

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.891

Барсук
Андрей Сергеевич

Математическое и программно-алгоритмическое обеспечение экспертной
системы для диагностики сердечно-сосудистых заболеваний человека

АВТОРЕФЕРАТ
на соискание степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Скудняков Ю.А.
к.т.н., доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

Сердечно-сосудистые заболевания (ССЗ) – это группа заболеваний сердца и сосудов, к которой относятся: ишемическая болезнь сердца (ИБС), болезнь сосудов головного мозга, болезнь периферических артерий, ревмокардит, врождённый порок сердца, тромбоз глубоких вен и эмболия лёгких. Различают две формы протекания ИБС: острую (инфаркт миокарда) и хроническую (приступы стенокардии).

ССЗ являются основной причиной смертности во всём мире. По данным ВОЗ на 2016 год от данных заболеваний умерло 17,9 миллиона человек, что составило 31% всех случаев смерти в мире. 85% этих смертей произошло в результате сердечного приступа и инсульта. Заболевания сердца и сосудов развиваются постепенно, поэтому люди, страдающие от них или подвергающиеся высокому риску таких заболеваний, нуждаются в раннем выявлении и оказании помощи путем консультирования и, при необходимости, приема лекарственных средств.

В связи с опасностью ССЗ для здоровья и жизни человека, а также их широкой распространённости, было разработано множество средств и методов медицинской диагностики и лечения данной группы заболеваний. Повышение эффективности методов и средств диагностики ССЗ, особенно на ранних стадиях развития заболевания – одна из приоритетных задач современной медицины. Одним из путей достижения данной задачи является разработка экспертных систем (ЭС) и систем поддержки принятия решений (СППР) для сопровождения процесса диагностики заболеваний сердца и сосудов.

Диссертационная работа посвящена разработке математических моделей и программно-алгоритмического обеспечения для ЭС диагностики ССЗ. В процессе работы в данном направлении были изучены методы и средства диагностики ССЗ, применяемые в современной медицине, а также был проведён анализ публикаций, посвящённых разработке и внедрению ЭС и СППР в области диагностики ССЗ. Предложенная ЭС не позволяет сопровождать процесс диагностики всех известных ССЗ, однако оставляет возможность постоянного пополнения и улучшения базы знаний.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель работы состоит в исследовании, разработке и применении математического и программно-алгоритмического обеспечения для диагностики сердечно-сосудистых заболеваний человека.

Исходя из цели, были определены следующие *задачи* исследования:

- анализ существующих методов и средств диагностики сердечно-сосудистых заболеваний;
- анализ существующих экспертных систем в области диагностики сердечно-сосудистых заболеваний человека;
- разработка математического и программно-алгоритмического обеспечения экспертной системы для диагностики сердечно-сосудистых заболеваний человека.

Объектом исследования являются математические модели и программно-алгоритмическое обеспечение процесса диагностики сердечно-сосудистых заболеваний. *Предметом* исследования является процесс диагностики сердечно-сосудистых заболеваний человека.

В настоящее время методы и средства, применяемые для диагностики сердечно-сосудистых заболеваний человека недостаточно совершенны и поэтому требуют своего дальнейшего развития. Недостатками существующих методов и средств является недостаточно высокая степень автоматизации процесса диагностики. Применение экспертных систем и последних достижений в области искусственного интеллекта может в значительной мере автоматизировать и тем самым ускорить процесс диагностики и принятия решений, что в свою очередь может способствовать повышению качества и доступности медицинских услуг. Развитие медицины и повышение качества медицинских услуг – приоритетное направление научных исследований и разработок во всём мире.

Некоторые результаты проведённой работы были апробированы на следующих конференциях:

- 24-я международная научно-техническая конференция, посвящённая 100-летию Нижегородской радиолоборатории, проходившей в Нижнем Новгороде в 2018 году. Материалы конференции опубликованы в сборнике “Информационные системы и технологии”;

- 3-я международная научно-практическая конференция, проходившая в г. Кропивницке 19-20 апреля 2019 года. Материалы конференции опубликованы в сборнике тезисов докладов “Информационная безопасность и компьютерные технологии”.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указано направление исследования, показана актуальность темы диссертационной работы, поставлена цель, а также дана краткая характеристика каждой главы.

