

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет (институт) _____ Кафедра _____
ЛФ МИ _____ Медицинская Кибернетика и Информатика _____
(наименование) (наименование)

Направление подготовки (специальность) 30.05.03 Медицинская кибернетика
(код и наименование)

ДИПЛОМНАЯ РАБОТА

на тему

Информационная система дифференциальной диагностики болезней почек

Студент Неурова 29.05.18 Неурова Юлия Владиславовна
(подпись, дата) (ф.и.о. полностью)

Руководитель/научный руководитель Гаврилов 30.05.18 В.П. Гаврилов
(подпись, дата) (фамилия, инициалы)

Консультанты (если имеются):

по _____
(наименование раздела) (подпись, дата) (фамилия, инициалы)

по _____
(наименование раздела) (подпись, дата) (фамилия, инициалы)

по _____
(наименование раздела) (подпись, дата) (фамилия, инициалы)

Нормоконтролёр _____ Алексеев 01.06.18 Н.Н. Якина
(подпись, дата) (фамилия, инициалы)

Рецензент инженер-фриланс ГБУЗ "ООО" Б.А. 13.06.18 С.С. Бухин
(должность, место работы) (подпись, дата) (фамилия, инициалы)

Работа допущена к защите (протокол заседания кафедры от 25.05.18 № 1)

Заведующий кафедрой МКИИ _____ Геращенко Р.А.
(подпись) (фамилия, инициалы)

Работа защищена с отметкой отлично (протокол заседания ГЭК от 22.06 № 19)
Секретарь ГЭК _____ Кривонозова С.В.
(подпись) (фамилия, инициалы)

Пенза 2018

«Утверждаю»

Заведующий кафедрой

«26» февраля 2018 г.

ЗАДАНИЕ на дипломное проектирование

1. Студент гр. 14.01 факультета лечебного специальности 30.05.03
Нусупова Юлия Владимировна
(фамилия, имя, отчество)
2. Руководитель проекта Рандеев Владимир Петрович
3. Время дипломного проектирования с 26 февраля 2018 г. по 5 июня 2018 г.
4. Место преддипломной практики МИТБУ кафедры «МКиУ»
5. Тема проекта Информационная система распознавания
болезней ног

Тема утверждена приказом ПГУ № 156 а/о от «01» марта 2018 г.

6. Техническое задание на проект (назначение устройства, условия применения, внешние воздействия, специальные требования и т. п.)

1. Информационная система предназначена для внедрения в диагностических центрах и в учебный процесс медицинского института.
2. Наименование распознаваемых заболеваний: нейропатии, телоневриты.
3. Метод распознавания заболеваний: статистический метод, метод Байеса.
4. Состав информационной системы: база данных пациентов диагностического центра, комплект программ распознавания болезней ног.
5. Система управления базой данных: Microsoft Access 2007.
6. Язык программирования: MATLAB.

7. Объем и содержание основной части проекта

7.1. Пояснительная записка (перечень вопросов, подлежащих разработке, расчетов, обоснований, описаний)

- Введение
1. Классификация заболеваний ног

2. Информационная система распознавания болезней почек

3. Обработка статистических данных результатов обследования

4. Методика распознавания болезней почек

Заключение

7.2. Графическая часть (перечень и содержание чертежей, плакатов)

1. Обоснование актуальности выбранной темы

2. Структура информационной системы распознавания болезней почек

3. Качественные признаки, заложенные в диагностическую модель

4. База данных

5. Методика распознавания заболеваний почек методом Байеса

6. Результаты работы

8. Консультанты и содержание дополнительных разделов (указывается конкретное содержание задания)

8.1. По конструкции устройства не копируется

Консультант _____

8.2. По технологии не копируется

Консультант _____

8.3. По экономике и организации производства не копируется

Консультант _____

8.4. По безопасности жизнедеятельности не планируется

Консультант _____

9. Календарный график по выполнению проекта

Наименование этапов работы	Объем работы	Срок выполнения	Подпись руководителя, консультанта
1. Введение	5%	26.06.18-10.03.18	
2. Диагностика учебного процесса	15%	11.03.18-05.04.18	
3. Информационная система распознавания багуней	15%	06.04.18-25.04.18	
4. Обработка статистических данных результатов обследований	25%	26.04.18-15.05.18	
5. Методика распознавания багуней	25%	16.05.18-25.05.18	
6. Заключение	5%	26.05.18-01.06.18	
7. Оформление ТЗ	5%	02.06.18-07.06.18	
8. Подготовка доклада и презентация	5%	08.06.18-21.06.18	

Примечание. Задание по п. 8 должно быть согласовано с руководителем проекта. В случае, если по отдельным дополнительным разделам (см. п. 8) консультанты не назначаются, необходимые задания на проектирование выдает руководитель проекта и вносит в пп. 8.1-8.4.

Дата выдачи задания «26» февраля 2018 г.

Руководитель дипломного проекта № ПГУ 30.05.03-12 ДР 12.05 ТЗ

Задание к исполнению принял «26» февраля 2018 г.

Дипломный проект к защите допустить

Декан факультета

Директор института

20__ г.

РЕФЕРАТ

Пояснительная записка к дипломной работе на тему «Информационная система дифференциальной диагностики болезней почек» содержит 98 листов, 61 рисунок, 30 таблиц, 28 источников информации, 1 приложение.

БОЛЕЗНИ ПОЧЕК, ОСТРЫЙ ГЛОМЕРУЛОНЕФРИТ, ОСТРЫЙ ПИЕЛОНЕФРИТ, РАСПОЗНАВАНИЕ ЗАБОЛЕВАНИЙ, ИНФОРМАЦИОННАЯ ДИАГНОСТИЧЕСКАЯ СИСТЕМА, БАЗА ДАННЫХ, СТАТИСТИЧЕСКИЙ АНАЛИЗ, ДИАГНОСТИЧЕСКАЯ МОДЕЛЬ, МЕТОД БАЙЕСА

Объект исследования: информационные технологии и системы распознавания болезней почек.

Цель работы: разработка элементов информационной системы дифференциальной диагностики болезней почек на основе статистических методов, обеспечивающих увеличение достоверности, снижение трудоемкости постановки диагноза и помощь специалисту в постановке правильного диагноза.

Назначение: создание модели базы данных пациентов с болезнями почек, диагностической модели пиелонефрита и гломерулонефрита применительно к периоду клинических проявлений, алгоритма и программного комплекса распознавания заболеваний почек по методу Байеса.

Область применения: информационная система предназначена для внедрения в диагностических центрах и в учебный процесс медицинского института.

					ПГУ 30.05.03 – 12ДР122.05 ПЗ				
Изм.	Лист	№ докум.	Подпись	Дата	Информационная система дифференциальной диагностики болезней почек	Лит.	Лист	Листов	
Разраб.		Юсупова Ю.В.	<i>Юсупова Ю.В.</i>	29.05.18					
Провер.		Фандеев В.П.	<i>Фандеев В.П.</i>	30.05.18			4	98	
Реценз.		Буянин С.С.	<i>Буянин С.С.</i>	30.05.18		Кафедра МКиИ, гр.12ЛК1			
Н. Контр.		Янкина Н.Н.	<i>Янкина Н.Н.</i>	01.06.18					
Утверд.		Янкина Н.Н.	<i>Янкина Н.Н.</i>	01.06.18					

СОДЕРЖАНИЕ

	с.
ВВЕДЕНИЕ.....	3
1 ДИАГНОСТИКА БОЛЕЗНЕЙ ПОЧЕК	6
1.1 Эпидемиология болезней почек	6
1.2 Этиология болезней почек	10
1.3 Патогенез болезней почек	12
1.4 Диагностические признаки болезней.....	16
1.5 Методика распознавания болезней почек (обоснование потребности в информационной системе).....	21
1.6 Основные результаты и выводы по разделу 1	23
2 ИНФОРМАЦИОННАЯ СИСТЕМА РАСПОЗНАВАНИЯ БОЛЕЗНЕЙ ПОЧЕК.....	25
2.1 Классификация и структура информационных систем распознавания болезней.....	25
2.2 Информационные системы распознавания болезней (примеры)....	30
2.3 Структура информационной системы распознавания болезней почек	34
2.4 База данных пациентов с болезнями почек.....	36
2.5 Пользовательский интерфейс базы данных	41
2.6 Основные результаты и выводы по разделу 2	43
3 ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ ОБСЛЕДОВАНИЯ В EXCEL.....	46
3.1 Методика обработки статистических данных.....	46
3.2 Построение вариационных рядов значений признаков	49
3.3 Вычисление характеристик распределений	53

3.4 Расчёт корреляции значений признаков	61
3.5 Вычисление вероятностей интервалов значений признаков.....	62
3.6 Основные результаты и выводы по разделу 3	65
4 МЕТОДИКА РАСПОЗНАВАНИЯ БОЛЕЗНЕЙ ПОЧЕК.....	67
4.1 Методы распознавания болезней	67
4.2 Диагностические модели болезней почек	70
4.3 Программный комплекс формирования диагностической модели	73
4.4 Программный комплекс распознавания болезней почек.....	76
4.5 Интерфейс пользователя	79
4.6 Вычислительный эксперимент по распознаванию болезней почек	83
4.7 Основные результаты и выводы по разделу 4	85
ЗАКЛЮЧЕНИЕ	88
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	92
ПРИЛОЖЕНИЕ А. ПРОГРАММНЫЙ КОМПЛЕКС ФОРМИРОВАНИЯ ДИАГНОСТИЧЕСКОЙ МОДЕЛИ И РАСПОЗНОВАНИЯ БОЛЕЗНЕЙ ПОЧЕК МЕТОДОМ БАЙЕСА	95

ВВЕДЕНИЕ

Актуальность

Одной из значимых проблем современной медицины являются заболевания почек, которые относятся к группе болезней мочеполовой системы. По данным Росстата заболеваемость болезнями почек на 100000 человек населения Пензенской области в 2011 году составляло 337,9 человек, а в 2016 году показатель снизился до 255 человек [1]. Заболеваемость населения жителей Пензенской области гломерулярными и тубулоинтерстициальными болезнями почек за 2014 год составляла 522432 человек.

Следовательно, меры по борьбе с болезнями мочеполовой системы, принимаемые государством, дают свои результаты. Но, несмотря на тенденцию к снижению заболеваемости, на данный момент она остается на достаточно высоком уровне.

Заболевания почек имеют тяжелые осложнения со стороны сердечно - сосудистой системы, а также требуют дорогостоящих методов лечения и поддержания жизни пациента в терминальной стадии, таких как гемодиализ и пересадка почки.

Поэтому заболевания почек необходимо своевременно диагностировать и лечить.

Диагностика заболеваний почек является трудоёмким процессом, что в условиях массового обращения пациентов, требует привлечения к диагностике большого числа квалифицированных специалистов. Совокупность данных факторов не исключает возникновения диагностических ошибок. Снижение трудоёмкости и повышение достоверности постановки диагноза, достигается разработкой и внедрением информационных систем дифференциальной диагностики заболеваний.

Проблема

Снижение трудоёмкости постановки диагноза достигается за счет автоматизации принятия решений. Повышение достоверности диагноза достигается за счёт принятия решений на основе опыта высококвалифицированных специалистов. Данный опыт собран и хранится в базе данных информационной системы, содержащей результаты обследований и достоверные диагнозы пациентов.

На данный момент достаточно большое количество разработчиков занимается исследованием и созданием информационных систем. Однако результаты исследований не доведены до методик проектирования систем распознавания заболеваний, в части проектирования базы данных диагнозов и результатов обследования пациентов, формирования вероятностной диагностической модели на основе базы данных, статистических методов распознавания.

Объект исследования - информационные технологии и системы распознавания болезней почек.

Предмет исследования – модели, методы и алгоритмы статистического распознавания болезней почек.

Цель работы - разработка элементов информационной системы дифференциальной диагностики болезней почек на основе статистических методов, обеспечивающих увеличение достоверности, снижение трудоемкости постановки диагноза и помощь специалисту в постановке правильного диагноза.

Задачи:

- 1) произвести нозологический анализ заболеваний почек;
- 2) произвести анализ историй болезней пациентов с заболеваниями почек, систематизировать полученные данные и создать базу данных пациентов с болезнями почек;
- 3) произвести обработку полученных данных в Excel и сформировать достоверную диагностическую модель;

4) реализовать методику распознавания заболеваний с использованием статистического метода Байеса, разработать необходимые алгоритмы и создать программный комплекс распознавания заболеваний почек.

1 ДИАГНОСТИКА БОЛЕЗНЕЙ ПОЧЕК

1.1 Эпидемиология болезней почек

По Международной классификации болезней десятого пересмотра (МКБ 10) все заболевания почек относятся к обширной группе болезней мочеполовой системы (N00-N99).

Классификация болезней мочеполовой системы по МКБ 10 представлена в таблице 1.1.

Таблица 1.1 - Классификация болезней мочеполовой системы по МКБ 10

Классы МКБ 10	Название блока
N00-N08	Гломерулярные болезни
N10-N16	Тубулоинтерстициальные болезни
N17-N19	Почечная недостаточность
N20-N23	Мочекаменная болезнь
N25-N29	Другие болезни почки и мочеточника
N30-N39	Другие болезни мочевыделительной системы
N40-N51	Болезни мужских половых органов
N60-N64	Болезни молочной железы
N70-N77	Воспалительные болезни женских тазовых органов
N80-N98	Невоспалительные болезни женских половых органов
N99-N99	Другие нарушения мочеполовой системы

К болезням мочеполовой системы относятся пять групп заболеваний почек: гломерулярные болезни (N00-N08), тубулоинтерстициальные болезни (N10-N16), почечная недостаточность (N17-N19), мочекаменная болезнь (N20-N23), а также другие болезни почки и мочеточника (N25-N29). Гломерулонефрит относится к группе гломерулярных болезней (N00-N08), а пиелонефрит – к группе тубулоинтерстициальных болезней (N10-N16).

Восемь подразделов гломерулярных болезней и 6 подразделов тубулоинтерстициальных болезней представлены в таблице 1.2.

Таблица 1.2 - Классификация гломерулярных и тубулоинтерстициальных болезней по МКБ 10.

Гломерулярные болезни (N00-N08)	Тубулоинтерстициальные болезни (N10-N16)
N00 Острый нефритический синдром	N10 Острый тубулоинтерстициальный нефрит
N01 Быстро прогрессирующий нефритический синдром	N11 Хронический тубулоинтерстициальный нефрит
N02 Рецидивирующая и устойчивая гематурия	N12 Тубулоинтерстициальный нефрит, не уточненный как острый или хронический
N03 Хронический нефритический синдром	N13 Обструктивная уропатия и рефлюкс-уропатия
N04 Нефротический синдром	N14 Тубулоинтерстициальные и тубулярные поражения, вызванные лекарствами и тяжелыми металлами
N05 Нефритический синдром неуточненный	N15 Другие тубулоинтерстициальные болезни почек
N06 Изолированная протеинурия с уточненным морфологическим	N16 Тубулоинтерстициальные поражения почек, классифицируемые в других рубриках
N07 Наследственная нефропатия, не классифицированная в других рубриках	
N08 Гломерулярные поражения при болезнях, классифицированных в других рубриках	

Заболеваемость населения Пензенской области болезнями почек представлена в таблице 1.3 и на рисунке 1.1[2].

Таблица 1.3 - Заболеваемость болезнями почек населения Пензенской области на 100000 человек

Год	2011	2012	2013	2014	2015	2016
На 100000 человек населения	337,9	318,4	340,1	273,1	320,4	255



Рисунок 1.1 – Заболеваемость болезнями почек жителей Пензенской области на 100000 человек населения

По графику видно, что заболеваемость болезнями почек с 2011 по 2016 год идет на спад. Следовательно, меры борьбы государства с болезнями почек дают результаты, но заболеваемость все еще остается на достаточно высоком уровне.

В нашей работе мы будем рассматривать острые формы пиелонефрита (тубулоинтерстициальные болезни, блок N10) и гломерулонефрита (гломерулярные болезни, блок N00).

Пиелонефрит является самым частым заболеванием среди болезней почек и достаточно распространенным заболеванием среди болезней с инфекционной этиологией. На основании данных о вскрытии пиелонефрит находят у каждого 10-12 пациента, не достигшего пожилого возраста и у каждого 5 пациента пожилого возраста. Практически в каждом 4 случае диагностируется острая форма пиелонефрита [3].

Заболеваемость острым пиелонефритом на 10000 населения составляет 15,7 случаев. Чаще пиелонефритом заболевают женщины. Это обусловлено коротким мочеиспускательным каналом, близким расположением прямой

кишки и половых путей. Так же влияние оказывает гормональный фон и прием пероральных контрацептивов.

Заболеваемость пиелонефритом имеет три возрастных пика (таблица 1.4) [4].

Таблица 1.4 - Три возрастных пика заболеваемости пиелонефритом

Возрастной период	Заболеваемость по возрастному периоду
Ранний детский возраст (до 3 лет)	Отмечается значительное преобладание заболеваемости среди девочек (8:1)
Активный репродуктивный возраст (18 – 35 лет)	Сохраняется преобладание заболеваемости у женщин (7:1)
Пожилой и старческий возраст	С 60 лет соотношение заболеваемости у мужчин и у женщин выравнивается, а после 70 лет преобладает у лиц мужского пола. Это связано с нарушением уродинамики, которое возникает из-за гипертрофических и опухолевых процессов в предстательной железе.

Гломерулонефрит среди заболеваний почек занимает третье место, является наиболее частой причиной развития хронической почечной недостаточности и выявляется чаще у людей моложе 40 лет. Гломерулонефритом страдают около 10- 15 человек на 100000 населения России.

Так как заболеваемость острым пиелонефритом составляет 15,7 на 10000 населения в год, а острым гломерулонефритом 10 - 15 на 10000 населения, то исходя из эпидемиологических данных о распространенности, априорная вероятность составляет 0,157 для острого пиелонефрита и от 0,1 до 0,15 для острого гломерулонефрита. Данные значения являются необходимыми для реализации методики минимизации рисков диагностической ошибки и построения диагностической модели.

Таким образом, несмотря на тенденцию к снижению заболеваемости населения болезнями почек, на данный момент она остается на достаточно высоком уровне (таблица 1.2). Большая распространенность этого заболевания

приводит к увеличению нагрузки на медицинский персонал, за счет массового обращения пациентов. Исходя из эпидемиологических данных о распространенности заболеваний почек, априорная вероятность для острого пиелонефрита составляет 0,157 и от 0,1 до 0,15 для острого гломерулонефрита.

Для разработки информационной системы дифференциальной диагностики болезней почек, необходимо выявление основных признаков заболевания, на определенной стадии развития болезни. Для этого нужно разобраться в этиопатогенезе заболеваний почек.

1.2 Этиология болезней почек

Пиелонефрит – неспецифическое инфекционное воспалительное заболевание почек, которое поражает паренхиму почки, лоханку и чашечки.

Наиболее частыми возбудителями пиелонефрита являются микроорганизмы, представляющие собой собственную микрофлору кишечника человека – ее облигатную или факультативную часть (таблица 1.5) [3].

Таблица 1.5 - Частота высеваемости возбудителей пиелонефрита (в процентах)

Возбудитель	Частота высеваемости (%)	
	Средняя	Разброс
Кишечная палочка	39,8	27,2 – 68
Энтерококки	21,1	4,6 – 36,7
Протей	16,3	6,5 – 41,6
Синегнойная палочка	10,6	2,1 – 18,7
Стафилококки	7,6	5,1 – 11,5
Клебсиелла	3,2	2,3 – 8
Ассоциации возбудителей	26,9	4,1 – 45,5

В качестве возбудителей госпитальной инфекции чаще всего выступают: *Escherichia coli*, виды *Klebsiella*, *Proteus*, *Staphylococcus aureus*, *Candida*. При гнойных поражениях почек чаще всего обнаруживают *Staphylococcus aureus*.

У части больных при посевах мочи выявить возбудитель не удастся, что не всегда свидетельствует о ремиссии заболевания. Микроорганизмы могут трансформироваться в нестабильные L-формы, которые при благоприятных для бактерии условиях претерпевают обратную трансформацию и продолжают поддерживать воспалительный процесс [4].

Гломерулонефрит – иммуновоспалительное заболевание с преимущественным поражением почечных клубочков. В 80 – 90% случаев острой формы гломерулонефрита удастся установить этиологический фактор. В остальных случаях причина остается неизвестной (таблица 1.6).

Таблица 1.6 - Основные этиологические факторы развития гломерулонефрита

Этиологический фактор	Примеры
Бактериальная инфекция	Стрептококковая, стафилококковая, туберкулез, малярия, сифилис
Вирусная инфекция	Гепатит В и С, герпесвирусы, ВИЧ
Токсические вещества и лекарственные средства	Алкоголь, наркотические вещества, органические растворители, соединения ртути, свинца
Эндогенные и экзогенные антигены	Опухолевые, ДНК, мочева́я кислота

Наиболее частой и наиболее изученной причиной развития гломерулонефрита считается стрептококковая инфекция [4]. Острый постстрептококковый гломерулонефрит является следствием перенесенной инфекции или развивается на ее фоне. Возбудителем чаще всего является β -гемолитический стрептококк группы А.

Таким образом, основным этиологическим фактором развития пиелонефрита и гломерулонефрита являются микроорганизмы. Для пиелонефрита – это представители собственной микрофлоры кишечника, чаще всего *Escherichia coli* (таблица 1.5). Наиболее частой причиной развития гломерулонефрита является стрептококковая инфекция, а возбудителем - β -гемолитический стрептококк группы А (таблица 1.6). Зная причину

возникновения заболевания можно предположить механизмы его развития в организме человека и клинические проявления.

1.3 Патогенез болезней почек

Пиелонефрит – инфекционное заболевание.

Инфицирование почек микроорганизмами может происходить несколькими путями (рисунок 1.2).

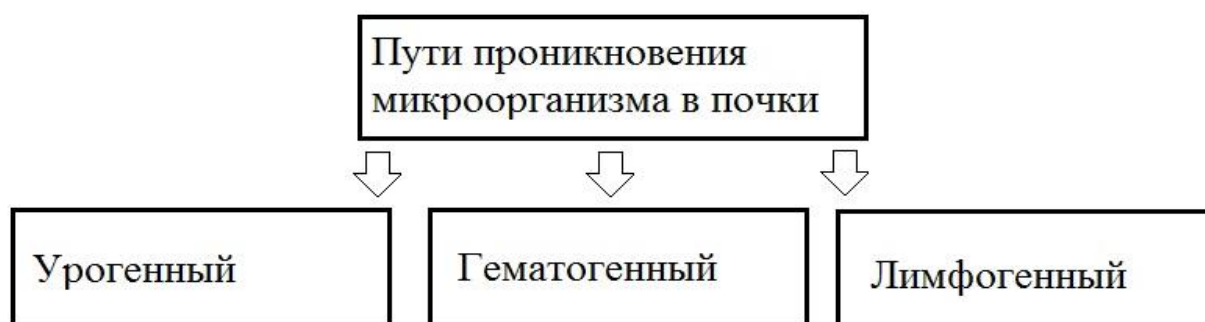


Рисунок 1.2 – Пути инфицирования почек микроорганизмами

При урогенном пути инфицирования возбудитель попадает в ткани почек из мочеиспускательного канала (восходящим путем). При гематогенном пути инфицирование происходит вторичным образом из уже имеющегося в организме человека источника инфекции. Лимфогенный путь достаточно редкий, например, при ретроперитонеальном абсцессе или кишечной инфекции [4].

