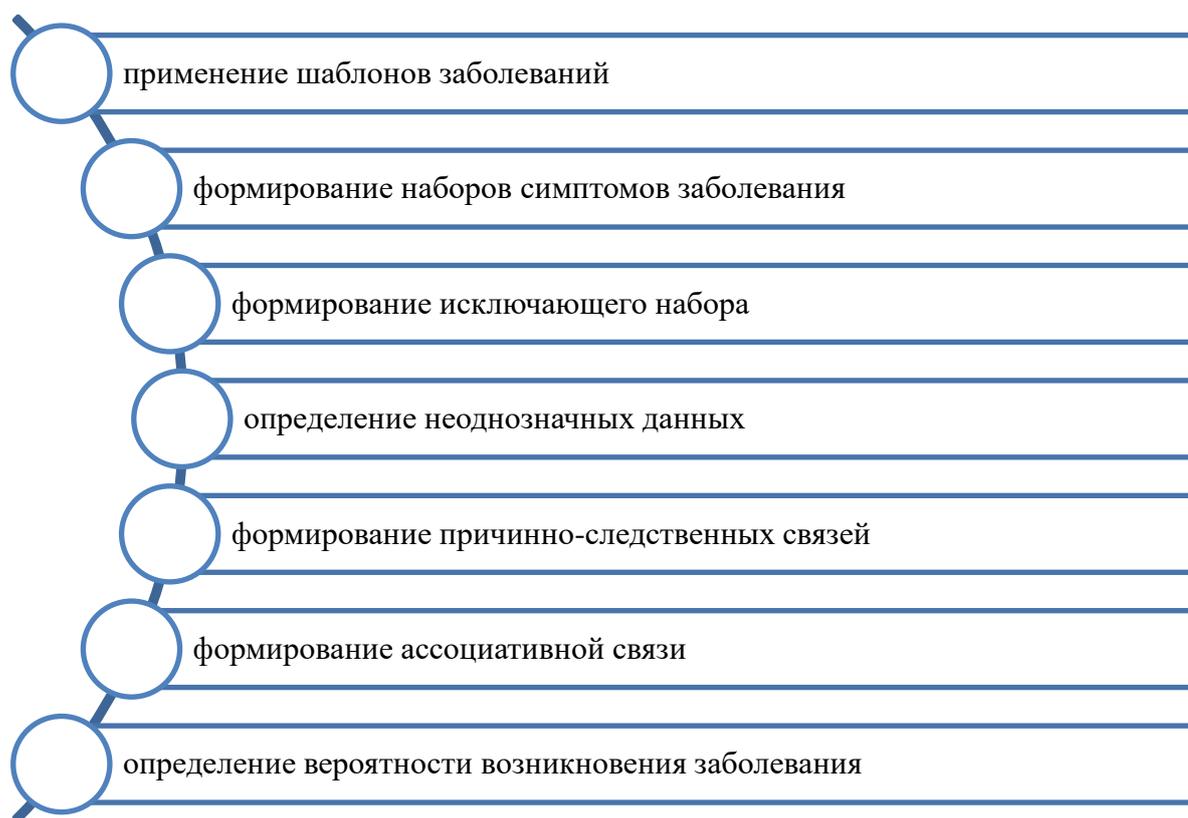


Рисунок 1.2. Способы представления данных при построении медицинских СППР

Практическая значимость систем поддержки принятия решений позволяет решить ряд проблем, которые связаны со следующими факторами:

- нехватка времени для принятия оперативного клинического решения при неотложных состояниях;
- недостоверность данных о клинических состояниях в условиях деятельности врачей скорой помощи;
- необходимость оперативного выбора дополнительных медицинских исследований в условиях угрозы жизни пациента.

Разработанные к настоящему времени системы поддержки принятия медицинских решений, в какой-то мере соответствуют вышеназванным характеристикам. Тем не менее, все еще существует необходимость разработок СППР, которые призваны комплексно решать проблемы с точки зрения клинической медицины.

1.2 Анализ реализации Концепции развития е-здравоохранения Республики Казахстан

1.2.1 Предпосылки развития электронного здравоохранения

В современном мире одним из основных факторов, определяющих динамику социально-экономического развития любого государства, в том числе здравоохранения, является информатизация общества и активное внедрение информационно-коммуникационных технологии во все отрасли жизнедеятельности человека. Главная цель информатизации здравоохранения в целом может быть сформулирована следующим образом: создание новых информационных технологий на всех уровнях управления здравоохранением и новых медицинских компьютерных технологий, повышающих качество лечебно-профилактической помощи и способствующих реализации основной функции охраны здоровья населения – увеличению продолжительности активной жизни [5].

Для того чтобы службы здравоохранения могли содействовать интеграции медицины и общественного здравоохранения, в них должны присутствовать три основные характеристики: ориентация на население и определенный географический район в контексте децентрализованной системы здравоохранения; попытка разработать организационные модели для поддержки процессов координации и интеграции; и правильное использование системы медико-санитарной информации.

Как мы знаем, в медицинской отрасли производится и накапливается очень большой объем медицинских и статистических данных. Все эти данные, в какой-то степени используется врачами, медицинскими работниками, а так же управляющими органами в целях правильной организации и улучшения качества медицинской помощи и поднятия общего уровня жизни населения.

Чтобы быть актуальным, информационная система здравоохранения должна вписываться в организацию системы здравоохранения, для которой она генерирует информацию. Основываясь на четко определенных функциях управления, определить информацию, необходимую для принятия соответствующих решений на каждом уровне управления, относительно легко.

Следующий вопрос - как получить эту информацию наиболее эффективным и действенным способом. Чтобы ответить на этот вопрос, важно понять структуру информационной системы здравоохранения.

Информационная система здравоохранения, как и любая система, имеет организованный набор взаимосвязанных компонентов, которые можно сгруппировать по двум объектам: информационный процесс и структура управления. Посредством информационного процесса необработанные данные (входные данные) преобразуются в информацию в «удобной» форме для принятия управленческих решений (выходных данных). Информационный процесс может быть разбит на следующие компоненты: (1) сбор данных, (2) передача данных, (3) обработка данных, (4) анализ данных и (5) представление информации для использования в сфере ухода за больными и здравоохранения

решения по управлению услугами.

От качественного использования медицинской информации в полной мере зависит не только здоровье населения, но и растет уровень развитие страны в целом. Таким образом, актуальной является использование постоянно растущих больших объемов медицинских данных при решении диагностических, терапевтических, статистических и управленческих задач для развития электронного здравоохранения в Казахстане.

Предпосылками развития электронного здравоохранения в Республике Казахстан являются:

- Кодекс РК «О здоровье народа и системе здравоохранения»;
- Государственная программа «Информационный Казахстан – 2020», утвержденная в январе 2013 г.;
- Концепция развития электронного здравоохранения РК на 2013-2020 г., утвержденная приказом в сентябре 2013г. и дорожная карта;
- Государственная программа развития здравоохранения РК "Денсаулық" на 2016 - 2019 годы, утвержденная в январе 2016г.;
- Закон Республики Казахстан от 24 ноября 2015 года № 418-V «Об информатизации».

Проект Всемирного банка и Правительства РК «Передача технологий и проведение институциональной реформы в секторе здравоохранения Республики Казахстан.

Список официально утвержденных нормативно-правовых актов, регулирующих развитие сферы электронного здравоохранения в Республике Казахстан; ряд важных медицинских автоматизированных систем, интегрированных в единую информационную систему здравоохранения изучена в работе [4]. Здесь проведен подробный обзор действующих автоматизированных информационных систем в медицинских организациях страны, при котором использованы методы системного и структурно-логического анализа, проведен SWOT анализ.

1.2.2 Основные направления развитие электронного здравоохранения в Казахстане

Одним из важнейших стратегических направлений развития системы здравоохранения является организация единого информационного пространства и его технологической инфраструктуры.

К сегодняшнему дню в казахстанском здравоохранении наблюдается переход к автоматизации обработки медицинской информации и ведения документооборота. Это предполагает повышение скорости обработки, тем самым улучшение качества обслуживания пациентов, облегчение работы врачебного и медицинского персонала.

В целях реализации электронного здравоохранения в Казахстане предполагается разработка и внедрение стандартов, которые дают возможность реализации сервис ориентированной архитектуры. Она в свою очередь обеспечит полную интероперабельность между информационными системами,

участвующими в поддержке процессов здравоохранения. Т.е. данная архитектура позволяет системе взаимодействовать и функционировать с другими продуктами или системами без каких-либо ограничений доступа и реализации.

Сегодня в стране активно внедряется информационно-технологическая платформа, так называемая Единая Информационная Система Здравоохранения (ЕИСЗ), основной целью которой является создание информационной структуры здравоохранения Республики Казахстан, соответствующей уровню экономического, социального, технического, технологического развития общества и обеспечивающей рациональное использование ресурсов здравоохранения при более качественном предоставлении медицинских услуг населению.

Создание ЕИСЗ предполагает выполнение ряда задач, основанных на разработке и внедрении единых стандартов обмена медицинскими данными, применении единой системы идентификации объектов учета и субъектов информационного взаимодействия в здравоохранении; а также централизованного ведения и открытого доступа к базе данных единых классификаторов, справочников и нормативов, в том числе к базе данных стандартов медицинской помощи, протоколов ведения больных, государственным реестрам лекарственных средств и изделий медицинского назначения [4].

В рамках Государственной программы развития здравоохранения Республики Казахстан «Денсаулық» на 2016-2020 годы продолжается развитие ЕИСЗ в части разработки единых стандартов, технологических спецификаций и характеристик различных информационных систем по требуемому функционалу [5]. Электронный паспорт здоровья станет центральным звеном, обеспечивающим взаимодействие медицинских информационных систем и поставщиков медицинских услуг за счет реализации стандартизированной модели медицинской информации. При этом будут предусмотрены механизмы информационной безопасности и защиты персональных и конфиденциальных данных.

Большинство медицинских учреждений в настоящее время развивают среду, используя медицинские информационные технологии, за счет внедрения интегрированного доступа к клинической информации. Здесь информационные технологии и их инструменты помогают по административным и финансовым вопросам, в научных исследованиях, в автоматизации делопроизводства, а так же помогает пациентам.

1.2.3 Информационные системы здравоохранения: архитектурные модели и управление

С функциональной точки зрения, электронное здравоохранение Казахстана поддерживает три основных уровня системы здравоохранения:

Централизованное управление на национальном и региональном уровнях: сюда входят возможности централизованного планирования, управление ресурсами, определение правил и процедур, которым надлежит следовать, общий контроль над финансовой деятельностью, контроль над качеством и безопасностью.

Первичная медико-санитарная помощь: этот уровень включает все системы, которые поддерживают услуги, предоставляемые гражданам на всей национальной или региональной территории. Она включает в себя все поставщики услуг, такие как врачи общей практики, местные обычаи и т. д.

Вторичное медицинское обслуживание: этот уровень относится главным образом к системам, которые поддерживают процессы здравоохранения среди поставщиков медицинских услуг.

Эти три уровня обычно взаимосвязаны только в отношении административных и учетных потоков, но потенциальный обмен данными между различными слоями делает ИКТ важными как для обмена, так и для обработки больших наборов клинических данных.

Функционал и архитектура информационных систем е-здравоохранения, а так же потребность в стандартизации определяется информационной моделью ЭПЗ (Электронный Паспорт Здоровья) (рисунок 1.3) [3].

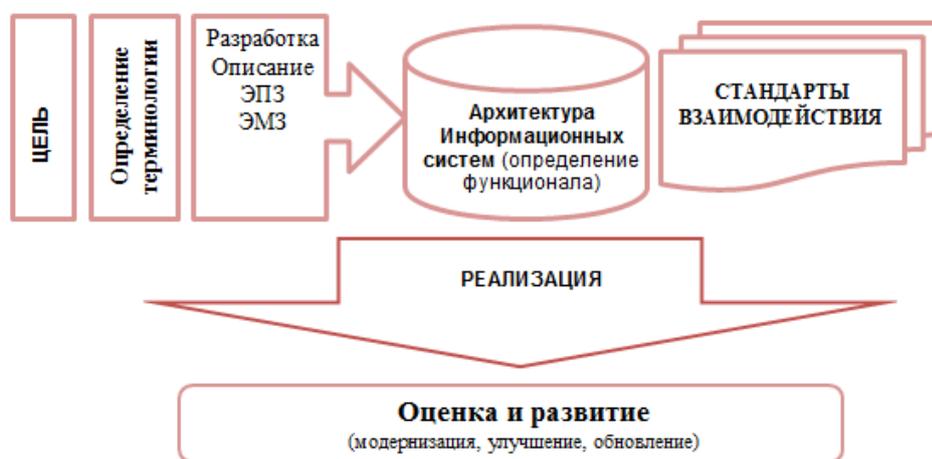


Рисунок 1.3 - Общая схема реализации и развития е-здравоохранения РК

Сегодняшняя архитектура е-здравоохранения Республики Казахстан содержит следующие 2 типа ИС и приложений:

- ИС, разработанные по принципу клиент-серверных настольных приложений. Это системы имеют сервера в пределах самого медицинского учреждения, и все пользователи работают с системами по локальной сети от своих рабочих станций. В течение ночи, все данные, которые обновились в течение дня, отправляются с локального сервера вверх (на уровень области

для обеспечения резервного копирования данных), где происходит синхронизация данных, а вниз передаются возможные изменения в справочниках, в структурах данных и в кодах ПО. Обмен оперативной информацией, необходимой для процесса оказания медицинской помощи, между централизованными системами и узлами систем ЕИСЗ осуществляется посредством обращения к сервисам путем обмена сообщениями через транспортную среду.

- Приложения, разработанные на веб-технологиях, в которых пользователи посредством веб-браузеров обращаются напрямую в центральный ЦОД МЗ РК, где размещены веб-сервера, приложения и базы данных.

Осуществляется переход на облачные технологии. Рассматривается наличие резервного центра обработки медицинских данных (ЦОД) и канала связи. В данной архитектуре существует две категории медицинских организаций: работающие в облаке и работающие с локальным сервером, причем в начале большинство организаций будет из второй категории (рисунок 1.4).

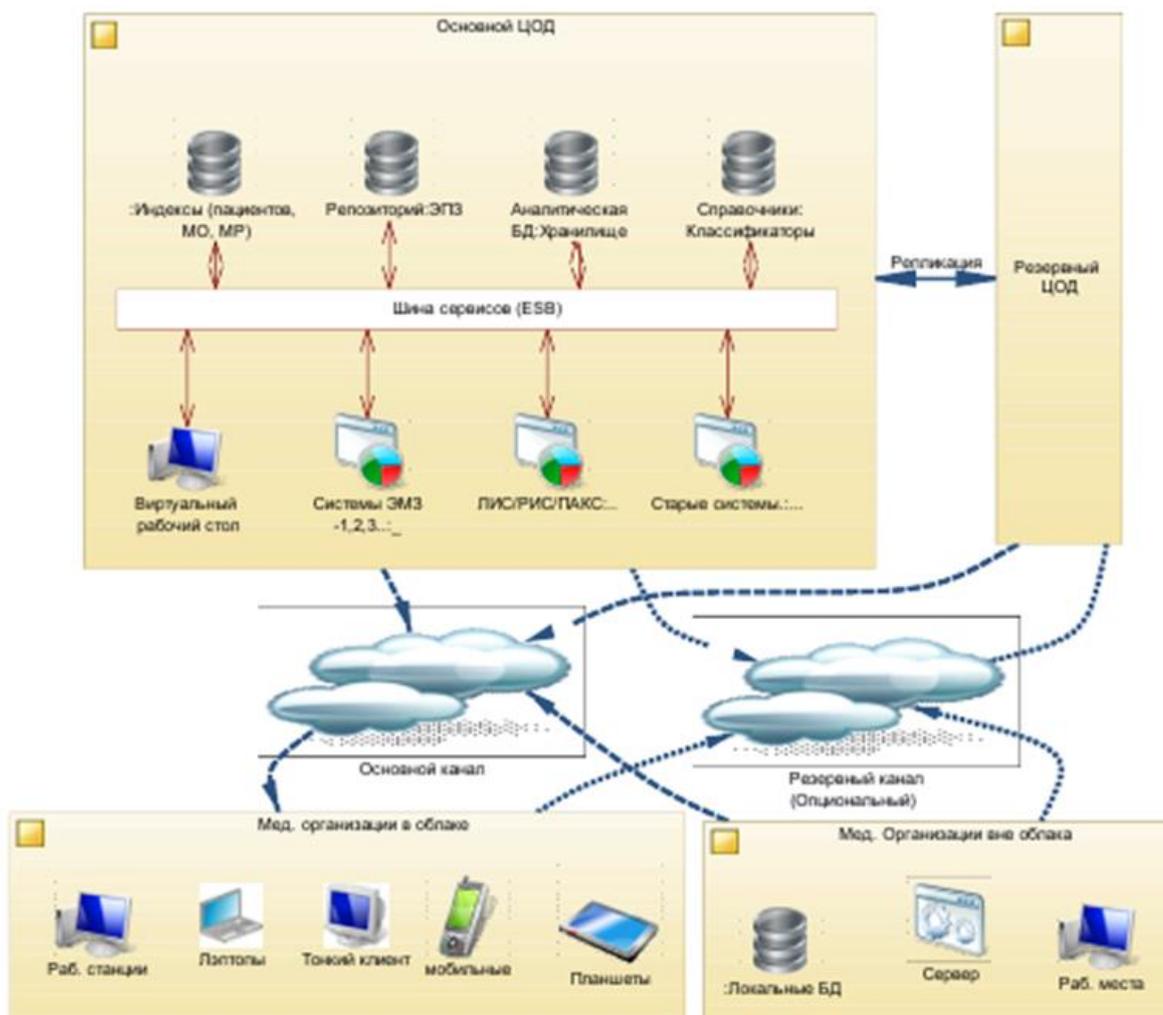


Рисунок 1.4 - Архитектура программно-аппаратных средств в е-здоровоохранении

Внутри ЦОД, храниться все БД, среди которых выделены следующие основные:

- Индекс пациентов (РМІ), который содержит основные идентификационные данные каждого пациента (при этом персональные данные будут храниться в зашифрованном виде в БД), демографию и другие основные данные;

- Электронный паспорт здоровья, в котором хранятся основные медицинские данные каждого пациента;

- Специализированные БД отдельных ЭМЗ и веб-приложений, а также централизованных информационных систем, работающих в облаке;

- Хранилище данных для статистических вычислений [3].

Сейчас все большее распространение получает сервис «Программное обеспечение как услуга» (SaaS). Следовательно, становятся все более популярными службы «облачных» вычислений. Часто они даже полностью вытесняют локальные системы и хранилища данных. Предполагается, что такие системы могут полагаться исключительно на службы SaaS и предоставлять все необходимые функции через тонкие клиенты, которые представляют собой сочетание недорого аппаратного обеспечения, операционной системы и веб-браузера.

Преимущества модели SaaS в сфере здравоохранения неоспоримы. Эти преимущества выражаются в следующих пунктах:

- 1) Наличие единого информационного пространства. Возможность быстрого доступа к вычислительным и к информационным ресурсам, когда это необходимо, в значительной степени является идеальным решением для многих больниц, медицинских клиник и кабинетов врача. Это должно предоставить им возможность улучшить услуги для своих пациентов.

- 2) Все данные системы хранятся в едином хранилище и обрабатываются на серверах системы. Проблема, по крайней мере, на данный момент, является сохранение конфиденциальности пациентов. Риски предоставления конфиденциальных данных о пациентах, особенно в общедоступных облачных инфраструктурах, по-прежнему сдерживают темпы внедрения облака.

- 3) Обновления и архивация проводятся автоматически. Распространение электронных медицинских карт, реализуемое на базе облачных технологий, позволяет автоматизировать не только регистрацию больного, но и вести централизованное отслеживание процесса предоставления медицинских услуг на более качественном уровне. Так же облегчается перевод статистических, административно-хозяйственных, финансовых данных в электронный формат, что значительно упрощает их обработку и оптимизирует бизнес-процессы.

- 4) Системные требования к рабочим местам пользователя минимальны, не требуется установка и настройка, тем самым достигается экономия средств и времени;

- 5) Возможный доступ из любого узла системы. Облако приносит мощные ИТ-ресурсы для медицинских работников, медицинских организаций, которые доступны в любой точке мира. Приложения мирового класса и

вычислительная инфраструктура доступны всем без значительных предварительных инвестиций.

Существуют следующие сферы применения модели SaaS в здравоохранении:

- ERP (Enterprise Resource Planning) система, которая позволяет хранить и обрабатывать большинство критически важных данных: документооборот, финансовый отчет, кадровый учет, закупки и управление запасами;

- Система медицинской нормативно-справочной информации;

- CRM (Customer Relationship Management) – программное обеспечение для организации и автоматизации работы с заказчиками. В здравоохранении к ним можно отнести: управление базой пациентов, управление текущей деятельностью объекта здравоохранения);

- Портальное решение. В целях повышения уровня информированности общественности создается информационный портал в рамках единой системы здравоохранения. Оно направлено на повышение осведомленности общественности, повышение прозрачности системы здравоохранения и повышение подотчетности и эффективности в этой области. Через этот портал должен осуществляться доступ граждан к достоверной и полной информации о: государственных программах здравоохранения, медицинских услугах, поставщиках услуг, медицинском оборудовании, медицинском персонале и их рабочем месте. Портал должен позволять гражданину проверить свой страховой статус и узнать, какие услуги ему доступны в рамках программ здравоохранения.

Таким образом, внедрение облачных технологий в медицинскую отрасль позволяет решить ряд проблем, связанных с неэффективностью оказания медицинских услуг. Большие массивы различных медицинских данных, которые преимущественно хранятся в бумажном виде, тяжело поддаются к обработке и анализу и препятствуют быстрому принятию клинических решений.

Хотя такой подход может быть в некоторых случаях рациональным, он не может быть универсальным, поскольку таким образом невозможно обеспечить удобство использования, полную интеграцию с производительными настольными системами, а также возможность работы в автономном режиме (без подключения к Интернету). Поэтому, считается целесообразным применять расширенный подход «программные продукты и интернет-службы» (S+S).

1.3 Обзор существующих методов и моделей поддержки принятия медицинских решений

На сегодняшний день медицинская сфера обладает большими объемами медицинской информации, которые требуют надлежащей обработки и дальнейшего применения в диагностике и лечении различных заболеваний. Развитие компьютерных технологий дает огромные возможности сбора, обработки, управления и исследования медицинской информации для лучшего понимания сложных биологических процессов жизни и содействия решению проблемы диагностирования и лечения в медицинских учреждениях.

Точная и беспроблемная диагностика и лечение, получаемая пациентами также является одной из главных проблем медицинского обслуживания. Поэтому системы поддержки принятия медицинских решений на основе различных методов анализа данных становятся популярными инструментами для медицинских исследователей, которые стремятся выявлять и использовать шаблоны и отношения между большим числом переменных и быть в состоянии предсказать определенное заболевание и результаты применяемого лечения.

Область биомедицинской информатики привлекает все большую популярность и внимание и быстро растет за последние два десятилетия. Благодаря новым достижениям в области медицинских технологий огромное количество медицинских данных исследований генерируется каждый день. Оцифровка важной медицинской информации, такой как лабораторные отчеты, записи пациентов, научные статьи и анатомические изображения также привели к большому количеству медицинских данных.

Биомедицинские исследователи и практические специалисты теперь сталкиваются с проблемой «переполнения информации». В настоящее время скорость накопления данных намного выше, чем скорость интерпретации данных. Эти данные должны быть эффективно организованы и проанализированы для того, чтобы быть полезными в медицинских учреждениях.

Для управления этими большими хранилищами медицинских данных и выявления полезных знаний с помощью различных моделей необходимо использование новых вычислительных и информационных технологий. В частности, в последние годы в сферу медицинского диагностирования внедряются методы управления знаниями, интеллектуального анализа данных.

Методы интеллектуального анализа применительно к медицинским данным, имеют большие возможности, особенно для выявления скрытых закономерностей в массивах медицинских данных. Созданные с их помощью модели с большим успехом используются для постановки диагнозов. Хотя, имеется большой объем накопленных медицинских данных, в данном направлении существуют некоторые трудности, касательно их неоднородности. И это является одним из относительно решаемых задач, которые рассматриваются в различных исследованиях данного направления.

При работе с большим объемом медицинских данных возникает

необходимость в использовании системы комплексного интеллектуального анализа данных, которая могла бы агрегировать и анализировать разнотипную информацию, поступающую от различных объектов медицинской организации. Из-за больших объемов поступающей информации такая система должна правильно использовать существующие технологии работы с большими данными. Анализируемые данные должны собираться в определенной организованной форме. С помощью информационных технологий собранные данные интегрируются в медицинскую информационную систему. Записи о ходе лечения миллионов пациентов могут храниться и компьютеризироваться, а медицинские учреждения будут иметь возможность анализировать эти данные. Тем самым, ускоряя процессы постановки диагноза и принятия терапевтических решений.

Клинические решения часто принимаются на основе интуиции и опыта врачей, а не на основе знаний богатых данных, скрытых в базе данных. Такая практика приводит к нежелательным перекосам, ошибкам и чрезмерным медицинским затратам, что сказывается на качестве обслуживания пациентов. Многие исследователи в своих трудах доказывают, что внедрение систем поддержки принятия решений, основанную на компьютеризированных записях медицинских данных, позволяют уменьшить количество медицинских ошибок, увеличить безопасность пациентов, что является основной задачей здравоохранения.

Система оказания медицинской помощи меняется во многих отношениях. Технологические достижения открывают возможности для оптимизации ухода за пациентами. КИС имеют потенциал для решения многих проблем, возникающих в области здравоохранения, а именно, управление большими объемами данных о пациентах и снижение затрат на здравоохранение, ведение научных исследований. Врач может дистанционно или непосредственно получить доступ к историям болезни пациентов, изучая которых в контексте соответствующих клинических исследований, могут принимать научно-обоснованные решения [16].

Методы интеллектуального анализа могут быть применены для построения моделей принятия решений для таких процедур, как прогнозирование, диагностика и планирование лечения, которые после проверки и оценки могут быть встроены в клинические ИС [17].

Существует ряд исследований, в которых подробно изложены работы авторов по использованию методов интеллектуального анализа данных в медицине [18-23]. В работе [22] делается не только обзор литературы в данном направлении, но и рассматриваются проблемы и предстоящие задачи применения методов интеллектуального анализа данных в медицине. В качестве элемента исследования методов интеллектуального анализа данных рассматриваются вопросы разработки методов ИАД на основе обзора литературы и классификации статей за последние десятилетия, так как этот период важен тем, что в течении этого времени наблюдается распространение методов ИАД, используемых в медицинской отрасли [24]. Сделан обзор

предыдущих попыток разработки компьютерных клинических диагностических инструментов и обсуждаются проблемы, возникающие при разработке, внедрении, оценке и поддержании систем поддержки принятия решений по клинической диагностике [25].

Авторы исследования [26] сделали следующие выводы:

1) улучшение точности применения систем поддержки принятия медицинских решений может быть возможным путем моделирования неопределенных и временных данных, исследований алгоритмов вывода, интеграции информации о пациентах из разных источников и улучшения алгоритмов профилирования генов;

2) для оценки характеристик методов требуется сравнительная оценка различных методов моделирования, которые могут помочь разработчикам в выборе техники для решения конкретной медицинской проблемы;

3) оценки применений систем поддержки принятия медицинских решений в клинических условиях необходимы для содействия использованию врачами этих вспомогательных средств для принятия решений.

Различные типы систем поддержки принятия решений были разработаны для оказания помощи врачам в принятии более эффективных или быстрых решений. Однако большинство из них сосредоточено на диагнозе [27-40], и прогнозировании исхода болезни [41-48].

Существуют также интегрированные системы, которые связывают решение о лечении с решением, о диагностике [49, 50]. Таким образом, система оказания медицинской помощи меняется во многих отношениях. Технологические достижения открывают возможности для оптимизации принятия решения по диагностированию заболевания и по уходу за пациентами. Клинические информационные системы имеют потенциал для решения многих проблем, возникающих в области здравоохранения, а именно, управление большими объемами данных о пациентах и ведение научных исследований на основе этих данных.

В работе [16] так же подробно описываются все преимущества и проблемы существующих клинических информационных систем. Основными преимуществами клинических информационных систем по словам авторов являются: прямой доступ к мгновенным обновлениям медицинской карты пациента, а также удаленный доступ к записям пациентов; ориентированное на пациентов принятие решений на основе лучших клинических данных; быстрая обработка данных обеспечивает быстрое принятие решения по диагностике заболевания; возможность разработки эффективного и интуитивно понятного программного обеспечения для обработки данных и инструментов биоинформатики; больше шансов провести потенциальное исследование, основанное на реальных данных; улучшение качества данных и анализа данных пациента путем объединения их с собственными знаниями врача. Наряду с преимуществами клинических информационных систем, также существуют требующие решения различные проблемы, такие как: необходимость разработки и интеграция подсистем; высокие первоначальные инвестиций при

низкой предполагаемой отдаче; безопасность и конфиденциальность; необходимость создания коммуникации между большим количеством врачей разных специальностей;

Работа [51] иллюстрирует необходимость многоуровневых междисциплинарных интегрированных интеллектуальных систем в медицине. Такая система не только снизит затраты, ускорит оказание медицинских услуг, предоставит точные статистические данные, позволит научное междисциплинарное и многоуровневое планирование, предоставит более точную медицинскую документацию или помочь в принятии критических решений в области здравоохранения.

Хотя все большее количество данных производится различными биомедицинскими и медицинскими системами, они еще не полностью воспользовались возможностями, которые эти данные предоставляют для трансформации. Применение методов, основанных на данных, для больших данных о здоровье может быть очень полезным в области биомедицины и здравоохранения, позволяя идентифицировать и извлекать соответствующую информацию и сокращать время, затрачиваемое биомедицинскими и медицинскими специалистами и исследователями, которые пытаются найти содержательные шаблоны и новые потоки знаний [16].

В статье [52] каждая работа изучается на основе шести медицинских задач: скрининг, диагностика, лечение, прогноз, мониторинг и управление. В каждой задаче рассматриваются пять подходов интеллектуального анализа данных: классификация, регрессия, кластеризация, ассоциация и гибридный подход.

Авторы обзора [53] описывают основные особенности интеллектуального анализа клинических данных и рассматривают два конкретных аспекта, представляющих особый интерес: методы, способные работать с временными данными, и усилия, предпринимаемые для перевода результатов молекулярной медицины в клинически полезные модели интеллектуального анализа данных.

В обзорной работе [54] авторы дают сравнительный анализ различных экспертных систем, где основной упор делается на рассмотрении использованных методов. Авторы утверждают, что точность и другие параметры экспертной системы зависят от базы знаний. Особое внимание следует уделять приобретению знаний - этапу, на котором происходит их накопление. Наиболее эффективными являются медицинские экспертные системы на основе методов теории нечетких множеств, нейронных сетей и на основе базы правил.

Большинство рассматриваемых работ сосредоточены на исследовании наборов данных по следующим конкретным заболеваниям: рак молочной железы [32, 55-60], пневмония [27, 61], артериальное давление [62], туберкулез [63], сердечно-сосудистые [64-66], рак [67-71], диабет [72-74], церебрально-васкулярная болезнь [75,76], аутизм [33], кожные заболевания [77,78], головные боли [79], болезнь Альцгеймера [80,81], эпилепсия [82,83].

Медицинские данные для исследований собираются из различных

источников, поэтому они могут содержать различные шумы, противоречия, пропущенные значения, предвзятость и избыточность. Важная информация может храниться в неструктурированном виде.

Поскольку существуют разные наборы данных для разных заболеваний, и наборы данных могут иметь разные характеристики, диапазон низкой достоверности должен быть определен конкретно. Кроме того, требуемый набор данных для проверки структуры системы должен состоять из клинических измерений, фактического класса, фактического осуществленного лечения и фактического эффекта от лечения [49].

При сборе данных для анализа представляется возможным, также, многоуровневая интеграция услуг здравоохранения, в которой достигается высокая эффективности хранения и использования входных данных о пациентах [53]. Характеристики собранных данных для анализа могут быть различны: измеримые клинические (например, количество лейкоцитов, нейтрофилов, гемоглобина и т.д.), наблюдаемые клинические (например, сонливость, стенокардия, кашель и т.д.), общепринятые показатели жизненно важных функций (например, частота дыхания, частота сердечных сокращений, насыщение кислородом и т.д.) [27].

Для создания набора данных, как правило, используются существующие хранилища данных, после извлечения которых, их необходимо нормализовать.

Дискретизация и нормализация - это две процедуры преобразования данных, которые помогают представлять данные и их взаимосвязи точно в табличном формате, что делает базу данных легкой для понимания и эффективной в работе. Это также уменьшает избыточность данных и повышает эффективность. Элементы данных могут быть нормализованы путем выделения уникального номера столбца для каждого возможного значения. Числовые поля данных дискретизируются путем взятия значений, которые находятся в некотором диапазоне, определяемом минимальным и максимальным пределами. В таких случаях можно разделить данный диапазон на несколько поддиапазонов и назначить уникальный номер столбца для каждого поддиапазона соответственно [84].

В работе [29] предварительная обработка данных включает пересчет нестратифицированной случайной выборки, метод синтетического меньшинства (SMOTE), очистку данных вне допустимого диапазона, и удаление дубликатов.

В работах [85] обсуждаются те основные трудности, которые возникают при создании систем поддержки принятия медицинских решений. К таким проблемам авторы относят: неоднородность медицинских данных; недостаточная связь между информационными технологиями и медицинским сообществом; правовые, этические и социальные аспекты, касательно медицинских данных.

1.4 Существующие системы диагностирования заболеваний крови

История медицинских систем по диагностированию заболеваний крови начинается с разработкой системы CADIAG, которая была системой в традициях работы биомедицинских исследователей в США после Второй мировой войны. Она использовала методы автоматической обработки данных для изучения корреляции признаков и симптомов с заболеваниями. Результатом работы системы стало быстрое накопления данных медицинских исследований.

Ученые M. Lipkin and J. D. Hardy попытались классифицировать и соотнести все медицинские данные при диагностировании гематологических заболеваний с использованием механического аппарата [75]. Для регистрации данных, на основе которых был поставлен диагноз гематологических заболеваний, проводились консультации по стандартным учебникам с выбором 26 заболеваний и с перечислениями всех характеристик каждого заболевания. Хранение и сортировка этих медицинских данных были впервые выполнены с использованием маргинальных перфокарт. Периферия этих карт была разделена на нумерованные пространства, и в каждом из них было пробито по одному отверстию. Каждое из 138 доступных мест имело значение элемента информации, и каждая карточка представляла определенный объем информации. Один край карточки состоял из данных, полученных из истории болезни, а второй - из данных, полученных в результате физического обследования. Информация, касающаяся исследования периферической крови, была отнесена к третьей стороне, а исследование костного мозга и другие лабораторные работы - к четвертой стороне. Таким образом, информация об определенном заболевании была представлена одной единственной картой.

CADUCEUS был медицинской экспертной системой, законченной в середине 1980-х годов профессорами Питтсбургского университета: Гарри Попла и Джек Мейерс, ведущими специалистами по внутренней медицине. Их мотивацией было стремление улучшить MYCIN, который сосредоточен на переносимых кровью инфекционных бактериях, для того, чтобы сосредоточиться на более комплексных вопросах, чем на такой узкой области как заражение крови.

В конечном итоге CADUCEUS мог диагностировать до 1000 различных заболеваний. Хотя CADUCEUS работал аналогичного MYCIN, с использованием механизма вывода, в него был внесен ряд изменений. Система CADUCEUS была описана как «самая наукоемкая экспертная система из существующих» [86].

Системы MYCIN и CADUCEUS поддерживают работу с большими объемами данных. Предметная область ограничивается только знаниями экспертов и возможностью трансформировать эти знания в четкие правила. Имеют ограничения в работе с данными, имеющими пропуски, не могут получать новые знания из экспериментальных данных.

В настоящее время разработаны, получили признание, хорошо документированы и доступны для использования системы построения баз

знаний от простых языковых средств представления знаний до законченных систем со средствами ввода-вывода. Среди них выделяется ЕМУСИН - фреймовая система, являющаяся дальнейшим развитием МУСИН, которая дополнительно содержит спецификацию знаний по заболеваниям крови и обладает специальным интерфейсом для пользователя КАС (Knowledge Acquisition System).

Название системы ЕМУСИН происходит от сокращения «Empty MYCIN» - пустая МУСИН, которая может использоваться для создания программ, консультационного характера, применительно к любой предметной области. К тому же, данная система обладает большим количеством инструментальных средств, которые позволяют проектировать и разрабатывать программы для решения задач экспертных систем.

Данная система особенно удобна для поддержки принятия систем диагностирования заболеваний, поскольку может работать с нечеткими данными, которые характерны для медицинских признаков, например, такие как результаты лабораторных анализов, симптоматика болезней, анамнез жизни больного.