Первая глава диссертационной работы посвящена методам и средствам диагностики ССЗ человека. В современной медицине применяется широкий спектр методов и средств диагностики ССЗ на разных стадиях развития заболеваний. Во-первых широко применяются специальные системы оценки вероятности развития ССЗ у того или иного пациента. Во-вторых для диагностики ССЗ используется множество других методов, которые можно разделить на 3 группы: лабораторные, неинвазивные и инвазивные.

Системы оценки вероятности развития ССЗ также называют сердечно-сосудистым риском. Всего в первой главе рассмотрено 3 системы оценки сердечно-сосудистого риска: QRISK, Фрамингемская шкала и шкала SCORE.

QRISK – это алгоритм прогнозирования развития ССЗ, разработанный в Великобритании на основании анализа данных большого количества пациентов и позволяющий оценить вероятность развития ССЗ в ближайшие 10 лет. Фрамингемская шкала разработана на основании обширного эпидемиологического исследования, проводившегося на протяжении более 12 лет в США и также позволяет оценить вероятность смерти от ССЗ в ближайшие 10 лет на основании ряда факторов риска. Шкала SCORE была создана в 2003 году экспертами Европейского общества кардиологов и во многом похожа на Фрамингемскую шкалу. Шкала SCORE активно применяется при оценке сердечно-сосудистого риска в странах Европы, в том числе в Республике Беларусь. Данная шкала представлена на рисунке 1.

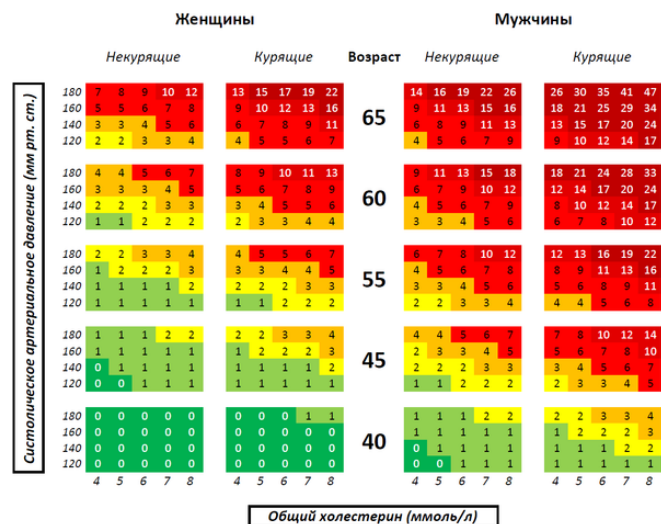


Рисунок 1 – Шкала SCORE для стран с высоким сердечно-сосудистым риском

При оценке сердечно-сосудистого риска на основании шкалы SCORE используются следующие факторы риска: возраст, пол, артериальное давление (АД), курение и общих холестерин. Сердечно-сосудистый риск, полученный с помощью данной шкалы, можно разделить на 4 уровня: низкий ($< 1\%$), умеренный (в пределах $\geq 1\%$ и $< 5\%$), высокий (в пределах $\geq 5\%$ и $< 10\%$), очень высокий ($\geq 10\%$).

Лабораторные методы диагностики ССЗ включают биохимический анализ крови и исследование функций почек. Биохимический анализ крови проводится в целях оценки нарушения липидного и углеводного обменов, исследования свёртываемости крови, оценки биохимических маркеров некроза миокарда, определения уровня мочевины для диагностики артериальной гипертензии (АГ). Исследование функций почек производится с целью обнаружения таких болезней, как АГ и хроническая сердечная недостаточность.