Однако, просто инфицирования микроорганизмами недостаточно для развития воспалительного процесса. Наиболее эффективно воспалительный процесс реализуется при одновременном проявлении микроорганизмом вирулетных свойств, неадекватного иммунного ответа организма, нарушении уродинамики или внутривисочечной гемодинамики [3].

Некоторые факторы способствующие развитию пиелонефрита приведены в таблице 1.7.

Таблица 1.7 - Факторы, способствующие развитию пиелонефрита

Фактор	Описание фактора
Обструкция мочевых путей	Нарушение выделения мочи, ведущее к появлению остаточной мочи в мочевом пузыре. Слизистая растягивается, кровообращение в слизистой оболочке нарушается. Снижается выработка факторов защиты слизистой, что приводит к облегчению ее инфицирования
Инструментальные исследования мочевых путей (катетеризация)	Прямой занос микроорганизмов на слизистую оболочку
Беременность	Косвенное влияние за счет гиперэстрогемии и расстройства цистидной функции мочеточников. Прямое влияние за счет сдавления мочевых путей увеличенной маткой
Сахарный диабет	Сопровождается нарушением почечного кровотока, гломерулосклерозом, нейрогенной дисфункцией мочевого пузыря
Неинфекционная патология почек	Например, гломерулонефрит. Вызывает также нарушение почечного кровотока и развитие гломерулосклероза
Иммуносупрессивная терапия	Изменения иммунитета и неспецифических факторов защиты мочевых путей и в целом всего организма

Классифицируют пиелонефрит по локализации, по характеру течения и по форме (рисунок 1.3).



Рисунок 1.3 – Классификация пиелонефрита по локализации, характеру течения и форме

Развитие заболеваний болезней почек происходит в три периода: период максимальной активности, период стихания, период ремиссии [13].

Острый пиелонефрит характеризуется переходом активной стадии заболевания в стадию ремиссии менее чем за шесть месяцев. При остром пиелонефрите микроорганизмы, попавшие в почку, оседают на почечных сосудах. Это приводит к развитию местной воспалительной реакции и притоку к месту воспаления лейкоцитов, а также дегенеративных процессов в эндотелии сосудов, что приводит к нарушению их проницаемости. Затем микроорганизмы и лейкоциты проникают в просвет почечных канальцев и выводятся с мочой, приводя к появлению бактериурии и лейкоцитурии.

В итоге, одним из первых признаков развития острого пиелонефрита служит появление бактериурии и лейкоцитурии. Поражение почечных сосудов и нарушение их проницаемости приводит к последующему появлению в моче пациента небольшого количества белка и эритроцитов, а также развития реакции организма на воспаление – лихорадки.

Гломерулонефрит – иммуновоспалительное заболевание.

Этиологический фактор, вызывающий развитие острого гломерулонефрита, может быть различным, но наиболее частой и наиболее изученной причиной возникновения этого заболевания является перенесенная стрептококковая инфекция (скарлатина, ангина, невмония). Основным возбудителем является β -гемолитического стрептококк группы А (12 штамм, называемый по его тропности к мембране клубочков нефритогенным) [6].

Классификация гломерулонефритов основывается на выделении ведущего синдрома, варианта течения или морфологических изменений (таблица 1.8).

Таблица 1.8 - Классификация гломерулонефритов на основе выделения ведущего синдрома, варианта течения и морфологических изменений

Ведущий синдром	Варианты течения	Морфологические изменения
Гематурический Гипертонический Нефротический	Острый Хронический Быстропрогрессирующий	Пролиферативный Непролиферативный

Рассмотрим патогенез гломерулонефрита на примере стрептококковой инфекции (рисунок 1.4) [5].



Рисунок 1.4 – Патогенез гломерулонефрита на примере стрептококковой инфекции

В патогенезе гломерулонефрита выделяют два основных компонента – иммунный и неиммунный. Иммунный компонент включает в себя иммунный ответ, иммунное воспаление, склеротические изменения в почках. Неиммунный компонент – это гемодинамические и метаболические механизмы.

Механизмы прогрессирования гломерулонефрита связаны как с поражением клубочкового аппарата почек, так и с поражением их интерстициального пространства. Повреждение канальцев почек в результате длительной протеинурии является дополнительным фактором аутоагрессии [5].

При остром гломерулонефрите происходит поражение почечных клубочков иммунными комплексами. Это приводит к увеличению проницаемости почечного фильтра и появлению в моче белка, эритроцитов и лейкоцитов. Нарушение клубочковой фильтрации приводит к снижению азотовыделительной функции почек и накоплению в крови креатинина.

Все приведенные механизмы приводят к развитию нефросклероза и хронической почечной недостаточности [4].

Таким образом, в основе патогенеза пиелонефрита и гломерулонефрита лежит воспалительная реакция. Развитие заболеваний болезней почек происходит в три периода: период максимальной активности, период стихания, период ремиссии. Острые формы обоих заболеваний характеризуются увеличением проницаемости почечного фильтра и появлением в моче лейкоцитов, эритроцитов и белка в период максимальной активности. Нарушение почечной работы фильтра также способствуют накоплению в крови азотистых шлаков, в том числе и креатинина. Заболевания почек имеют тяжелые осложнения, которые требуют дорогостоящих методов лечения и поддержания жизни пациента в терминальной стадии, таких как гемодиализ и пересадка почки. Поэтому заболевания почек необходимо своевременно диагностировать и лечить.

1.4 Диагностические признаки болезней

Обследование больного при заболеваниях почек начинается с расспроса пациента и изучения его жалоб.

При сборе анамнеза необходимо обратить внимание на следующие ситуации и состояния: переохлаждение, стрептококковые инфекции (ангины, скарлатины, пневмонии), вирусные инфекции, воздействие токсических веществ и лекарственных средств. Также необходимо обратить внимание на наличие беременности или сопутствующих заболеваний, таких как: сахарный диабет, мочекаменная болезнь, опухоли или гиперплазии предстательной

железы у мужчин. Все эти ситуации и состояния являются предрасполагающими факторами к развитию пиелонефрита и гломерулонефрита [6].

Большинство жалоб пациентов при пиелонефрите и гломерулонефрите достаточно неспецифичны – это слабость, лихорадка, артралгии и миалгии. Также присутствуют симптомы более специфичные для заболеваний мочевыделительной системы – это боли в пояснице, изменения в мочеотделении и отечность лица (таблица 1.9) [4].

Таблица 1.9 - Основные жалобы пациентов с острым пиелонефритом и гломерулонефритом

Пиелонефрит	Гломерулонефрит
Лихорадка, сопровождающаяся ознобом	
Слабость	
Артралгии и миалгии	
Тошнота и рвота	
Боли в пояснице	
Частые позывы к мочеиспусканию иногда болезненность при мочеиспускании	Уменьшение количества мочи или ее отсутствие
-	Припухлость век или лица по утрам
Головные боли	
-	Учащение сердцебиения
-	Изменение цвета мочи

Среди физических методов при заболеваниях почек используются осмотр, пальпация и перкуссия (таблица 1.10) [6].

Таблица 1.10 - Результаты физических методов обследования при остром пиелонефрите и гломерулонефрите

Заболевание	Пиелонефрит	Гломерулонефрит
Осмотр		Отеки. Лицо бледное, одутловатое, с припухшими веками
Пальпация	Болезненные ощущения в области почек	
Перкуссия	Болезненные ощущения в области почек	

Более информативными методами в диагностике болезней почек являются лабораторные и инструментальные методы. Инструментальные методы диагностики при остром пиелонефрите и гломерулонефрите представлены в таблице 1.11 [4].

Таблица 1.11 - Основные инструментальные методы диагностики острого пиелонефрита и гломерулонефрита

Метод	Пиелонефрит	Гломерулонефрит
Ультразвуковое исследование	При остром пиелонефрите плотность почек неравномерно снижается из-за очагового воспаления почечной ткани. Снижение подвижности почек сочетается с их увеличением. Процесс обычно односторонний. Возможно увеличение почки	Почки могут быть не изменены. Иногда наблюдается небольшое увеличение размеров, отечность паренхимы, признаки ишемии
Рентгенологические исследования	Экскреторная урография. Определение ограничения подвижности почек на стороне поражения	Рентгенография органов грудной клетки. Признаки застоя в малом круге кровообращения
Компьютерная томография	Дифференциальная диагностика с опухолевыми процессами. Уточнение особенностей почечных лоханок, паренхимы, сосудов, паранефральной клетчатки	
ЭКГ	Норма	Изменение положения электрической оси сердца, признаки перегрузки левых отделов сердца, легочной гипертензии
Биопсия	Обычно не проводится из-за очагового характера заболевания	Проводится для определения морфологического типа гломерулонефрита

Среди инструментальных методов наиболее распространенным для диагностики заболеваний почек является ультразвуковое исследование. Диагноз гломерулонефрита с морфологической точки зрения можно поставить только после проведения биопсии почки.

Лабораторные методы диагностики являются основными и наиболее информативными методами в диагностике острого пиелонефрита и гломерулонефрита (таблица 1.12) [4].

Таблица 1.12 - Основные лабораторные методы диагностики острого пиелонефрита и гломерулонефрита

Метод	Пиелонефрит	Гломерулонефрит
Общий анализ крови	Повышение СОЭ, лейкоцитоз со сдвигом лейкоцитарной формулы влево	Повышение СОЭ, снижение концентрации гемоглобина (разведение крови)
Биохимический анализ крови	Повышение С – реактивного белка, возможно повышение азотистых шлаков	Повышение концентрации азотистых шлаков (креатенина, мочевины, мочевой кислоты). Клубочковая фильтрация снижена
Иммунологические тесты	Норма	Повышение титров антител к антигенам стрептококка. Снижение общей активности комплемента и особенно концентрации С3-фракции
Общий анализ мочи	Моча имеет щелочную реакцию, лейкоцитурия, протеинурия, бактериурия	Протеинурия, гематурия, цилиндрурия, протеинурия. На начальных стадиях лейкоцитурия
Анализ мочи по Нечипоренко	Позволяет уточнить степень выраженности гематурии, цилиндрурии и лейкоцитурии	
Проба Зимницкого	Снижение относительной плотности мочи и преобладание ночного диуреза	На стадии выздоровления возможно снижение плотности мочи
Бактериологическое исследование	Позволяет выявить возбудитель заболевания и подобрать необходимый антибиотик	Норма

Так как общий анализ мочи входит в перечень исследований общей диагностики, позволяет быстро получить результаты, а изменение показателей этого анализа встречаются как при пиелонефрите, так и при гломерулонефрите, то в своей работе мы будем использовать именно общий анализ мочи, в частности, количество лейкоцитов, эритроцитов.

Гематурия часто встречается как при гломерулонефрите, так и при пиелонефрите, но при гломерулонефрите носит более выраженный и стойкий характер. Лейкоцитурия более выражена при пиелонефрите [6]. Резкое повышение креатинина более характерно для больных гломерулонефритом, у больных с пиелонефритом креатинин увеличивается не так резко.

Таблица 1.13 - Нормальные значения некоторых показателей общего анализа мочи и биохимического анализа крови

Показатель	Нормальные значения	При пиелонефрите	При гломерулонефрите
Лейкоциты	0 – 6 в поле зрения	Резко выражена лейкоцитурия	Менее выражена лейкоцитурия
Эритроциты	1 – 3 в поле зрения	Слабо выражена эритроцитурия	Более выражена эритроцитурия
Креатинин	Мужчины: 62 – 115 мкмоль/л. Женщины: 53 – 97 мкмоль/л	Менее выраженное повышение	Более выраженное повышение

При анализе литературы для данных показателей законы распределения и их характеристики выявлены не были. Следовательно, требуется провести анализ историй болезней пациентов с заболеваниями почек с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками.

Таким образом, основными методами диагностики острого пиелонефрита и гломерулонефрита являются лабораторные методы. Так как общий анализ мочи входит в перечень исследований общей диагностики и позволяет быстро получить результаты, то в создании информационной системы мы будем использовать именно общий анализ мочи, в частности, количество лейкоцитов и эритроцитов, а также уровень креатинина в крови, получаемый при помощи биохимического анализа крови (таблица 1.13) Эти показатели удобны для создания информационной системы, так как являются количественными признаками и появляются при обоих заболеваниях. При анализе литературы для данных показателей законы распределения и их характеристики выявлены не были. Следовательно, требуется провести анализ историй болезней пациентов с заболеваниями почек с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками.

1.5 Методика распознавания болезней почек (обоснование потребности в информационной системе)

Заболеваниями болезней почек занимается врач-нефролог или врач-уролог. Этапы дифференциальной диагностики заболеваний представлены в таблице 1.14 [7].

Таблица 1.14 – Этапы дифференциальной диагностики заболеваний

Название этапа	Описание этапа
Первый	Выделение одного или нескольких очевидных синдромов, подтвержденных субъективным и объективным обследованием
Второй	Формирование перечня заболеваний, при которых данный синдром встречается. Чем более образован и опытен врач, тем этот список больше
Третий	Проведение дополнительных обследований для исключения заболеваний из составленного списка
Четвертый	Начинают лечение наиболее вероятного заболевания
Аутопсия	Посмертное обследование тела пациента. Позволяет сформировать окончательное представление о болезни и выявить диагностические ошибки врача

Для врачей существуют нормативные документы, согласно которым врач осуществляет свою деятельность. Согласно приказу Министерства здравоохранения Российской Федерации от 19.12.2016 г. № 973н "Об утверждении типовых отраслевых норм времени на выполнение работ, связанных с посещением одним пациентом врача-кардиолога, врача-эндокринолога, врача-стоматолога-терапевта" норма времени для врача-терапевта 15 минут [24].

В условиях большого потока пациентов, время на прием каждого человека очень ограничено, что требует максимально быстро принять верное диагностическое решение.

Этапы распознавания болезней почек представлены на рисунке 1.5.

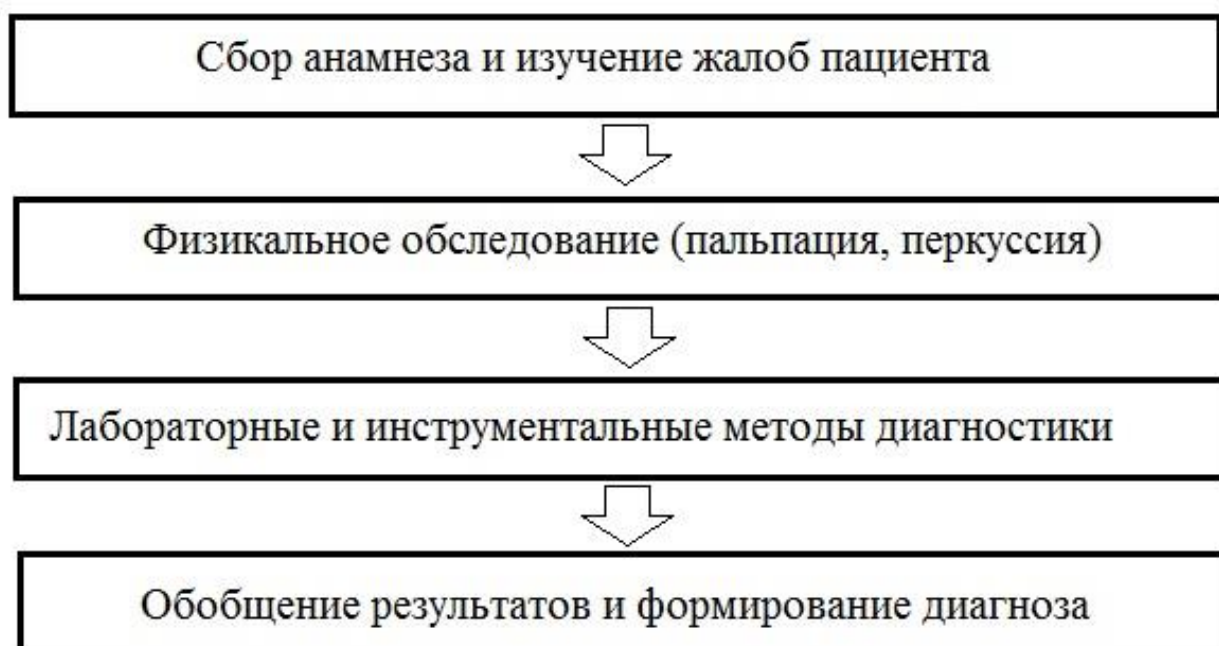


Рисунок 1.5 – Этапы распознавания заболеваний почек

Этап обобщения результатов и постановки диагноза основывается на клиническом мышлении врача. Клиническое мышление зависит от знаний, опыта и врачебной интуиции. Чем выше квалификация врача, тем больше вероятность постановки им правильного диагноза.

Таким образом, распознавание болезней почек является трудоемким процессом, который происходит в условиях дефицита времени и зависит от опыта и квалификации врача - специалиста. Все эти факторы не исключают возможности возникновения диагностических ошибок. Уменьшение трудоёмкости и увеличение достоверности постановки диагноза, может быть достигнуто путем разработки и внедрения информационных систем дифференциальной диагностики заболеваний.

1.6 Основные результаты и выводы по разделу 1

1 Несмотря на тенденцию к снижению заболеваемости населения болезнями почек, на данный момент она остается на достаточно высоком уровне (таблица 1.2). Большая распространенность этого заболевания приводит к увеличению нагрузки на медицинский персонал, за счет массового обращения пациентов. Исходя из эпидемиологических данных о распространенности заболеваний почек, априорная вероятность острого пиелонефрита составляет 0,157 и от 0,1 до 0,15 для острого гломерулонефрита.

2 Основным этиологическим фактором развития пиелонефрита и гломерулонефрита являются микроорганизмы. Для пиелонефрита – это представители собственной микрофлоры кишечника, чаще всего *Escherichia coli* (таблица 1.5). Наиболее частой причиной развития гломерулонефрита является стрептококковая инфекция, а возбудителем - β -гемолитический стрептококк группы А (таблица 1.6). Зная причину возникновения заболевания можно предположить механизмы его развития в организме человека и клинические проявления.

3 В основе патогенеза пиелонефрита и гломерулонефрита лежит воспалительная реакция. Развитие заболеваний болезней почек происходит в три периода: период максимальной активности, период стихания, период ремиссии. Острые формы обоих заболеваний характеризуются увеличением проницаемости почечного фильтра и появлением в моче лейкоцитов, эритроцитов и белка в период максимальной активности. Нарушение работы почечного фильтра также сопровождается накоплением в крови азотистых шлаков, в том числе и креатинина.

4 Основными методами диагностики острого пиелонефрита и гломерулонефрита являются лабораторные методы. Так как общий анализ мочи входит в перечень исследований общей диагностики и позволяет быстро получить результаты, то в создании информационной системы мы будем использовать именно общий анализ мочи, в частности, количество

лейкоцитов, эритроцитов и белка (таблица 1.13). Эти показатели удобны для создания информационной системы, так как являются количественными признаками и появляются при обоих заболеваниях.

5 При анализе литературы для данных показателей законы распределения и их характеристики выявлены не были. Следовательно, требуется провести анализ историй болезней пациентов с заболеваниями почек с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками.

6 Распознавание болезней почек является трудоемким процессом, который происходит в условиях дефицита времени и зависит от опыта и квалификации врача - специалиста. Все эти факторы не исключают возможности возникновения диагностических ошибок. Уменьшение трудоёмкости и увеличение достоверности постановки диагноза, может быть достигнуто путем разработки и внедрения информационных систем дифференциальной диагностики заболеваний.

2 ИНФОРМАЦИОННАЯ СИСТЕМА РАСПОЗНАВАНИЯ БОЛЕЗНЕЙ ПОЧЕК

2.1 Классификация и структура информационных систем распознавания болезней

Информационная система — система, предназначенная для хранения, поиска и обработки информации, и соответствующие организационные ресурсы (человеческие, технические, финансовые и т. д.), которые обеспечивают и распространяют информацию.

Информационные системы могут быть использованы в различных областях – геологии, химии, математике, здравоохранении.

В нашей стране существует стандарт организации СТО МОСЗ 91500.16.0002–2004 «Информационные системы в здравоохранении. Общие требования» [14]. В данном стандарте вводится следующая функциональная классификация информационных систем в здравоохранении (рисунок 2.1).

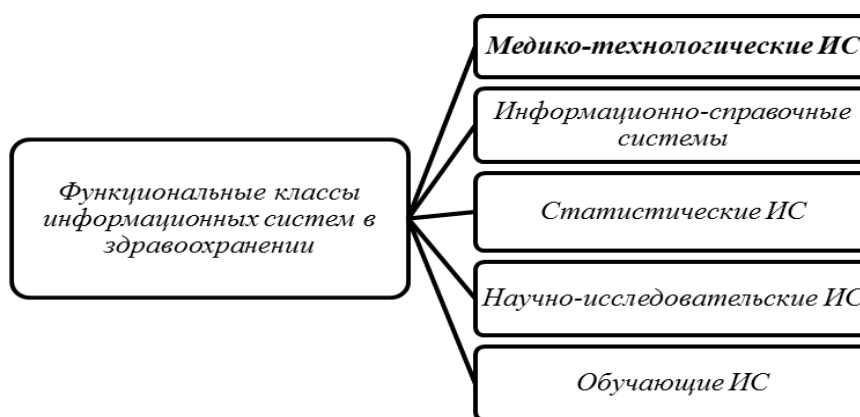


Рисунок 2.1 – Функциональные классы информационных систем
в здравоохранении

Медико-технологические ИС в зависимости от объектов описания, решаемых социальных задач, потребностей пользователей и степени агрегирования выходной информации подразделяются на виды (рисунок 2.2).

