Министерство образования и науки Российской Федерации

федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт Институт кибернетики
Направление подготовки Прикладная информатика
Кафедра Оптимизации систем управлении

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Исследование и апробация ключевых метрик для прецедентного анализа, применяемых на ограниченном наборе формализованных данных

УДК _____

Студент

Группа	ФИО	Подпись	Дата
8КМ41	Алексеева Алина Александровна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент кафедры ОСУ	Фофанов О.Б.	к.т.н		

КОНСУЛЬТАНТЫ:

По разделу «Основная часть»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент каф. ОСУ	Тараник М.А..			

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. Менеджмента	Антонова И.С.	к.э.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ЭБЖ	Пустовойтова М.И.	к.х.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
ОСУ	Иванов М. А.	к.т.н		

Томск – 2016 г.

Реферат

Выпускная квалификационная работа содержит 76 страниц, 23 рисунка, 26 таблиц, 47 формул.

Ключевые слова: прецедентный подход, база знаний, метрика, нечеткая логика

Объект исследования - ряд медицинский показателей определяющий патологии сердечно-сосудистой системы

Цель исследования - исследование и апробация существующих методов вывода формализованных данных, а также разработка информационной системы повышения эффективности алгоритмов и точности из результатов.

В процессе выполнения работы применялись методы логического ввода – Евклидова метрика, метрика Махаланобиса, Метрика Хемминга

Результатом выполнения выпускной квалификационной работы является разработанная гибридная модель системы вывода по прецедентам.

Содержание

РЕФЕРАТ **2**

ГЛАВА 1. ИССЛЕДОВАНИЕ МЕТОДА ЛОГИЧЕСКОГО ВЫВОДА, ОСНОВАННОГО НА ПРЕЦЕДЕНТАХ **6**

1.1. Вывод на основе прецедентного подхода 7

1.2. Основные методы извлечения прецедентов 10

1.3. Отбор ключевых метрик для прецедентного анализа 15

ГЛАВА 2. АПРОБАЦИЯ КЛЮЧЕВЫХ МЕТРИК **17**

2.1. Проектирование приложения для сбора данных 17

2.1.1. Исходные данные 17

2.1.2. Логическое представление БД 18

2.1.3. Построение промежуточной ИС 21

ГЛАВА 3. ПРОЕКТИРОВАНИЕ ГИБРИДНОЙ МОДЕЛИ **29**

3.1. Нечеткая логика 29

3.2. Описание алгоритма обработки данных гибридного метода 39

3.3. Разработка функциональной модели системы 43

3.4. Проектирование гибридной модели 46

ГЛАВА 4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ **48**

4.1. Предпроектный анализ 48

4.2. Инициация проекта 54

4.3. Планирование управления научно-техническим проектом 56

4.5. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования 70

ЗАКЛЮЧЕНИЕ **75**

СПИСОК ЛИТЕРАТУРЫ **82**

Введение

В настоящее время повышение качества оказываемой медицинской помощи является одной из наиболее сложных и значимых задач. С ростом заболеваемости возникает ряд проблем, в частности рационального использования врачебного времени. Так как 50% рабочего времени врача уходит на поиск нужной информации, возросла необходимость в использование диагностических систем. Использование данных систем позволит почти в 4 раза сократить время поиска нужной информации, на 25% сократить время для определения диагноза и на 10-20% увеличить поток принятых врачом пациентов [1]. Таким образом, очевидно, что использование современных медицинских информационных систем позволит повысить доступность и качество медицинской помощи. Различают несколько направления в развитии логического вывода знаний при использовании медицинских систем поддержки принятия решений:

- развитие систем логического вывода, основанного на правилах;
- развитие систем логического вывода, основанного на прецедентах.

В основу системы логического вывода, основанного на правилах, закладывалась совокупность правил вида "*если...то...*", согласно которым на основании входных данных генерировалось то или иное заключение по интересующей проблеме. Идея вывода по правилам подразумевает наличие четко формализованной задачи, для которой существуют научные методы, доказавшие свою применимость и позволяющие получить решение, не требующее доказательств.

В случае если задача слабо формализована, и найти подходящее решение становится достаточно трудоемкой задачей, применяются системы логического вывода, основанного на прецедентах.[4] Широкое применение данная система получила в области медицинской диагностики [10]. В основе технологии вывода, основанного на прецедентах, лежит накопленный опыт прошлых ситуаций.

Актуальность проблемы обусловлена многочисленностью слабо формализованных задач, и практической потребностью найти в «сырых» данных хотя бы одно достоверно верное решение

Целью работы является исследование и апробация существующих методов вывода формализованных данных, а также разработка информационной системы повышения эффективности алгоритмов и точности из результатов.

Для достижения поставленной цели, необходимо решить следующие задачи:

1. Исследование предметной области;
2. Ознакомление с основными типами метрик, которые могут быть использованы в задачах поиска близких прецедентов;
3. Проведение сравнительного анализа путем апробации выделенных метрик;
4. Применение аппарата нечёткой логики;
5. Построение гибридной системы.

Глава 1. Исследование метода логического вывода, основанного на прецедентах

Необходимость в ускорении принятия решения, в обработке и упорядочивании больших объемов разнородной информации, с которой приходится иметь дело в процессе классификации и распознавания того или иного заболевания привело к развитию компьютерной интеллектуализации процедуры диагностирования заболеваний [3].

Компьютерная диагностика тесно связана с накоплением и обработкой разнородной информации. Следовательно, от качества методов и способов работы с данной информацией зависит надежность диагностики, а в дальнейшем эффективность лечения. Очень часто при определении диагноза возникает важная проблема - недостаток опыта врача-диагноста. В результате чего, диагноз ставится по отдельным признакам, без учета других, по мнению врача, менее значимых, а так же без учета общего фона, на котором развивается заболевание и особенностей организма пациента.

Для повышения качества диагностики есть потребность в накоплении и статистическом анализе данных, получаемых путем наблюдения большого количества пациентов. Формирование базы знаний позволит учесть все показатели, выявленные у пациента и наиболее точно поставить диагноз.

Существуют различные инструменты распознавания образов и принятия решений при выработке клинических диагнозов [11]:

- искусственные нейронные сети
- генетические алгоритмы,
- метод группового учета аргументов,
- логические методы,

В диссертационной работе ограничимся рассмотрением методов распознавания образов и принятия решений, в основу которых положены средства логического вывода.

В качестве логического метода анализа данных рассмотрим подход вывода по прецедентам.

1.1. Вывод на основе прецедентного подхода

В области медицины существует огромное количество слабо формализованных задач, в связи с этим очень трудно получить достоверно верное решение – диагноз. Основной трудностью в определении диагноза является существование огромного количества возможных заболеваний в каждом конкретном случае. Прецедентный подход стал успешным методом для разработки медицинских систем диагностики. Данный метод является решением существующих проблем в части постановки диагноза.

Прецедентный подход – это процесс (методология) решения новой задачи (проблемы) путем повторного использования и адаптации (при необходимости) решений, которые были ранее получены при решении подобных задач.

Основным понятием является прецедент – это структурированное представление накопленного опыта в виде данных и знаний, обеспечивающее его последующую автоматизированную обработку при помощи специализированных программных систем[5]. Как правило, прецедент состоит из описания проблемной ситуации (результатов лабораторных и инструментальных методов исследования) и совокупности действий, предпринимаемых для решения данной проблемной ситуации (постановка диагноза).

Основные характеристики прецедентов:

• прецедент представляет особое знание, привязанное к контексту, что позволяет использовать знания на прикладном уровне;

• прецеденты могут принимать различную форму (вид): охватывая разные по продолжительности промежутки времени; связывая решения с описаниями проблем; результаты с ситуациями и т.д.;

• прецедент фиксирует только тот опыт, который может быть полезным, фиксируемые прецеденты потенциально могут помочь специалисту (ЛПР) достичь цели, облегчить ее формулирование в будущем или предупредить его о возможной неудаче или непредвиденной проблеме.

Схема прецедентного подхода имеет следующий вид рис.1.

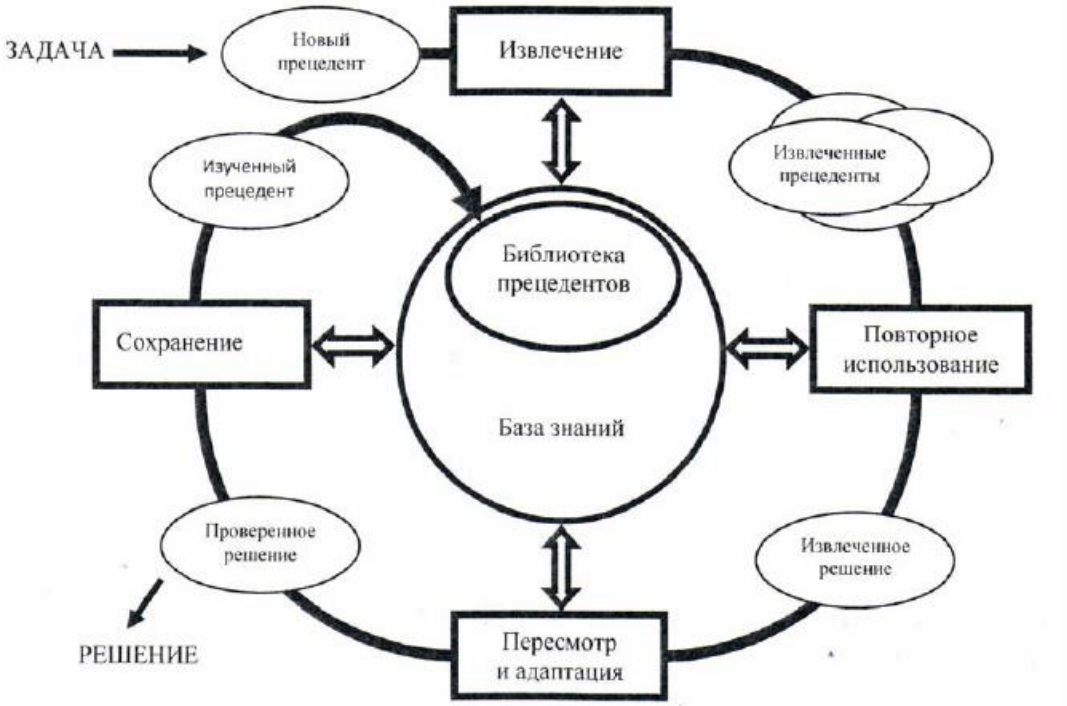


Рис. 1 Схема CBR-цикла

Решение проблемы начинается с представления ее пользователем в виде нового прецедента. На стадии восстановления (извлечения) из базы прецедентов выбирается одни или несколько прецедентов, которые считаются полезными для решения новой проблемы. Прецеденты отбираются, основываясь на подобии (близости) новой проблемы к проблемам, описание которых содержат прецеденты из базы прецедентов. Предполагается, что близкие проблемы имеют подобные решения. Точная формализация подобия очень сильно влияет на качество работы системы Возможность повторного использования в ряде случаев достигается путем адаптации (преобразования) восстановленных решений. На стадии сохранения происходит обучение CBR системы, суть которого заключается в добавлении пересмотренного прецедента в базу прецедентов. Таким образом,

опыт решения новой проблемы становится доступным для повторного использования в будущем[5].

Основной недостаток данного подхода состоит в том, что при росте базы прецедентов обычно наблюдается уменьшение эффективности извлечения прецедентов, так как приходиться анализировать все большее количество прецедентов. В этом случае для более эффективного извлечения предлагается применять индексирование прецедентов.

Для уточнения понятия близости обычно используются понятие метрики.

Метрика - это правило, по которому в данном пространстве определяется расстояние между двумя точками[5].

Перечислим основные типы метрик, которые могут быть использованы в задачах поиска близких прецедентов[5]:

1. Евклидова метрика;
2. Мера сходства Хемминга;
3. Вероятностная мера сходства;
4. Мера сходства Роджерса-Танимото;
5. Манхэттенская метрика;
6. Расстояние Чебышева;
7. Метрики Махаланобиса;
8. Метрика Брея-Кертиса;
9. Метрика Чекановского;
10. Метрика Жаккара;
11. Обобщенное расстояние Евклида-Махаланобиса.

Далее рассмотрим каждую метрику в подробности.

1.2. Основные методы извлечения прецедентов

Евклидова метрика

Евклидова метрика - это геометрическое расстояние между двумя точками в многомерном пространстве. Данная величина применяется в методе "ближайшего соседа".

Расстояние Евклида показывает, как далеко друг от друга находятся два вектора, тем самым характеризуя их причастность к тому или иному классу, определенному вектором средних.

Расчет метрики Евклида (De) проводится по классической формуле:

$$De = \sqrt{\sum_{i=1}^{k}(x_i - \mu_i)^2} \quad (1)$$

где k – мощность инварианта (количество характеристик);

x, μ – сравниваемые векторы.

Евклидово расстояние (и его квадрат) вычисляется по исходным, а не по стандартизованным данным.

Мера сходства Хемминга

Расстояние Хэмминга является также расстоянием редактирования (определенной для строк одинаковой длины) с единственной допустимой операцией – заменой.

В более общем случае расстояние Хэмминга применяется для строк одинаковой длины любых k-ичных алфавитов и служит метрикой различия (функцией, определяющей расстояние в метрическом пространстве) объектов одинаковой размерности.

$$\mu_{ij}^{N} = \frac{n_{ik}}{N} \quad (2)$$

Где n_{ik} - число совпадающих признаков у образцов X_i и X_k;

Расстояние Хэмминга между двумя словами подсчитывается с помощью прохода по строкам и подсчёта количества несовпадающих символов на соответствующих позициях.

Результатом работы сети является нахождение образа с наименьшим расстоянием.

Вероятностная мера сходства

$$P_j = exp\left(\frac{-\sum_{i=1}^{N}\left(w_{ij} - x_i\right)^2}{\sigma_j^2}\right) \quad (3)$$

где j-номер эталона, X_i (i=1,2,…,N) - элемент неизвестного входного образца, w_{ij}– значение весового коэффициента, соответствующее математическому ожиданию i-го элемента (признака) j-го эталона. Величина среднеквадратичного отклонения σ_j- находится в результате экспериментов для каждого эталона;

Мера сходства Роджерса-Танимото

$$\mu_{ik}^{R-T} = n_{ik}(n_i + n_k - n_{ik}) \quad (4)$$

где:

n_{ik}- число совпадающих единичных признаков у образцов X_i и X_k

n_i, n_k- общее число единичных признаков у образцов X_i и X_k соответственно;

Манхэттенская метрика

Расстояние городских кварталов - метрика, введённая Германом Минковским. Согласно этой метрике, расстояние между двумя точками равно сумме модулей разностей их координат.

Расстояние между двумя точками, измеренное вдоль осей под прямым углом.

$$d_{ik} = \sum_{j=1}^{N}\left|x_{ij} - x_{kj}\right| \quad (5)$$

Данная метрика уменьшает влияние отдельных больших разностей между одноименными координатами точек, так как при вычислении расстояния эти разности не возводятся в квадрат (в отличие от евклидовой метрики).

Евклидова и манхеттенская метрики приводят к близким результатам;

Расстояние Чебышева

Принимает значение наибольшего модуля разности между значениями соответствующих свойств (признаков) объектов

$$d_{ik} = arg \max_{1 \le j \le N} |x_{ij} - x_{kj}| \quad (6$$

где N - количество переменных (признаков) i и j номера объектов

Частично используется в нечетких нейронных сетях в виде минимаксных критериев. Недостаток - кластеры, полученные с помощью расстояния Чебышева, «склеиваются» друг с другом.

Данная метрика может оказаться полезной, когда нужно определить два объекта как «различные», если они различаются по какой-либо одной координате (каким-либо одним измерением).

Метрика Махаланобиса

Расстояние Махаланобиса – это мера расстояния между векторами случайных величин, обобщающая понятие расстояния Евклида. Оно отличается от него тем, что учитывает корреляции между переменными и инвариантно к масштабу

Формально расстояние Махаланобиса (Dm) от многомерного вектора

x = (x1, x2, x3, …, xN) T до множества с вектором среднего μ = (μ1, μ2, μ3, …, μN) T и матрицей ковариации S определяется по нижеприведенной формуле:

$$D_m = \sqrt{(x-\mu)^T S^{-1} (x-\mu)} \quad (7)$$

где x – вектор характеристик входного текста; μ – вектор средних для некоторого класса текстов; S – объединенная ковариационная матрица.

Метрика Махаланобиса не применима, если выборочная дисперсия хотя бы одного из параметров равна нулю.