Неинвазивные методы диагностики ССЗ достаточно разнообразны. В первой главе были рассмотрены проба с 6-минутной ходьбой, измерение АД методом Короткова, электрокардиография, холтеровское мониторирование, магнитно-резонансная томография, эхокардиография, ультразвуковое ангиосканирование и доплерография. Данные методы как правило требуют наличия специальных технических средств для проведения диагностики.

Инвазивные методы диагностики включают в себя катетеризацию сердца и коронарную ангиографию (КАГ), а также внутрисосудистое звуковое исследование (ВСУЗИ) коронарных артерий. КАГ – это инвазивное диагностическое исследование, проводимое для оценки состояния коронарных артерий и коронарного кровотока, обнаружения патологических изменений артерий, а также для качественной и количественной характеристики пораженных сегментов артерий. Метод ВСУЗИ позволяет оценить структуру сосудистой стенки, состав и гемодинамическую значимость атеросклеротической бляшки, выявить осложнённые и структурно нестабильные бляшки и на основании результатов выбрать наиболее подходящий метод лечения. Также метод ВСУЗИ позволяет оценить эффективность хирургического вмешательства.

Первая глава завершается рассмотрением профилактических мер и ранней диагностики ССЗ. Основным звеном патогенеза ИБС является атеросклероз коронарных артерий. Важным фактором риска развития атеросклероза является гиперлипидемия (дислипидемия) – аномально повышенный уровень липидов и/или липопротеинов в крови человека. Существует два подхода к нормализации уровня липидов в крови: медикаментозный и немедикаментозный. Первый подход предполагает приём специальных лекарственных препаратов, а второй – изменение образа жизни. Данные подходы могут использоваться как по

отдельности, так и совместно. В данном разделе первой главы рассматривается алгоритм принятия решений по выбору тех или иных профилактических мер для пациента. На рисунке 2 представлена схема выбора тактики ведения пациента с гиперлипидемией без сахарного диабета и клинических проявлений атеросклероза в зависимости от категории риска.

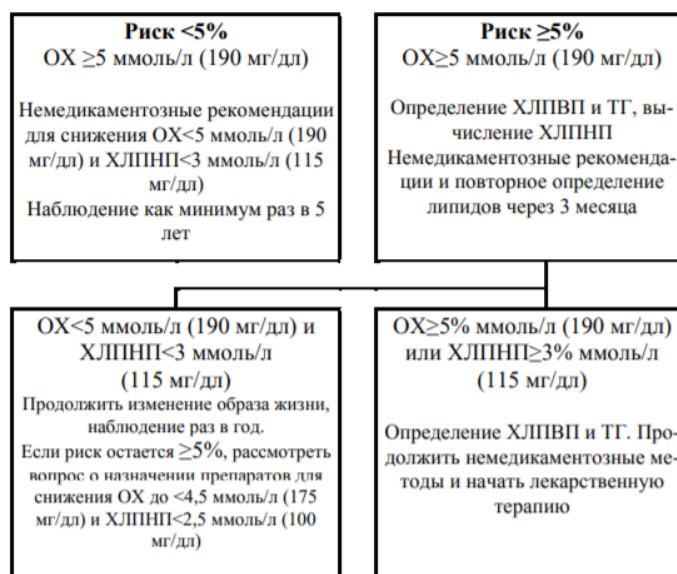


Рисунок 2 – Тактика ведения лиц с гиперлипидемией без сахарного диабета и клинических проявлений атеросклероза в зависимости от категории риска

Во **второй главе** производится анализ публикаций, посвящённых разработке и внедрению интеллектуальных систем в области диагностики ССЗ. Помимо анализа их возможностей, производится анализ их реализации. Рассмотренные системы можно разделить на следующие группы: системы на основе нечёткой логики, системы на основе искусственных нейросетей, нейро-нечёткие системы и системы на основе специально разработанных алгоритмов. Далее даётся краткая характеристика каждого из перечисленных подходов, а также рассматривается их применение на примере существующих, либо разрабатываемых интеллектуальных систем в области кардиологии. Рассматриваются как преимущества данных систем и подходов к их разработке, так и их недостатки.

