

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

Факультет (институт)

Кафедра

ЛФ МИ МКиИ
(наименование) (наименование)

Направление подготовки (специальность) 300503 Медицинская кибернетика
(код и наименование)

ДИПЛОМНАЯ РАБОТА

на тему

Информационная система распознавания сахарного диабета

Студент

(подпись, дата)

(фио полностью)

Руководитель/научный руководитель

(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 _____

(подпись, дата)

(фамилия, инициалы)

Рецензент _____

(должность, место работы)

(подпись, дата)

(фамилия, инициалы)

Работа допущена к защите (протокол заседания кафедры от _____ № ____)

Заведующий кафедрой _____

(подпись)

(фамилия, инициалы)

Работа защищена с отметкой _____ (протокол заседания ГЭК от _____ № ____)

Секретарь ГЭК

(подпись)

(фамилия, инициалы)

Пенза 2017

Задание 1 лист

Задание 2 лист

РЕФЕРАТ

Пояснительная записка на тему «Информационная система распознавания сахарного диабета» содержит 151 лист, 60 рисунков, 24 таблицы, 41 источников информации, 2 приложения.

САХАРНЫЙ ДИАБЕТ, РАСПОЗНАВАНИЕ ЗАБОЛЕВАНИЙ, ИНФОРМАЦИОННАЯ ЭКСПЕРТНАЯ СИСТЕМА, БАЗА ДАННЫХ ПАЦИЕНТОВ И БОЛЕЗНЕЙ, ДИАГНОСТИЧЕСКАЯ СТАТИСТИЧЕСКАЯ МОДЕЛЬ, МЕТОД БАЙЕСА

Цель работы – разработка элементов информационной системы распознавания сахарного диабета статистическими методами, обеспечивающей повышение достоверности и снижение продолжительности и трудоемкости постановки диагноза.

Результатами работы являются: модель базы данных пациентов с сахарным диабетом, диагностическая модель двух форм сахарного диабета применительно к периоду клинических проявлений, алгоритм и программный комплекс распознавания сахарного диабета по методу Байеса.

Информационная система предназначена для внедрения в диагностических центрах и в учебный процесс медицинского института.

					ПензГУ 300503 – 12ДР11.1.09. ПЗ				
Изм.	Лист	№ докум.	Подпись	Дата	Информационная система распознавания сахарного 4 диабета	Лит.	Лист	Листов	
Разраб.		Тимербулатов							
Провер.		Фандеев В.П.					4	151	
Реценз.						Кафедра МКиИ, гр.11ЛК1			
Н. Контр.		Янкина Н.Н.							
Утверд.		Янкина Н.Н.							

СОДЕРЖАНИЕ

	с.
ВВЕДЕНИЕ	8
1 ДИАГНОСТИКА САХАРНОГО ДИАБЕТА.....	11
1.1 Эпидемиология болезни.....	11
1.2 Этиология болезни	18
1.3 Патогенез болезни	22
1.4 Диагностические признаки болезни.....	29
1.5 Методика распознавания болезни (обоснование потребности в информационной системе)	34
1.6 Основные результаты и выводы	36
2 ИНФОРМАЦИОННАЯ СИСТЕМА РАСПОЗНАВАНИЯ САХАРНОГО ДИАБЕТА.....	40
2.1 Назначение, классификация и структура информационных систем распознавания болезней.....	40
2.2 Информационные системы распознавания болезней (примеры)	46
2.3 Структура информационной системы распознавания сахарного диабета.....	49
2.4 База данных пациентов с заболеванием сахарным диабетом	54
2.5 Пользовательский интерфейс базы данных	63
2.6 Основные результаты и выводы	66
3 ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ ОБСЛЕДОВАНИЯ В EXCEL	70
3.1 Построение вариационных рядов значений признаков.....	70
3.2 Вычисление характеристик распределений	78
3.3 Расчёт корреляции значений признаков	86

3.4 Вычисление вероятностей интервалов значений признаков	88
3.5 Основные результаты и выводы	91
4 МЕТОДИКА РАСПОЗНАВАНИЯ САХАРНОГО ДИАБЕТА	93
4.1 Методы распознавания болезней.....	93
4.2 Диагностические модели сахарного диабета.....	99
4.3 Методика формирования диагностической модели	102
4.4 Методика распознавания сахарного диабета методом Байеса.....	106
4.5 Интерфейс пользователя	111
4.6 Вычислительный эксперимент по распознаванию сахарного диабета	120
4.7 Основные результаты и выводы	122
ЗАКЛЮЧЕНИЕ	126
ПРИЛОЖЕНИЕ А.РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ВЫБОРОК	135
ПРИЛОЖЕНИЕ Б.ПРОГРАММНЫЙ КОМПЛЕКС ФОРМИРОВАНИЯ ДИАГНОСТИЧЕСКОЙ МОДЕЛИ И РАСПОЗНОВАНИЯ САХАРНОГО ДИАБЕТА МЕТОДОМ БАЙЕСА	144

ПЕРЕЧЕНЬ ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ

БЗ – База знаний;

ВОЗ – Всемирная организация здравоохранения;

ГКФ – Главная кнопочная форма;

ИЗСД – Инсулинозависимый сахарный диабет;

ИНСД – Инсулинонезависимый сахарный диабет;

ИМС – Информационная медицинская система;

ЛПУ – Лечебно-профилактические учреждения;

НИИ – Научно-исследовательский институт;

СД – Сахарный диабет;

СД 1 – Сахарный диабет 1 типа;

СД 2 – Сахарный диабет 2 типа;

СУБД – Система управления базами данных;

ЭВМ – Электронно-вычислительная машина;

ЭС – Экспертная система;

ADA – Американская диабетологическая ассоциация;

HbA1c – Гликозилированный гемоглобин;

IDF – Международный фонд диабета.

ВВЕДЕНИЕ

Среди всего разнообразия задач, возникающих перед практическими врачами, остро стоит вопрос о качественной диагностике сахарного диабета (СД), т. к. среди болезней эндокринных органов по частоте СД занимает первое место.

Согласно последним данным IDF 8,3% взрослого населения – 382 млн. человек – страдают СД. По прогнозам, менее чем через 25 лет количество людей с этим заболеванием достигнет 592 млн. Распространенность сахарного диабета – существенный фактор роста числа сердечно-сосудистых болезней, которые развиваются у большинства больных сахарным диабетом. Сахарный диабет в настоящее время является главной причиной слепоты. В группе больных диабетом гангрена встречается в 20–30 раз чаще, чем среди лиц, не страдающих этим заболеванием. Среди причин смерти сахарный диабет занимает третье место после сердечно-сосудистых и онкологических заболеваний (в 2013 г. СД стал причиной 5,1 млн. смертей). Человеческие и финансовые потери от СД непомерно высоки, затраты здравоохранения на это заболевание в 2013 г. составили 548 млрд. долларов США (11% от общемировых затрат) [1]. Ежегодные прямые расходы Российской Федерации, связанные с сахарным диабетом, составляют 375 миллиардов рублей (12,4 миллиардов долларов).

Диагностика сахарного диабета является трудоёмким процессом, в условиях массового распространения требующим привлечения большого числа высококвалифицированных специалистов, не исключающим ошибки. Снижение трудоёмкости и повышение достоверности постановки диагноза, ограничение требований к квалификации специалистов, достигается разработкой и внедрением информационных технологий распознавания заболеваний.

Повышение достоверности диагноза достигается, прежде всего, за счёт принятия решений на основе опыта постановки диагноза высококвалифицированных специалистов, накопленного в базе данных, содержащей результаты обследований и достоверные диагнозы пациентов, и обучения информационной системы, применения методов распознавания образов. Снижение трудоёмкости постановки диагноза достигается оптимизацией обследования (выбора состава диагностических признаков и очередности обследования) пациентов, автоматизацией принятия решений. Ограничение требований к опыту специалистов достигается за счёт принятия решений, основанных на опыте высококвалифицированных специалистов. Ограничение требований к квалификации специалистов в области компьютерных технологий достигается за счёт применения «дружественного» к специалисту графического интерфейса пользователя.

Ведущими учеными в области ИС являются С.А. Гаспарян, Г.А. Хай, Г.С. Лебедев, Ю.Ю. Мухин, А.П. Столбов, Б.А. Кобринский и др. в работах которых подробно рассматриваются вопросы, касающиеся ИС распознавания болезней [2,3,4,5]. Созданием медицинских информационных систем занимаются в настоящее время довольно большое количество разработчиков (СП.АРМ, БАРС-групп, КОРУС-Консалтинг, «Акросс-Инжиниринг») [5,6].

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.

Объектом исследования являются информационные технологии и системы распознавания сахарного диабета.

Предмет исследования – модели, методы и алгоритмы статистического распознавания сахарного диабета.

Цель работы – разработка элементов информационной системы, реализующей информационную технологию распознавания сахарного диабета статистическими методами, обеспечивающей снижение трудоёмкости, повышение достоверности постановки диагноза и ограничение требований к квалификации специалиста.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) осуществить нозологический анализ сахарного диабета;
- 2) осуществить разработку элементов базы данных, позволяющей накопить опыт квалифицированных специалистов, систематизировать достоверные факты о сахарном диабете, упорядочить результаты обследований для удобства их обработки;
- 3) на основе данных, содержащихся в базе данных, сформировать достоверную диагностическую модель;
- 4) реализовать методику распознавания заболеваний с использованием статистического метода Байеса, разработать необходимые алгоритмы и создать программный комплекс распознавания сахарного диабета.

1 ДИАГНОСТИКА САХАРНОГО ДИАБЕТА

1.1 Эпидемиология болезни

Формирование базы данных информационной системы распознавания болезней и диагностической модели основываются на сборе и анализе сведений о заболеваниях. Сведения о распознаваемых патологиях структурируются следующим образом:

- 1) перечень (номенклатура) болезней;
- 2) описание болезней;
- 3) перечень симптомов заболеваний.

По Международной классификации болезней десятого пересмотра (МКБ 10) [8], сахарный диабет (E10–E14) входит в группу «Болезни эндокринной системы, расстройства питания и нарушения обмена веществ» (E00–E90) и включают в себя следующие рубрики (таблица 1.1):

Таблица 1.1 – Классификация сахарного диабета по МКБ 10

Название заболевания	Код заболевания
Инсулинозависимый сахарный диабет (сахарный диабет 1 типа)	E10
Инсулиннезависимый сахарный диабет (сахарный диабет 2 типа)	E11
Сахарный диабет, связанный с недостаточностью питания	E12
Другие уточненные формы сахарного диабета	E13
Сахарный диабет неуточненный	E14

Перечень распознаваемых заболеваний представлен в таблице 1.2.