Стоит отметить важные особенности и характеристики системы ЕМУСИН:

- язык представления правил в системе ЕМУСИН немного отличается от языка системы МУСИН, и является более понятным и структурированным;

- схемы применения правил имеют возможность группировать существующие правила, где основными критериями группирования правил являются ссылки на эти правила;

- структура управления правилами характеризуется в виде дерева, в котором данные представляются как листья этого дерева. Запрашиваемые пользователем данные хранятся в виде таблицы;

- существует два интерфейса данной системы: а) интерфейс консультационной программы, который управляет запросами системы на получение различных данных по критериям пользователя; б) интерфейс системы с разработчиками, где управляется процесс редактирования правил и базы знаний, которые как обычно представлены в виде таблиц, а также тестирование правил.

Система СЛОТ, предназначенная для помощи врачам в определении нарушений свертывания крови у пациентов действует на основе выполнения заданных правил. Данная система применяется врачами для консультаций по проблемам кровотечения. СЛОТ имеет концептуальную структуру консультации, которая включает в себя первоначальную оценку нарушений кровотечения на основе клинических данных, анализ ферментативных и тромбоцитных показателей, окончательную оценку нарушений кровотечения и составление окончательного диагноза. База знаний СЛОТ содержит 63 правила, 50 параметров консультаций; 1 консультация требует 5 мин. работы за интерактивным терминалом. Данная автоматизированная система была разработана с помощью независимой от области применения консультирующей

системы EMYCIN, ранее применявшейся для диагностических целей и лечения инфекционных болезней. Здесь рассмотрены возможности использования системы EMYCIN в качестве базы для представления знаний в другие системы.

В работе [87] даются оценки временных затрат на разработку CLOT, а также показаны подробные примеры формулировок правил и ведения консультаций системой CLOT.

Описываются 3 основные стадии процесса комплектования базы знаний для этих систем:

1) начальная фаза процесса формализации знаний; предусматривается ознакомление специалистов с предметной областью, функциями и задачами системы, со средствами представления знаний; формируется концептуальная структура ведения консультаций;

2) средняя фаза, при которой детализируются причинно-следственные связи и разрабатываются основные логические правила базы знаний; проводится моделирование функционирования программ;

3) завершающая фаза, предусматривающая обнаружение и устранение дефектов системы; могут быть включены новые параметры и добавлены отдельные правила для обработки неясных случаев.

Существующая на сегодняшний день автоматизированная система гематологической диагностики «Атлант-гемо» является одной из передовых систем, предназначенных для диагностирования острых лейкозов, которые определяются на основе кластных структур. Система была создана совместными усилиями группы ученых Российского онкологического центра имени Н. Блохина РАМН, Гематологического научного центра РАМН и Российской медицинской академии. Данные для работы системы получаются из медицинских оборудований, с помощью которых снимаются цифровые изображения компонентов крови и сохраняются в базах данных..

Существует также множество интернет сервисов расшифровки различных анализов крови и постановки предварительного диагноза.

Например, на сайте Polismed.com обрабатываются персональные медицинские данные пользователей, после которого ставится предварительный диагноз [88]. Окно web-страницы сервиса Polismed.com представлено на рисунке 1.5.

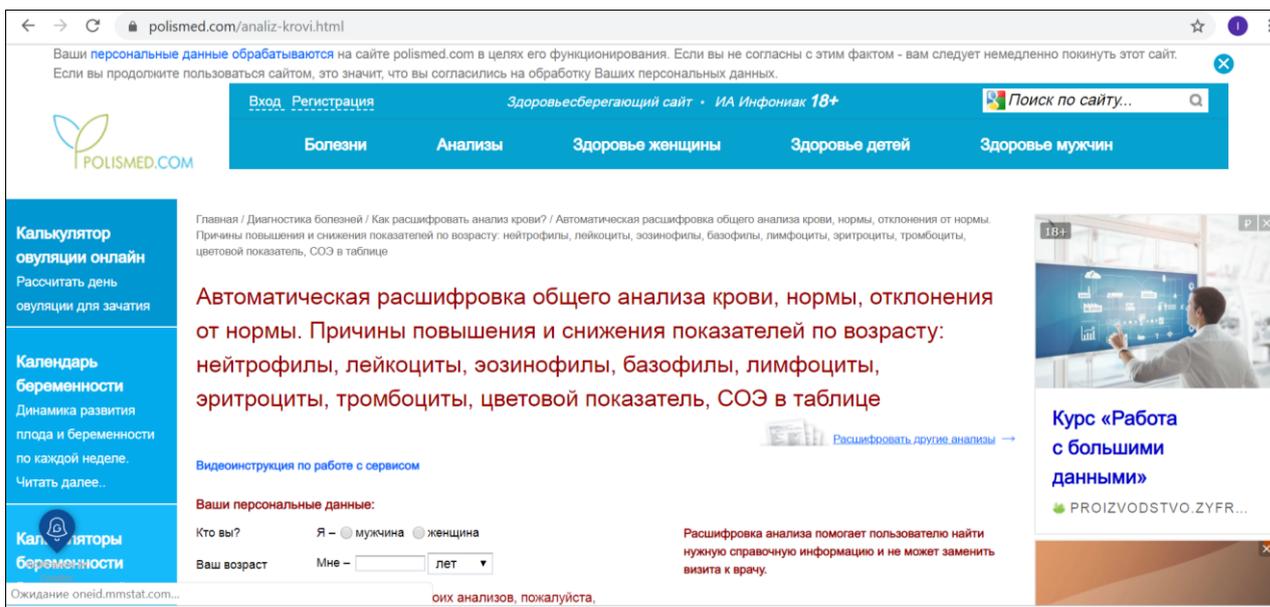


Рисунок 1.5 – Окно web-страницы сервиса Polimed.com

Данная система позволяет автоматически расшифровывать данные лабораторных анализов и выявлять отклонения от нормы.

Вводя лабораторные данные анализа в соответствующие поля, пользователь может получать информацию о возможных причинах повышения или понижения значений показателей биохимического анализа крови, что дает возможность постановки предварительного диагноза. Система также рассматривает функцию диагностирования заболевания по введенным данным пользователя в онлайн режиме.

Аналогичный интернет сервис TestResult.org, web-страницы которого представлены на рисунке 1.6, также предлагает список анализов, по которым предполагается постановка предварительного диагноза [89].

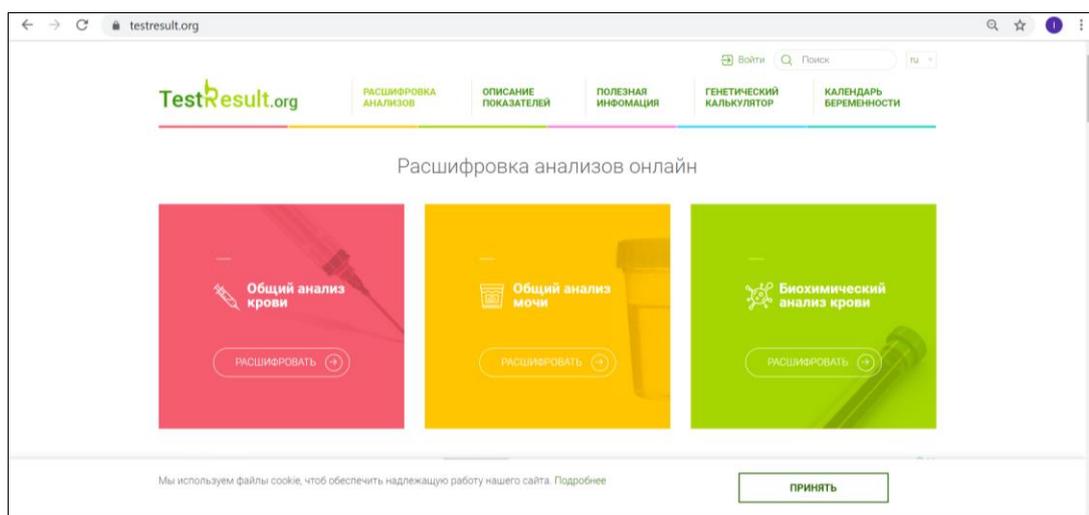


Рисунок 1.6 - Окно web-страницы сервиса TestResult.org

На данном сайте пользователь вводит не числовые значения показателей анализа, в том числе и биохимического, а указывает значения: ниже нормы, норма или выше нормы. По введенным данным выдается результат в виде списка заболеваний с их описаниями, которыми предположительно страдает пользователь сервиса.

Анализируя существующие практические решения можно сделать вывод, что любой интернет сервис по онлайн диагностированию заболеваний предупреждает, что результаты онлайн-диагностики не должны считаться официальным диагнозом или заменять посещение врача. Интернет сервисы выполняют исключительно информационные функции и при подозрении на заболевание советуют как можно скорее обращаться к специалистам.

1.5 Выводы по первому разделу

1. В результате анализ систем поддержки принятия решений в медицине было выявлено основные четыре направления медицинских систем: дифференциальное диагностирование; прогнозирование хода заболевания и определение клинической картины; наблюдение и регистрация текущего состояния пациента; оценка степени выраженности синдрома болезни в режиме «реального времени».

2. Проанализированы предпосылки и основные направления развитие электронного здравоохранения в Казахстане. В рамках принятых документов по здравоохранению в Республике Казахстан осуществляется повсеместное внедрение и развитие единой национальной информационной системы здравоохранения в части разработки единых стандартов, технологических спецификаций и характеристик различных информационных систем по требуемому функционалу. И это является предпосылками для разработок систем поддержки принятия решений для диагностирования различных заболеваний, которые призваны повысить эффективность деятельность медицинских работников путем внедрения различных методов анализа данных и медицинских данных.

3. Изучены и описаны архитектурные модели и модели управление информационными системами здравоохранения.

4. Актуальность исследования и разработки систем поддержки принятия медицинских решений для диагностирования болезней крови было обосновано обзором существующих методов и моделей поддержки принятия медицинских решений.

2 МЕТОДЫ И МОДЕЛИ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ

2.1 Постановка задачи и разработка функциональной модели диагностирования клинико-гематологических синдромов

Кровеносная система – одна из главных систем организма человека, которая регулирует все его функции. Кровь – это важнейшая жидкая среда организма. Она непрерывно циркулирует по разветвленной сети сосудов и выполняет целый ряд жизненно важных функций – защитную, транспортную, дыхательную и др. [90].

Среди болезней органов кроветворения анемии являются наиболее распространенными патологиями. Заболевания анемии включают различные патологические состояния, общим для которых является малокровие [91]. Анемии неоднородны по причинам и механизмам развития, по клиническим проявлениям, что затрудняет диагностику.

Задачи диагностики состояния объекта можно формализовать следующим образом (2.1):

$$MDia = \langle M, B, SB, J, U, O \rangle \quad (2.1)$$

где: M – знания экспертной группы; $B: M \rightarrow SB$ – проектирование базы знаний диагностирования клинико-гематологических синдромов; $U: SD, J \rightarrow O$ – реализация алгоритма морфологической классификации показателей анемического синдрома на основе клинико-гематологических показателей; $SB = \{A, R, SC, K, D\}$ – знания экспертов, хранящиеся в базе знаний; $A = \{A_i\}$ – группы диагнозов клинико-гематологических синдромов диагностируемого объекта; $R = \{R_i\}$ – множество признаков дифференциальной диагностики клинико-гематологических синдромов; $SC = \{\langle \{R_j^i, RF_j^i\}, A_i \rangle\}$ – множество комплексов признаков, при котором для диагноза A_i выставляется экспертная оценка R у которой сила проявления равна RF_j^i ; $M = \{\langle \{A_i^q\}, K_q \rangle\}$ – множество набора диагнозов, в котором A_i^q принадлежит в набору K_q ; $G = \{\langle \{S_j^p\}, D_p \rangle\}$ – множество наборов признаков, где R_j^p принадлежит набору D_p .

Состояние объекта дифференциальной диагностики клинико-гематологических синдромов описываются набором признаков k , значения которых измеряются по шкале $I = \{\langle R_k, x_k \rangle\}$ и набором диагностических решений h , значения которых измеряются по шкале $O = \{\langle A_h, y_h \rangle\}$. Параметр вероятности возникновения признаков и отнесения его к к определённому классу диагнозов лежит в диапазоне $[0.0; 1.0]$.

На рисунке 2.1 представлена концептуальная модель дифференциальной диагностики клинико-гематологических синдромов

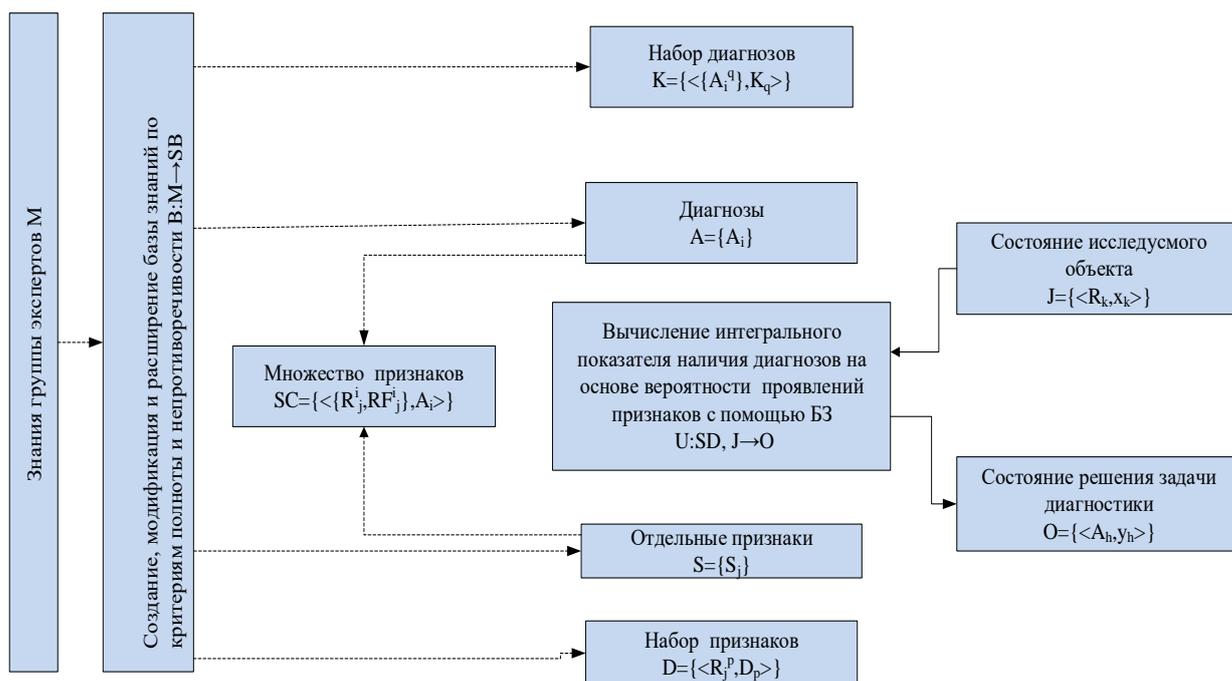


Рисунок 2.1 - Концептуальная модель дифференциальной диагностики клиничко-гематологических синдромов

Постановка задачи, которая описана выше, основывается на данных базы знаний, в которой хранятся полные и непротиворечивые знания экспертов. Эта база знаний применяется для установления диагноза состояния исследуемого объекта дифференциальной диагностики клиничко-гематологических синдромов.

В рамках данного исследования осуществляется разработка медицинской диагностической системы диагностирования клиничко-гематологических симптомов, в частности заболеваний анемии.

Анемия – это клиничко-гематологический синдром, для которого характерны уменьшение содержания гемоглобина в единице объема крови, чаще при одновременном уменьшении количества эритроцитов, приводящие к развитию кислородного голодания тканей [91]. Термин «анемия» без детализации не определяет конкретного заболевания, а указывает на изменение в анализах крови, т. е. анемию следует считать одним из симптомов различных патологических состояний крови.

Для того, чтобы спроектировать ИС, на начальных этапах создания необходимо понять, какие задачи будут решаться с помощью создаваемой ИС, каковы цели ее создания и функции, которые она будет выполнять. Также очень важно определить каков состав функциональных подсистем разрабатываемой системы, и как они могут взаимодействовать между собой. Для наглядного описания деятельности системы необходимо построить ее модель, которая дает отображение предметной области и всех участников реализуемых бизнес-процессов.

Для описания процесса диагностирования клинико-гематологических симптомов была разработана его функциональная модель, которая отображает основные бизнес-процессы, из которых состоит процесс и взаимосвязь между ними, а также участников процесса.

Целью построения функциональной модели диагностирования клинико-гематологических симптомов является точная спецификация всех операций и действий, осуществляемых в процессе диагностирования, а также характера взаимосвязей между ними. Модель обеспечивает полное представление, как о функционировании исследуемого процесса работы, так и обо всех имеющих в нём место потоках информации и материалов [92]. В качестве CASE средства для графического отображения разработанной модели была использована система IDEF0/EMTool [93], которая обеспечивает поддержку на этапе анализа и проектирования, являющегося одним из наиболее трудоёмких, а так же обеспечивает визуальное представление информации. Назначение модели заключается в том, что она графически представляет основные этапы работы, функциональные связи, потоки объектов, информации. Преимуществами использования IDEF0 является простота графики, используемой при построении диаграмм, постепенность представления деталей процессов, использование декомпозиции блоков диаграммы.

В результате проведенного анализа основным процессом был выделен процесс «Диагностирование клинико-гематологических симптомов». Контекстная модель представлена ниже на рисунке 2.2.

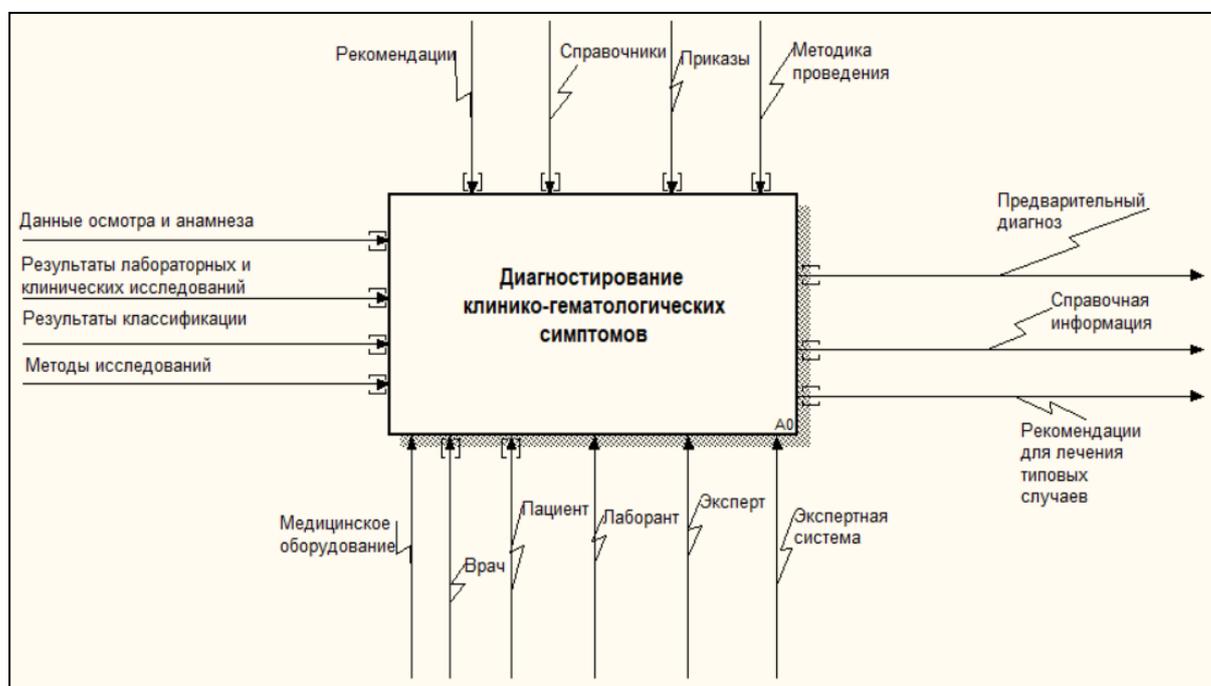


Рисунок 2.2 - Контекстная модель процесса диагностирования клинико-гематологических симптомов

Процесс «Диагностирование клинико-гематологических симптомов»

является процессом верхнего уровня и был декомпозирован следующие подпроцессы:

- «Первичный осмотр пациента и сбор анамнеза»;
- «Проведение лабораторных исследований»;
- «Установление наличия и степени выраженности анемического синдрома»;
- «Определение характера анемии в соответствии с морфологической классификацией»;
- «Подтверждение и уточнение типа анемии».

Графическая схема функциональной модели представлена ниже на рисунке 2.3.

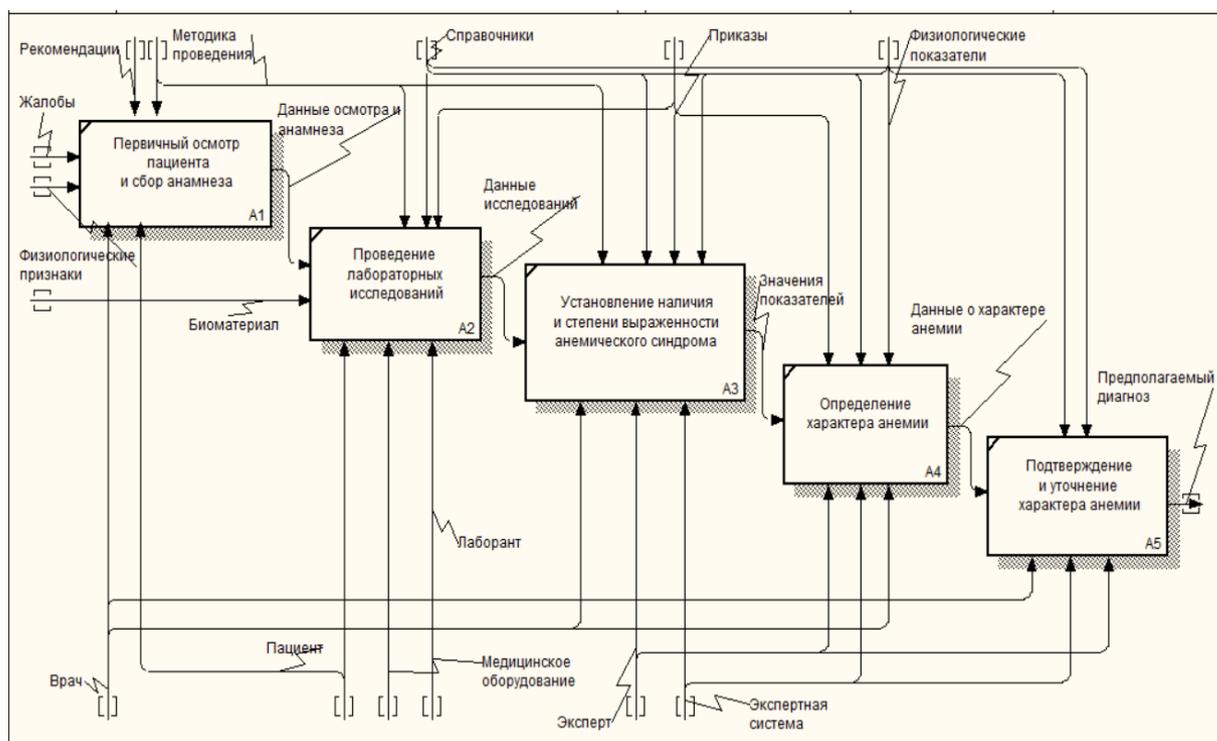


Рисунок 2.3 - Функциональная модель процесса диагностирования клиничко-гематологических симптомов

Первым этапом диагностирования клиничко-гематологических симптомов является прием пациента и его первичный осмотр, в результате которого врач собирает анамнез. Входящей информацией для данного процесса являются жалобы пациента, а также наличие физиологических признаков. Участниками процесса являются пациент и врач, непосредственно проводящий прием и осмотр пациента. При проведении осмотра и сбора анамнеза врач руководствуется существующей методикой проведения и рекомендациями, а также медицинскими справочниками.

После проведения осмотра пациента врач на основании собранного анамнеза назначает проведение лабораторных исследований крови. В качестве

лабораторных исследований крови могут быть назначены общий анализ и биохимический анализ крови. В данном процессе участниками выступают непосредственно сам пациент и врач-лаборант, проводящий исследования. В качестве входящего ресурса для проведения исследования используется биоматериал пациента, а именно кровь. Исследования крови пациента проводятся с использованием специального медицинского оборудования, а также с применением методики проведения лабораторных исследований и справочников, которыми руководствуется врач-лаборант.

Данные полученные в результате исследования крови выступают в качестве входящих ресурсов для следующего этапа, на котором проводится установление наличия анемического синдрома и степени его выраженности, если он имеет место. Данные исследования обычно представляются в виде гемограммы, в которой указаны основные количественные показатели крови. Для установления наличия анемического синдрома проверяется соответствие значений показателей крови их нормам, указанным медицинских нормативно-справочных документах и в клинических протоколах, утвержденных Министерством здравоохранения Республики Казахстан (МЗ РК). Сам процесс осуществляется на основе алгоритма расчета с использованием экспертной системы.

После этапа установления и подтверждения наличия анемического синдрома, а также степени его выраженности проводится следующий этап, на котором определяется характер анемии в соответствии с морфологической классификацией. Классификация анемии по морфологическим признакам проводится врачом при участии эксперта на основе значений показателей, полученных на предыдущих этапах, при помощи алгоритма классификации с использованием экспертной системы. В качестве регламентирующего документа для определения характера анемии используются клинические протоколы, утвержденные МЗ РК и медицинские справочники. Данный этап «Определение характера анемии в соответствии с морфологической классификацией» был рассмотрен более и детально и декомпозирован на следующие операции:

- «Определение среднего содержания гемоглобина в эритроцитах и среднего объема эритроцита»;
- «Определение принадлежности к микроцитарному типу» (Приложение Д);
- «Определение принадлежности к макроцитарному типу» (Приложение Е);
- «Определение принадлежности к нормоцитарному типу» (Приложение Ж).

Графически декомпозиция подпроцесса представлена на рисунке 2.4.

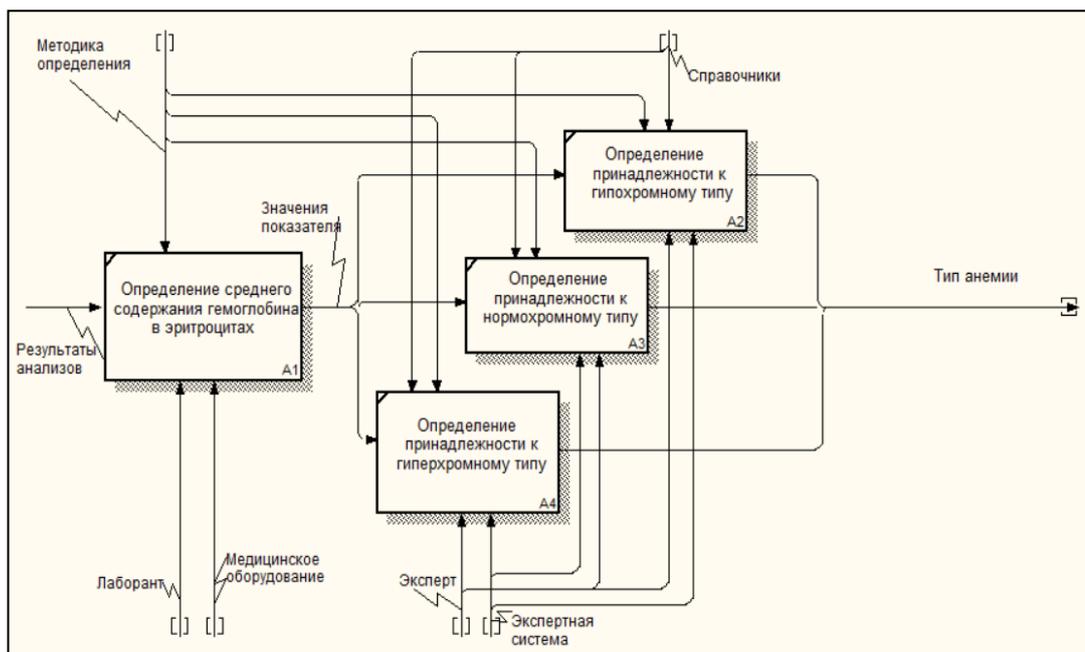


Рисунок 2.4 - Графическая схема декомпозиция процесса «Определение характера анемии в соответствии с морфологической классификацией»

На последнем этапе «Подтверждение и уточнение типа анемии» на основе входящих данных по классификации анемии и значений показателей лабораторных исследований крови, полученных на предыдущих этапах, определяется тип анемии. В качестве участников процесса выступают врач и эксперт, которые при помощи сформулированной базы правил определяют непосредственно тип анемии и по имеющимся данным в результате выносят предварительный диагноз.

2.2 Концептуальная модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови

Предложенная в работе концептуальная модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови (ППРДЗК) может быть представлена в виде (2.2):

$$KM = \langle OP, BZ, IP, PLV, E, KA \rangle \quad (2.2)$$

где: *OP* – модуль оперативной памяти; *BZ* – модуль базы знаний; *IP* – модуль интерфейса пользователя; *PLV* – модуль процедуры логического вывода; *IR* – модуль интерпретации решения; *NBZ* – модуль наполнения и проверки базы знаний.

Общая структурная модель информационной технологии ППРДЗК представлена на рисунке 3.

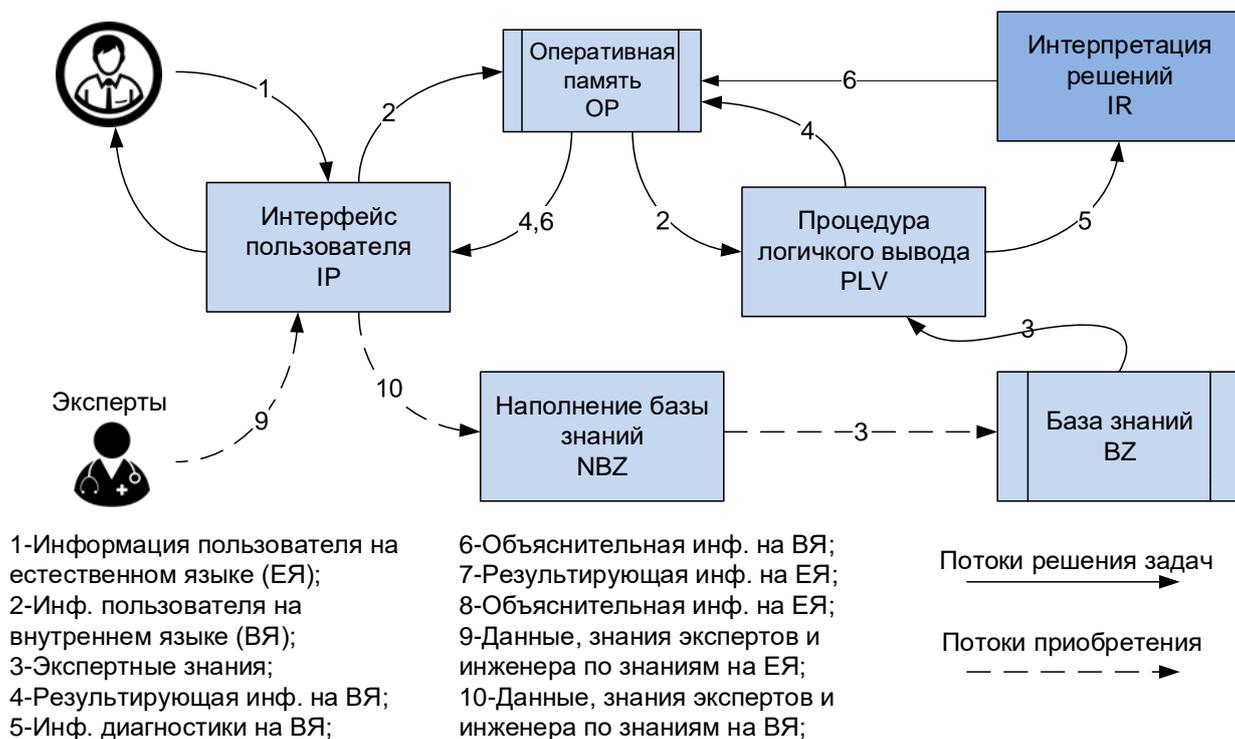


Рисунок 2.5 – Структура информационной технологии ППРДЗК

Модуль оперативной памяти (OP) используется для хранения данных дифференциальной диагностики клинко-гематологических синдромов, вводимой пользователями, и результатов дифференциальной диагностики клинко-гематологических синдромов.

Для системы поддержки принятия решений при диагностировании заболеваний крови предложена модель представления знаний, которая интегрирует структуру фрейма (frame) и нечетко-продукционную модель. Фрейм имеет определенную внутреннюю структуру, состоящую из множества элементов, называемых слотами, которым также присваиваются имена. За слотами следуют шпации, в которые помещают данные, представляющие текущие значения слотов [94]. Для описания групп показателей клинко-гематологических синдромов используется фреймовая модель представления знаний, а для описания причинно-следственных отношений, связывающих диагнозы с показателями клинко-гематологических синдромов.

Модель представления знаний можно представить в виде модели (2.3):

$$KM = \langle \{FrK, FrD, FrGD, FrP, FrGP, FrKP, \{FD_i\}, \{FEGD_i\}, \{FEP_i\}, \{FEGP_i\}, \{FCP_i\}\} \rangle \quad (2.3)$$

где: *FrK*- Frame-класс; *FrD*- протофрейм диагноза;
FrGD - протофрейм группы диагнозов;
FrP- протофрейм отдельного признака;
FrGP - протофрейм группы признаков;

$FrKP$ - протофрейм комплекса признаков;
 $\{FD_i\}$ - множество экзофрейм диагнозов;
 $\{FEGD_i\}$ – множество экзофрейм групп диагнозов;
 $\{FEP_i\}$ – множество экзофрейм отдельных признаков;
 $\{FEGP_i\}$ - множество экзофрейм групп признаков;
 $\{FCP_i\}$ - множество фреймов-комплексов признаков для диагнозов.

Структура Frame-класса представлена в таблице 2.1.

Таблица 2.1 - Структура Frame-класса

Название модели	Модель	Показатели модели
Frame-класс	$FrK = \left\langle \left\langle NS, TS, VS, \{PDS_i\} \right\rangle, \{PMS_j\} \right\rangle$ (2.4)	где: NFC – имя Frame-класс; NS – слота; TS – тип данных; VS – значение; $\{PDS_i\}$ – набор демонов слота; $\{PMS_j\}$ – набор включенных процедур.
Протофрейм диагнозов	$FrD = \langle NFSD, AKO, LVOD, CVO, \{\omega_i\}, SX, AE, NFIM \rangle$ (2.5)	где: FSD – имя Frame-прототип диагноза; AKO – слот, с именем фрейм Frame-класса; $LVOD$ – слот входной лингвистической переменной признака; CVO – слот количественной оценки наличия диагноза; $\{\omega_i\}$ – слот с набором коэффициентов нечетко-продукционных правил, определяющих группу весов для диагноза; SX – слот, хранящий пол пациента; AE – слот с диапазоном возраста пациента; $NFIM$ – слот имени экзофрейма набора диагнозов
Протофрейм отдельных признаков	$FrP = \langle NFSS, AKO, LVIS, CVI, \{SF_i\}, SG \rangle$ (2.6)	где: $NFSS$ – имя фрейма-прототипа признака; AKO – слот Frame-класса; $LVIS$ – слот входной лингвистической переменной признака; CVI – слот количественной оценки проявления признака; $\{SF_i\}$ – слот набора экспертных оценок силы проявления признаков в наборе признаков для диагнозов; SG – слот с именем экзофрейма групп признаков.
Протофрейм групп признаков	$FrKP = \langle NFSC, AKO, NFID, \{NFIS_i\}, \{FR_j\}, \{SFI_k\} \rangle$ (2.7)	где: $NFSC$ – имя фрейма-комплекса признаков для определено диагноз клинико-гематологических синдромов; AKO – имя слота Frame-класс; $NFID$ – слот с именем экзофрейма диагноза; $\{NFIS_i\}$ – слот набора имен экзофреймов признаков в комплексе признаков; $\{FR_j\}$ – слот, задающий набор НПП для диагноза; $\{SFI_k\}$ – слот, задающий множество слотов экзофреймов диагноза $NFID$ и признаков группы $\{NFIS_i\}$.

Frame-класс это фрейм верхнего уровня для представления совокупности фреймов образцов. Протофрейм используется для описания объектов диагностики заболеваний крови, в котором вложены протофрейм диагнозов, протофрейм групп диагнозов, протофрейм отдельных признаков, протофрейм групп признаков, протофрейм комплексов признаков для диагнозов.