Рассмотренные системы диагностики ССЗ и поддержки принятия решений ориентированы на такие заболевания, как АГ, ИБС, коронарная недостаточность, сердечная недостаточность, аритмия, стенокардия и т. д. Также ряд работ уделяют внимание оценке сердечно-сосудистого риска и выявлению некоторых факторов риска развития ССЗ.

Можно выделить два наиболее распространённых типа систем: системы на основе нечёткой логики и системы на основе искусственных нейронных сетей.

Характерной чертой современных интеллектуальных систем в области кардиологии является их узкая направленность на диагностику одного заболевания, либо на небольшую группу заболеваний, либо на некоторый класс задач (оценка риска развития ССЗ и оценка факторов риска). На данном этапе развития ЭС и СППР в области кардиологии есть попытки создания комплексных решений, способных диагностировать одновременно целый комплекс заболеваний и проблем сердечно-сосудистой системы, тем не менее данные системы всё ещё не получили широкого распространения и имеют ряд недостатков, связанных с методологией их проектирования и разработки.

По результатам анализа было принято решение, что в проектируемой в рамках исследования экспертной системе основное внимание будет уделяться профилактике развития ССЗ при помощи оценки сердечно-сосудистого риска, анализу образа жизни пациента, а также выявлению скрытых факторов риска, таких как нарушение липидного и белкового обмена. На основании состояния пациента, система должна предлагать подходящие меры профилактики.

Третья глава посвящена разработке собственной ЭС в области диагностики ССЗ на основании рассмотренных в первой главе медицинских средств, методов и алгоритмов диагностики, а также с учётом рассмотренных во второй главе возможностей и недостатков существующих решений в данной предметной области.

В первую очередь были определены требования к системе. ЭС должна оценивать сердечно-сосудистый риск, осуществлять поддержку диагностики гиперлипидемии, давать рекомендации по ведению данного пациента. Также система должна быть легко интегрируемой с программным обеспечением, используемым в учреждениях здравоохранения, медицинскими онлайн-сервисами и т. д. Были определены задачи, которые необходимо решить в рамках разработки математического и программно-алгоритмического обеспечения ЭС.

Первая задача – разработка архитектуры ЭС. Основными компонентами ЭС являются: пользовательский интерфейс, средство объяснения, база знаний, рабочая память, машина логического вывода (МЛВ), рабочий список правил, система приобретения знаний.

Вторая задача – выбор моделей представления знаний. Выбор той или иной модели зависит от сферы применения ЭС, специфики решаемой задачи и т. д. Для разрабатываемой ЭС рассматривается применение следующих распространённых моделей представления знаний: продукционная модель, семантическая модель, фреймы и управляемые образцами модули (УОМ). Для диагностики гиперлипидемии целесообразно использование продукционной модели представления знаний. В продукционной модели знания представляются в виде правил, которые можно описать следующим образом:

$$(i) Q; P; A \Rightarrow B; N; \quad (1)$$

где i – это имя продукционной модели знаний или ее порядковый номер,
 Q – сфера применения правила,
 $A \Rightarrow B$ – ядро продукции, представляющая условную конструкцию ”ЕСЛИ-ТО”,
 P – условие применимости ядра продукции,
 N – постусловие продукции.

Для представления знаний по оценке сердечно-сосудистого риска был выбран матричный способ. Было определено две матрицы: матрица признаков и матрица выводов. Количество столбцов матрицы признаков соответствует количеству входных параметров предметной области. Количество столбцов матрицы выводов соответствует количеству выходных параметров данной области. Обе матрицы имеют одинаковое количество строк. Таким образом для каждого набора признаков может быть определён соответствующий набор выходных значений.