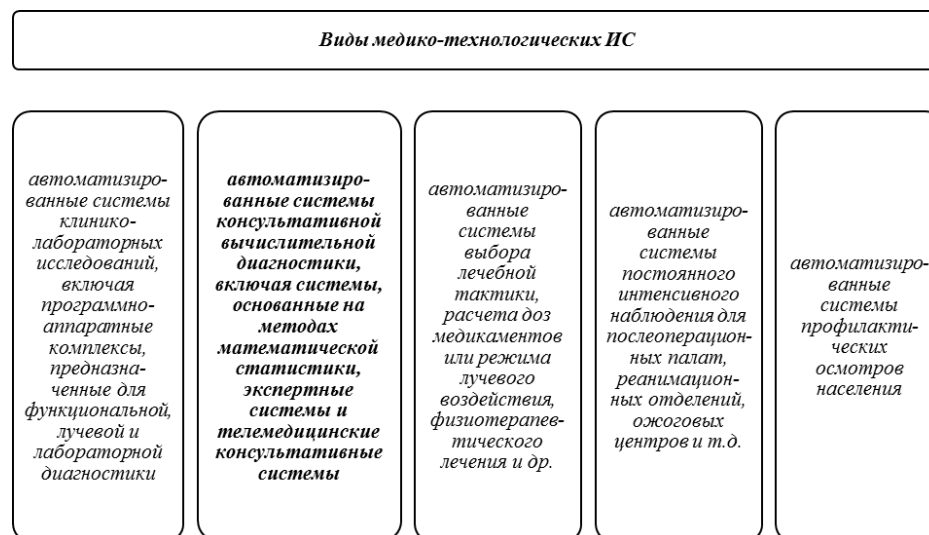


Рисунок 2.2 – Виды медико-технологических информационных систем

Классификация медицинских информационных систем по иерархическому принципу представлена на рисунке 2.3 [9].



Рисунок 2.3 – Классификация медицинских информационных систем по иерархическому принципу

Информационная система дифференциальной диагностики болезней почек предназначена для помощи врачу – нефрологу, поэтому она относится к

информационным медицинским системам базового уровня. Общая классификация информационных систем представлена в таблице 2.1.

Таблица 2.1 — Классификация информационных систем

Классификационные признаки	Классификационные группировки
Наличие элементов искусственного интеллекта (например, базы знаний)	Системы искусственного интеллекта, в том числе экспертные системы Традиционные информационные системы
Область применения	Медицинские Географические Геологические и т. п.
Назначение	Диагностика Проектирование Обучение Системы управления Информационно-поисковые
Медицинские ИС	
Уровень	Базовый ЛПУ Территориальный Государственный
Выполняемые функции	Медико-технологические Статистические Информационно-справочные и т. п.
Медико-технологические ИС	
Назначение	Вычислительной диагностики и др. Системы клинико-лабораторных исследований Системы выбора лечебной тактики

По области применения разрабатываемая информационная система относится к медицинским системам, используемым для диагностики заболеваний почек и обучения студентов медицинских ВУЗов. Так как информационная система дифференциальной диагностики болезней почек предназначена для помощи врачу – нефрологу, поэтому она относится к информационным медицинским системам базового уровня. По выполняемым функциям разрабатываемая информационная система относится к медико-технологическим информационным системам, а по назначению к системам вычислительной диагностики.

Интеллектуальная информационная система – это взаимосвязанная совокупность средств, методов и персонала, имеющая возможность хранения, обработки и выдачи информации.[8].

Экспертные системы - это направление исследований по созданию вычислительных систем, которые умеют принимать решения, похожие на решения экспертов заданной предметной области [10].

Так как информационная система дифференциальной диагностики заболеваний почек необходима для решения практических задач в достаточно узкоспециализированной области, в которой ведущую роль играют опыт и знания врача-нефролога, то данная информационная система относится к экспертным системам.

Типовая структура экспертных систем представлена на рисунке 2.4 [9].



Рисунок 2.4 – Структура экспертной системы

База данных хранит форматированные данные – конкретные факты и константы. В базе данных накапливаются данные, собранные от экспертов, например, данные по диагнозам и результатам обследований пациентов из историй болезней (знания экспертов). Знания экспертов представляются в базе знаний в форме, например, диагностических моделей. Фундаментальные знания извлекаются, например, из медицинских справочников. Машина вывода играет роль интерпретатора, она применяет знания подходящим образом, для получения результата (диагноза) по определённым алгоритмам. Система приобретения знаний – это механизм пополнения базы знаний и обучения информационной системы. Реализуется этот механизм с помощью эксперта и машинного обучения. Система интерфейса преобразует запросы пользователя и выдает ему результат. Система объяснения решения – это механизм, протоколирующий работу системы интерпретации знаний и предоставляющий пользователю результат в виде цепочки логических выводов [9].

Таким образом, по области применения разрабатываемая информационная система относится к медицинским системам, используемым для диагностики заболеваний почек и обучения студентов медицинских ВУЗов. Так как информационная система дифференциальной диагностики болезней почек предназначена для помощи врачу – нефрологу, то она относится к информационным медицинским системам базового уровня. По выполняемым функциям разрабатываемая информационная система относится к медико-технологическим информационным системам, а по назначению к системам вычислительной диагностики.

Структура разрабатываемой информационной системы будет состоять из базы данных, извлекаемых из историй болезней, базы знаний, содержащей диагностические модели заболеваний, решателя, содержащего алгоритмы принятия решения о диагнозе по результатам обследования пациента и интерфейса пользователя.

2.2 Информационные системы распознавания болезней (примеры)

MYCIN. Экспертная система диагностики и лечения инфекционных заболеваний, разработанная в 1970 годах в Стэнфордском университете. Система работает на базе знаний, состоящей примерно из 600 правил. Программа задает пользователю ряд вопросов, а затем выдает результат о предположительном возбудителе болезни, диагнозе, а также рассчитывает необходимого количества антибиотиков в зависимости от массы тела пациента. Программа никогда не использовалась на практике из-за этических и правовых моментов, а так же необходимости вводить данные в систему вручную. Это занимало около 30 минут времени пользователя (врача) и было крайне неудобно в условиях клиники [12].

EasyDiagnosis. Экспертная система, содержащая список и клиническое описание наиболее вероятных заболеваний, основанных на анализе симптомов пациента. Данная система является бесплатным англоязычным сервисом, работающим онлайн. Программа просит выбрать симптомы и признаки болезни из предлагаемого списка (рисунок 2.5).

Click on the desired module below.

Symptoms and Signs

- CHEST PAIN
- CHEST PAIN (PRO VERSION)
- CONSTIPATION
- COUGH
- FAINTING (SYNCOPE)
- CHRONIC FATIGUE
- HEADACHE
- BACK PAIN
- VERTIGO (DIZZINESS)
- UPPER ABDOMINAL PAIN AND INDIGESTION

Conditions

- ALCOHOLISM
- DEPRESSION
- GALL BLADDER DISEASE
- HEARING LOSS
- OSTEOPOROSIS
- SEIZURES AND EPILEPSY
- SEIZURES AND EPILEPSY (PRO VERSION)
- ABDOMINAL CONDITIONS

Рисунок 2.5 –Предлагаемые симптомы и признаки экспертной системы EasyDiagnosis

В списке мы видим такие пункты как: грудная боль, головная боль, кашель, запор, алкоголизм, потеря слуха, остеопороз.

После выбора симптома, программа предупреждает пользователя о том, что ввод неправильной информации приведет к ошибочным результатам.

Также программа предупреждает о том, что она не ставит диагноза (это может сделать только врач), а только рассчитывает вероятности на основе ответов пользователя.

Далее программа просит ввести свой возраст и пол, а затем ответить на вопросы о характере боли, ее местоположении, сопутствующих заболеваниях, вредных привычках. Количество вопросов зависит от выбранного симптома (рисунок 2.6).

1.

Where is the chest pain located.

☐

A. Under the sternum (breastbone) or between the shoulder blades

☐

B. Left chest in front ("under the heart")

☐

C. "Pit of stomach" or in mid upper abdomen

☐

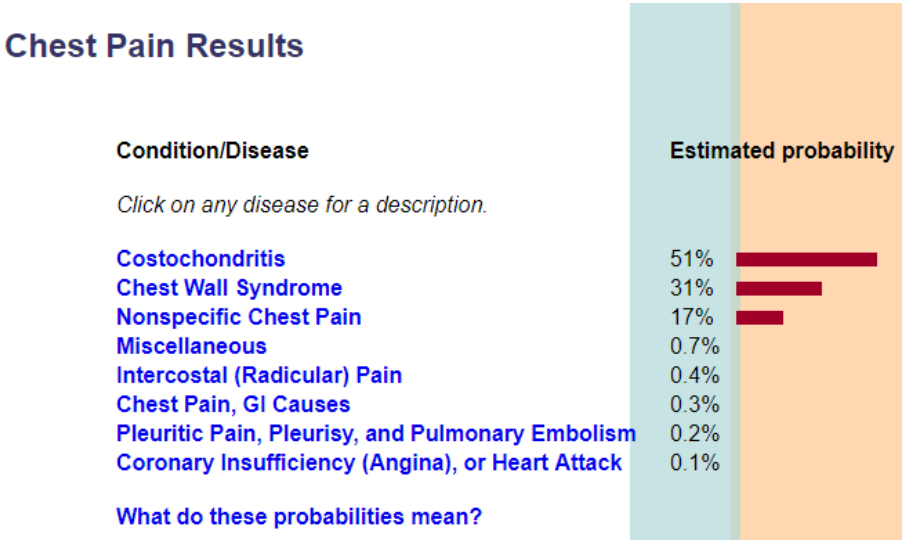
D. None of the above

☐

E. Unknown/not applicable

Рисунок 2.6 – Пример вопроса экспертной системы EasyDiagnosis

В результате прохождения теста появляется диаграмма с наиболее вероятными заболеваниями (рисунок 2.7) [11].



Домашний доктор. Русскоязычная медицинская экспертная система. Определение заболевания происходит на основе теста, который проходит пользователь. База знаний программы включает около 100 самых распространенных заболеваний (рисунок 2.8) [10].

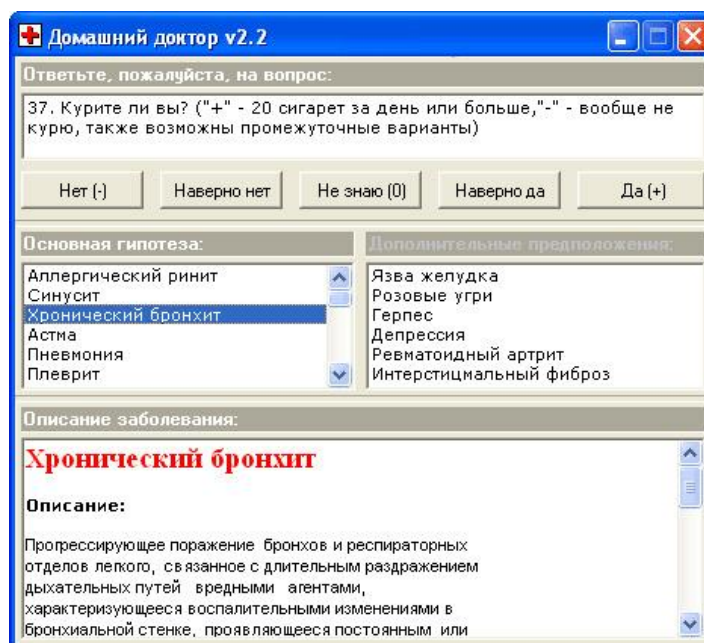


Рисунок 2.8 – пример интерфейса программы «Домашний доктор»

DXplain – система поддержки принятия решений, разработанная в Бостоне в 1986. Данная система используется до сих пор.

В настоящее время база знаний системы включает в себя более 2400 заболеваний и более 5000 клинических данных (симптомы, признаки, эпидемиологические сведения, а также лабораторные и инструментальные данные). В DXplain используется интерактивный формат для сбора клинической информации и измененная форма байесовской логики для вывода заключения.

DXplain применяется в ряде больниц и медицинских образовательных учреждениях для обучения и образовательной помощи в решении клинических проблем. Данная интеллектуальная система не дает

окончательного медицинского заключения и не может использоваться в качестве замены для врача [15].

AIMedica – это интернет-услуга по компьютерной диагностике заболеваний на основе введенных пользователем жалоб (субъективных симптомов), лабораторных анализов и половозрастных данных.

Пользователь вводит эту информацию в произвольном порядке и отвечает на вопросы, уточняющие возможные диагнозы. Результатом диагностики является перечень из 10 наиболее вероятных диагнозов болезни с указанием процента вероятности, рекомендации обратиться к врачам соответствующей специальности для уточнения диагноза и назначения лечения, а также разнообразная справочная информация, связанная с заболеванием.

Основу сервиса составляет база знаний, разработанная путем анализа больших массивов слабоструктурированных данных, содержащихся в более чем 22 млн. медицинских источниках [17]. Сравнение основных характеристик информационных систем представлено в таблице 2.2.

Таблица 2.2 – Сравнение основных характеристик информационных систем распознавания болезней

Информационная система	Производитель (страна)	Целевая аудитория	Входные данные	Количество нозологий
DXplain	США	Врачи, студенты медицинских вузов	Симптомы, лабораторные показатели	2400
MYCIN	США	Врачи, студенты медицинских вузов	Симптомы	600
EasyDiagnosis	США	пациенты	Субъективные жалобы	Основные нозологии
AIMedica	Россия	пациенты	Субъективные жалобы	225 (тестовая версия)
Домашний доктор	Россия	пациенты	Субъективные жалобы	100

В рассмотренных информационных системах входными данными являются субъективные жалобы пациентов (EasyDiagnosis, AIMedica, Домашний доктор), это является недостатком данных систем, в связи с низкой достоверностью субъективных симптомов. Российский рынок информационных медицинских систем на данный момент находится на стадии становления, информационные системы для диагностики заболеваний почек представлены единичными разработками, так же эти системы не используют данные лабораторных анализов, хотя эти данные являются объективными и играют важную роль в диагностике заболеваний почек.

Таким образом, в анализируемых экспертных системах базы знаний не содержат данных результатов обследований и диагнозов пациентов. Входными данными являются субъективные жалобы пациентов, что имеет более низкую информативность по сравнению с данными лабораторных анализов. В рассматриваемых системах не заложена методика построения вероятностной диагностической модели на основе базы данных. В связи с этим разработка информационной системы, содержащей базу данных результатов обследования и диагнозов пациентов с заболеваниями почек, а также решатель для формирования вероятностной диагностической модели является актуальной задачей.

2.3 Структура информационной системы распознавания болезней почек

Структура информационной системы распознавания болезней почек состоит из четырех основных элементов: базы данных, базы знаний, решателя и интерфейса пользователя (рисунок 2.9).

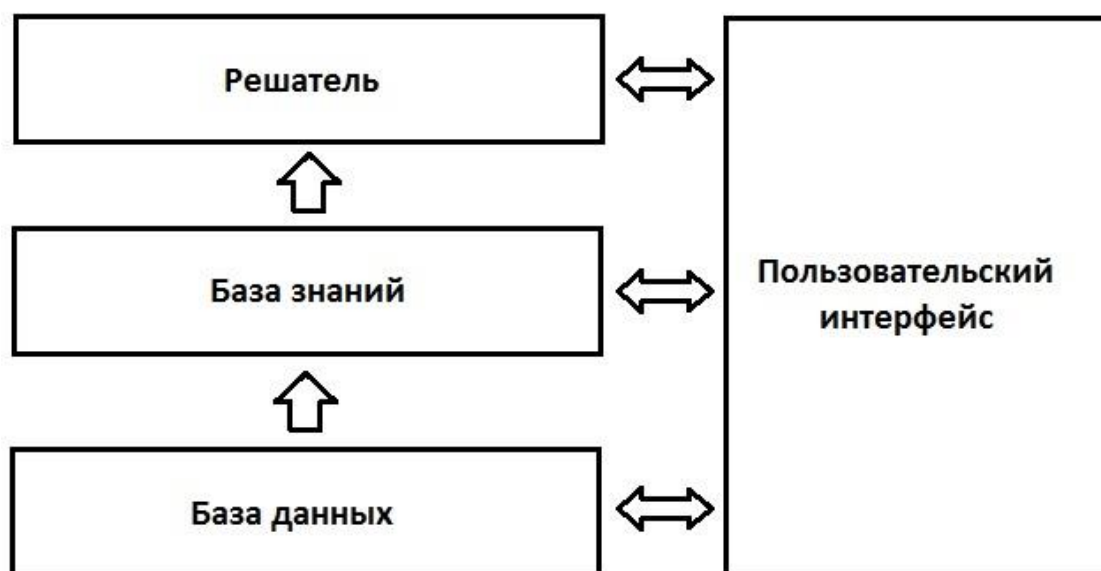


Рисунок 2.9 – Структура информационной системы распознавания болезней почек

База данных реализуется в MS Access. Она необходима для накопления и хранения данных пациентов, результатов лабораторных анализов и установленных диагнозов заболеваний почек. В базе данных формируются исходные данные для разработки диагностической модели. Исходными данными являются данные о результатах обследования и диагнозах пациентов с заболеваниями почек, собранные на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического отделения ГБУЗ КМРБ, а также данные обследования здоровых пациентов, сгенерированные в среде MS Excel. Из базы данных также формируется запрос о результатах лабораторных анализов, который объединяет в себе данные из различных таблиц. Этот запрос экспортируется в Excel.

База знаний включает в себя диагностические модели, построенные на основе данных из карт пациентов с заболеваниями почек и неразрывно связана с базой данных и решателем. Диагностические модели формируются в среде Matlab.

Решатель необходим для статистической обработки данных обследований пациентов с заболеваниями почек в программе MS Excel. В результате статистической обработки мы получаем общие статистические

характеристики, закон распределения признаков, корреляцию признаков. Также решатель необходим оценивания апостериорных вероятностей методом Байеса в Matlab.

Интерфейс пользователя реализуется посредством среды GUIDE программы Matlab, посредством макросов программы MS Excel, посредством макросов и кнопочных форм программы MS Access.

Таким образом, структура информационной системы распознавания болезней почек состоит из трех основных элементов: базы данных, реализуемой с помощью MS Access; решателя, который необходим для статистической обработки данных в MS Excel, формирования диагностической модели и оценивания апостериорных вероятностей методом Байеса в Matlab; интерфейса пользователя, реализуемого посредством среды GUIDE программы Matlab, макросов программы MS Excel, макросов и кнопочных форм программы MS Access.

2.4 База данных пациентов с болезнями почек

База данных необходима для накопления результатов обследования и диагнозов пациентов с заболеваниями почек.

Создание базы данных происходит в несколько этапов.

На первом этапе необходимо выявить и подобрать данные, которые будут храниться в базе. В нашем случае – это данные о результатах обследования и диагнозах пациентов с заболеваниями почек, собранные на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического отделения ГБУЗ КМРБ.

На втором этапе необходимо определить сущности базы данных, и представить каждую из сущностей в виде таблицы базы данных. Каждая таблица будет заполняться определенными значениями, а так же содержать первичный ключ. Первичный ключ — в реляционной модели данных один из

потенциальных ключей отношения, выбранный в качестве основного ключа (рисунок 2.10 – 2.13).

	Имя поля	Тип данных
🔑	Код пациента	Счетчик
	Фамилия	Текстовый
	Имя	Текстовый
	Отчество	Текстовый
	Пол	Текстовый
	Дата рождения	Дата/время

Рисунок 2.10– Таблица «Пациенты»

	Имя поля	Тип данных
	Код обследования	Счетчик
	Код пациента	Числовой
	Код ОАМ	Числовой
	Код БАК	Числовой

Рисунок 2.11 – Таблица «Обследования»

	Имя поля	Тип данных
🔑	Код диагноза	Счетчик
	Код пациента	Числовой
	Код заболевания	Числовой
	Код специалиста	Текстовый

Рисунок 2.9 – Таблица «Диагнозы»

	Имя поля	Тип данных
🔑	Код специалиста	Счетчик
	Специальность	Текстовый

Рисунок 2.10 – Таблица «Специалисты»

	Имя поля	Тип данных
🔑	Код заболевания	Числовой
	Код диагноза по МКБ 10	Текстовый
	Название заболевания	Текстовый

Рисунок 2.11 – Таблица «Заболевания»

Имя поля	Тип данных
Код БАК	Счетчик
Креатенин мкмоль/л	Текстовый

Рисунок 2.12 – Таблица «БАК»

Имя поля	Тип данных
Код ОАМ	Счетчик
Белок г/л	Текстовый
Эритроциты в пз	Текстовый
Лейкоциты в пз	Текстовый

Рисунок 2.13 – Таблица «ОАМ»

Теперь, для обеспечения целостности данных при их добавлении в таблицы, упрощения создания форм, отчетов, запросов необходимо связать таблицы базы данных между собой. Тип связи – один ко многим.

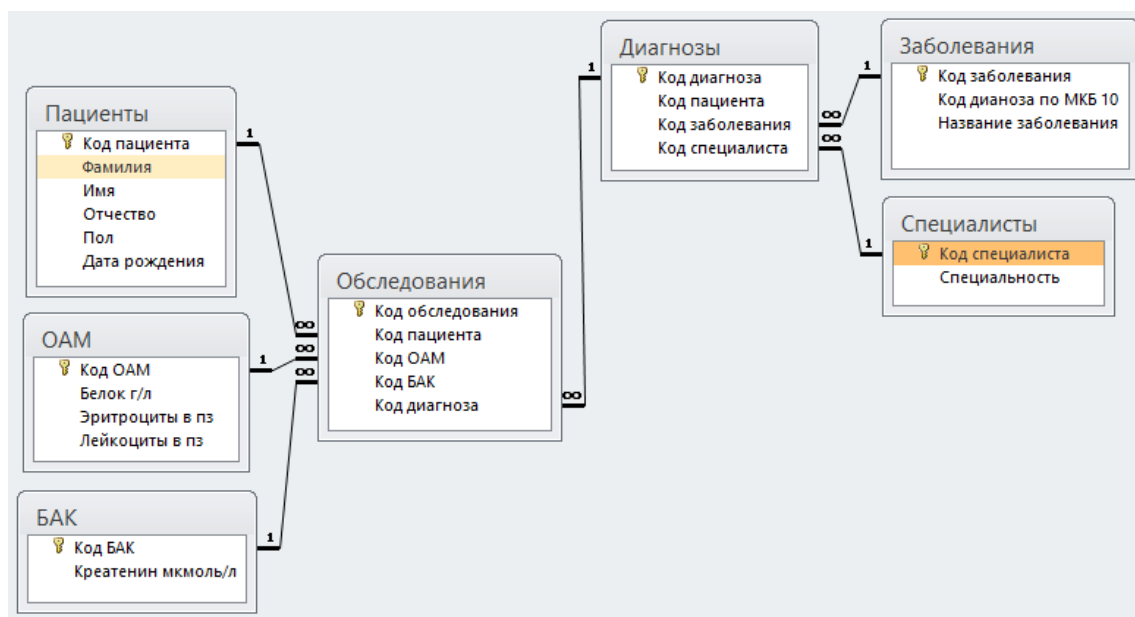


Рисунок 2.14 – Схема данных

Для уменьшения временных затрат на ввод данных в таблицы, создаем аналогичные им формы (рисунок 2.15). Формы используются для ввода, корректировки и управления данными, хранящимися в базе.