Frame имеет определенную внутреннюю структуру, состоящую из множества элементов, называемых слотами, которым также присваиваются имена [94].

Нечетко-продукционное правило задается следующим образом (2.8):

$$FR = \langle NFR, \{\langle LVS_i, LTS_i \rangle\} \rightarrow \langle LVD, LTD \rangle, \omega \rangle \quad (2.8)$$

где: NFR – название нечетко-продукционного правила; LVS_i, LTS_i – лингвистическая переменная диагноза клинико-гематологических синдромов заключения нечетко-продукционного правила; ω – вес нечетко-продукционного правила.

Для обучение нечетко-продукционных базы знаний необходимо спроектировать общую структуру множества нечетко-продукционных правил, представленные в таблице 2.2.

Таблица 2.2 – Общая структура множества нечетко-продукционных правил

№	ЕСЛИ			ТО		
	X_1	... X_j ...	X_n	Y		ω
[1,1]	$s_1^{1,1}$	$s_j^{1,1}$	$s_n^{1,1}$	D_1	d_1^1	ω_1^1
[1,2]...	$s_1^{1,2}$	$s_j^{1,2}$	$s_n^{1,2}$		d_1^2	ω_1^2
[1, k_1]	s_1^{1,k_1}	s_j^{1,k_1}	s_n^{1,k_1}		$d_1^{k_1}$	$\omega_1^{k_1}$
[i ,1]	$s_1^{i,1}$	$s_j^{i,1}$	$s_n^{i,1}$	D_i	d_i^1	ω_i^1
[i ,2]...	$s_1^{i,2}$	$s_j^{i,2}$	$s_n^{i,2}$		d_i^2	ω_i^2
[i , k_i]	s_1^{i,k_i}	s_j^{i,k_i}	s_n^{i,k_i}		$d_i^{k_i}$	$\omega_i^{k_i}$
[m ,1]	$s_1^{m,1}$	$s_j^{m,1}$	$s_n^{m,1}$	D_m	d_m^1	ω_m^1
[m ,2]...	$s_1^{m,2}$	$s_j^{m,2}$	$s_n^{m,2}$		d_m^2	ω_m^2
[m , k_m]	s_1^{m,k_m}	s_j^{m,k_m}	s_n^{m,k_m}		$d_m^{k_m}$	$\omega_m^{k_m}$

где: $1, k_1$ – номер правила для диагноза D_1 ; $X_j, j = 1..n$ – входные показатели признаков; $Y = D_i, i = 1..m$ – выходные показатели диагнозов; s_j^{i, k_i} и $d_i^{k_i}$ – ЛТ входной переменной X_j и выходной переменной $Y = D_i$ в правиле с номером k_1 ; $\omega_i^{k_i}$ – весовой коэффициент правил с номером k_1 в соответствии с переменной диагноза $Y = D_1$.

При проектировании frame-базы создаются протофрейм, затем

экзофреймы, можно представить следующим образом:

- фрейм «ЖДА»: АКО – Фрейм-прототип диагноза; LVOD – <ЖДА, ...>; CVO – 0.65; { ω_i } – {0.8; ...}; SX – «Мужской и Женский»; AE – [0, 60]; FIM – «Гематолог»;

- фрейм «Слоятся ногти»: АКО – Фрейм-прототип признака; LVIS – <Слоятся ногти, ...>; CVI – 0.75; { SF_i } = {0.9; ...}; SG – «Внешние симптомы».

В качестве примера правил в комплексе признаков для диагноза «ЖДА» можно определить следующим образом:

Если (< X_i = "Слоятся ногти высокое_проявление"> И <Дисфагия – высокое_проявление > И ...)
ТО <Y= "ЖДА - высокое_обнаружение"> _с_ $\omega_i=0.8$

На рисунке 2.6 представлен фрагмент frame-структура базы знаний.

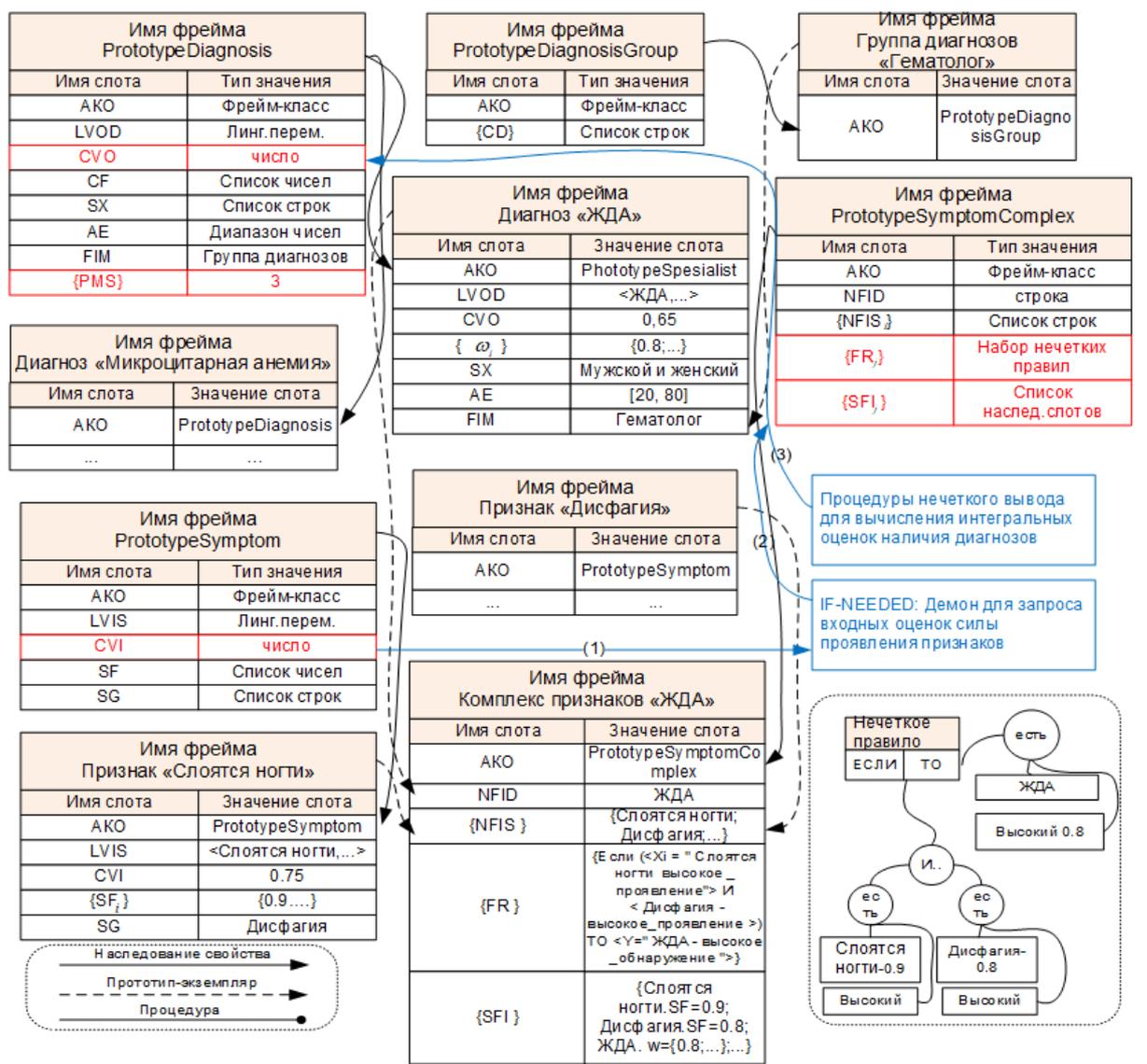


Рисунок 2.6 -Фрагмент frame-структура базы знаний

Для заполнения базы знаний эксперт выявляет связь основных групп заболеваний крови с признаками, показателями клинко-гематологических синдромов, группами признаков для диагнозов и т.д. в соответствии с алгоритмом на рисунке 2.7.

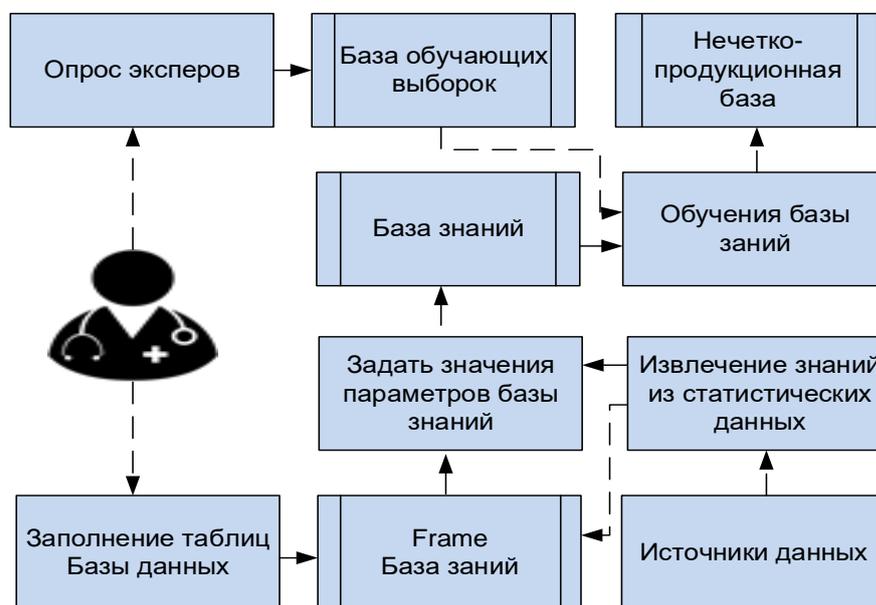


Рисунок 2.7 – Алгоритм заполнения базы знаний

Далее необходимо определить правила диагностирования клинко-гематологических синдромов с указанием входных и выходных параметров на основе метода нечеткой логики. В таблице 2.3 представлены значения интервалов функции принадлежности.

Таблица 2.3 – Соответствие числовых интервалов

№	Степень принадлежности	Числовой интервал
1	Низкий	$[0.0;0.2]$
2	Ниже среднего	$[0.2;0.37]$
3	Средний	$[0.37;0.63]$
4	Выше среднего	$[0.63;0.8]$
5	Высокий	$[0.8;1.0]$

Причинно-следственные отношения между лингвистическим переменными, которые формализуются в виде нечетких правил следующего типа:

«*R: ЕСЛИ A, ТО B (ω)*»,

где A – предпосылка, B – заключение, ω – весовой коэффициент правила.

Предложенная в диссертационном исследовании технология поддержки

принятия решений при диагностировании заболеваний крови система проверки базы знаний (СБЗ) спроектирована для выявления противоречий. Данная система проверки анализирует экспертные данные по алгоритму представленный на рисунке 2.8.

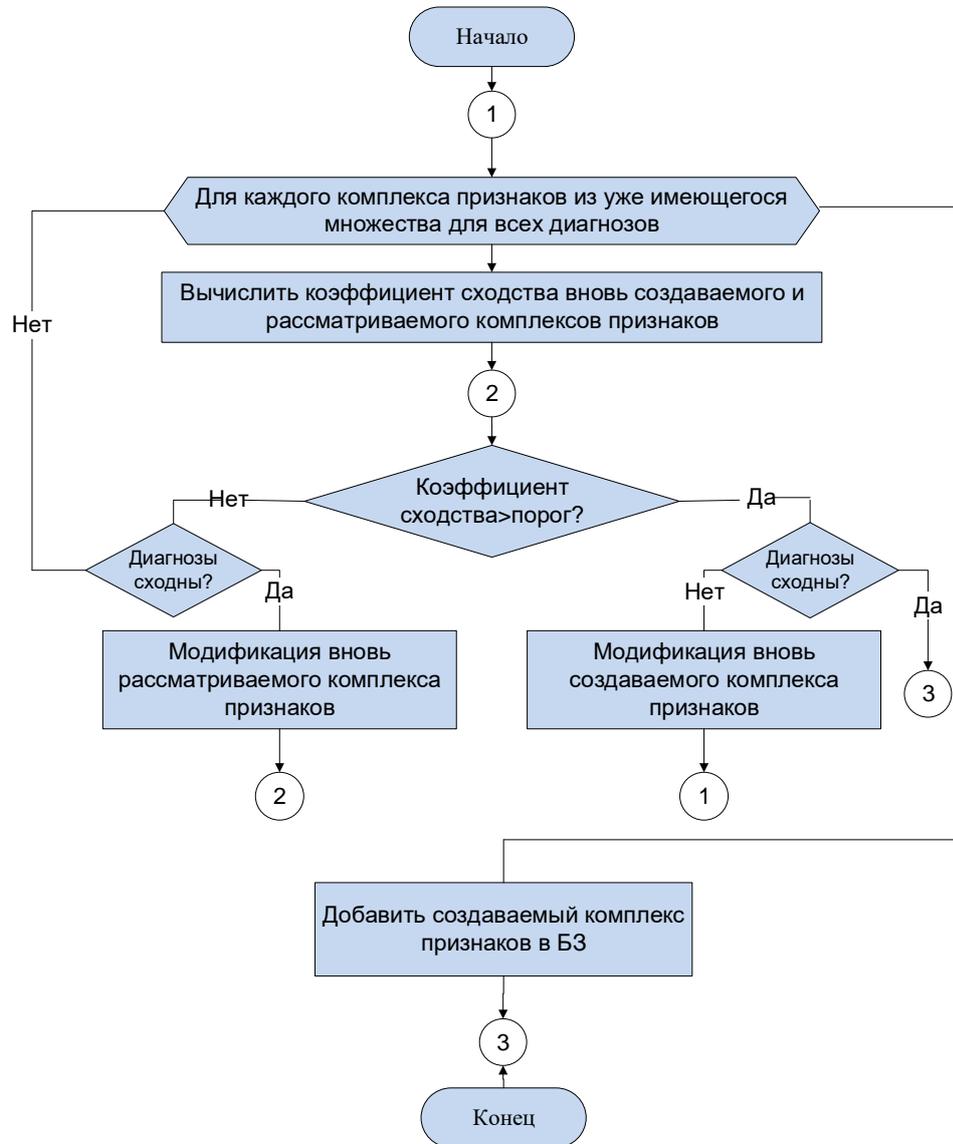


Рисунок 2.8 - Алгоритм поиска и исключения противоречий в базе знаний

Коэффициент сходства вычисляется по формуле (2.8):

$$K = [1/\sum_i^p(A_i) + 1/\sum_i^q(B_i)] \sum_i^r \min(A_i, B_i) / 2 \quad (2.8)$$

где: A_i – оценка, выставленная экспертом на проявления i -го признака в наборе признаков A ; B_i – оценка, выставленная экспертом на проявления i -го признака в наборе признаков B ; p – общее количество признаковых значений в наборе признаков A ; q – общее количество признаковых значений в наборе признаков B ; r – количество признаковых значений, общих в наборе признаков A и B .

Механизм вывода диагностического решения предназначен для получения диагностического заключения на основании входных оценок выявленных признаков [95]. Комбинированный механизм вывода решения реализуется следующим образом: дополнение базы показателей клинко-гематологических синдромов; подготовка заключение диагностирования. Дополнение базы показателей клинко-гематологических синдромов осуществляется по алгоритму, указанного на рисунке 2.9.

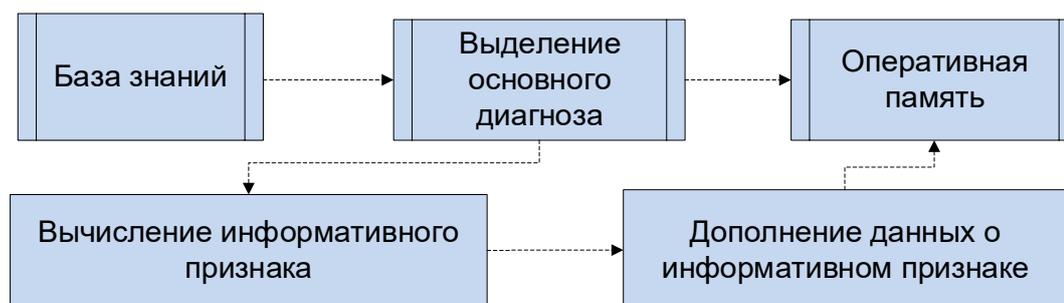


Рисунок 2.9 - Схема генерации дополнительных вопросов

Далее реализуется интерпретация решение на основе анализа протокола вывода. В этом модуле каждая запись соответствует диагностическому заключению на основании проявляющихся признаков. IR задается следующей моделью (2.9):

$$EM = \langle \{HR_i\}, FE \rangle \quad (2.9)$$

где: $\{HR_i\}$ – книга протоколов вывод, хранящий набор записей; FE – регистрация сведений о поведении системы на определенном шаге ее функционирования.

Запись книги протоколов вывода задается следующей моделью (2.10):

$$HR = \langle NR, \{SR_i\}, \{DHR_j\}, \{ARR_k\} \rangle \quad (2.10)$$

где: NR – номер записи книги протоколов вывода; $\{SR_i\}$ – массив проявляющихся признаков клинко-гематологических синдромов; $\{DHR_j\}$ – массив диагностических гипотез; $\{ARR_k\}$ – набор предложенных диагнозов клинко-гематологических синдромов.

Интерфейс пользователя ведется на ограниченном естественном языке, а также выполняет распределение ролей участников диалога между пользователями и ППРДЗК, и организацию их взаимодействия в процессе постановки диагноза, а также приобретения знаний.

Интерфейс пользователя задается следующим образом (2.11):

$$UI = \langle CDUS, CDSU, \{ \langle DG_i, RU_i \{ PD_j^i \} \} \rangle \rangle \quad (2.11)$$

где: *CDUS* – конвертирование входных данных пользователей в сообщения программного комплекса ППРДЗК; *CDSU* – конвертирование сообщений программного комплекса ППРДЗК в выходные данные пользователей; *DG_i*– диалог между *i*-го пользователем и программным комплексом ППРДЗК; *RU_i*– роль *i*-го пользователя; $\{PD_j^i\}$ – набор функций *i*-го пользователя системы.

Таким образом, предложенная технология поддержки принятия решений при диагностировании заболеваний крови описана в виде концептуальной модели, которая основана на оперативной памяти; базе знаний; интерфейсе пользователя; процедуре логического вывода, механизме объяснения решения; модуле приобретение знаний.

2.3 Разработка комплекса методов вычислительной ИТ-инфраструктуры технологии диагностирования заболеваний крови

Предлагаемая в диссертации информационная технология поддержки принятия решений при диагностировании заболеваний крови основана на ИТ-инфраструктуре диагностирования клиничко-гематологических синдромов. ИТ-инфраструктура включает вычислительный алгоритм морфологической классификации, метод оценки информативности диагностических показателей и гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клиничко-гематологических синдромов. Вычислительный алгоритм морфологической классификации основан на алгоритме диагностики анемий, классифицирующий показатели клиничко-гематологических синдромов по морфологическим признакам и предоставляющий количественную характеристику выраженности патологического процесса. Метод оценки информативности диагностических показателей основан на показателях информативности, доверии эксперта к набору признаков и размерностью клиничко-гематологических синдромов. Гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клиничко-гематологических синдромов используется для решения аналитических задач технологического процесса и получения более точных результатов

2.3.1 Вычислительный алгоритм морфологической классификации клинико-гематологических синдромов

Вычислительный алгоритм морфологической классификации показателей клинико-гематологических синдромов включает 3 основных блока, представленные на рисунке 2.10.



Рисунок 2.10. Общая структура алгоритма морфологической классификации показателей клинико-гематологических синдромов

На первом блоке производится проверка наличия и определение степени выраженности клинико-гематологических синдромов. Оценка наличия и степени выраженности анемического синдрома проводится по следующим показателям клинико-гематологических синдромов:

- уровень гемоглобина (HGB);
- количество эритроцитов (RBC);
- гематокрит (HCT);
- среднее содержание гемоглобина в эритроците (MCH);
- средний объем эритроцита (MCV);
- средняя концентрация гемоглобина в эритроците (MCHC).

В таблице 1 представлены описание и формулы вычисления некоторых показателей клинико-гематологических синдромов.

Таблица 2.4 - Математический аппарат проверки наличия и определение степени выраженности клинико-гематологических синдромов

Показатель	Описание	Формула
Гемоглобин (HGB)	В качестве нормального уровня гемоглобина, при котором отсутствует анемия, для мужчин и женщин определен уровень ≥ 125 g/l. В данном случае соответствие патологии равно 0. Патологические отклонения имеют место быть при уменьшении содержания HGB < 125 g/l, которые при значениях HGB ≤ 115 g/l представляют собой манифестированные формы анемии. Тогда соответствие патологии приравнивается 1,0.	В интервалах от 125 до 115 g/l принадлежность к патологии оценивается по формуле (2.23): $m_{HGB} = \frac{125 - X_p}{125 - 115}$ (2.23) где X_p – результат пробы пациента.
Гематокрит (HCT)	В качестве нормального уровня гематокрита для мужчин и женщин, при котором анемия отсутствует, определен уровень $\geq 0,38$. В этом случае соответствие патологии равно 0. При снижении HCT $< 0,38$ возможны патологические изменения, которые при значениях HCT $\leq 0,30$ представляют собой манифестированные формы анемии. Степень соответствия патологии приравнивается 1,0.	При значении показателя в интервале от 0,38 до 0,30 принадлежность к патологии оценивается по формуле (2.24): $m_{HCT} = \frac{0,38 - X_p}{0,38 - 0,3}$ (2.24)
Средняя концентрация гемоглобина в эритроците (MCHC)	В качестве нормального уровня значений, при котором отсутствует анемия - MCHC ≥ 32 g/dl. В этом случае степень соответствия патологии равен 0. Патологические отклонения имеют место быть при снижении MCHC ≤ 32 g/dl, которые при значениях MCHC = 28g/dl представляют собой манифестированные формы анемии. В данном случае соответствие патологии равно 1,0.	При значениях в интервале от 32 g/dl до 28 g/dl принадлежность к патологии оценивается по формуле (2.25): $m_{MCHC} = \frac{32 - X_p}{32 - 28}$ (2.25)

Продолжение таблицы 2.4.

Среднее содержание гемоглобина в эритроците (MCH)	В качестве нормального интервала определен интервал значений MCH от 27 до 34 pg. При значении, входящих в данный интервал степень соответствия патологии равен 0. При уменьшении значения MCH < 27 pg появляется степень принадлежности к патологии, которая приравнивается 1,0 при значениях MCH ≤ 18,5 pg.	При значениях в интервале от 27 pg до 18,5 pg принадлежность к патологии оценивается по формуле (2.26): $m_{MCH} = \frac{27 - X_p}{27 - 18,5} \quad (2.26)$
		При увеличении значения MCH выше 34 pg появляется степень принадлежности к патологии, степень которой равна 1,0 при значениях MCH 36,4 pg и более. В интервале от 34 pg до 36,4 pg степень принадлежности к патологии оценивается по формуле (2.27): $m_{MCH} = \frac{X_p - 34}{36,4 - 34} \quad (2.27)$
Средний объем эритроцита (MCV)	Степень соответствие патологии равна 0, если значение MCV находится в пределах нормального интервала, который определен зоной 80-95 fl. При уменьшении MCV < 80 fl появляется степень принадлежности к патологии, которая достигает 1,0 при значениях MCV ≤ 64 fl.	При значения в интервале от 80 fl до 64 fl степень принадлежности к патологии оценивается по формуле (2.28): $m_{MCV} = \frac{80 - X_p}{80 - 64} \quad (2.28)$
		При увеличении уровня среднего объема эритроцита более 95 fl появляется степень принадлежности к патологии. Степень принадлежности к патологии равна 1,0 при значениях MCV 129 fl и более. При значениях в интервале от 95 fl до 129 fl принадлежность к патологии оценивается по следующей формуле (2.29): $m_{MCV} = \frac{X_p - 95}{129 - 95} \quad (2.29)$
Эритроциты (RBC)	При уменьшении RBC < 4,0 × 10 ¹² /л появляется степень принадлежности к патологии (m), которая достигает 1,0 при RBC ≤ 3,5 × 10 ¹² /л.	В интервалах от 4,0 × 10 ¹² /л до 3,5 × 10 ¹² /л принадлежность к патологии оценивается по формуле (2.30): $m_{RBC} = \frac{4,0 - X_p}{4,0 - 3,5} \quad (2.30)$

После расчета степени принадлежности к патологии всех показателей производится расчет степени принадлежности к анемическому синдрому (M), который рассчитывается по следующей формуле (2.12):

$$M = 0,5 \cdot m_{HGB} + 0,1 \cdot m_{HCT} + 0,1 \cdot m_{MCHC} + 0,1 \cdot m_{MHC} + 0,1 \cdot m_{MCV} + 0,1 \cdot m_{RBC} \quad (2.12)$$

При значении $M \geq 0,5$ можно определить соответствие результатов анализа состоянию анемии. Значения $M > 0,2$ но $< 0,5$ соответствуют подозрению на анемию. Если значения M находятся в интервале от 0 до 0,2, то результаты анализа в норме [96].

На втором блоке реализации алгоритма определяется характер анемий в соответствии с морфологической классификацией. В таблице 2.5 представлены описание и формулы принадлежности показателей эритрона к микроцитарным анемиям.

Таблица 2.5 - Математический аппарат определения принадлежности показателей эритрона к микроцитарным анемиям

Показатель	Описание	Формула
Среднее содержание гемоглобина в эритроците (MCH)	При уменьшении MCH менее 27 pg появляется степень принадлежности к патологии, которая достигает 1,0 при значениях MCH 18,5 pg и менее.	В интервале от 27 pg до 18,5 pg принадлежность к патологии оценивается по формуле (2.32): $m_{MCH} = \frac{27 - X_p}{27 - 18,5} \quad (2.32)$
Средний объем эритроцита (MCV)	При уменьшении MCV менее 80 fl появляется степень принадлежности к патологии, которая достигает 1,0 при значениях MCV 64 fl и менее.	В интервале от 80 fl до 64 fl принадлежность к патологии оценивается по формуле (2.33): $m_{MCV} = \frac{80 - X_p}{80 - 64} \quad (2.33)$

Постановка диагноза дефицита железа и/или патологией его обмена ($M_{\text{микро}}$) возможна при многокритериальной оценки гемограммы в том случае, если значений вышеуказанных показателей с учетом их степени значимости и рассчитывается по формуле (2.13):

$$M_{\text{микро}} = 0,5 \cdot m_{MHC} + 0,5 \cdot m_{MCV} \quad (2.13)$$

где $M_{\text{микро}}$ - общий критерий соответствия; m_{MHC} , m_{MCV} - критерий соответствия отдельных показателей; 0,5 – степень значимости показателей.

При $M_{\text{микро}} > 0,5$ анемия является микроцитарной.

Далее определяется принадлежность показателей эритрона к макроцитарным анемиям таким как нарушения эритропоэза,

связанные с дефицитом витамина В12 и фолиевой кислоты.

В таблице 2.6 представлены описание и формулы принадлежности показателей эритрона к макроцитарным анемиям.

Таблица 2.6 - Математический аппарат определения принадлежности показателей эритрона к макроцитарным анемиям

Показатель	Описание	Формула
Среднее содержание гемоглобина в эритроците (MCH)	При увеличении $MCH > 34 \text{ pg}$ появляется степень принадлежности к патологии, которая достигает 1,0 при значениях $MCH = 36,4 \text{ pg}$.	В интервале от 34 pg до 36,4 pg принадлежность к патологии оценивается по формуле (2.13): $m_{MCH} = \frac{X_p - 34}{36,4 - 34} \quad (2.14)$
Средний объем эритроцита (MCV)	При увеличении $MCV > 95 \text{ fl}$ появляется степень принадлежности к патологии, которая достигает 1,0 при значениях $MCV \geq 129 \text{ fl}$.	В интервале от 95 fl до 129 fl принадлежность к патологии оценивается по формуле (2.14): $m_{MCV} = \frac{X_p - 95}{129 - 95} \quad (2.14)$

Постановка диагноза, связанного с дефицитом витамина В12 и фолиевой кислоты ($M_{\text{макро}}$) возможна при многокритериальной оценки гемограммы в том случае, если значений вышеуказанных показателей с учетом их степени значимости и рассчитывается по формуле (2.15):

$$M_{\text{макро}} = 0,5 \cdot m_{MCH} + 0,5 \cdot m_{MCV} \quad (2.15)$$

где $M_{\text{макро}}$ - общий критерий соответствия.

При $M_{\text{макро}} > 0,5$ анемия является к макроцитарной.

Далее определяется принадлежность показателей эритрона к нормоцитарным анемиям, таким как нарушения эритропоэза, связанные с низкой продукцией эритропоэтина, резистентность эритроидных клеток к эритропоэтину, перераспределительным железодефицитом.

В таблице 2.7 представлены описание и формулы принадлежности показателей эритрона к нормоцитарным анемиям.

Таблица 2.7 - Математический аппарат определения принадлежности показателей эритрона к нормоцитарным анемиям

Показатель	Описание	Формула
Среднее содержание гемоглобина в эритроците (MCH)	Показатель MCH при нормоцитарной анемии находится в пределах нормы, т.е. значения показателя 27-34 pg полностью соответствуют данному виду патологии (принадлежности к патологии 1,0).	В интервалах от 27 pg до 18,5 pg и от 34 до 38,0 принадлежность к патологии оценивается по формулам (2.16)-(2.17): $m_{MCH} = \frac{X_p - 18,5}{27 - 18,5} \quad (2.16)$ $m_{MCH} = \frac{36,4 - X_p}{36,4 - 34} \quad (2.17)$
Средний объем эритроцита (MCV)	Показатель MCV при нормоцитарной анемии находится в пределах нормы, т.е. значения показателя 90-95 fl полностью соответствуют данному виду патологии (принадлежности к патологии 1,0).	В интервалах от 90 fl до 80 fl и от 95 fl до 100 fl принадлежность к патологии оценивается по формулам (2.18)-(2.19): $m_{MCV} = \frac{X_p - 80}{90 - 80} \quad (2.18)$ $m_{MCV} = \frac{100 - X_p}{100 - 95} \quad (2.19)$

Постановка диагноза, связанного с анемиями при хронических заболеваниях (Мнормо), возможна при многокритериальной оценки гемограммы в том случае, если значений вышеуказанных показателей с учетом их степени значимости и рассчитывается по формуле (2.20):

$$M_{нормо} = 0,5 \cdot m_{MHC} + 0,5 \cdot m_{MCV} \quad (2.21)$$

где $M_{нормо}$ - общий критерий соответствия.

Т.о. представленный математический способ многокритериальной оценки результатов биохимического анализа крови позволяет количественно охарактеризовать их принадлежность к различным видам анемий, оценка может быть проведена при смешанных формах патологии. На анемии смешанного характера указывают не менее двух $M > 0,4$. Метод дает возможность выделить группу пациентов с подозрением на анемию ($0,2 > M > 0,5$) [96].

В третьем блоке алгоритма морфологической классификации клинико-гематологических синдромов производится подтверждение и уточнение характера анемии с помощью методов клинической биохимии и иммуноферментного анализа.

Данные, которые были получены методами клинической биохимии и иммуноферментного анализа, позволяют уточнить наличие и степень выраженности анемии, а также уточнить характер патологии, выявить наиболее распространенные виды анемий. Это относится в первую очередь для диагнозов ЖДА из группы микроцитарных анемий. Для группы

макроцитарных анемий выявляются В12-дефицитные анемии. Для группы нормоцитарных анемий устанавливается вероятность АХЗ. Для повседневного использования рекомендуется тесты доступные лабораториям различного уровня и входящие в большинство известных алгоритмов диагностики. Такими лабораторными показателями являются определение ферритина (Fer), и определение витамина В12 [97]. Расчет степени принадлежности значений этих показателей к патологии при микроцитарной, макроцитарной и нормоцитарной анемиях должен проводиться по формулам, представленных в таблице 2.8.

Таблица 2.8. Расчет степени принадлежности значений этих показателей к патологии при микроцитарной, макроцитарной и нормоцитарной анемиях

Тип анемии	Показатель	Формула
Для анемии при хронических заболеваниях (нормоцитарные)	Ферритин (Fer)	При значениях Fer >60 мкг/л принадлежность к патологии – 1, в интервале от 40 до 60 мкг/л принадлежность к патологии оценивается по формуле (2.22): $M_{АХЗ} = \frac{X_p - 40}{60 - 40} \quad (2.22)$
Для железодефицитной анемии (микро- и нормоцитарные)	Ферритин (Fer)	При значениях Fer >40 мкг/л принадлежность к патологии – 0, при значениях Fer <20 мкг/л принадлежность к патологии – 1, в интервале от 20 до 40 мкг/л принадлежность к патологии оценивается по формуле (2.23): $M_{ЖДА} = \frac{40 - X_p}{40 - 20} \quad (2.23)$
Для В12-дефицитных анемий (макроцитарные)	Витамин В12	При значениях В12 > 400 нг/мл принадлежность к патологии – 0, при значениях В12 < 100 нг/мл принадлежность к патологии – 1, в интервале от 100 до 400 нг/мл принадлежность к патологии оценивается по формуле (2.24): $M_{В12} = \frac{400 - X_p}{400 - 100} \quad (2.24)$

Таким образом, предложенный в диссертационном исследовании, вычислительный алгоритма морфологической классификации клинко-гематологических синдромов выполняет две основные функции:

- классификация анемий по морфологическим признакам;
- количественная оценка характера выраженности патологического процесса.

Данный вычислительный алгоритма морфологической классификации клинко-гематологических синдромов позволяет с высокой долей вероятности выявить ряд следующих часто встречающихся анемий: ЖДА, АХЗ, В12-дефицитная, гемолитическая анемия. Использованием данного

алгоритма позволит отказаться от выполнения дорогостоящих биохимических и иммуноферментных исследований показателей клинико-гематологических синдромов. Этапы диагностики определяются на основе количественной оценки принадлежности состояния эритрона к анемии, определенным видам патологий по морфологическими показателям. При значениях соответствия анемии $\geq 0,5$ постановка наличия анемии является достоверным. Заключение о ее морфологическом характере диагноза, т.е. отнесение его к микроцитарной, макроцитарной, нормоцитарной или смешанной возможно при критериях соответствия той или иной форме анемии.

2.3.2 Метод оценки информативности диагностических показателей клинико-гематологических синдромов

При проектировании диагностических систем возникает вопросы подбора адекватного набора признаков, характеризующих альтернативные классы. Для этого пусть задано набор клинико-гематологических синдромов $A: \{a_1, a_2, \dots, a_n\}$, каждый элемент которого характеризуются множеством признаков $a \rightarrow \{v_1, v_2, \dots, v_m\}$. Каждый признак является номинальной величиной и может принимать одно из предопределенных значений (2.25):

$$v_{i(r)} \in V_i: \{v_{i(1)}, v_{i(2)}, \dots, v_{i(s_i)}\} \quad (2.25)$$

где i — номер признака; r — номер значения признака; V_i — множество уникальных значений признака v_i , которые встречаются в множестве элементов A ; s_i — количество элементов в множестве V_i ; $v_{i(r)}$ — значение номер r признака V_i .

Вероятность появления значения $v_{i(r)}$ в произвольном элементе составляет (2.26):

$$p_{i(r)} = \frac{n_{i(r)}}{n} \quad (2.26)$$

где $n_{i(r)}$ — количество элементов, у которых признак имеет значение $v_{i(r)}$. Отметим, что для всех i верны утверждения (2.27):

$$\sum_{r=1}^{s_i} n_{i(r)} = n \quad \sum_{r=1}^{s_i} p_{i(r)} = 1 \quad (2.27)$$

Энтропию каждого признака можно представить выражением следующего вида (2.28):

$$H_i = - \sum_{r=1}^{s_i} p_{i(r)} \log_2 p_{i(r)} \quad (2.28)$$

Значение энтропии H_i будет находиться в некотором диапазоне $0 \leq H_i \leq$

H_{max} , причем $H_i = 0$ соответствует ситуации с нулевой дисперсией, когда v_i признак у всех элементов имеет одинаковое значение. Своего максимального значения $H(i = H_{max_i})$ энтропия достигает в случае с максимальной дисперсией, т.е. когда каждое значение признака встречается равное количество раз [98].

Далее необходимо определить взаимное влияние признаков. Для каждой пары признаков и строим двумерную переменную, представленную в таблице 2.9, где МР – маргинальное распределение.