Таблица 1.2 – Перечень форм сахарного диабета

Название заболевания	Код заболевания
Инсулинозависимый сахарный диабет (сахарный диабет 1 типа)	E10
Инсулиннезависимый сахарный диабет (сахарный диабет 2 типа)	E11
Следующие четвертые знаки используются с рубриками E10–E11	
С комой	.0
С кетоацидозом	.1
С поражением почек	.2
С поражением глаз	.3
С неврологическими осложнениями	.4
С нарушениями периферического кровообращения	.5
С другими уточненными осложнениями	.6
С множественными осложнениями	.7
С неуточненными осложнениями	.8
Без осложнений	.9

Описания заболеваний составлены по следующим критериям:

- 1) определение;
- 2) эпидемиология;
- 3) этиология и классификация;
- 4) патогенез;
- 5) диагностические признаки;
- 6) методика распознавания болезни.

В последние годы практически во всех странах мира отмечается неуклонный рост заболеваемости и распространенности сахарного диабета, что позволило зарубежным авторам квалифицировать эти процессы как новую эпидемию неинфекционного характера конца XX - начала XXI века - эпидемию диабета [9]. Согласно последним данным IDF (Всемирной Федерации Диабета) 8,3% взрослого населения – 382. млн. человек – страдают СД. По прогнозам, менее чем через 25 лет количество людей с этим заболеванием достигнет 592 млн. Количество больных сахарным диабетом постоянно увеличивается в связи с ростом численности и возраста населения, урбанизации территории, увеличения распространенности ожирения и малоподвижного образа жизни.

Согласно опубликованным данным IDF и ВОЗ, Россия входит в пятерку стран-лидеров по темпам роста и распространённости диабета (рисунок 1.1 – 1.2) [10].

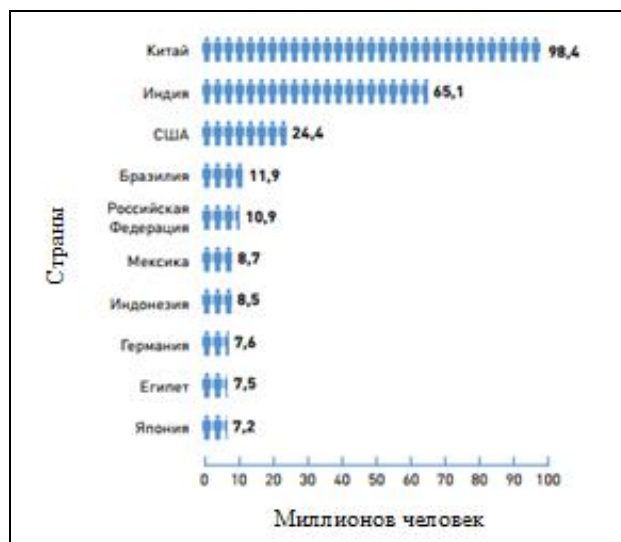


Рисунок 1.1 – Страны с наибольшим числом людей, страдающих сахарным диабетом (возраст 20 – 79 лет), 2013 г. [1]



Рисунок 1.2 – Рост распространённости сахарного диабета в России за период 2000 – 2015 (Данные ГосРегистра СД 2000 – 2015) [11]

Заболеваемость сахарным диабетом в России продолжает увеличиваться. Так рост численности больных в период с 2000 по 2011 гг. составил 1,525 млн. человек. По прогнозам экспертов ЭНЦ МЗ РФ к 2030 году ожидается рост численности больных до 5,81 млн. человек. Прирост по сравнению с 2000 г. составит 3,743 млн. человек, т.е. увеличиться в 2,81 раза (рисунок 1.3).



Рисунок 1.3 – Прогноз распространённости сахарного диабета в Российской Федерации [24]

Клинико-эпидемиологический мониторинг сахарного диабета в Российской Федерации осуществляется посредством Государственного регистра больных СД, методологическим и организационным референс-центром которого является ФГБУ Эндокринологический научный центр Минздрава России. На основании данных государственного регистра СД хотелось бы упомянуть о том, что в Пензенской области сахарный диабет наблюдается у 3266 на 100 тыс. человек, что превышает среднюю величину по РФ, которая составляет 2724 на 100 тыс. человек. К тому же, для Пензенского региона характерны высокие показатели смертности при СД, в рейтинге субъектов область занимает пятое место (рисунок 1.4).

Рейтинг	Субъект федерации	Все типы СД	СД 1 типа	СД 2 типа
	В среднем по РФ	74,05	2,72	70,81
1	Нижегородская область	149,51	3,49	143,92
2	Алтайский край	133,27	3,70	127,49
3	Воронежская область	132,56	5,73	124,12
4	Белгородская область	131,91	4,16	125,00
5	Пензенская область	129,55	4,06	124,96
6	Брянская область	126,37	4,81	120,97
7	Оренбургская область	126,31	5,15	120,60
8	Свердловская область	116,03	4,01	111,76
9	Ивановская область	115,67	4,26	108,76
10	Пермский край	108,72	2,92	105,18

Рисунок 1.4 – Показатели смертности при сахарном диабете на 100 тыс. населения в регионах РФ, 2015 г. (Данные ГосРегистра СД 2000 – 2015) [11]

Распространенность сахарного диабета – существенный фактор роста числа сердечно-сосудистых болезней, которые развиваются у большинства больных сахарным диабетом. Сахарный диабет в настоящее время является главной причиной слепоты. В группе больных диабетом гангрена встречается в 20–30 раз чаще, чем среди лиц, не страдающих этим заболеванием. Среди причин смерти сахарный диабет занимает третье место после сердечно-сосудистых и онкологических заболеваний (в 2013 г. СД стал причиной 5,1 млн. смертей) [1].

Финансовые потери от СД непомерно высоки, затраты здравоохранения в 2013 г. на это заболевание составили 548 млрд. долларов США (11% от общемировых затрат) [1]. Ежегодные прямые расходы Российской Федерации, связанные с сахарным диабетом, составляют 375 миллиардов рублей (12,4 миллиардов долларов). Для СД 1 типа прямые расходы для взрослых составили 82,4 млрд. рублей (2,764 млрд. долларов); для СД 2 типа – 291 млрд. рублей (9,7 млрд. долларов).

Для проектирования диагностической модели информационной системы распознавания СД необходимы данные об априорных вероятностях сахарного диабета 1 и 2 типа. Для этого проведен анализ литературных

источников по эпидемиологии данных заболеваний. Краткие сведения по эпидемиологии сахарного диабета приведены в таблице 1.3.

Таблица 1.3 – Эпидемиология сахарного диабета 1 и 2 типа

Группы населения	E10 (сахарный диабет 1 типа)	E11 (сахарный диабет 2 типа)
По нозологическим формам	На СД 1 приходится порядка 1,5 – 2% всех случаев диабета. По данным государственного регистра больных СД за 2015 год в РФ инсулинозависимым сахарным диабетом болеют 340 462 человека, что составляет 5,6 % от общей заболеваемости СД [12].	По данным государственного регистра больных СД за 2015 год в РФ сахарным диабетом 2 типа болеют 4 045 564 человека, что составляет 92,2 % от общей заболеваемости СД [12].
По возрастным группам	Возрастной пик манифестации СД1 составляет 10 – 13 лет. В подавляющем большинстве случаев манифестирует до 40 лет [12].	По данным Международной Диабетической Федерации, в мире подавляющее большинство пациентов с СД 2 находится в возрастной категории от 40 до 59 лет, в Российской Федерации – от 30 до 50 лет [13]. По данным третьего пересмотра национального регистра здоровья США (NHANES III), распространенность СД 2 составляет около 8% в возрасте 60 лет и достигает максимальных значений (22–24%) в возрасте старше 80 лет. В возрасте старше 60 распространённость составляет 16%.
По половым группам	Заболевание чаще регистрируется у лиц мужского пола.	Распространенность СД 2 среди женщин в 2,5 раза выше, чем среди мужчин. Как показали скрининговые исследования, эти различия существенно меньше и, возможно, обусловлены более частым определением сахара в крови, так как женщины более часто попадают в поле зрения врача [15].

Продолжение таблицы 1.3

Тип сахарного диабета	E10 (сахарный диабет 1 типа)	E11 (сахарный диабет 2 типа)
По этническим группам	<p>Наиболее высокая заболеваемость в Финляндии и Швеции (30 – 35 случаев на 100 000 населения в год), самая низкая в Японии, Китае, Корее (0,5 – 2 случая в год).</p> <p>Наследственная предрасположенность популяционный риск составляет 0,2%. Чем больше случаев СД1 в семье, тем риск заболеть выше для родственников. Риск развития диабета для членов семей, имеющих больных СД1, в среднем составляет 5%. Если в семье больны два ребенка, то риск для третьего составляет 9,5%. Если больны двое родителей, то риск развития СД1 для ребенка уже увеличивается до 34%. Кроме того, риск развития СД1 зависит от возраста манифестации заболевания у пробанда (заболевший в семье СД1). Чем раньше заболел ребенок в семье, тем риск для sibса выше. Так, если манифестация заболевания произошла у пробанда до 20 лет, то риск для sibса – 6,4%, а если манифестация заболевания – старше 20 лет, то риск для sibса – 1,2%. [16].</p>	<p>Сахарный диабет 2 типа крайне редко встречается среди некоторых популяций Меланезии, Восточной Африки и Южной Америки, а также среди коренных жителей Севера. В популяциях европейского происхождения показатели распространенности составляют 3 – 15%. В группах эмигрантов из Индии, Китая, а также американцев испанского происхождения они несколько выше (15 – 20%) [16].</p> <p>Наследственная предрасположенность Популяционный риск СД 2 очень зависит от изучаемой популяции; в большинстве популяций этот риск составляет от 1 до 5%, хотя в США он равен 6-7%. Если пациент имеет больного sibса, риск возрастает до 10%; наличие больного sibса и другого родственника первой степени родства повышает риск до 20%; если болен монозиготный близнец, риск поднимается до 50-100% [17].</p>

Для разработки информационной системы распознавания сахарного диабета необходимы сведения о вероятности СД 1 и 2 типа. Вероятность заболевания сахарным диабетом второго типа превышает вероятность заболевания сахарным диабетом первого типа в несколько раз. Частота заболеваний сахарным диабетом второго типа составляет от 80 до 90 %, а первого типа – от 10 до 20 %. Априорная вероятность СД 1 составляет 0,2%, СД 2 – от 1 до 5%. Вероятность развития СД 2 увеличивается с возрастом, так, для лиц в возрасте 60 лет априорная вероятность СД 2 составляет 8 %; старше 60 лет – 16%, для лиц старше 80 лет 22 – 24%.

Частота сахарного диабета 1 типа у родственников I степени родства составляет до 10%. Частота сахарного диабета 2 типа у родственников I степени родства составляет около 20%.

Для разработки ИС распознавания СД необходимо иметь сведения о наиболее информативных симптомах, для этого необходимо провести анализ этиопатогенеза рассматриваемой эндокринной патологии.

1.2 Этиология болезни

Сахарный диабет – это группа метаболических (обменных) заболеваний, характеризующихся хронической гипергликемией, которая является результатом нарушения секреции инсулина, действия инсулина или обоих этих факторов. Хроническая гипергликемия при СД сопровождается повреждением, дисфункцией и недостаточностью различных органов, особенно глаз, почек, нервов, сердца и кровеносных сосудов [17].