Пациенты

Код пациента

1

Фамилия

Дубенцева

Имя

Елена

Отчество

Николаевна

Пол

Ж

Дата рождения

17.02.1977

Рисунок 2.15 – Форма «Пациенты»

Через формы заполняем каждую таблицу данными (рисунок 2.16).

Пациенты						
	Код пациен	Фамилия	Имя	Отчество	Пол	Дата рождения
+	1	Дубенцева	Елена	Николаевна	Ж	17.02.1977
+	2	Иванова	Нина	Максимовна	Ж	05.03.1938
+	3	Заводчикова	Антонани	Сергеевна	Ж	30.05.1938
+	4	Зенюкова	Эльвира	Шамильевна	Ж	19.12.1965
+	5	Земскова	Людмила	Викторовна	Ж	26.11.1955
+	6	Рогачева	Александра	Владимировн	Ж	27.11.1956
+	7	Сапожникова	Валентина	Александров	Ж	14.04.1940
+	8	Камбаренц	Анжела	Михайловна	Ж	22.05.1959
+	9	Подвалова	Светлана	Владимировн	Ж	01.09.1963
+	10	Паноренова	Галина	Андреевна	Ж	24.02.1951
+	11	Савельева	Валентина	Васильевна	Ж	25.04.1953
+	12	Пельнина	Екатерина	Николаевна	Ж	13.10.1946
+	13	Федотова	Любовь	Степановна	Ж	31.08.1956
+	14	Любанова	Галина	Андреевна	Ж	21.12.1948
+	15	Леванов	Николай	Александров	М	03.07.1949
+	16	Шарибгиланов	Рушан	Равилович	М	06.01.1983
+	17	Бганова	Светлана	Владимирона	Ж	21.04.1969
+	18	Клюев	Виктор	Васильевич	М	21.03.1956
+	19	Веретенникова	Валентина	Викторовна	Ж	03.04.1951
+	20	Клюев	Иван	Михайлович	М	09.09.1980
+	21	Шаннина	Ольга	Алексеевна	Ж	06.12.1955
+	22	Шушпанова	Валентина	Ивановна	Ж	23.12.1940
+	23	Ларионова	Марина	Ивановна	Ж	15.04.1985
+	24	Земцова	Марина	Константинов	Ж	18.04.1961
+	25	Кондрашева	Надежда	Васильевна	Ж	13.02.1955

Рисунок 2.16 –Таблица «Пациенты» с внесенными данными

Для экспорта данных из базы в решатель необходимо создать простой запрос, который будет включать в себя данные из различных таблиц – код

пациента, название заболевания и результаты обследований (белок, лейкоциты, эритроциты в моче и уровень креатинина в крови) (рисунок 2.17).

Результаты обследований					
Код пациен	Название заболевания	Белок г/л	Эритроциты в пз	Лейкоциты в пз	Креатинин мкмоль/л
2	Острый пиелонефрит	0,041	1	55	99
3	Острый пиелонефрит	0,03	2	35	118
4	Острый пиелонефрит	0,046	3	73	97
5	Острый пиелонефрит	0,032	1	11	64
6	Острый пиелонефрит	0,059	1	70	90
7	Острый пиелонефрит	0,044	4	65	104
8	Острый пиелонефрит	0	3	27	68
9	Острый пиелонефрит	0,278	0	88	84
10	Острый пиелонефрит	0,118	2	70	53
11	Острый пиелонефрит	0,02	2	28	81
12	Острый пиелонефрит	0,03	3	14	64
13	Острый пиелонефрит	0	2	45	71
14	Острый пиелонефрит	0,033	3	25	133
15	Острый пиелонефрит	0	1	36	98
16	Острый пиелонефрит	0,034	2	33	81
17	Острый пиелонефрит	0,41	1	28	90
18	Острый пиелонефрит	0,031	0	32	64
19	Острый пиелонефрит	0,03	2	81	81
20	Острый пиелонефрит	0	5	46	138
21	Острый пиелонефрит	0,044	6	48	89
22	Острый пиелонефрит	0	1	55	78
23	Острый пиелонефрит	1,01	1	33	97
24	Острый пиелонефрит	0	3	35	64
25	Острый пиелонефрит	0,033	2	78	64
26	Острый пиелонефрит	1,02	5	27	91

Рисунок 2.17 – Запрос «Результаты обследований»

Для экспорта таблицы «Результаты обследований» в MS Excel необходимо создать макрос. Макрокомандой этого макроса будет экспорт с форматированием (рисунок 2.18).

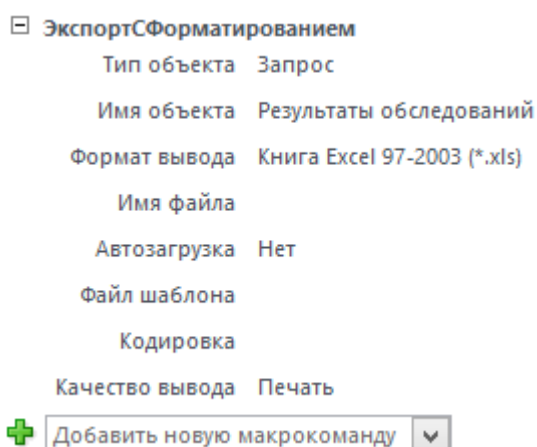


Рисунок 2.18 – Структура макроса для экспорта данных в MS Excel

Таким образом, была сформирована база данных, содержащая результаты обследования и диагнозы пациентов с заболеваниями почек, собранные на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического отделения ГБУЗ КМРБ. Принятие диагностических решений на основе опыта высококвалифицированных специалистов, ведет к повышению достоверности постановки диагноза. Данный опыт собран и хранится в базе данных информационной системы, содержащей результаты обследований и достоверные диагнозы пациентов.

2.5 Пользовательский интерфейс базы данных

Для удобства использования базы данных необходимо создать управляющую форму, состоящую из главной формы и подчиненных форм. В нашем случае предполагается одна главная форма и четыре подчиненные (рисунок 2.19, 2.20).



Рисунок 2.19 – Структура управляющей формы

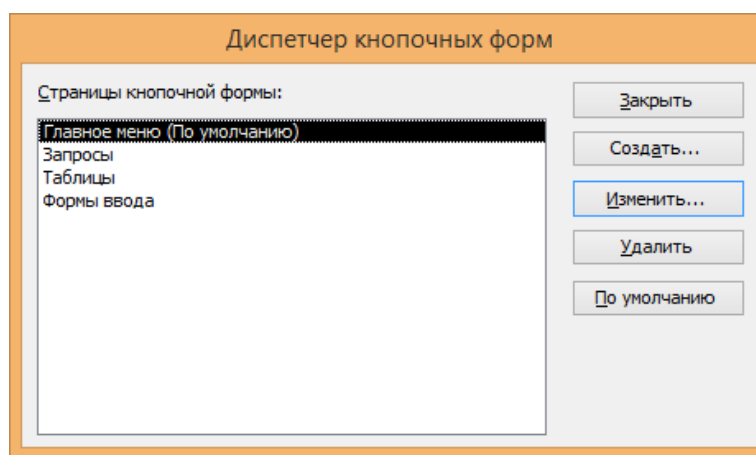


Рисунок 2.20 – Основные страницы кнопочной формы

Управляющая форма предоставляет пользователю возможность вносить новые данные в таблицы через раздел «Формы ввода», редактировать уже заполненные таблицы через раздел «Таблицы», получать результаты анализов пациентов через раздел «Запросы», экспортировать полученные результаты в MS Excel через раздел «Экспорт результатов обследований», возвращаться к главной кнопочной форме с побочных форм с помощью раздела «К главному меню», а также закрывать базу данных с помощью раздела «Выход» (рисунок 2.21, 2.22).

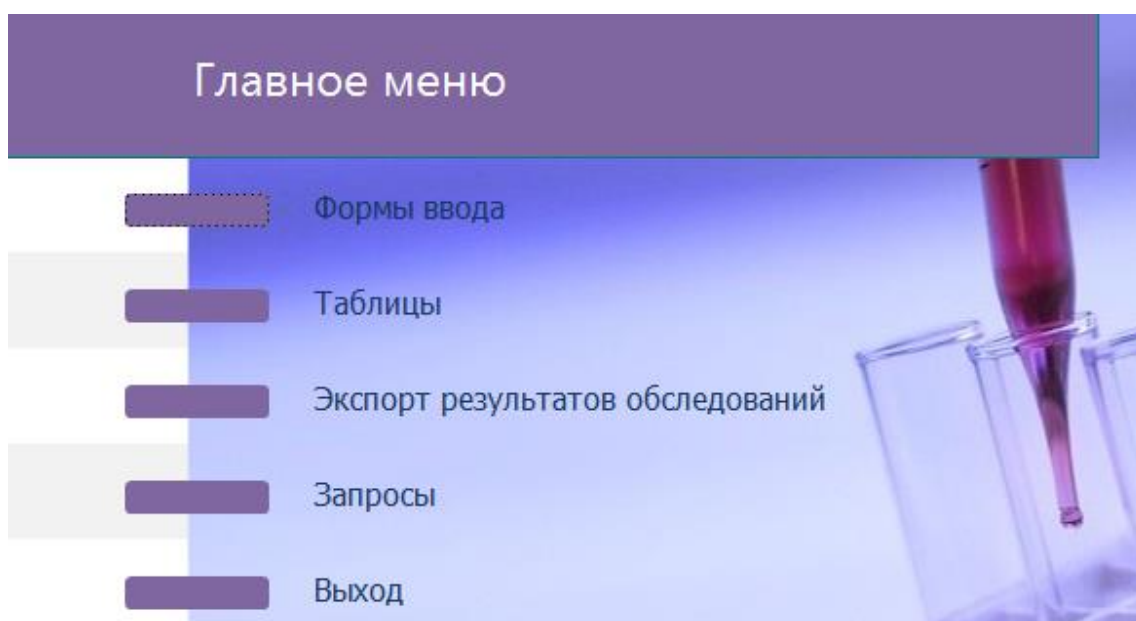


Рисунок 2.21 – Главное меню управляющей формы

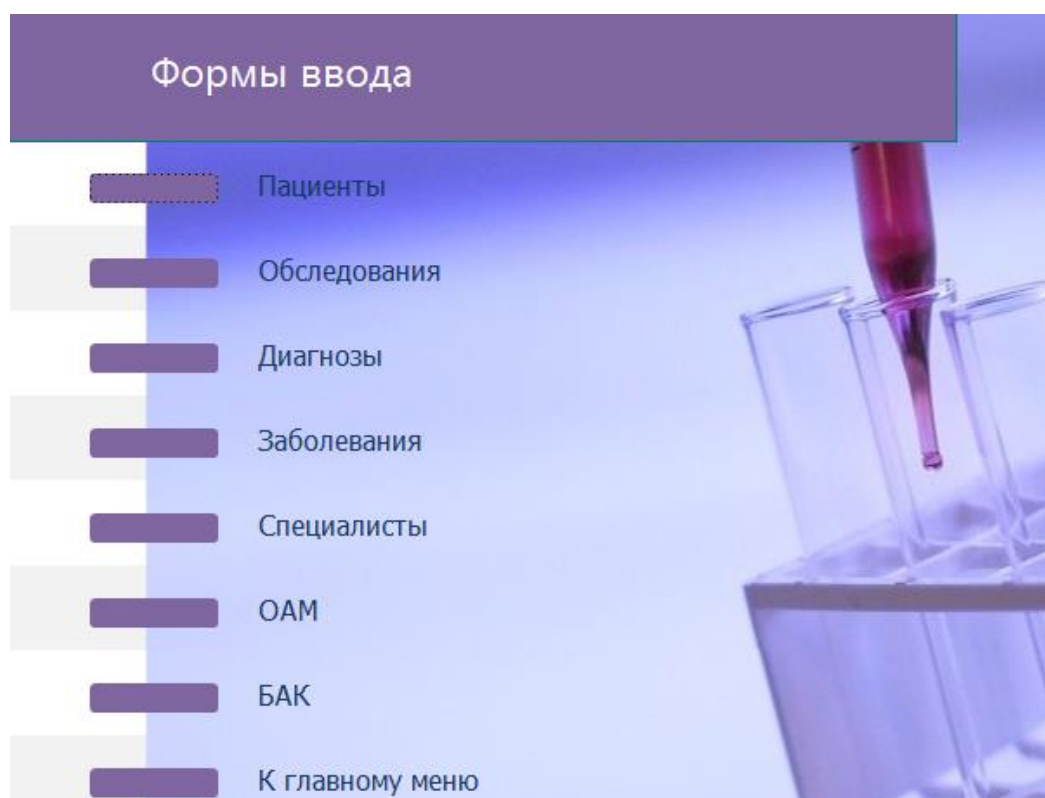


Рисунок 2.22 – Подчиненная страница управляющей формы «Формы ввода»

Таким образом, кнопочные формы MS Access обеспечивают простоту и доступность интерфейса созданной базы данных. Главное меню кнопочной формы позволяет быстро найти необходимые пользователю элементы базы данных для ввода, изменения и экспорта диагностической информации (рисунок 2.21, 2.22). Простой пользовательский интерфейс позволяет снизить требования к квалификации медицинских работников в области информационных технологий.

2.6 Основные результаты и выводы по разделу 2

1 Разрабатываемая информационная система по области применения относится к медицинским системам, используемым для диагностики заболеваний почек и обучения студентов медицинских ВУЗов. Так как информационная система дифференциальной диагностики болезней почек

предназначена для помощи врачу – нефрологу, то она относится к информационным медицинским системам базового уровня. По выполняемым функциям разрабатываемая информационная система относится к медико-технологическим информационным системам, а по назначению к системам вычислительной диагностики. Структура разрабатываемой информационной системы будет состоять из трех основных элементов – базы данных, с диагнозами пациентов и результатами лабораторных обследований, решателя и интерфейса пользователя.

2 В анализируемых экспертных системах базы знаний не содержат данных результатов обследований и диагнозов пациентов. Входными данными являются субъективные жалобы пациентов, что имеет более низкую информативность по сравнению с данными лабораторных анализов. В рассматриваемых системах не заложена методика построения вероятностной диагностической модели на основе базы данных. В связи с этим разработка информационной системы, содержащей базу данных результатов обследования и диагнозов пациентов с заболеваниями почек, а также решатель для формирования вероятностной диагностической модели является актуальной задачей.

3 Структура информационной системы распознавания болезней почек состоит из трех основных элементов: базы данных, реализуемой с помощью MS Access; решателя, который необходим для статистической обработки данных в MS Excel, формирования диагностической модели и оценивания апостериорных вероятностей методом Байеса в Matlab; интерфейса пользователя, реализуемого посредством среды GUIDE программы Matlab, макросов программы MS Excel, макросов и кнопочных форм программы MS Access.

4 Была сформирована база данных, содержащая результаты обследования и диагнозы пациентов с заболеваниями почек, собранные на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического

отделения ГБУЗ КМРБ. Принятие диагностических решений на основе опыта высококвалифицированных специалистов ведет к повышению достоверности постановки диагноза. Данный опыт собран и хранится в базе данных информационной системы, содержащей результаты обследований и достоверные диагнозы пациентов.

5 Простота и доступность интерфейса созданной базы данных, обеспечивается за счет кнопочных форм MS Access . Главное меню кнопочной формы позволяет быстро найти необходимые пользователю элементы базы данных для ввода, изменения и экспорта диагностической информации (рисунок 2.21, 2.22). Простой пользовательский интерфейс позволяет снизить требования к квалификации медицинских работников в области информационных технологий.

3 ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ ОБСЛЕДОВАНИЯ В EXCEL

3.1 Методика обработки статистических данных

Одним из этапов формирования вероятностной диагностической модели является статистическая обработка лабораторных данных из карт пациентов с заболеваниями почек. Статистическая обработка данных происходит в несколько этапов (рисунок 3.1).

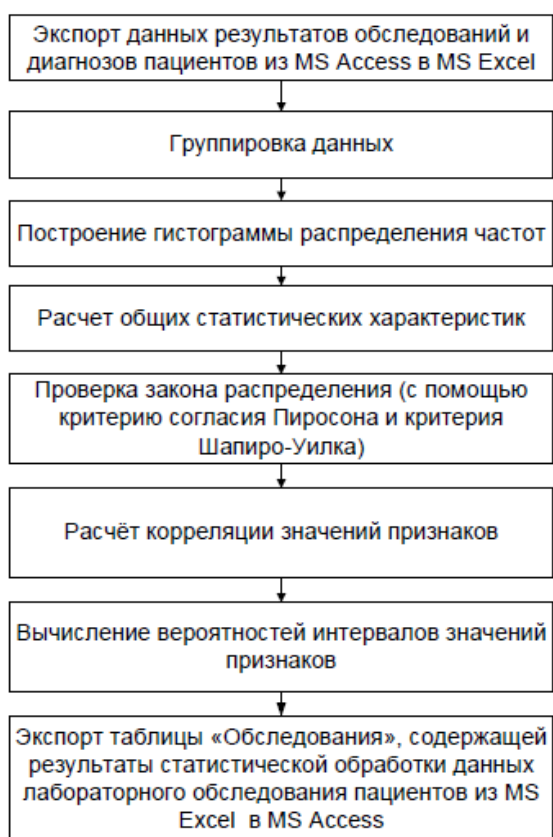


Рисунок 3.1 – Этапы методики статистической обработки данных в
решателе MS Excel

Экспорт данных результатов обследований и диагнозов пациентов с пиелонефритом и гломерулонефритом из MS Access в решатель MS Excel осуществляется с помощью макроса.

Данные пациентов с установленными диагнозами пиелонефрита и гломерулонефрита получены из историй болезней на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического отделения ГБУЗ КМРБ.

Данные третьей группы пациентов (здоровые пациенты) получены с помощью генерации случайных чисел в MS Excel, при помощи пакета «Анализ данных». У здорового человека каждый из показателей может варьироваться в определенном диапазоне (таблица 3.1).

Таблица 3.1 – Нормальные результаты лабораторных анализов для количества эритроцитов в моче, лейкоцитов и уровня креатинина в крови

Название показателя	Количество эритроцитов в моче (клеток в поле зрения)	Количество лейкоцитов в моче (клеток в поле зрения)	Уровень креатинина в крови (мкмоль/л)
Нормальные показатели	0 - 3	0 - 6	Для женщин: 53 – 97. Для мужчин: 62 – 115

В среде Excel, с использованием пакета «Анализ данных», генерируем выборку с нормальным распределением признаков, так как нормальное распределение встречается наиболее часто (таблица 3.2):

Таблица 3.2 – Параметры генерации выборок для здоровых пациентов

Параметр генерации	Эритроциты	Лейкоциты	Креатинин
Число переменных	1		
Число случайных чисел	30		
Распределение	Нормальное		
Среднее	2	3	84
Стандартное отклонение	0,7	1,5	23

Так как эритроциты и лейкоциты – это количество клеток в поле зрения, то эти показатели не могут быть дробными. Поэтому округляем получившиеся

при генерации значения до 0 чисел после запятой. Результат генерации чисел для здоровых пациентов показан на рисунке 3.2:

Креатинин ЗД	Лейкоциты ЗД	Эритроциты ЗД
85	4	1
85	2	1
110	4	0
105	3	1
89	5	3
86	5	0
71	1	1
104	2	2
64	6	2
86	5	2
64	4	3
82	3	2
114	2	2
80	4	2
69	2	2
81	7	1
80	2	2
44	5	2
89	2	2
78	2	2
103	7	1
72	1	2
105	4	3
98	2	0
79		

Рисунок 3.2 – Генерация выборок для пациентов с нормальными показателями анализов крови и мочи

Каждую выборку пациентов размещаем на отдельном листе. При этом получается 9 листов (количество эритроцитов, лейкоцитов и уровень креатинина в крови у здоровых пациентов, у пациентов с гломерулофритом и пиелонефритом).

Таким образом, модельные данные результатов лабораторных исследований у здоровых пациентов были сгенерированы в среде Excel по параметрам нормального распределения признаков. Методика статистической обработки данных включает в себя экспорт результатов анализов

из MS Access в MS Excel, группировку данных, построение гистограммы распределения частот, расчет общих статистических характеристик, определение законов распределения, оценку корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.1).

3.2 Построение вариационных рядов значений признаков

Теперь необходимо систематизировать все собранные данные. Для этого необходимо их сгруппировать - объединить по какому-либо признаку в относительно однородные группы.

Вариационный ряд – соотношение ранжированных значений признака с частотой их встречаемости в данной выборке.

Статистическая совокупность может распределяться в безынтервальный или интервальный вариационный ряд в зависимости от того, как и в каком диапазоне варьирует признак. Тип вариационного ряда можно определить по формуле:

$$\lambda = x_{max} - x_{min} / k, \quad (3.1)$$

где λ – ширина классового интервала;

x_{max} – максимальная варианта совокупности;

x_{min} – минимальная варианта совокупности;

k – число классов, на которые следует разбить вариацию признака, рассчитывается по формуле Стерджеса:

$$k = 1 + 3.32 \lg(n), \quad (3.2)$$

где n – объем выборки.

Если ширина классowego интервала $\lambda = 1$ или $\lambda \approx 1$, то строится безынтервальный ряд, если $\lambda \neq 1$, то строится интервальный ряд [18].

Для построения интервального ряда необходимо определить определения нижнюю границу первого классowego интервала. Для этого используют формулу:

$$x_n = x_{min} - \lambda / 2, \quad (3.3)$$

где x_n – нижняя граница первого классowego интервала;

x_{min} – минимальная варианта совокупности;

λ – величина классowego интервала.

При этом минимальная варианта совокупности попадает примерно в середину первого классowego интервала.

После определения классowych интервалов, определяют частоты каждого класса.

Для этого последовательно прибавляют к x_n λ , а затем уменьшают верхние границы каждого класса на величину, равную точности, принятой при измерении признака. Это необходимо для разграничения классowych интервалов.

Наиболее точно центральную величину классowego интервала можно получить по формуле [19]:

$$x_i = (x_n - x_k) / 2, \quad (3.4)$$

где x_i – центральная величина классowego интервала;

x_n – нижняя граница первого классowego интервала;

x_k – конечная точка интервала.

Рассмотрим обработку данных для пациентов с пиелонефритом для признака количества эритроцитов (обработка остальных выборок будет проходить аналогично).

На листе «Эритроциты ПН» расположим соответствующие значения признака в столбец А (ячейки А2:А30). В столбце В найдем минимальную и максимальную варианты признака. Для этого в ячейке В2 запишем формулу «=МИН(А:А)», а в ячейке В4 – «=МАКС(А:А)». Рассчитаем величину классового интервала по формуле (3.1).

Определим нижнюю границу первого классового интервала по формуле (3.3). Для этого в ячейку В8 запишем формулу «=В2–0,5*В6» (рисунок 3.3).