Метрика Брея-Кертиса

Метрика на основе коэффициента различия Брэя-Кёртиса применяется как оценка сходства образцов в анализе многомерных данных. Благодаря своей робастности и надежной оценке различия между векторными данными

метрика получила широкое распространение в задачах классификации в экологии. Данная метрика рассматривает пространство в виде таблицы

$$d_{ik} = \frac{\sum_{j=1}^{N} |x_{ij} - x_{kj}|}{\sum_{j=1}^{N} |x_{ij} + x_{kj}|} \quad (8)$$

В этом случае значения заключены между 0 и 1. Обычно перед использованием этой метрики данные стандартизуют. Данные после стандартизации должны быть неотрицательными. Значение 1 означает полное совпадение двух данных записей в N-мерном пространстве. Результат считается неопределенным, когда переменных двух объектов имеют значение 0.

Метрика Чекановского

$$d_{ik} = \frac{b + c}{2a + b + c} \quad (9)$$

Коэффициенты a, b, c, и d берутся из таблицы (матрицы) ассоциативности, построенной для двух объектов i и k, в которой 1 указывает на наличие признака у объекта, 0 – на его отсутствие. Проще всего рассмотреть эти коэффициенты, обратившись к таблице (матрице) ассоциативности размера 2 x 2:

	1	0
1	a	b
0	c	d

Метрика Жаккара

Мера Жаккара - бинарная мера сходства. Коэффициент сходства (также мера сходства, индекс сходства) - безразмерный показатель, применяемый в биологии для количественного определения степени сходства биологических объектов. Также известен под названиями «мера ассоциации», «мера подобия» и др. Показатель подобия Жаккара для таблиц сопряженности 2x2 вычисляется по следующей формуле:

$$d_{ik} = \frac{a}{a + b + c} \quad (10)$$

где a — общие число общих видов,

b — число видов в 1-м списке,

c — число видов во 2-м списке.

Как и в случае метрики Чекановского, коэффициенты a, b, c, и d берутся из таблицы ассоциативности.

Коэффициент Жаккара может иметь значение от 1 до 100% (или от 0 до 1)

Обобщенное расстояние Евклида-Махаланобиса

Для определения расстояния от точки, координаты которой представляют собой параметры наблюдаемого объекта, до класса n сходных объектов обычно пользуются метриками Евклида и Махаланобиса. Каждая из этих метрик имеет свои преимущества и недостатки.

Метрика Евклида, используемая для определения расстояния между точками, x1, x2

$$R_E^2(x_1, x_2) = (x_1 - x_2)^T\ (x_1 - x_2) \tag{11}$$

Метрика Махаланобиса неприменима, если выборочная дисперсия хотя бы одного из параметров равна нулю:

$$\lim_{D_i \to 0} R_M^2\ (x, X) = \infty \tag{12}$$

Метрика Махаланобиса совпадает с Евклидовой в случае, если класс представляет собой вектор реализаций нормированных (дисперсии Di=1, i=1,…, n) независимых (ковариации Kij=0, i,j=1,…, n, i≠j) случайных величин. Если дисперсии больше 1, то расстояние Махаланобиса меньше Евклидова, если меньше, то $R_M^2 > R_E^2$. Проверка аксиом расстояния затруднена тем, что метрика используется для определения расстояния между разнородными объектами. Расстояние между двумя точками почти всегда бесконечно велико.

1.3. Отбор ключевых метрик для прецедентного анализа

Выбор соответствующей метрики достаточно трудоемкая задача, от успешного решения которой непосредственно зависит результативность поиска и извлечения прецедентов. В каждом конкретном случае этот выбор производится по-разному, в зависимости от целей пользователя системы, физической и статистической природы используемой информации при управлении сложным объектом и других ограничений и факторов, влияющих на процесс поиска решения.

Неформально прецедент, хранящийся в базе данных, имеет вид: <ОПИСАНИЕ СОСОТОЯНИЯ> <РЕШЕНИЕ>. Так, в диагностике медицинских показателей «описание состояния» включает результаты лабораторных и инструментальных исследований при выявлении патологии сердечно-сосудистой системы, решением данной системы является постановка диагноза. Ситуация (состояние), для которой ищется решение, называется обычно проблемной ситуацией, или проблемой, и её неформальное описание имеет вид <ПРОБЛЕМА> <РЕШЕНИЕ>

Учитывая вышесказанное основными фазами метода диагностики пациента на основе прецедентов будет, являются:

1. идентификация текущего состояния пациента;
2. поиск подходящего прецедента;
3. использование найденного прецедента для постановки диагноза и как следствие выработки последовательности лечебно-диагностических мероприятий. Данная задача называется также задачей адаптации (имеющегося решения к новой проблеме).

Для поиска подходящего прецедента были отобраны наиболее распространенные и простые в реализации метрики:

- Евклидово расстояние
- Метрика Махалонобиса
- Расстояние Хеминга

Для проведения анализа и поиска наиболее популярных метрик был выбран электронный ресурс sciencedirect. Отбор осуществлялся в процентном соотношении, за основу были взяты статьи, в которых описывались применение и эффективность вывода при использовании данных метрик.

Глава 2. АПРОБАЦИЯ КЛЮЧЕВЫХ МЕТРИК

2.1. Проектирование приложения для сбора данных

2.1.1. Исходные данные

В области медицины существует огромное количество слабо формализованных задач, в связи с этим очень трудно получить достоверно верное решение - диагноз, так как существует много возможных заболеваний в каждом конкретном случае. Прецедентный подход стал успешным методом для разработки медицинских систем диагностики. Данный метод является решением существующих проблем в части постановки диагноза.

В качестве объекта исследования был выбран ряд медицинский показателей определяющий патологии сердечно-сосудистой системы.

Основой данных для предметной области служат лабораторные и инструментальные исследования, которые проводят при сердечно - сосудистых заболеваний, таких как:

- Ишемическая болезнь сердца;
- Кардиосклероз;
- Нарушения ритма сердечных сокращений — аритмия;
- Сердечная недостаточность;
- Эндокардит;
- Миокардит;
- Перикардит;
- Приобретённый порок сердца;
- Врождённый порок сердца;
- Атеросклероз;
- Фибрилляция;
- Экстрасистолия.

Информация включает в себя пол, возраст пациента, а так же ряд показателей лабораторных и инструментальных методов исследований:

- исследование уровня креатинкиназы в крови;
- экспресс-исследование уровня тропонина в крови;
- коагулограмма (ориентировочное исследование системы гемостаза;
- общий (клинический) анализ крови развернутый;
- анализ крови биохимический общетерапевтический;
- анализ крови по оценке нарушений липидного обмена биохимический;
- эхокардиография;
- регистрация электрокардиограммы.

Для обработки данных показателей и оперативного определения заболевания пациента, было решено применить прецедентный подход.

2.1.2. **Логическое представление БД**

Хранение информации в современном обществе сложно представить без баз данных и СУБД, позволяющих вести работу с ними.

База данных — организованная в соответствии с определёнными правилами и поддерживаемая в памяти компьютера совокупность данных, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователей.

Логическое проектирование - преобразование требований к данным в структуры данных, на выходе которого получаем СУБД-ориентированную структуру базы данных и спецификации прикладных программ.

Модели данных помогают визуализировать структуру данных, обеспечивая эффективный процесс организации, управления и администрирования таких аспектов деятельности предприятия, как уровень сложности данных, технологий баз данных и среды развертывания.

В данной дипломной работе процесс логического представления выполнен с помощью AllFusion ERwin Data Modeler, CASE-средства позволяющего проектировать и документировать базы данных. Логическая структура БД состоит из 8 сущностей и представлена на рис.2 .

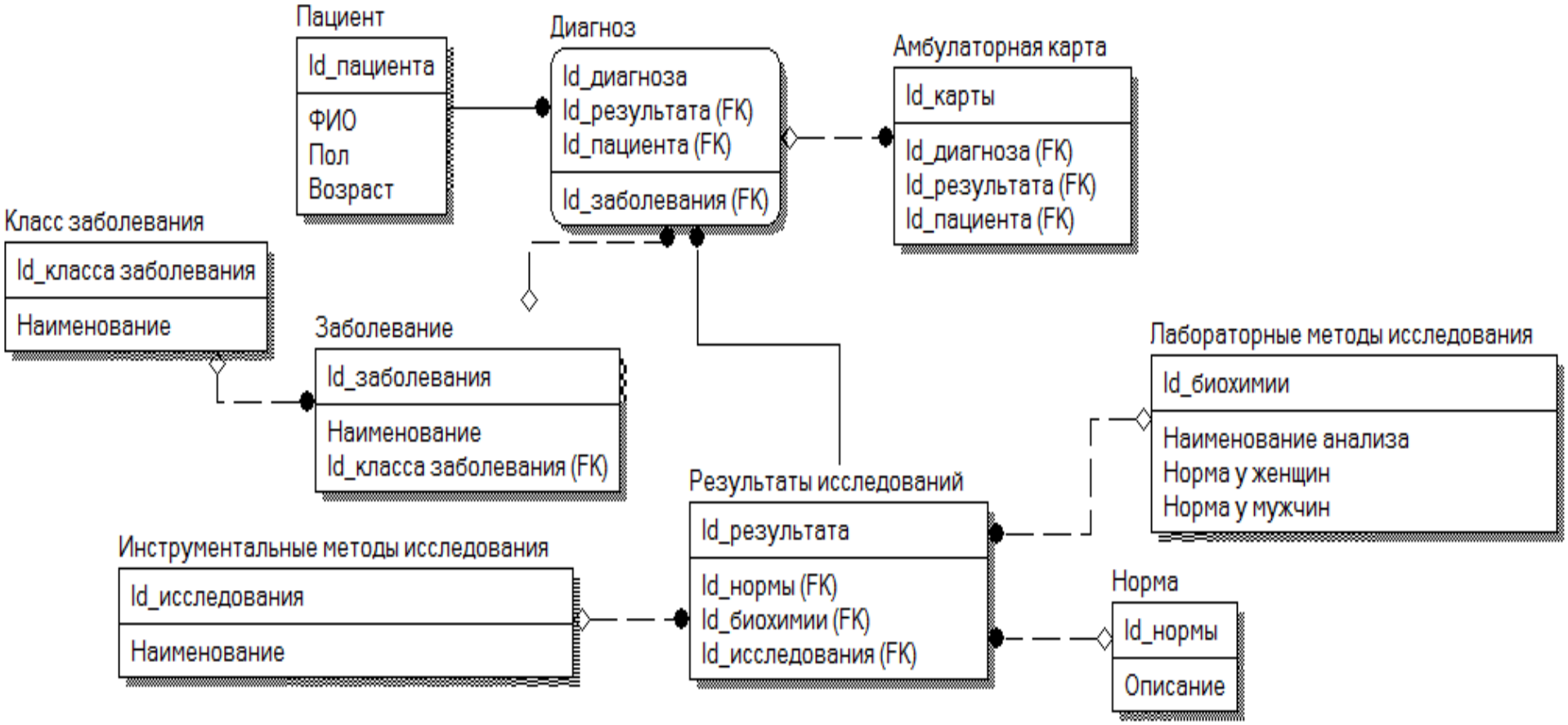


Рис. 2 Логическая структура базы данных ИС

2.1.3. Построение промежуточной ИС

Повышения качества диагностики можно достичь с помощью накопления данных, получаемых путем наблюдения большого количества пациентов, доскональном отслеживании их состояния на протяжении длительного времени.

Информацию о ранее накопленных данных пользователь системы может получить на вкладке пациент, рис. 3. Для удобства добавлена поисковая система.

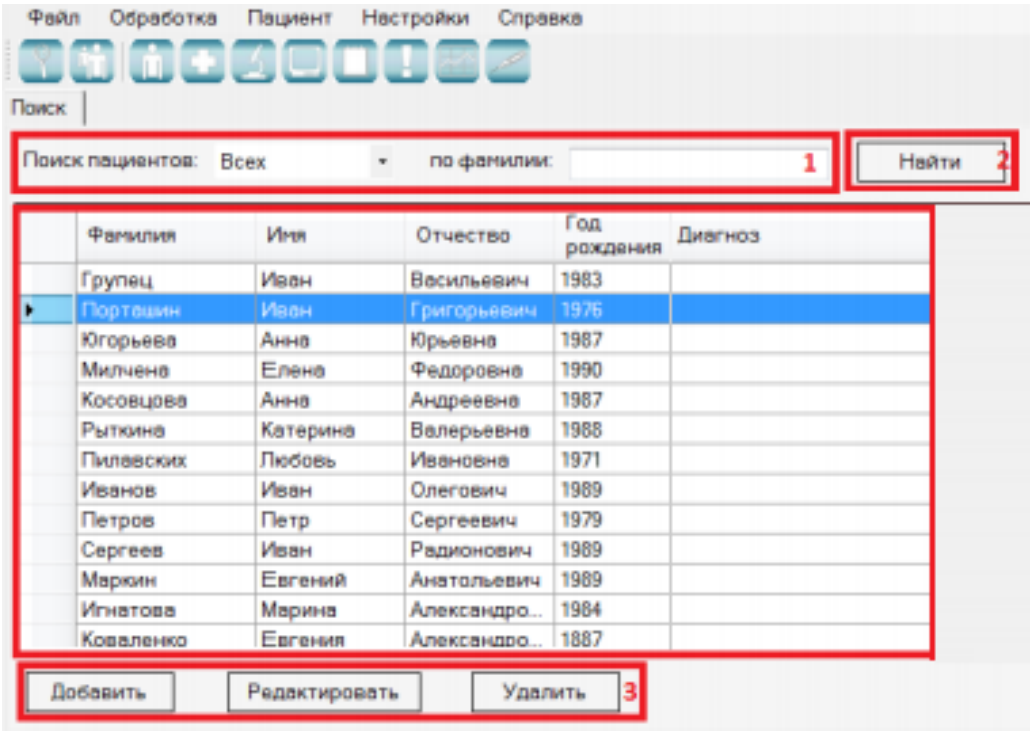


Рис. 3 Интерфейс вкладки Пациент

Для открытия медицинской карточки пациента, необходимо выбрать ее двойным щелчком мыши, рис. 4. В амбулаторной карточке хранятся все проведенные исследования, а так же информация о поставленном диагнозе. Данная форма позволяет добавлять новые истории болезни пациента.

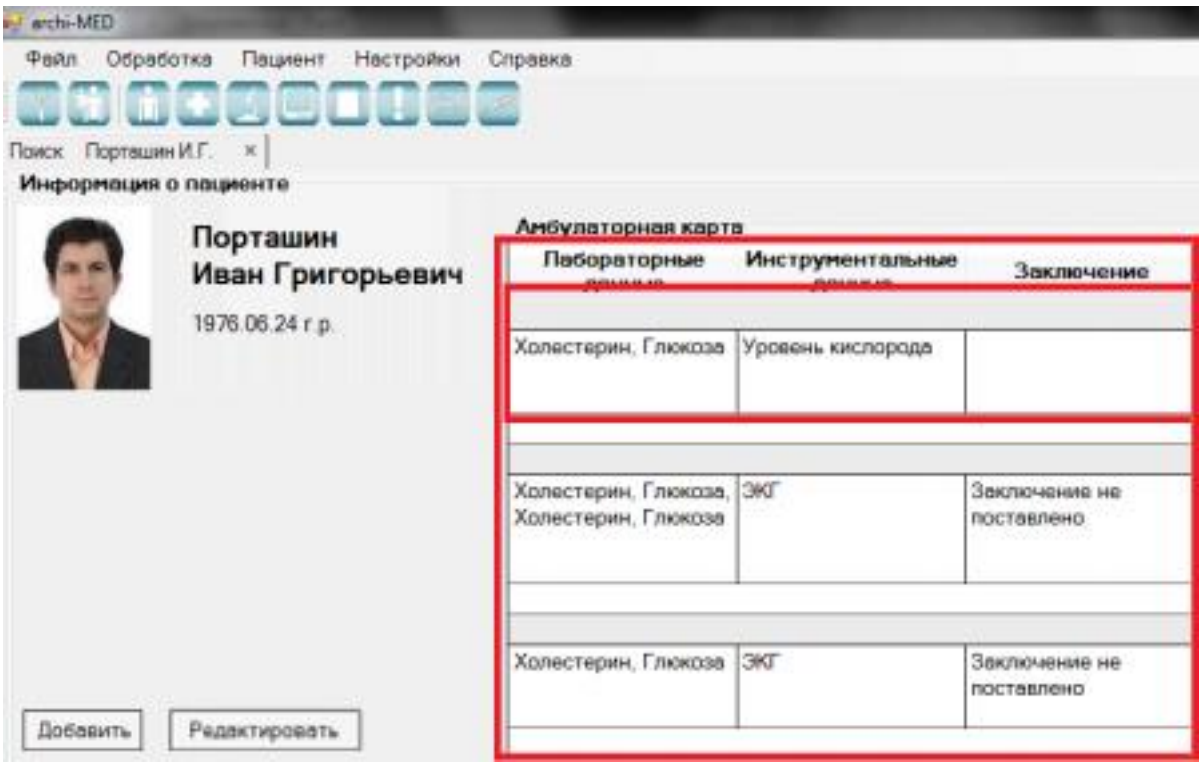


Рис. 4 Форма «Амбулаторная карта пациента»