Термин «сахарный диабет 1 типа» применяется к обозначению группы заболеваний, которые развиваются вследствие прогрессирующего разрушения бета-клеток поджелудочной железы, что приводит к дефициту синтеза проинсулина и гипергликемии, требует заместительной гормональной терапии.

В настоящее время сахарный диабет 1 типа можно рассматривать как полигенное многофакторное заболевание [15]. СД1 типа развивается на фоне генетической предрасположенности (определяет развитие заболевания примерно на 1/3). При СД 1 наблюдается аутоиммунная деструкция бета-клеток поджелудочной железы. При этом гуморальные маркеры аутоиммунного поражения бета-клеток, к которым относятся антитела к островкам поджелудочной железы, антитела к глутамат-декарбоксилазе (GAD65) и антитела к тирозин-фосфатазе (IA-2 и IA-2бета), обнаруживаются у 85-90 % пациентов. Тем не менее, основное значение в деструкции бета-клеток придается факторам клеточного

иммунитета. Сахарный диабет 1 типа ассоциирован с такими гаплотипами HLA, как DQA и DQB. С повышенной частотой сахарный диабет 1 типа сочетается с другими аутоиммунными эндокринными (аутоиммунный тиреоидит, болезнь Аддисона) и неэндокринными заболеваниями, такими как алоpecia, витилиго, болезнь Крона, ревматические заболевания [12].

Для реализации генетической предрасположенности необходимы факторы внешней среды, выступающие в роли триггера аутоиммунного поражения бета-клеток поджелудочной железы. К таким факторам относят вирусы, ряд ингредиентов пищевых продуктов, химические вещества.

Ведущая роль среди всех средовых факторов отводится вирусной инфекции. Тропными к поджелудочной железе человека являются вирусы Коксаки В, эпидемического паротита, ветряной оспы, кори, мононуклеоза, цитомегаловируса и др. Ряд химических веществ являются токсичными для бета-клеток, например, противоопухолевые препараты стрептозацин, вакор. Среди пищевых ингредиентов такими же свойствами обладают некоторые компоненты коровьего молока – бычий альбумин; соя, глютен, нитраты/нитриты в составе пищевых красителей и консервантов [15].

Термин «сахарный диабет 2 типа» относится к заболеванию, развивающемуся у лиц с избыточным накоплением жировой ткани, имеющих инсулинорезистентность, вследствие чего наблюдается избыточный синтез проинсулина, инсулина и амилина бета-клетками поджелудочной железы, возникает так называемый «относительный дефицит» [18].

Сахарный диабет 2 типа является многофакторным заболеванием с наследственной предрасположенностью. Большинство пациентов с сахарным диабетом 2 типа указывают на наличие сахарного диабета 2 типа у ближайших родственников. Какого-то одного гена, полиморфизм которого определяет предрасположенность к сахарному диабету 2 типа, не обнаружено. Большое значение в реализации наследственной

предрасположенности к сахарному диабету 2 типа играют факторы окружающей среды, в первую очередь, особенности образа жизни.

Факторами риска развития сахарного диабета 2 типа являются:

- 1) ожирение, особенно висцеральное;
- 2) этническая принадлежность (особенно при смене традиционного образа жизни на западный);
- 3) сахарный диабет 2 типа у ближайших родственников;
- 4) малоподвижный образ жизни;
- 5) особенности диеты (высокое потребление рафинированных углеводов и низкое содержание клетчатки);
- 6) артериальная гипертензия.

Классификация сахарного диабета

По степени тяжести (таблица 1.4):

- 1) лёгкая форма;
- 2) средняя форма;
- 3) тяжёлая форма.

Таблица 1.4 – Классификация сахарного диабета по степени тяжести

Степень тяжести	Определение
СД легкого течения	СД 2-го типа, компенсированный на диетотерапии, без микро- и макрососудистых осложнений
СД среднетяжелого течения	СД 1-го и 2-го типа на сахароснижающей терапии, без осложнений или при наличии начальных стадий осложнений: диабетическая ретинопатия, (непролиферативная стадия), диабетическая нефропатия (стадия микроальбуминурии), диабетическая нейропатия
СД тяжелого течения	Лабильное течение СД (частые гипогликемии и /или кетоацидотические состояния) СД 1 -го и 2-го типа с тяжелыми сосудистыми осложнениями

По степени компенсации углеводного обмена (таблица 1.5):

- 1) фаза компенсации;
- 2) фаза субкомпенсации;
- 3) фаза декомпенсации.

Таблица 1.5 – Критерии компенсации сахарного диабета*

Показатели		Компенсация	Субкомпенсация	Декомпенсация
HbA1c, %		< 7.0	7.0 – 7.5	> 7.5
Глюкоза плазмы, ммоль/л (мг/дл)	Натощак / перед едой	6.5(<117)	6.5 – 7.5	> 7.5 (> 135)
	Через 2 часа после еды	<8.0(< 144)	8.0 – 10.0 (144 – 180)	> 10 (> 180)

*Данные целевые значения не относятся к детям и подросткам, беременным женщинам, пожилым людям с ожидаемой продолжительностью жизни менее 5 лет.

Таким образом, сахарный диабет – полигенное многофакторное заболевание. При сахарном диабете 1 типа к этиологическим факторам риска относятся: наследственность, воздействие агентов внешней среды (вирусы, ряд ингредиентов пищевых продуктов, химические вещества). При сахарном диабете 2 типа факторами риска являются: наследственность, этническая принадлежность, возраст старше 45 лет, наличие в анамнезе гестационного сахарного диабета, избыточная масса тела, гиподинамия, особенности питания, нарушения липидного обмена, артериальная гипертензия (таблица 1.6).

Таблица 1.6 – Этиологические факторы (признаки) риска сахарного диабета

Признаки	Форма сахарного диабета	
	СД 1	СД 2
Генетическая предрасположенность	+	+
Этническая принадлежность	+	+
Возраст	+(пик 10 – 13 лет)	+(от 45 лет)
Холодный климат	+	+/-
Вирусная инфекция (вирус Коксаки В, эпидемического паротита, ветряной оспы, кори, мононуклеоза, цитомегаловируса)	+	–
Особенности питания	+(искусственное вскармливание)	+(высокое потребление рафинированных углеводов и низкое употребление клетчатки)
Химические вещества	+	–
Гестационный сахарный диабет	–	+

Продолжение таблицы 1.6

Гиподинамия (низкая физическая активность)	–	+
Генетическая предрасположенность	+	+
Избыточная масса тела и ожирение	–	+
Нарушения липидного обмена (холестерин ЛПВП 0,9 ммоль/л и/или триглицериды $\geq 2,82$ ммоль/л)	–	+
Артериальная гипертензия	–	+

Однако, для постановки диагноза недостаточно сведений из анамнеза и физикального осмотра, необходимы диагностические признаки, полученные при других исследованиях. Для того чтобы понять, какие же признаки наиболее информативны необходимо знание патогенеза.

1.3 Патогенез болезни

При сахарном диабете 1 типа ведущая причина развития нарушений со стороны различных органов и систем – эффекты, обусловленные абсолютным дефицитом инсулина. Дефицит инсулина является следствием аутоиммунной деструкции бета-клеток, которая проходит несколько стадий до клинической манифестации СД 1 типа (рисунок 1.5, таблица 1.7).

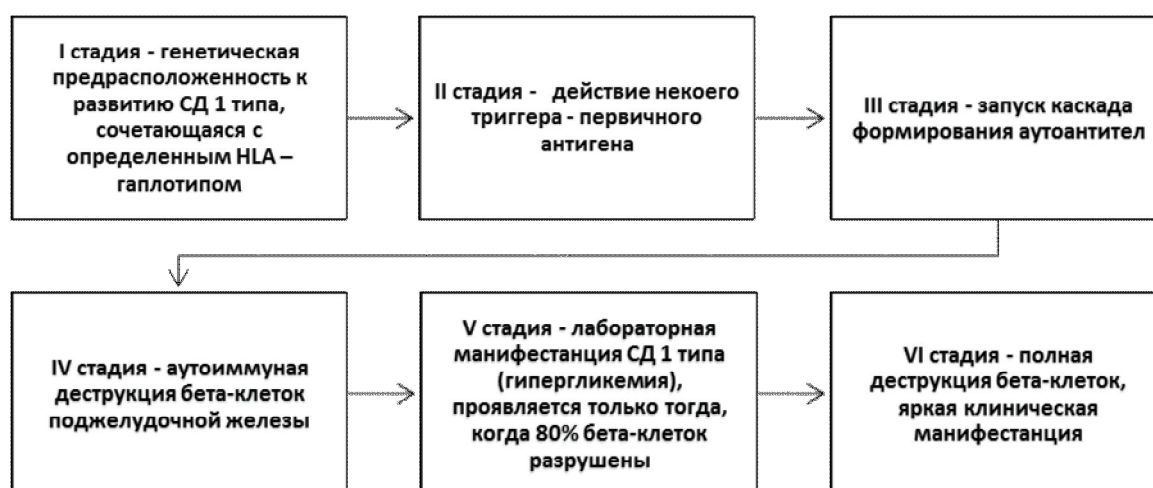


Рисунок 1.5 – Этапы развития сахарного диабета 1 типа

Таблица 1.7 – Характеристики стадий патогенеза сахарного диабета первого типа

Стадия заболевания	Характеристика стадии	Диагностические признаки
1. Генетическая предрасположенность	Наличие генетической предрасположенности	HLA – типирование: обнаружение специфических антигенов системы HLA (B8, B15, DR3, DR4 и др.).
2. Триггирование (запуск) аутоиммунных процессов	Гипотетический пусковой момент. Повреждение β – клеток различными диабетогенными факторами и триггирование иммунных процессов	Данные анамнеза: о перенесенной вирусной инфекции, о воздействии или приеме ряда химических веществ, об искусственном вскармливании, о перенесённом стрессе. Иммунологический анализ крови: определяются аутоантитела к инсулину (IAA), цитоплазме островковых клеток (ICA), глутаматдекарбоксилазе (GAD), тирозин-фосфатазе островковых клеток (IA2) в небольшом титре.
3. Стадия активных иммунных процессов	Отмечаются нарушения со стороны иммунной системы - появление аутоантител к антигенам β -клеток, инсулину, тирозинфосфатазе	Иммунологический анализ крови: определяются АТ к бета-клеткам поджелудочной железы, к глутаматкарбоксилазе, к инсулину. Внутривенный глюкозотолерантный тест: отсутствует первая фаза выработки инсулина
4. Прогрессирующее снижение секреции инсулина, стимулированного глюкозой	Характеризуется серьезными иммунными сбоями, а именно, стремительно снижается секреция инсулина из-за развития инсулита, нарушается резистентность к глюкозе, при этом уровень сахара в крови остается в пределах нормы	Иммунологический анализ крови: определяются АТ к бета-клеткам поджелудочной железы, к глутаматкарбоксилазе, к инсулину. Снижена секреция инсулина. Пероральный глюкозотолерантный тест: выявляется повышение уровня глюкозы натощак и/или через 2 часа после теста. Биохимический анализ крови: возможен повышенный уровень глюкозы натощак.