Минимальное значение
=МИН(А:А)
Максимальное значение
=МАКС(А:А)
λ
$= (B4 - B2) / (1 + 3,32 * \text{LOG10}(\text{СЧЁТ}(А:А)))$
X_n
$= B2 - 0,5 * B6$
k
$= 1 + 3,32 * \text{LOG10}(30)$

Рисунок 3.3 – Расчет минимальных и максимальных вариантов выборки, величины классового интервала и нижней границы первого класса в среде Excel

В столбцах С и D произведем разбиение ряда на классовые интервалы (рисунок 3.4):

Xd	Xu
=B\$8	=C3-0,01
=C2+B\$6	=C4-0,01
=C3+B\$6	=C5-0,01
=C4+B\$6	=C6-0,01
=C5+B\$6	=C7-0,01
=C6+B\$6	=C8-0,01
=C7+B\$6	=C9-0,01
=C8+B\$6	

Рисунок 3.4 – Разбиение ряда на классовые интервалы в среде Excel

Рассчитаем центральные величины классовых интервалов по формуле (3.4) (рисунок 3.5):

X_i
=(C2+D2)/2
=(C3+D3)/2
=(C4+D4)/2
=(C5+D5)/2
=(C6+D6)/2
=(C7+D7)/2
=(C8+D8)/2

Рисунок 3.5 – Расчет центральных величин классовых интервалов в Excel

Для построения безынтервального вариационного ряда выборку упорядочивают по возрастанию ее элементов, а затем определяют частоту каждого признака.

Для более наглядного предоставления закономерности варьирования количественных признаков, изобразим вариационный ряд в виде гистограммы. По оси абсцисс откладываем срединные значения классов, по оси ординат – частоты. Частоты показывают, сколько раз в совокупности встречаются данные значения признака. Высота перпендикуляров, восставляемых по оси абсцисс, соответствует частотам классов.

Расчёт частот в среде Excel приведен на рисунке 3.6.

fi
=СЧЁТЕСЛИМН(A:A;">="&C2;A:A;"<="&D2)
=СЧЁТЕСЛИМН(A:A;">="&C3;A:A;"<="&D3)
=СЧЁТЕСЛИМН(A:A;">="&C4;A:A;"<="&D4)
=СЧЁТЕСЛИМН(A:A;">="&C5;A:A;"<="&D5)
=СЧЁТЕСЛИМН(A:A;">="&C6;A:A;"<="&D6)
=СЧЁТЕСЛИМН(A:A;">="&C7;A:A;"<="&D7)
=СЧЁТЕСЛИМН(A:A;">="&C8;A:A;"<="&D8)
=СУММ(E1:E8)

Рисунок 3.6 – Расчёт частот в Excel

Для построения гистограммы распределения частот по оси абсцисс откладывают границы классовых интервалов, по оси ординат – частоты интервалов [18]. Гистограмма распределения частот количества эритроцитов в моче у пациентов с пиелонефритом приведена на рисунке 3.7.



Рисунок 3.7 – Гистограмма распределения частот количества эритроцитов в моче у пациентов с пиелонефритом

3.3 Вычисление характеристик распределений

Для полного описания варьирующих объектов используют статистические обобщающие характеристики (таблица 3.3).

Таблица 3.3 – Общие статистические характеристики

Характеристики распределения	Определения
Объем выборки	Число наблюдений, образующих выборку
Выборочное среднее	Средняя оценка изучаемого в эксперименте признака
Дисперсия	Средний квадрат отклонений индивидуальных значений признака от их средней величины
Среднеквадратическое отклонение (СКО)	В теории вероятностей и статистике наиболее распространенный показатель рассеивания значений случайной величины относительно её математического ожидания
Медиана	Число, характеризующее выборку. Если все элементы выборки различны, то медиана – это такое число выборки, что ровно половина из элементов выборки больше него, а другая половина меньше него
Мода	Это варианта, которая чаще всего встречается в изучаемой совокупности. Она соответствует точке наибольшего подъема графика. Распределение может иметь не одну моду или не иметь моды вообще
Асимметрия	Мера несимметричности распределения. Чем больше показатель отклоняется от 0, тем больше асимметрия. Величина асимметрии может быть положительной (для правосторонней асимметрии) и отрицательной (для левосторонней асимметрии).
Эксцесс	Измеряет остроту пика распределения. Положительный эксцесс говорит об островершинном распределении, а отрицательный – об плосковершинном. [18].

Для расчета данных характеристик используем готовые формулы среды Excel. Расчет обобщающих характеристик приведен на рисунке 3.8.

Объем выборки	=СЧЁТ(A:A)
Выборочное среднее	=СРЗНАЧ(A:A)
Дисперсия	=ДИСП(A:A)
СКО	=СТАНДОТКЛОН(A:A)
Медиана	=МЕДИАНА(A:A)
Мода	=МОДА(A:A)
Эксцесс	=ЭКСЦЕСС(A:A)
Асимметрия	=СКОС(A:A)

Рисунок 3.8 – Расчет обобщающих характеристик в Excel

Аналитическое выражение для нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad (3.5)$$

где μ - математическое ожидание (среднее значение);

σ – стандартное отклонение;

σ^2 – дисперсия.

Общие статистические характеристики для модельных диагностических признаков у здоровых пациентов представлены в таблице 3.4.

Таблица 3.4 – Общие статистические характеристики для модельных диагностических признаков у здоровых пациентов

Диагностические признаки	Здоровое состояние		
	Количество эритроцитов (клеток в поле зрения)	Количество лейкоцитов(клеток в поле зрения)	Уровень креатинина (мкмоль/л)
Объем выборки	30		
Выборочное среднее	1,57	3,3	85,97
Дисперсия	0,74	2,4	269,3
Стандартное отклонение σ	0,86	1,56	16,4
Медиана	2	3	85
Мода	2	2	85
Экцесс	-0,48	-0,53	-0,02
Ассиметрия	-0,05	0,05	-0,27

Общие статистические характеристики для диагностических признаков у пациентов с заболеваниями почек (гломерулонефритом и пиелонефритом) представлены в таблице 3.4.

Таблица 3.4 – Общие статистические характеристики для диагностических признаков болезней почек

Диагностические признаки	Пиелонефрит			Гломерулонефрит		
	Количество эритроцитов (клеток в поле зрения)	Количество лейкоцитов (клеток в поле зрения)	Уровень креатинина (мкмоль/л)	Количество эритроцитов (клеток в поле зрения)	Количество лейкоцитов (клеток в поле зрения)	Уровень креатинина (мкмоль/л)
Объем выборки	30			30		
Выборочное среднее	2,1	47,4	87,4	4,6	4,6	333,4
Дисперсия	2,3	570,4	543,5	5,89	9,08	39217,97
Стандартное отклонение σ	1,52	23,9	23,3	2,4	3,01	198,03
Медиана	2	40,5	85	4	4	314,7
Мода	2	70	64	4	3	-
Экссесс	0,49	-1,2	0,62	-0,96	-0,39	-0,028
Ассиметрия	0,77	0,34	0,94	0,24	0,48	0,6

Нормальное распределение наиболее часто встречающееся. Оно характерно для непрерывных случайных величин, зависящих от многочисленных факторов. При нормальном распределении значений признака наибольшие и наименьшие значения появляются сравнительно редко, а значения, которые ближе к среднему арифметическому, встречаются чаще [18]. При этом распределение становится похожим на колокол. Проверка распределения на нормальность осуществляется по критериям, представленным в таблице 3.6.

Таблица 3.6 – Достоинства и недостатки критериев нормальности

Название критерия	Достоинства	Недостатки
Критерий Пирсона, или критерий χ^2	Универсален, так как применяется при проверке как непрерывных, так и дискретных законов распределения [18]	При сравнительно малом объеме выборки наблюдается потеря части первоначальной информации. В связи с этим рекомендуется дополнять проверку соответствия распределений по критерию χ^2 другими критериями. [22]. Необходимо, чтобы частоты значений признака в крайних классах были ≥ 5 [18]
Критерий Шапиро-Уилка	Считается наиболее мощным критерием, если объем выборки небольшой ($8 \leq n \leq 50$) [18]	Малые выборки $n \leq 8$ при обнаружении отклонений от нормального распределения не дают достоверных результатов [20]
Критерий Колмогорова-Смирнова	Обладает большей мощностью для непрерывных переменных [18]	Выводимые значения вероятности основаны на предположении, что среднее и стандартное отклонение нормального распределения известны априори и не оцениваются из данных. В последние годы предпочтительнее становится критерий нормальности Шапиро-Уилка [23]

Для проверки распределения на нормальность выбирается критерий согласия Пирсона и критерий Шапиро-Уилка, наиболее подходящий для малых выборок.

Критерий согласия Пирсона может быть применим для интервальных и безынтервальных функций, даже если не известны их параметры.

Использование критерия χ^2 подразумевает деление выборки на интервалы и определения частоты каждого интервала. При делении выборки на интервалы происходит потеря части первоначальной информации. Это является недостатком данного критерия.

Поэтому критерий согласия Пирсона необходимо дополнять другими критериями. Особенно это необходимо при сравнительно малом объеме выборки.

Гипотеза о нормально распределённой величине по критерию Пирсона проверялась с использованием MS Excel (рисунок 3.9, 3.10).

Проверка на нормальность по критерию Пирсона						
F(x_i_min)	F(x_i_max)	p_i	n*p_i	p_i%-n*p_i	(p_i%-n*p_i)^2/n*p_i	
=НОРМРАСП(C2;\$H\$4;\$H\$8;1)	=НОРМРАСП(D2;\$H\$4;\$H\$8;1)	=J3-I3	=K3*\$H\$2	=E2-L3	=M3^2/L3	
=НОРМРАСП(C3;\$H\$4;\$H\$8;1)	=НОРМРАСП(D3;\$H\$4;\$H\$8;1)	=J4-I4	=K4*\$H\$2	=E3-L4	=M4^2/L4	
=НОРМРАСП(C4;\$H\$4;\$H\$8;1)	=НОРМРАСП(D4;\$H\$4;\$H\$8;1)	=J5-I5	=K5*\$H\$2	=E4-L5	=M5^2/L5	
=НОРМРАСП(C5;\$H\$4;\$H\$8;1)	=НОРМРАСП(D5;\$H\$4;\$H\$8;1)	=J6-I6	=K6*\$H\$2	=E5-L6	=M6^2/L6	
=НОРМРАСП(C6;\$H\$4;\$H\$8;1)	=НОРМРАСП(D6;\$H\$4;\$H\$8;1)	=J7-I7	=K7*\$H\$2	=E6-L7	=M7^2/L7	
=НОРМРАСП(C7;\$H\$4;\$H\$8;1)	=НОРМРАСП(D7;\$H\$4;\$H\$8;1)	=J8-I8	=K8*\$H\$2	=E7-L8	=M8^2/L8	
=НОРМРАСП(C8;\$H\$4;\$H\$8;1)	=НОРМРАСП(D8;\$H\$4;\$H\$8;1)	=J9-I9	=K9*\$H\$2	=E8-L9	=M9^2/L9	
=НОРМРАСП(C9;\$H\$4;\$H\$8;1)					Хи-квадрат расчетное	=СУММ(N3:N9)
					Хи-квадрат критическое	=ХИ2ОБР(0,05;5)

Рисунок 3.9 – Проверка выборки количества эритроцитов в моче у пациентов с пиелонефритом на нормальность по критерию Пирсона (режим формул)

Проверка на нормальность по критерию Пирсона						
F(x_i_min)	F(x_i_max)	p_i	n*p_i	p_i%-n*p_i	(p_i%-n*p_i)^2/n*p_i	
0,042739163	0,14542834	0,102689177	3,080675315	0,919324685	0,274341756	
0,146939448	0,349693944	0,202754496	6,082634885	0,917365115	0,13835431	
0,352138598	0,611775801	0,259637203	7,78911609	1,21088391	0,188242135	
0,614300028	0,829970243	0,215670215	6,470106455	-0,47010645	0,034157101	
0,831633766	0,947827561	0,116193795	3,485813863	-2,48581386	1,772690913	
0,948527272	0,989111705	0,040584434	1,217533011	0,782466989	0,502864878	
0,989299549	0,998483333	0,009183784	0,275513506	0,724486494	1,905099638	
0,998515519					Хи-квадрат расчетное	4,81575
					Хи-квадрат критическое	11,0705

Рисунок 3.10 - Проверка выборки количества эритроцитов в моче у пациентов с пиелонефритом на нормальность по критерию Пирсона (результаты расчетов)

Проверка гипотезы о нормально распределённой величине по критерию Шапиро - Уилка выполняется поэтапно.

Сначала формируется ранжированный ряд значений в порядке убывания.

Затем по выборочным данным вычисляется сумма квадратов центральных отклонений:

$$C^2 = \sum (X_i - \bar{X})^2, \quad (3.6)$$

где C^2 – обозначение суммы квадратов центральных отклонений; X_i – значение из ранжированного ряда; $\bar{X}$ – среднее значение выборки.

Определяется величина k . Для четных выборок она равна половине её объема, в противном случае вычисляется по формуле:

$$k = \frac{(n-1)}{2}, \quad (3.7)$$

где k – искомая величина; n – объем выборки.

Вычисляется вспомогательная величина B с использованием табличных коэффициентов a_{n-i+1} [20]:

$$B = \sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i). \quad (3.8)$$

Сомножители, стоящие в скобках, представляют собой разности между значениями, расположенными симметрично относительно концов ранжированного ряда.

Затем вычисляется величина W по формуле:

$$W = \frac{B^2}{C}. \quad (3.9)$$

Значение W служит статистикой для проверки гипотезы о нормальности распределения. Если W превышает критические значения W_{st} , то распределение допустимо рассматривать как нормальное. Критические значения W_{st} находят в зависимости от уровня значимости и объема выборки из специальных таблиц [18].

Во всех выборках, кроме выборки с количеством белка в общем анализе мочи, значение критерия Пирсона больше критического, а Шапиро-Уилка меньше критического (таблица 3.7), следовательно, гипотеза, состоящая в том,

что генеральная совокупность подчиняется нормальному закону распределения, принимается. Выборка с количеством белка в моче у пациентов с пиелонефритом и гломерулонефритом не подчиняется нормальному закону распределения.

Таблица 3.7 – Результаты проверки выборки на нормальность по критериям Пирсона и Шапиро-Уилка

	Пиелонефрит			Гломерулонефрит		
	Количество эритроцитов в моче	Количество лейкоцитов в моче	Уровень креатинина в крови	Количество эритроцитов в моче	Количество лейкоцитов в моче	Уровень креатинина в крови
Критерий Пирсона (n=30). Критическое значение 0,927						
Расчетное значение	4,82	3,82	5,5	0,929	0,94	0,93
Критерий Шапиро-Уилка (n=30). Критическое значение 11,07						
Расчетное значение	0,931	1,65	0,931	3,5	3,6	2,5

Таким образом, значения количества лейкоцитов и эритроцитов в моче, а также уровень креатинина в крови подчиняются нормальному распределению (таблица 3.7). Произведено вычисление обобщающих статистических характеристик, а именно, расчет объема выборки, выборочного среднего, дисперсии, стандартного отклонения, моды, медианы, асимметрии и эксцесса (таблица 3.4). Наибольшей изменчивостью из рассматриваемых признаков обладает уровень креатинина у больных гломерулонефритом (198 мкмоль/л).

3.4 Расчёт корреляции значений признаков

Коэффициент корреляции - количественная мера взаимосвязи (совместной изменчивости) двух переменных.

При корреляционной связи результативный признак y не всецело зависит от факторного признака x , а только частично. Возможно влияние других факторов ε :

$$y = \varphi | x | + \varepsilon . \quad (3.10)$$

Проявляются корреляционные связи не в единичных случаях, а в массе.

Коэффициент корреляции может изменяться от -1 до $+1$.

По силе связи коэффициент корреляции можно разделить на три группы: $r < 0,3$ — слабая связь; $0,3 < r < 0,7$ — умеренная связь ; $r > 0,7$ — сильная связь (50% и более от общей доли дисперсии).

Предпочтительным методом для расчета коэффициента корреляции является метод Пирсона [5]:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} , \quad (9)$$

где r_{xy} — обозначение коэффициента корреляции; x_i — значения, принимаемые переменной x ; y_i - значения, принимаемые переменной y ; $\bar{x}$ — средняя по x ; $\bar{y}$ — средняя по y .

Чем ближе к единице величина r_{xy} , тем связь сильнее, чем ближе к нулю, тем связь слабее. Если коэффициент корреляции имеет положительное значение, то связь прямая, если отрицательное — связь обратная.

В среде Excel существует готовая функция КОРРЕЛ для вычисления данного показателя. Расчеты для пар признаков у пациентов с болезнями почек и для здоровых пациентов приведены на рисунках 3.10, 3.11:

Коэффициент корреляции признаков у пациентов с пиелонефритом			
Лейкоциты и эритроциты	Лейкоциты и креатинин	Эритроциты и креатинин	
=КОРРЕЛ('Лейкоциты ПН'!A2:A31;'Эритроциты ПН'!A2:A31)	=КОРРЕЛ('Лейкоциты ПН'!A2:A31;'Креатинин ПН'!A2:A31)	=КОРРЕЛ('Эритроциты ПН'!A2:A31;'Креатинин ПН'!A2:A31)	

Рисунок 3.10 – Формулы в Excel для расчета коэффициента парной корреляции признаков у больных пиелонефритом

Коэффициент корреляции признаков у здоровых людей				
Лейкоциты и эритроциты	Лейкоциты и креатинин	Эритроциты и креатинин		
0,281	-0,271	0,041		
Коэффициент корреляции признаков у пациентов с пиелонефритом				
Лейкоциты и эритроциты	Лейкоциты и креатинин	Эритроциты и креатинин		
-0,180	-0,052	0,130		
Коэффициент корреляции признаков у пациентов с гломерулонефритом				
Лейкоциты и эритроциты	Лейкоциты и креатинин	Эритроциты и креатинин		
-0,169	-0,057	-0,003		

Рисунок 3.11 – Результаты расчета коэффициентов корреляции признаков

Таким образом, между рассматриваемыми диагностическими признаками наблюдается слабая теснота связи (коэффициент парной корреляции находится в пределах от 0,1 до 0,3). Следовательно, связь между диагностическими признаками незначительная и не представляет практического интереса.

3.5 Вычисление вероятностей интервалов значений признаков

Для составления диагностической модели необходимо разделить значения количественных признаков на интервалы. Число классов, на которые следует разбить вариацию признака, рассчитывается по формуле Стерджеса (3.2). Также необходимо определить границу первого классового интервала (3.3), а также верхнюю и нижнюю границы каждого классового интервала (рисунок 3.12).

Лейкоциты	Минимальное значение	Xd	Xu	fi	Значение интервала
70	0	-5,94277	5,932774	46	u11
55	Максимальное значение	5,942774	17,81832	17	u12
35	89	17,82832	29,70387	6	u13
73	λ	29,71387	41,58942	6	u14
11	11,88554863	41,59942	53,47497	3	u15
70	Xn	53,48497	65,36052	3	u16
65	-5,942774317	65,37052	77,24607	4	u17
27	k	77,25607	89,13161	5	u18
88	7,488085131	89,14161		90	

Рисунок 3.12 – Разделение на классовые интервалы признака «Количество лейкоцитов в моче»

Каждый признак у здоровых пациентов и пациентов с заболеваниями почек был разделен на восемь классовых интервалов.

Каждое значение признака попадает в определенный классовый интервал. Для определения принадлежности признака к какому-либо интервалу производим расчет с помощью логической функции «Если» в Excel: =ЕСЛИ(И(C2>=Лейкоциты!\$C\$2;C2<=Лейкоциты!\$D\$2),"u11";ЕСЛИ(И(C2>=Лейкоциты!\$C\$3;C2<=Лейкоциты!\$D\$3),"u12";ЕСЛИ(И(C2>=Лейкоциты!\$C\$4;C2<=Лейкоциты!\$D\$4),"u13";ЕСЛИ(И(C2>=Лейкоциты!\$C\$5;C2<=Лейкоциты!\$D\$5),"u14";ЕСЛИ(И(C2>=Лейкоциты!\$C\$6;C2<=Лейкоциты!\$D\$6),"u15";ЕСЛИ(И(C2>=Лейкоциты!\$C\$7;C2<=Лейкоциты!\$D\$7),"u16";ЕСЛИ(И(C2>=Лейкоциты!\$C\$8;C2<=Лейкоциты!\$D\$8),"u17";ЕСЛИ(И(C2>=Лейкоциты!\$C\$9;C2<=Лейкоциты!\$D\$9),"u18")))))))).

Функция сравнивает значение уровня лейкоцитов в моче (ячейка C2) с границами каждого классового интервала. При попадании признака в классовый интервал ячейке присваивается значение признака (например, u18).

Принадлежность значений признаков к классовым интервалам представлена на рисунке 3.13.

Название заболевания	Лейкоциты	Значение	Эритроциты	Значение	Креатинин	Значение
Пиелонефрит	70	u17	0	u21	146	u32
Пиелонефрит	55	u16	1	u22	99	u32
Пиелонефрит	35	u14	2	u23	118	u32
Пиелонефрит	73	u17	3	u23	97	u32
Пиелонефрит	11	u12	1	u22	64	u31
Пиелонефрит	70	u17	1	u22	90	u31
Пиелонефрит	65	u16	4	u24	104	u32
Пиелонефрит	27	u13	3	u23	68	u31
Пиелонефрит	88	u18	0	u21	84	u31
Пиелонефрит	70	u17	2	u23	53	u31
Пиелонефрит	28	u13	2	u23	81	u31
Пиелонефрит	14	u12	3	u23	64	u31
Пиелонефрит	45	u15	2	u23	71	u31
Пиелонефрит	25	u13	3	u23	133	u32
Пиелонефрит	36	u14	1	u22	98	u32
Пиелонефрит	33	u14	2	u23	81	u31
Пиелонефрит	28	u13	1	u22	90	u31

Рисунок 3.13 – Принадлежность значений признаков к различным классовым интервалам

Последний этап статистической обработки представляет собой экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.14).

Код интервала	Код пациен	Название заболевания	Лейкоциты	Значение
1	1	Пиелонефрит	70	u17
2	2	Пиелонефрит	55	u16
3	3	Пиелонефрит	35	u14
4	4	Пиелонефрит	73	u17
5	5	Пиелонефрит	11	u12
6	6	Пиелонефрит	70	u17
7	7	Пиелонефрит	65	u16
8	8	Пиелонефрит	27	u13
9	9	Пиелонефрит	88	u18
10	10	Пиелонефрит	70	u17
11	11	Пиелонефрит	28	u13
12	12	Пиелонефрит	14	u12
13	13	Пиелонефрит	45	u15
14	14	Пиелонефрит	25	u13
15	15	Пиелонефрит	36	u14
16	16	Пиелонефрит	33	u14
17	17	Пиелонефрит	28	u13

Рисунок 3.14 – Результат экспорта таблицы из MS Excel в MS Access

Таким образом, каждый количественный признак (количество лейкоцитов в моче, количество эритроцитов в моче, уровень креатинина в крови) был разделен на 8 классовых интервалов. В результате получены необходимые данные для формирования вероятностной диагностической модели.