Таблица 2.9 - Значения двумерной номинальной переменной

Значения признака v_i	Значения признака v_j				МР v_j
	$v_{j(1)}$	$v_{j(2)}$...	$v_{j(s_j)}$	
$v_{i(1)}$	$n_{i(1),j(1)}$	$n_{i(1),j(2)}$...	$n_{i(1),j(s_j)}$	$n_{i(1)}$
$v_{i(2)}$	$n_{i(2),j(1)}$	$n_{i(2),j(2)}$...	$n_{i(2),j(s_j)}$	$n_{i(2)}$
...
$v_{i(s_j)}$	$n_{i(s_j),j(1)}$	$n_{i(s_j),j(2)}$...	$n_{i(s_j),j(s_j)}$	$n_{i(s_j)}$
МР v_i	$n_{j(1)}$	$n_{j(2)}$...	$n_{j(s_j)}$	n

Вероятность появления в элементе комбинации значений будем рассчитывать по формуле (2.29):

$$p_{i(r),j(q)} = \frac{n_{i(r),j(q)}}{n} \quad (2.29)$$

где $n_{i(r),j(q)}$ – количество элементов, у которых признаки i и j имеют значения $v_{i(r)}$ и $v_{j(q)}$. Соответственно энтропию для пары этих признаков можно вычислить по формуле (2.30):

$$H_{ij} = - \sum_{r=1}^{s_i} \sum_{q=1}^{s_j} p_{i(r),j(q)} \log_2 p_{i(r),j(q)} \quad (2.30)$$

Интервал значений лежит в диапазоне $[0; H_i + H_j]$. В случае полной связи, т.е. когда $s_i = s_j$ и когда каждому признаку $v_{i(r)}$ соответствует строго определенный признак $v_{j(q)}$, имеем по формуле (2.31):

$$H_{ij} = H_i = H_j \quad (2.31)$$

При статистической независимости, когда связь между признаками отсутствует, получаем: (2.32):

$$H_{ij} = H_i + H_j \quad (2.32)$$

Условная энтропия $(H_{j;i})$ (2.33) показывает, какая часть энтропии, остается, если становится известно значение признака v_i .

$$H_{j;i} = H_{ij} - H_i \quad (2.3354)$$

В случае, когда выполняется условие $H_{ij} = H_i = H_j$ условная энтропия равна нулю. Это означает, что вся информация о значениях признака v_j содержится в признаке v_i . При независимости признаков, т.е. если выполняется условие $H_{ij} = H_i + H_j$, условная энтропия $H_{j;i}$ будет равна всей энтропии H_i признака v_j . Оценить часть энтропии признака v_i , которая будет объясняться значением признака v_j , позволяет следующее выражение (2.34) [98]:

$$H_i - H_{j;i} = H_i + H_j - H_{ij} = H_i - H_{i;j} = h_{ij} \quad (2.34)$$

Относительный частный коэффициент влияния v_i на v_j можно тогда определить как отношение (2.35):

$$\frac{h_{ij}}{H_i} \quad (2.35)$$

Аналогично, частный коэффициент влияния v_j на v_i будет соответствовать (2.36):

$$\frac{h_{ij}}{H_j} \quad (2.36)$$

Рассчитаем средневзвешенное значение этого параметра для признака v_i относительно всех прочих признаков, иными словами для всех $j \neq i$, по формуле (2.37):

$$\left\{ \frac{h_{ij}}{H_{ij}} \right\} = \frac{\sum_{j \neq i} \frac{h_{ij}}{H_{ij}} H_{ij}}{\sum_{j \neq i} H_{ij}} = \frac{\sum_{j \neq i} h_{ij}}{\sum_{j \neq i} H_{ij}} = I_i \quad (2.37)$$

На рисунке 2.11 представлен алгоритм реализации метода оценки информативности диагностических показателей клинико-гематологических синдромов.

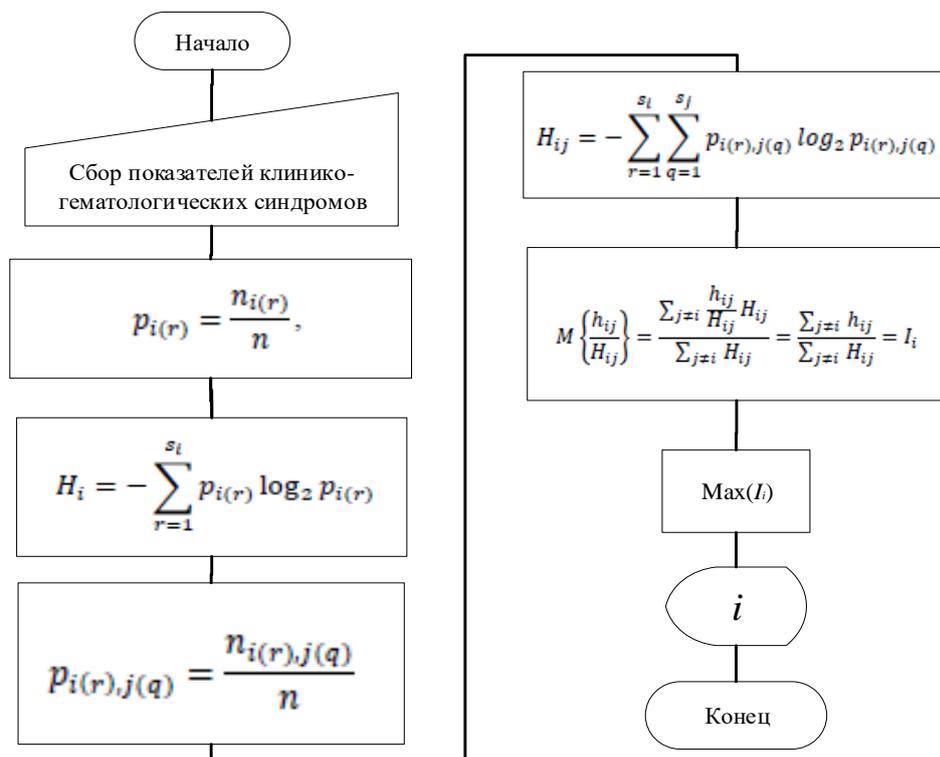


Рисунок 2.11. Алгоритм реализации метода оценки информативности диагностических показателей клинико-гематологических синдромов

Таким образом, полученная величина I_i может служить мерой информативности признака v_i относительно остальных признаков. Чем выше значение I_i , тем больший объем совокупной информации несет в себе признак v_i о значениях признаков $v_{j \neq i}$. Признак, обладающий наибольшим значением I_i , является наиболее информативным признаком.

2.3.3 Гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов

Алгоритм Stacking является одним из видов ансамблирования методов интеллектуального анализа данных. Применение ансамблей моделей к решению различных задач анализа открывает широкие возможности для повышения эффективности моделей Data Mining. В связи с этим сейчас проводятся активные исследования по использованию большого количества различных методов и алгоритмов формирования ансамблей. Самыми распространёнными являются следующие методы ансамблирования моделей: - bagging; - stacking; - boosting [99].

Алгоритм Stacking не основывается на математической модели. Его идея заключается в использовании в качестве базовых моделей различных классификационных алгоритмов, обучаемых на одинаковых данных. Затем, мета-классификатор обучается на исходных данных, дополненных результатами прогноза базовых алгоритмов. Иногда мета-классификатор

использует при обучении не результаты прогноза базовых алгоритмов, а полученные ими оценки параметров распределения, например, оценки вероятностей каждого класса [100].

Обобщенная схема реализации алгоритма stacking представлена на рисунке 2.12.

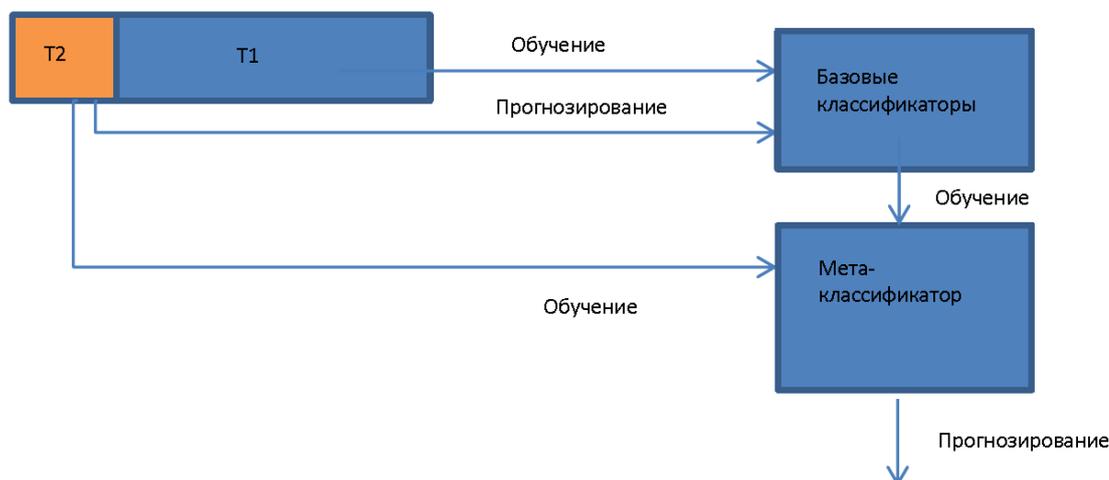


Рисунок 2.12 - Схема реализации алгоритма stacking

Идея Stacking основана на мета-алгоритме классификации базовых алгоритмов по принципу доверия на определенных областях входных данных [99].

Математическое описание алгоритма Stacking для задачу классификации показателей клиничко-гематологических синдромов имеет следующее описание: X — множество описаний объектов, Y — конечное множество номеров (имён, меток) классов. Существует неизвестная "целевая зависимость" — отображение $f: X \rightarrow Y$ значения которой известны только на объектах конечной обучающей выборки $(X, Y) = \{(x_1, y_1), \dots, (x_{|(X,Y)|}, y_{|(X,Y)|})\}$.

Требуется построить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$. Используются обозначения, представленные в таблице 2.10.

Идея стэкинга состоит в том, чтобы обучить метаклассификатор M на (1) исходных признаках, матрице X , и (2) на предсказаниях (метапризнаках), полученных с помощью базовых классификаторов. Метапризнаки, полученные с помощью классификатора A для выборки X будем обозначать $MF(X, A)$.

На рисунке 2.13 представлена схема реализации алгоритма stacking.

Таблица 2.10 – Обозначения и их описания

Обозначение	Описание
(X_0, Y_0)	валидационная выборка
A	базовый классификатор, использующийся для построения метапризнака
$A.fit(X, Y)$	функция обучения классификатора A на (X, Y)
$A.predict(X)$	функция, предсказывающая целевую переменную для X классификатором A
M	некоторый метаклассификатор
$MF(X, A)$	метапризнак, полученный классификатором A для выборки X
P	финальное предсказание стекинга для валидационной выборки
$concatV(X_i, X_j)$	операция конкатенирования X_i и X_j по столбцам;
$concatH(X_i, X_j)$	операция конкатенирования X_i и X_j по строкам

Далее представлен простейший алгоритм stacking для предсказания выборки P, основанный на делении обучающей выборки (X, Y) на 2 части: (X_1, Y_1) и (X_2, Y_2) .

$A.fit(X_1, Y_1)$
 $MF(X_2; A) := A.predict(X_2)$
 $MF(X_0; A) := A.predict(X_0)$
 $M.fit(concatV(X_2; MF(X_2; A)); Y_2)$
 $P := M.predict(concatV(X_0; MF(X_0; A)))$

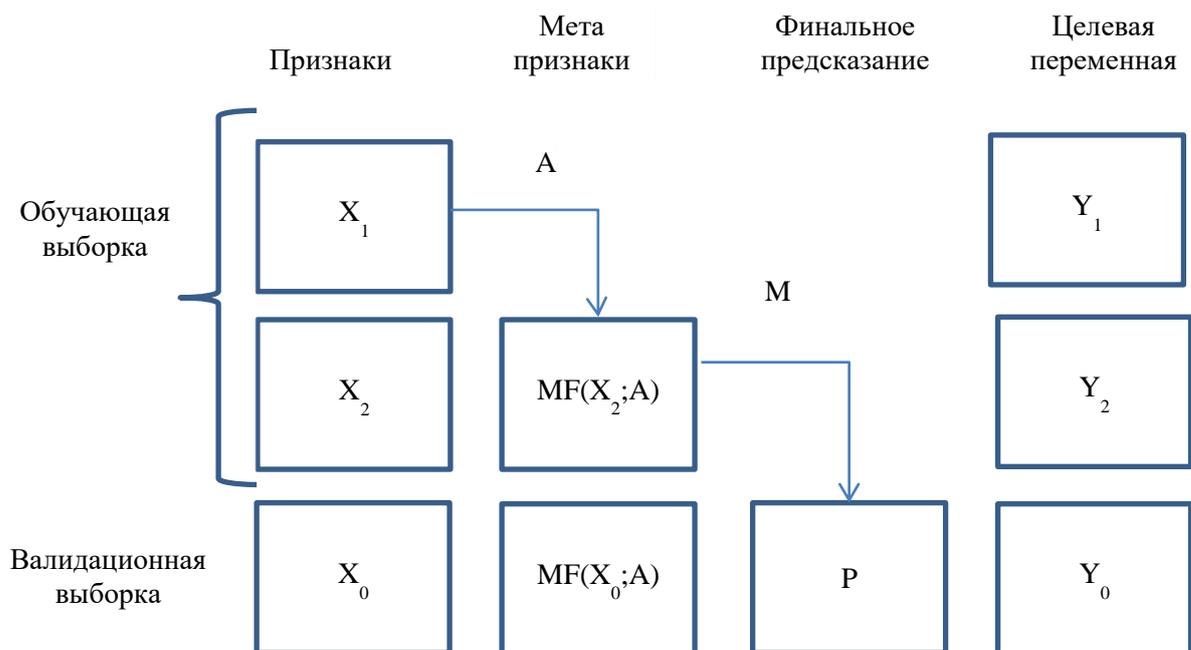


Рисунок 2.13 - Схема реализации алгоритма stacking-ансамблирования

Данный алгоритм является менее эффективным, из-за того что метаклассификатор обучается лишь на второй части выборки P . Для повышения эффективности предсказания можно применить этот алгоритм несколько раз, используя различные разбиения и затем усреднить полученные предсказания. Алгоритм данного подхода выглядит следующим образом:

Для $n=1 \dots N$
 $A_n := A$
 $A_n.fit(X_{n1}; Y_{n1})$
 $MF(X_{n2}; A_n) := A_n.predict(X_{n2})$
 $MF(X_0; A_n) := A_n.predict(X_0)$
 $M_n := M$
 $M_n.fit(concatV(X_{n2}; MF(X_{n2}; A_n)); Y_2)$
 $P_n := M_n.predict(concatV(X_0; MF(X_0; A_n)))$

$$P := \frac{1}{N} \sum_{n=1}^N P_n$$

Далее представим модификацию второго алгоритма, позволяющий добавить новый базовый классификатор для обучения множества A . Для этого необходимо получить мета-признак $MF(X, A)$ для всей обучающей выборки. Алгоритм данного подхода выглядит следующим образом:

$A1 := A$
 $A1.fit(X1; Y1)$
 $MF(X2; A1) := A1.predict(X2)$
 $MF(X0; A1) := A1.predict(X0)$
 $A2 := A$
 $A2.fit(X2; Y2)$
 $MF(X1; A2) := A2.predict(X1)$
 $MF(X0; A2) := A2.predict(X0)$

 $MF(X; A) := concatH(MF(X1; A2); MF(X2; A1))$
 $MF(X0; A) := (MF(X0; A1) + MF(X0; A2)) / 2$

 $M.fit(concatV(X; MF(X; A)); Y)$
 $P := M.predict(concatV(X0; MF(X0; A)))$

Таким образом, представленный модифицированный алгоритм Staking, который позволяет метаклассификатору M использовать для обучения всю выборку X . Принятым также является представление стекинга в виде многоуровневой схемы, где признаки обозначаются как "уровень 0", метапризнаки, полученные с помощью обучения базовых классификаторов на признаках, как "уровень 1", и так далее.

Для построения ансамбля моделей на данных клинико-гематологических синдромов был использован алгоритм Stacking, который агрегировал следующие 3 типа интеллектуальных моделей: алгоритм Байеса (БК), алгоритм дерева решений (ДР) и нейронная сеть (НС). Обучение производилось на единственном наборе данных. Данные метауровня, используемые для обучения метамодели, будут представлять собой результаты предсказания моделей для каждого из полей. Для возможности использования в качестве метамодели БК, включим в данные метауровня поля класса числом для БК, ДР и НС. Таким образом, алгоритм stacking пытается обучить каждый классификатор, используя алгоритм метаобучения, который позволяет обнаружить лучшую комбинацию выходов базовых моделей.

Структурная схема ансамбля на основе stacking представлена на рисунке 2.14.

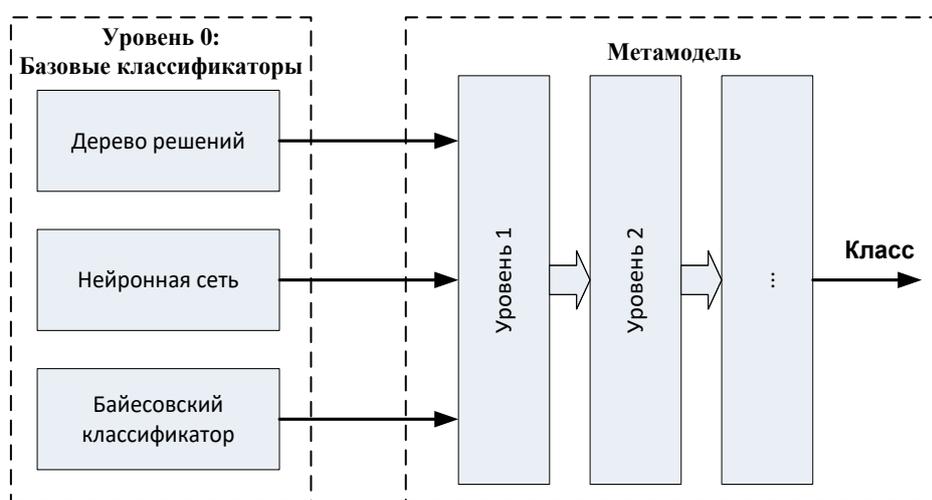


Рисунок 2.14 - Структурная схема ансамбля на основе stacking

Базовые модели образуют уровень 0. На вход метамодели, которая также называется моделью уровня 1, подаются результаты с выходов базовых моделей. Пример уровня 1 имеет столько атрибутов, сколько существует моделей уровня 0, а сами значения атрибутов выходы моделей нулевого уровня. Затем на основе результатов, полученных моделью уровня 1, строится модель уровня 2 и т. д., пока не будет выполнено какое либо из условий остановки обучения [100].

На рисунке 2.15 представлена схема алгоритма ансамблирования моделей БК, ДР и НС на основе stacking. Наиболее общим способом уменьшить перекосы, вызванные неудачным сэмплингом, является повторение всего процесса обучения и тестирования несколько раз при различных случайных выборках. На каждой итерации определенная доля данных (например, 2/3) случайно отбирается для обучения, а оставшиеся используются для тестирования. Для этого необходимо определить число итераций разбиения исходных данных. Далее задается вектор, в который собираются данные для обучения метамодели.

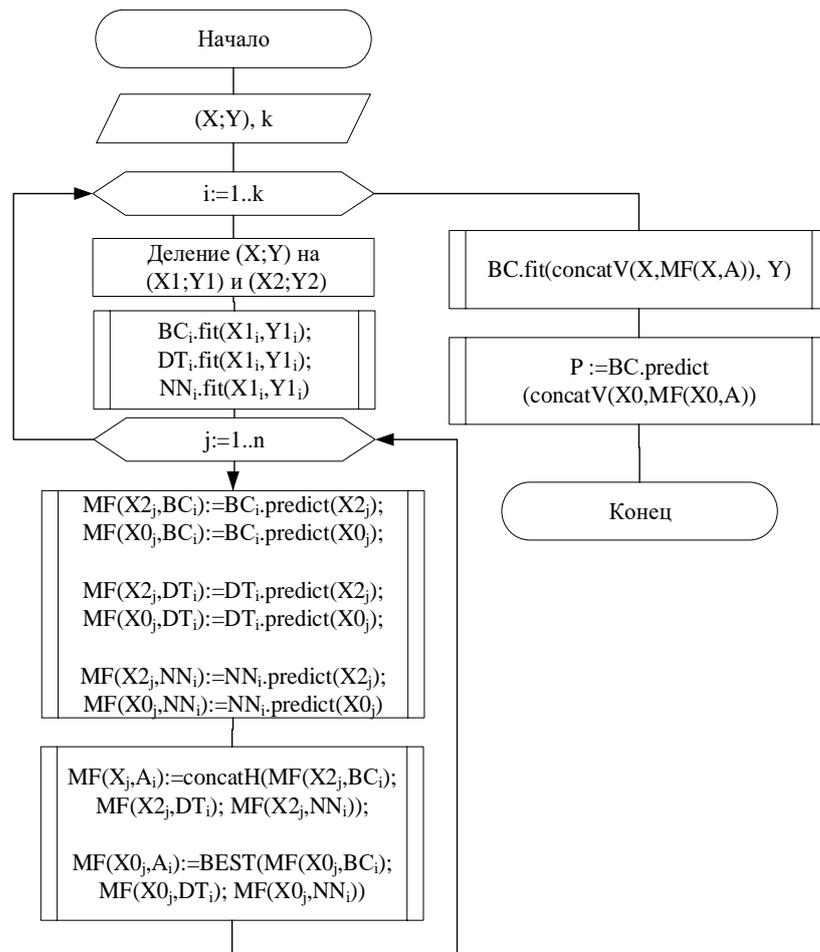


Рисунок 2.15 - Схема алгоритма ансамблирования моделей БК, ДР и НС на основе stacking

Для каждой итерации необходимо произвести следующие шаги:

Шаг 1: Поделить исходные данные на 2 части: первая часть предназначена для обучения моделей, вторая - для проверки прогноза моделей. Пропорция деления исходных данных выбирается исходя из того, чтобы как все исходные данных после всех итераций попали либо в 1-ю либо во 2-ю часть. Для реализации этого сделал несколько структур интеллектуального анализа в Analysis Services (в названии добавлен номер варианта). Вариативность выборки задается через свойство HoldoutSeed. Оно разное у разных вариантов.

Шаг 2: Обучаем три исходные модели на первой части данных.

Шаг 3: Для каждой строки данных второй части необходимо:

- реализовать прогноз по каждой модели;
- из прогнозных значений выбрать наиболее точный;
- записать строку с данными и прогнозом в вектор. Если в векторе уже имеется эта строка данных, то для нее записывается более точный прогноз из имеющегося и полученного из шага 3.

Шаг 4. После реализации всех итераций, выбирается алгоритм построения метамодели. Обучение метамодели производится на основе данных, собранные в векторе.

Вычислительный эксперимент ИТ-инфраструктуре диагностирования клинико-гематологических синдромов реализован на данных клинико-гематологических синдромов полученных из MIMIC-III (Medical Information Mart for Intensive care III) [101], которая является большой и свободно доступной базой данных, содержащая идентифицированные данные о здоровье, связанные с более чем сорока тысячами пациентов, которые оставались в отделениях интенсивной терапии Медицинского центра Beth Israel Deaconess с 2001 по 2016 год.

Достоверность проведенных исследований и полученных результатов будет обеспечена аргументацией выдвигаемых положений, обоснованным и корректным применением математического аппарата, результатами вычислительного эксперимента и результатами использования разработок.

2.4 Выводы по второму разделу

1. Разработана функциональная модель диагностирования клинико-гематологических симптомов является точная спецификация всех операций и действий, осуществляемых в процессе диагностирования, а также характера взаимосвязей между ними. Модель обеспечивает полное представление, как о функционировании исследуемого процесса работы, так и обо всех имеющих в нём место потоках информации и материалов. функциональная модель описывает потоки данных следующих основных подпроцессов: «Установление наличия и степени выраженности анемического синдрома», «Определение характера анемии в соответствии с морфологической классификацией»; «Подтверждение и уточнение типа анемии».

2. Разработана концептуальная модель информационной технологии поддержки принятия решений при диагностировании заболеваний крови, основанная на модуле оперативной памяти, базы знаний, модуль интерфейса пользователя, процедуры логического вывода, а также модуле наполнения и проверки базы знаний.

3. Разработана ИТ-инфраструктура диагностирования клинико-гематологических синдромов, которая включает вычислительный алгоритм морфологической классификации, метод оценки информативности диагностических показателей и гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов.

4. Разработанный вычислительный алгоритм морфологической классификации основан на алгоритме диагностики анемий, позволяющий классифицировать анемии по морфологическим признакам и дать количественную характеристику выраженности патологического процесса. При помощи предложенных алгоритмов можно с высокой долей вероятности выявить ряд наиболее часто встречающихся анемий (железодефицитные анемии, анемии хронических заболеваний, В12-дефицитную анемию). Вычислительный алгоритм включает 3 основных этапа диагностики анемии:

установление наличия и степени выраженности анемического синдрома; определение характера анемий в соответствии с морфологической классификацией; подтверждение и уточнение характера анемии методами клинической биохимии и иммуноферментного анализа.

5. Метод оценки информативности диагностических показателей клинико-гематологических синдромов основан на вычислении энтропийных показателей множества данных клинико-гематологических синдромов и значения их информативности. Показатель, имеющий максимальное значение, является наиболее информативным по отношению к остальным.

6. Гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов используется для решения аналитических задач технологического процесса и получения более точных результатов диагностирования. Алгоритм объединяет усилия нескольких интеллектуальных методов на основе создания ансамбля моделей-предикторов. Для построения ансамбля моделей на данных показателей КГС был использован алгоритм Stacking, который агрегировал следующие 3 типа интеллектуальных моделей: алгоритм Байеса (БК), алгоритм дерева решений (ДР) и нейронная сеть (НС).

3 АРХИТЕКТУРА И РЕАЛИЗАЦИЯ ПРОГРАММНОГО КОМПЛЕКСА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ

3.1 Общее описание и обоснование программного инструментария системы поддержки принятия решений

Предложенная в рамках диссертационного исследования, информационная технология, была реализована в виде программного комплекса поддержки принятия решений при диагностировании заболеваний крови (ППРДЗК), обеспечивающий следующие основные возможности:

- интегрирование frame структуры и НПВ для представления знаний по дифференциальной диагностике заболеваний крови;
- проектировать, изменять и масштабировать базу знаний на основе комплексного использования экспертных методов;
- предоставлять диагностические решения по дифференциальной диагностике клиничко-гематологических синдромов;
- реализация вычислительного алгоритма морфологической классификации клиничко-гематологических синдромов;
- оценка информативности диагностических показателей клиничко-гематологических синдромов;
- взаимодействие основных потребителей услуг дифференциальной диагностики клиничко-гематологических синдромов в режиме удаленного доступа;
- разграничение прав доступа пользователей основных потребителей услуг дифференциальной диагностики и администратора.

При проектировании ППРДЗК использовался комплекс программного инструментария, включающий оболочки экспертных систем; языки программирования и инструментальные системы разработки программ.

В результате изучения существующих ИТ-решений были выявлены основные недостатки и преимущества. К сожалению, существующие ИТ-решения не могут обеспечить вышеуказанный функционал, позволяющий повысить эффективность бизнес-процессов дифференциальной диагностики показателей клиничко-гематологических синдромов.

Для эффективной реализации предложенной информационной технологии было решено проектировать программный комплекс на основе WEB-приложений с использованием языка программирования высокого уровня. WEB-приложение клиент-серверное приложение, в котором клиент взаимодействует с веб-сервером при помощи браузера. Логика веб-приложения распределена между сервером и клиентом, хранение данных осуществляется, преимущественно, на сервере, обмен информацией происходит по сети. Одним из преимуществ такого подхода является тот факт, что клиенты не зависят от конкретной операционной системы пользователя, поэтому WEB-приложения являются межплатформенными службами. Для реализации такого подхода

используются протокол передачи данных HTTP и стандарты описания интерфейса HTML/CSS с применением схемы «запрос-ответ».

В качестве технологии разработки предложена WEB-приложения выбрана ASP.NET (Active Server Pages для .NET). ASP.NET— платформа разработки веб-приложений, в состав которой входит: веб-сервисы, программная инфраструктура, модель программирования от компании Майкрософт [102]. ASP.NET входит в состав платформы .NET Framework[2] и является развитием более старой технологии Microsoft ASP.

Для развертывания WEB-приложения выбрана платформа ASP.NET MVC. ASP.NET MVC Framework — фреймворк для создания веб-приложений, который реализует шаблон Model-view-controller [103].

В качестве СУБД выбрана система Microsoft SQL Server, позволяющая управлять реляционными базами данных и использующий язык запросов — Transact-SQL. MS SQL Server – это платформа для решения критически важных задач в масштабе предприятия, обладающая высокой доступностью, повышенной производительностью и безопасностью [104]. Решение представляет собой хорошо масштабируемый, полностью реляционный, быстродействующий сервер, способный обрабатывать большие объемы данных для клиент-серверных приложений.

Для реализации доступа потребителей к WEB-страницам в программном комплексе ППРДЗК используются службы интернета IIS, являющейся проприетарным набором серверов для нескольких служб Интернета от компании Microsoft.

Таким образом, для проектирования ППРДЗК предложено использовать среду Microsoft Visual Studio 2019 и язык программирования C#. Для хранения и обработки данных предложен сервер БД Microsoft SQL Server 2019. Приложения реализованы с помощью технологий платформ .NET Framework и веб-сервера IIS.

3.2 Проектирование архитектуры системы поддержки принятия решений

Для проектирования комплекса поддержки принятия решений при диагностировании заболеваний крови предложена архитектура, основанная на интеграции трех архитектурных шаблонов [105]:

- архитектурная модель object-oriented, объединяющая набор взаимодействующих объектов, в котором объект содержит данные и необходимое поведение, а коммуникация между объектами происходит через открытые интерфейсы путём отправки и приёма сообщений;
- архитектурная модель three-tier, объединяющая три звена: клиента, сервера приложений (к которому подключено клиентское приложение) и сервера баз данных (с которым работает сервер приложений);
- компонентная архитектурная модель, основанная на применении

компонентов, обладающих следующими свойствами: пригодность для повторного использования; замещаемость; независимость от контекста; расширяемость; инкапсуляция; независимость.

Для проектирования ППРДЗК трех шаблонная архитектура дает возможности, представленные на рисунке 3.1.

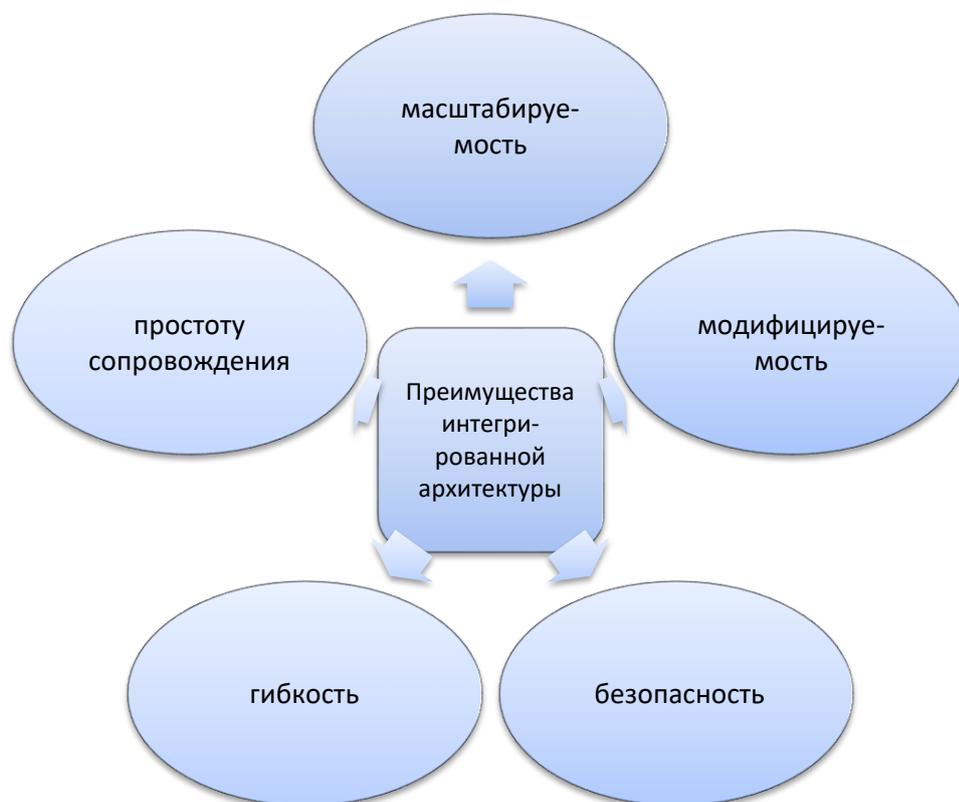


Рисунок 3.1 - Преимущества интегрированной архитектуры

Схема архитектурного решения программного комплекса ППРДЗК представлена на рисунке 3.2.

Web-браузеры осуществляют функцию сбора для дальнейшей обработки, а также для визуализации проанализированных данных дифференциальной диагностики показателей клинико-гематологических синдромов. В качестве узлов сбора данных могут выступать любые ИКТ устройства (компьютеры, ноутбуки, планшеты, сотовые телефоны и др.)

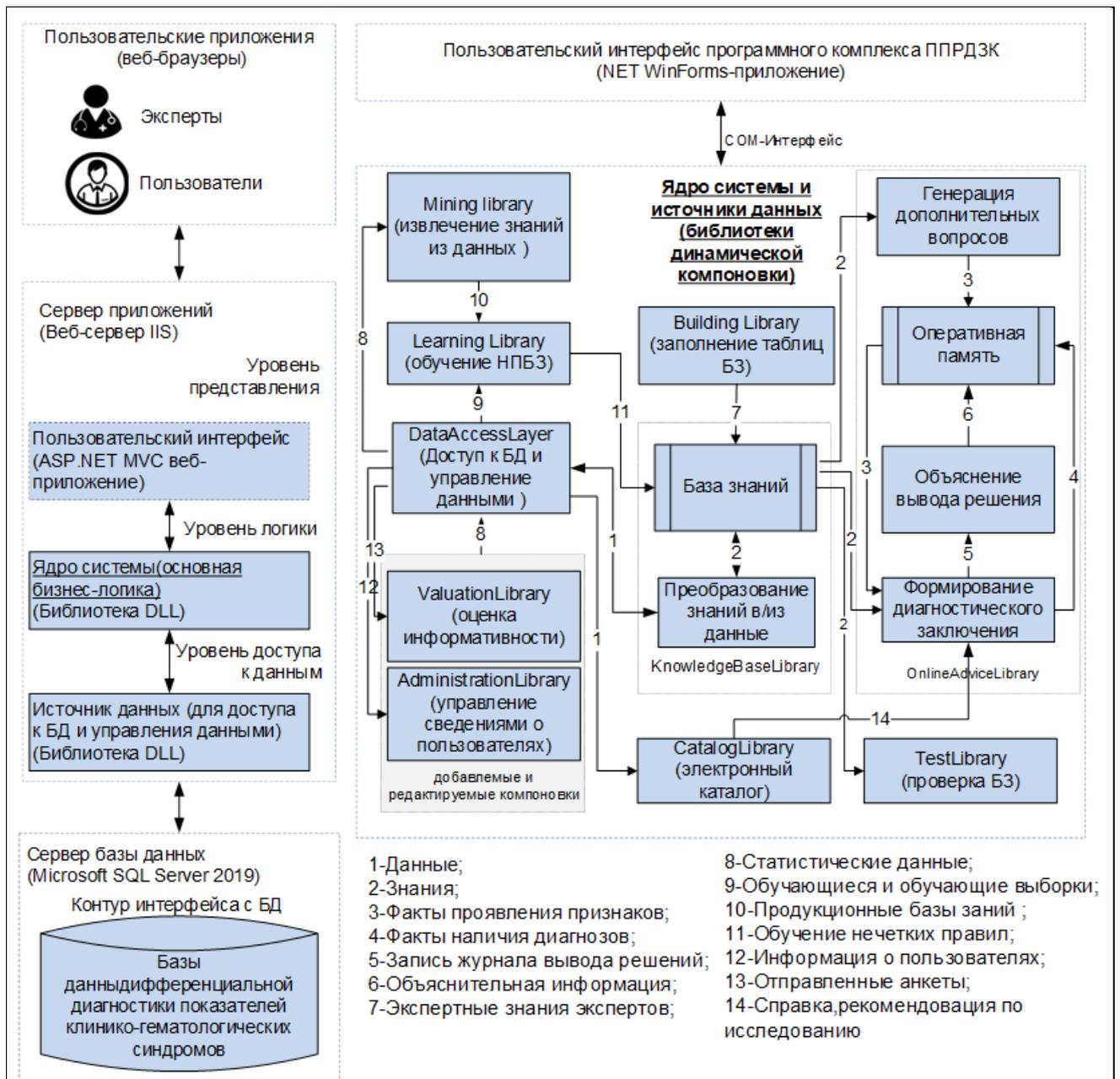


Рисунок 3.2 – Схема архитектуры программного комплекса системы поддержки принятия решений при диагностировании заболеваний крови

При разработке сервера приложений учитываются 3 уровня доступа, представленные на рисунке 3.3.