На вкладке «обработка» врач вводит результаты исследования пациента. Для корректного заполнения результатов, представлен справочник показателей рис. 5, рис. 6

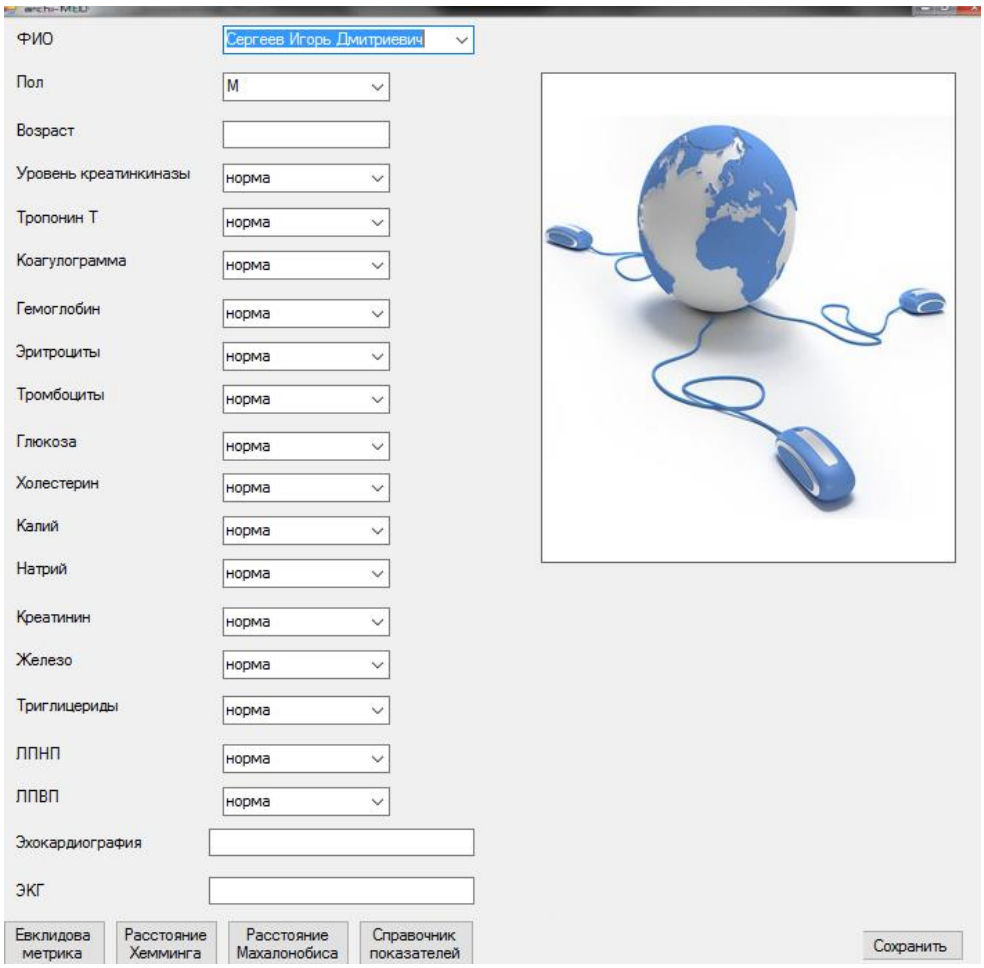
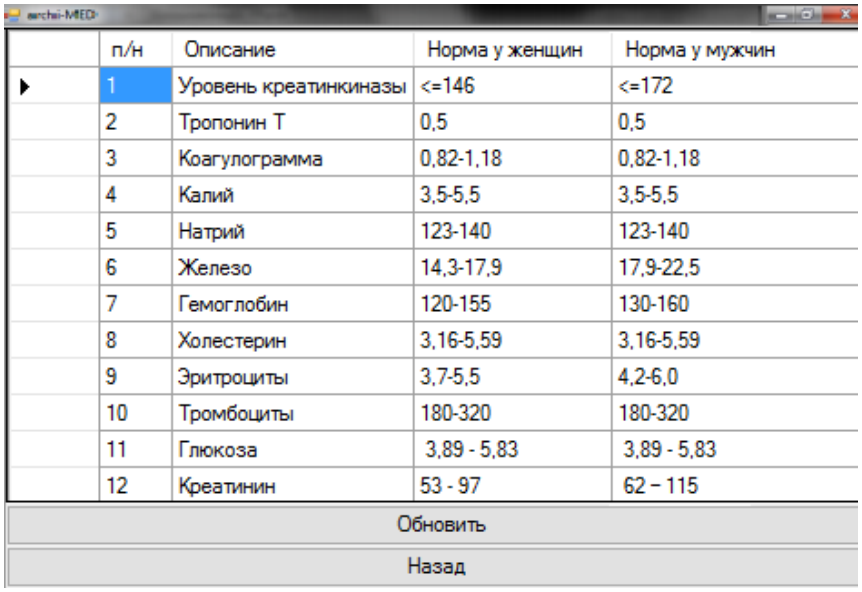


Рис. 5 Форма обработки результатов исследования.



п/н	Описание	Норма у женщин	Норма у мужчин
1	Уровень креатинкиназы	<=146	<=172
2	Тропонин Т	0,5	0,5
3	Коагулограмма	0,82-1,18	0,82-1,18
4	Калий	3,5-5,5	3,5-5,5
5	Натрий	123-140	123-140
6	Железо	14,3-17,9	17,9-22,5
7	Гемоглобин	120-155	130-160
8	Холестерин	3,16-5,59	3,16-5,59
9	Эритроциты	3,7-5,5	4,2-6,0
10	Тромбоциты	180-320	180-320
11	Глюкоза	3,89 - 5,83	3,89 - 5,83
12	Креатинин	53 - 97	62 – 115

Обновить

Назад

Рис. 6 Справочник показателей

В зависимости от выбора метода извлечения показателей, система будет определять класс заболевания пациента путем анализа введенных

данных. В процессе извлечения данных, система может получить либо точное совпадение с существующим в базе прецедентов случаем (заболеванием), либо ближайший случай (заболевание), который имеет наименьшее расстояние с введенными показателями. Если система не находит соответствий, данный диагноз записывается в базу прецедентов.

Для проведения анализа работы каждой метрики в качестве входных параметров возьмем одни и те же показатели исследований:
Пациентка - Астафьева Ирина Николаевна, 1954г.р.

Лабораторные исследования:

- Калий – выше нормы (>5,5 ммоль/л)
- Натрий - выше нормы (>140 ммоль/л)
- Железо – выше нормы (>30,43 мкмоль/л)
- Уровень тромбоцитов - выше нормы (>320 х 10^9)
- Гемоглобин – выше нормы (>155 мкмоль/л)
- Эритроциты – выше нормы (>5,5 ммоль/л)

Инструментальные исследования

- Изменение ЧСС;
- Нарушение проводимости;
- Изменение массы миокарда;
- Изменение объема сердца в состоянии покоя.

Результат работы метрики Евклида представлен на рис 7:

В результате обработки введенных данных система выводит список пациентов, которые относятся к первому классу заболевания. К данному классу заболевания относятся пациенты результаты исследований которых совпадают с входными данными. Как видно из результатов обработки, определить к какому-либо заболеванию относится данных пациент достаточно сложно. Метрика нуждается в доработке.

	п/н	ФИО	Наименование заболевания	Класс заболевани	Результаты исследований
	1	Баринова Наталь...	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	2	Беркинблит Миха...	Сердечная недостат...	2	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	3	Порташин Иван Г...	Перикардит	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	4	Барг Александр ...	Экстрасистолия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	5	Астафьева Ирина...	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	6	Андреева Регина ...	Фибриляция	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	7	Юрьева Анна Юрь...	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	8	Игнатова Марина ...	Экстрасистолия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	9	Рыткина Екатери...	Фибриляция	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...
	10	Групец Иван Васи...	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Гемогл...

Назад

Рис. 7 Вывод метрики Евклида

Результат работы метрики Хемминга представлен на рис 8:

Данная метрика так же не дает однозначно поставить диагноз, близка по результатам к Евклидовой метрики. Нуждается в доработке.

	п/н	ФИО	Наименование заболевания	Класс заболеван	Результаты исследований
	2	Устинова Виктория Леонид...	Сердечная недост...	2	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	6	Тимохов Геннадий Андреев...	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	9	Рыткина Екатерина Валерь...	Фибриляция	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	3	Порташин Иван Григорьевич	Перикардит	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	8	Игнатова Марина Алексан...	Экстрасистолия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	10	Групец Иван Васильевич	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	7	Белов Владимир Николаевич	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	1	Баринова Наталья Евгенье...	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	4	Барг Александр Михайлович	Экстрасистолия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
	5	Астафьева Ирина Николае...	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...

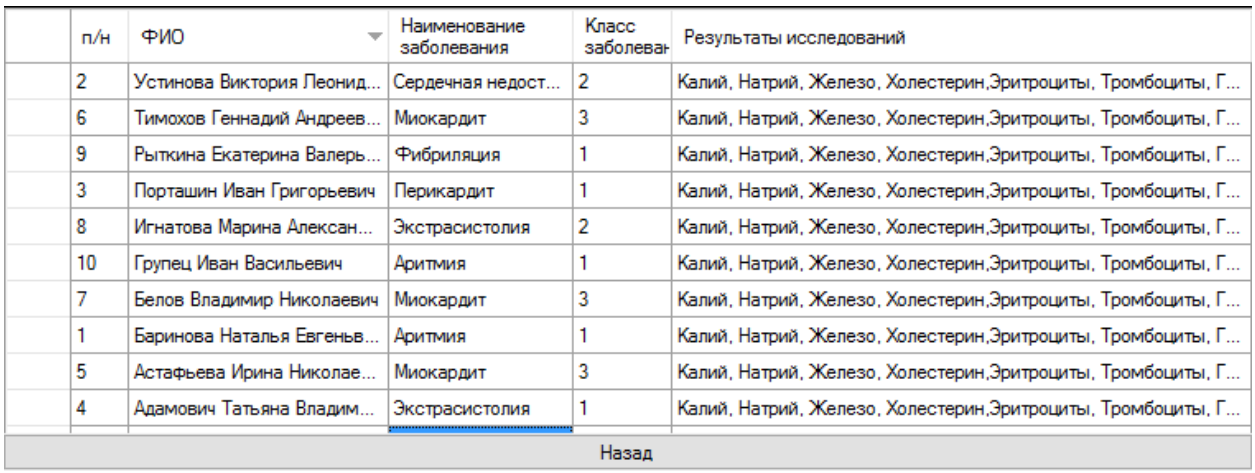
Назад

Рис. 8 Вывод метрики Хемминга

Результат работы метрики Махаланобиса представлен на рис 9:

Метрика Махадонобиса выводит разбросанные результаты, что так же усложняет процесс распознавания того или иного заболевания Результаты достаточно далеки от предыдущих методов вывода.

Необходима доработка метода



п/н	ФИО	Наименование заболевания	Класс заболевани	Результаты исследований
2	Устинова Виктория Леонид...	Сердечная недост...	2	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
6	Тимохов Геннадий Андреев...	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
9	Рыткина Екатерина Валерь...	Фибриляция	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
3	Порташин Иван Григорьевич	Перикардит	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
8	Игнатова Марина Алексан...	Экстрасистолия	2	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
10	Групец Иван Васильевич	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
7	Белов Владимир Николаевич	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
1	Баринова Наталья Евгеньв...	Аритмия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
5	Астафьева Ирина Николае...	Миокардит	3	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...
4	Адамович Татьяна Владим...	Экстрасистолия	1	Калий, Натрий, Железо, Холестерин,Эритроциты, Тромбоциты, Г...

Назад

Рис. 9 Вывод метрики Махаланобиса

Недостаток всех этих методов состоит в том, что они могут порождать разные решения в результате простого переупорядочения объектов в матрице расстояний и, кроме того, их результаты изменяются, если некоторые объекты исключаются из рассмотрения, в данном случае результаты исследований. Так же было установлено, что при больших значениях входных параметров работа алгоритма становится ресурсоемкой задачей и в случае неполноты исходных данных методы становятся неэффективными. Для удобства сравнительной оценки методов как инструмента применяемого для решения задачи компьютерного диагностирования заболеваний сведем данные в единую таблицу 1:

Таблица 1 Сравнительный анализ методов распознавания образов

Характеристики	Наименование метрики		
	Евклидова	Хемминга	Махаланобиса
1	2	3	4
Устойчивость к погрешностям	Неустойчивость к погрешностям	Неустойчивость к погрешностям	Неустойчивость к погрешностям
Простота реализации	Простая	Простая	Простая
Обработка большого количества данных	При неудачном выборе произвольных начальных точек может получиться неоптимальная кластеризация	Ресурсоемкая задача	Плохо работает, если ковариационная матрица высчитывается на всем множестве входных данных
Тип признаков	Количественные	Номинальные(качественные)	Количественные

1	2	3	4
Входные данные	Число кластеров необходимо знать заранее	Одинаковая длина строк	Конкретная группа данных, т. е. внутри кластера. Данная мера расстояния показывает хорошие результаты
Преимущества	-Решения не уникальны для конкретной ситуации, возможно их использование для других случаев - простота реализации	-простота реализации	-эффективна для анализа количественных данных -простота реализации
Недостатки	-необходимость полного перебора данных -вычисление задачи небольшой размерности -Высокая вычислительная трудоемкость - Отсутствие обобщения	- требование одинаковой длины строк - Высокая вычислительная - Отсутствие обобщения	-Высокая вычислительная трудоемкость

Таким образом, основными выявленными недостатками рассмотренных методов вывода, требующими устранения, являются следующие:

1. отсутствие однозначного решения;
2. длительность получения результатов диагностики сердца;
3. ограничение набора результатов исследований, по которым может работать система;
4. большая сложность, трудоемкость и громоздкость проведения динамического анализа лабораторных данных по множеству измеряемых параметров.

Анализ преимуществ и недостатков приведенных методов показал необходимость применения комплексирования методов, направленных на решение задачи диагностики состояния пациента.

Для построения гибридной системы необходимо учесть следующее:

1. Целесообразно использовать метод, обладающий минимальными затратами и вычислительной мощности;
2. В случае наличия признаков характерных для нескольких классов заболеваний, необходимо использовать метод, дающий количественные оценки, т.е. характеризующие принадлежность к конкретному заболеванию

Таким образом, дальнейшая работа по исследованию метрик вывода по прецедентам будет посвящена разработке гибридных моделей с использованием аппарата нечеткой логики. Применение аппарата нечеткой логики позволит повысить эффективность решения задач в реальных условиях неопределенности, а также сделать методы вывода более универсальными.

Глава 3. Проектирование гибридной модели

3.1. Нечеткая логика

Базовыми понятиями в теории нечётких множеств и нечёткой логики являются понятия функции принадлежности и лингвистической переменной . Под нечётким множеством A понимается совокупность упорядоченных пар, составленных из элементов x, универсального множества X и соответствующих степеней принадлежности $\mu_A(x)$ [13]

$$A = \{(x, \mu_A(x)) | \, x \in X \}, \tag{13}$$

Где $\mu_A(x) \rightarrow$ [0,1] - отображение множества X в единичный отрезок [0, 1] – функция принадлежности нечёткого множества A. Значение функции принадлежности $\mu_A(x)$ для конкретного элемента x ∈ X называют степенью принадлежности.