3.6 Основные результаты и выводы по разделу 3

1 Модельные данные результатов лабораторных исследований у здоровых пациентов были сгенерированы в среде Excel по параметрам нормального распределения признаков. Методика статистической обработки данных включает в себя экспорт результатов анализов из MS Access в MS Excel, группировку данных, построение гистограммы распределения частот, расчет общих статистических характеристик, определение законов распределения, оценку корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.1).

2 Значения количества лейкоцитов и эритроцитов в моче, а также уровень креатинина в крови подчиняются нормальному распределению. (таблица 3.7). Произведено вычисление обобщающих статистических характеристик, а именно, расчет объема выборки, выборочного среднего, дисперсии, стандартного отклонения, моды, медианы, асимметрии и эксцесса (таблица 3.4). Наибольшей изменчивостью из рассматриваемых признаков обладает уровень креатинина у больных гломерулонефритом (198 мкмоль/л).

3 Между рассматриваемыми диагностическими признаками наблюдается слабая теснота связи (коэффициент парной корреляции находится в пределах от 0,1 до 0,3). Следовательно, связь между диагностическими признаками незначительная и не представляет практического интереса.

4 Каждый количественный признак (количество лейкоцитов в моче, количество эритроцитов в моче, уровень креатинина в крови) был разделен на 8 классовых интервалов. В результате получены необходимые данные для формирования вероятностной диагностической модели.

4 МЕТОДИКА РАСПОЗНАВАНИЯ БОЛЕЗНЕЙ ПОЧЕК

4.1 Методы распознавания болезней

В процессе развития теории распознавания возникало множество различных методов, которые позволяли достигать определенных целей, поэтому классификация методов распознавания заболеваний является условной и неоднозначной. Все методы распознавания подразделяются на две большие группы по типу моделей, положенных в их основу. В системах, основанных на формальных моделях, априорной информации достаточно для формирования набора признаков, необходимых для решения поставленных задач. По характеру информации о признаках данная группа подразделяется на две подгруппы – детерминистские и вероятностные методы (рисунок 4.1).

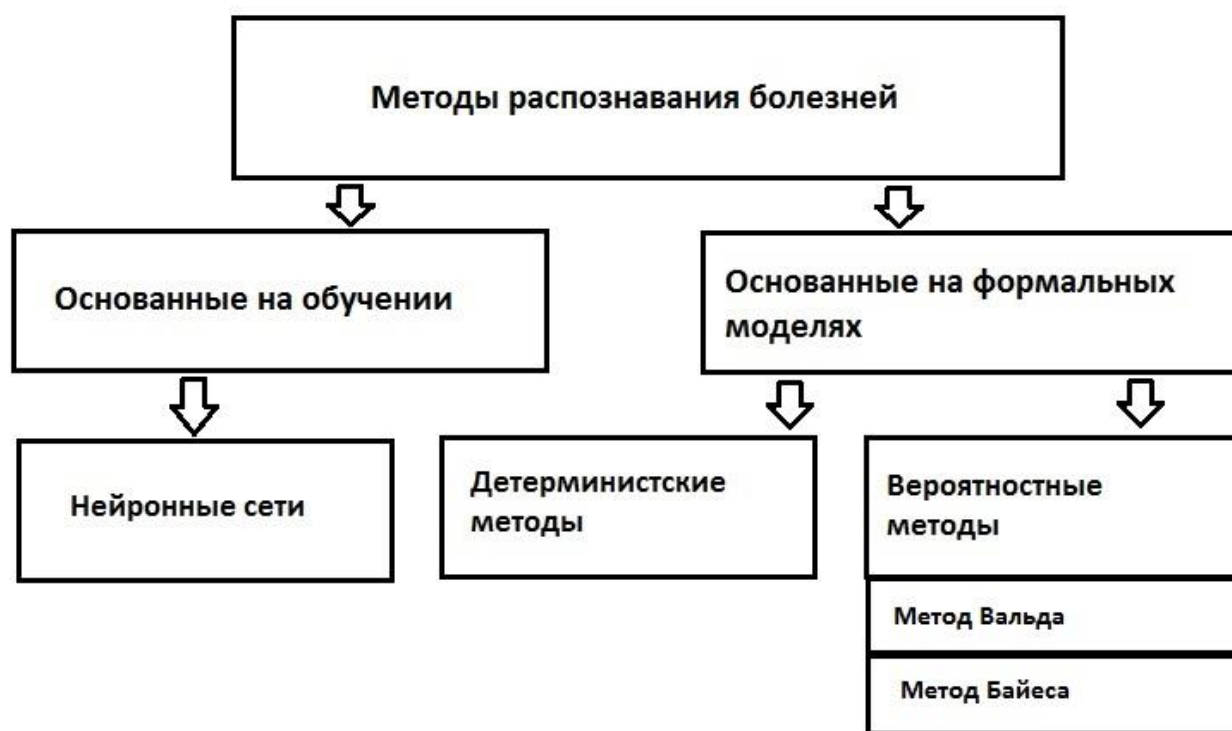


Рисунок 4.1 – Классификация методов распознавания болезней

Каждый из представленных методов имеет свои достоинства и недостатки (таблица 4.1)

Таблица 4.1 – Достоинства и недостатки различных методов распознавания образов

Метод распознавания	Достоинства	Недостатки
Нейронные сети	Могут решать не одну, а сразу несколько задач. Не программируется, а обучается на решение определенной задачи. Обладают высокой надежностью и устойчивостью к небольшим изменениям и повреждениям сети.	Сложны в реализации.
Детерминистские методы	Однозначно и достаточно просто определяют, что делать с объектом, если у него наблюдаются определенные внешние параметры.	При неспецифичных признаках не дают достоверной информации. Практически не работают со сложными динамическими многопараметрическими объектами.
Вероятностные методы	Подходят для работы с достаточно сложно устроенными объектами, например, человеком.	Необходим перебор всей выборки, достаточно чувствительны к артефактам.

Нейронные сети очень сложны в реализации, а детерминистские модели не позволяют описывать такой сложный объект исследования, как человек [26].

Для заболеваний почек многие симптомокомплексы и диапазоны возможных значений отдельных диагностических признаков совпадают, различаясь только частотами, с которыми при тех или иных заболеваниях встречаются отдельные признаки, их сочетания или количественные значения. Поэтому для информационной системы дифференциальной диагностики болезней почек больше подходят вероятностные методы. Область применения и ограничения вероятностных методов распознавания состояний здоровья представлены в таблице 4.2.

Таблица 4.2 – Область применения и ограничения вероятностных методов распознавания состояний здоровья

Метод	Область применения	Ограничения
Метод Байеса	Удобен для распознавания объектов или явлений в условиях неизменного признакового пространства, стабильном описании классов.	Необходим большой объем изначальной статистической информации, возможно «угнетение» редко встречающихся.
Метод последовательного анализа (метод Вальда)	Применяется для дифференциальной диагностики состояний. Число необходимых обследований не устанавливается заранее, а корректируется по мере необходимости для принятия решения с минимальной степенью риска. В методе последовательного анализа рассматриваемые отношения вероятностей признаков составляют не сразу, а в последовательном порядке; поэтому требуется меньшее число обследований [26].	Ограничением метода является необходимость дихотомического рассмотрения при решении многоальтернативных задач.

Особое место среди вероятностных методов распознавания заболеваний имеет метод Байеса. При достаточном объеме необходимой статистической информации его целесообразно использовать как один из наиболее надежных и эффективных методов [26].

Методикой распознавания состояния здоровья, основанной на методе Байеса, предусматривается сбор статистических данных результатов лабораторных исследований и установленных диагнозов, проектирование базы данных, статистическая обработка данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.2)

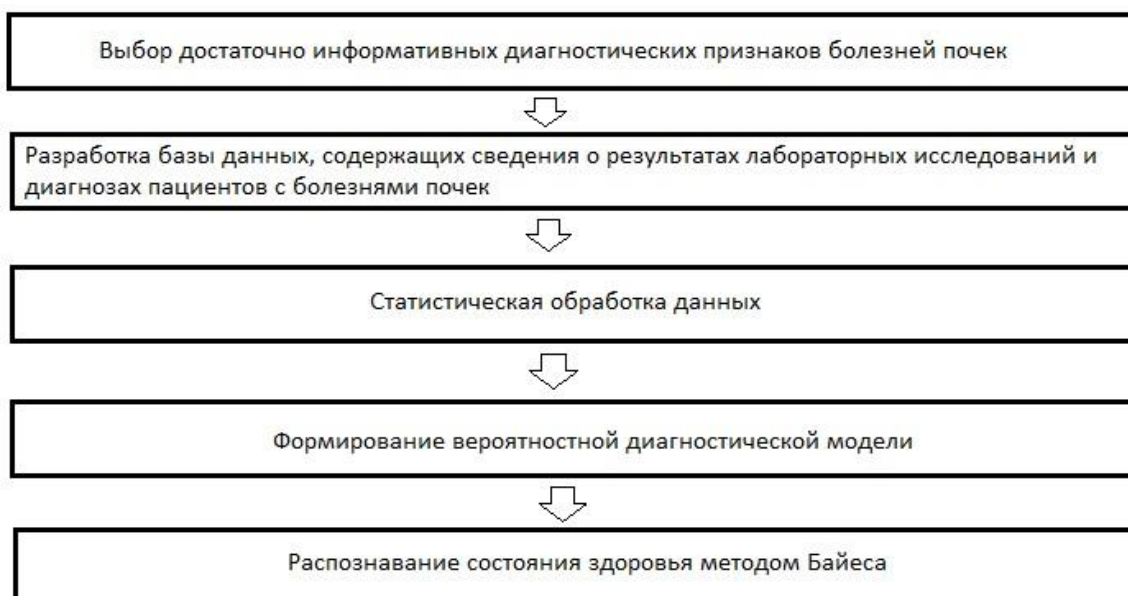


Рисунок 4.2 – Методика распознавания болезней почек

Таким образом, среди вероятностных методов наиболее подходящим является метод Байеса. При достаточном объеме статистических данных он является наиболее надежным и эффективным. Методика распознавания болезней почек, основанная на методе Байеса, предусматривает сбор статистических данных из карт пациентов с болезнями почек, выбор наиболее информативных количественных признаков заболеваний, проектирование базы данных, статистическую обработку данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья.

4.2 Диагностические модели болезней почек

Решение задач распознавания состояний здоровья по вероятностным диагностическим признакам решается на основе вероятностных диагностических моделей с помощью статистических методов.

Классы состояния здоровья пациентов - это случайные события, вероятности которых составляют множество:

$$P_E = \{p(E_i)\}, i = 1, 2, 3, \dots, N, \quad (4.1)$$

где P_E – множество вероятностей классов состояния здоровья;

$p(E_i)$ – вероятность одного из классов состояния здоровья;

E_i – один из классов состояния здоровья;

N – число классов состояния здоровья.

Состояние здоровья проявляется значениями вероятностных диагностических признаков. Диапазон значений вероятностного диагностического признака делится на интервалы. События попаданий значений диагностических признаков в разные интервалы несовместные. Класс состояния здоровья и попадание значения признака в интервал – совместные и зависимые события. Вероятности зависимых событий составляют множество:

$$P_U = \{p(u_{jl} / E_i)\}, j = 1, 2, \dots, K, l = 1, 2, \dots, L_j, i = 1, 2, \dots, N, \quad (4.2)$$

где P_U – множество вероятностей;

$p(u_{jl}/E_i)$ – вероятность принятия признаком значений из интервала u_{jl} в состоянии здоровья E_i ;

K – число признаков;

L_j – число интервалов значений признака;

N – число классов состояния здоровья.

Вероятностная диагностическая модель состояния здоровья задается взвешенным графом, вершинам и ребрам которого сопоставляются числовые характеристики (веса):

$$G = (E, U_{KL}, \varphi, P_E, P_U), \quad (4.3)$$

где G – взвешенный граф;

E – множество вершин, сопоставляемых классам состояния здоровья;

U_{KL} – множество вершин, сопоставляемых интервалам значений признаков;

φ – множество ребер, задающих отображение множества E в множество U_{KL} ;

P_E – множество вероятностей классов состояния здоровья;

P_U – множество вероятностей принятия признаками значений из выбранных интервалов. Модель представляется в табличной форме.

Вычисление вероятностей классов состояния здоровья и вероятностей принятия признаками значений из выбранных интервалов происходит по статистическим данным результатов обследований и диагнозов пациентов. Плотность распределения вероятностей значений диагностического признака определяется по данным из историй болезни. Вероятность значений признака из выбранного интервала вычисляется по формуле:

$$p(u_{jl} / E_i) = \int_{u_{j/n}}^{u_{j/k}} f(u_i / E_i) du_j \quad (4.4)$$

где $p(u_{jl}/E_i)$ – вероятность принятия признаком значений из интервала u_{jl} от $u_{j/n}$ до $u_{j/k}$; в состоянии здоровья E_i ;

$f(u_i / E_i)$ – плотность распределения вероятностей значений признака u_j в состоянии здоровья E_i .

Статистические оценки вероятностей диагнозов, вероятностей и условных вероятностей интервалов значений признаков вычисляются по следующим формулам:

$$p(E_i) = M_i / M; \quad (4.5)$$

$$p(u_{jl}) = \frac{1}{M} \sum_{i=1}^N M_{ij} = M_{jl} / M; \quad (4.6)$$

$$p(u_{jl} / E_i) = M_{ijl} / M_i, \quad (4.7)$$

где $p(E_i)$ – оценка вероятности диагноза E_i ;

M_i – число больных с диагнозом E_i ;

M – суммарное число больных;

$p(u_{jl})$ – оценка вероятности значений признака из интервала u_{jl} ;

M_{ijl} – число больных с диагнозом E_i , у которых обнаружены значения признака из интервала u_{jl} ;

M_{jl} – суммарное число больных, у которых обнаружены значения признака из интервала u_{jl} ;

$p(u_{jl}/E_i)$ – оценка условной вероятности значений признака из интервала u_{jl} при диагнозом E_i [25].

Таким образом, вероятностная диагностическая модель будет представлять собой таблицу, содержащую результаты вычислений статистических оценок вероятностей диагнозов и условных вероятностей интервалов значений признаков. Представление вероятностной диагностической модели в форме таблицы позволяет без затруднений реализовать математический аппарат распознавания методом Байеса на языке MatLab.

4.3 Программный комплекс формирования диагностической модели

Перед формированием диагностической модели в программе Matlab, необходимо создать запрос на выборку количества пациентов, относящихся к каждому классовому интервалу. Создаем перекрестный запрос через «Мастер запросов». В качестве наименований строк будут использоваться названия заболеваний, а в качестве названия столбцов – интервалы признаков (рисунок 4.3).

Название заболевания ▾	Итоговое значение ▾	u11 ▾	u12 ▾	u13 ▾	u14 ▾	u15 ▾	u16 ▾	u17 ▾	u18 ▾
Гломерулонефрит	30	19	11						
Обследование (Здоров)	30	27	3						
Пиелонефрит	30	3	6	6	3	3	4	5	

Рисунок 4.3 – Запрос с выборкой количества пациентов, относящихся к каждому классовому интервалу

Формирование вероятностной диагностической модели состоит из нескольких последовательных этапов, описанных в предыдущих разделах. Это: сбор данных лабораторных исследований и установленных диагнозов у пациентов с болезнями почек, формирование базы данных (раздел 2), Статистическая обработка данных (раздел 3), импорт данных и вычисление условных вероятностей интервалов значений признаков и вероятностей диагнозов (рисунок 4.4).

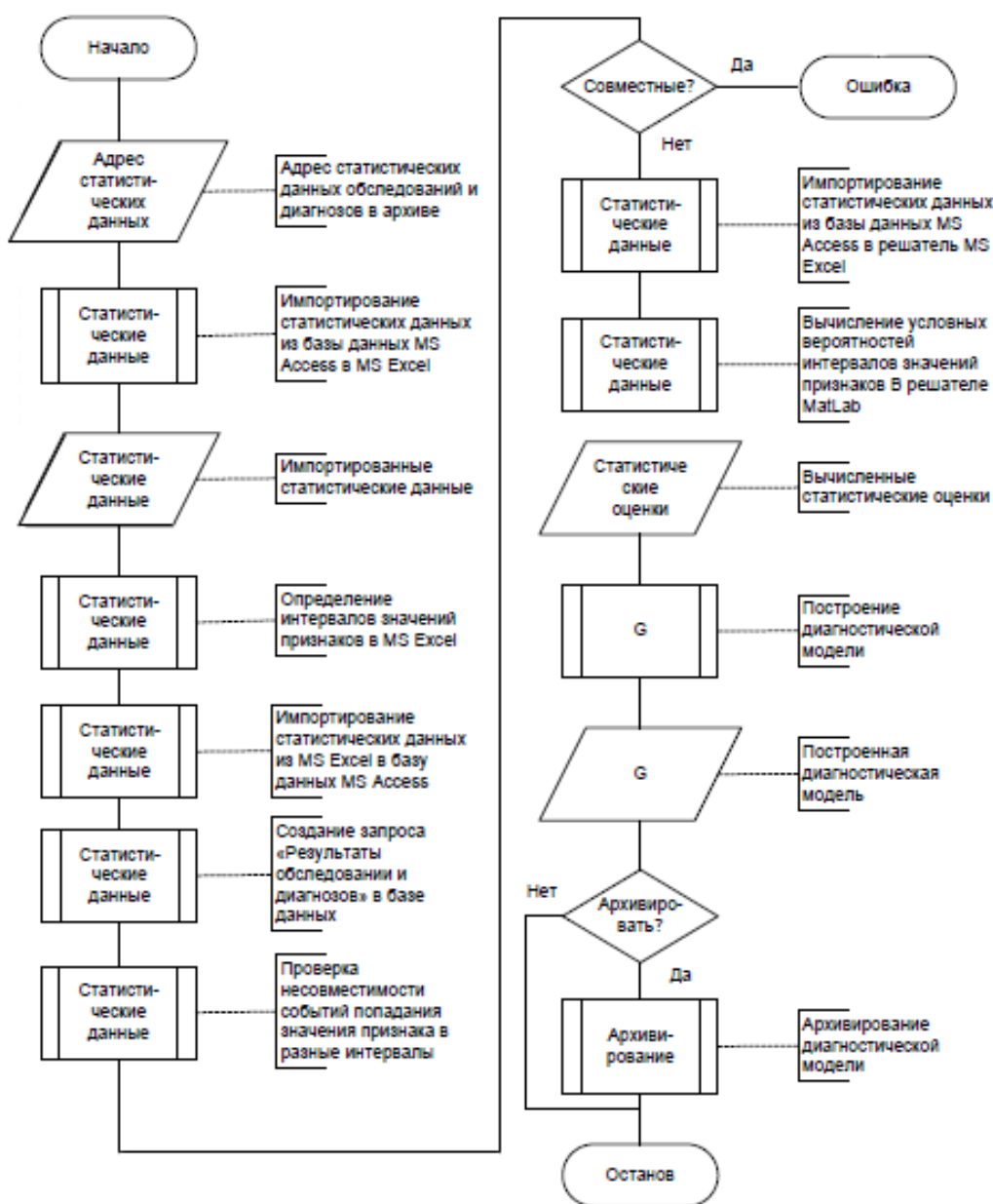


Рисунок 4.4 – Схема методики построения диагностической модели

Полный текст программного комплекса формирования диагностической модели представлен в приложении А.

Для того чтобы сформировать диагностическую модель, необходимо импортировать данные из MS Excel в Matlab. Импортируемые данные представляют собой матрицу Q.

Результат импорта данных из MS Excel в Matlab показан на рисунке 4.5.

		u11	u12	u13	u14	u15	u16	u17
Обследование(здоров)	30	27	3	0	0	0	0	0
Пиелонефрит	30	0	3	6	6	3	3	4
Гломерулонефрит	30	19	11	0	0	0	0	0

Рисунок 4.5 – Таблица, содержащая импортированные данные

Формирование вероятностной диагностической модели происходит в Matlab. Текст программы представлен в таблице 4.3.

Таблица 4.3 – Текст программы формирования вероятностной диагностической модели

Операторы	Комментарии
Оценка условных вероятностей интервалов значений признаков, $P(u_{jl}/E_i)$	
<code>n=size(Q,1)</code>	%определение количества строк матрицы Q
<code>k=size(Q,2);</code>	%определение количества столбцов матрицы Q
<code>for i =1:n</code>	%начало цикла от 1 до конца вектора n с шагом в единицу
<code>for j =2:k</code>	%начало цикла от 1 до конца вектора k с шагом в единицу
<code>Q(i,j)=Q(i,j)/Q(i,1)</code>	%расчет условных вероятностей интервалов значений признаков $p(u_{jl}/E_i)$
<code>end</code>	%окончание цикла
«Оценка вероятности диагноза $p(E_i)$ »	
<code>d=sum(Q(:,1))</code>	%расчет суммы пациентов (знаменатель формулы $p(E_i)=M_i/M$)
<code>for i =1:n;</code>	%начало цикла от 1 до n с шагом в единицу
<code>Q(i,1) = Q(i,1)/d</code>	%расчет оценки вероятности диагноза, $p(E_i)=M_i/M$
<code>end</code>	%окончание цикла

Результат работы программы формирования вероятностной диагностической модели представлен на рисунке 4.6.

		u11	u12	u13	u14	u15	u16
Обследование(здоров)	0.3333	0.9000	0.1000	0	0	0	0
Пиелонефрит	0.3333	0	0.1000	0.2000	0.2000	0.1000	0.1000
Гломерулонефрит	0.3333	0.6333	0.3667	0	0	0	0

Рисунок 4.6 – Результаты формирования вероятностной диагностической модели

Таким образом, формирование вероятностной диагностической модели состоит из нескольких последовательных этапов: сбора данных лабораторных исследований и установленных диагнозов у пациентов с болезнями почек, формирования базы данных и последующего формирования диагностической модели в MatLab. Формирование диагностической модели включает в себя вычисление условных вероятностей значений признака и оценку вероятностей диагнозов (рисунок 4.5). Диагностическая модель необходима для дальнейшего распознавания заболеваний почек методом Байеса.