Продолжение таблицы 1.7

Стадия заболевания	Характеристика стадии	Диагностические признаки
5. Клинически явный или манифестный сахарный диабет	Характеризуется выраженными клиническими проявлениями, так как три четверти β -клеток к этому моменту разрушены. Сохранена лишь остаточная секреция С-пептида.	Клинические проявления (раздел 1.4). Иммунологический анализ крови: определяются АТ к бета-клеткам поджелудочной железы, к глутаматкарбоксилазе, к инсулину. Инсулин снижен или не определяется, сохранена остаточная секреция С-пептида. Биохимический анализ крови: повышен уровень глюкозы, гликированных белков (гемоглобина). Анализ мочи: обнаружение кетоновых тел.
6. Полная деструкция бета-клеток	Тотальная гибель β -клеток. С-пептид не выявляется, титры антител снижаются. Эта стадия иначе именуется тотальным диабетом	Клинические проявления (раздел 1.4). Иммунологический анализ крови: снижение титра АТ к бета-клеткам поджелудочной железы, к глутаматкарбоксилазе, к инсулину. Инсулин и С-пептид не определяется. Биохимический анализ крови: повышен уровень глюкозы, гликированных белков (гемоглобина), возможно повышение уровня лактата. Анализ мочи: обнаружение белков, кетоновых тел, глюкозы.

Патогенетически сахарный диабет 2 типа представляет собой гетерогенную группу нарушений обмена веществ, именно это и определяет его значительную клиническую неоднородность. В основе его патогенеза лежит инсулинорезистентность (снижение опосредованной инсулином утилизации глюкозы тканями), которая реализуется на фоне секреторной дисфункции β -клеток. Таким образом, происходит нарушение баланса чувствительности к инсулину и инсулиновой секреции. Секреторная дисфункция β -клеток заключается в замедлении «раннего» секреторного выброса инсулина в ответ на увеличение уровня глюкозы в крови.

При этом первая (быстрая) фаза секреции, которая заключается в опорожнении везикул с накопленным инсулином, фактически отсутствует; вторая (медленная) фаза секреции осуществляется в ответ на стабилизирующуюся гипергликемию постоянно, в тоническом режиме, и, несмотря на избыточную секрецию инсулина, уровень гликемии на фоне инсулинорезистентности не нормализуется.

Следствием гиперинсулинемии является снижение чувствительности и числа инсулиновых рецепторов, а также подавление пострецепторных механизмов, опосредующих эффекты инсулина (инсулинорезистентность). Содержание основного транспортера глюкозы в мышечных и жировых клетках (GLUT-4) снижено на 40 % у лиц с висцеральным ожирением и на 80 % – у лиц с СД 2. Вследствие инсулинорезистентности гепатоцитов и портальной гиперинсулинемией происходит гиперпродукция глюкозы печенью, и развивается гипергликемия натощак, которая выявляется у большинства пациентов с СД-2, в том числе и на ранних этапах заболевания. Сама по себе гипергликемия неблагоприятно влияет на характер и уровень секреторной активности β -клеток (глюкозотоксичность). Длительно, на протяжении многих лет и десятилетий существующая гипергликемия в конечном счете приводит к истощению продукции инсулина β -клетками и у пациента могут появиться некоторые симптомы дефицита инсулина – похудение, кетоз при сопутствующих инфекционных заболеваниях. Тем не менее, остаточная продукция инсулина, которой оказывается достаточно для предотвращения кетоацидоза, при СД 2 практически всегда сохраняется.

При СД 1 и СД 2 типа наблюдается абсолютный или относительный дефицит инсулина, который приводит к нарушению метаболизма жиров, белков и углеводов. Патологические изменения органов и тканей представлены в таблице 1.8.

Таблица 1.8 – Патофизиология сахарного диабета

Показатель	Изменения при сахарном диабете
Углеводный обмен	<p>Механизмами развития гипергликемии являются:</p> <ol style="list-style-type: none"> 1) Вследствие дефицита инсулина снижается поглощение глюкозы мышцами (при обоих типах СД) и жировой тканью (при I типе СД) после еды. 2) В гепатоцитах, миоцитах и липоцитах низкие уровни инсулина приводят к угнетению гексокиназной реакции, что ведет к понижению продукции глюкозо-6-фосфата, что обуславливает угнетение гликолиза и пентозного цикла. 3) Вследствие активации фосфоорилазы и ингибирования гликогенсинтетазы в гепатоцитах усиливается гликогенолиз и глюкоза поступает в кровь. 4) Глюкагон активирует ферменты глюконеогенеза и обеспечивает глюконеогенез субстратами (усиливает липолиз, тормозит синтез белка на рибосомах). <p>Глюкозурия является следствием гипергликемии происходит в результате нарушения реабсорбции глюкозы в почечных канальцах.</p> <p>Гиперлактатацидемия развивается вследствие торможения катаболизма лактата в цикле Кребса, нарушения ресинтеза гликогена из лактата.</p>
Жировой обмен	<p>В условиях недостатка глюкозы в клетках жиры становятся основным источником энергии. Нарушения жирового обмена при СД проявляются гиперлипидемией, кетонемией, кетонурией.</p> <p>Гиперлипидемия – увеличение содержания в крови липидов выше нормы (более 8 г/л). Это обусловлено активацией липолиза, торможением транспорта и утилизации липидов клетками, снижением активности ЛПЛазы.</p> <p>Кетонемия – повышение концентрации в крови кетоновых тел (КТ) выше нормы (более 2,5 мг%). К КТ относят ацетон, ацетоуксусную и β-оксимасляную кислоты. Кетонемия, как правило, развивается при дефиците инсулина и обусловлена активацией окисления ВЖК в клетках с образованием КТ.</p> <p>Кетонурия – выделение КТ из организма с мочой – симптом неблагоприятного течения СД. Причина: высокая концентрация в крови КТ, которые хорошо фильтруются в почках.</p>
Белковый обмен	<p>Нарушения белкового обмена при СД характеризуются гиперазотемией и азотурией.</p> <p>Гиперазотемия – увеличение содержания в крови небелковых азотистых соединений. Гиперазотемия обусловлена усилением катаболизма белка, который при дефиците глюкозы в клетках служит источником энергии. Небелковый азот представлен азотом мочевины, аминокислот, мочевой кислоты, креатинина, аммиака (остаточный азот).</p> <p>Азотурия – повышение экскреции азотистых соединений с мочой вследствие гиперазотемии.</p> <p>Белковый дефицит приводит к задержке физического и умственного развития детей, снижению регенерации тканей (замедление заживления ран, сращения костей), развитию вторичного иммунодефицита (увеличение чувствительности к инфекциям).</p>

Продолжение таблицы 1.8

Показатель	Изменения при сахарном диабете
Водно-электролитный обмен	<p>Гиперосмотическая гипогидратация– развивается в результате полиурии, которая обусловлена гипергликемией и кетонемией.</p> <p>Гипонатриемия – возникает на ранней стадии, из-за выхода воды из клеток по градиенту концентрации.</p> <p>Гиперкалиемия и гиперфосфатемия– возникают на ранней стадии, из-за выхода калия и фосфора из клеток в связи с распадом тканей, ацидозом и дефицитом инсулина.</p> <p>Гипернатриемия – возникает вследствие обезвоживания (потери воды превышают потери электролитов).</p> <p>Гипокалиемия, гипофосфатемия, гипомагниемия – происходят из-за потери соответствующих ионов при осмотическом диурезе, затем при леченииинсулином [19].</p>
Кислотно-щелочное равновесие	<p>Характерно развитие метаболического ацидоза:</p> <p>Кетонемический метаболический ацидоз (связан с накоплением кетоновых тел).</p> <p>Лактацидемический метаболический ацидоз (связан с накоплением молочной кислоты).</p>

Течение сахарного диабета условно можно разделить на следующие периоды: латентный, продромальный, период клинических проявлений, период осложнений. Для каждого периода характерны свои диагностические признаки (таблица 1.9).

Таблица 1.9 – Диагностические признаки по периодам течения сахарного диабета

Период СД	Диагностические признаки	
	СД 1	СД 2
Латентный	Наличие в анамнезе этиологических факторов риска (см. раздел 1.2). Исследование генетических маркеров СД 1: обнаружение специфических антигенов системы HLA (B8, B15, DR3, DR4 и др.).	Наличие в анамнезе этиологических факторов риска (см. раздел 1.2).

Продолжение таблицы 1.9

Период СД	Диагностические признаки	
	СД 1	СД 2
Продромальный	<p>Иммунологический анализ крови: наличие АТ к бета-клеткам поджелудочной железы, к глутаматкарбоксилазе, к инсулину.</p> <p>Внутривенный глюкозотолерантный тест: отсутствие первой фазы выработки инсулина.</p> <p>Пероральный глюкозотолерантный тест: выявляется повышение уровня глюкозы натощак и/или через 2 часа после теста.</p> <p>Биохимический анализ крови: возможен повышенный уровень глюкозы натощак.</p>	<p>Биохимический анализ крови: гипергликемия, повышен уровень гликированного гемоглобина. Уровень гликемии превышает норму в течение длительного периода. Заболевание остается нераспознанным из-за отсутствия видимых проявлений.</p>
Период клинических проявлений	<p>Жалобы на полидипсию, полифагию, полиурию, сухость во рту, астенический синдром, похудание. Симптомы нарастают быстро.</p> <p>Биохимический анализ крови: гипергликемия, повышен уровень гликированного гемоглобина. Повышенный уровень кетоновых тел в крови.</p> <p>Анализ мочи: глюкозурия, кетонурия, протеинурия.</p>	<p>Жалобы на полидипсию, полифагию, полиурию, сухость во рту, астенический синдром. Симптомы нарастают медленно.</p> <p>Биохимический анализ крови: гипергликемия, повышен уровень гликированного гемоглобина.</p> <p>Нарушение липидного обмена (гипертриглицеридемия, гиперхолестеринемия, повышение липопротеидов низкой плотности, снижение липопротеидов высокой плотности).</p> <p>Анализ мочи: глюкозурия, протеинурия.</p>
Период осложнений	<p>Те же, что и при стадии клинических проявлениях, с появлением признаков поздних и острых осложнений.</p>	<p>Те же, что и при стадии клинических проявлениях, с появлением признаков поздних и острых осложнений.</p>

Таким образом, латентный период обеих форм сахарного диабета характеризуется воздействием этиологических факторов риска, перечисленных в таблице 1.6 (раздел 1.2). К диагностическим признакам

продромального периода сахарного диабета 1 типа относятся наличие антител к инсулину, бета-клеткам поджелудочной железы, глутаматкарбоксилазе, сниженный уровень инсулина. При сахарном диабете 2 типа в продромальный период выявляется гипергликемия, при этом клинические проявления заболевания длительно отсутствуют. Период клинических проявлений обеих форм сахарного диабета характеризуется появлением следующих признаков: жалоб на полидипсию, полифагию, полиурию, сухость во рту, астенический синдром; наличием гипергликемии и глюкозурии. Однако, сахарный диабет 1 типа имеет бурное клиническое начало с потерей массы тела, склонностью к кетоацидозу (диагностические признаки – кетонурия, кетонемия). Сахарный диабет 2 типа протекает медленно, помимо нарушения углеводного обмена сопровождается нарушением липидного обмена (гипертриглицеридемия, гиперхолестеринемия, повышение липопротеидов низкой плотности, снижение липопротеидов высокой плотности).