Согласно Л. Заде [12], степень принадлежности $\mu_A(x)$ является субъективной мерой того, насколько элемент x ∈ X соответствует понятию, смысл которого формализуется нечётким множеством A. Обычные множества составляют подкласс класса нечетких множеств. Функцией принадлежности обычного множества B ⊂ U является функция:

$$\mu_{\mathrm{B}}(u) = \begin{cases} 1, u \in B \\ 0, u \notin B \end{cases} \tag{14}$$

Определим используемые в дальнейшем две основные операции над нечеткими множествами, а именно, объединения и пересечения[14]: Объединением нечетких множеств A и B на универсальном множестве U является наибольшее нечеткое множество A ∨ B с функцией принадлежности вида:

$$\mu_{A \vee B}(u) = \max\big(\mu_A(u), \mu_B(u)\big), u \in U \tag{15}$$

Объединение соответствует функции логического элемента ИЛИ, что отвечает с точки зрения лингвистики союзу «ИЛИ». Таким образом, если X и Y – символы нечетких множеств, то

$$X \text{ ИЛИ } Y = X \cup Y \tag{16}$$

Пересечением нечетких множеств A и B на универсальном множестве U является наименьшее нечеткое множество A ∧ B с функцией принадлежности вида:

$$\mu_{A \vee B}(u) = \min\left(\mu_A(u), \mu_B(u)\right), u \in U \tag{17}$$

Пересечение соответствует функции логического элемента И, что отвечает с точки зрения лингвистики союзу «И». Таким образом, если X и Y – символы нечетких множеств, то

$$X \text{ И } Y = X \cap Y \tag{18}$$

На практике для принятия решения на основе нечётких множеств необходимо наличие функций принадлежности, с помощью которых лингвистическая информация превращается в форму, удобную для обработки на ЭВМ. Задача построения функций принадлежности ставится следующим образом. Даны два множества: множество термов

$$L = \{l_1, l_2, \dots, l_m\} \tag{19}$$

и универсальное множество:

$$U = \{u_1, u_2, \dots, u_n\} \tag{20}$$

Нечеткое множество A_j , которым описывается лингвистический терм l_j ,$j = \overline{1,m}$, на универсальном множестве U представляется в виде:

$$A_j = \left(\frac{\mu_{lj}(u_1)}{u_1}, \frac{\mu_{lj}(u_2)}{u_2}, \dots \frac{\mu_{lj}(u_n)}{u_n}\right) \tag{21}$$

Существуют следующие основные методы построения функций принадлежности нечётких множеств [15]:

- на основе матрицы парных сравнений (метод Саати);
- с использованием статистических данных;
- построение функций принадлежности на основе экспертных оценок;

- параметрический метод;
- построение функций принадлежности на основе интервальных оценок.

Методику нечеткого вывода о состоянии пациента можно представить в виде нескольких последовательно выполняемых действий.

1. Формирование функций принадлежностей.

На первом этапе определяют и строят функции принадлежности лингвистических переменных результатов лабораторного и инструментального исследований, характеризующих многомерный образ состояния $A_{NB_iB_j}(t)$ - мгновенное значение $A_N(t)$, находящиеся на пересечения двух и более областей заболеваний B_i . В связи с этим, все результаты исследования, определенные в N-мерном пространстве задают на нечетком множестве. В качестве входных переменных математической модели нечеткого вывода выбраны следующие показателя состояния пациента: лабораторные исследования L, инструментальных исследований I, возраст T.

В качестве методики построения функций принадлежности выбран метод с использованием статистических данных. Он предусматривает наличие списка результатов лабораторных и инструментальных исследований с их количественными показателями частоты встреч для конкретных заболеваний. Шкала лабораторных инструментальных и исследований разбивается на фиксированные интервалы (например, шкала исследования возраст – разбивается на интервалы (25– 30,…, 50–55, … 80–85)).

Наблюдая за n объектами (пациентами) в течение некоторого времени устанавливается, что по некоторому интервалу $i(i = \overline{1,r})$ исследования $j(j = \overline{1,z})$ k раз зафиксирован диагноз «Класс 1» (имеет место следующий набор диагнозов: «Класс 1», «Класс 2», «Класс 3»,…, «Класс 12»). По итогам

наблюдения эксперт фиксирует частоту попадания определенного интервала исследования в заболевание:

$$p_i = \frac{k_i}{n} \tag{22}$$

На основании этих статистических данных строится гистограмма рис.5

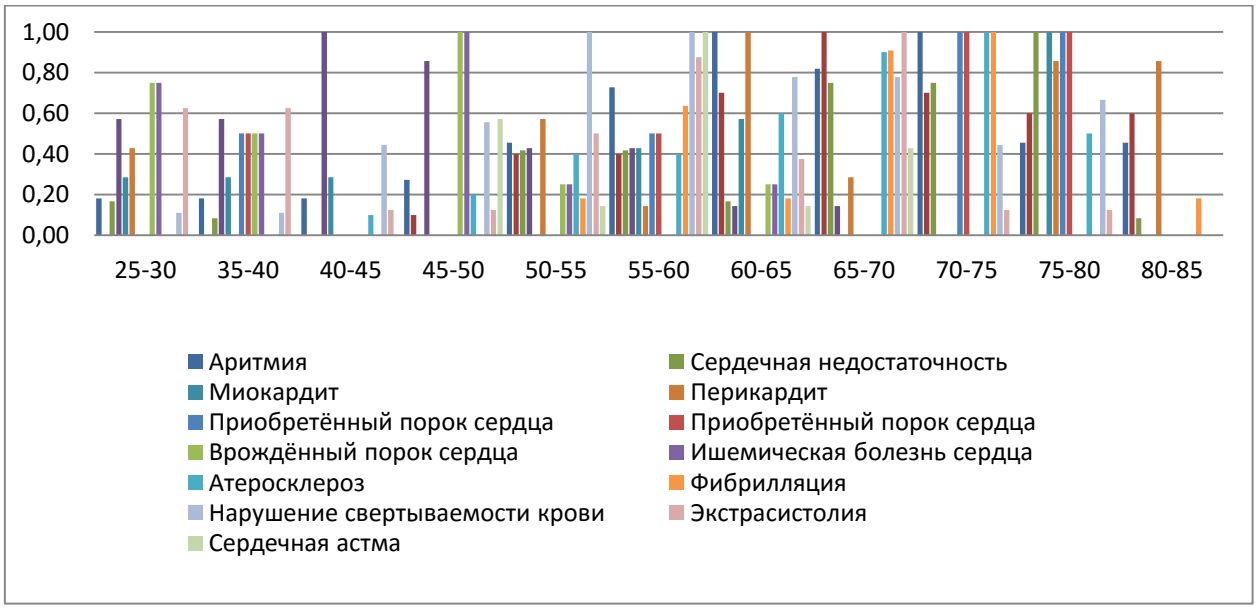


Рис. 10 Гистограмма статистических данных по соотношению заболевание-возраст

Матрица оценки показаний имеет вид, представленный в таблице 2.

Таблица 2 Количество попаданий результатов исследования в определенный интервал в заболеваниях. (Матрица оценки показаний)

Возраст	25-30	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
Аритмия	2	2	2	3	5	8	11	9	11	5	5
Сердечная недостаточность	2	1	0	0	5	5	2	9	9	12	1
Миокардит	2	2	2	0	0	3	4	0	0	7	0
Перикардит	3	0	0	0	4	1	7	2	0	6	6
Приобретённый порок сердца	0	1	0	0	0	1	0	0	2	2	0
Врождённый порок сердца	3	2	0	4	1	0	1	0	0	0	0
Ишемическая болезнь сердца	2	2	0	3	3	5	9	8	11	10	6
Атеросклероз	0	0	1	2	4	4	6	9	10	5	0
Фибрилляция	0	0	0	0	2	7	2	10	11	0	2
Нарушение свертыва	1	1	4	5	9	9	7	7	4	6	0
Экстрасистолия	5	5	1	1	4	7	3	8	1	1	0
Сердечная астма	0	0	0	4	1	7	1	3	0	0	0

На универсальной шкале [0,1] размещены значения интервалов по каждому результату исследования. Тогда степень принадлежности некоторого значения вычисляется как отношение числа экспериментов, в

которых оно встречалось в определенном интервале шкалы, к максимальному для этого значения числу экспериментов по всем интервалам:

$$c_{i\,max} = max\; c_{ij}, \qquad i = \overline{1{,}12}, j = \overline{1{,}11} \tag{23}$$

$$\mu_i = \frac{c_{ij}}{c_{imax}}, \qquad i = \overline{1{,}12}, j = \overline{1{,}11} \tag{24}$$

где c_{ij} – элементы матрицы оценки показаний. Значения функций принадлежности μ_{ij} приведены в таблице 3

Таблица 3Матрица значений функций принадлежности по результатам исследования возраст

Возраст	25-30	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85
μ1т	0,18	0,18	0,18	0,27	0,45	0,73	1	0,82	1	0,45	0,45
μ2т	0,17	0,08	0	0	0,42	0,42	0,17	0,75	0,75	1	0,08
μ3т	0,29	0,29	0,29	0	0	0,43	0,57	0	0	1	0
μ4т	0,43	0	0	0	0,57	0,14	1	0,29	0	0,86	0,86
μ5т	0	0,5	0	0	0	0,5	0	0	1	1	0
μ6т	0,75	0,5	0	1	0,25	0	0,25	0	0	0	0
μ7т	0,18	0,18	0	0,27	0,27	0,45	0,82	0,73	1	0,91	0,55
μ8т	0	0	0,1	0,2	0,4	0,4	0,6	0,9	1	0,5	0
μ9т	0	0	0	0	0,18	0,64	0,18	0,91	1	0	0,18
μ10т	0,11	0,11	0,44	0,56	1	1	0,78	0,78	0,44	0,67	0
μ11т	0,63	0,63	0,13	0,13	0,5	0,88	0,38	1	0,13	0,13	0
μ12т	0	0	0	0,57	0,14	1	0,14	0,43	0	0	0

По найденным значениям μij строятся функции принадлежностей рис.11

Рис. 11 Функции принадлежностей по результатам исследования возраст

По количеству попаданий результатов лабораторного исследования в определенный интервал в заболевании (табл.4) определяются значения соответствующих функций принадлежностей (табл. 5) и строятся сами функции принадлежностей (рис. 12)

Таблица 4 Количество попаданий лабораторных показателей в определенный интервал в заболеваниях. (Матрица оценки показаний)

Лаб-ные исследования	Коагулограмма	Тропонин Т	Креатинкиназа	Биохимия									ОАК		
				Калий	Натрий	Железо	Холестерин	Глюкоза	Креатинин	ЛПНП	ЛПВП	Триглицериды	Эритроциты	Тромбоциты	Гемоглобин
Аритмия	0	0	0	10	10	0	10	0	10	6	6	1	1	1	1
Сердечная недостаточность	0	0	0	0	0	8	7	10	10	4	4	0	0	10	10
Миокардит	7	10	10	0	0	0	0	0	0	0	0	0	10	0	10
Перикардит	0	0	0	0	0	2	10	0	0	6	6	0	10	2	2
Приобретённый порок сердца	4	1	1	4	3	3	2	1	1	2	2	4	10	10	10
Врождённый порок сердца	3	1	1	2	2	2	5	1	1	4	4	2	10	10	10
Ишемическая болезнь сердца	1	1	1	0	0	0	0	1	0	0	0	1	10	10	10
Атеросклероз	0	1	1	0	0	0	10	10	0	4	4	0	4	0	5
Фибрилляция	1	1	1	0	0	0	0	0	0	0	0	0	10	10	10
Нарушение свертываемости крови	0	0	2	1	1	1	0	1	1	0	0	0	10	10	10
Экстрасистолия	0	0	0	10	10	0	10	0	10	4	4	1	1	1	1
Сердечная астма	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Таблица 5 Матрица значений функций принадлежности по лабораторным показателям

Лаб-ные исследования	Коагулог	Тропони	Креатин	Биохимия									ОАК		
	Коагулограмма	Тропонин Т	Креатинкиназа	Калий	Натрий	Железо	Холестерин	Глюкоза	Креатинин	ЛПНП	ЛПВП	Триглицериды	Эритроциты	Тромбоциты	Гемоглобин
μ1L	0	0	0	1	1	0	1	0	1	0,6	0,6	0,1	0,1	0,1	0,1
μ2L	0	0	0	0	0	0,8	0,7	1	1	0,4	0,4	0	0	1	1
μ3L	0,7	1	1	0	0	0	0	0	0	0	0	0	1	0	1
μ4L	0	0	0	0	0	0,2	1	0	0	0,6	0,6	0	1	0,2	0,2
μ5L	0,4	0,1	0,1	0,4	0,3	0,3	0,2	0,1	0,1	0,2	0,2	0,4	1	1	1
μ6L	0,3	0,1	0,1	0,2	0,2	0,2	0,5	0,1	0,1	0,4	0,4	0,2	1	1	1
μ7L	0,1	0,1	0,1	0	0	0	0	0,1	0	0	0	0,1	1	1	1
μ8L	0	0,1	0,1	0	0	0	1	1	0	0,4	0,4	0	0,4	0	0,5
μ9L	0,1	0,1	0,1	0	0	0	0	0	0	0	0	0	1	1	1
μ10L	0	0	0,2	0,1	0,1	0,1	0	0,1	0,1	0	0	0	1	1	1
μ11L	0	0	0	1	1	0	1	0	1	0,4	0,4	0,1	0,1	0,1	0,1
μ12L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 12 Функции принадлежностей по лабораторным исследованиям

По количеству попаданий инструментальных исследований в определенный интервал в заболевании (табл. 6) определяются значения соответствующих функций принадлежностей (табл. 7) и строятся сами функции принадлежностей (рис. 13)

Таблица 6 Количество попаданий инструментальных исследований в определенный интервал в заболеваниях. (Матрица оценки показаний)

Инст-ные исследования	ЭКГ				ЭХГ				
	Изменения ЧСС	Нерегулярный ритм	Несинусовый ритм	Нарушения проводимости	Изменение массы миокарда	Изменение объем в состоянии покоя	Изменения размера в состоянии покоя	Изменение размер в состоянии сокращения	Изменение толщины стенки
Аритмия	10	9	10	10	2	2	1	1	1
Сердечная недостаточность	5	7	1	10	1	2	4	4	6
Миокардит	5	4	4	7	10	9	9	8	9
Перикардит	10	1	10	4	3	3	2	1	1
Приобретённый порок сердца	10	9	9	9	4	5	5	4	4
Врождённый порок сердца	10	9	9	9	4	5	5	4	4
Ишемическая болезнь сердца	10	10	10	7	2	1	1	2	4
Атеросклероз	4	5	4	5	7	10	10	10	9
Фибрилляция	5	5	5	3	10	3	3	4	2
Нарушение свертываемости крови	0	0	0	0	0	0	0	0	0
Экстрасистолия	4	4	5	5	9	10	9	8	8
Сердечная астма	10	10	10	10	1	1	1	1	1

Таблица 7 Матрица значений функций принадлежности по инструментальным исследованиям

Инст-ные исследования	ЭКГ				ЭХГ				
	Изменения ЧСС	Нерегулярный ритм	Несинусовый ритм	Нарушения проводимости	Изменение массы миокарда	Изменение объем в состоянии покоя	Изменения размера в состоянии покоя	Изменение размер в состоянии сокращения	Изменение толщины стенки
μ1I	1	0,9	1	1	0,,2	0,2	0,1	0,1	0,1
μ2I	0,5	0,7	0,1	1	0,1	0,2	0,4	0,4	0,6
μ3I	0,5	0,4	0,4	0,7	1	0,9	0,9	0,8	0,9
μ4I	1	0,1	1	0,4	0,3	0,3	0,2	0,1	0,1
μ5I	1	0,6	0,9	0,9	0,4	0,5	0,5	0,4	0,4
μ6I	1	0,9	0,9	0,9	0,4	0,5	0,5	0,4	0,4
μ7I	1	1	1	0,7	0,2	0,1	0,1	0,2	0,4
μ8I	0,4	0,5	0,4	0,5	0,7	1	1	1	0,9
μ9I	0,5	0,5	0,5	0,3	1	0,3	0,3	0,4	0,2
μ10I	0	0	0	0	0	0	0	0	0
μ11I	0,4	0,4	0,5	0,5	0,9	1	0,9	0,8	0,8
μ12I	1	1	1	1	0,1	0,1	0,1	0,1	0,1

Рис. 13 Функции принадлежностей инструментальных исследований к заболеваниям

2. Нечеткая композиция.

Этап нечеткой композиции состоит из двух операций:

- операция логической конъюнкции,
- операция логической дизъюнкции.

Первая операция используется для того, чтобы определить степень принадлежности точки $A_{NB_iB_j}(t)$, характеризующей состояние пациента, к i-той области заболевания (μ_{min}). При этом необходимо выполнить операцию логической конъюнкцию по полученным значениям относительно каждой области заболевания в соответствии со следующим соотношением:

$$\mu B_i = min\{\mu_{iT}, \mu_{iL}, \mu_{iI}\} \quad (25)$$

Выполнив операцию логической конъюнкции для всех значений i = 1,2,3 выбранных входных переменных математической модели нечеткого вывода в соответствии с (25), получаем следующую совокупность количественных оценок $\mu_{i\,min}$:

$$\mu_{1min} = min\{\mu_{1T}, \mu_{1L}, \mu_{1I}\} \quad (26)$$

$$\mu_{2min} = min\{\mu_{2T}, \mu_{2L}, \mu_{2I}\} \quad (27)$$

$$\mu_{3min} = min\{\mu_{3T}, \mu_{3L}, \mu_{3I}\} \quad (28)$$

$$\dots$$

$$\mu_{12min} = min\{\mu_{12T}, \mu_{12L}, \mu_{12I}\} \quad (29)$$

Вторая операция, а именно, операция логической дизъюнкции, используется для определения максимального значения оценки степени принадлежности образа состояния $A_{NB_iB_j}(t)$ к соответствующей области заболевания Bi , что осуществляется путем сопоставительного анализа полученных совокупностей оценок (27). Тогда для всех значений i = 1,2,3,..,12 выбранных входных переменных математической модели нечеткого вывода получим

$$\mu_{max} = max\{\mu_{1min}, \mu_{2min}, \dots, \mu_{12min}\} \quad (30)$$

3. Разработка продукционной модели базы знаний в виде нечетких правил.