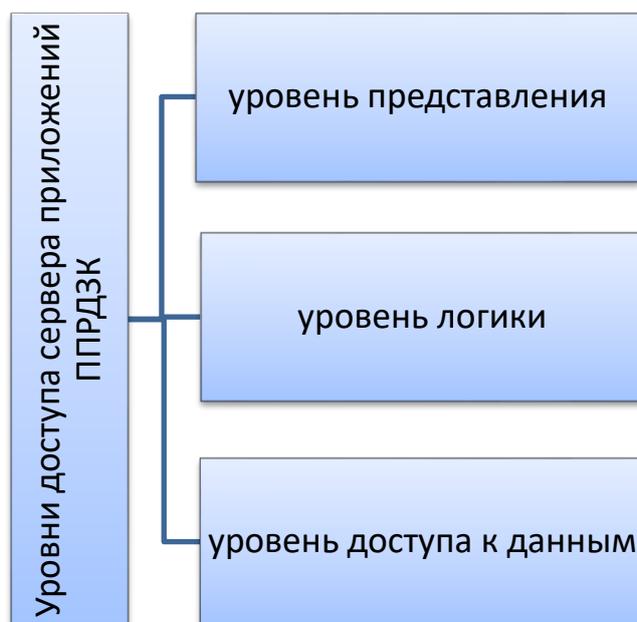


Рисунок 3.3 - Уровни доступа сервера приложений программного комплекса ППРДЗК

Для реализации функций ППРДЗК используется уровень логики, поддерживающий архитектурную модель object-oriented.

Ядро программного комплекса ППРДЗК создано в виде модулей компонентной технологии и функционирует по следующим принципам:

- 1) Реализация и тестирование функциональных модулей комплекса в отдельных библиотеках;
- 2) Создание в ASP.NET MVC объекта COM.

Для реализации вышеуказанных шагов программное ядро комплекса ППРДЗК должен содержать несколько DLL (динамически подключаемых библиотек), представленных в таблице 3.1.

Для обеспечения функционала программного комплекса ППРДЗК и прав их доступа необходимо обеспечить доступ в серверу БД. Источник данных о показателях клинико-гематологических синдромов можно использовать DLL, называемая DataAccessLayer. Данная динамически подключаемая библиотека обеспечивает следующий функционал программного комплекса ППРДЗК:

- обеспечение доступа к базе данных программного комплекса ППРДЗК с применением технологии ADO.NET;
- обеспечение управлением данными клинико-гематологических синдромов, которые хранятся в БД и применением функционал-блоков Data Access Application Block.

Таблица 3.1 - Динамически подключаемые библиотеки программного комплекса ППРДЗК

№	DLL	Функционал DLL
1	BaseKnowledgeLibrary	представление, преобразование, обработка знаний
2	BuildLibrary	Заполняемость данными базы знаний
3	DataMiningLibrary	Выборка нужных знаний из показателей клинико-гематологических синдромов
4	KnowledgeLibrary	обучение моделей ИАД на основе алгоритма Stacking-ансамблирования
5	TestLibrary	тестирование базы знаний на наличие противоречий
6	ValuationLibrary	оценка информативности диагностических показателей клинико-гематологических синдромов
7	HelpLibrary	предоставление справочной информации
8	AdminLibrary	учет пользователей программного комплекса ППРДЗК и прав их доступа

Представленная в архитектуре программного комплекса ППРДЗК динамически подключаемая библиотека DataAccessLayer предоставляет возможности, перечисленные на рисунке 3.4.

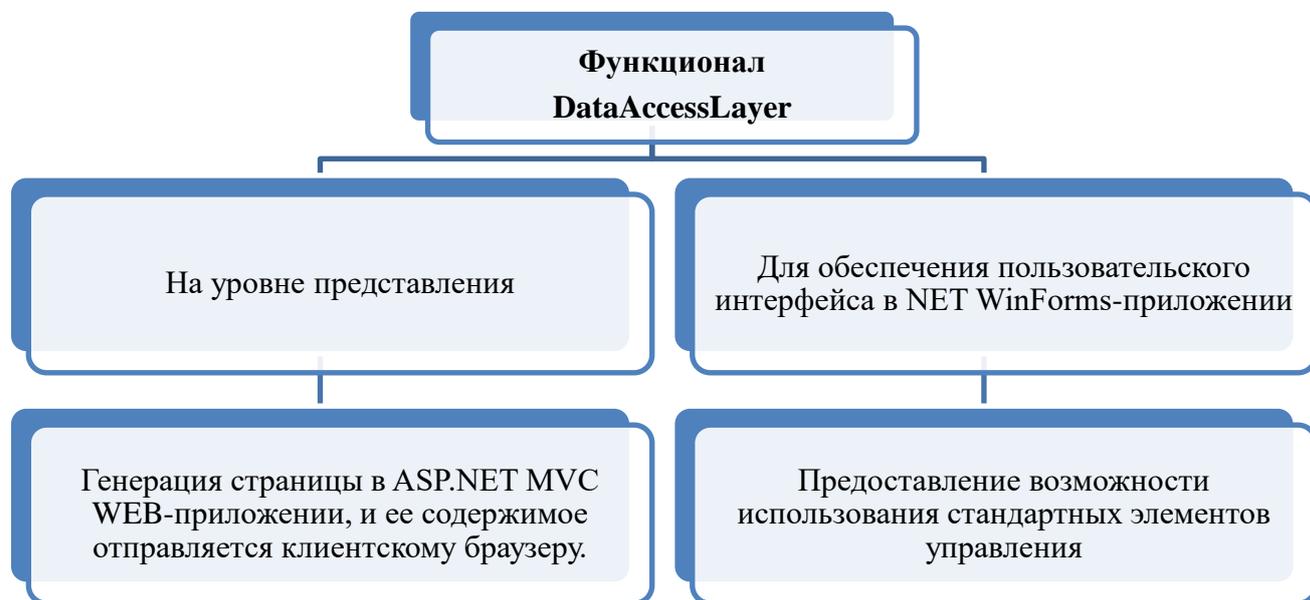


Рисунок 3.4 - Функционал динамически подключаемой библиотеки DataAccessLayer

Сервер БД обеспечен системой управления базами данных MS SQL Server 2019. Этот модуль программного комплекса используется для проектирования БД информационной технологии ППРДЗК.

Таким образом, программный комплекс ППРДЗК основан на гибридной архитектурой модели, интегрирующий архитектурную модель object-oriented и three-tier, а также компонентную модель.

3.3 Модель классов системы поддержки принятия решений

Для проектирования программных модулей комплекса ППРДЗК потребуется разработка модель классов, основанные на объектно-ориентированном языке. Для этого используется языка моделирования UML, обеспечивающий объектно-ориентированный подход к анализу, проектированию и программированию [106].

Как было описано в 3.2 архитектурное решение программного комплекса ППРДЗК интегрирует следующие три уровня доступа (рисунок 3.2): уровень представления, уровень логики, уровень доступа к данным.

Диаграмма классов комплекса ППРДЗК на уровне представления знаний представлена на рисунке 3.5.

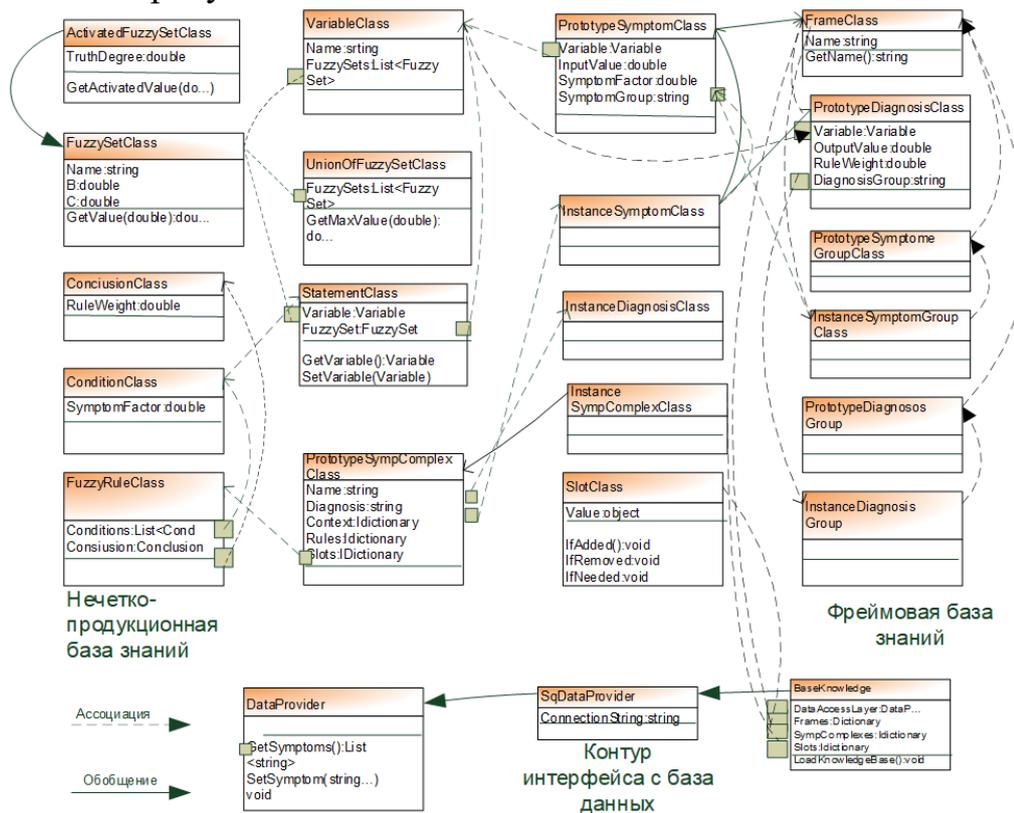


Рисунок 3.5 – Уровень представления знаний

Описание классов уровня представления знаний архитектуры программного комплекса ППРДЗК представлены в таблице 3.2.

Таблица 3.2 - Описание классов уровня представления знаний архитектуры программного комплекса ППРДЗК

Название класса	Описание функционала класса
Класс BaseKnowledge	Реализует БД, содержит метод, осуществляющий загрузку данных и представление знаний
Класс SlotClass	Содержит методы, выполняемые при обращении к его значению
Классы InsDiagnosis, InstDiagnosisGroup, InsSymptom, InsSymptomGroup и InstSympComplex	Реализуют экземпляры конкретных диагнозов, групп диагнозов, отдельных признаков, групп признаков и комплексов признаков для диагнозов в соответствующем порядке. Данные классы наследуются от классов-прототипов InsDiagnosis, InstDiagnosisGroup, InsSymptom, InsSymptomGroup и InstSympComplex соответственно
Класс ProtSympComplexClass	Содержит множество правил описания причинно-следственных отношений между признаками и диагнозами
Класс FuzzyRuleClass	Реализует правила, включающие условия, реализуемого классом Condition, и заключения – классом Conclusion
Класс StatementClass	Является основным для условия и заключения продукционных правил
Класс VariableClass	Реализует переменную «признаков и диагнозов», терм которых реализуются классом FuzzySet
Класс ActivatedFuzzySetClass	Реализует нечеткое множество и наследуется от класса FuzzySet
Класс UnionOfFuzzySetClass	Объединяет нечеткое множество
Класс SqlDataProviderClass	Является потомком класса DataProvider и содержит методы по доступу к базе данных

Описание классов уровня вывода диагноза архитектуры программного комплекса ППРДЗК представлены в таблице 3.3.

Таблица 3.3 - Описание классов уровня вывода диагноза архитектуры программного комплекса ППРДЗК

Название класса	Описание функционала класса
ClassInferenceEngine	реализует механизм вывода диагностического решения на БЗ, реализуемой классом KnowledgeBase, и РП – классом WorkingMemory
ClassBackwardChain	вывод решения начинается с генерации дополнительных вопросов, реализуемой классом BackwardChain, и заканчивается формированием диагностического заключения, реализуемого классом ForwardChain
ClassBackwardChain	содержит методы по выделению основных диагнозов, реализуемых классом Candidate, и запросу оценки силы проявления ведущего признака
ClassForwardChain	содержит методы по сопоставлению, разрешению конфликта, срабатыванию, действию для получения диагностических гипотез
ClassStaking	реализует алгоритм Staking-ансамблирования данных
ClassExplanationFacility	реализует описание решения и содержит метод включения в книгу протокола вывода
ClassBuilding	реализует пополнение базы знаний и задачи удаления, изменения, добавления новых данных по клинико-гематологическим синдромам
ClassMining	реализует извлечение знаний из данных показателей клинико-гематологических синдромов, содержит метод по первой обработке данных
ClassLearning	реализует обучение нечетко-продукционной базы знаний
ClassInitialization	реализует инициализацию параметров и коэффициентов базы знаний
ClassTesting	реализует проверку базы знаний и методы проверки комплекса признаков и добавления создаваемого комплекса признаков в базу знаний
ClassAdministration	реализует задачи администрирования и учета прав пользователей

Диаграмма классов комплекса ППРДЗК на уровне вывода диагноза представлена на рисунке 3.6.

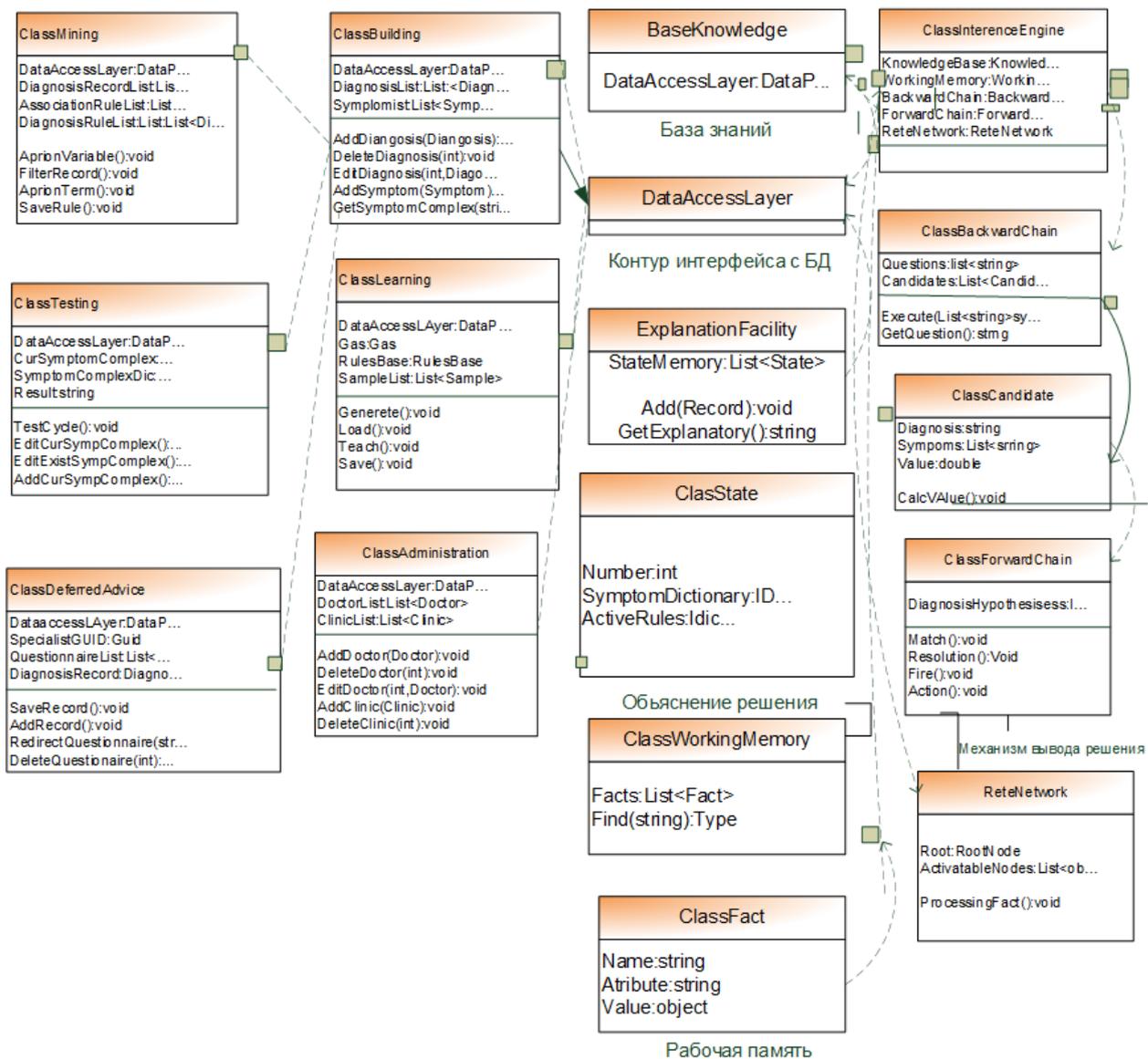


Рисунок 3.6 – Уровень вывода диагноза

Таким образом, для проектирования программных модулей комплекса ППРДЗК разработана модель классов, основанная на объектно-ориентированном языке.

3.4 Проектирование базы данных системы поддержки принятия решений

База данных дифференциальной диагностики показателей клинико-гематологических синдромов состоит из таблиц, которые можно разделить на четыре основных категорий: экспертные данные; статистические данные; ассоциативные правил; обучающиеся; данные пользователей.

Содержание категории «Экспертные данные» представлены в таблице 3.4.

Таблице 3.4 - Описание и содержание категории «Экспертные данные»

Название таблицы	Описание таблицы	Содержание таблицы
TableDiagnosis	содержит список диагнозов	название, описание, диапазон допустимых оценок, группа диагнозов
TableDiagnosisGroup	содержит список групп диагнозов	название, описание
TableSymptom	содержит список отдельных признаков	название, описание, диапазон допустимых оценок, группа признаков
TableSymptomGroup	таблица содержит список групп признаков	название, описание
TableSymptomComplex	содержит список комплексов признаков для диагнозов	диагноз, отдельный признак, оценка силы проявления признака при наличии диагноза, пользователь-эксперт задающий комплекс признаков для конкретного диагноза

Содержание категории «Статистические данные» представлены в таблице 3.5.

Таблице 3.5 - Описание и содержание таблиц категории «Статистические данные»

Название таблицы	Описание таблицы	Содержание таблицы
TDiagnosisRecord	содержит список диагнозов в статистической записи для извлечения знаний	ЛТ переменной «диагноза»
TSymptomRecord	содержит список признаков в статистической записи для извлечения знаний	ЛТ переменной «диагноза», ЛТ переменной «признака»
TDiagnosisRule	содержит список диагнозов из заключения АП, описывающих зависимости между ЛТ переменных	ЛТ переменной «диагноза», достоверность, поддержка, весовой коэффициент НПП сгенерированного АП после обучения
TSymptomRule	содержит список признаков из предпосылки АП, описывающих зависимости между ЛТ переменных	диагноз из заключения АП, ЛТ переменной «признака»
TDiagnosisTerm	содержит список ЛТ переменных «диагнозов»	диагноз, название ЛТ, параметры ФП после обучения
TSymptomTerm	содержит список ЛТ переменных «признаков»	признак, название ЛТ, параметры ФП после обучения
TRulesBase	содержит список обучающихся НПБЗ	оценка приспособленности
TDiagnosisParams	содержит список параметров ФП ЛТ переменных «диагнозов»	обучающаяся НПБЗ, диагноз, ЛТ переменной диагноза, параметры ФП
TSymptomParams	содержит список параметров ФП ЛТ переменных «признаков»	обучающаяся НПБЗ, ЛТ переменной признака, параметры ФП
TRuleCertainty	содержит список сгенерированных НПП	обучающаяся НПБЗ, диагноз из заключения АП, весовой коэффициент НПП

Содержание категории «Данные пользователей» представлены в таблице 3.6.

Таблице 3.6 - Описание и содержание таблиц категории «Данные пользователей»

Название таблицы	Описание таблицы	Содержание таблицы
Users	содержит список пользователей	Ф.И.О., контактные данные
VirtualHelper	Для предотвращения конфликтов с другими модулями веб-приложения	

На рисунке 3.7 представлена схема базы данных дифференциальной диагностики показателей клинико-гематологических синдромов.

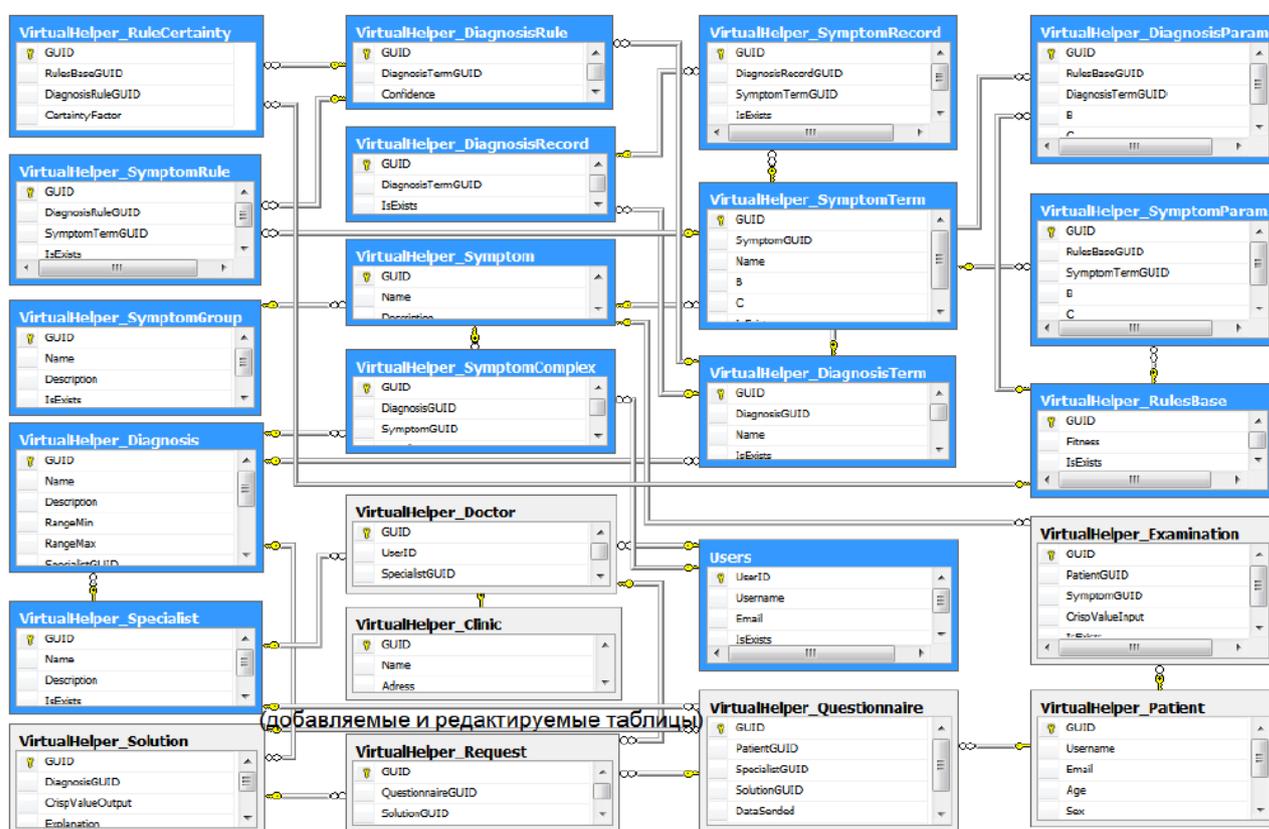


Рисунок 3.7 – Схема базы данных дифференциальной диагностики показателей клинико-гематологических синдромов

Взаимодействие таблиц различных групп описано в таблице 3.7.

Таблице 3.7 - Взаимодействие таблиц различных групп

Таблица	Совместная таблица	Функционал таблиц
Diagnosis и Symptom	DiagnosisTerm, SymptomTerm	определение ЛП «диагнозов и признаков»
Spesialist (DiagnosisGroup)	Diagnosis	определение группы диагнозов
SymptomGroup	Symptom	определение группы признаков
SymptomComplex	Diagnosis и Symptom	определение комплекса признаков для диагноза
DiagnosisRule, SymptomRule	DiagnosisTerm и SymptomTerm	определение АП
DiagnosisRecord и SymptomRecord	DiagnosisTerm и SymptomTerm	определение статистической записи
RulesBase	SymptomParams и DiagnosisParams, RuleCertainty	определение обучающихся НПБЗ

Для управления данными и обеспечения целостности базы данных используются хранимые процедуры. Хранимые процедуры написаны на языке T-SQL. Использование хранимых процедур имеет ряд преимуществ, как например, возможность модульного проектирования и налаженная поддержка, позволяющая осуществить перекомпиляции на стороне приложения при изменениях в таблицах БД.

Таким образом, для проектирования программных модулей комплекса ППРДЗК потребуется модель классов, основанная на объектно-ориентированном языке моделирования UML.

3.5 Выводы по третьему разделу

1. Рассмотрены основные требования к программному комплексу ППРДЗК и описан следующий инструментарий разработки программного комплекса ППРДЗК: язык программирования C#; среда разработки Microsoft Visual Studio; сервер баз данных Microsoft SQL Server; платформа .NET Framework; веб-сервер IIS.

2. Спроектирована архитектура системы поддержки принятия решений, основанная на гибридной архитектурой модели, интегрирующая архитектурные модели object-oriented и three-tier, а также компонентную модель.

3. Построена модель классов системы поддержки принятия решений.

4 ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИАГНОСТИРОВАНИИ ЗАБОЛЕВАНИЙ КРОВИ

4.1 Описание экспериментального исследования

4.1.1 *Общее описание экспериментального исследования*

Экспериментальное исследование основывалось на технологическом процессе диагностирования клинико-гематологических синдромов, позволяющий классифицировать анемии по морфологическим признакам и дать количественную характеристику выраженности патологического процесса. При помощи предложенных алгоритмов можно с высокой долей вероятности выявить ряд наиболее часто встречающихся анемий (железодефицитные анемии, анемии хронических заболеваний, В12-дефицитную анемию).

Технология программного модуля включает 3 основных этапа диагностики анемии:

- установление наличия и степени выраженности анемического синдрома;
- определение характера анемий в соответствии с морфологической классификацией;
- подтверждение и уточнение характера анемии методами клинической биохимии и иммуноферментного анализа.

На этом этапе устанавливается наличие и степени выраженности анемического синдрома. Оценка наличия и степени выраженности анемического синдрома проводится по шести диагностическим признакам: уровню гемоглобина (HGB), количеству эритроцитов (RBC), гематокриту (HCT), среднему содержанию гемоглобина в эритроците (MCH), среднему объему эритроцита (MCV) и средней концентрации гемоглобина в эритроците (MCHC).

На втором этапе определяется характер анемий в соответствии с морфологической классификацией. Функция принадлежности показателей эритроцита к микроцитарным анемиям (нарушения эритропоэза, связанные с дефицитом железа и/или патологией его обмена).

Таким образом, предлагаемый математический способ многокритериальной оценки результатов исследования крови позволяет количественно охарактеризовать их принадлежность к различным видам анемий, оценка может быть проведена при смешанных формах патологии.

На третьем этапе производится подтверждение и уточнение характера анемии методами клинической биохимии и иммуноферментного анализа.

Использование данных, полученных методами клинической биохимии и иммуноферментного анализа позволяет не только уточнить наличие и степень выраженности анемии, но и существенным образом уточнить характер патологии, выявить наиболее распространенные виды анемий. В первую очередь речь идет о диагностике железодефицитной (ЖДА) в группе микроцитарных анемий.

В группе макроцитарных анемий являются В12-дефицитные анемии. Из числа нормоцитарных анемий устанавливается вероятность наличия анемии при хронических заболеваниях (АХЗ). Для повседневного использования рекомендуется тесты доступные лабораториям различного уровня и входящие в большинство известных алгоритмов диагностики. Такими лабораторными показателями являются определение ферритина (Fer), и определение витамина В12.

4.1.2 Описание источника данных экспериментального исследования

Исходные данные для проведения экспериментального исследования предложенной информационной технологии исходные были получены из базы данных MIMIC-III (Medical Information Mart for Intensive care III) [101]. MIMIC-III - это большая, свободно доступная база данных, содержащая идентифицированные данные о здоровье, связанные с более чем сорока тысячами пациентов, которые оставались в отделениях интенсивной терапии (ОИТ) Медицинского центра Beth Israel Deaconess с 2001 по 2012 год.

База данных включает в себя информацию, такую как демографические данные, измерения жизненно важных признаков жизни, сделанные у постели больного, результаты лабораторных испытаний, процедуры, лекарства, заметки воспитателя, отчеты о визуализации и смертности.

MIMIC поддерживает широкий спектр аналитических исследований, охватывающих эпидемиологию, улучшение клинических правил принятия решений и разработку электронных инструментов. Он отличается тремя факторами:

- свободно доступен для исследователей по всему миру;
- охватывает разнообразную и очень большую популяцию пациентов ОИТ;
- содержит данные с высоким временным разрешением, включая результаты лабораторных исследований, электронную документацию, тенденции и сигналы прикроватного монитора.

База данных MIMIC-III периодически обновляется по мере поступления новых данных, совершенствования методов увязки и извлечения данных и предоставления сообществом обратной связи по содержанию базы данных. Для обеспечения прозрачности этого процесса обновления базы данных производятся в пакетном режиме, а номер версии базы данных MIMIC-III обновляется.

Текущая версия базы данных-v1.4. был выпущен 2 сентября 2016 года. Это был крупный релиз, повышающий качество данных и предоставляющий большое количество дополнительных данных для пациентов Metavision.

Основной политикой интернет ресурса является стремление включать данные, которые поддерживают исследования, имеющие отношение к углублению знаний о неотложной помощи. MIMIC в настоящее время содержит такие данные, как:

- лабораторные измерения внутри больницы (т. е. стационара) и из

клиник (т. е. амбулаторно);

- графические наблюдения во время пребывания пациента в отделении интенсивной терапии;

- примечания относительно пребывания пациента в отделении интенсивной терапии, включая примечания по уходу, примечания врача и резюме выписки;

- отчеты по эхокардиографии;

- отчеты электрокардиограммы.

Эти данные в MIMIC периодически дополняются данными, и особый интерес вызывают следующие направления:

- изображения, соответствующие любой медицинской визуализации, выполненной во время пребывания пациента;

- клинические измерения, записанные во время присутствия пациента в операционной;

- измерения аварийного отдела;

- рассматриваются и другие области данных, которые имеют решающее значение для исследования пользователей и в настоящее время недоступны в базе данных, о котором может сообщить пользователь.

По структуре MIMIC - это реляционная база данных, содержащая таблицы данных о пациентах, которые находились в отделениях интенсивной терапии в Медицинском центре Beth Israel Deaconess. Таблица представляет собой структуру хранения данных, которая подобна электронной таблице: каждый столбец содержит согласованную информацию (например, идентификаторы пациента), и каждая строка содержит экземпляр этой информации (например, строка может содержать целое число 340 в столбце идентификатора пациента, что означает, что идентификатор пациента строки равен 340).

Таблицы связаны идентификаторами, которые обычно имеют суффикс "ID". Например, HADM_ID относится к уникальной госпитализации и SUBJECT_ID относится к уникальному пациенту.

Таблицы, предварительно зафиксированные с помощью "D_", являются словарями и предоставляют определения идентификаторов. Например, каждая строка OUTPUTEVENTS связана с одним ITEMID, который представляет понятие измеряется, но он не содержит фактическое название препарата. Объединив OUTPUTEVENTS и d_items on ITEMID, можно определить, какое понятие ITEMID представляет данные. Для определения и отслеживания пребывания пациентов используются метаданные, представленные в таблице 4.1 [101].

Таблица 4.1 – Метаданные базы данных MIMIC-III, хранящие информацию о пациентах

№	Название таблицы	Описание
1	ADMISSIONS	Каждая уникальная госпитализация для каждого пациента в базе данных (определяет HADM_ID)
2	CALLOUT	Информация о том, когда пациент был пройден этап выписки из ОИТ и когда пациент был фактически выписан
3	ICUSTAYS	каждое уникальное ОИТ пребывание в базе данных (определяет ICUSTAY_ID)
4	PATIENTS	каждый уникальный пациент в базе данных (определяет SUBJECT_ID)
5	SERVICES	клиническая служба, в рамках которой зарегистрирован пациент
6	TRANSFERS	движение пациента от кровати к кровати в пределах больницы, включая входы и выходы ОИТ

Каждому ICUSTAY_ID соответствует один HADM_ID и один SUBJECT_ID. Каждый HADM_ID соответствует одному SUBJECT_ID. Один SUBJECT_ID может соответствовать нескольким HADM_ID (несколько госпитализаций одного и того же пациента), и несколько ICUSTAY_ID (несколько ICU остается либо в пределах одной и той же госпитализации, или через несколько госпитализаций, или оба). В таблице 4.2 содержатся данные, собранные в ОИТ.

Таблица 4.2 – Метаданные базы данных MIMIC-III

№	Название таблицы	Данные в таблице
1	CAREGIVERS	каждый обслуживающий персонал, который записал данные в базу данных (определяет CGID)
2	CHARTEVENTS	все графические наблюдения за пациентами
3	DATETIMEEVENTS	все записанные наблюдения, которые являются датами
4	INPUTEVENTS_CV	приемы пациентов, контролируемых с помощью системы Philips CareVue в отделении интенсивной терапии
5	INPUTEVENTS_MV	приемы пациентов, контролируемых с помощью системы Imdsoft Metavision в том числе и в ОИТ
6	NOTEEVENTS	примечания относительно пребывания пациента в отделении интенсивной терапии, включая примечания по уходу, примечания врача, рапорты ЭКГ, отчеты по визуализации и сводки о выписке
7	OUTPUTEVENTS	выходные данные по пациентам в ОИТ
8	PROCEDUREEVENTS_MV	процедуры пациентов для подмножества пациентов, которые наблюдались в ОИТ с использованием системы Imdsoft MetaVision

Следующие таблицы содержат данные, собранные в системе больничных записей (таблица 4.3).

Таблица 4.3 – Метаданные базы данных MIMIC-III о больничных записях

№	Название таблицы	Данные таблицы
1	CPTEVENTS	процедуры, записанные как коды текущей процедурной терминологии
2	DIAGNOSES_ICD	поставленные диагнозы, закодированные с использованием системы Международной статистической классификации заболеваний и связанных с ними проблем со здоровьем (МКБ)
3	DRGCODES	данные, которые используются больницей для выставления счетов
4	LABEVENTS	лабораторные измерения для пациентов как больницы, так и амбулаторных учреждений
5	MICROBIOLOGYEVENTS	микробиологические измерения и чувствительность из базы данных больницы
6	PRESCRIPTIONS	выписанные пациенту лекарства
7	PROCEDURES_ICD	процедуры пациентов, закодированные с использованием системы Международной статистической классификации болезней и связанных с ними проблем со здоровьем (МКБ)

Справочные метаданные базы данных MIMIC-III представлены в таблице 4.4.

Таблица 4.4 - Справочные метаданные базы данных MIMIC-III

№	Название таблицы	Описание
	D_CPT	высокоуровневый словарь кодов текущей процедурной терминологии
	D_ICD_DIAGNOSES	словарь Международной статистической классификации болезней и связанных с ними проблем со здоровьем (МКБ), коды связанные с диагнозами
	D_ICD_PROCEDURES	словарь Международной статистической классификации болезней и связанных с ними проблем со здоровьем (МКБ) коды, касающиеся процедур
	D_ITEMS	словарь ITEMIDs, появляющийся в базе данных MIMIC, за исключением тех, которые относятся к лабораторным испытаниям
	D_LABITEMS	словарь ITEMIDs в базе данных лаборатории, которые относятся к лабораторным испытаниям

База данных MIMIC-II содержит также множество производных таблиц, которые упрощают использование базы данных. Но, в то же время, создатели MIMIC-III приняли сознательное решение не включать любые производные таблицы или вычисляемые параметры, насколько это возможно. Вместо этого они рекомендуют сообществу создавать сценарии, которые можно запускать для создания этих таблиц или параметров, и обмениваться ими. Это имеет много преимуществ: он сохраняет различие между необработанными данными и вычисляемыми данными, поощряет пользователей проверять скрипты, которые получают данные, и позволяет использовать столько скриптов, сколько возможно, не загромождая базу данных для всех пользователей.

4.2 Реализация модели нечеткого логического вывода

Характерной особенностью анемий является то, что в доступном для наблюдения пространстве признаков диагностические классы имеют достаточно сильное пересечение с нечетко выраженными границами, а сами признаки сгруппированы в подсистемы разнородных признаков, характеризующих функциональное состояние различных подсистем организма.

Характер признаков таков, что их следует определять с использованием теории нечетких множеств, что связано как с возможными грубыми ошибками при их измерении, так и с высокой их динамичностью, не связанной с диагностируемой нозологией. Каждому такому признаку может быть поставлена в соответствие функция принадлежности, которая характеризует экспертную уверенность наличия заболевания, связанную с конкретным значением данного признака [107].