Разрабатываемая информационная система распознавания сахарного диабета предназначена для выявления сахарного диабета на стадии клинических проявлений.

Знание патогенеза и патофизиологии сахарного диабета позволяет выявить диагностические признаки. В разделе 1.4 приведена классификация основных симптомов сахарного диабета, имеющих код по МКБ.

1.4 Диагностические признаки болезни

Симптомы сахарного диабета, имеющие код по МКБ, их характеристики и классификация рассмотрены в таблице 1.10. Классификация признаков осуществлена на основе следующих критериев:

- 1) способ выявления (субъективные, объективные);
- 2) сроки проявления и выявления (ранние, поздние);

3) диагностическое значение с позиций патогенетического истолкования (неспецифичные, специфичные, патогномоничные);

4) использование в дифференциальном диагностировании болезней с помощью компьютера (вероятностные, детерминирующие).

Таблица 1.10 – Классификация и характеристики признаков сахарного диабета, имеющих код по МКБ 10

Признаки	Код МКБ	Описание	Формы заболеваний	
			СД 1, E10.9	СД 2, E11.9
			Классификационные группировки признаков	
Сбор анамнеза (жалобы)				
Полиурия	R35	Суточное выделение мочи у взрослого человека более 1800–2000 мл	Субъективный Поздний Специфичный Вероятностный Наиболее выраженный	Субъективный Поздний Специфичный Вероятностный Менее выраженный
Полидипсия	R63.1	Суточное потребление воды взрослым человеком более 2 литров	Субъективный Поздний Специфичный Вероятностный Наиболее выраженный	Субъективный Поздний Специфичный Вероятностный Менее выраженный
Полифагия	R63.2	Чрезмерный аппетит. Переедание	Субъективный Поздний Специфичный Вероятностный Наиболее выраженный	Субъективный Поздний Специфичный Вероятностный Менее выраженный
Зуд кожи	L23.8		Субъективный Поздний Специфичный Вероятностный Более выраженный, чем при СД2	Субъективный Поздний Специфичный Вероятностный Менее выраженный, чем при СД1
Недомогание и утомляемость	R53	Постепенно нарастающая слабость	Субъективный Ранний Специфичный Вероятностный	
Физикальное обследование				
Аномальная потеря массы тела	R63.4	Интенсивное немотивированное похудание при сохранении аппетита	Объективный Поздний Специфичный Вероятностный Более выраженный, чем при СД 2 В сочетании с полифагией и полидипсией у молодого человека почти патогномоничный симптом начала СД1	Объективный Поздний Специфичный Вероятностный Менее выраженный, чем при СД 1
Аномальная прибавка массы тела	R63.5		Маловероятный	Субъективный Поздний Специфичный Вероятностный

Продолжение таблицы 1.10

Признаки	Код МКБ	Описание	Формы заболеваний	
			СД 1, E10.9	СД 2, E11.9
			Классификационные группировки признаков	
Сухой рот неуточнённый	R68.2		Субъективный Ранний Специфичный Вероятностный	
Сухость кожи	R23.8		Субъективный Ранний Специфичный Вероятностный	
Лабораторные исследования				
Гипергликемия неуточнённая	R73.9	Содержание глюкозы в цельной капиллярной крови натощак $\geq 6,1$ ммоль/л	Объективный Поздний Специфичный Вероятностный	Объективный Ранний Специфичный Вероятностный
		Содержание глюкозы в венозной плазме натощак ≥ 7 ммоль/л		
Отклонения результатов теста толерантности к глюкозе нормы на к	R73.0	Содержание глюкозы в цельной капиллярной крови через 2 часа после ПГТТ ≥ 7 ммоль/л	Объективный Поздний Специфичный Вероятностный	Объективный Ранний Специфичный Вероятностный
		Отклонение от нормы содержания глюкозы в венозной плазме через 2 часа после ПГТТ $\geq 11,1$ ммоль/л		
Другие уточненные отклонения от нормы белков плазмы	R77.8	Повышение в крови уровня гликолизированных белков (гемоглобина, альбумина, фруктозамина). Уровень гликозилированного гемоглобина более 6%.	Объективный Поздний Специфичный Вероятностный	Объективный Ранний Специфичный Вероятностный
Гликозурия	R81	Концентрация глюкозы в первой утренней порции мочи превышает 1,7 ммоль/л	Объективный Поздний Специфичный Вероятностный	
Смешанная гиперлипидемия	E78.2	Высокий уровень липопротеинов и липидов в крови	Объективный Поздний Специфичный Вероятностный	Объективный Поздний Специфичный Вероятностный
Ацетонурия	R82.4	Тест на ацетон в моче	Объективный Поздний Специфичный Вероятностный	Объективный Поздний Специфичный Вероятностный
Изолированная протеинурия	K82	Наличие белка в моче	Объективный Поздний Специфичный Вероятностный	
Высокий титр антител	R76.0	Наличие антител к островкам поджелудочной железы, к глутамат-декарбоксилазе (GAD65) и к тирозин-фосфатазе (IA-2 и IA-2бета)	Объективный Ранний Специфичный Вероятностный	Маловероятный

Продолжение таблицы 1.10

Признаки	Код МКБ	Описание	Формы заболеваний	
			СД 1, E10.9	СД 2, E11.9
			Классификационные группировки признаков	
Инструментальные исследования				
Отклонения от нормы, выявленные при получении диагностического изображения в ходе исследования других отделов пищеварительного тракта	R93.3	Ультразвуковое исследование поджелудочной железы на первых этапах развития заболевания акустические размеры железы остаются в пределах нормы, не отличаются от нормальных параметров эхогенность и зернистость ткани. С течением болезни (более 5-6 лет) размеры железы уменьшаются, железа – резко истончается, приобретая «лентовидное» отображение. На фоне уменьшения размеров железы растет эхогенность ткани в сочетании с уменьшением ее зернистости.	Объективный Поздний Специфичный Вероятностный	
Отклонения от нормы, выявленные при проведении функциональных исследований сердечно-сосудистой системы	R94.3	Электрокардиографические признаки ишемической болезни сердца (деформация конечной части желудочкового комплекса, отклонение электрической оси сердца влево, рубцовые изменения миокарда после перенесенного инфаркта, нарушения ритма и внутрижелудочковой проводимости).	Объективный Поздний Специфичный Вероятностный	

Диагностические признаки, полученные при лабораторном обследовании, обладают достаточной информативностью для постановки диагноза сахарного диабета, а также являются удобными для дальнейшей работы, поскольку принимают определенные количественные значения, т.е. могут подвергаться статистической обработке. К таким признакам относятся: отклонения результатов нормы теста на толерантность к глюкозе (R73.0), отклонения от нормы белков плазмы, а именно гликогемоглобина (R77.8), смешанная гиперлипидемия (E78.2). Для данных показателей известны также законы распределения и их характеристики. Законы распределения и их характеристики получены в результате анализа литературных источников и анализа историй болезней (таблица 1.11, приложение А). Признак ацетонурия (R82.4) оценивается по бинарной шкале (положительный и отрицательный тест)

и чаще встречается у пациентов с сахарным диабетом 1 типа, поэтому также будет использоваться в разрабатываемой информационной системе.

Таблица 1.11 – Параметры нормального распределения вероятностей значений признаков сахарного диабета (данные литературных источников)

Признак	Распределение	Норма		СД1		СД2	
		μ	σ	μ	σ	μ	σ
Постприандиальная гликемия, ммоль/л [20]	Нормальное [20]	4,74	0,13	11,1	0,34	11,6	1,1
Гликированный гемоглобин (HbA1C), % [21]	Нормальное [35,36]	4,58	0,44	7,44	0,91	7,85	1,55
Триглицериды, ммоль/л [22]	Нормальное	1,18	0,84	1,46	0,81	3,2	1,01
μ – математическое ожидание; σ – среднеквадратичное отклонение							

Таблица 1.12 – Параметры нормального распределения вероятностей значений признаков сахарного диабета (данные из истории болезней)

Признак	Распределение	СД1 (n=27)		СД2 (n=28)	
		μ	σ	μ	σ
Постприандиальная гликемия, ммоль/л	Нормальное	13,5	3,7	11,41	4,22
Гликированный гемоглобин (HbA1C), %	Нормальное	10,79	2,99	10,51	2,71
Триглицериды, ммоль/л	Нормальное	1,32	0,73	3,65	1,86
μ – математическое ожидание; σ – среднеквадратичное отклонение					

В результате анализа литературы были выявлены диагностические признаки сахарного диабета, позволяющие установить диагноз данного заболевания и провести дифференциальную диагностику двух форм. К таким признакам относятся: уровень гликированного гемоглобина, уровень постприандиальной гликемии, уровень триглицеридов и наличие ацетонурии. Данные признаки подчиняются нормальному распределению. Для них также известны характеристики распределения: математическое ожидание и стандартное отклонение (таблица 1.11– 1.12). Однако, сведений, полученных из литературы о распределении и его параметрах для рассматриваемых диагностических признаков недостаточно, поскольку отсутствуют работы, в которых данные признаки рассматривались вместе в группе пациентов с сахарным диабетом. К тому же, анализ

литературы не позволил выявить зависимость между признаками. Следовательно, требуется провести анализ историй болезней пациентов с сахарным диабетом 1 и 2 типа с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками.

1.5 Методика распознавания болезни (обоснование потребности в информационной системе)

Распознавание сахарного диабета осуществляется врачом-терапевтом и представляет собой процесс принятия клинических решений. Процесс распознавания состоит из следующих этапов:

Этапы принятия клинического решения (распознавания болезни):

- 1) изучение жалоб путем клинического исследования (сбор анамнеза и физикальное обследование);
- 2) назначение диагностических обследований, каждый из которых обладает определенной достоверностью и информативностью;
- 3) обобщение клинических находок и результатов лабораторных исследований для разработки возможных вариантов диагноза;
- 4) сравнительная оценка риска и преимущества альтернативных вариантов дальнейших исследований;
- 5) на основании собранных данных формируется заключительный диагноз.

Одна из методик принятия решений, и именно распознавания сахарного диабета приведена на рисунке 1.5.