Нечеткая база знаний используется для взаимодействия между входными и выходными параметрами и осуществления поддержки принятия решений. Нечеткая база знаний представляет собой конечное множество нечетких правил (НП). Применительно к иллюстрируемым классам диагнозов и на основе всех значений i = 1,2,3,…,12 выбранных входных переменных математической модели нечеткого вывода можно сформировать следующие продукционные правила[14]:

1. НП 1: Если «$\mu_{1min} = \mu_{max}$» то «Класс 1»; (31)
2. НП 2: Если «$\mu_{2min} = \mu_{max}$» то «Класс 2»;
3. НП 3: Если «$\mu_{3min} = \mu_{max}$» то «Класс 3»;
4. НП 4: Если «$\mu_{4min} = \mu_{max}$» то «Класс 4»;
5. НП 5: Если «$\mu_{5min} = \mu_{max}$» то «Класс 5»;
6. НП 6: Если «$\mu_{6min} = \mu_{max}$» то «Класс 6»;
7. НП 7: Если «$\mu_{7min} = \mu_{max}$» то «Класс 7»;
8. НП 8: Если «$\mu_{8min} = \mu_{max}$» то «Класс 8»;
9. НП 9: Если «$\mu_{9min} = \mu_{max}$» то «Класс 9»;
10. НП 10: Если «$\mu_{10min} = \mu_{max}$» то «Класс 10»;
11. НП 11: Если «$\mu_{11min} = \mu_{max}$» то «Класс 11»;
12. НП 12: Если «$\mu_{12min} = \mu_{max}$» то «Класс 12.

Таким образом, решение задачи базируется на использовании теории нечетких множеств (формулы (13) (30)), которая позволяет осуществить количественную оценку степени принадлежности текущего значения каждого из параметров исследования к каждому из заболеваний в соответствующей точке, принадлежащей области пересечения B_{2i} . Приведенные продукционные правила формируются на основе выбранных входных переменных математической модели нечеткого вывода, но являются универсальными и могут быть расширены в зависимости от наличия большего количества входных переменных.

3.2. Описание алгоритма обработки данных гибридного метода

Применение аппарата нечеткой логики позволит количественно оценить значение степени принадлежности текущего состояния пациента соответствующему образу B_{2i}. Анализируя степени принадлежности точки многомерного виртуального образа $A_N(t)$ к областям заболеваний B_i делаем вывод, что заболевание с наибольшей степенью принадлежности и будет указывать на область(класс) заболевания, в которой находится точка

многомерного виртуального образа $A_N(t)$. В случае, когда степень принадлежности точки многомерного виртуального образа $A_N(t)$ к областям заболеваний B_i одинаковые, происходит отказ от принятия медицинского решения и возникает необходимость на дополнительный ввод параметров.

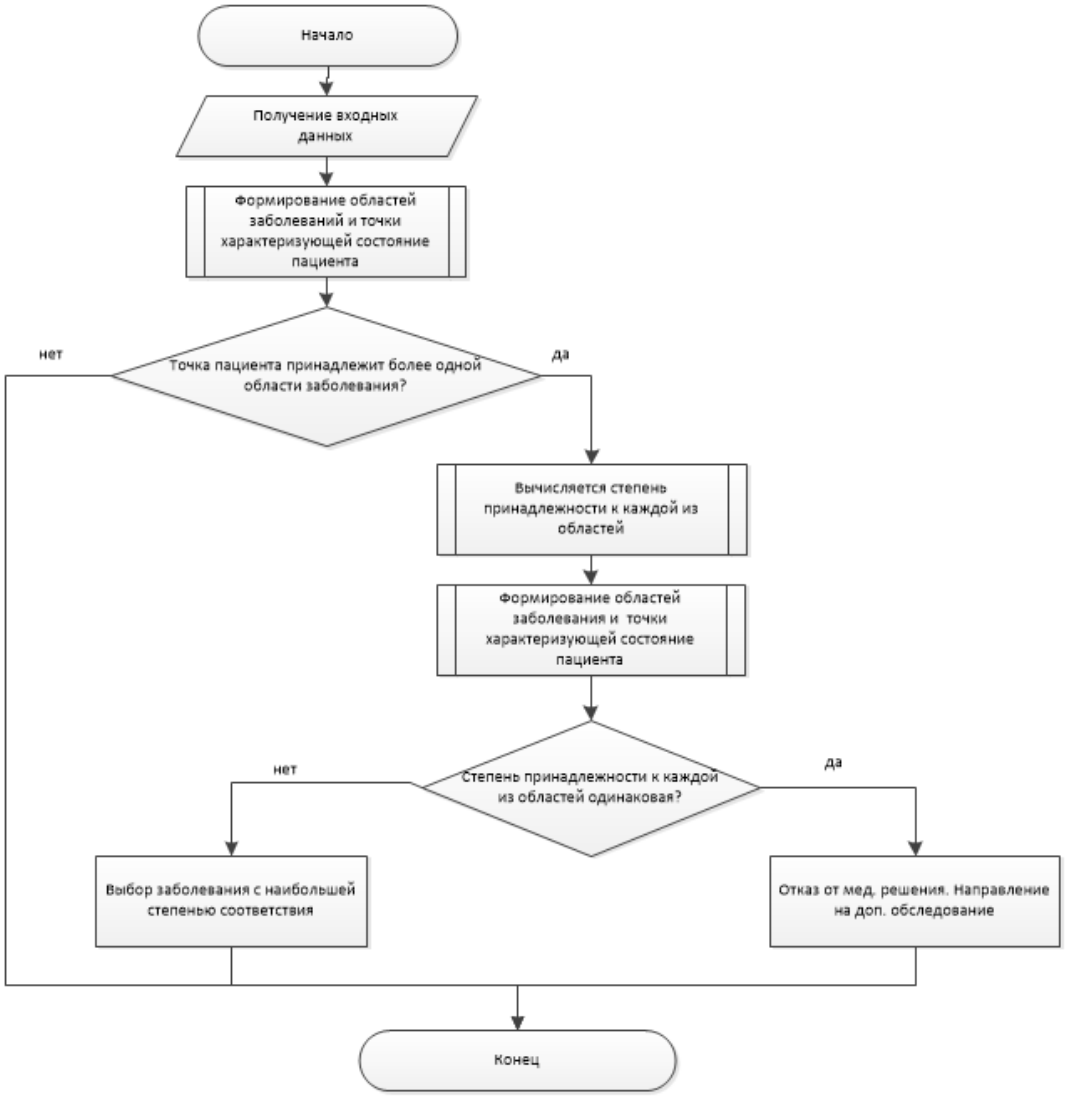


Рис. 14 Схема алгоритма диагностики состояния пациента

Применение предлагаемого способа, при обработке исследований, позволит оперативно контролировать и анализировать закономерности течения болезней, следовательно, в конечном итоге повысит точность диагностики болезней, а в некоторых случаях даст возможность решать прогностические задачи. Рассмотрим более подробно алгоритмизацию метода нечетких множеств для оценки степеней принадлежности точки, характеризующей состояние пациента, к областям заболеваний B_{2i} (рис. 15).

Рис. 15 Алгоритм метода нечетких множеств при неоднозначности принятия о медицинского решения

Предлагаемый способ с использованием теории нечетких множеств направлен на оценку степени принадлежности точки, характеризующей состояние пациента, к каждому заболеванию. На первом шаге строятся функции принадлежности и определяются степени принадлежности по каждому исследованию(лабораторному и инструментальному) к заболеванию в точке, характеризующей состояние пациента

$$\mu B_{ij}(C_j) \tag{32}$$

где $\mu B_{ij}(C_j)$ – степень принадлежности j-того исследования(лабораторного и инструментального) точки, характеризующей состояние пациента, к i-тому заболеванию, $i = \overline{1,m}, j = \overline{1,n}$

На втором шаге оценивается степень принадлежности точки, характеризующей состояние пациента, для каждого заболевания

$$\mu B_i = min\{\mu B_{ij}(C_j)\} \quad (33)$$

где $\mu B_{ij}(C_j)$ – степень принадлежности j-того исследования(лабораторного и инструментального) точки, характеризующей состояние пациента, к i-тому заболеванию, $i = \overline{1,m}$, $j = \overline{1,n}$, μB_i – степень принадлежности точки, характеризующей состояние пациента, к i-тому заболеванию

На третьем шаге вызывается процедура отображения областей заболеваний из многомерного признакового пространства. После чего, на экране отображаются степени принадлежности точки, характеризующей состояние пациента, к каждому заболеванию. Далее, система предлагает возможный вариант диагноза, который соответствует заболеванию с наибольшей степенью принадлежности точки, характеризующей состояние пациента

$$\mu B = max\{\mu B_i\} \quad (34)$$

где μB_i – степень принадлежности точки, характеризующей состояние пациента, для i-того заболевания, μB – степень принадлежности заболевания, с наибольшим значением.

В результате применения этого способа эксперту-врачу предоставляется медицинское решение, на основе которого он может сделать заключительный диагноз. Применительно к классам диагнозов и на основе всех значений выбранных входных переменных математической модели нечеткого вывода можно сформировать следующее общее продукционное правило:

$$\text{НП: Если } "\mu_{i\,min}=\mu_{max}" \text{ то "Класс } i" \quad (35)$$

Таким образом, решение задачи базируется на использовании теории нечетких множеств, которая позволяет осуществить количественную оценку степени принадлежности текущего значения каждого из параметров

состояния к каждому из заболеваний в соответствующей точке, принадлежащей области пересечения

3.3. Разработка функциональной модели системы

Для проектирования модели предметной области будет использоваться функциональный подход [16]. Построение диаграммы IDEF0 позволяет описать все необходимые процессы с точностью, достаточной для однозначного моделирования. Для построения общей модели информационных потоков определим все входящие и исходящие документы, методы и объекты. Входными объектами являются результаты лабораторных и инструментальных исследований пациента и его личные данные. Выходными объектами системы является предположительный диагноз (управляющее медицинское решение). Управляющими объектами являются Нормативные и внутренние документы, а также федеральные законы.

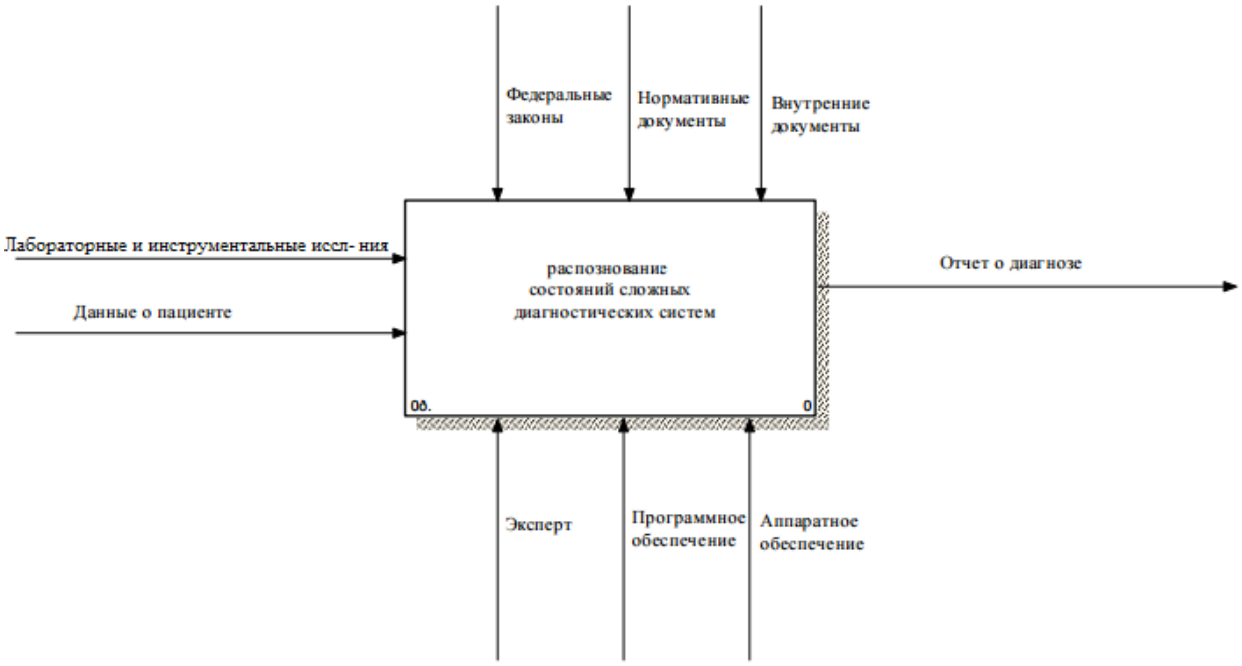


Рис. 16 Функциональная модель распознавания состояний сложных диагностических систем

Результатом работы системы является вывод ряда областей заболеваний, точки, характеризующей состояние конкретного пациента и

определение заболевания у пациента путем исследования топологии взаимного расположения характерных для пациента симптомов, относительно отображенных областей заболеваний. Деятельность осуществляется сотрудниками, имеющим квалификацию врача и являющимися пользователями системы. Часть функций выполняет программное обеспечение, которое подразделяется на общее (операционная система, драйверы и др.) и специализированное (разрабатываемая информационная система и др.) На рис. 17 представлена декомпозиция функциональной модели распознавания состояний сложных диагностических систем.

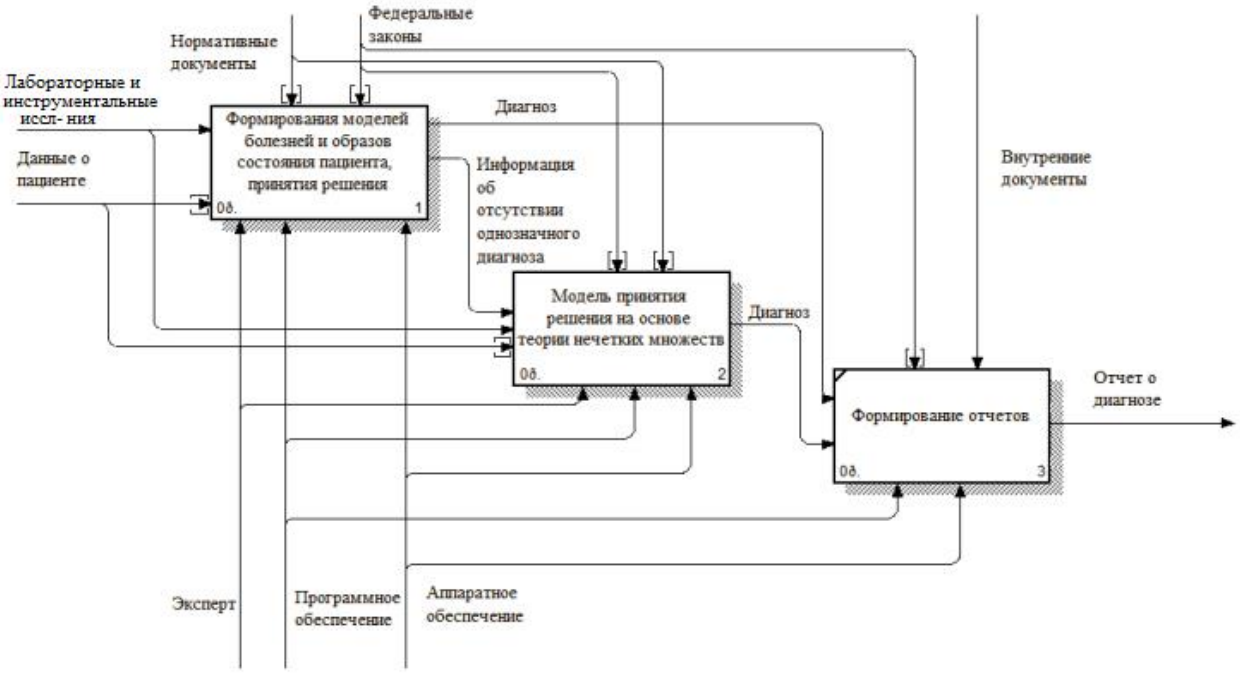


Рис. 17 Декомпозиция функциональной модели

На рис. 18 представлена декомпозиция второго блока функциональной модели «Модель принятия решения на основе теории нечетких множеств». В рамках данного блока реализуется подход на основе теории нечетких множеств