Поскольку представление знаний для первой задачи существенно отличается от представления знаний для второй задачи, то размещать их следует в отдельных модулях, каждый из которых будет обладать следующими свойствами: уникальное наименование модуля, описание модуля, сфера применения, приоритет выполнения, функция оценки применимости модуля в данной ситуации, функция активации (будет вызвана, если модуль применим), ссылка на следующий модуль (после срабатывания функции активации), набор оригинальных данных модуля.

Третья задача – выбор алгоритмов логического вывода. Был разработан и описан алгоритм логического вывода для матричного представления знаний. Алгоритм осуществляет поиск наиболее близкого набора признаков и извлекает соответствующий ему набор выходных значений. Для диагностики гиперлипидемии был выбран алгоритм Мамдани, включающий следующие этапы: формирование базы правил, фаззификация, агрегирование подусловий, активизация подзаключений, аккумуляирование заключений и дефаззификация.

Четвёртая задача – разработка модели ЭС. Модель модуля оценки сердечно-сосудистого риска включает следующие входные параметры: возраст, пол, курение, АД, общий ХС. Для параметров “возраст”, “АД” и “общий ХС” определены функции приближения. Функции приближения позволяют определить, к какому диапазону относится значение того или иной параметр и осуществить замену первоначального значения на соответствующее этому диапазону число. Параметры “пол” и “курение” – строковые, поэтому их будут

преобразованы в числовые эквиваленты. количество строк матрицы признаков, достаточное для покрытия всех возможных комбинаций, которые образуют данные признаки, равно 400, что соответствует количеству ячеек в шкале SCORE.

Модуль диагностики гиперлипидемии представляет собой систему нечёткого вывода. Для данной системы был определён набор из 4-х входных лингвистических переменных: “Risk” (сердечно-сосудистый риск), “Cholesterol” (общий ХС), “LDL” (ХС ЛПНП) и “RepRecept” (является ли приём повторным). Также было определено 3 лингвистические переменные: “RepRecept” (интервал повторного приёма), “NotMedical” (немедицинские рекомендации), “Medical” (медикаментозное лечение). Для работы с данными лингвистическими переменными был разработан набор из 9-и продукционных правил. Для разработки и тестирования модели диагностики гиперлипидемии была применена утилита “Fuzzy Logic” среды MATLAB R2017b.

Общая структура базы знаний включает 2 основных модуля, описанных выше, а также 3 вспомогательных. Общая схема взаимодействия модулей базы знаний между собой представлена на рисунке 3.