Для перехода к нечеткой логике принятия решений в каждой из подсистем признаков четкий вектор состояния объекта классификации подвергается фаззификации (размыванию) с помощью функций принадлежности [108].

В условиях неполноты и неточности информации построение точной математической модели оказывается проблематичным. При этом создание модели сложных, плохо формализуемых объектов является трудно выполнимым. В таких случаях наиболее эффективными являются нечеткие методы моделирования, которые в значительной степени основаны на знаниях экспертов, на основании которых могут быть получены позитивные результаты. К числу сложных, слабо формализуемых объектов, функционирующих в значительной степени в условиях неопределенности, следует отнести и медицину.

Нечеткий вывод занимает центральное место в нечеткой логике и системах нечеткого управления. Процесс нечеткого вывода представляет собой некоторую процедуру или алгоритм получения нечетких заключений на основе нечетких условий или предпосылок с использованием понятий нечеткой логики. Этот процесс соединяет в себе все основные концепции теории

нечетких множеств: функции принадлежности, лингвистические переменные, нечеткие логические операции, методы нечеткой импликации и нечеткой композиции [109].

Как описано в статье [110] логический вывод осуществляется в четыре этапа:

1. Введение нечеткости (фаззификация). Функции принадлежности, определенные на входных переменных, применяются к их фактическим значениям для определения степени истинности каждой предпосылки каждого правила.

2. Логический вывод. Вычисленное значение истинности для предпосылок каждого правила применяется к заключениям каждого правила. Это приводит к одному нечеткому подмножеству, которое будет назначено каждой переменной вывода для каждого правила. В качестве правил логического вывода обычно используются только операции \min (МИНИМУМ) или prod (УМНОЖЕНИЕ). В логическом выводе МИНИМУМА функция принадлежности вывода «отсекается» по высоте, соответствующей вычисленной степени истинности предпосылки правила (нечеткая логика «И»). В логическом выводе УМНОЖЕНИЯ функция принадлежности вывода масштабируется при помощи вычисленной степени истинности предпосылки правила.

3. Композиция. Все нечеткие подмножества, назначенные к каждой переменной вывода (во всех правилах), объединяются вместе, чтобы формировать одно нечеткое подмножество для каждой переменной вывода. При подобном объединении обычно используются операции \max (МАКСИМУМ) или sum (СУММА). При композиции МАКСИМУМА комбинированный вывод нечеткого подмножества конструируется как поточечный максимум по всем нечетким подмножествам (нечеткая логика «ИЛИ»). При композиции СУММЫ комбинированный вывод нечеткого подмножества конструируется как поточечная сумма по всем нечетким подмножествам, назначенным переменной вывода правилами логического вывода.

4. Приведение к четкости (дефаззификация). Используется, когда полезно преобразовать нечеткий набор выводов в четкое число. Имеется большое количество методов приведения к четкости.

Построим модель нечеткого логического вывода применительно к медицинской системе диагностирования клинко-гематологических симптомов, разрабатываемой в рамках данного исследования. Для этого необходимо определить все входные и выходные переменные, участвующих в построении системы и дать им нечеткую характеристику.

Определение входных и выходных параметров. Как уже было описано в предыдущих разделах, диагностирование клинко-гематологических симптомов, а именно заболевания анемией осуществляется по 3 морфологическим классам: микроцитарные, макроцитарные и нормоцитарные анемии.

К классу микроцитарных анемий относятся железодефицитная анемия (ЖДА), анемия при хронических заболеваниях (АХЗ), анемии, связанные с нарушением синтеза порфиринов (анемии НСП), талассемия. Главной задачей является по входным параметрам определить тип анемий данного класса.

Изучение и анализ процесса диагностирования заболеваний анемией позволил обосновать количество значимых факторов, влияющих на целевые функции, и таким образом, сформировать в качестве рабочей базы для дальнейших исследований редуцированную базу.

В качестве универсума предпосылок примем множество факторов или наиболее релевантных причин, влияющих на качество образовательных характеристик $X = \{x_1, x_2, x_3, x_4\}$, где x_1 – содержание сывороточного железа в крови (SFe), x_2 – общая железосвязывающая способность сыворотки крови (TFCS), x_3 – содержание ферритина в сыворотке крови (FS), x_4 – количество ретикулоцитов в крови (RTC).

В качестве второго универсума рассмотрим множество заключений или типы анемий данного класса $Y = \{y_1, y_2, y_3, y_4, y_5\}$, где y_1 – железодефицитная анемия (ЖДА), y_2 – анемии, связанные с нарушением синтеза порфиринов (анемии НСП), y_3 – отсутствие анемии, y_4 – анемии при хронических заболеваниях (АХЗ), y_5 – талассемия.

Введение нечеткости (фаззификация). Дальнейшее решение поставленной проблемы требует нечеткой характеристики всех переменных, участвующих в построении системы принятия решений. Известно, что нечеткие свойства представимы двумя понятиями и их свойствами: нечеткой переменной и лингвистической переменной.

Нечеткая переменная характеризуется тройкой (α, X, A) , где α – наименование переменной; X – универсальное множество (область определения α); A – нечеткое множество на X , описывающее ограничения (т.е. $\mu_A(x)$) на значения нечеткой переменной α .

Лингвистической переменной называется набор (β, T, X, G, M) , где β – наименование лингвистической переменной; T – множество ее значений (терм-множество), представляющих собой наименования нечетких переменных, областью определения каждой из которых является множество X .

Множество T называется *базовым терм-множеством* лингвистической переменной; G – синтаксическая процедура, позволяющая оперировать элементами терм-множества T , в частности, генерировать новые термы (значения). Множество $TUG(T)$, где $G(T)$ – множество сгенерированных термов, называется *расширенным терм-множеством* лингвистической переменной; M – семантическая процедура, позволяющая превратить каждое новое значение лингвистической переменной, образуемое процедурой G , в нечеткую переменную, т.е. сформировать соответствующее нечеткое множество.

Так для фактора x_1 – содержание сывороточного железа в крови, определяются следующие нечеткие характеристики:

- лингвистическая переменная $\beta = \langle \text{SFe} \rangle$;

- область определения $X=[0;2]$;
- терм-множество значений лингвистической переменной $T=\{\text{«снижено»}, \text{«норма»}, \text{«повышено»}\}$.

Для каждого компонента терм-множества T , представляющего нечеткую переменную α_i ($i = 1, 2, 3$), следует построить нечеткое множество A_i . Компонентами этого множества являются возможные значения нечеткой переменной α_i . Принадлежность этих значений множеству, определяемому семантикой термина α_i , задается функцией принадлежности. Таким образом, функция принадлежности представляет собой отображение $\mu_{\tilde{A}}(x) \rightarrow [0, 1], x \in X$ [111].

Построим функции принадлежности для лингвистической переменной («SFe», T, X), где $T=\{\text{«снижено»}, \text{«норма»}, \text{«повышено»}\}$, $X=[0;2]$. Для этой цели был использован пакет прикладных программ Fuzzy Logic Toolbox, входящий в состав среды MatLab. Он позволяет создавать системы нечеткого логического вывода и нечеткой классификации в рамках среды MatLab, с возможностью их интегрирования в Simulink. Базовым понятием Fuzzy Logic Toolbox является FIS-структура - система нечеткого вывода (Fuzzy Inference System). FIS-структура содержит все необходимые данные для реализации функционального отображения «входы-выходы» на основе нечеткого логического вывода согласно схеме, приведенной на рисунке 4.1.

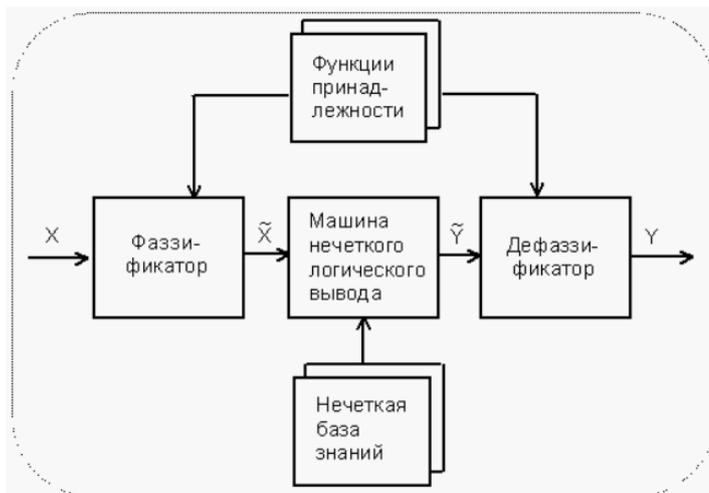


Рисунок 4.1 - Нечеткий логический вывод в среде MatLab

В общем виде модель нечеткого логического вывода проектируемой диагностической системы, построенная в среде MatLab имеет следующий вид (рисунок 4.2).

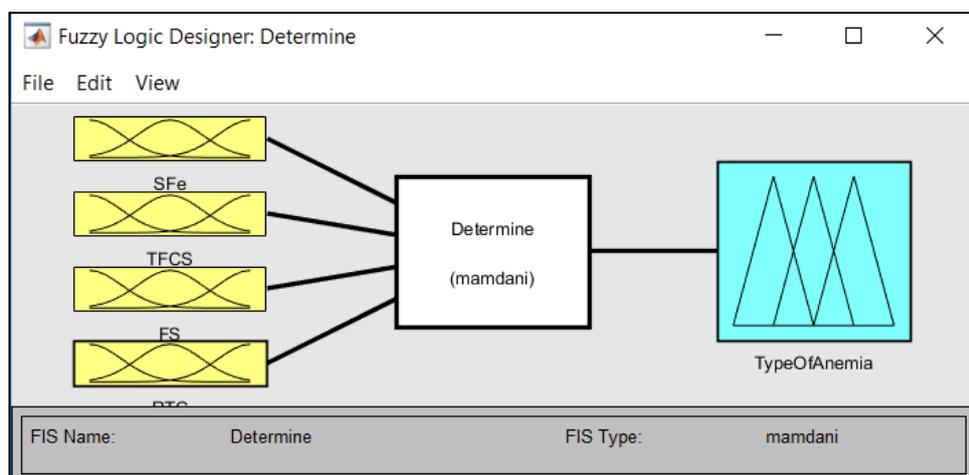


Рисунок 4.2 - FIS-структура моделируемой системы в среде MatLab

Функции принадлежности, построенные для лингвистической переменной («SFe», T , X), представлены на рисунке 4.3.

На рисунке 4.3 видно, что для значения переменной $T=\{\text{снижено}\}$ область определения равна $X=[0;0.4]$. Значение переменной $T=\{\text{норма}\}$ находится в диапазоне $X=[0.4;1.6]$, а для значения $T=\{\text{повышено}\}$ область определения равна $X=[0.4;1.6]$.

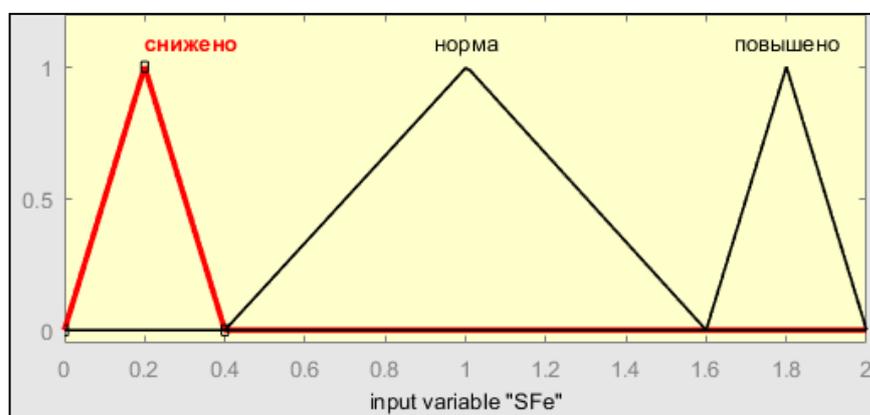


Рисунок 4.3 - Функции принадлежности лингвистической переменной «SFe»

Аналогичным образом строятся функции принадлежности для остальных элементов множества факторов. Следует заметить, что процедура построения функций принадлежности представляет собой этап фаззификации множества предпосылок, конкретизированные значения которых определяют значения следствий, выводимых в процедуре нечеткого логического вывода.

Для фактора x_2 – общая железосвязывающая способность сыворотки крови, определяются следующие нечеткие характеристики:

- лингвистическая переменная $\beta = \langle \text{TFCS} \rangle$;
- область определения $X=[0;8]$;
- терм-множество значений лингвистической переменной $T=\{\langle \text{«снижено»},$

«норма», «повышено»}.

На рисунке 4.4 представлены функции принадлежности лингвистической переменной («TFCS», T , X).

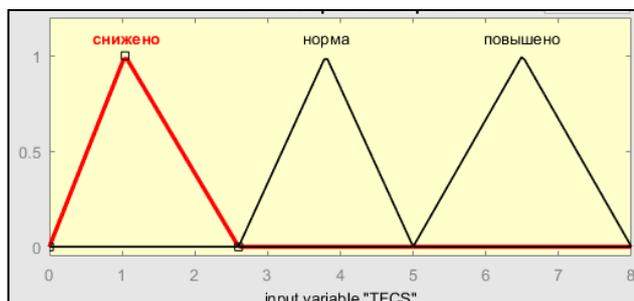


Рисунок 4.4 - Функции принадлежности лингвистической переменной «TFCS»

На представленном выше рисунке 4.4 видно, что для значения переменной $T=\{\text{снижено}\}$ область определения равна $X=[0;2.6]$. Значение переменной $T=\{\text{норма}\}$ находится в диапазоне $X=[2.6;5]$, а область определения $X=[0.4;1.6]$ соответствует значению переменной $T=\{\text{повышено}\}$.

Для фактора x_3 – содержание ферритина в сыворотке крови, определяются следующие нечеткие характеристики: лингвистическая переменная $\beta = \text{«FS»}$; область определения $X=[0;200]$; терм-множество значений лингвистической переменной $T=\{\text{«снижено»}, \text{«норма»}, \text{«повышено»}\}$.

На рисунке 4.5 представлены функции принадлежности лингвистической переменной («FS», T , X).

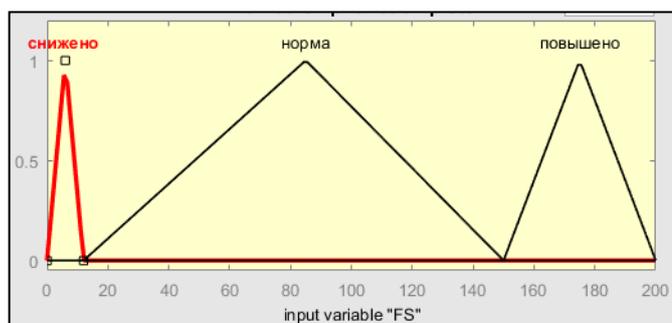


Рисунок 4.5 - Функции принадлежности лингвистической переменной «FS»

На рисунке 4.5 видно, что для значения переменной $T=\{\text{снижено}\}$ область определения равна $X=[0;12]$. Значение переменной $T=\{\text{норма}\}$ находится в диапазоне $X=[12;150]$, а область определения $X=[150;200]$ соответствует значению переменной $T=\{\text{повышено}\}$.

Для фактора x_4 – количество ретикулоцитов в крови определяются следующие нечеткие характеристики:

- лингвистическая переменная $\beta = \text{«RTC»}$;
- область определения $X=[0;2]$;

- терм-множество значений лингвистической переменной $T = \{\text{«снижено»}, \text{«норма»}, \text{«повышено»}\}$.

Функции принадлежности лингвистической переменной («RTC», T , X) представлены на рисунке 4.6.

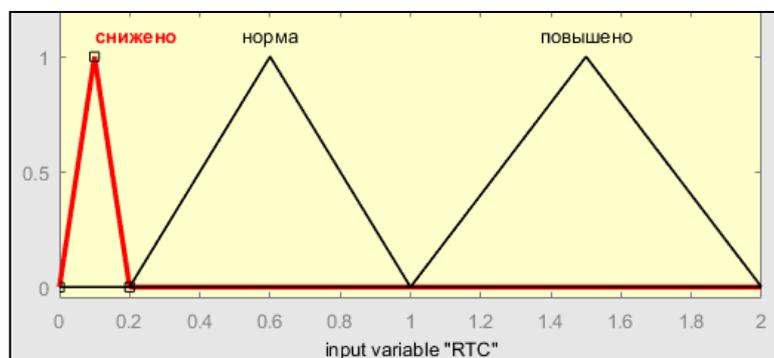


Рисунок 4.6 - Функции принадлежности лингвистической переменной «RTC»

На рисунке 4.6 видно, что для значения переменной $T = \{\text{«снижено»}\}$ область определения равна $X = [0; 0.2]$. Значение переменной $T = \{\text{«норма»}\}$ находится в диапазоне $X = [0.2; 1]$, а область определения $X = [1; 2]$ соответствует значению переменной $T = \{\text{«повышено»}\}$.

На рисунке 3.15 представлены значения, заданные для выходной переменной y – тип анемии, для которой определяются следующие нечеткие характеристики:

- лингвистическая переменная $\beta = \text{«Тип анемии»}$;
- область определения $X = [0; 100]$;
- терм-множество значений лингвистической переменной $T = \{\text{«ЖДА»}, \text{«Анемии НСП»}, \text{«Норма»}, \text{«АХЗ»}, \text{«Талассемия»}\}$.

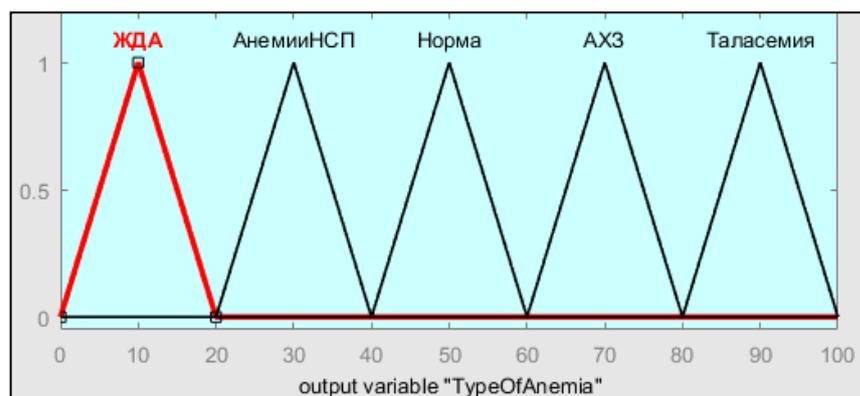


Рисунок 4.7 - Функции принадлежности лингвистической переменной «Тип анемии»

На рисунке 4.7 видно, что для значения переменной $T = \{\text{«ЖДА»}\}$ область определения равна $X = [0; 20]$. Значение переменной $T = \{\text{«Анемии НСП»}\}$

находится в диапазоне $X=[20;40]$, а область определения $X=[40;60]$ соответствует значению переменной $T=\{\text{Норма}\}$. Значение переменной $T=\{\text{АХЗ}\}$ находится в диапазоне $X=[60;80]$, значение переменной $T=\{\text{Анемии НСП}\}$ находится в диапазоне $X=[80;100]$.

Реализация нечеткого вывода основана на нечетких продукционных правилах. Нечеткие правила наиболее близки к логическим моделям, но, что очень важно, они адекватно отражают знания экспертов конкретной предметной области.

Согласно [50] обобщенная форма представления нечеткой продукции имеет вид:

$$i : A \Rightarrow B; F,$$

где i - имя нечеткой продукции, $A \Rightarrow B$ - ядро продукции, в которой A - условная часть ядра (антецедент), B - заключение (консеквент); \Rightarrow - оператор логического следования; F - коэффициент достоверности продукции. Ядро продукции имеет обычную форму словесной интерпретации *если A то B* .

В нечеткой логике, вывод базируется на множестве возможных фактов, появление которых определяется функцией принадлежности. Так, обобщенное правило *modus ponens* (Fuzzy Modus Ponens — FMP) для нечетких систем имеет следующий вид:

$$\frac{A', A \supset B}{B'}$$

Множества A' и A необязательно совпадают. Если A' и A близки друг другу, то можно их сопоставить и получить вывод B' .

В общем случае взаимосвязь между антецедентной и консеквентной частями продукции представляет собой некоторое бинарное отношение на декартовом произведении универсумов соответствующих высказываний, т. е. отражает нечеткое причинное отношение посылки и заключения. Если предположить, что известна функция принадлежности посылки $\mu_{A'}(x)$ и некоторым образом определено бинарное нечеткое отношение на декартовом произведении универсумов

$$R = \{\langle x, y \rangle, \mu_R(x, y)\}, \quad x \in X, y \in Y,$$

то функция принадлежности $\mu_{B'}(y)$ заключения B' представляется в виде композиции [52]:

$$B' = A' \circ R = A' \circ A \rightarrow B.$$

Вывод B' определяется из свертки *max-min* нечеткого множества A' и отношения R :

$$\mu_{B'}(y) = \bigcup_{x \in X} \mu_{A'}(x) \wedge \mu_R(x, y) = \bigcup_{x \in X} (\mu_{A'}(x) \wedge \mu_A(x)) \wedge \mu_B(y).$$

Операции нечеткой конъюнкции и нечеткой дизъюнкции нечетких множеств определим как получение минимума и максимума степеней принадлежности составного высказывания:

$$\mu_A(x) \wedge \mu_B(x) = \min\{\mu_A(x), \mu_B(x)\},$$

$$\mu_A(x) \vee \mu_B(x) = \max\{\mu_A(x), \mu_B(x)\}.$$

Выражение (2) можно интерпретировать как векторно-матричное логическое умножение, причем алгебраической сумме будет соответствовать взятие максимума в соответствии с (4), а умножению - взятие минимума (2).

Перейдем к формированию правил конкретной предметной области, анализируемой в данном исследовании. При составлении правил применяется правило нечеткой конъюнкции, представленные ниже.

Правило 1. ЕСЛИ SFe снижено И TFCS повышено И FS снижено И RTC норма ТО ЖДА.

Правило 2. ЕСЛИ SFe повышено И TFCS снижено И FS повышено И RTC норма ТО Анемии НСП.

Правило 3. ЕСЛИ SFe снижено И TFCS снижено И FS повышено И RTC повышено ТО АХЗ.

Правило 4. ЕСЛИ SFe повышено И TFCS снижено И FS повышено И RTC повышено ТО Талассемия.

Правило 5. ЕСЛИ SFe норма И TFCS норма И FS норма И RTC норма ТО отсутствие анемии.

В диссертационном исследовании реализация нечеткого логического вывода была выполнена по алгоритму Мамдани. Алгоритм Мамдани является одним из первых, который нашел применение в системах нечеткого вывода. По своей сути этот алгоритм содержит рассмотренные выше этапы, поскольку в наибольшей степени соответствует их параметрам [112].

В рассматриваемой ситуации он математически может быть описан следующим образом [112]:

1) Введение нечеткости. Находятся степени истинности для предпосылок каждого из правил $A_1(x_0)$, $A_2(x_0)$, $B_1(y_0)$, $B_2(y_0)$.

2) Логический вывод. Находятся уровни «отсечения» для предпосылок каждого из правил (с использованием операции min):

$$\alpha_1 = A_1(x_0) \wedge B_1(y_0); \alpha_2 = A_2(x_0) \wedge B_2(y_0),$$

где через « min » обозначена операция логического минимума (min). Затем находятся «усеченные» функции принадлежности:

$$C'_1(z) = (\alpha_1 \wedge C_1(z)); C'_2(z) = (\alpha_2 \wedge C_2(z));$$

3) Композиция. Производится объединение найденных усеченных функций с использованием операции min , далее обозначаемой как « \vee », что приводит к получению итогового нечеткого подмножества для переменной вывода z с функцией принадлежности: $\mu_{\gamma}(z) = C(z) = C'_1(z) \vee C'_2(z) = (\alpha_1 \wedge C_1(z)) \vee (\alpha_2 \wedge C_2(z))$.

4) Приведение к четкости. Проводится для нахождения z_0 центроидным методом.

В данной работе реализация логического вывода выполнена с помощью пакета прикладных программ Fuzzy Logic Toolbox, в среде MatLab и представлена ниже на рисунке 4.8.

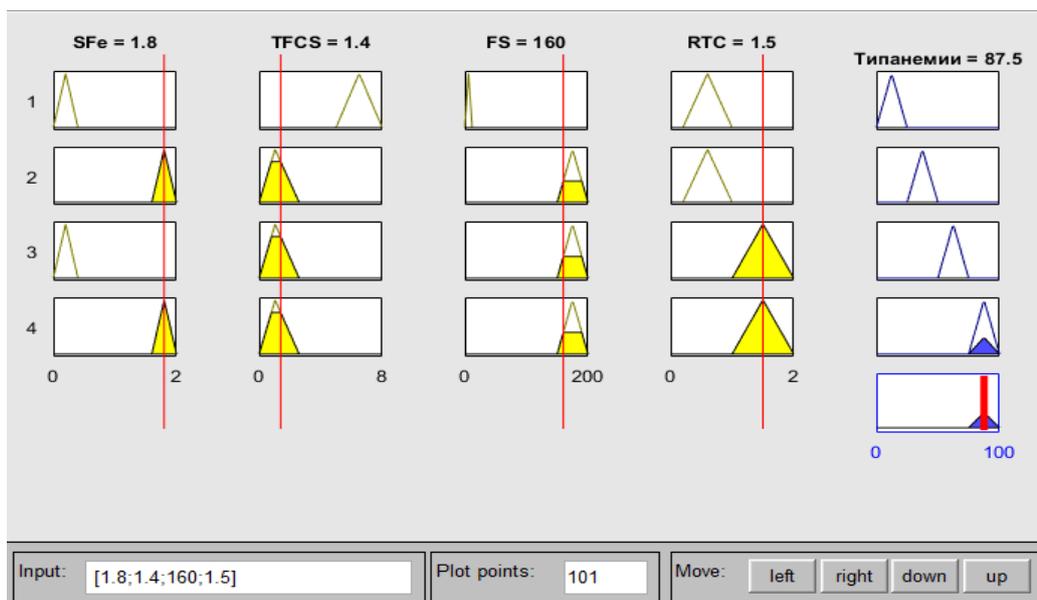


Рисунок 4.8 - Реализация логического вывода

На рисунке 3.16 видно, что для входных переменных заданы конкретные значения: «SFe» = 1.8, «TFCS» = 1.4, «FS» = 160, «RTC» = 1.5.

На первом этапе, на основании данных значений, и исходя из функций принадлежности переменных, находятся степени истинности $\alpha(\text{правило1})$, $\alpha(\text{правило2})$, $\alpha(\text{правило3})$, $\alpha(\text{правило4})$ для предпосылок каждого из четырех правил.

На втором этапе происходит «отсекание» функций принадлежности заключений правил на уровнях $\alpha(\text{правило1})$, $\alpha(\text{правило2})$, $\alpha(\text{правило3})$, $\alpha(\text{правило4})$. На рисунке эти области выделены желтым цветом.

На третьем этапе рассматриваются функции принадлежности, усеченные на предыдущем этапе, и производится их объединение с использованием операции \max , в результате чего получается комбинированное нечеткое подмножество, описываемое функцией принадлежности μ_{γ} (Тип анемии) и соответствующее логическому выводу для выходной переменной «Тип анемии».

На четвертом этапе находится четкое значение выходной переменной «Тип анемии», с применением центроидного метода. В данном случае значение выходной переменной «Тип анемии» равен 87.5, что соответствует области значения переменной $T = \{\text{Анемии НСП}\}$, который находится в диапазоне $X = [80; 100]$.

Таким образом, можно сделать вывод, что при указанных значения входных параметров SFe = 1.8, TFCS = 1.4, FS = 160, RTC = 1.5, которые соответствуют показателям анализа крови: содержание сывороточного железа в

крови (SFe), общая железосвязывающая способность сыворотки крови (TFCS), содержание ферритина в сыворотке крови (FS) и количество ретикулоцитов в крови (RTC), можно определить наличие анемического синдрома, который в данном случае, соответствует типу «Анемии НСП». Другими словами, на основании показателей анализа крови при помощи нечеткого логического вывода определяется тип анемии.

Аналогичным образом строятся нечеткие модели логического вывода для определения типа анемий нормоцитарных и макроцитарных классов анемического синдрома.

4.3 Программная реализация вычислительного алгоритма морфологической классификации клинко-гематологических синдромов

После запуска программного модуля диагностирования клинко-гематологических синдромов на экране выводится окно, представленное на рисунке 4.9.

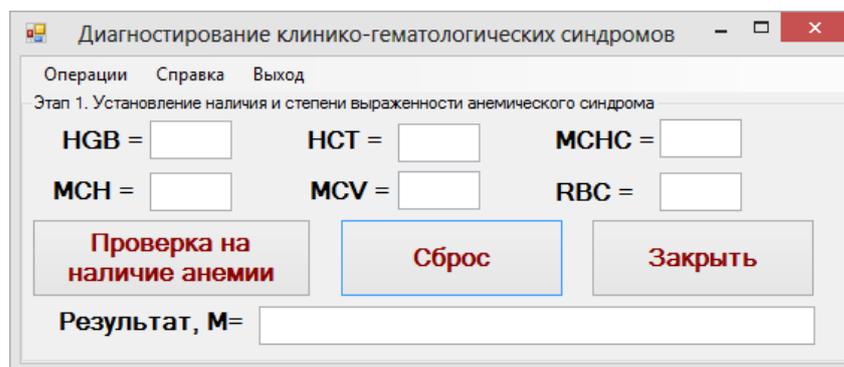


Рисунок 4.9 - Главное окно программы

Данное окно разделено на 3 области: в первой «Установление анемического синдрома» находятся 6 полей для ввода значений анализируемых параметров гемограммы пациента (гемоглобин, гематокрит, MCHC, MCH, MCV, количество эритроцитов), а так же окно с результатом расчета критерия принадлежности. Когда значение критерия находится в пределах от 0 до 0,2 рядом с окном результата появляется заключение: «Анемия отсутствует» (рисунок 4.10).



Рисунок 4.10 - Окно программы с результатом «Анемия отсутствует»

При значении критерия от 0,2 до 0,5 выводится заключение «Подозрение на наличие анемии» (рисунок 4.11).

При значении показателя более 0,5 выводится заключение «Состояние анемии» и активируются вторая и третья области («Морфологический характер анемии» и «Методы клинической биохимии и иммуноферментного анализа» соответственно).

Операции	Справка	Выход			
Этап 1. Установление наличия и степени выраженности анемического синдрома					
HGB =	120	HCT =	0,41	MCHC =	32
MCH =	28	MCV =	90	RBC =	4,5
Проверка на наличие анемии			Сброс		Закреть
Результат, M=			Подозрение на наличие анемии (M=0,25)		
2 этап. Определение характера анемий в соответствии с морфологической классификацией					
Микроцитарная	<input type="checkbox"/>	X			Определение характера анемии
Макроцитарная	<input type="checkbox"/>	X			Лечение
Нормоцитарная	<input type="checkbox"/>	X			

Рисунок 4.11 - Окно программы с результатом «Подозрение на наличие анемии»

Во второй области окна программы отображаются результаты расчета принадлежности к типам анемий (микроцитарная, макроцитарная, нормоцитарная, смешанная).

В третьей области программы активируются поля для ввода соответствующих биохимических и/или иммуноферментных анализов.

Операции	Справка	Выход			
Этап 1. Установление наличия и степени выраженности анемического синдрома					
HGB =	120	HCT =	0,3	MCHC =	25
MCH =	19	MCV =	90	RBC =	4,5
Проверка на наличие анемии			Сброс		Закреть
Результат, M=			Состояние анемии (M=0,54)		
2 этап. Определение характера анемий в соответствии с морфологической классификацией					
Микроцитарная	0,47	X	Смешанный тип анемии		Определение характера анемии
Макроцитарная	0	X			Лечение
Нормоцитарная	0,53	V			

Рисунок 4.12 - Анемия с определением ее морфологического характера

Критерии принадлежности по морфологическим признакам анемии рассчитываются во всех случаях, при наличии соответствующих признаков

делается заключение о наличии микро-, макро-, нормоцитарной анемий или анемии смешанного характера. При невозможности идентификации морфологического характера анемии дается количественная оценка признакам микро-, макро-, нормоцитарной анемий (рисунок 4.13).

Диагностирование клиничко-гематологических синдромов

Операции Справка Выход

Этап 1. Установление наличия и степени выраженности анемического синдрома

HGB = 115 HCT = 0,4 MCHC = 25
MCH = 25 MCV = 80 RBC = 3,5

Проверка на наличие анемии Сброс Закреть

Результат, M= Состояние анемии (M =0,72)

2 этап. Определение характера анемий в соответствии с морфологической классификацией

Микроцитарная 0,12 X Смешанный тип анемии
Макроцитарная 0 X
Нормоцитарная 0,38 V

Определение характера анемии Лечение

Рисунок 4.13 - Количественная оценка признакам микро-, макро-, нормоцитарной анемий

При идентификации морфологического характера анемии производится ввод результатов биохимического и/или иммуноферментного анализов, позволяющие с высокой долей вероятности идентифицировать железодефицитную анемию (ЖДА), анемию хронических заболеваний (АХЗ) и В12-дефицитную анемию. При значениях данных показателей выше 0,5 ставится заключение о наличии ЖДА или АХЗ (Рисунок 4.14).

а

б

Рисунок 4.14 – а) Железодефицитная анемия;
б) Переход к следующему результату анализа

4.4 Реализация алгоритма Staking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов

Для реализации экспериментального исследования проверки алгоритма Staking-ансамблирования интеллектуальной обработки показателей клинико-гематологических синдромов были определены 2 следующие задачи дифференциальной диагностики: 1) анализ данных показателей клинико-гематологических синдромов для прогнозирования микроцитарной анемии; 2) анализ данных показателей клинико-гематологических синдромов для прогнозирования нормоцитарной анемии.

Исследование было проведено на данных из базы данных MIMIC-III (Medical Information Mart for Intensive care III) [101].

Алгоритм Staking-ансамблирования предполагает, что берутся данные для обучения и делятся случайным образом на 2 части. Первая часть используется для обучения исходных моделей 2 часть используется для прогнозирования данных на основе исходных моделей и на основе этого прогноза строится метамодель. Данная операция проводится несколько раз - так чтобы все исходные данные в разных разбиениях попали во 2 часть.

Для реализации данного алгоритма были созданы несколько структур интеллектуального анализа в Analysis Services. Вариативность выборки задается через свойство HoldoutSeed. Оно разное у разных вариантов.

Для каждого вариант вычисляется вероятность наступления события по каждой модели в ансамбле. На основе вероятностей по каждой проверочной строке - формируются вектор, который будут использоваться для обучения метамодели. Для просмотра полученных данных для обучения метамодели в MS SQL Server Management Studio вызывается одна из следующих процедур:

- CALL SPMining.SPMining.SPortalMiningProc.GenerateTableGPA()
- CALL SPMining.SPMining.SPortalMiningProc.GenerateTableEmployment()

Для создания метамодели используется следующий алгоритм:

```
CREATE MINING MODEL [Стэкинг – Диагностирование микроцитарной анемии]  
(  
  [Ид] LONG KEY,  
  [HGB] TEXT DISCRETE,  
  [RBC] LONG DISCRETE,  
  [HCT] LONG DISCRETE,  
  [MCH] LONG DISCRETIZED,  
  [MCV] TEXT DISCRETE,  
  [MCHC Обучения] TEXT DISCRETE,  
  [Класс] TEXT DISCRETE PREDICT  
)  
USING Microsoft_Naive_Bayes(MINIMUM_DEPENDENCY_PROBABILITY=0.01)
```

Далее данную модель необходимо обучить. Для обучения определяются классы, которые выделяются в зависимости от вероятности.

Классы задаются строкой при обучении модели: <ИмяКласса1>=<Вероятность1>;<ИмяКласса2>=<Вероятность2>;... Выделение классов производится экспертным путем (обучение с учителем).

Для построения метамодели будем использовать алгоритмы БК, ДР и НС. На рисунке 4.15 представлено окно подсистемы анализа данных ООО на основе алгоритма Staking-ансамблирования.

Использование нескольких моделей дает следующие преимущества:

- реализовывать предсказание выходного показателя клинико-гематологических синдромов как в численном виде (НС) так и в виде классификации (ДР и БК), что определяет разнообразие возможных методов визуализации результатов;

- согласованность результатов по всем моделям позволяет сделать вывод об их достоверности.

Код программной реализации алгоритма Staking-ансамблирования представлен в приложении И.

Обучение на основе базовых классификаторов для анализа данных первой задачи дифференциальной диагностики показало следующие результаты.