Сахарный
диабет 1
типа

Сахарный
диабет 2
типа

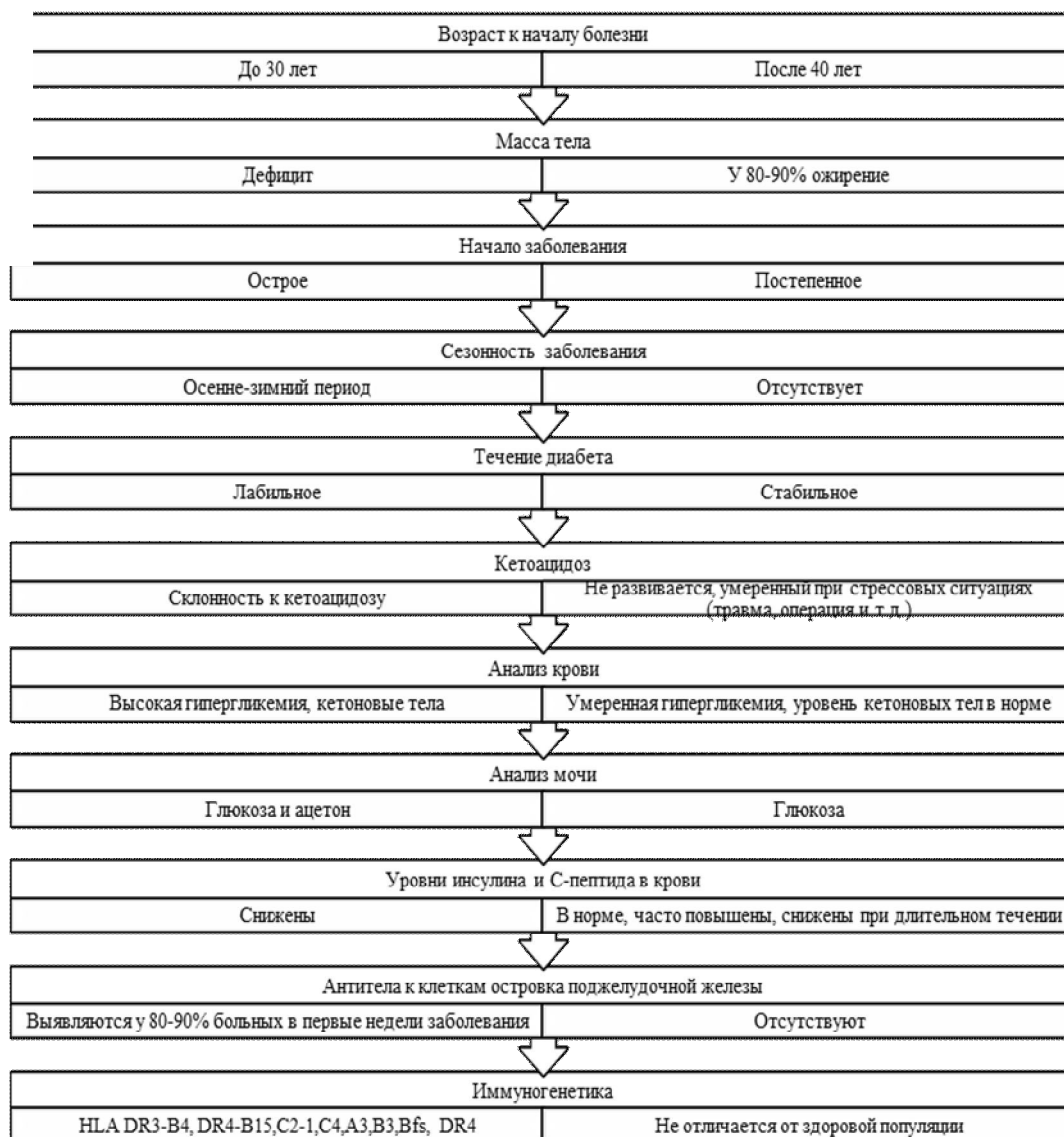


Рисунок 1.5 – Методика распознавания сахарного диабета

Принятие клинических решений, основывается на клиническом мышлении врача-эксперта. Клиническое мышление основывается на таких факторах, как опыт и обучение, индуктивное и дедуктивное мышление, интерпретация фактов, воспроизводимость и ценность которых непостоянны, и интуиция, которую бывает трудно определить.

Свою деятельность врач осуществляет по установленным нормам нагрузки и времени. 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-эндокринолога составляет 19 минут, для врача-терапевта 15 минут [23]. Норматив посещений (пациентов в час) для врача-терапевта составляет 4 условных единиц трудоемкости.

Таким образом, распознавание сахарного диабета – трудоемкий процесс, поскольку является коллективным интеллектуально сложным процессом принятия решений в условиях неопределённости и ограничений по времени, достоверность которого зависит от квалификации и опыта специалистов. Принятие решения о диагнозе социально-значимого заболевания в условиях высокой заболеваемости населения требует большого числа специалистов, осуществляется специалистом за ограниченное время при дефиците информации о состоянии пациента.

1.6 Основные результаты и выводы

1 Сахарный диабет относится к социально-значимым заболеваниям и занимает первое место среди болезней эндокринных органов, среди причин смерти – третье место после сердечно-сосудистых и онкологических заболеваний. Каждые 5 секунд в мире один человек заболевает сахарным диабетом, а каждые 7 секунд один человек умирает от этой болезни, которая получила статус неинфекционной эпидемии XXI века. Последние данные указывают на то, что основную тяжесть эпидемии несут на себе люди в странах с низким и средним уровнем дохода, и что эта болезнь затрагивает гораздо больше людей трудоспособного возраста, чем считалось ранее. В настоящее время число заболевших диабетом приближается к 400 миллионам, половина из которых в возрасте между 20 и 60 годами. Согласно опубликованным данным IDF и ВОЗ, Россия входит в пятерку стран-лидеров по темпам роста и распространённости диабета. Заболеваемость сахарным диабетом в России продолжает увеличиваться.

Так рост численности больных в период с 2000 по 2011 гг. составил 1,525 млн. человек. По прогнозам экспертов ЭНЦ МЗ РФ к 2030 году ожидается рост численности больных до 5,81 млн. человек. Прирост по сравнению с 2000 г. составит 3,743 млн. человек, т.е. увеличиться в 2,81 раза. Повышение эффективности диагностики сахарного диабета, в условиях широкого распространения данного заболевания и продолжающегося роста числа больных СД, ранней инвалидизации, высокой смертности от осложнений является актуальной задачей, стоящей перед практическим здравоохранением.

2 Вероятность заболевания сахарным диабетом второго типа превышает вероятность заболевания сахарным диабетом первого типа в несколько раз. Частота заболеваний сахарным диабетом второго типа составляет от 80 до 90 %, а первого типа – от 10 до 20 %. Априорная вероятность СД 1 составляет 0,2%, СД 2 – от 1 до 5%. Вероятность развития СД 2 увеличивается с возрастом, так, для лиц в возрасте 60 лет априорная вероятность СД 2 составляет 8 %; старше 60 лет – 16%, для лиц старше 80 лет 22–24%. Частота сахарного диабета 1 типа у родственников I степени родства составляет до 10%. Частота сахарного диабета 2 типа у родственников I степени родства составляет около 20%.

3 Сахарный диабет – полигенное многофакторное заболевание. При сахарном диабете 1 типа к этиологическим факторам риска относятся: наследственность, воздействие агентов внешней среды (вирусы, ряд ингредиентов пищевых продуктов, химические вещества). При сахарном диабете 2 типа факторами риска являются: наследственность, этническая принадлежность, возраст старше 45 лет, наличие в анамнезе гестационного сахарного диабета, избыточная масса тела, гиподинамия, особенности питания, нарушения липидного обмена, артериальная гипертензия (таблица 1.6). Однако, для постановки диагноза недостаточно признаков, полученных из анамнеза и физикального обследования, необходимы данные,

полученные при других исследованиях. Для того чтобы понять, какие же признаки наиболее информативны необходимо знание патогенеза.

3 Латентный период обеих форм сахарного диабета характеризуется воздействием этиологических факторов риска, перечисленных в таблице 1.6 (раздел 1.2). К диагностическим признакам продромального периода сахарного диабета 1 типа относятся наличие антител к инсулину, бета-клеткам поджелудочной железы, глутаматкарбоксилазе, сниженный уровень инсулина. При сахарном диабете 2 типа в продромальный период выявляется гипергликемия, при этом клинические проявления заболевания длительно отсутствуют. Период клинических проявлений обеих форм сахарного диабета характеризуется появлением следующих признаков: жалоб на полидипсию, полифагию, полиурию, сухость во рту, астенический синдром; наличием гипергликемии и глюкозурии. Однако сахарный диабет 1 типа имеет бурное клиническое начало с потерей массы тела, склонностью к кетоацидозу (диагностические признаки – кетонурия, кетонемия). Сахарный диабет 2 типа протекает медленно, помимо нарушения углеводного обмена сопровождается нарушением липидного обмена (гипертриглицеридемия, гиперхолестеринемия, повышение липопротеидов низкой плотности, снижение липопротеидов высокой плотности).

4 В результате анализа литературы были выявлены диагностические признаки сахарного диабета, позволяющие установить диагноз данного заболевания и провести дифференциальную диагностику двух форм. К таким признакам относятся: уровень гликированного гемоглобина, уровень постпрандиальной гликемии, уровень триглицеридов и наличие ацетонурии. Данные признаки подчиняются нормальному распределению. Для них также известны характеристики распределения: математическое ожидание и стандартное отклонение (таблица 1.11 – 1.12). Однако, сведений, полученных из литературы о распределении и его параметрах для рассматриваемых диагностических признаков недостаточно,

поскольку отсутствуют работы, в которых данные признаки рассматривались вместе в группе пациентов с сахарным диабетом. К тому же, анализ литературы не позволил выявить зависимость между признаками. Следовательно, требуется провести анализ историй болезней пациентов с сахарным диабетом 1 и 2 типа с целью выявления распределения и его параметров, а также зависимости между диагностическими признаками.

6 Распознавание сахарного диабета – трудоемкий процесс, поскольку является коллективным интеллектуально сложным процессом принятия решений в условиях неопределённости и ограничений по времени, достоверность которого зависит от квалификации и опыта специалистов. Принятие решения о диагнозе социально-значимого заболевания в условиях высокой заболеваемости населения требует большого числа специалистов, осуществляется специалистом за ограниченное время при дефиците информации о состоянии пациента.

7 Повышение достоверности диагноза достигается, прежде всего, за счёт принятия решений на основе опыта постановки диагноза высококвалифицированных специалистов, накопленного в базе данных, содержащей результаты обследований и достоверные диагнозы пациентов, и обучения информационной системы, применения методов распознавания образов. Снижение трудоёмкости постановки диагноза достигается оптимизацией обследования (выбора состава диагностических признаков и очередности обследования) пациентов, автоматизацией принятия решений. Ограничение требований к опыту специалистов достигается за счёт принятия решений, основанных на опыте высококвалифицированных специалистов. Ограничение требований к квалификации специалистов в области компьютерных технологий достигается за счёт применения «дружественного» к специалисту графического интерфейса пользователя.