4.4 Программный комплекс распознавания болезней почек

Распознавание заболеваний методом Байеса основывается на формуле Байеса (4.8), которая выводится как следствие из теоремы умножения вероятностей и формулы полной вероятности.

$$p(E_i / u_{j_l}) = \frac{p(E_i) p(u_{j_l} / E_i)}{\sum_{i=1}^N p(E_i) p(u_{j_l} / E_i)}, \quad (4.8)$$

где $p(E_i / u_{j_l})$ – апостериорная вероятность состояния здоровья E_i (после обнаружения значения признака из интервала u_{j_l} при обследовании пациента);

$p(E_i)$ – априорная вероятность состояния здоровья E_i (до обследования пациента);

$p(u_{ji}/E_i)$ – вероятность принятия признаком u_j значений из интервала I в состоянии здоровья E_i [25].

Распознавание заболеваний почек методом Байеса возможно по одному признаку (таблица 4.4) или по совокупности нескольких признаков.

Таблица 4.4 – Текст программы распознавания болезней почек методом Байеса по одному признаку на примере количества лейкоцитов в моче

Операторы	Комментарии
<pre> if b>=0 & b<=5; Q1=Q(:,2) end if b>=6 & b<=17; Q1=Q(:,3) end if b>=18 & b<=29; Q1=Q(:,4) end if b>=30 & b<=41; Q1=Q(:,5) end if b>=42 & b<=53; Q1=Q(:,6) end if b>=54 & b<=65; Q1=Q(:,7) end if b>=66 & b<=77; Q1=Q(:,8) end if b>=78 & b<=89; Q1=Q(:,9) end </pre>	%Определение принадлежности введенного пользователем значения к определенному интервалу. Значение сравнивается с нижней и верхней границей каждого классового интервала. При попадании в классовый интервал переменной Q1 присваивается значение определенного столбца
<pre> D1= Q(1,1)*Q1(1,:); D2= Q(2,1)*Q1(2,:); D3= Q(3,1)*Q1(3,:) </pre>	%Расчет числителя формулы Байеса для трех возможных состояний
<pre> for i=1:n r1=Q1(1,:)*Q(i,1); R1 = R1+r1 end for i=1:n r2=Q1(2,:)*Q(i,1); R2 = R2+r2 end for i=1:n r3=Q1(3,:)*Q(i,1); R3 = R3+r3 end </pre>	%Расчет знаменателя формулы Байеса для трех возможных состояний
<pre> F1=D1/R1; F2=D2/R2; F3=D3/R3 </pre>	%Расчет апостериорных вероятностей методом Байеса
<pre> if F1>F2 & F1>F3; e=msgbox ('Наиболее вероятным является состояние "Здоров" для введенного значения,лейкоциты в моче,кл в пз' ,'Результат','custom',myicon); end if F2>F1 & F2>F3; e=msgbox ('Наиболее вероятным является состояние "Пиелонефрит" для введенного значения,лейкоциты в моче,кл в пз','Результат','custom',myicon1); end if F3>F2 & F3>F1; e=msgbox ('Наиболее вероятным является состояние "Гломерулонефрит" для введенного значения,лейкоциты в моче,кл в пз','Результат','custom',myicon1); end </pre>	%Сравнение апостериорных вероятностей между собой и принятие решения о диагнозе

Определение апостериорных вероятностей по признакам количества эритроцитов в моче и креатинина в крови, а также определение апостериорных вероятностей по совокупности признаков производятся аналогичным способом.

На рисунке 4.7 представлены апостериорные вероятности диагнозов методом Байеса по одному признаку.

	Здоров	Пиелонефрит	Гломерулонефрит
$p(E_i)$	0	0.3333	0

Рисунок 4.7 – Результат расчета апостериорных вероятностей методом Байеса по одному признаку (признак – количество лейкоцитов в моче, введенное значение -60)

На рисунке 4.8 представлены апостериорные вероятности диагнозов методом Байеса по нескольким признакам

	Здоров	Пиелонефрит	Гломерулонефрит
$p(E_i)$	0.9446	0	0.0554

Рисунок 4.8 – Результат расчета апостериорных вероятностей методом Байеса по нескольким признакам (количество лейкоцитов в моче – 1, количество эритроцитов в моче – 1, уровень креатинина в крови – 80)

Таким образом, методика распознавания болезней почек методом Байеса состоит из ввода значения признака, определения интервала к которому относится введенный признак, определения апостериорных вероятностей трех возможных диагнозов по формуле Байеса, сравнение вероятностей и принятие диагностического решения (таблица 4.4). Использование нескольких диагностических признаков дает более высокую вероятность и, как следствие, более достоверные результаты. Для удобства использования разработанной программы необходимо создать понятный для пользователя интерфейс.

4.5 Интерфейс пользователя

Интерфейс пользователя разрабатывается при помощи среды GUIDE MatLab и состоит из панели с кнопками и таблицами для вывода результатов.

Кнопочная панель позволяет загрузить данные (кнопка «Загрузка»), произвести расчеты (кнопки: «Оценка условных вероятностей и интервалов значений признаков», «Оценка вероятности диагноза», «Метод Байеса по нескольким признакам», а также совокупность кнопок «Метод Байеса по одному признаку»), сохранить результаты (кнопка «Сохранить») (рисунок 4.9).

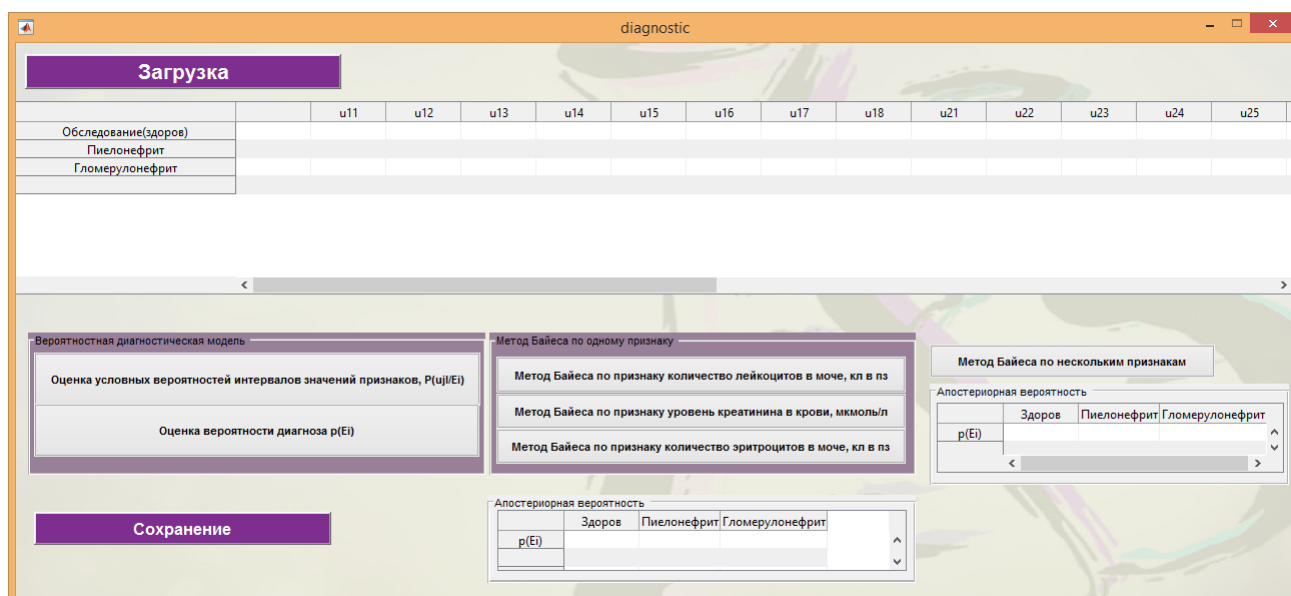


Рисунок 4.9 – Интерфейс пользователя

При нажатии на кнопку «Загрузка» пользователю предлагается выбрать таблицу с исходными данными (рисунок 4.10).

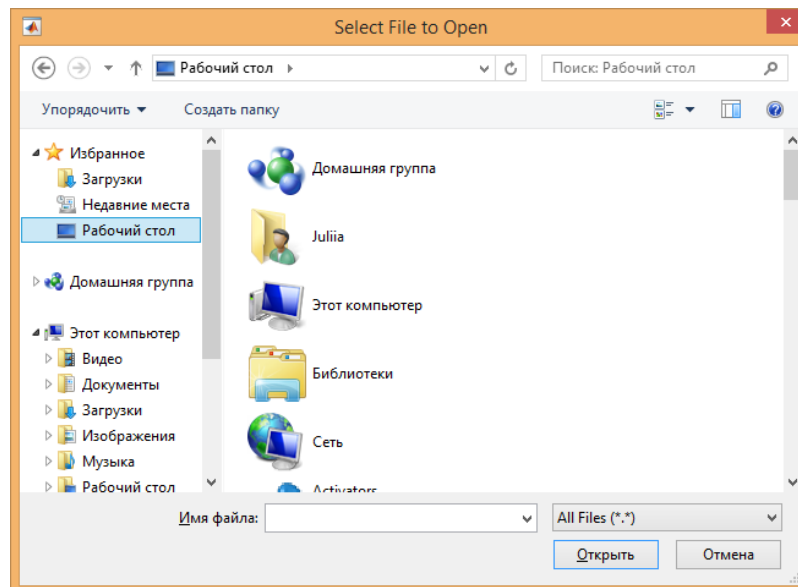


Рисунок 4.10 – Окно выбора таблицы исходных данных (кнопка «Загрузка»)

Результат загрузки данных появляется в таблице (рисунок 4.11).

Загрузка

	u11	u12	u13	u14	u15	u16	u17	u18	u21	u22	u23	u24	u25
Обследование(здоров)	30	27	3	0	0	0	0	0	21	9	0	0	0
Пилонефрит	30	0	3	6	6	3	4	5	19	10	1	0	0
Гломерулонефрит	30	19	11	0	0	0	0	0	4	5	6	5	5

Вероятностная диагностическая модель

Оценка условных вероятностей интервалов значений признаков, $P(u|E_i)$

Оценка вероятности диагноза $p(E_i)$

Метод Байеса по одному признаку

- Метод Байеса по признаку количество лейкоцитов в моче, кл в пз
- Метод Байеса по признаку уровень креатинина в крови, мкмоль/л
- Метод Байеса по признаку количество эритроцитов в моче, кл в пз

Метод Байеса по нескольким признакам

Апостериорная вероятность

	Здоров	Пилонефрит	Гломерулонефрит
$p(E_i)$			

Сохранение

Апостериорная вероятность

	Здоров	Пилонефрит	Гломерулонефрит
$p(E_i)$			

Рисунок 4.11 – Результат загрузки исходных данных в программу

Результат расчета вероятности диагноза и оценки условных вероятностей интервалов значений признаков представлен на рисунке 4.12.

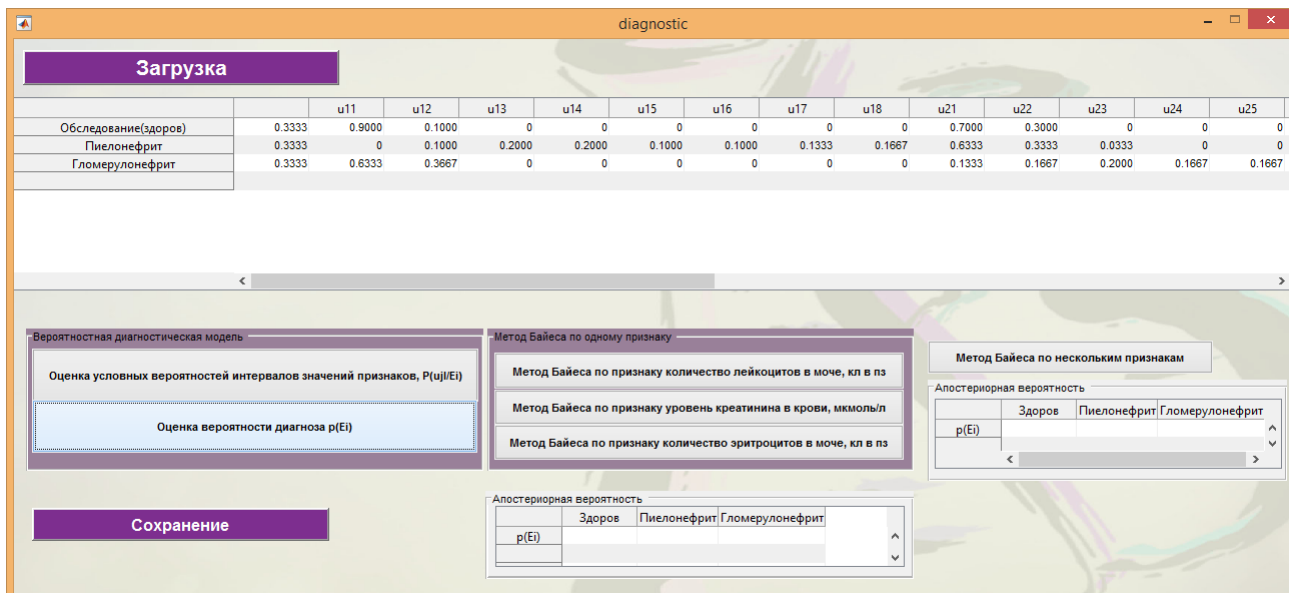


Рисунок 4.12 - Результат расчета вероятности диагноза и оценки условных вероятностей интервалов значений признаков

При нажатии на одну из группы кнопок «Метод Байеса по одному признаку» появляется диалоговое окно, которое просит ввести значение признака (рисунок. 4.13).

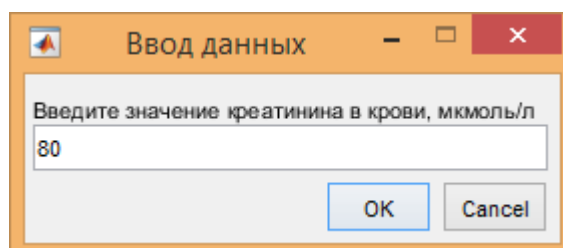


Рисунок 4.13 – Диалоговое окно ввода данных (кнопка «Метод Байеса по признаку уровень креатинина в крови»)

В результате получаем диагностическое заключение о здоровье пациента (рисунок 4.14) и апостериорные вероятности.

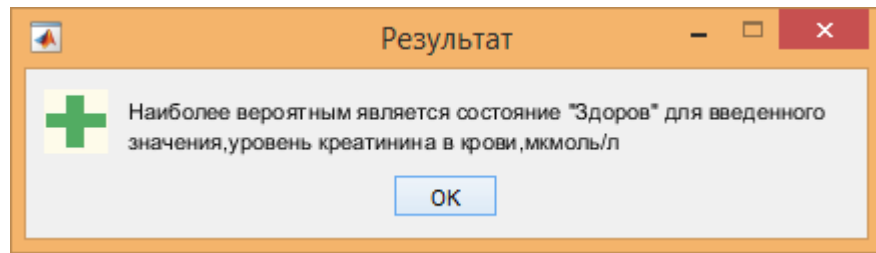


Рисунок 4.14 – Диагностическое заключение для введенного значения уровня креатинина в крови

При нажатии на кнопку «Метод Байеса по нескольким признакам», появляется диалоговое окно ввода данных для трех признаков (рисунок 4.15).

A screenshot of a Windows-style dialog box titled "Ввод данных" (Data Input). It contains three text input fields. The first is labeled "Введите значение лейкоцитов в моче, кл в пз" and contains the value "1". The second is labeled "Введите значение креатинина в крови, мкмоль/л" and contains the value "80". The third is labeled "Введите значение эритроцитов в моче, кл в пз" and contains the value "1". At the bottom right are "OK" and "Cancel" buttons.

Рисунок 4.15 – Диалоговое окно для ввода нескольких признаков

Диагностическим заключением для введенных значений является «Здоров».

Интерфейс пользователя после выполнения расчетов представлен на рисунке 4.16.

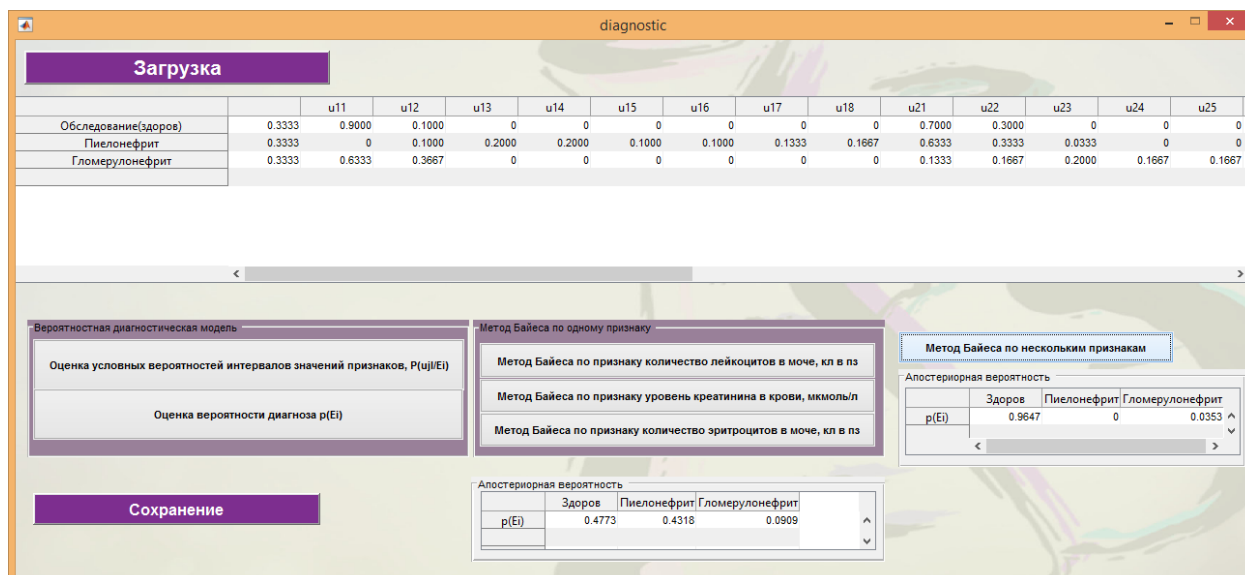


Рисунок 4.16 – Интерфейс пользователя после выполнения расчетов

Таким образом, создание интерфейса позволяет упростить взаимодействие пользователя с программой. Ввод всей необходимой информации для расчетов производится посредством диалоговых окон, расчеты производятся после нажатия пользователем соответствующих кнопок. Результаты проведенных расчетов выводятся в форме диагностического заключения при помощи диалоговых окон, а также в таблицы в виде вероятностей (таблица 4.15). Все это сводит к минимуму время необходимое для обучения пользователя.

4.6 Вычислительный эксперимент по распознаванию болезней почек

Проведем вычислительный эксперимент, в котором протестируем программно-методический комплекс распознавания болезней почек, на основе которого можно будет сделать вывод об эффективности информационной системы диагностики болезней почек. В качестве данных для эксперимента будем использовать данные лабораторных обследований здоровых пациентов и пациентов с болезнями почек, использованные нами для создания базы данных.

Результаты тестирования представлены в таблице 4.5.

Таблица 4.5 – Тестирование программно-методического комплекса по сочетанию диагностических признаков

Название болезни	Количество верно установленных диагнозов	Количество неверно установленных диагнозов
Пиелонефрит	26	4
Гломерулонефрит	28	2
Обследование (здоров)	30	0
Итого:	84	6

По проведенному эксперименту получилось, что было выявлено 54 истинно положительных результата заболеваний почек, 6 ложно отрицательных результата о здоровье пациента с болезнью почек, 30 истинно отрицательных результатов об отсутствии заболеваний у здоровых пациентов и 0 ложно положительных результатов.

Оценка эффективности предложенной методики проводилась путем вычисления прогностической значимости результатов по формулам (4.9) – (4.13) (таблица 4.6):

$$Ч = (ИП / (ИП + ЛО)) * 100\%; \quad (4.9)$$

$$С = (ИО / (ИО + ЛП)) * 100\%; \quad (4.10)$$

$$ПЦПТ = (ИП / (ИП + ЛП)) * 100\%; \quad (4.11)$$

$$ПЦОТ = (ИО / (ИО + ЛО)) * 100\%; \quad (4.12)$$

$$ПТ = ((ИП + ИО) / N) * 100\%, \quad (4.13)$$

где N – число наблюдений;

ИП – истинно положительные результаты;

ИО – истинно отрицательные результаты;

ЛП – ложноположительные результаты;

ЛО – ложноотрицательные результаты;

Ч – чувствительность;

С – специфичность;

ПЦПТ – предсказуемая ценность положительного теста;

ПЦОТ – предсказуемая ценность отрицательного теста;

ПТ – предсказуемая точность.

Таблица 4.6 – Оценка эффективности программно-методического комплекса распознавания болезней почек

Показатель	Значение, %
Чувствительность	90
Специфичность	100
Предсказуемая ценность положительного теста	100
Предсказуемая ценность отрицательного теста	83,3
Предсказуемая точность	93,3

Таким образом, анализ результатов расчетов (таблица 4.6) позволяет утверждать об эффективности информационной системы распознавания болезней почек. В связи с чем, может быть внедрена в медицинские Вузы с целью обучения студентов, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам. Повышение достоверности распознавания болезней почек возможно за счет увеличения количества используемых в системе диагностических признаков.

4.7 Основные результаты и выводы по разделу 4

1 Среди вероятностных методов наиболее подходящим является метод Байеса. При достаточном объеме статистических данных он является наиболее надежным и эффективным. Методика распознавания болезней почек, основанная на методе Байеса, предусматривает сбор статистических данных из карт пациентов с болезнями почек, выбор наиболее информативных количественных признаков заболеваний, проектирование базы данных, статистическую обработку данных обследований и диагнозов, формирование

диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.2).

2 Вероятностная диагностическая модель будет представлять собой таблицу, содержащую результаты вычислений статистических оценок вероятностей диагнозов и условных вероятностей интервалов значений признаков. Представление вероятностной диагностической модели в форме таблицы, позволяет без затруднений реализовать математический аппарат распознавания методом Байеса на языке MatLab.

3 Формирование вероятностной диагностической модели состоит из нескольких последовательных этапов: сбора данных лабораторных исследований и установленных диагнозов у пациентов с болезнями почек, формирования базы данных и последующего формирования диагностической модели в MatLab. Формирование диагностической модели включает в себя вычисление условных вероятностей значений признака и оценку вероятностей диагнозов (рисунок 4.5). Диагностическая модель необходима для дальнейшего распознавания заболеваний почек методом Байеса.

4 Методика распознавания болезней почек методом Байеса состоит из ввода значения признака, определения интервала к которому относится введенный признак, определения апостериорных вероятностей трех возможных диагнозов по формуле Байеса, сравнение вероятностей и принятие диагностического решения (таблица 4.4). Использование нескольких диагностических признаков дает более высокую вероятность и, как следствие, более достоверные результаты. Для удобства использования разработанной программы необходимо создать понятный для пользователя интерфейс.