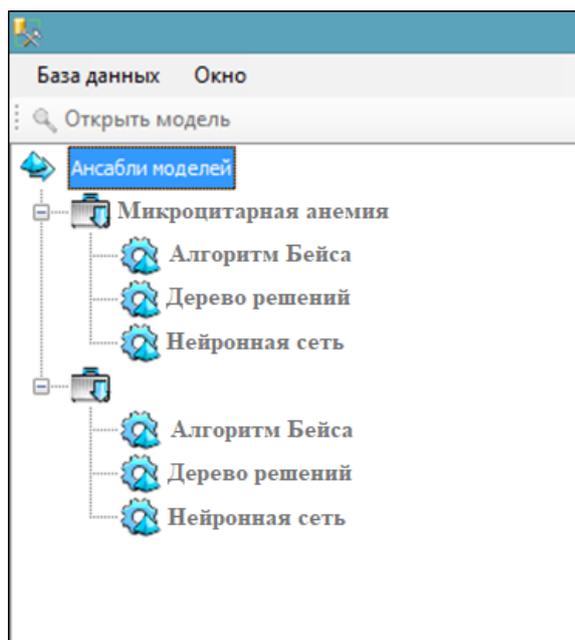


Рисунок 4.15 - Главное меню модуля Staking-ансамблирования

На основе данных показателей клинико-гематологических синдромов были получены результаты обучения базового классификатора на основе алгоритма БК для анализа данных принадлежности к микроцитарному типу и представлены в виде сети зависимостей на рисунке 4.16.

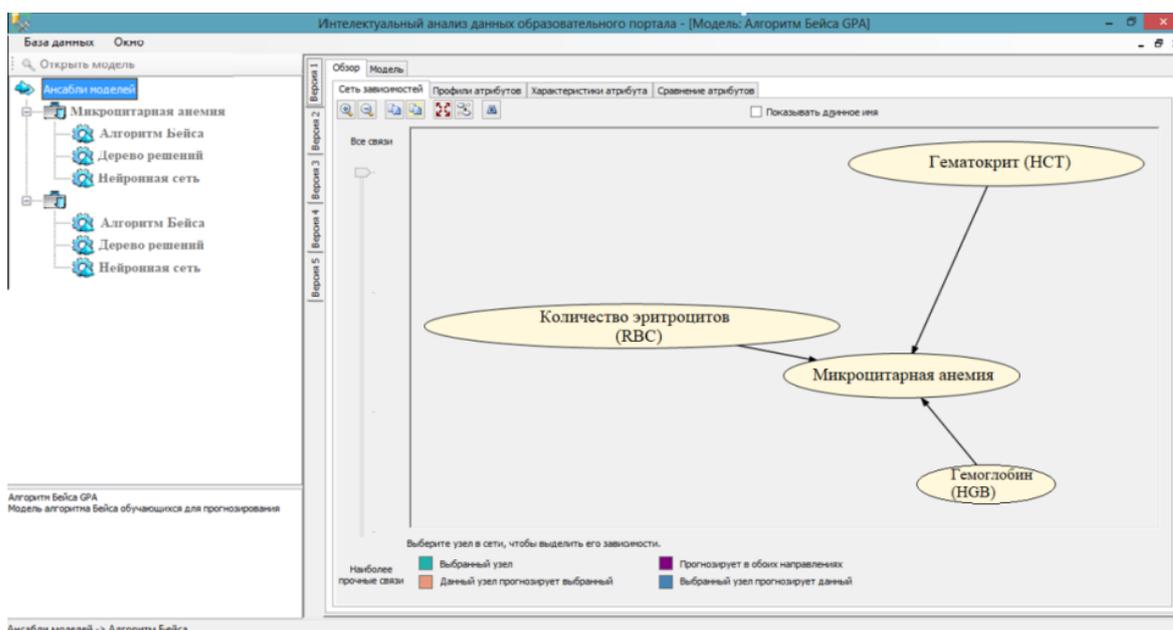


Рисунок 4.16 – Окно модуля с результатами обучения базового классификатора на основе алгоритма БК для анализа данных показателей клинко-гематологических синдромов

Результаты обучения базового классификатора на основе алгоритма БК показал, что набор «Микроцитарная анемия» имеет зависимость со следующими показателями клинко-гематологических синдромов: Гемоглобин (HGB); количество эритроцитов (RBC); Гематокрит (HCT).

4.5 Анализ эффективности применения предложенной информационной технологии в задачах диагностирования заболеваний крови

Эффективность применения предложенной модели инфраструктуры диагностирования клинко-гематологических синдромов в целях ППР, основанной на наблюдении, контроле, анализе данных распределенных объектов с помощью разработанных моделей, алгоритмов и методов анализируется с точки зрения достижения требуемого результата - обеспечение эффективного управления. Эффективность управления, главным образом, определяется своевременным принятием адекватных решений.

Для оценивания значимости предложенных в работе методических и алгоритмических средств информационно-аналитической поддержки принятия решений и целесообразности их применения в обеспечении наиболее эффективного управления используется метод анализа иерархий Т. Саати [113]. Метод анализа иерархий (МАИ) совмещает в себе достоинства аналитических и экспертных методов, обеспечивает реализацию наиболее эффективного способа оценки количественно неизмеримых, но важных факторов для принятия

обоснованных решений.

Согласно методу Т. Саати, построена иерархическая система оценки значимости подходов для обеспечения эффективного управления (рисунок 4.17), в которой заданы цель иерархии, критерии первого уровня, критерии второго уровня и альтернативы. К критериям первого уровня относятся основные этапы принятия решений в процессе управленческой деятельности: подготовка, формирование и реализация решения.



Рисунок 4.17- Иерархическая система оценки значимости подходов для обеспечения эффективного управления

В результате исследования особенностей задач дифференциальной диагностики, методов и алгоритмов поддержки процессов принятия решений, в качестве критериев второго уровня выделены следующие факторы: адекватность; актуальность; системность; компетентность; оперативность; гибкость; наглядность; полнота; конструктивность; структуризация; эргономичность; интерактивность; совместимость.

В качестве альтернатив выбраны варианты поддержки принятия решений: комплексное применение предложенных средств ППР; частичное применение предложенных средств ППР; применение других существующих решений ППР; отсутствие специализированных средств ППР.

Элементы иерархии сравниваются попарно по отношению к их воздействию на общую для них характеристику, с целью выявления относительных весов. Процедура парных сравнений приводит к результату, который представляется в виде обратно симметричной матрицы. Элементом матрицы сравнений $a(i, j)$ является интенсивность проявления элемента

иерархии i относительно элемента j , оцениваемая по шкале отношений Саати, где оценки имеют следующий смысл: 1 — равная важность; 3 - умеренное превосходство одного над другим; 5 — существенное превосходство одного над другим; 7 - значительное превосходство одного над другим; 9 - очень сильное превосходство одного над другим; 2, 4, 6, 8 - соответствующие промежуточные значения.

На основе парных сравнений рассчитываются коэффициенты значимости. Расчет сводится к вычислению главного собственного вектора матрицы попарного сравнения и его нормализации. С целью определения степени согласованности данных определяется отношение согласованности (ОС). Значение ОС большее, чем 0,10, считается неприемлемым. Результаты попарного сравнения элементов первого уровня по отношению к их воздействию на общую цель отражены в таблице 4.5.

Таблица 4.5 - Матрица сравнений для критериев первого уровня

Эффективное управление	Подготовка решения	Принятие решения	Реализация решения	Нормализованные оценки вектора приоритета
Подготовка решения	1	1/2	3	0,31
Принятие решения	2	1	5	0,58
Реализация решения	1/3	1/5	1	0,11
ОС=0,0031850				

Как следует из таблицы 4.5, наибольший вес по влиянию на обеспечение эффективного управления имеет элемент «принятие решения». Результаты попарного сравнения критериев второго уровня по отношению к их воздействию на критерии первого уровня представлены в таблицах 4.6 - 4.8.

Таблица 4.6 — Матрица сравнений критериев второго уровня по отношению к их воздействию на подготовку решений

	Подготовка решения	1	2	3	4	5	6	7	Нормализованные оценки вектора приоритета
1	Оперативность	1	1/3	1/4	5	2	1/6	1/7	0,06
2	Адекватность	7	1	3	2	4	6	5	0,37
3	Системность	4	1/3	1	3	5	1/3	1/2	0,13
4	Гибкость	1/5	1/2	1/3	1	6	4	1/7	0,08
5	Наглядность	1/2	1/4	1/5	1/6	1	1/6	1/5	0,03
6	Компетентность	2	1/6	3	1/4	6	1	1/2	0,10
7	Актуальность	7	1/5	2	7	5	2	1	0,23
ОС= 0,2656881									

Таблица 4.7 — Матрица сравнений критериев второго уровня по отношению к их воздействию на принятие решения

	Принятие решения	1	2	3	4	5	6	7	Нормализованные оценки вектора приоритета
1	Адекватность	1	6	7	2	3	1/2	4	0,24
2	Полнота	1/6	1	1/3	1/4	1/5	1/6	3	0,04
3	Наглядность	1/7	4	1	1/3	1/4	1/5	1/3	0,04
4	Компетентность	1/2	3	4	1	1/2	1/6	5	0,11
5	Актуальность	1/3	5	4	2	1	1/4	5	0,15
6	Конструктивность	2	6	5	6	4	1	7	0,37
7	Структуризация	1/4	1/3	3	1/5	1/5	1/7	1	0,04
OS = 0,0910800									

Таблица 4.8 - Матрица сравнений критериев второго уровня по отношению к их воздействию на реализацию решения

	Реализация решения	1	2	3	4	5	6	Нормализованные оценки вектора приоритета
1	Оперативность	1	2	3	4	3	8	0,38
2	Эргономичность	1/2	1	1/4	5	1/3	2	0,11
3	Гибкость	1/3	4	1	3	2	4	0,23
4	Структуризация	1/4	1/5	1/3	1	1/3	2	0,06
5	Интерактивность	1/3	3	1/2	3	1	3	0,17
6	Совместимость	1/8	1/2	1/4	1/2	1/3	1	0,05
OS= 0,0913197								

Из таблиц 4.6 — 4.8 следует, что на этапе подготовки решения наибольшую значимость имеют такие факторы, как: адекватность, актуальность, компетентность, системность, оперативность; на этапе формирования решения - конструктивность, адекватность, актуальность, компетентность, наглядность; на этапе документационной поддержки реализации решения - оперативность, гибкость, интерактивность, эргономичность. Влияние остальных факторов существенно ниже. Значения показателей клинико-гематологических синдромов свидетельствуют о достаточно хорошей согласованности оценок.

Сравнение альтернатив по отношению к критериям второго уровня иерархии выполняется на основе анализа характеристик качества подходов, методов, моделей и реализующих их программных средств поддержки процессов подготовки, формирования и реализации управленческих решений.

Результаты попарного сравнения альтернатив по критериям второго уровня иерархии представлены в таблице приложения К.

Альтернатива «отсутствие специализированных средств ППР» существенно уступает в значимости, поскольку ни один из заданных критериев в полной мере не обеспечивается. Дальнейший анализ связан с обработкой матриц сравнения. В результате синтеза матриц рассчитывается глобальный вектор приоритетов. Итоговые значения представлены в таблице 4.9.

Таблица 4.9 - Итоговые значения глобального приоритета

Критерии	Значения вектора приоритета	Комплексное применение предложенных методов	Частичное применение предложенных методов	Применение других существующих методик	Отсутствие специализированных методов
Адекватность	0,231	0,517	0,298	0,139	0,046
Актуальность	0,124	0,593	0,251	0,114	0,042
Системность	0,028	0,593	0,256	0,083	0,068
Компетентность	0,149	0,509	0,329	0,120	0,042
Оперативность	0,051	0,606	0,203	0,143	0,048
Гибкость	0,024	0,535	0,316	0,096	0,053
Наглядность	0,051	0,644	0,204	0,094	0,058
Полнота	0,045	0,558	0,245	0,154	0,043
Конструктивность	0,189	0,587	0,253	0,108	0,051
Структуризация	0,047	0,602	0,255	0,106	0,037
Эргономичность	0,020	0,551	0,225	0,159	0,065
Интерактивность	0,024	0,561	0,288	0,110	0,041
Совместимость	0,016	0,471	0,346	0,136	0,046
Глобальный приоритет		0,558	0,272	0,121	0,047

Результаты синтеза показали, что наибольший вес среди факторов получили адекватность (0,2307), конструктивность (0,1888), компетентность (0,149), актуальность (0,124) и оперативность (0,051). Среди альтернатив наибольший вес имеют «Комплексное применение предложенных средств ППР» (0,5830) и «Частичное применение предложенных средств ППР» (0,2509). Таким образом, установлено, что комплексная поддержка процессов подготовки, формирования и реализации управленческих решений на основе предложенных в работе метода оценивания, средств формирования управляющих рекомендаций и оперативной генерации документов имеет наибольшую значимость для достижения требуемого результата - обеспечения эффективного управления.

4.6 Выводы по четвертому разделу

1. Представлено общее описание данных экспериментального исследования информационной технологии поддержки принятия решений при диагностировании заболеваний крови. Исходные данные для проведения экспериментального исследования предложенной информационной технологии исходные были получены из базы данных MIMIC-III. MIMIC поддерживает широкий спектр аналитических исследований, охватывающих эпидемиологию, улучшение клинических правил принятия решений и разработку электронных инструментов.

2. Проведено экспериментальное исследование модели нечеткого логического вывода, реализованного на основе алгоритма Мамдани.

3. Описаны результаты программной реализации вычислительного алгоритма морфологической классификации клинико-гематологических синдромов.

4. Проведено экспериментальное исследование алгоритма Stacking-ансамблирования методов интеллектуальной обработки показателей клинико-гематологических синдромов, в котором были определены задачи дифференциальной диагностики - анализ данных показателей клинико-гематологических синдромов для прогнозирования микроцитарной анемии и нормоцитарной анемии.

5. Проведен анализ эффективности применения предложенной информационной технологии в задачах диагностирования заболеваний крови.

ЗАКЛЮЧЕНИЕ

По результатам проведенного диссертационного исследования можно сделать следующие выводы.

Проанализированы предпосылки и основные направления развитие электронного здравоохранения в Казахстане. В диссертационной работе разработана информационная технология поддержки принятия решений при диагностировании заболеваний крови, основанная на ИТ-инфраструктуре диагностирования клиничко-гематологических синдромов, интегрирующая вычислительный алгоритм морфологической классификации, метод оценки информативности диагностических показателей и гибридный алгоритм Stacking-ансамблирования методов интеллектуальной обработки показателей клиничко-гематологических синдромов. В диссертации предложена архитектура и реализация модулей программного комплекса поддержки принятия решений при диагностировании заболеваний крови.

Разработанная, в рамках исследования, архитектура программного комплекса основана на интеграции трехуровневой, объектно-ориентированной и компонентной архитектурных шаблонов. Предложенная комбинированная архитектура позволяет повысить эффективность процессов масштабирования, обеспечения безопасности, гибкости системы, а также является удобным для сопровождения.

Представлено общее описание данных экспериментального исследования информационной технологии поддержки принятия решений при диагностировании заболеваний крови. Проведено экспериментальное исследование модели нечеткого логического вывода, реализованного на основе алгоритма Мамдани. Описаны результаты программной реализации вычислительного алгоритма морфологической классификации клиничко-гематологических синдромов. Проведено экспериментальное исследование алгоритма Stacking-ансамблирования методов интеллектуальной обработки показателей клиничко-гематологических синдромов. Проведен анализ эффективности применения предложенной информационной технологии в задачах диагностирования заболеваний крови.

Программный модуль оценки информативности признаков клиничко-гематологических синдромов был внедрен в Казахский Институт онкологии и радиологии г. Алматы, программный модуль диагностирования клиничко-гематологических синдромов был внедрен в городской ревматологический центр г. Алматы, что подтверждается актами внедрения, которые представлены в Приложении А,Б.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Кобринский Б. А. Системы поддержки принятия решений в здравоохранении и обучении // Врач и информационные технологии. 2010. №2. URL: <https://cyberleninka.ru/article/n/sistemy-podderzhki-prinyatiya-resheniy-v-zdravoohranenii-i-obuchenii> (дата обращения: 24.08.2019).
2. Спатаев Е. М., Романова Ж. В., Есенбаев Б. С. Современные подходы к формированию национальной стратегии электронного здравоохранения // Вестник Казахского Национального медицинского университета. – 2018. – №. 2.
3. Приказ Министра здравоохранения Республики Казахстан от 3 сентября 2013 года №498 "Об утверждении Концепции развития электронного здравоохранения Республики Казахстан на 2013-2020 годы" URL: <http://dsm.gov.kz/ru/pages/prikaz-ministra-zdravoohraneniya-respubliki-kazahstan-ot-3-sentyabrya-2013-goda-no498-ob> (дата обращения: 24.08.2019).
4. А.Р. Шопабаева, Р.М. Блатов, С.Б. Сыдыков К.С. Жакипбеков, К.М. Елшибекова. Информационно-коммуникационные технологии в системе здравоохранения Республики Казахстан: проблемы и перспективы развития // Здравоохранение. – 2016. – Т. 13. – №. 9. – С. 6.
5. Постановление Правительства Республики Казахстан от 12 декабря 2017 года № 827 «Об утверждении Государственной программы «Цифровой Казахстан» (с изменениями и дополнениями по состоянию на 29.07.2019 г.) URL: https://online.zakon.kz/document/?doc_id=37168057 (дата обращения: 24.08.2019)
6. Kylyshkanov M.K., Belginova S.A., Uvaliyeva I.M. Informational and analytical system of the Biochemical analysis of blood in children // Материалы международной научной конференции «Информатика и прикладная математика» («Computer science and Applied Mathematics»), посвященной 25-летию Независимости РК и 25-летию Института информационных и вычислительных технологий. Алматы – 2016. – С. 8-19.
7. Uvaliyeva I., Belginova, S., Ismukhamedova, A. Decision support system for diagnosing anemia // 2018 4th International Conference on Computer and Technology Applications (ICCTA 2018) – Istanbul, Turkey. 2018. – pp. 211-215.
8. Uvaliyeva I., Belginova, S., Ismukhamedova, A. Informational and analytical system to diagnose anemia // Proceedings of the Fourth International Conference on Engineering & MIS 2018 (ICEMIS '18), ACM. - Istanbul, Turkey. 2018. – pp. 1-8 doi:10.1145/3234698.3234712018
9. Kalimoldayev, M., Uvaliyeva I., Belginova, S., Ismukhamedova, A. IT infrastructure of e-health of the republic of Kazakhstan // Communications in Computer and Information Science. – Springer, Cham, 2018. – №998 – pp. 54-63.
10. Uvaliyeva I., Belginova, S., Ismukhamedova, A. Development and implementation of the algorithm of differential diagnostics // Proceedings of the

International Conference «Application of Information and Communication Technologies-AICT 2018». – Almaty, Kazakhstan. 2018. – pp. 276-281

11. Uvaliyeva I., Belginova, S., Ismukhamedova, A. Design and implementation of the diagnostic of algorithms complex for clinical and hematological symptoms // Proceedings of the 5th International Conference on Engineering and MIS 2019 (ICEMIS '19), ACM. - Astana, Kazakhstan. 2019. – pp. 13-18 (DOI: 10.1145/3330431.3330444)

12. Rustamov, S., Uvaliyeva I., Belginova, S., Ismukhamedova, A. Algorithm diagnosis of anemia on the basis of the method of the synthesis of the decisive rules // Proceedings of the International Conference «Application of Information and Communication Technologies-AICT 2019». – Baku, Azerbaijan. 2019. – pp. 8-12.

13. Назаренко Г. И., Гулиев Я. И., Ермаков Д. Е. Медицинские информационные системы: теория и практика /Подред. Г.И. Назаренко, Г.С. Осипова. — М.: ФИЗМАТЛИТ, 2005. — 320 с. — ISBN 5-9221-0594-9

14. Королук И.П.. Медицинская информатика : Учебник / - 2 изд., перераб. и доп. - Самара : ООО «Офорт» : ГБОУ ВПО «СамГМУ».2012.- 244 с. 2012

15. Кодекс Республики Казахстан от 18 сентября 2009 года № 193-IV «О здоровье народа и системе здравоохранения» (с изменениями и дополнениями по состоянию на 28.12.2018 г.)

16. Islam, Md. M., & Poly, T. N. (2018) Recent Advancement of Clinical Information Systems: Opportunities and Challenges. Yearbook of medical informatics 27(01):083-090. DOI: 10.1055/s-0038-1667075

17. Bellazzi, R., & Zupan, B. (2008) Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics 77(2):81-97. DOI: 10.1016/j.ijmedinf. 2006.11.006

18. Gozali, E., Rahimi, B., Sadeghi, M., & Safdari, R. (2017) Diagnosis of diseases using data mining. Medical Technologies Journal, Vol. 1, № 4. DOI: 10.26415/2572-004X-vol1iss4p120-121

19. Sen, I., & Khandelwal, K. (2018) Data Mining in Healthcare. Project: RSC Market Analysis. DOI: 10.13140/RG.2.2.22189.38887. Retrieved from https://www.researchgate.net/publication/322754945_DATA_MINING_IN_HEALTH_CARE

20. Niaksu, O., & Kurasova, O.(2012) Data Mining Applications in Healthcare Theory vs Practice. In Proceedings of Conference: Baltic DB&IS 2012. Vilnius, Lithuania: Vol-924

21. Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J., & Hua, L. (2011) Data Mining in Healthcare and Biomedicine: A Survey of the Literature. Journal of Medical Systems 36(4):2431-48. DOI: 10.1007/s10916-011-9710-5

22. Iavindrasana,J., Cohen, G., Depeursinge, A., Miller, H., Meyer, R., & Geissbuhler, A.(2009) Clinical Data Mining: a Review. Methods of Information in Medicine. DOI: 10.1055/s-0038-1638651

23. Chen, L-Y., & Fawcett T. Using Data Mining Strategies in Clinical

Decision Making: A Literature Review. *CIN Computers Informatics Nursing* 34(10):1. DOI: 10.1097/CIN.0000000000000282

24. Jothi, N., Abdul Rashid, N., & Husain W. (2015) Data Mining in Healthcare – A Review. *Procedia Computer Science* 72:306-313. DOI:10.1016/j.procs.2015.12.145

25. Kala, R., Shukla, A., Tiwari, R. (2010) Hybrid Intelligent Systems for Medical Diagnosis. In book: *Intelligent Medical technologies and Biomedical Engineering: Tools and Applications*. Chapter: 9, Publisher: IGI GlobalEditors

26. Wagholikar, K., Sundararajan, V., & Deshpande, A. (2011) Modeling Paradigms for Medical Diagnostic Decision Support: A Survey and Future Directions *Journal of Medical Systems* 36(5):3029-49. DOI: 10.1007/s10916-011-9780-4

27. Naydenova, E., Tsanas, A., Howie, S., Casals-Pascual, C., & De Vos, M. (2016) The power of data mining in diagnosis of childhood pneumonia. *Journal of The Royal Society Interface* 13(120):20160266. DOI:10.1098/rsif.2016.0266

28. Nalluri, M.R., Kannan, K., Manisha, M., & Roy, D.S. (2017) Hybrid Disease Diagnosis Using Multiobjective Optimization with Evolutionary Parameter Optimization. *Journal of Healthcare Engineering*. <https://doi.org/10.1155/2017/5907264>

29. Wiharto, W., Kusnanto, H., & Herianto, H. (2016) Intelligence System for Diagnosis Level of Coronary Heart Disease with K-Star Algorithm. *Healthcare Informatics Research* 22(1):30. DOI: 10.4258/hir.2016.22.1.30

30. Alizadeh, B., Safdari, R., Zolnoori, M., & Bashiri, A. (2015) Developing an Intelligent System for Diagnosis of Asthma Based on Artificial Neural Network. *Acta Informatica Medica* 23(4):220. DOI: 10.5455/aim.2015.23.220-223

31. Vandewiele, G., De Backere, F., Lannoye, K., Berghe, M., Janssens, O., Hoecke, S., Keereman, V., Paemeleire, K., Ongenaes F., & De Turck, F. (2018) A decision support system to follow up and diagnose primary headache patients using semantically enriched data. *BMC Medical Informatics and Decision Making* 18(1). DOI: 10.1186/s12911-018-0679-6

32. Selvathi, D., & AarthyPoornila, A. (2017) Performance analysis of various classifiers on deep learning network for breast cancer detection. *Proceedings of Conference: 2017 International Conference on Signal Processing and Communication (ICSPC)*. DOI: 10.1109/CSPC.2017.8305869

33. Kosmicki, JA., Sochat, V., Duda, M., & Wall, DP. (2015) Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry* 5(2):e514. DOI: 10.1038/tp.2015.7

34. Cabral, G., & Inacio de Oliveira, AL. (2014) One-class Classification for heart disease diagnosis. *Proceedings of Conference: 2014 IEEE International Conference on Systems, Man and Cybernetics – SMC*. DOI:10.1109/SMC.2014.6974311

35. Sepandi, M., Taghdir, M., Rezaianzadeh, A., & Rahimikazerooni, S. (2018) Assessing Breast Cancer Risk with an Artificial Neural Network. *Asian Pacific Journal of Cancer Prevention*. DOI:10.22034/APJCP.2018.19.4.1017

36. Shafique, U., Majeed, F., Qaiser, H., & Mustafa, I. (2015) Data Mining

in Healthcare for Heart Diseases. *International Journal of Innovation and Applied Studies*. ISSN 2028-9324 Vol. 10No. 4Mar.2015, pp. 1312-1322© 2015

37. Afifi, Sh.M., Gholamhosseini, H., & Sinha, R. (2018) Dynamic hardware system for cascade SVM classification of melanoma. *Neural Computing and Applications*. DOI: 10.1007/s00521-018-3656-1

38. Huang, M-W., Chen, Ch-W., Lin, W-Ch., Ke, Sh-W., & Tsai, Ch-F. (2017) SVM and SVM ensembles in breast cancer prediction. *PLoS ONE* 12(1):e0161501. DOI: 10.1371/journal.pone.0161501

39. Panyavaraporn, J., & Horkaew, P. (2018) Classification of Alzheimer's disease in PET scans using MFCC and SVM. DOI: 10.18517/ijaseit.8.5.6503

40. Jirka, J., Prauzek, M., Krejcar, O., & Kuca, K. (2018) Automatic epilepsy detection using fractal dimensions segmentation and GP-SVM classification. *Neuropsychiatric Disease and Treatment*, Volume 14:2439-2449. DOI:10.2147/NDT.S167841

41. Pereira, J., Peixoto, H., Machado, J., & Abelha, A. (2017) A Data Mining Approach for Cardiovascular Diagnosis. *Open Computer Science*. 2017; 7:36–40. DOI:10.1515/comp-2017-0007

42. Kaur, G., Arora, A.Sh., & Jain, V.K. (2017) Comparative Analysis of Hybrid Models for Prediction of BP Reactivity to Crossed Legs. *Journal of Healthcare Engineering*, Volume 2017, Article ID 2187904, 13 pages. <https://doi.org/10.1155/2017/2187904>

43. Jatav, Sh., & Sharma, V. (2018) An Algorithm for Predictive Data Mining Approach in Medical Diagnosis. *International Journal of Computer Science & Information Technology (IJCSIT)* Vol 10, No 1. DOI: 10.5121/ijcsit.2018.10102

44. Cichosz P., Jagodzinski, D., Matysiewicz, M., Neumann, L., Nowak, R., Okuniewski, R., & Oleszkiewicz, W. (2016) Novelty Detection for Breast Cancer Image Classification. *Proceedings of Conference: XXXVIII-th IEEE-SPIE Joint Symposium on Photonics, Web Engineering, Electronics for Astronomy and High Energy Physics Experiments*, At: Wilga, PolandVolume: 10031. DOI: 10.1117/12.2249183

45. Chen, Y-Ch., Ke, W-Ch., Chiu, H-W. (2014) Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Computers in Biology and Medicine* 48C(1):1-7. DOI:10.1016/j.compbimed.2014.02.006

46. Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y., & De Moor, B. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics* 22(14):e184-90. DOI:10.1093/bioinformatics/btl230

47. Masethe, H.D., & Masethe, M.A. (2014) Prediction of Heart Diseases using Classification Algorithms. *Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II. WCECS 2014, 22-24 October, 2014, San Francisco, USA*. ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

48. De Canete, J.F., Gonzalez-Perez, S., & Ramos-Diaz, J.C. (2011) Artificial neural networks for closed loop control of in silico and ad hoc type 1

diabetes. *Computer Methods and Programs in Biomedicine* 106(1):55-66. DOI:10.1016/j.cmpb.2011.11.006

49. Li, Y., Chou, Ch-A., & Wang, H. (2015) A Machine Learning Based Intelligent System for Diagnosis and Treatment. Conference: IIE Annual Conference & Expo 2015, At: Nashville, USA

50. Hamouda, SKM., Abo El Ezz HR., & Wahed ME. (2017) Intelligent System for Predicting, Diagnosis and Treatment of Breast Cancer. *International Journal of Biomedical Data Mining* 06(02). DOI: 10.4172/2090-4924.1000128

51. Hamouda, SKM., Abo El Ezz HR., & Wahed ME. (2017) Intelligent System for Predicting, Diagnosis and Treatment of Breast Cancer. *International Journal of Biomedical Data Mining* 06(02). DOI: 10.4172/2090-4924.1000128

52. Esfandiari. N., Babavalian, M., Moghadam, A-M., & Tabar, V. (2014) Knowledge discovery in medicine: Current issue and future trend *Expert Systems with Applications* 41(9):4434–4463. DOI:10.1016/j.eswa.2014.01.011

53. Sagi, A., Sabo, A., Kuljic, B., & Szakall, T. (2015) Multilevel Interdisciplinary Intelligent Systems in Medicine. Proceedings of Conference: International Symposium on Intelligent Systems and Informatics. DOI:10.1109/SISY.2015.7325356

54. Bellazzi, R., Ferrazzi, F., & Sacchi, L. (2011) Predictive data mining in clinical medicine: A focus on selected methods and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1(5):416-430. DOI:10.1002/widm.23

55. Xu, X., Zhang, Y., Zou, L., Wang, M., & Li, W. (2012) A gene signature for breast cancer prognosis using support vector machine. Proceedings of Conference: 5th International Conference on BioMedical Engineering and Informatics (BMEI). DOI: 10.1109/BMEI.2012.6513032

56. Azar, A., & El-Metwally, Sh. (2013) Decision tree classifiers for automated medical diagnosis. *Neural Computing and Applications* 23(7-8):2387-2403. DOI:10.1007/s00521-012-1196-7

57. Cichosz P., Jagodzinski, D., Matysiewicz, M., Neumann, L., Nowak, R., Okuniewski, R., & Oleszkiewicz, W. (2016) Novelty Detection for Breast Cancer Image Classification. Proceedings of Conference: XXXVIII-th IEEE-SPIE Joint Symposium on Photonics, Web Engineering, Electronics for Astronomy and High Energy Physics Experiments, At: Wilga, Poland Volume: 10031. DOI: 10.1117/12.2249183

58. P., Lequerica-Fernandez, P., Villalain, L., Pena, I., Sanchez-Lasheras, F., & de Vicente, J. (2013) Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines. *Expert Systems with Applications* 40(12):4770–4776. DOI:10.1016/j.eswa.2013.02.032

59. Bocharé, A., Gangopadhyay, A., Yesha, Yelena, Joshi, A., Yesha, Yaacov, Brady, M., Grasso, M., & Rishe, N. (2014) Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer. *International Journal of Medical Engineering and Informatics* 6(2):87 – 99.

DOI:10.1504/IJMEI.2014.060245

60. Li, Y., Chou, Ch-A., & Wang, H. (2015) A Machine Learning Based Intelligent System for Diagnosis and Treatment. Conference: IIE Annual Conference & Expo 2015, At: Nashville, USA

61. Iakovidis, D., & Smailis, Ch. (2012) A semantic model for multimodal data mining in healthcare information systems. Studies in health technology and informatics. DOI: 10.3233/978-1-61499-101-4-574

62. Kourou, K., Exarchos, Th., Exarchos, K., Karamouzis, M., & Fotiadis, D. (2014) Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal 13(C). DOI: 10.1016/j.csbj.2014.11.005

63. Mamiya H. et al. Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes //Journal of biomedical informatics. – 2015. – Т. 53. – С. 237-242.

64. Groenhof T. K. J. et al. The effect of computerized decision support systems on cardiovascular risk factors: a systematic review and meta-analysis //BMC medical informatics and decision making. – 2019. – Т. 19. – №. 1. – С. 108.

65. Williams P. A. et al. Usability testing and adaptation of the pediatric cardiovascular risk reduction clinical decision support tool //JMIR human factors. – 2016. – Т. 3. – №. 1. – С. e17.

66. Meric-Bernstam F. et al. A decision support framework for genomically informed investigational cancer therapy //JNCI: Journal of the National Cancer Institute. – 2015. – Т. 107. – №. 7.

67. Артеменко М. В., Бабков А. С. Информационная поддержка принятия решений в скрининг диагностики рака желудка на основе анализа крови //Научный вестник. – 2016. – №. 1. – С. 115-131.

68. Ганцев Ш. Х., Франц М. В. Искусственный интеллект как инструмент поддержки в принятии решений по диагностике онкологических заболеваний //Медицинский вестник Башкортостана. – 2018. – Т. 13. – №. 4 (76).

69. Бабков А., Артеменко М., Кирюткин М. В. Превентивная диагностика онкологических заболеваний желудка при скрининге с помощью системы поддержки принятия классификационных решений //Научное обозрение. Биологические науки. – 2017. – №. 2. – С. 5-13.

70. Berry D. L. et al. Evaluating Clinical Implementation Approaches for Prostate Cancer Decision Support //Urology Practice. – 2019. – Т. 6. – №. 2. – С. 93-99.

71. Beauchemin M. P. Supporting Clinical Decision Making in Cancer Care Delivery : дис. – Columbia University, 2019.

72. Шестерникова О. П. и др. Интеллектуальная система прогнозирования развития сахарного диабета у больных хроническим панкреатитом //Искусственный интеллект и принятие решений. – 2015. – №. 4. – С. 12-50.

73. Судаков О. В. и др. Разработка информационного комплекса

поддержки принятия врачебных решений на основе нейросетевого моделирования //Сборник материалов международной научной конференции. – 2018. – С. 23.

74. Кошербаева Л. К. и др. Анализ участия пациентов в программе управления заболеванием по сахарному диабету //Вестник Казахского Национального медицинского университета. – 2018. – №. 2.

75. Evans R. S. et al. Clinical decision support to efficiently identify patients eligible for advanced heart failure therapies //Journal of cardiac failure. – 2017. – Т. 23. – №. 10. – С. 719-726.

76. Barbieri C. et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients //Kidney international. – 2016. – Т. 90. – №. 2. – С. 422-429.

77. Купеева И. А., Разнатовский К. И., Раводин Р. А. Разработка интеллектуальной системы поддержки принятия врачебных решений в дерматовенерологии //Проблемы медицинской микологии. – 2015. – Т. 17. – №. 3. – С. 27-31.

78. Гаврилов Д. А. и др. Автоматизированная диагностика меланомы кожи на основе математической модели искусственной сверточной нейронной сети //Research'n Practical Medicine Journal. – 2018. – Т. 5. – №. 3.

79. Vandewiele G. et al. A decision support system to follow up and diagnose primary headache patients using semantically enriched data //BMC medical informatics and decision making. – 2018. – Т. 18. – №. 1. – С. 98.

80. Каменских Е. М. и др. Применение нейросети для оценки моторных нарушений при болезни паркинсона с использованием технологии видеозахвата в виртуальной реальности //Ульяновский медико-биологический журнал. – 2018. – №. 4.

81. Выучейская М. В., Крайнова И. Н., Грибанов А. В. Нейросетевые технологии в диагностике заболеваний (обзор) //Журнал медико-биологических исследований. – 2018. – Т. 6. – №. 3.

82. Шалькевич Л. В., Жевнеронок И. В. Прогнозирование развития эпилепсии у детей при однократном судорожном припадке //Журнал Гродненского государственного медицинского университета. – 2019. – Т. 17. – №. 5. – С. 587-591.

83. Okazaki E. M. et al. Usage of EpiFinder clinical decision support in the assessment of epilepsy //Epilepsy & Behavior. – 2018. – Т. 82. – С. 140-143.

84. Asha, T., Natarajan, S., & Murthy, K.N.B. (2012) Data Mining Techniques in the Diagnosis of Tuberculosis. In book: Understanding Tuberculosis - Global Experiences and Innovative Approaches to the Diagnosis. DOI: 10.5772/30504

85. Щекина Е. Н. Использование системного подхода для создания систем поддержки принятия решений в медицине (обзор литературы) //Вестник новых медицинских технологий. Электронное издание. – 2017. – Т. 11. – №. 2.