2 ИНФОРМАЦИОННАЯ СИСТЕМА РАСПОЗНАВАНИЯ САХАРНОГО ДИАБЕТА

2.1 Назначение, классификация и структура информационных систем распознавания болезней

Информационная система – это организационно упорядоченная совокупность документов (массивов документов) и информационных технологий, в том числе с использованием средств вычислительной техники и связи, реализующих информационные процессы. Информационные системы предназначены для сбора, хранения, обработки, поиска, распространения, передачи и предоставления информации.

Во второй половине XX в., в период начала бурного развития компьютерных технологий, всеобщая информатизация не могла не захватить такую жизненно важную область, как медицина. Как в России, так и за рубежом стали разрабатываться информационные медицинские системы (ИМС).

Основными задачами, решаемыми с помощью ИМС, являются:

- 1) информационная поддержка оказания медицинской помощи населению;
- 2) информационная поддержка управления отраслью здравоохранения.

Определений ИМС в современной специальной литературе много. Рассмотрим одно из них. Информационная медицинская система – это совокупность информационных, организационных, программных и технических средств, предназначенных для автоматизации медицинских процессов и (или) организаций [24].

Медицинские информационные системы используются на различных уровнях в структуре здравоохранения:

- 1) базовом (клиническом) уровне (врачами разных профилей);
- 2) уровне учреждений (поликлиники, стационары, диспансеры, скорая

помощь);

3) территориальном уровне (специализированные службы и региональные органы управления);

4) федеральном уровне (федеральные учреждения и органы управления).

Разрабатываемая информационная система диагностики сахарного диабета относится к системам базового уровня, поскольку предназначена для информационного обеспечения принятия решений в профессиональной деятельности врачей.

Вопросы классификации медицинских информационных систем рассматриваются в отечественной и зарубежной литературе с конца семидесятых годов двадцатого века. Наиболее значимой работой, выполненной в нашей стране по рассматриваемой тематике, является разработка стандарта организации СТО МОСЗ 91500.16.0002–2004 «Информационные системы в здравоохранении. Общие требования» [25] в разработке которого принимали участие Лебедев Г.С., Емелин И.В., Столбов А.П. и другие ведущие специалисты в области медицинской информатики. В данном стандарте вводится следующая функциональная классификация информационных систем в здравоохранении (рисунок 2.1).

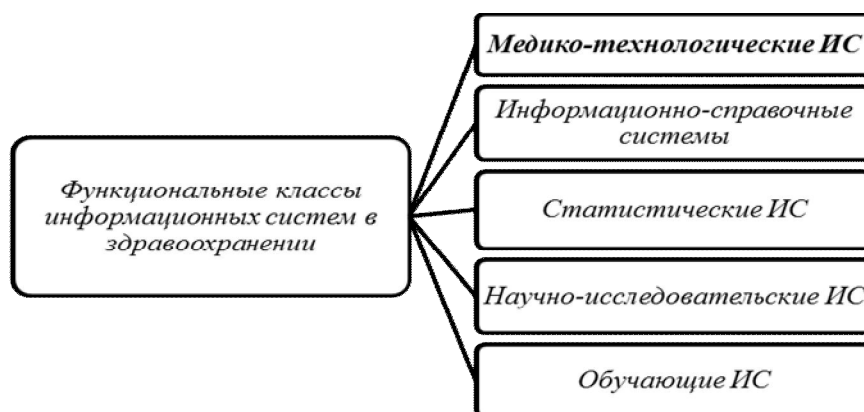


Рисунок 2.1 – Функциональные классы информационных систем в здравоохранении

Медико-технологические информационные системы (ИС), предназначены для информационного обеспечения процессов диагностики, лечения, реабилитации и профилактики пациентов в лечебно-профилактических учреждениях.

Информационно-справочные системы, содержат банки медицинской информации для информационного обслуживания медицинских учреждений и служб управления здравоохранением.

Статистические ИС органов управления здравоохранением.

Научно-исследовательские ИС, предназначены для информационного обеспечения медицинских исследований в клинических научно-исследовательских институтах (НИИ).

Обучающие ИС, предназначены для информационного обеспечения процессов обучения в медицинских учебных заведениях.

Поскольку для диагностики болезней используются медико-технологические ИС, рассмотрим их более подробно.

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

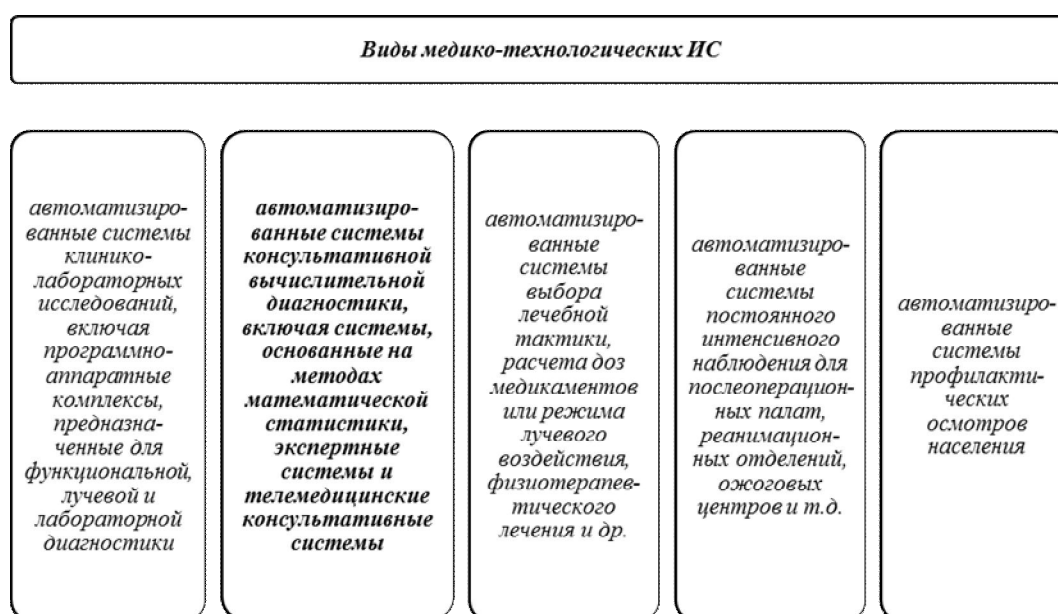


Рисунок 2.2 – Виды медико-технологических информационных систем

Разрабатываемая информационная система диагностики сахарного диабета относится к автоматизированным системам консультативной вычислительной диагностики, а именно к экспертным системам.

Экспертная система (ЭС) – это программное средство, использующее экспертные знания для обеспечения высокоэффективного решения неформализованных задач в узкой предметной области. Основу ЭС составляет база знаний о предметной области, которая накапливается в процессе построения и эксплуатации ЭС.

Общепринятая классификация экспертных систем отсутствует, однако наиболее часто экспертные системы различают по назначению, предметной области, динамичности, типу ЭВМ и по степени интеграции с другими программами. Классификационные признаки разрабатываемой ЭС отражены на рисунке 2.3.



Рисунок 2.3 –Классификация экспертных систем

В структуру экспертных систем входят:

- 1) база знаний;
- 2) решатель;
- 3) интеллектуальный редактор баз знаний;
- 4) интерфейс пользователя;
- 5) подсистема объяснений;
- 6) рабочая память.

Схема основных блоков статической экспертной системы приведена на рисунке 2.4.

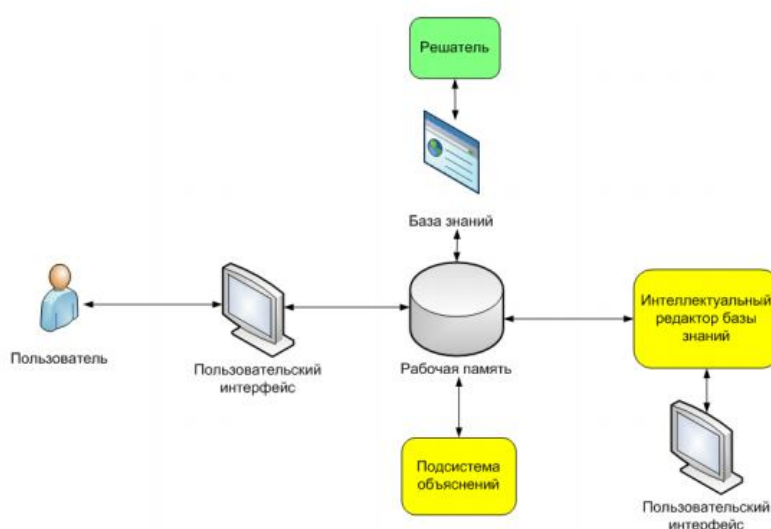


Рисунок 2.4 – Основные блоки экспертной системы [26]

Пользовательский интерфейс – комплекс программ, реализующих диалог пользователя с ЭС как на стадии ввода информации, получения результатов и «объяснения» решения.

База знаний (БЗ) – ядро ЭС, совокупность формализованных знаний предметной области, записанная на машинный носитель в форме, понятной эксперту и инженеру по знаниям. Параллельно такому представлению существует БЗ во внутреннем "машинном" представлении.

Решатель (машина вывода, интерпретатор) – программа, моделирующая ход рассуждений эксперта на основании формализованных

знаний, имеющихся в БЗ и исходных данных (фактах), получаемых от пользователя.

Подсистема объяснений – программа, протоколирующая работу решателя в виде «цепочки логических выводов». Она позволяет пользователю получить ответы на вопросы; "Как была получена та или иная рекомендация?" и "Почему система приняла такое решение?".

Редактор БЗ – программа, представляющая инженеру по знаниям возможность создавать и пополнять БЗ в диалоговом режиме. Осуществляет ввод формализованных знаний [26].

Таким образом, разрабатываемая информационная система распознавания сахарного диабета относится к системам базового уровня, поскольку предназначена для информационного обеспечения принятия решений в профессиональной деятельности врачей. Согласно функциональной классификации ИС разрабатываемая информационная система относится к классу медико–технологических, а именно к классу информационных экспертных систем. Основными составными частями экспертной системы являются: база знаний, решатель, пользовательский интерфейс. Разрабатываемая информационная экспертная система предназначена для поддержки принятия решений специалистам в области диагностики эндокринной патологии – сахарного диабета, а также для обучения студентов медицинских вузов. Данная экспертная система предназначена для работы на персональном компьютере, по степени интеграции относится к гибридным, поскольку представляет программный комплекс из следующих прикладных программ: Microsoft Access, Microsoft Excel, MATLAB.

2.2 Информационные системы распознавания болезней (примеры)

DXplain

DXplain – система поддержки принятия решений, разработанная в Бостоне в 1986 и используемая до сих пор.