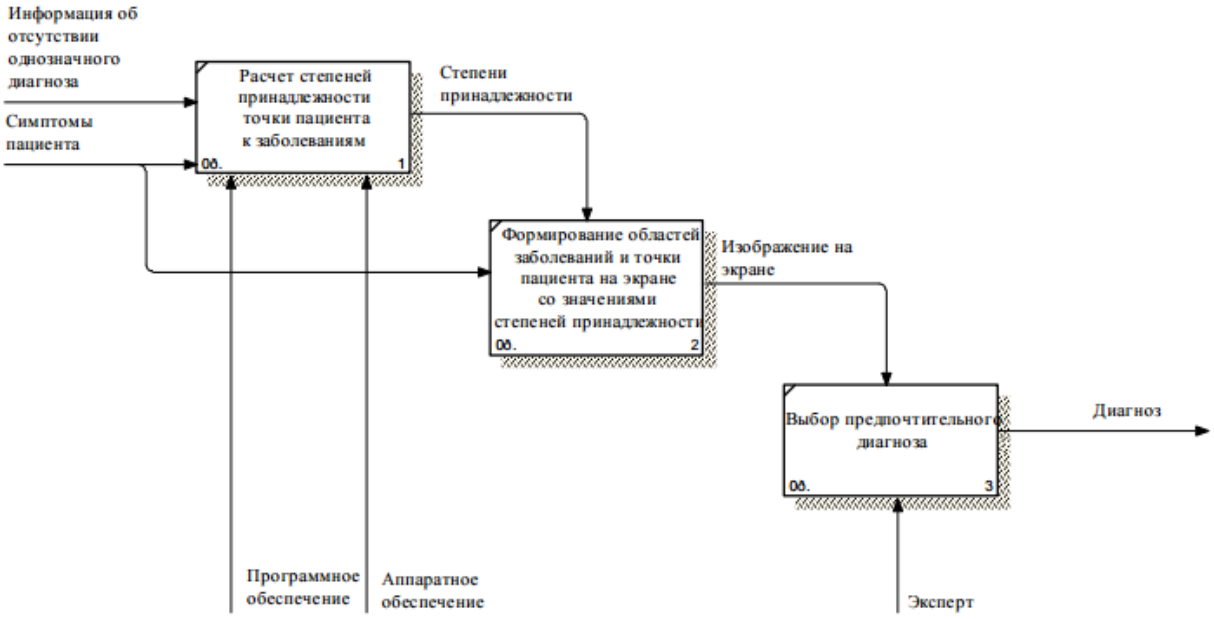


Рис. 18Декомпозиция блока «Модель принятия решения на основе теории нечетких множеств».

3.4. Проектирование гибридной модели

На форму обработки результатов исследования была добавлена кнопка «Гибридный метод», которая реализует алгоритм нечеткой логике рис. 19

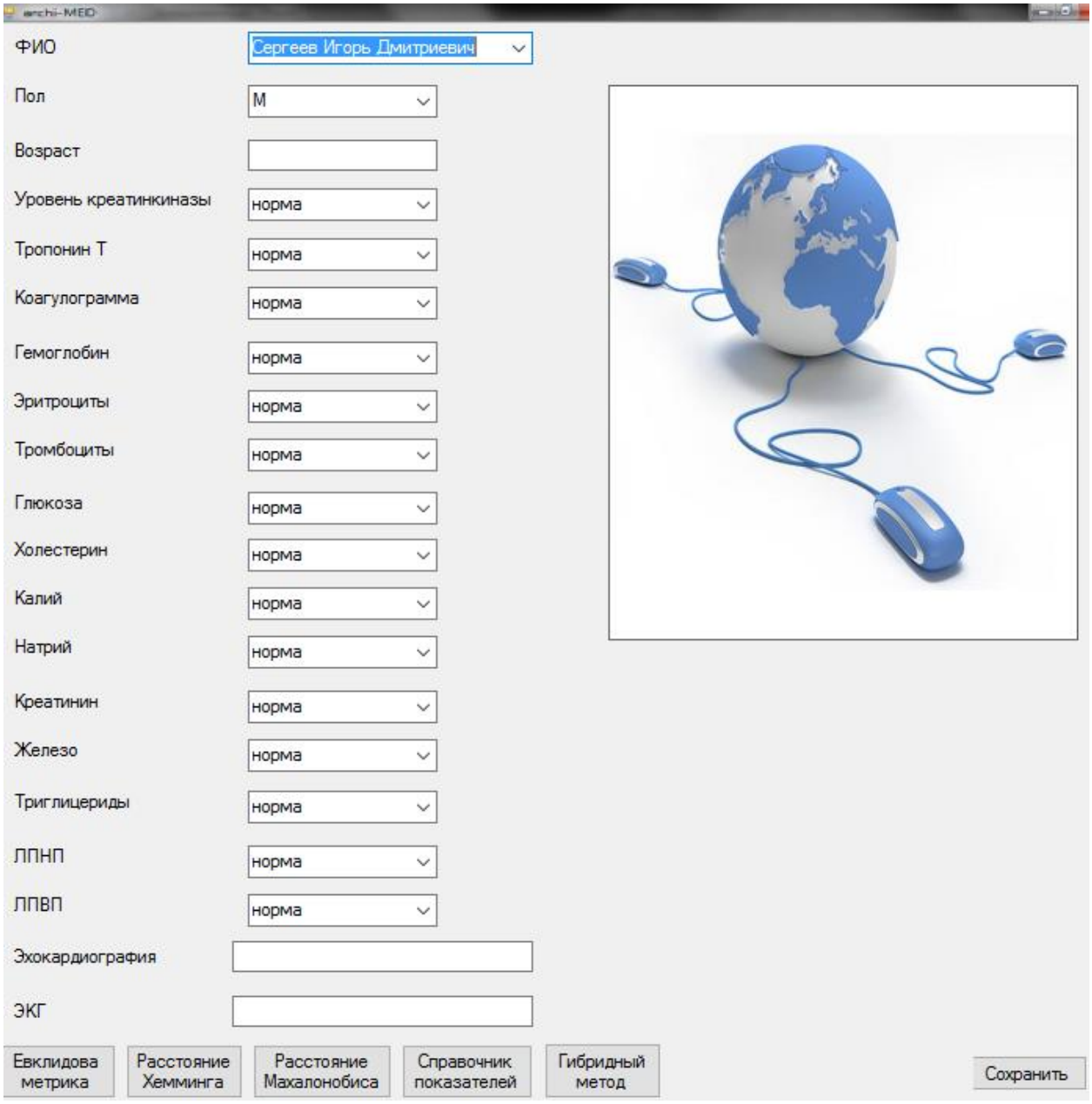


Рис. 19 Форма обработки результатов исследования

При нажатии на кнопку «Гибридный метод» система реализует алгоритм разграничения областей заболеваний путем применения нечетких множеств и дает процентную оценку степени принадлежности пациента к тому или иному заболеванию

Рис. 20 Вывод гибридного метода

Разработанная информационная система предназначена для сбора, структурирования и осуществления анализа имеющейся информации в БД по конкретному пациенту. Программный модуль выработки однозначного медицинского решения осуществляет вывод результатов принадлежности данного набора исследований (т.е. точки, описывающей состояние пациента по заданному набору исследований) к каждому из заболеваний и приводятся в процентном соотношении, таким образом, заболевание с наибольшим процентом (100%) предполагается истинным

Данные по медицинскому решению сохраняются в БД как заключительный диагноз и будут использоваться для записи в кардиологическую историю пациента. Структурированная информация в виде амбулаторной карты позволяет хранить в базе пациентов различное множество диагнозов, тем самым гарантируя нахождение единственно-верного решения.

ГЛАВА 4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

4.1. Предпроектный анализ

4.1.1. Потенциальные потребители результатов исследования

В настоящее время повышение качества оказываемой медицинской помощи является одной из наиболее сложных и значимых задач. С ростом заболеваемости возникает ряд проблем, в частности рационального использования врачебного времени. Так как 50% рабочего времени врача уходит на поиск нужной информации, возросла необходимость в использование диагностических систем. Использование данных систем позволит почти в 4 раза сократить время поиска нужной информации, на 25% сократить время для определения диагноза и на 10-20% увеличить поток принятых врачом пациентов

В качестве объекта исследования были выбраны результаты медицинской диагностики, определяющие патологии сердечно-сосудистой системы.

Основой данных для предметной области служат лабораторные и инструментальные исследования, которые проводят при сердечно - сосудистых заболеваний точка.

Целевая аудитория – это медицинские учреждения г. Томска

Результатом выполнения диссертационной работы стала разработанная информационная система, основанная на реализации алгоритма нечеткой логики. Данная ИС использует инструментальные средства в виде совокупности методов для анализа их многомерных топологических свойств и особенностей для выработки однозначного медицинского решения точка.

Конечными пользователями являются сотрудники медицинского учреждения-врачи, для которых важен удобный интерфейс, простота функционирования, быстродействие и точность результатов.

4.1.2. Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения

Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения.

Для оценки конкурентоспособности составим таблицу с перечнем наиболее важных технических показателей (таблица 3)

Таблица 8. Оценочная карта для сравнения конкурентных технических решений (разработок)

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		$Б_ф$	$Б_{к1}$	$Б_{к2}$	$К_ф$	$К_{к1}$	$К_{к2}$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Технические критерии оценки ресурсоэффективности							
1.Простота в использовании	0,1	4	4	2	0,4	0,4	0,2
2.Надежность хранения данных	0,2	5	1	5	1	0,2	1
3.Гибкость системы	0,1	5	2	4	0,5	0,2	0,4
4. Функциональная мощность (представляемые возможности)	0,16	5	1	5	0,8	0,16	0,8
5.Повышение производительности труда пользователя	0,2	4	5	4	0,8	1	0,8
Экономические критерии оценки эффективности							
1. Конкурентоспособность продукта	0,08	3	4	2	0,24	0,32	0,16
2. Цена	0,04	3	5	1	0,12	0,20	0,04
3. Предполагаемый срок эксплуатации	0,04	4	4	4	0,16	0,16	0,16
4. Послепродажное обслуживание	0,03	4	3	5	0,12	0,09	0,15
5. Финансирование научной разработки	0,05	1	2	5	0,05	0,10	0,25
Итого	**1**	**38**	**31**	**37**	**4,19**	**2,83**	**3,96**

Исходя из результатов оценочной карты, можно сделать вывод, что результат разработки можем считать конкурентоспособным.

4.1.3. SWOTанализ

Использование данной методологии в магистерской диссертации позволит выявить сильные и слабые стороны, а также угрозы и возможности, разрабатываемой ИИС.

Таблица 9. SWOT анализ

	Сильные стороны проекта: С1. Неограниченные возможности в плане проектирования и реализации продукта С2. Создание универсальной системы С3. Бесперебойная работа ИС и свободное хранение С4. Принципиально новая разработка ИИС по диагностированию патологий ССС	Слабые стороны проекта: Сл1. Отсутствие у разработчиков опыта работы в данной теме Сл2. Высокая стоимость разработки базы СЛ3. Небольшой функционал. СЛ4. Не испытана в работе при большом количестве входных данных
Возможности: В1. Использование инновационной инфраструктуры ТПУ В2. Повышение стоимости конкурентных разработок В3. Появление дополнительных финансовых ресурсов.	В1В3С1С2С4 Стать основным поставщиком ПП. В1В3С3 Повышение уровня надежности разрабатываемой ИИС В2С2С3С4Стать основным поставщиком по проектированию ИИС по диагностированию патологий ССС	Сл1В1 Повышение квалификации разработчика Сл3Сл4В1В3 Усовершенствование системы диагностирования патологий ССС
Угрозы: У1. Отсутствие спроса на новые технологии производства У2. Появление конкурентов на рынке У3. Увеличение срока выхода на рынок при неудовлетворительных результатах испытания	У1С2С Создание универсальной системы увеличивает спрос на данную разработку У2С2С3С4 Снижение цен за счет увеличения объемов	У1У2Сл1Сл3Изменение направления проектирования ИС У1У2Сл2Закрытие проекта

На основе приведенного выше анализа можно сделать выводы о высоком потенциале разрабатываемого проекта, для того чтобы его развить

необходимо умело использовать имеющиеся возможности и сильные стороны, а так же учесть угрозы и устранить слабые стороны.

Таким образом, в целях увеличения объемов продаж ПП возможны следующие рекомендации:

Для устранения слабых сторон необходимо использовать возможности:

1.1. Разработчика следует направить на повышение квалификации.

1.2. Усовершенствовать разработанную ИИС путем использования инновационной инфраструктуры ТПУ и привлечением дополнительных финансовых ресурсов

Для закрепления сильных сторон, необходимо:

1. Использовать развитую инфраструктуру ТПУ для повышение уровня надежности разрабатываемой ИИС

Для устранения угроз необходимо:

1. Устранить конкуренцию по разработке интеллектуальной информационной системы с использованием интеллектуального анализа данных при диагностировании патологий ССС, путем создания принципиально новой, бесперебойной ИИС

По проведенному SWOT анализу, можно сделать вывод, что проект по разработке интеллектуальной информационной системы по диагностированию патологий ССС занимает высокое место на данном рынке, имеет сильные стороны и стремиться устранить все виды угроз.

4.1.4. Оценка готовности проекта к коммерциализации

На любой стадии жизненного цикла научной разработки полезно оценить степень ее готовности к коммерциализации и выяснить уровень собственных знаний для ее завершения. Таблица 10 содержит показатели, говорящие о степени проработанности проекта с позиции коммерциализации и компетенциям разработчика научного проекта.

Таблица 10. Бланк оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1.	Определен имеющийся научно – технический задел	5	5
2.	Определены перспективные направления коммерциализации научно – технического задела	5	5
3.	Определены отрасли и технологии (товары, услуги) для предложения на рынке	4	2
4.	Определена товарная форма научно – технического задела для представления на рынок	4	5
5.	Определены авторы и осуществлена охрана их прав	3	2
6.	Проведена оценка стоимости интеллектуальной собственности	4	4
7.	Проведены маркетинговые исследования рынков сбыта	5	5
8.	Разработан бизнесплан коммерциализации научной разработки	1	2
9.	Определены пути продвижения научной разработки на рынок	2	1
10.	Разработана стратегия (форма) реализации научной разработки	5	5
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	1
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	1	1
13.	Проработаны вопросы финансирования коммерциализации научной разработки	1	1
14.	Имеется команда для коммерциализации научной разработки	1	1
15.	Проработан механизм реализации научного проекта	5	5
	ИТОГО БАЛЛОВ	47	45

Значение $Б_{сум}$ позволяет говорить о мере готовности научной разработки и ее разработчика к коммерциализации.

В данном случае значения $Б_{сум}$ входят в диапазон от 47 до 45. Можно сделать вывод, что перспективность проекта выше среднего. Для дальнейшего улучшения проекта, может возникнуть потребность в привлечение новых специалистов в команду.

4.1.5. Методы коммерциализации результатов научно-технического исследования

Коммерциализация представляет собой процесс разработки и реализации ряда мероприятий, с помощью которых результаты научных исследований и опытно-конструкторских разработок можно предложить на рынках товаров и услуг с коммерческими целями.

В процессе коммерциализации самым важным моментом для инновационных проектов является выбор способа коммерциализации.

В качестве метода коммерциализации инновационной продукции была выбрана лицензионная торговля, которая осуществляется на основе лицензионных соглашений. При заключении лицензионного соглашения собственник передает права на нематериальные активы покупателю на определенный срок и за известную плату. Объектом лицензионного соглашения могут быть: патенты, изобретения, формулы, процессы, конструкции, схемы, торговые марки, франшизы, программы и т. д. При этом соглашение обычно обязывает лицензиара (продавца технологии) предоставить техническую информацию и помощь, а лицензиата (покупателя) – использовать полученные права.

В зависимости от объема прав, предоставляемых лицензионным соглашением покупателю и продавцу, различают простую, исключительную и полную лицензии. Простая лицензия оставляет за продавцом право предоставлять аналогичные лицензии третьей стороне. Исключительная лицензия предоставляет покупателю монопольное право на использование данной технологии, а продавец в этом случае не может ни самостоятельно использовать запатентованную технологию, ни передавать ее третьим лицам

на данной территории. В случае полной лицензии продавец полностью лишается права использовать объект лицензии.