5 Создание интерфейса позволяет упростить взаимодействие пользователя с программой. Ввод всей необходимой информации для расчетов производится посредством диалоговых окон, расчеты производятся после нажатия пользователем соответствующих кнопок. Результаты проведенных расчетов выводятся в форме диагностического заключения при помощи

диалоговых окон, а также в таблицы в виде вероятностей (таблица 4.15). Все это сводит к минимуму время необходимое для обучения пользователя.

6 Анализ результатов расчетов (таблица 4.6) позволяет утверждать об эффективности информационной системы распознавания болезней почек. В связи с чем, может быть внедрена в медицинские Вузы с целью обучения студентов, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам. Повышение достоверности распознавания болезней почек возможно за счет увеличения количества используемых в системе диагностических признаков.

ЗАКЛЮЧЕНИЕ

1 Несмотря на тенденцию к снижению заболеваемости населения болезнями почек, на данный момент она остается на достаточно высоком уровне (таблица 1.2). Большая распространенность этого заболевания приводит к увеличению нагрузки на медицинский персонал, за счет массового обращения пациентов. Основным этиологическим фактором развития пиелонефрита и гломерулонефрита являются микроорганизмы. (таблица 1.5, 1.6). Развитие заболеваний болезней почек происходит в три периода: период максимальной активности, период стихания, период ремиссии. Острые формы обоих заболеваний характеризуются увеличением проницаемости почечного фильтра и появлением в моче лейкоцитов, эритроцитов. Нарушение работы почечного фильтра также сопровождается накоплением в крови азотистых шлаков, в том числе и креатинина. Основными методами диагностики острого пиелонефрита и гломерулонефрита являются лабораторные методы. Так как общий анализ мочи входит в перечень исследований общей диагностики и позволяет быстро получить результаты, то в создании информационной системы мы количество лейкоцитов и эритроцитов, а также биохимический анализ крови, в частности, уровень креатинина (таблица 1.13). Эти показатели удобны для создания информационной системы, так как являются количественными признаками и появляются при обоих заболеваниях. При анализе литературы для данных показателей законы распределения и их характеристики выявлены не были. Следовательно, требуется провести анализ историй болезней пациентов с заболеваниями почек с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками. Распознавание болезней почек является трудоемким процессом, который происходит в условиях дефицита времени и зависит от опыта и квалификации врача - специалиста. Все эти факторы не исключают возможности возникновения диагностических ошибок.

Уменьшение трудоёмкости и увеличение достоверности постановки диагноза, может быть достигнуто путем разработки и внедрения информационных систем дифференциальной диагностики заболеваний.

2 Разрабатываемая информационная система по области применения относится к медицинским системам, используемым для диагностики заболеваний почек и обучения студентов медицинских ВУЗов, а по назначению к системам вычислительной диагностики. В существующих экспертных системах входными данными чаще всего являются субъективные жалобы пациентов, что имеет более низкую информативность по сравнению с данными лабораторных анализов. В рассматриваемых системах не заложена методика построения вероятностной диагностической модели на основе базы данных. В связи с этим разработка информационной системы, содержащей базу данных, а также решатель для формирования диагностической модели является актуальной задачей. Структура информационной системы распознавания болезней почек состоит из трех основных элементов: базы данных, решателя и интерфейса пользователя. Сформированная база данных, содержит результаты обследования и диагнозы пациентов с заболеваниями почек, собранные на базе нефрологического отделения ПОКБ им. Н.Н. Бурденко и урологического отделения ГБУЗ КМРБ. Принятие диагностических решений на основе опыта высококвалифицированных специалистов ведет к повышению достоверности постановки диагноза. Простота и доступность интерфейса созданной базы данных, обеспечивается за счет кнопочных форм MS Access (рисунок 2.21, 2.22). Простой пользовательский интерфейс позволяет снизить требования к квалификации медицинских работников в области информационных технологий.

3 Методика статистической обработки данных включает в себя экспорт результатов анализов, группировку данных, построение гистограммы распределения частот, расчет общих статистических характеристик,

определение законов распределения, оценку корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.1). Значения количества лейкоцитов и эритроцитов в моче, а также уровень креатинина в крови подчиняются нормальному распределению (таблица 3.7). Произведено вычисление обобщающих статистических характеристик (таблица 3.4). Наибольшей изменчивостью из рассматриваемых признаков обладает уровень креатинина у больных гломерулонефритом (198 мкмоль/л). Между рассматриваемыми диагностическими признаками наблюдается слабая теснота связи (коэффициент парной корреляции находится в пределах от 0,1 до 0,3). Следовательно, связь между диагностическими признаками незначительная и не представляет практического интереса. Каждый количественный признак был разделен на 8 классовых интервалов. В результате получены необходимые данные для формирования вероятностной диагностической модели.

4 Среди вероятностных методов наиболее подходящим является метод Байеса. При достаточном объеме статистических данных он является наиболее надежным и эффективным. Методика распознавания болезней почек, основанная на методе Байеса, предусматривает сбор статистических данных из карт пациентов с болезнями почек, выбор наиболее информативных количественных признаков заболеваний, проектирование базы данных, статистическую обработку данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.2). Вероятностная диагностическая модель представляет собой таблицу, содержащую результаты вычислений статистических оценок вероятностей диагнозов и условных вероятностей интервалов значений признаков. Формирование диагностической модели включает в себя вычисление условных вероятностей значений признака и оценку вероятностей диагнозов (рисунок 4.5). Методика

распознавания болезней почек методом Байеса состоит из ввода значения признака, определения интервала к которому относится введенный признак, определения апостериорных вероятностей трех возможных диагнозов по формуле Байеса, сравнение вероятностей и принятие диагностического решения (таблица 4.4). Использование нескольких диагностических признаков дает более высокую вероятность и, как следствие, более достоверные результаты. Наличие интерфейса позволяет упростить взаимодействие пользователя с программой и минимизировать время необходимое для его обучения (таблица 4.15). Анализ результатов вычислительного эксперимента (таблица 4.6) позволяет утверждать об эффективности информационной системы распознавания болезней почек. В связи с чем, может быть внедрена в медицинские Вузы с целью обучения студентов, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам. Повышение достоверности распознавания болезней почек возможно за счет увеличения количества используемых в системе диагностических признаков.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Суринов, А.Е. Федеральная служба государственной статистики (Росстат). Российский статистический ежегодник за 2017 год. Статистический сборник / А.Е. Суринов, Э.Ф. Баранов, Т.С. Безбородова, Н.С. Бугакова, М.И. Гальвановский. Москва, 2017. 688 с.
- 2 Кокорина, Е.П. Заболеваемость всего населения России в 2014 году. Статистические материалы / Е.П.Кокорина, В.И. Стародубов. Москва, 2015. 142 с.
- 3 Тареева, И.Е. Нефрология. Руководство для врачей / И.Е. Тареева. – Москва. Медицина, 2000. 688 с.
- 4 Мухина, Н.А. Внутренние болезни. Том 1 / Н.А. Мухина – Москва. Медицина, 2008. 671 с.
- 5 Князева Л.И. Внутренние болезни / Л.И. Князева, Л.А. Князева, И.И. Горайнов. КГМУ, 2013. 624 с.
- 6 Мухин Н.А. Пропедевтика внутренних болезней /Н.А. Мухин, В.С. Моисеев. – Москва. ГЭОТАР – Медиа, 2008. 857 с.
- 7 Батюшин, М.М. Нефрология. Ключи к трудному диагнозу / М.М. Батюшин. – Элиста. ЗАОр НПП «Джангар», 2009. 168 с.
- 8 Козлов, А. Н. Интеллектуальные информационные системы / А.Н. Козлов. — Пермь. ФГБУО Пермская ГСХА, 2013. — 278 с.
- 9 Научный портал о проблемах биомедицинской инженерии. Классификация информационных систем в медицине: [Электронный ресурс] // URL: <http://labdata.ru/node/13> (дата обращения: 19.03.2018)
- 10 Каталог статей по основным направлениям исследований в области искусственного интеллекта: [Электронный ресурс] // URL: <http://www.aiportal.ru/articles/> (дата обращения: 19.03.2018).
- 11 EasyDiagnosis: [Электронный ресурс] // URL: <http://www.easydiagnosis.com/> (дата обращения 20.03.2018).

12 Биотехнические вычислительно–измерительные системы. Экспертные системы для классификации болезней в медицинской диагностике [Электронный ресурс] // URL:<https://cyberleninka.ru/article/v/ekspertnye-sistemy-dlya-klassifikatsii-bolezney-v-meditsinskoy-diagnostike> (дата обращения 20.03.2018).

13 Аверьянова, Н.И. Пиелонефрит и гломерулонефрит у детей / Н.И. Аверьянова, И.А. Шипулина, А.Е. Жуйков, Н.Ю. Зарницына, Л.А. Кичигина, Е.А. Вельдер – Ухта: УГТУ, 2001. 179 с.

14 Жариков О. Г., Литвин А. А., Ковалёв В. А. Экспертные системы в медицине: [Электронный ресурс] // URL:<http://mednovosti.by/journal.aspx?article=4013> (дата обращения: 22.03.2018).

15 DXplain – система поддержки принятия решений: [Электронный ресурс] // URL: <http://www.mghlcs.org/projects/dxplain> (дата обращения 2.04.2018).

16 Экспертные системы. Классификация экспертных систем. Разработка простейшей экспертной системы: [Электронный ресурс] // URL: <http://ref.by/refs/67/14877/1.html> (дата обращения: 2.04.2018).

17 AIMedica – интернет-услуга по компьютерной диагностике заболеваний: [Электронный ресурс] // URL: <http://aimedica.ru/about/aimedica> (дата обращения: 2.04.2018).

18 Сиделев С. И. Математические методы в биологии и экологии: введение в элементарную биометрию: учебное пособие / С.И. Сиделев. – Ярославль. ЯрГУ, 2012. – 140 с.

19 Лакин Г. Ф. Биометрия: учебное пособие для биологических специальных вузов / Г.Ф.Лакин. Москва. Высшая школа, 2009. – 352 с.

20 Государственный стандарт Российской Федерации. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения: [Электронный ресурс] // URL: <http://docs.cntd.ru/document/gost-r-iso-5479-2002> (дата обращения: 4.02.2018).

21 Гиниятуллин К.Г. Статистическая обработка результатов научных исследований. Краткий конспект лекций / К.Г. Гиниятуллин. Казань. 2014, 218 с.

22 Библиотека методов и моделей; описательные статистики и статистические характеристики; критерии качества; критерий согласия Пирсона: [Электронный ресурс] // URL: http://help.prognoz.com/ru/mergedProjects/Lib/05_statistics/uimodelling_chitest.htm (дата обращения: 7.04.2018).

23 Портал знаний. Критерий Колмогорова – Смирнова: [Электронный ресурс] // URL: <http://statistica.ru/glossary/general/kriteriy-kolmogorova-smirnova/> (дата обращения 8.04.2018).

24 Приказ Министерства здравоохранения Российской Федерации от 19.12.2016 г. № 973н "Об утверждении типовых отраслевых норм времени на выполнение работ, связанных с посещением одним пациентом врача-кардиолога, врача-эндокринолога, врача-стоматолога-терапевта": [Электронный ресурс] // URL: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=LAW&n=210955&fld=134&dst=100001,0&rnd=0.3568275418827398#0> (дата обращения 18.04.2018).

25 Фандеев В.П. Статистические методы распознавания состояния здоровья: метод. указания к практическим занятиям по дисциплине «Теоретические основы кибернетики» / В.П. Фандеев. Пенза. ПГУ, 2016. 44 с.

26 Дмитриенко А.Г. Диагностика. Оценка состояния и прогнозирование остаточного ресурса технически сложных объектов / А. Г. Дмитриенко, А. В. Блинов, Д.В. Волков. Пенза, 2013. 62 с.

27 Лепский А.Е. Математические методы распознавания образов: курс лекций / А.Е. Лепский, А.Г. Броневиц. Таганрог: ТТИ ЮФУ, 2009. 155 с.

ПРИЛОЖЕНИЕ А

ПРОГРАММНЫЙ КОМПЛЕКС ФОРМИРОВАНИЯ ДИАГНОСТИЧЕСКОЙ МОДЕЛИ И РАСПОЗНОВАНИЯ БОЛЕЗНЕЙ ПОЧЕК МЕТОДОМ БАЙЕСА

Программный код загрузки исходных данных (кнопка «Загрузка»)

```
%объявление переменных глобальными
global Q T
T=imread('zmz1.jpg');
%диалоговое окно для открытия файла с содержимым текущего
каталога
[FName, PName] = uigetfile('*.*)
if ~ iQequal(FName, 0)
%Если пользователь выбрал файл, то происходит его считывание
FullName = Qtrcat(PName, FName);
%импортирование данных в виде матрицы
Q=importdata(FullName)
end
%Замена значений "NaN" на "0"
Q(iQnan(Q))=0
%вывод матрицы в таблицу
Qet(handleQ.uitable1, 'ViQible','on');
Qet(handleQ.uitable1, 'Data',Q);
```

Программный код формирования диагностической модели (кнопка
«Оценка условных вероятностей интервалов значений признаков $P(u_{jl}/E_i)$ »)

```
%объявление переменных глобальными
global Q n k
%определение количества строк
n=Qize(Q,1);
%определение количества столбцов
k=Qize(Q,2);
%расчет условных вероятностей интервалов значений признаков
p(ujl/Ei)
for i =1:n
for j =2:k
Q(i,j)=Q(i,j)/Q(i,1)
end
end
%вывод матрицы в таблицу
Qet(handleQ.uitable1, 'ViQible','on');
Qet(handleQ.uitable1, 'Data',Q);
%кнопка "Оценка вероятности диагноза p(Ei)"
function puQhbutton3_Callback(hObject, eventdata, handleQ)
%объявление переменных глобальными
global Q n k
%расчет суммы пациентов, знаменателя для формулы  $p(E_i)=M_i/M$ 
```

```

d=Qum(Q(:,1));
%расчет оценки вероятности диагноза,  $p(E_i)=M_i/M$ 
for i =1:n;
Q(i,1) = Q(i,1)/d;
end
%вывод матрицы в таблицу
Qet(handleQ.uitable1, 'ViQible','on');%вывод матрицы в
таблицу
Qet(handleQ.uitable1, 'Data',Q);

```

Программный код распознавания методом Байеса по одному признаку
(кнопка «Метод Байеса по признаку количество лейкоцитов в моче, кл в пз»)

```

%объявление переменных глобальными
global Q n k T cuQtom T1
%диалоговое окно ввода данных
w = inputdlg({'Введите значение количество лейкоцитов в
моче, кл в пз'})
a = Qtr2num(w{1})
%определение принадлежности введенного значения к
определенному
%интервалу
%интервал u11
if a>=0 & a<=5
S1=S(:,2)
end
%интервал u12
if a>=6 & a<=17
S1=S(:,3)
end
%интервал u13
if a>=18 & a<=29
S1=S(:,4)
end
%интервал u14
if a>=30 & a<=41
S1=S(:,5)
end
%интервал u15
if a>=42 & a<=53
S1=S(:,6)
end
%интервал u16
if a>=54 & a<=65
S1=S(:,7)
end
%интервал u17
if a>=66 & a<=77
S1=S(:,8)
end
%интервал u18

```

```

if a>=78 & a<=89
    S1=S(:,9)
end
%расчет числителя  $p(E_i) * p(u_{j1}/E_i)$  для состояния "Здоров"
D1= S(1,1)*S1(1,:);
%расчет числителя  $p(E_i) * p(u_{j1}/E_i)$  для состояния "Пиелонефрит"
D2= S(2,1)*S1(2,:);
%расчет числителя  $p(E_i) * p(u_{j1}/E_i)$  для состояния
"Гломерулонефрит"
D3= S(3,1)*S1(3,:);
R1=0;
R2=0;
R3=0;
%расчет знаменателя в формуле  $p(E_i/u_{j1})$  для состояния
"Здоров"
for i=1:n
    r1=S1(1,:)*S(i,1);
    R1 = R1+r1
end
%расчет знаменателя в формуле  $p(E_i/u_{j1})$  для состояния
"Пиелонефрит"
for i=1:n
    r2=S1(2,:)*S(i,1);
    R2 = R2+r2
end
%расчет знаменателя в формуле  $p(E_i/u_{j1})$  для состояния
"Гломерулонефрит"
for i=1:n
    r3=S1(3,:)*S(i,1);
    R3 = R3+r3
end
F1=D1/R1
F2=D2/R2
F3=D3/R3
F1(isnan(F1))=0
F2(isnan(F2))=0
F3(isnan(F3))=0
U11=[F1,F2,F3]
clc
%вывод значений апостериорных вероятностей в таблицу
set(handles.uitable9, 'Visible','on');
set(handles.uitable9, 'Data',U11);
%принятие решение о диагнозе
%если апостериорная вероятность здорового состояния больше
апостериорных вероятностей болезней почек, то диагноз "Здоров"
if F1>F2 & F1>F3
    e=msgbox ('Наиболее вероятным является состояние "Здоров"
для введенного значения,лейкоциты в моче,кл в пз'
, 'Результат', 'custom',myicon)
end

```

```

% если апостериорная вероятность пиелонефрита больше
апостериорных вероятностей гломерулонефрита и здорового
состояния, диагноз "пн"
if F2>F1 & F2>F3
myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние
"Пиелонефрит" для введенного значения,лейкоциты в моче,кл в
пз','Результат','custom',myicon1)
end
% если апостериорная вероятность гломерулонефрита больше
апостериорных вероятностей пиелонефрита и здорового состояния,
диагноз "гн"
if F3>F2 & F3>F1
myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние
"Гломерулонефрит" для введенного значения,лейкоциты в моче,кл в
пз','Результат','custom',myicon1)
end

```

Программный код распознавания методом Байеса по одному признаку для значений количества эритроцитов в моче и креатинина в крови аналогичен.

Программный код распознавания методом Байеса по сочетанию признаков (кнопка «метод Байеса по нескольким признакам»)

```

%объявление глобальных переменных
global Q n k T T1;
% диалоговое окно ввода данных
w = inputdlg({'Введите значение количества лейкоцитов в
моче,кл в пз' 'Введите значение эритроцитов в моче, кл в пз'
'Введите значение креатинина, мкмоль/л'}, 'Ввод данных')
% ввод значения количества лейкоцитов в моче
a = Qtr2num(w{1})
%определение принадлежности введенного значения к
определенному
%интервалу
%интервал u11
if a>=0 & a<=5
S1=S(:,2)
end
%интервал u12
if a>=6 & a<=17
S1=S(:,3)
end
%интервал u13
if a>=18 & a<=29
S1=S(:,4)

```

```

end
%интервал u14
if a>=30 & a<=41
    S1=S(:,5)
end
%интервал u15
if a>=42 & a<=53
    S1=S(:,6)
end
%интервал u16
if a>=54 & a<=65
    S1=S(:,7)
end
%интервал u17
if a>=66 & a<=77
    S1=S(:,8)
end
%интервал u18
if a>=78 & a<=89
    S1=S(:,9)
end
%ввод значения креатинина в крови, мкмоль/л
a1 = str2num(w{2
%определение принадлежности введенного значения к
определенному интервалу
%интервал u21
if a1>=0 & a1<=96.56
    S2=S(:,10)
end
% интервал u22
if a1>=96.57 & a1<=201.7
    S2=S(:,11)
end

%интервал u23
if a1>=201.71 & a1<=306.84
    S2=S(:,12)
end
% интервал u24
if a1>=306.85 & a1<=411.9813
    S2=S(:,13)
end
%интервал u25
if a1>=411.99 & a1<=517.1216
    S2=S(:,14)
end
% интервал u26
if a1>=517.132 & a1<622.26
    S2=S(:,15)
end
% интервал u27
if a1>=622.27 & a1<=727.4

```

```

        S2=S(:,16)
    end
    %интервал u28
    if a1>=727.41 & a1<=832.54
        S2=S(:,17)
    end
    %ввод значения количества эритроцитов в моче, кл в пз
    a2 = str2num(w{3})
    %определение принадлежности введенного значения ТГ к
определенному интервалу
    %интервал u31
    if a2>=0 & a2<=0.57
        S3=S(:,18)
    end
    %интервал u32
    if a2>=0.6 & a2<=1
        S3=S(:,19)
    end
    %интервал u33
    if a2>=2 & a2<=3
        S3=S(:,20)
    end
    %интервал u34
    if a2>=4 & a2<=4.2
        S3=S(:,21)
    end
    %интервал u35
    if a2>=4.2 & a2<=5.1
        S3=S(:,22)
    end
    %интервал u36
    if a2>=5.2 & a2<=6.1
        S3=S(:,23)
    end
    %интервал u37
    if a2>=6.6 & a2<=7.1
        S3=S(:,24)
    end
    %интервал u38
    if a2>=7.51 & a2<=9.1
        S3=S(:,25)
    end
    %Расчет числителя обобщенной формулы Байеса p(Ei/Uj1)
    X1=S(1,1)*S1(1,1)*S2(1,1)*S3(1,1)
    X2=S(2,1)*S1(2,1)*S2(2,1)*S3(2,1)
    X3=S(3,1)*S1(3,1)*S2(3,1)*S3(3,1)
    Y1=0
    %Расчет знаменателя обобщенной формулы Байеса p(Ei/Uj1)
    for i=1:n
        y1=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
        Y1 = Y1+y1
    end

```

```

W1=X1/Y1
Y2=0
for i=1:n
    y2=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
Y2 = Y2+y2
end
W2=X2/Y2
Y3=0
for i=1:n
    y3=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
Y3 = Y3+y3
end
W3=X3/Y3
W=[W1,W2,W3]
W(isnan(W))=0
%вывод рассчитанных апостериорных вероятностей в таблицу
set(handles.uitable6, 'Visible','on');
set(handles.uitable6, 'Data',W);
%принятие решение о диагнозе
if W1>W2 & W1>W3 % если апостериорная вероятность здорового
состояния больше апостериорных вероятностей болезней почек,
диагноз "Здоров"
    e=msgbox ('Наиболее вероятным является состояние "Здоров" '
, 'Результат', 'custom',myicon);
end
if W2>W1 & W2>W3 % если апостериорная вероятность
пиелонефрита больше апостериорных вероятностей гломерулонефрита и
здорового состояния, диагноз "Пиелонефрит"
    myicon1=imread('zmkr.jpg');
    e=msgbox ('Наиболее вероятным является состояние
"Пиелонефрит" ', 'Результат', 'custom',myicon1);
end
if W3>W2 & W3>W1 % если апостериорная вероятность
гломерулонефрита больше апостериорных вероятностей пиелонефрита
и здорового состояния, диагноз "Гломерулонефрита"
    myicon1=imread('zmkr.jpg');
    e=msgbox ('Наиболее вероятным является состояние
"Гломерулонефрит" ', 'Результат', 'custom',myicon1);
end
end

```