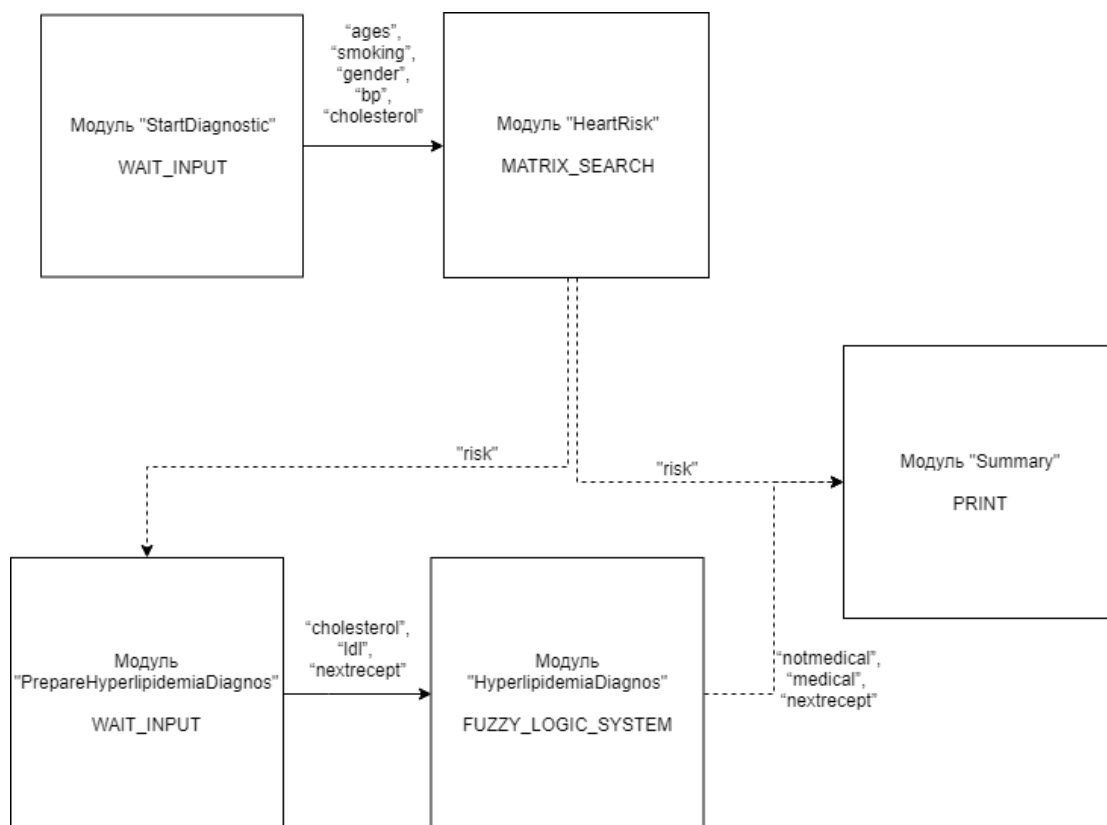


Рисунок 3 – Общая схема связи модулей базы знаний

ЗАКЛЮЧЕНИЕ

Основные результаты диссертации

1. Изучены методы медицинской диагностики ССЗ. Данные методы можно разделить на 3 группы: лабораторные, неинвазивные и инвазивные. Также были изучены методы ранней диагностики ССЗ и системы оценки риска развития ССЗ и выявления факторов риска, влияющих на развитие заболеваний сердца и сосудов. Было изучено несколько систем оценки сердечно-сосудистого риска: QRISK, Фрамингемская шкала и шкала SCORE. Шкала SCORE была выбрана в качестве системы оценки сердечно-сосудистого риска для дальнейшего использования в исследованиях, поскольку данная система рекомендована к использованию в странах с высоким уровнем риска, к которым относится Республика Беларусь. Также было уделено внимание диагностике гиперлипидемии как к одному из важнейших факторов развития атеросклероза и ИБС.

2. Был произведён обзор существующих публикаций в области разработки интеллектуальных систем диагностики ССЗ. Данный обзор показал, что современные интеллектуальные системы в области кардиологии уделяют внимание диагностике таких заболеваний, как АГ, ИБС, коронарная недостаточность, сердечная недостаточность, аритмия, стенокардия и т. д. Также ряд работ уделяют внимание оценке сердечно-сосудистого риска и выявлению некоторых факторов риска развития ССЗ. По результатам обзора было установлено, что наиболее популярными типами диагностических систем являются системы на основе нечёткой логики и системы на основе искусственных нейронных сетей. Также было выявлено, что в рассмотренных системах, в особенности в системах оценки сердечно-сосудистого риска и факторов риска развития ССЗ, уделяется больше внимания анализу образа жизни пациента, однако мало внимания уделено анализу скрытых факторов, таких как нарушение белкового и липидного обмена, ведущих к развитию атеросклероза, который в свою очередь ведёт к развитию ИБС. На этом основании было принято разработать такую ЭС, которая выявляла бы также гиперлипидемию у пациентов и помогала бы сформировать рекомендации по предотвращению развития атеросклероза и ИБС.

3. Были определены требования к будущей ЭС. Система решает две задачи: оценка сердечно-сосудистого риска по шкале SCORE и диагностики гиперлипидемии.

4. Была определена архитектура разрабатываемой ЭС. Основными компонентами системы являются: база знаний, рабочая память, МЛВ, рабочий

список правил, система приобретения знаний, пользовательский интерфейс, подсистема объяснений.