Далее в проекте будет рассматриваться продажа исключительной лицензии. Стоимость одной лицензии составит 65 тыс. руб.

4.2. Инициация проекта

4.2.1 Цели и результат проекта.

Заинтересованными сторонами проекта являются руководитель проекта, исполнитель, а также организация, технические средства которой были использованы во время выполнения проекта – НИ ТПУ. Заинтересованные стороны и их ожидания представлены в таблице 11.

Таблица 11. Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Руководитель проекта	Реализация научных задумок
Исполнитель проекта	Написанная магистерская диссертация
НИ ТПУ	Созданный проект, обладающий научной новизной
Медицинские учреждения города Томска	Созданный проект, а также применение, внедрение данного программного продукта с целью повышения эффективности работы труда.
Пациенты	Внедрение системы по повышению качества диагностирования патологий ССС.
Инвесторы	Готовый проект, ожидание полной окупаемости проекта в кратчайшие сроки (до года)

В таблице 12 представлена информация о иерархии целей проекта и критериях достижения целей.

Таблица 12. Цели и результат проекта

Цель проекта:	Создание интеллектуальной информационной системы по диагностированию патологий ССС.
Ожидаемые результаты проекта:	Повышение качества оказываемой медицинской помощи, в частности рационального использования врачебного времени по диагностированию патологий ССС
Критерии приемки результата проекта:	Бесперебойная работа ИС
Требования к результату проекта:	ИИС, способная хранить и обрабатывать большое количество информации
	Интуитивно понятный интерфейс.
	Возможность пополнения системы (базы знаний)
	ИИС вывода результатов принадлежности пациента к каждому из заболеваний в процентном соотношении
	Быстрая обработка входных параметров при диагностировании патологи ССС

4.2.2. Организационная структура проекта.

В таблице 13 представлена информация о составе рабочей группы проекта и описаны роль, функции и трудозатраты каждого из ее участников.

Таблица 13. Рабочая группа проекта

№ п/п	ФИО	Роль в проекте	Функции	Трудозатраты, час.
1.	Фофанов Олег Борисович	Руководитель проекта	Управление проектом, координирование и анализ выполнения проекта.	308
2.	Алексеева Алина Александровна	Исполнитель	Реализация проекта	524
Итого:				832

4.2.3. Ограничения и допущения проекта

Ограничивающие факторы проекта представлены в таблице 14

Таблица 14. Ограничения проекта

Фактор	Ограничения/ допущения
3.1. Бюджет проекта:	631,9 тыс. руб.
3.1.1. Источник финансирования	Инвстор
3.2. Сроки проекта:	
3.2.1. Дата утверждения плана управления проектом	11.01.2016
3.2.2. Дата завершения проекта	20.05.2016
3.3. Прочие ограничения и допущения	
3.3.1. Ограничения по времени работы участников проекта	Ограниченное время работы.

4.3. Планирование управления научно-техническим проектом

4.3.1. Иерархическая структура работ проекта

Иерархическая структура работ (ИСР) – детализация укрупненной структуры работ. В процессе создания ИСР структурируется и определяется содержание всего проекта. На рис. 21 представлена иерархическая структура работ по проекту разработки программного продукта.

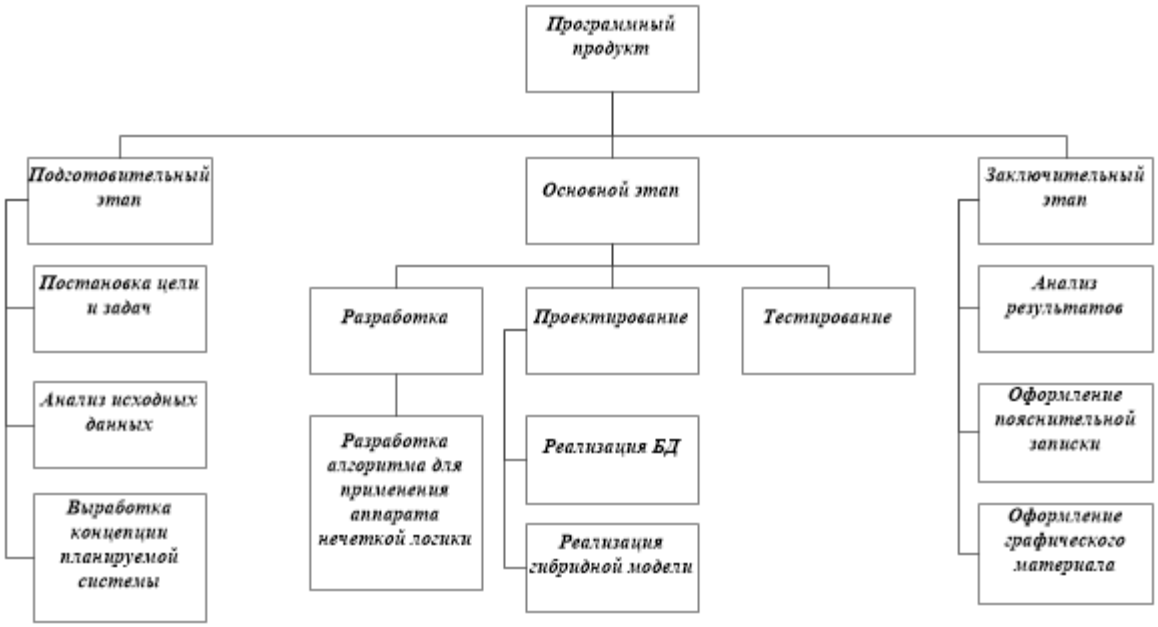


Рис. 21 Иерархическая структура работ по проекту разработки

4.3.2. Контрольные события проекта

При организации процесса реализации проекта необходимо оптимально планировать сроки проведения работ. В таблице 15 определены ключевые события проекта, их даты и результаты, которые должны быть получены по состоянию на эти даты.

Таблица 15. Контрольные события проекта

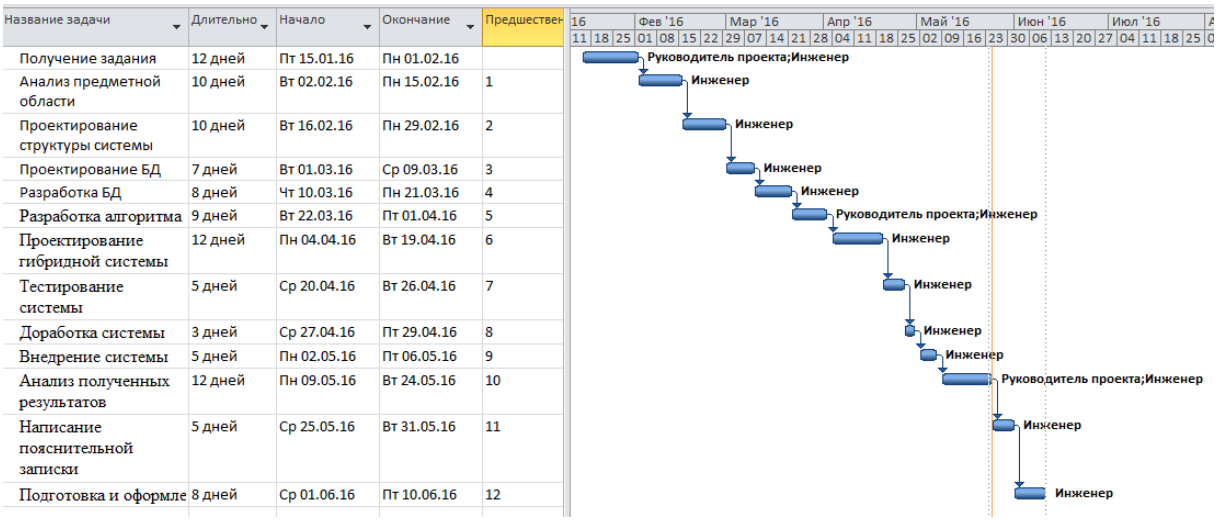
№ п/п	Контрольное событие	Дата	Результат
Подготовительный этап			
1	Получение задания	15.01.2016	Отчет
2	Анализ предметной области	01.02.2016	Отчет
Основной этап			
4	Проектирование структуры системы	12.02.2016	Отчет
5	Проектирование базы данных системы	25.02.2016	База данных
6	Разработка базы данных системы	04.03.2016	База данных
7	Разработка алгоритма	15.03.2016	Алгоритм

8	Проектирование гибридной системы	25.03.2016	Программа на языке C#
9	Тестирование системы	11.04.2016	Отчет
10	Доработка системы	15.04.2016	Программа на языке C#
11	Внедрение системы	19.04.2016	Программа на языке C#, Отчет
Заключительный этап			
12	Анализ полученных результатов	25.04.2016	Отчет
13	Подготовка пояснительной записки	10.05.2016	Пояснительная записка
14	Подготовка образца программного продукта для инвестора	16.05.2016	Программный продукт

4.3.3. План проекта

Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

График построен с разбивкой по месяцам и декадам за период времени разработки проекта по диагностированию патологий ССС (рис. 22).



Название задачи	Длительно	Начало	Окончание	Предшествен
Получение задания	12 дней	Пт 15.01.16	Пн 01.02.16	
Анализ предметной области	10 дней	Вт 02.02.16	Пн 15.02.16	1
Проектирование структуры системы	10 дней	Вт 16.02.16	Пн 29.02.16	2
Проектирование БД	7 дней	Вт 01.03.16	Ср 09.03.16	3
Разработка БД	8 дней	Чт 10.03.16	Пн 21.03.16	4
Разработка алгоритма	9 дней	Вт 22.03.16	Пт 01.04.16	5
Проектирование гибридной системы	12 дней	Пн 04.04.16	Вт 19.04.16	6
Тестирование системы	5 дней	Ср 20.04.16	Вт 26.04.16	7
Доработка системы	3 дней	Ср 27.04.16	Пт 29.04.16	8
Внедрение системы	5 дней	Пн 02.05.16	Пт 06.05.16	9
Анализ полученных результатов	12 дней	Пн 09.05.16	Вт 24.05.16	10
Написание пояснительной записки	5 дней	Ср 25.05.16	Вт 31.05.16	11
Подготовка и оформле	8 дней	Ср 01.06.16	Пт 10.06.16	12

Рис. 22 Диаграмма Ганта

4.3.4. Бюджет научного исследования

Определение затрат на выполнение ВКР производится путем составления калькуляции по отдельным статьям затрат всех видов необходимых ресурсов. Калькуляция является основным документом, на основании которого осуществляется планирование и учет затрат на научные исследования.

Калькуляция плановой себестоимости реализации проекта составляется по следующим статьям затрат:

- затраты на оплату труда работников, непосредственно участвующих в проекте;
- страховые взносы;
- расчет расходов на оборудование
- работы, выполняемые сторонними организациями;
- спецоборудование для научных и экспериментальных работ;
- прочие прямые расходы;

Затраты на проектирование

1. Материалы и покупные изделия

К этой статье относится стоимость всех материалов. Стоимость материалов формируется, исходя из цены их приобретения. Расчет затрат приводится в таблице 16.

На статью «Материалы» относятся следующие затраты:

1. Сырьё, основные и вспомогательные материалы;
2. Покупные полуфабрикаты и комплектующие изделия;
3. Электроэнергия на технологические цели.

Затраты на электроэнергию рассчитываются по формуле

$$С_{эл} = Ц_{эл} \cdot Р \cdot F_{об} \quad (36)$$

где:

Цэл – тариф на промышленную электроэнергию (2,93 руб/кВт·ч по одноставочному тарифу);

Р – мощность оборудования, кВт;

$F_{об}$ – время использования оборудования, ч.

$С_{эл}$ = 2,93 * 0,4кВт * (106д. * 8ч) = 993,86 руб.

Таблица 16. Материалы, комплектующие изделия и покупные полуфабрикаты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Сумма, руб.
Офисная бумага (формат А4)	Упаковка	1	190	190
Картридж для принтера (черный)	Штук	1	953	953
Картридж для принтера (цветной)	Штук	1	1095	1095
Канцелярские товары	Комплект	1	391	391
Электроэнергия				994
Итого:				**3 623**

2. Основная заработная плата

Основная заработная плата рассчитывается по формуле:

$$С_{зп} = З_{ос} + З_{доп} \quad (37)$$

где $З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата.

Основная заработная плата ($З_{осн}$) рассчитывается по следующей формуле:

$$З_{ос} = З_{дн} \cdot Т_{раб} \quad (38)$$

где $З_{осн}$ – основная заработная плата одного работника;

$Т_{раб}$ – продолжительность работ, выполняемых научнотехническим работником;

$З_{дн}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$З_{дн} = \frac{З_м \cdot М}{F_д} \quad (39)$$

где $З_м$ – месячный должностной оклад работника, руб.;

М – количество месяцев работы без отпуска в течение года:

$F_д$ – действительный годовой фонд рабочего времени научно технического персонала.

Таблица 17. Баланс рабочего времени

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней выходные дни праздничные дни	116	116
Потери рабочего времени отпуск невыходы по болезни	24	24
Действительный годовой фонд рабочего времени, час	1800	1800

Месячный должностной оклад работника:

$$З_м = З_д \cdot (1 + k_{пр} + k_p) \quad (40)$$

где $З_б$ – базовый оклад, руб.;

$k_{пр}$ – премиальный коэффициент;

$k_д$ – коэффициент доплат и надбавок;

$k_р$ – районный коэффициент, равный 1,3 (для Томска).

Затраты на оплату труда студента магистранта могут определяться как оклад инженера кафедры (учебно вспомогательный персоналу) в соответствии с квалификационным уровнем профессиональной квалификационной группы. Данный показатель равен 7 800 руб. Затраты на оплату труда руководителя возьмем равными 15 860 руб.

Таблица 18. Расчет основной заработной платы

Исполнители	Зб, руб.	*k*пр	*k*д	*k*р	Зм, руб	Здн, руб.	Тр, раб. дн.	Зосн, руб
Руководитель	28900	0,2	0,3	1,3	56355	3021	24	72517
Инженер	8800	0,2	0,3	1,3	17160	920	106	97520

3. Дополнительная заработная плата

В данную статью включается сумма выплат, предусмотренных законодательством о труде, например, оплата очередных и дополнительных отпусков; оплата времени, связанного с выполнением государственных и общественных обязанностей; выплата вознаграждения за выслугу лет и т.п.

Дополнительная заработная плата рассчитывается исходя из 10 15% от основной заработной платы работников, непосредственно участвующих в выполнение темы:

$$З_{доп} = k_{доп} + З_{осн} \quad (41)$$

где $З_{доп}$ – дополнительная заработная плата, руб.;

$k_{доп}$ – коэффициент дополнительной зарплаты;

$З_{осн}$ – основная заработная плата, руб.

В таблице 19 приведена форма расчёта основной и дополнительной заработной платы.

Таблица 19. Заработная плата исполнителей НТИ

Заработная плата	**Руководитель**	**Инженер**
Основная зарплата	72 517	97 520
Дополнительная зарплата	8 702	11 702
По статье *С*зп	81 219	109 222
Итого по статье *С*зп:	190 441	

4. Страховые взносы

Статья включает в себя отчисления во внебюджетные фонды и определяется по следующей формуле:

$$С_{внеб} = k_{внеб} \cdot (З_{осн} \cdot З_{доп}) \quad (42)$$

где $K_{внеб}$ – коэффициент, учитывающий размер отчислений из заработной платы. Данный коэффициент составляет 30% от затрат на заработную плату и включает в себя:

1) отчисления в пенсионный фонд;
2) на социальное страхование;
3) на медицинское страхование.

Отчисления из заработной платы составили:

$$С_{внеб} = 0,3 * 190\ 441 = 57\ 132$$

5. Расчет расходов на оборудование

Для реализации интеллектуальной ИС по диагностированию патологий ССС использовался один персональных компьютера (ПК) и принтер. Статья расходов на специальное оборудование в данном случае будет определяться в виде амортизационных отчислений, которые рассчитываются следующим образом:

$$С_А = \frac{Ф_{пер} \cdot Н_А}{100\%} \quad (43)$$

где:

$Ф_{пер}$ – первоначальная стоимость оборудования, рубли;

$Н_А$ – годовая норма амортизации.

Норма амортизации рассчитывается по формуле:

$$Н_А = \frac{1}{Т_{сл}} \quad (44)$$

где:

$Т_{сл}$ – срок службы оборудования, лет.

Специальное оборудование, используемое в ВКР, приведено в таблице 20.

Таблица 20. Амортизационные отчисления

Наименование	*Количество, шт*	$Ф_{пер.}$ *руб.*	$Т_{сл}$ *, лет*	$Н_А$*, %*	$С_А$*, руб.*
Компьютер	1	27 730	4	25	6 932,5
Принтер	1	3 500	2	50	1 750
Итого:		**31 230**			**8 682,5**

Таким образом, амортизационные отчисления будут составлять 8682,5рублей в год. За время выполнения проекта – 3 месяца, амортизационные отчисления, пропорционально, составят 2170,62 руб.