86. Banks, G (1986). Artificial intelligence in medical diagnosis: the INTERNIST/CADUCEUS approach. Critical Reviews in Medical Informatics. 1 (1):

87. Bennett, J.S. (1981). On the structure of the acquisition process for rule-based systems. *Mach. Intell.* - Maidenhead, 1981. - P189-203 . - ISBN 008-028556-2
88. Web-портал сервиса Polised.com. URL: <https://www.polised.com/analiz-krovi.html> (дата обращения: 24.08.2019).
89. Web-портал сервиса TestResult.org. URL: <https://testresult.org/> (дата обращения: 24.08.2019).
90. Левина А. А. и др. Дифференциальная диагностика анемий //Клиническая лабораторная диагностика. – 2015. – Т. 60. – №. 12.
91. Смирнова Л. А. Анемии: дифференциально-диагностические аспекты //Медицинские новости. – 2013. – №. 2 (221).
92. Копелиович Д. И., Рыженков Д. А. Функциональное моделирование процесса мониторинга данных //Мониторинг. Наука и технологии. – 2016. – №. 1. – С. 49-53.
93. Waissi G. R. et al. Automation of strategy using IDEF0—A proof of concept //Operations Research Perspectives. – 2015. – Т. 2. – С. 106-113.
94. Дерябкин В. П. Концепция построения базы знаний на основе использования унифицированных фреймовых структур //Традиции и инновации в строительстве и архитектуре. Строительные технологии. – 2017. – С. 589-592.
95. Виен Л. Н. Дистанционная диагностическая система на основе гибридных моделей знаний : дис. – Виен Ле Нгуен.–Волгоград: ВолГТУ, 2015.–123 с, 2015.
96. Надыров Э.А., Прокопович А.С., Никонович С.Н., Смирнова Л.А. Алгоритм диагностики анемий. Инструкция по применению. Белорусская медицинская академия последипломного образования. Гомель, 2013
97. Бова А. А., Валувевич В. В., Рудой А. С. Железодефицитная анемия: диагностика, лечение, профилактика. – 2017.
98. Зубков Е. В. Алгоритмы и методики интеллектуального анализа событий информационной безопасности в сетях и системах телекоммуникаций : дис. – автореф. дис. на соиск. учен. степ. канд. техн. наук (05.12. 13)/ФГБОУ ВО, 2016.
99. Graczyk M. et al. Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal //Asian conference on intelligent information and database systems. – Springer, Berlin, Heidelberg, 2010. – С. 340-350.
100. Увалиева И.М. Разработка распределенной информационно-аналитической системы образовательной статистики : Диссертация на соискание ученой степени доктора философии (PhD) (6D070300 – Информационные системы)/КазНИТУ им.К.Сатпава, 2015.
101. MIMIC-III Critical Care Database URL: <https://mimic.physionet.org/about/mimic/> (дата обращения: 24.08.2019).
102. Mosa M. J., Albatish I., Abu-Naser S. S. ASP. NET-Tutor: Intelligent Tutoring System for leaning ASP. NET. – 2018.
103. Troelsen A., Japikse P. Introducing ASP. NET MVC //Pro C# 7. –

Apress, Berkeley, CA, 2017. – С. 1179-1221.

104. Mukherjee S. SQL Server Development Best Practices //International Journal of Innovative Research in Computer and Communication Engineering. – 2019. – Т. 10.

105. Ядгарова Ю. В., Таратухин В. В. Методика выбора шаблона программной архитектуры на основе анализа параметров качества системы //Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2019. – Т. 12. – №. 2. – С. 28-38.

106. Буч Г., Якобсон И., Рамбо Д. Язык UML. Руководство пользователя. – Litres, 2017.

107. Кобринский Б. А. Триединство факторов уверенности в задачах медицинской диагностики //Искусственный интеллект и принятие решений. – 2018. – №. 2. – С. 62-72.

108. Ремезова Е. М. Модели поддержки принятия решений при подготовке проекта внедрения КИС на основе многоагентных систем и аппарата нечетких множеств. – 2017.

109. Корневский Н. А. Использование нечеткой логики принятия решений для медицинских экспертных систем //Медицинская техника. – 2015. – Т. 49. – №. 1. – С. 33-5.

110. Сагдатуллин А. М., Муравьева Е. А., Емекеев А. А. Классическая теория нечеткого управления в контексте методов дефаззификации //Современные проблемы науки и образования в техническом вузе. – 2015. – С. 198-204.

111. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. — СПб.: БХВ Петербург, 2005. — 736 с.

112. Deera P., Sivakumar R. Synthesis of heuristic control strategies for liquid level control in spherical tank //2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). – IEEE, 2017. – С. 316-319.

113. Волокобинский М. Ю., Пекарская О. А., Рази Д. А. Принятие решений на основе метода анализа иерархий //Финансы: Теория и Практика. – 2016. – №. 2 (92).

ПРИЛОЖЕНИЕ А

Акт внедрения результатов диссертационного исследования

«УТВЕРЖДАЮ»
Директор
Городского ревматологического центра
Калембаева А.Б.
2019 г



АКТ
внедрения на базе ГКП на ПХВ «Городской ревматологический центр г.Алматы»
результатов диссертационной работы «Разработка информационной технологии
поддержки принятия решений при диагностировании заболеваний крови»

Наименование предложения: Программный модуль диагностирования клинко-гематологических синдромов

Форма внедрения: методика

Область применения: медицина, ревматология

Ответственные за внедрение: Увалиева И.М., Бельгинова С.А., Исмухамедова А.М.

Основное содержание внедрения, разработанного в рамках научного исследования:

Технология программного модуля диагностирования клинко-гематологических синдромов основана на алгоритме диагностики анемий, позволяющий классифицировать анемии по морфологическим признакам и дать количественную характеристику выраженности патологического процесса. При помощи предложенных алгоритмов можно с высокой долей вероятности выявить ряд наиболее часто встречающихся анемий (железодефицитные анемии, анемии хронических заболеваний, В12-дефицитную анемию).

Технология программного модуля включает 3 основных этапа диагностики анемии:
- установление наличия и степени выраженности анемического синдрома;
- определение характера анемий в соответствии с морфологической классификацией;
- подтверждение и уточнение характера анемии методами клинической биохимии и иммуноферментного анализа.

Таким образом, предлагаемый математический способ многокритериальной оценки результатов исследования крови позволяет количественно охарактеризовать их принадлежность к различным видам анемий, оценка может быть проведена при смешанных формах патологии.

Эффективность внедрения: Использование данной методики позволяет:

- 1) ускорить процесс постановки диагноза;
- 2) облегчить работу врача, тем самым повысить производительность труда медицинского работника;
- 3) повысить качество предоставляемых медицинских услуг;
- 4) сократить затраты на проведение работ по обслуживанию пациентов

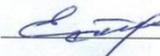
Предложения, замечания учреждения, осуществляющего внедрение: Замечаний нет.

Сроки внедрения: в течение 2019 года

Председатель комиссии  Ахметова Ж.С.

Члены комиссии:  Калыкова М.Б.

 Батырханова Н.М.

 Ефанова Е.С.

ПРИЛОЖЕНИЕ Б

Акт внедрения результатов диссертационного исследования

УТВЕРЖДАЮ
АО «Казахский институт онкологии и радиологии»
И.о. Председателя Правления
Д.Р.Кайдарова
2019 г.



АКТ
внедрения на базе АО «Казахский институт онкологии и радиологии»
результатов диссертационной работы «Разработка информационной
технологии поддержки принятия решений при диагностировании
заболеваний крови»

Наименование предложения: Программный модуль «Оценка информативности диагностических показателей клинико-гематологических синдромов»

Форма внедрения: программный продукт

Область применения: медицина, гематология, лаборатория

Ответственные за внедрение: Увалиева И.М., Бельгинова С.А.

Результаты диссертации, в качестве программного модуля оценки информативности диагностических показателей клинико-гематологических синдромов были внедрены в процесс деятельности медицинского учреждения АО «Казахский институт онкологии и радиологии» для исследований биохимического анализа крови пациентов.

Программный модуль вычисления информативности диагностических показателей позволяет повысить эффективность диагностирования заболеваний крови за счет структурирования исходных данных биохимического анализа крови и сокращения количества сущностей, требующих экспертного анализа. Алгоритм отражает особенности внутренней структуры и скрытых взаимосвязей в клинико-гематологических синдромах.

Применение разработанного программного модуля позволяет извлекать эту информацию и предоставлять ее в распоряжение специалиста, тем самым давая возможность соотносить между собой различные группы клинико-гематологических синдромов.

Председатель комиссии:

Д.Р. Кайдарова

Члены комиссии:



Н.Т.Балтабеков
Г.Б. Турсунова
Ж.С. Кайдарова
Ж.Ж Касымбекова

ПРИЛОЖЕНИЕ В

Свидетельство о государственной регистрации прав на объект авторского права «Программный модуль диагностирования клинико-гематологических синдромов» (программа для ЭВМ)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО
О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ
№ 4737 от «1» августа 2019 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):
УВАЛШЕВА ИНЛПА МАХМУТОВНА, БЕЛГШИНОВА САУЛЕ АСКЕРБЕКОВНА, ИСМУХАМЕДОВА
АЙГЕРИМ МЭЛСАТОВНА

Вид объекта авторского права: программа для ЭВМ

Название объекта: Программный модуль диагностирования клинико-гематологических синдромов

Дата создания объекта: 08.07.2019





Курорт, тундасқайсылығы: <http://www.kazpatent.kz> қолданылған;
"Авторлық құқық" бағамында: Телдереуші: <https://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте kazpatent.kz
в разделе «Авторское право»: <https://copyright.kazpatent.kz>

Подписано ЭЦП Оспанов Е.К.

ПРИЛОЖЕНИЕ Г

Скриншот страниц сайта международных баз цитирования Scopus, Clarivate Analytics (Web of Science) с указанием h-index докторанта

Belginova, Saule Просмотр потенциальных соответствий авторов

<http://orcid.org/0000-0002-7238-6016>

Организация (и): [D.Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan](#) [Смотреть больше](#)

Отрасль знаний: [Computer Science](#) [Mathematics](#) [Medicine](#) [Decision Sciences](#)

Документы автора: 6

Общее количество цитирований: 6 по 4 документам

h-индекс: 2

Документ и тенденции цитирования:

Год	Документы	Цитирования
2018	1	4
2019	2	2

6 ДОКУМЕНТОВ Цитирования в 4 документах Соавторов: 5 Темы

Пользователи, работающие в режиме предварительного просмотра, могут видеть последние 10 документов автора. [View 160 references](#)

Задать оповещение о документе

Название документа	Авторы	Год	Источник	Цитирования
Design and implementation of the diagnostic algorithms complex for clinical and hematological symptoms	Uvaliyeva, I., Belginova, S., Alimbayeva, Z., Ismukhamedova, A.	2019	ACM International Conference Proceeding Series	0
The application of data mining methods for the process of diagnosing diseases	Belginova, S., Uvaliyeva, I., Rustamov, S.	2019	Journal of Theoretical and Applied Information Technology	0
IT infrastructure of e-health of the republic of kazakhstan	Kalimoldayev, M., Belginova, S., Uvaliyeva, I., Ismukhamedova, A.	2019	Communications in Computer and Information Science	0
Development and implementation of the algorithm of differential diagnostics	Uvaliyeva, I., Belginova, S., Ismukhamedova, A.	2018	IEEE 12th International Conference on Application of Information and Communication Technologies, AICT 2018 - Proceedings	0
Decision support system for diagnosing anemia	Belginova, S., Uvaliyeva, I., Ismukhamedova, A.	2018	2018 4th International Conference on Computer and Technology Applications, ICCTA 2018	4
Informational and analytical system to diagnose anemia	Indira, U., Belginova, S., Ismukhamedova, A.	2018	ACM International Conference Proceeding Series	2

Пользователи, работающие в режиме предварительного просмотра, могут видеть последние 10 документов автора. [^](#) Верх страницы

Web of Science [v.5.34] - Отчет | [apps.webofknowledge.com/CitationReport.do?product=WOS&search_mode=CitationReport&SID=E3nHPMvLVPkxabqBxg&page=1&cr_pqid=1...](#)

Поиск Возврат к результатам поиска Инструменты Поисковые запросы и оповещения История поиска Список отмеченных публикаций

Отчет по цитированию 2 результаты из Web of Science Core Collection между 1975 и 2020 [Выполнить](#)

Вы искали: АВТОР: (belginova) ...Больше

Данный отчет отражает цитирования источников, проиндексированных в Web of Science Core Collection. Выполните поиск по приставной библиографии, чтобы включить цитирования документов, не проиндексированных в Web of Science Core Collection.

Экспорт данных: [Сохранить в фай...](#)

Всего публикаций 2 Анализировать	h-index 1 Среднее число цитирований документа 0,5	Суммарное количество цитирований 1 без самоцитирования 0	Цитирующие статьи 1 Анализировать без самоцитирования 0
--------------------------------------------	---------------------------------------------------------------------------------	---------------------------------------------------------------------------------	--------------------------------------------------------------------------------

ПРИЛОЖЕНИЕ Д

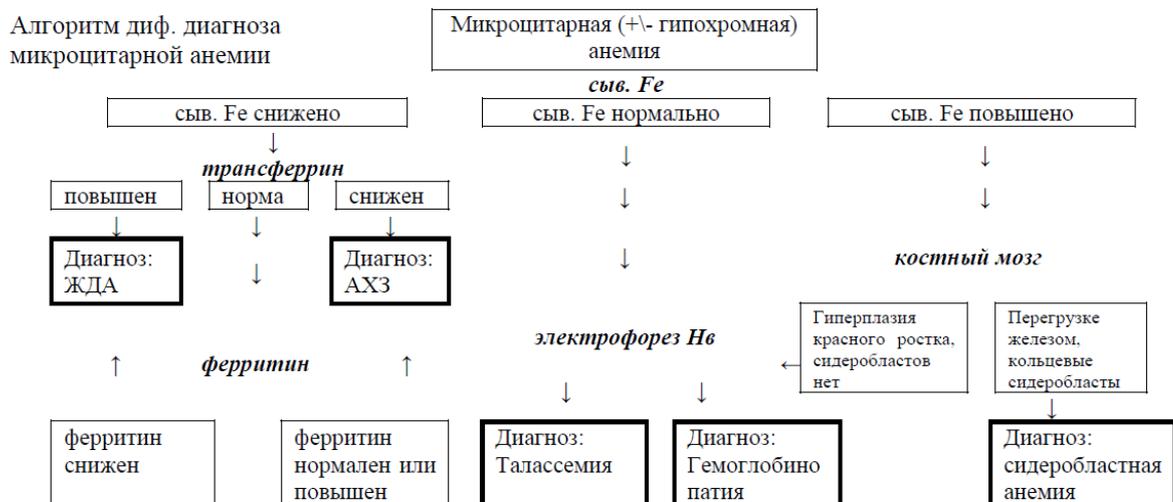


Рисунок Д.1. Дифференциальный диагноз микроцитарных анемий

Таблица Д.1 - Дифференциальные признаки микроцитарных анемий

	ЖДА *	АХЗ *	Талассемия и аномалии Hb	Сидеробластная
Особенности морфологии эритроцитов, отмечаемые врачом-лаборантом:				
• Анизоцитоз (RDW > 15%)	+++	+/-	+/-	++
• Базофильная пунктация и клетки - мишени	-	-	++	++
	-/+	-/+	+	+
Исследование запасов железа:				
Сывороточное железо	↓	↓	N или ↑	↑
Трансферрин	↑ или N	↓	↓ или N	↓
КНТ	< 16% (< 5% !)	< 16%		↑
Ферритин (нг/мл)	< 30	> 100	> 100	> 100
Гемосидерин в костном мозге (пункция или трепанобиопсия)	резко снижен	норма	повышен	резко повышен, кольцев. сидеробл.
Увеличение уровней СРБ, фибриногена, ускорение СОЭ	+/-	+++	+/-	+/-
Характерные жалобы (кроме астенических)	Слоятся ногти, выпадают волосы, появление пристрастия к необычной еде и запахам (риса), дисфагия		Характерных жалоб нет	
Характерные признаки при осмотре	Изменения ногтевых пластин, атрофия сосочков языка		Характерных признаков нет	
Темп развития анемии	постепенное углубление (месяцы и годы)	довольно быстрое (недели)	анемия стойкая, многолетняя, во всех ан. грови MCV и MCH снижены однотипно	может быть врожденной или приобретенной

ПРИЛОЖЕНИЕ Е

Алгоритм диф. диагноза
микроцитарной анемии



Рисунок Е.1. Дифференциальный диагноз макроцитарных анемий

Таблица Е.1 - Дифференциальные признаки макроцитарных анемий

	Мегалобластные		Немегалобластные
	В-12 дефицитная	Фолиев.дефицит.	
Морфологические признаки мегалобластоидного кроветворения			
• Выраженность макроцитоза	MCV > 110		MCV < 110
• Кол-во ретикулоцитов *)	↓		N или ↑
• Гиперсегментоз нейтрофилов	+++		-
• Панцитопения	+		+/-
• Мегалобласты и гигантские метамиелоциты в костном мозге	+		-
Характерные жалобы (кроме астенических)	жжение и пощипывание языка, снижение аппетита, парестезии	нет	нет
Характерные признаки при осмотре	«лакированный» язык, субиктеричность	признаки недостаточного питания	нет
Заболевания и состояния, на фоне которых развилась анемия	отсутствие фактора Кастла (пернициозная анемия); гастрэктомия и резекция желудка; фистулы, анастомозы, дивертикулезы и слепые петли тонкого кишечника; заболев. поджелудочной железы;	неадекватное питание, алкоголизм и цирроз, глютеновая энтеропатия и спру, беременность, медикаменты	алкоголизм, забол. печени, МДС, гипотиреоз, ХОБЛ, гемолиз, о. кровопотеря

ПРИЛОЖЕНИЕ Ж



Рисунок Ж.1. Дифференциальный диагноз нормоцитарных анемий

Таблица Ж.1 - Дифференциальные признаки нормоцитарных (нормохромных) анемий

	Гемолитическая анемия	Острая постгеморрагическая анемия	Анемия при болезни почек и т.д.	Анемия при аплазии или опухоли костного мозга
Особенности морфологии периферической крови и костного мозга				
• уровень ретикулоцитов	↑	↑	N или ↓	N или ↓
• уровень лейкоцитов	N или ↑	N или ↑	м. б. различными	м. б. различными
• уровень тромбоцитов	N или ↓	N или ↑		↓
• расширение красного ростка в костном мозге	+	+	+/-	-
• картина аплазии или опухоли костного мозга				+
Биохимическое исследование крови				
Билирубин	↑ (преимущественно непрямой)	N		
ЛДГ	↑	N		
Сыв. Fe	↑	N или ↓		
гаптоглобин	N или ↓	N		
Увеличение селезенки	+/-	-	-/+	-/+

ПРИЛОЖЕНИЕ И

Фрагмент кода программной реализации алгоритма Staking-ансамблирования интеллектуальных методов анализа данных

```
namespace SPMining
{
    /// <summary>
    /// Вспомогательный класс
    /// </summary>
    public static class SPortalMiningProc
    {
        /// <summary>
        /// Считать структуру исходных данных
        /// </summary>
        /// <param name="strName">Имя
        структуры</param>
        /// <returns>Таблица</returns>
        internal static DataTable
        CreateTable(string strName, string
        keyFieldName)
        {
            AdomdCommand cmd = new
            AdomdCommand("SELECT top 1 * FROM [" +
            strName + "].CASES WHERE IsTestCase()");
            DataTable dt = new DataTable();

            using(AdomdDataReader reader =
            cmd.ExecuteReader())
            {
                reader.Read();
                for(int i = 0; i <
                reader.FieldCount; i++)
                {
                    DataColumn dc = new
                    DataColumn(reader.GetName(i),
                    reader.GetFieldType(i));
                    dt.Columns.Add(dc);
                    if (dc.ColumnName ==
                    keyFieldName)
                    {
                        dt.PrimaryKey = new
                        DataColumn[] { dc };
                    }
                }

                return dt;
            }

            /// <summary>
            /// Считать проверочные данные
            /// </summary>
            /// <param name="dt">Таблица</param>
            /// <param name="strName">Имя
            структуры</param>
            internal static void
            ReadTable(DataTable dt, string strName)
            {
                AdomdCommand cmd = new
                AdomdCommand("SELECT * FROM [" + strName +
                "].CASES WHERE IsTestCase()");
                string keyName =
                dt.PrimaryKey[0].ColumnName;

                int keyIndex = -1;
                using(AdomdDataReader reader =
                cmd.ExecuteReader())
                {
                    while(reader.Read())
                    {
                        if (keyIndex == -1)
                        {
                            for(int i = 0; i < reader.FieldCount; i++)
                            {
                                if (reader.GetName(i) == keyName)
                                {
                                    keyIndex = i;
                                    break;
                                }
                            }
                        }
                        object keyValue =
                        reader[keyIndex];
                        if
                        (dt.Rows.Find(keyValue) == null)
                        {
                            DataRow rw =
                            dt.NewRow();
                            rw.BeginEdit();
                            for(int i = 0; i <
                            reader.FieldCount; i++)
                            {
                                if (reader[i] !=
                                DBNull.Value)
                                {
                                    rw[reader.GetName(i)] = reader[i];
                                }
                            }
                            rw.EndEdit();
                            dt.Rows.Add(rw);
                        }
                    }
                }

                /// <summary>
                /// Получение прогноза по моделям
                /// </summary>
                /// <param name="dt">Таблица с
                данными</param>
                /// <param name="dc">Колонка, куда
                записывается вероятность</param>
                /// <param
                name="structure">Структура</param>
                /// <param
                name="predictField">Прогнозируемое
                поле</param>
                /// <param
                name="value">Прогнозируемое значение</param>
                internal static void
                Prediction(DataTable dt, DataColumn dc,
                MiningStructure structure, string
                predictField, string value)
            }
        }
    }
}
```

```

        {
            foreach (MiningModel m in
structure.MiningModels)
            {
                AdomdCommand cmd = new
AdomdCommand();

                cmd.CommandText = "SELECT
T.[" + dt.PrimaryKey[0].ColumnName + "], T.["
+ predictField +
                "], PredictProbability(["
+ m.Name + "].[" + predictField + "], '" +
value + "') FROM [" +
m.Name + "] NATURAL PREDICTION JOIN (SELECT *
FROM [" +
                structure.Name + "].CASES
WHERE IsTestCase()) AS T";

                using(AdomdDataReader reader
= cmd.ExecuteReader())
                {
                    while(reader.Read())
                    {
                        object id = reader[0];
                        string pVal = reader[1].ToString();
                        decimal v = reader.GetDecimal(2);

                        DataRow rw = dt.Rows.Find(id);
                        if (rw[dc] == DBNull.Value)
                            {
                                rw[dc] = v;
                                continue;
                            }

                        if (rw.Field<string>(predictField) == value)
                            {
                                rw[dc] =
Math.Max(rw.Field<decimal>(dc), v);
                            }
                        else
                            {
                                rw[dc] = Math.Min(rw.Field<decimal>(dc), v);
                            }
                    }
                }

                /// <summary>
                /// Получить список вариантов для
ансамбля
                /// </summary>
                /// <param name="AssemblyName">Имя
ансамбля</param>
                /// <returns>Список
вариантов</returns>
                internal static List<MiningStructure>
GetAssembly(string AssemblyName)
                {
                    List<MiningStructure> r = new
List<MiningStructure>();
                    foreach(MiningStructure ms in
Context.MiningStructures)

```

```

                {
                    if (ms.Name.StartsWith(AssemblyName))
                        {
                            r.Add(ms);
                        }
                }
                return r;
            }

            /// <summary>
            /// Генерация для запроса обновления
списка столбцов
            /// </summary>
            /// <param name="cc">Список столбцов
в исходной таблице</param>
            /// <param name="cols">Список
столбцов в модели</param>
            /// <returns>Строка со списком
столбцов для запроса</returns>
            internal static string
GenerateColumnInsert(DataColumnCollection cc,
MiningModelColumnCollection cols)
            {
                string r = "";
                foreach (DataColumn c in cc)
                {
                    if (r.Length > 0)
                        {
                            r += ",";
                        }
                    if (cols.Find(c.ColumnName) == null)
                        {
                            r += "SKIP";
                        }
                    else
                        {
                            r += "[" + c.ColumnName + "];";
                        }
                }
                return r;
            }

            /// <summary>
            /// Формирование списка классов из
строки
            /// </summary>
            /// <param name="cls">Строка со
списком классов</param>
            /// <returns>Список классов</returns>
            internal static
List<KeyValuePair<string, decimal>>
GetClasses(string cls)
            {
                List<KeyValuePair<string,
decimal>> lst = new List<KeyValuePair<string,
decimal>>();

                NumberFormatInfo nf = new
NumberFormatInfo()
                { NumberDecimalSeparator = "." };
                foreach(string s in cls.Split(';'))
                {
                    string[] c = s.Split('=');

```

```

        lst.Add(new
KeyValuePair<string, decimal>(c[0],
Convert.ToDecimal(c[1], nf)));
    }

    return lst;
}

/// <summary>
/// Получить данные для обучения
метамодели по алгоритму стэкинга по набору
переводного балла GPA
/// </summary>
/// <returns>Таблица с
данными</returns>
public static DataTable
GenerateTableGPA()
{
    List<MiningStructure> msl =
GetAssembly("Макроцитарная анемия ");
    DataTable dt =
CreateTable(msl[0].Name, "Ид");
    DataColumn dc = new
DataColumn("Вероятность", typeof(decimal));
    dt.Columns.Add(dc);
    foreach (MiningStructure ms in msl)
    {
        ReadTable(dt, ms.Name);
        Prediction(dt, dc, ms, "
Макроцитарная анемия ", "Нет");
    }
    return dt;
}

/// <summary>
/// Обучение метамодели по набору
переводного балла GPA
/// </summary>
/// <param name="classInfo"></param>
public static void StackingGPA(string
classInfo)
{
    List<KeyValuePair<string, decimal>> cls =
GetClasses(classInfo);

    DataTable dt = GenerateTableGPA();
    DataColumn dc = new
DataColumn("Класс", typeof(string));
    dt.Columns.Add(dc);
    foreach(DataRow rw in dt.Rows)
    {
        foreach(KeyValuePair<string, decimal> c in
cls)
        {
            if (rw.Field<decimal>("Вероятность") >=
c.Value)
            {
                rw.BeginEdit();

                rw.SetField<string>(dc, c.Key);
                rw.EndEdit();
                break;
            }
        }
    }
}

```

```

    }
}

AdomdCommand cmdClear = new
AdomdCommand("DELETE FROM MINING MODEL [" +
Context.MiningModels["Стэкинг
- Макроцитарная анемия "].Parent.Name + "]);
cmdClear.ExecuteNonQuery();

AdomdCommand cmd = new
AdomdCommand();
cmd.CommandText = " INSERT INTO
MINING MODEL [Стэкинг - Макроцитарная
анемия](" +

GenerateColumnInsert(dt.Columns,
Context.MiningModels["Стэкинг - Макроцитарная
анемия "].Columns) +
") @Table";
cmd.Parameters.Add("Table", dt);
cmd.ExecuteNonQuery();
}

/// <summary>
/// Получить данные для обучения
метамодели по алгоритму стэкинга по
трудоустройству
/// </summary>
/// <returns></returns>
public static DataTable
GenerateTableEmployment()
{
    List<MiningStructure> msl =
GetAssembly("Нормоцитарная анемия");
    DataTable dt =
CreateTable(msl[0].Name, "Ид");
    DataColumn dc = new
DataColumn("Вероятность", typeof(decimal));
    dt.Columns.Add(dc);
    foreach (MiningStructure ms in
msl)
    {
        ReadTable(dt, ms.Name);
        Prediction(dt, dc, ms, "
Нормоцитарная анемия", "Не обнаружено");
    }
    return dt;
}

/// <summary>
/// Обучение метамодели по
трудоустройству
/// </summary>
/// <param name="classInfo">Список
классов</param>
public static void
StackingEmployment(string classInfo)
{
    List<KeyValuePair<string,
decimal>> cls = GetClasses(classInfo);

    DataTable dt =

```

```

GenerateTableEmployment();
    DataColumn dc = new
DataColumn("Класс", typeof(string));
    dt.Columns.Add(dc);
    foreach (DataRow rw in dt.Rows)
    {
        foreach (KeyValuePair<string,
decimal> c in cls)
        {
            if
(rw.Field<decimal>("Вероятность") >= c.Value)
            {
                rw.BeginEdit();
            }
        }
    }
}

```

```

        AdomdCommand cmdClear = new
AdomdCommand("DELETE FROM MINING MODEL [" +
        Context.MiningModels["Стэкинг
- Нормоцитарная анемия"].Parent.Name + "]);
        cmdClear.ExecuteNonQuery();

        AdomdCommand cmd = new
AdomdCommand();
        cmd.CommandText = " INSERT INTO
MINING MODEL [Стэкинг - Нормоцитарная
анемия](" +
GenerateColumnInsert(dt.Columns,
Context.MiningModels["Стэкинг - Нормоцитарная
анемия"].Columns) +
        ") @Table";
        cmd.Parameters.Add("Table", dt);
        cmd.ExecuteNonQuery();
    }
}

```

ПРИЛОЖЕНИЕ К

Скриншоты некоторых окон программной реализации



Рисунок К.1. Главное окно программы

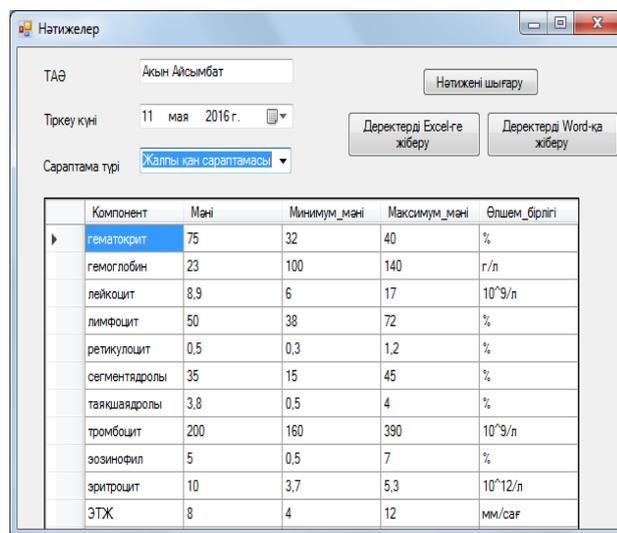


Рисунок К.3. Окно показателей анализа крови пациента

Компонент	МинимумМәні	МаксимумМәні	ӨлшемБірілігі
аланинаминотрансфераза	0	40	мккат/л
альбумин	37	55	г/л
аспартатаминотрансфераза	0	40	мккат/л
билирубин жалпы	3,4	20,7	ммоль/л
билирубин тура	0,83	3,4	ммоль/л
билирубин тура емес	2,56	17,3	ммоль/л
глюкоза	3,3	6,1	ммоль/л
жалпы кальций	2,3	2,87	ммоль/л
жалпы нәруыз	62	82	г/л
калий	3,6	5,1	ммоль/л
креатинин	35	110	ммоль/л
липаза	0	183	мккат/л
магний	0,7	1,2	ммоль/л
мыс	11	25	мккат/л
натрий	132	156	ммоль/л
несепнәр	4,3	7,3	ммоль/л
темір	9,3	32	мккат/л
холестерин	3,7	6,5	ммоль/л

Рисунок К.2. Окно нормативных значений показателей

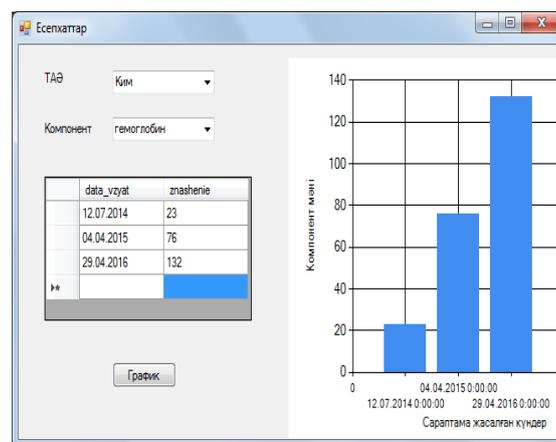


Рисунок К.4. Окно графика показателей анализа крови

ПРИЛОЖЕНИЕ Л

Таблица Л.1 - Матрица сравнений альтернатив по отношению к их воздействию на критерии второго уровня

	1	2	3	4	Нормализованные оценки
Адекватность					
Комплексное применение предложенных средств ППР	1	2	4	9	0,5166516
Частичное применение предложенных средств ППР	1/2	1	2	8	0,298289
Применение других существующих средств ППР	1/4	1/2	1	3	0,1387946
Отсутствие специализированных средств ППР	1/9	1/8	1/3	1	0,0462649
ОС = 0,0119379					
Актуальность					
Комплексное применение предложенных средств ППР	1	4	6	9	0,5925265
Частичное применение предложенных средств ППР	1/3	1	3	7	0,2514021
Применение других существующих средств ППР	1/5	1/2	1	3	0,1143865
Отсутствие специализированных средств ППР	1/9	1/7	1/3	1	0,0416849
ОС = 0,1513685					
Системность					
Комплексное применение предложенных средств ППР	1	3	6	8	0,5929064
Частичное применение предложенных средств ППР	1/3	1	3	5	0,2559399
Применение других существующих средств ППР	1/6	1/3	1	1	0,0830955
Отсутствие специализированных средств ППР	1/8	1/5	1	1	0,0680582
ОС = 0,0217249					
Компетентность					
Комплексное применение предложенных средств ППР	1	2	5	8	0,5089161
Частичное применение предложенных средств ППР	1/2	1	4	7	0,329159
Применение других существующих средств ППР	1/5	1/4	1	5	0,1203258
Отсутствие специализированных средств ППР	1/8	1/7	1/5	1	0,0415991
ОС = 0,0542466					
Оперативность					
Комплексное применение предложенных средств ППР	1	4	5	8	0,6061111
Частичное применение предложенных средств ППР	1/4	1	2	4	0,2026655
Применение других существующих средств ППР	1/5	1/2	1	5	0,1433061
Отсутствие специализированных средств ППР	1/8	1/4	1/5	1	0,0479173
ОС = 0,0583603					
Гибкость					
Комплексное применение предложенных средств ППР	1	3	4	8	0,5352499
Частичное применение предложенных средств ППР	1/3	1	5	7	0,3160279
Применение других существующих средств ППР	1/4	1/5	1	2	0,0961587
Отсутствие специализированных средств ППР	1/8	1/7	1/2	1	0,0525635
ОС = 0,0898737					

Наглядность					
Комплексное применение предложенных средств ППР	1	5	6	8	0,6440232
Частичное применение предложенных средств ППР	1/5	1	3	4	0,203658
Применение других существующих средств ППР	1/6	1/3	1	2	0,0944687
Отсутствие специализированных средств ППР	1/8	1/4	1/2	1	0,05785
OC = 0,0596626					
Полнота					
Комплексное применение предложенных средств ППР	1	3	4	9	0,5583383
Частичное применение предложенных средств ППР	1/3	1	2	6	0,2449382
Применение других существующих средств ППР	1/4	1/2	1	5	0,1539967
Отсутствие специализированных средств ППР	1/9	1/6	1/5	1	0,0427268
OC = 0,0352435					
Конструктивность					
Комплексное применение предложенных средств ППР	1	3	6	8	0,587083
Частичное применение предложенных средств ППР	1/3	1	3	5	0,2534261
Применение других существующих средств ППР	1/6	1/3	1	3	0,1082857
Отсутствие специализированных средств ППР	1/8	1/5	1/3	1	0,0512051
OC = 0,0341906					
Структуризация					
Комплексное применение предложенных средств ППР	1	4	6	9	0,6015277
Частичное применение предложенных средств ППР	1/4	1	4	7	0,2552212
Применение других существующих средств ППР	1/6	1/4	1	5	0,1060064
Отсутствие специализированных средств ППР	1/9	1/7	1/5	1	0,0372447
OC = 0,1118618					
Эргономичность					
Комплексное применение предложенных средств ППР	1	3	3	8	0,5510326
Частичное применение предложенных средств ППР	1/3	1	2	3	0,2249581
Применение других существующих средств ППР	1/3	1/2	1	3	0,1590694
Отсутствие специализированных средств ППР	1/8	1/3	1/3	1	0,0649398
OC = 0,0207783					
Интерактивность					
Комплексное применение предложенных средств ППР	1	3	5	9	0,5609801
Частичное применение предложенных средств ППР	1/3	1	4	7	0,2876559
Применение других существующих средств ППР	1/5	1/4	1	4	0,110058
Отсутствие специализированных средств ППР	1/9	1/7	1/4	1	0,0413061
OC = 0,0672790					
Совместимость					
Комплексное применение предложенных средств ППР	1	2	3	8	0,4713051
Частичное применение предложенных средств ППР	1/2	1	4	7	0,3463569
Применение других существующих средств ППР	1/3	1/4	1	4	0,1360541
Отсутствие специализированных средств ППР	1/8	1/7	1/4	1	0,0462839
OC = 0,0541615					