5. Были рассмотрены несколько моделей представления знаний: продукционная, семантическая, фреймовая и УОМ. Для подзадачи диагностики гиперлипидемии была выбрана продукционная модель представления знаний. Для оценки сердечно-сосудистого риска была разработана модель представления знаний на основе матриц, где комбинации признаков хранятся в матрице признаков и для каждой строки в данной матрице есть строка в матрице выводов.

6. Для ЭС было разработано представление знаний в виде модулей. Каждый модуль представляет собой независимый блок, который реализует свой алгоритм логического вывода для какой-либо модели представления знаний. Модули могут объединяться в сети и передавать управление друг другу, а могут выбираться МЛВ на основании состояния рабочей памяти. Данная модель сочетает в себе особенности представления знаний как в семантической модели, так и в модели на основе УОМ. Благодаря использованию данной модели можно сочетать большое количество различных способов представления знаний в рамках одной системы и предметной области.

7. Был разработан алгоритм логического вывода для матричной модели представления знаний.

8. Был разработан алгоритм работы МЛВ разрабатываемой ЭС.

9. Была разработана модель для модуля оценки сердечно-сосудистого риска. Модель определяет параметры, их связь со столбцами матриц признаков и выводов, а также функции преобразований и приближений. Также модель определяет размеры матриц и их источники данных для заполнения матриц.

10. Была разработана модель системы нечёткого вывода для модуля диагностики гиперлипидемии. Данная модель включает в себя 4 входные лингвистические переменные и 3 выходные, а также набор из 9 продукционных правил. Данная модель была протестирована при помощи математического пакета MATLAB. Результаты тестов представлены в данной диссертации.

11. Были определены 2 основных и 3 вспомогательных модуля базы знаний разрабатываемой ЭС. Для каждого модуля было разработано описание его структуры и принципов функционирования. Также была разработана схема взаимодействия между модулями.

Рекомендации по практическому использованию диссертации

1. Полученные результаты исследований детально описывают организацию ЭС для диагностики ССЗ. На основании этих данных можно разработать программные компоненты и интерфейсы для МЛВ, базы знаний,

рабочей памяти и других компонентов системы. Основные компоненты ЭС могут быть реализованы как монолитное ядро, а функции проверки применимости и функции активации желательно размещать в динамически подключаемых библиотеках, чтобы облегчить расширяемость системы. Этого можно достичь, если все модули ЭС будут поддерживать единый интерфейс взаимодействия.

2. Для описания модулей рекомендуется использовать текстовые конфигурационные файлы, либо XML-файлы, чтобы их легко могли редактировать обычные пользователи, тем самым имея возможность расширять базу знаний и решать более сложные комплексы задач.

3. Для повышения доступности системы и облегчения интеграции с различными медицинскими системами управления рекомендуется разработать веб-сервис со специальным API (application programming interface) для обслуживания внешних запросов. Веб сервис может быть размещён либо в локальной сети медицинского учреждения, либо в сети Интернет, что позволит одновременно большому количеству пользователей взаимодействовать с ЭС.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Скудняков, Ю. А. Проектирование экспертной системы диагностики сердечно-сосудистых заболеваний / Ю. А. Скудняков, А. С. Барсук // Информационные системы и технологии (ИСТ-2018) : материалы 24-й международ. науч.-технич. конференц., посвящ. 100-летию Нижегород. радиолоборатории, Нижний Новгород, 2018 / Нижегородский государственный технический университет им. Р. Е. Алексеева. – Нижний Новгород, 2018. – С. 826 – 831.

2. Барсук, А. С. Экспертная система диагностики сердечно-сосудистых заболеваний / А. С. Барсук ; науч. рук. Ю. А. Скудняков // Информационная безопасность и компьютерные технологии : материалы 3-й международ. науч.-практ. конф., Кропивницкий, 19-20 апреля 2018 / Центральноукраинский национальный технический университет. – Кропивницкий, 2018. – С. 288-290.