6. Прочие прямые расходы

В эту статью включаются затраты на управление и хозяйственное обслуживание, которые могут быть отнесены непосредственно на конкретную тему. Кроме того, сюда относятся расходы по содержанию, эксплуатации и ремонту оборудования, производственного инструмента и инвентаря, зданий, сооружений и др. В расчетах эти расходы принимаются в размере 50 % от суммы основной заработной платы научно-производственного персонала данной научно-технической организации.

$$С_{пр} = k_{пр} \cdot (З_{осн} \cdot З_{доп}) \quad (45)$$

где $k_{пр}$ – коэффициент прямых расходов.

Таким образом, прочие прямые расходы составили:

$С_{пр}$ = 0,5 * 190441 = 95 220,50

7. Смета затрат на разработку

На основании полученных данных по отдельным статьям затрат составляется калькуляция плановой себестоимости НТИ. Полная смета затрат приведена в таблице 21.

Таблица 21. Полная смета затрат

Наименование статьи	Затраты, руб.
1. Сырье, материалы, покупные изделия и полуфабрикаты.	3 623
2. Специальное оборудование для научных работ	2170,62
3. Основная заработная плата	170 037
4. Дополнительная заработная плата	20 404
5. Отчисления на социальные нужды	57 132
6. Научные и производственные командировки	
7. Оплата работ, выполняемых сторонними организациями и предприятиями	
8. Накладные расходы	133 309
Бюджет проекта	**384 505**

4.3.5. Организационная структура проекта

В практике используется несколько базовых вариантов организационных структур: функциональная, проектная, матричная. Проектную структура данная диссертационной работы имеет следующий вид (рис.23):

Рис. 23 Структура проекта

4.3.6. Матрица ответственности

Для распределения ответственности между участниками проекта сформирована матрица ответственности, представленная в таблице 13.

Степень участия в проекте может характеризоваться следующим образом:

- Ответственный (О) – лицо, отвечающее за реализацию этапа проекта и контролирующее его ход.

- Исполнитель (И) – лицо (лица), выполняющие работы в рамках этапа проекта.
- Согласующее лицо (С) – лицо, осуществляющее анализ результатов проекта и участвующее в принятии решения о соответствии результатов этапа требованиям.

Таблица 22. Матрица ответственности

Этапы проекта	Руководитель проекта	Инженер	Инвесторы
Подготовительный этап			
Получение задания	О,И		С
Анализ предметной области	О	И	
Основной этап			
Проектирование структуры системы	О,И	И	
Проектирование базы данных системы		И	
Разработка базы данных системы	О	И	
Разработка алгоритма	О	И	
Проектирование гибридной системы	О	И	
Тестирование системы	О	И	
Доработка системы		И	С
Внедрение системы	О	О,И	
Заключительный этап			
Анализ полученных результатов	О,И	И	С
Написание пояснительной записки		И	У
Подготовка образца программного продукта для инвестора		О,И	С

4.3.7. План управления коммуникациями проекта

План управления коммуникациями для проекта представлен в таблице 23. Он отражает требования к коммуникациям со стороны участников проекта.

Таблица 23. План управления коммуникациями

№ п/п	Какая информация передается	Кто передает информацию	Кому передается информация	Когда передает информацию
1	Текущее состояние проекта	Исполнитель проекта	Руководителю проекта	Раз в две недели
2	Обмен информации о текущем статусе проекта	Исполнитель проекта	Инженеру учебного отдела ИК по работе с магистрантами	Не позже дня, указанного в учебном плане
3	О выполнении этапа проекта	Исполнитель проекта	Руководителю проекта	Не позже дня контрольного события по календарному графику

4.3.8. Реестр рисков проекта

Идентифицированные риски проекта включают в себя возможные неопределенные события, которые могут возникнуть в проекте и вызвать последствия, которые повлекут за собой нежелательные эффекты. Причиной возникновения рисков являются неопределенности, существующие в каждом проекте.

Таблица 24 Реестр рисков

№	Риск	Потенциальное воздействие	Вероятность наступления	Влияние риска	Уровень риска	Способы смягчения риска	Условия наступления
1	Прекращение (сокращение) финансирования	Отсутствие возможности оплаты расходов	2	4	Средний	Привлечение сторонних инвесторов, планирование финансовых ресурсов	Спонсор становится банкротом либо убеждается в экономической неэффективности проекта
2	Потеря всех наработок (уничтожение жесткого диска)	Потеря временных и финансовых ресурсов	1	5	Средний	Хранение наработок в облачном хранилище	ЧП (пожар, землетрясение), поломка оборудования
3	Созданный программный продукт не отвечает	Отсутствие спроса на продукт	3	3	Средний	Тщательно описанные требования к ПП, проверка на их выполнение после	Неясно описанные требования, несогласованность выполнения работ

	требованиям					каждого этапа	
4	Невыполнение запланированных работ, не достижение плановых объемов производства	Недоработан-ный функционал программного продукта	5	4	Высокий	Тщательно продуманный календарный план, продолжительность этапов включает небольшой запас	Затянувшийся по времени этап, неправильно составленный график

4.5. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

4.5.1 Оценка абсолютной эффективности исследования

В основе проектного подхода к инвестиционной деятельности предприятия лежит принцип денежных потоков (cash flow). Особенностью является его прогнозный и долгосрочный характер, поэтому в применяемом подходе к анализу учитываются фактор времени и фактор риска.

Степень устойчивости проекта по отношению к возможным изменениям условий реализации может быть охарактеризована показателями границ безубыточности. Уровнем безубыточности $УБ_t$ в периоде t называется отношение "безубыточного" объема продаж (производства) к проектному на этом шаге. Под "безубыточным" понимается объем продаж, при котором чистая прибыль становится равной нулю. Оценка уровня безубыточности для проекта представлена в таблице 25.

Таблица 25. Расчет безубыточности

№	Показатель, тыс.руб	Номер шага (периода)расчета (t)				
		1	2	3	4	5
Операционная деятельность						
1	Выручка без НДС	975	1300	1625	1950	2275
2	Полные текущие издержки, в том числе:	878	1060	1196	1306	1503
3	прямые материальные затраты	20	31	39	43	49
4	ФОТ основных рабочих, включая взносы во внебюджетные фонды	351	420	476	512	598
5	Силовая энергия	91	114	135	147	164
6	Общепроизводственные расходы	247	285	316	335	400
7	Общехозяйственные расходы	150	195	214	247	269
8	Коммерческие расходы	30	35	39	46	50
9	Прочие расходы	9	11	16	19	22
10	Условно переменные издержки	492	600	689	748	861

11	Уровень безубыточности УБt	0,80	0,66	0,54	0,46	0,45

Расчет точки безубыточности:

Точка безубыточности показывает сколько продукции надо продать, чтобы доходы покрывали расходы:

$$Q = \frac{FC}{P - AVC} \quad (46)$$

где:

Q – точка безубыточности;

FC – постоянные затраты;

P – цена единицы товара;

AVC – переменные затраты.

$$Q = \frac{878}{427} = 2{,}06$$

Таким образом, нужно продавать, как минимум, 3 готовых продукта в год, чтобы получить доход, который покроет все расходы.

Выручка считается с тем учетом, что лицензия за один программный продукт составляет 65 тыс. руб. В таблице 26 представлен план денежных потоков

Таблица 26. План денежных потоков

№	Показатель, тыс. руб	Номер шага (периода)расчета (t)					
		0	1	2	3	4	5
Операционная деятельность							
1	Выручка без НДС	0	975	1300	1625	1950	2275
2	Полные текущие издержки, в том числе:	0	-878	-1060	-1196	-1306	-1503
3	прямые материальные затраты	0	-20	-31	-39	-43	-49
4	ФОТ основных рабочих, включая взносы во внебюджетные фонды	0	-351	-420	-476	-512	-598
5	Силовая энергия	0	-91	-114	-135	-147	-164
6	Общепроизводственные расходы	0	-247	-285	-316	-335	-400
7	Общехозяйственные расходы	0	-150	-195	-214	-247	-269

8	Коммерческие расходы	0	-30	-35	-39	-46	-50
9	Прочие расходы	0	-9	-11	-16	-19	-22
10	Денежный поток от производственной (операционной) деятельности (п.1п.2)	0	97	240	429	644	772
Инвестиционная деятельность							
11	Поступление инвестиций	0	0	0	0	0	0
12	Капиталовложения, обслуживание инвестиций	-413	-190	0	0	0	0
13	Сальдо от инвестиционной деятельности (п.11+п.12)	413	190	0	0	0	0
14	Сальдо суммарного потока (п.10+п.13)	-413	-93	240	429	644	772
15	Сальдо накопленного потока	413	506	266	163	807	1579
16	Коэффициент дисконтирования при ставке дохода 10%	1	0,909	0,826	0,751	0,683	0,621
17	Дисконтированное сальдо суммарного потока (стр.14*.стр.16)	-413	-84,54	198,24	322,18	439,85	479,41
18	Дисконтированные инвестиции (стр.12*стр.16)	-413	-172,7	0	0	0	0

Чистым доходом (ЧД) называется накопленный эффект (сальдо денежного потока) за расчетный период, где суммирование распространяется на все шаги расчетного периода. ЧД рассчитываемого проекта равен 1579 тысяч рублей.

Другим важнейшим показателем эффективности проекта является чистый дисконтированный доход (ЧДД) накопленный дисконтированный эффект за расчетный период. ЧДД проекта за пять лет при норме дисконта E=10%, приводя поток к шагу 0 (t_0=0). Чистый дисконтированный доход определяется суммированием строки 17: ЧДД = 942,14 тысяч рублей. Следовательно, проект является эффективным.

Внутренняя норма доходности (ВНД) характеризует внутреннюю норму рентабельности инвестиционного проекта. Внутренней нормой

доходности называется такое положительное число $E_{внутр}$, при котором в случае установления нормы дисконта равняющейся $E_{внутр}$, чистый дисконтированный доход проекта обращается в 0. В результате ВНД=32%. Это еще раз подтверждает эффективность проекта, так как ВНД>Е.

Потребность в дополнительном финансировании (ПФ) – максимальное значение абсолютной величины отрицательного накопленного сальдо от инвестиционной и производственной (операционной) деятельности. Величина ПФ показывает минимальный объем внешнего финансирования проекта, необходимый для обеспечения его финансовой реализуемости. Потребность в финансировании рассматриваемого проекта равна 413, тысяч рублей.

Сроком окупаемости (payback period) называется продолжительность периода от момента разработки проекта до момента окупаемости. Он определяется путем сопоставления произведенных капитальных вложений с величиной доходов от реализации проекта. Количество времени от начала периода до момента окупаемости определяется по формуле:

$$t = \frac{-C_{н2}}{C_{сум3}} \quad (47)$$

Где $C_{сум3}$ сальдо суммарного потока 4 периода,

$C_{н2}$ сальдо накопленного потока 3 периода.

Срок окупаемости, отсчитанный от начала нулевого периода, составляет 3,62 года, если же отсчитывать его от начала операционной деятельности, он окажется равным 2,62 года.

При оценке эффективности инвестиционных проектов так же часто используются индекс доходности инвестиций (ИД) отношение суммы доходов от производственной (операционной) деятельности к абсолютной величине капитальных вложений. Он равен увеличенному на единицу отношению ЧД к накопленному объему инвестиций.

ИД проекта =1+ (1579/ 603) = 3,62

Для определения дисконтированного индекса доходности (ИДД) необходимо найти сумму дисконтированных инвестиций (ДК). ДК=865,95 тыс.руб. Тогда

ИДД =1+ (942,14 / 585,70)= 2,61

Из проведенных расчетов можно заключить, что устойчивость проекта особых подозрений не вызывает.

Вывод

В экономической части данной работы были определены основные и требуемые технико-экономические показатели, такие как количество исполнителей, сроки исполнения, научно–технический эффект и эффективность проект, а так же была использована технология SWOT – анализ. В результате проведённых исследований можно сделать выводы о необходимости разработки данного проекта для осуществления финансовой эффективности научно – исследовательской деятельности.

Заключение

Результатом выполнения диссертационной работы стала разработанная информационная система, основанная на реализации алгоритма нечеткой логики. Данная ИС использует инструментальные средства в виде совокупности методов для анализа их многомерных топологических свойств и особенностей для выработки однозначного медицинского решения.

Первоначально была исследована предметная область, детально рассмотрено направление в развитии логического вывода знаний основанного на прецедентах. Были рассмотрены основные типы метрик, которые могут быть использованы в задачах поиска близких прецедентов. Выбор ключевых метрик основан на анализе применения метрик в данной предметной области. Были определены достоинства и недостатки каждой из них. В ходе анализа методов был сделан вывод о необходимости в разработке новых комплексных способов, направленных на решение задачи диагностики состояния пациента.

Была спроектирована база данных на MS SQL. Основой для базы стали результаты лабораторных и инструментальных исследований проводимых при выявлении патологий сердечно-сосудистой системы.

Итогом применения разработанного алгоритма обработки медицинских данных стала разработанная гибридная модель медицинской информационно-диагностической системы, требующего минимальной вычислительной мощности и минимальных временных затрат, так же позволяющая вырабатывать однозначное управляющее медицинское решение и повысить точность диагноза до 90%.

Список литературы

1. Вагин В.Н., Еремеев А.П. Некоторые базовые принципы построения интеллектуальных систем поддержки принятия решений реального времени // Известия РАН. Теория и системы управления. 2001. № 6. – С. 114-123.

2. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. 2-е издание // Под редакцией В.Н. Вагина, Д.А. Поспелова. – М.: ФИЗМАТЛИТ, 2008. – 712 с. 4.

3. Финн В.К. Об интеллектуальном анализе данных // Новости искусственного интеллекта. – 2004. – №3. – С. 3–18.

4. Карпов Л.Е., Юдин В.Н. Методы добычи данных при построении локальной метрики в системах вывода по прецедентам, М.: ИСП РАН, препринт № 18, 2006.

5. Осипов Г.С. Методы искусственного интеллекта М.: ФИЗМАТЛИТ, 2011-296с..

6. Информационно-аналитический материал «Актуальные вопросы информатизации при реализации программ модернизации здравоохранения субъектов Российской Федерации» [Электронный ресурс]. – Режим доступа: http://federalbook.ru/

7. Ефремова О.А. Редкие болезни, симптомы и синдромы в кардиологической практике: учеб. пособие / О.А. Ефремова, С.И. Логвиненко, Е.П. Погурельская, М.А. Руднева. – Белгород:ИПК НИУ БелГУ, 2011. – 132 с

8. Петри А. Наглядная статистика в медицине / А. Петри, К. Сэбин: пер. с англ. – М.: ГЭОТАР-МЕД, 2003. – 144 с

9. Пат. 55266 Российская Федерация, МПК A61B5/0402. Комплекс для экспресс диагностики сердца / А.С. Сулла, М.П. Рева, Д.А. Прилуцкий, Д.В. Архиреев; заявитель и патентообладатель ООО «Медицинские

Компьютерные Системы» (RU) – №2006106644/22, заявл. 06.03.2006; опубл. 10.08.2006.

10. Вапник В.Н. Теория распознавания образов (статистические проблемы обучения) / В.Н. Вапник, А.Я. Червоненкис. – М.: «Наука», 1974. – 416 с

11. Лепский А.Е. Математические методы распознавания образов: Курс лекций / А.Е. Лепский, А.Г. Броневич. – Т.: Изд-во ТТИ ЮФУ, 2009. – 155 с.

12. Заде Л. Основы нового подхода к анализу сложных систем и процессов принятия решений / Л. Заде: пер. с англ. // Математика сегодня. – 1974. – №7. – С. 5–49.

13. Макеева А.В. Основы нечеткой логики. Учебное пособие для вузов / А.В. Макеева. – Н.Новгород:ВГИПУ, 2009. – 59 с

14. Рыбин В.В. Основы теории нечетких множеств и нечеткой логики: учеб. пособие / В.В. Рыбин – М.: Изд-во МАИ, 2007. – 96 с

15. Каид В.А.. Методы построения функций принадлежности нечетких множеств / В.А.А. Каид // Известия ЮФУ. Технические науки.– 2013. – №2 (139). – С. 144–153

16. Гайдес, М. А. Общая теория систем (системы и системный анализ) / М. А. Гайдес. – Винница: Глобус-пресс, 2005. – 201 с.

17. Криницына З.В., Видяев И.Г. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение: Учебное пособие. – Томск: ТПУ, 2014. – 73с.