В настоящее время база знаний DXplain 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 [27].

ILIAD – экспертная система, использующая Байесовские методы для анализа возможного диагноза на основе текущих симптомов. Система разрабатывалась для постановки диагноза в области общей медицины. В данный момент база знаний содержит описание свыше 1500 диагнозов. Оболочка Iliad также может быть использована для создания систем поддержки медицинской диагностики в других областях. Активно используется в качестве системы обучения в медицинских колледжах в США [28].

Онлайн-версии медицинских экспертных систем

В настоящее время существуют онлайн-версии медицинских экспертных систем, где любой пользователь может установить диагноз с той или иной долей вероятности. Примером таких систем являются: EasyDiagnosis, AIMedica, Домашний доктор.

EasyDiagnosis

Данная система является бесплатным англоязычным сервисом, где предоставляется возможность установить диагноз болезни или причины недомогания при помощи теста. Выбрав одну из 19 категорий, например, "Cough" (кашель), пользователю для продолжения необходимо согласиться с тем, что данный сервис не обеспечивает профессионального диагностирования, и предпочтительно обращение к врачу; кроме того, сервис предупреждает пользователя что при вводе неверных данных, диагноз не может быть определен верно. После принятия условий сервиса пользователем, открывается окно с тестом, где необходимо указать свой пол и возраст и затем приступить к ответам на вопросы. После того как все вопросы пройдены (их количество варьируется в зависимости от болезни), появляется диаграмма, где в процентных соотношениях предоставляется вероятность наличия той или иной болезни [29].

AIMedica

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 автоматически генерируемые 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.

AIMedica позволяет также проверить возможный диагноз путем проверки наличия характерных для этого заболевания симптомов и получить альтернативный список вероятных диагнозов.

Основу сервиса составляет база знаний, разработанная путем анализа (data-mining) больших массивов слабоструктурированных данных (bigdata), содержащихся в более чем 22 млн. медицинских источниках [30].

Домашний доктор

Домашний доктор –пример простой медицинской экспертной системы, разработанной в России. Она определяет характер заболевания, основываясь на ответах пользователя, полученных в результате диалога. База знаний включает порядка 100 распространенных заболеваний.

Таблица 2.1 – Сравнение основных характеристик информационных систем распознавания болезней

Информационная система	Производитель (страна)	Метод логического вывода	Целевая аудитория	Входные данные	Количество нозологий
DXplain	США	Bayesian Network	Врачи, студенты медицинских вузов	Симптомы, лабораторные показатели	2400
ILIAD	США	Bayesian Network	Врачи, студенты медицинских вузов	Симптомы, лабораторные показатели	1500
EasyDiagnosis	США	Bayesian Network	пациенты	Субъективные жалобы	Основные нозологии
AIMedica	Россия	Fuzzy logic	пациенты	Субъективные жалобы	225 (тестовая версия)
Домашний доктор	Россия		пациенты	Субъективные жалобы	100

Для задач диагностирования используются методы логического вывода: байесовские сети (Bayesian Network), нечеткая логика (Fuzzy logic), метод опорных векторов (Support Vector Machine) [31]. Среди анализируемых систем наиболее часто применяемым методом является байесовские сети (Bayesian Network). В рассмотренных информационных системах входными данными являются субъективные жалобы пациентов (EasyDiagnosis, AIMedica, Домашний доктор), что является существенным недостатком

данных систем и ограничивает их использование в лечебных учреждениях, в связи с низкой достоверностью. DXplain и ILIAD относятся к зарубежным продуктам, что является своего рода их недостатком. К сожалению, российский рынок ИМС находится на стадии становления, ИС для диагностики СД представлены единичными разработками, к тому же они не оперируют лабораторными показателями, несмотря на то, что эти данные играют весомую роль при диагностике СД.

Таким образом, в анализируемых экспертных системах базы знаний не содержат в своей структуре данных результатов обследования и диагнозов пациентов. В экспертных системах для задач распознавания используются субъективные жалобы, которые имеют более низкую информативность по сравнению с результатами лабораторного обследования. 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, а также решатель для формирования вероятностной диагностической модели является актуальной задачей.

2.3 Структура информационной системы распознавания сахарного диабета

Информационная система распознавания сахарного диабета образуется взаимодействующими между собой и с пользователем базой данных форм и моделей сахарного диабета, решателем MS Excel и решателем MatLab (рисунок 2.5).

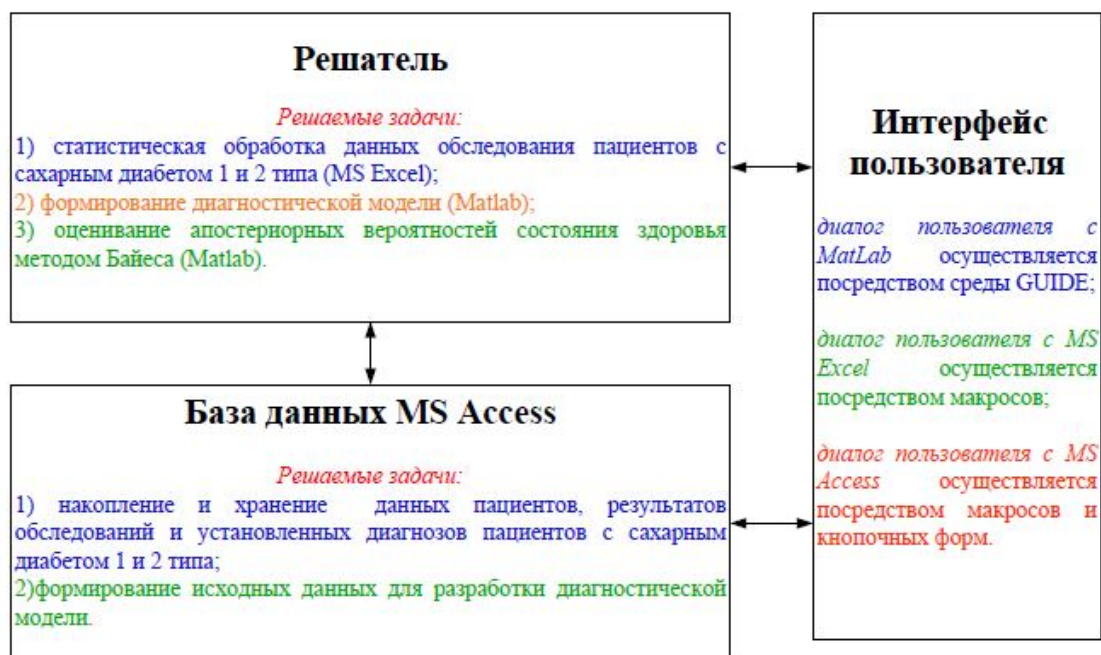


Рисунок 2.5 – Структура информационной системы распознавания сахарного диабета

2.3.1 База данных разрабатываемой системы

База данных – ядро информационной системы, совокупность знаний предметной области. Для разрабатываемой информационной системы она должна содержать данные для формирования диагностической модели – лабораторные показатели сахарного диабета, а также, сведения о пациентах (ФИО, пол, дата рождения, дата установления диагноза, дата обследования) и врачах (ФИО, специальность, категория).

Для разработки БД используется реляционная система управления базами данных (СУБД) – Microsoft Access. Данная СУБД выбрана, поскольку имеет ряд преимуществ перед другими:

- 1) распространенность, которая обусловлена тем, что Access является продуктом компании Microsoft;
- 2) полная совместимость с операционной системой Windows;

3) широкие возможности по импорту/экспорту данных в различные форматы, от таблиц Excel и текстовых файлов, до практически любой серверной СУБД через механизм ODBC;

4) наличие большого количества Мастеров, которые выполняют основную работу за пользователя при работе с данными и разработке приложений, помогают избежать рутинных действий и облегчают работу;

5) наличие встроенного языка макрокоманд;

6) ориентированность на пользователя с разной профессиональной подготовкой, что выражается в наличии большого количества Мастеров, развитую систему справки и понятный интерфейс.

2.3.2 Решатель MS Excel

Назначение решателя:

Статистическая обработка данных лабораторного обследования здоровых пациентов и пациентов с сахарным диабетом 1 и 2 типа: экспорт данных из MS Access в MS Excel, группировка, построение гистограммы распределения частот, расчет общих статистических характеристик, определение законов распределения, оценка корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков.

2.3.3 Решатель Matlab

Назначение решателя:

1) формирование диагностической модели на основе сведений, содержащихся в базе данных;

2) оценивание апостериорных вероятностей состояния здоровья методом Байеса.

Решатель для разрабатываемой информационной системы реализуется в среде Matlab. Выбор данной среды обусловлен тем, что по сравнению с традиционными языками программирования (C/C++, Java, Pascal, FORTRAN) Matlab позволяет на порядок сократить время решения типовых

задач и значительно упрощает разработку новых алгоритмов. Перечислим лишь некоторые преимущества Matlab:

1) открытость кода, что дает возможность пользователям разбираться в запрограммированных алгоритмах и при необходимости изменять их;

2) простой встроенный язык программирования системы позволяет легко создавать собственные алгоритмы;

3) визуальная среда GUIDE предназначена для написания приложений с графическим интерфейсом пользователя;

4) символьные вычисления в MATLAB основаны на библиотеке, являющейся ядром пакета Maple. Решение уравнений и систем, интегрирование и дифференцирование, вычисление пределов, разложение в ряд и суммирование рядов, поиск решения дифференциальных уравнений и систем, упрощение выражений – вот далеко не полный перечень возможностей MATLAB для проведения аналитических выкладок и расчетов;

5) обширная и удобная справочная система MATLAB.

Алгоритм постановки диагноза в информационной системе распознавания сахарного диабета построен на основе байесовского вероятностного подхода. Для его реализации необходима вероятностная диагностическая модель, содержащая данные об априорных вероятностях заболеваний и условных вероятностях симптомов. Данная модель формируется путем экспорта диагностической модели из базы данных MS Access.

2.3.4 Пользовательский интерфейс

Для работы с информационной системой, используется удобный пользовательский интерфейс: в среде MS Access с помощью макросов применяется главная кнопочная форма, в среде MS Excel с помощью макросов производятся расчеты, в Matlab для создания графического интерфейса – пакет GUIDE.

2.3.5 Взаимодействие между элементами информационной системы

Взаимодействие между блоками разрабатываемой информационной системы происходит следующим образом:

1) из блока базы данных результаты обследования пациентов экспортируются в решатель MS Excel;

2) в решателе MS Excel значения количественных признаков подвергаются статистической обработке, целью которой является группировка данных, определение закона распределения и его характеристик, оценка корреляции между признаками;

3) из решателя MS Excel, таблица, содержащая значения количественных признаков и рассчитанные интервалы, экспортируются в базу данных MS Access;

4) на основе загруженных данных в базе данных формируется запрос, в строках которого отражаются формы сахарного диабета, в столбцах – интервалы значений признаков, в ячейках – количество пациентов, у которых класс состояния здоровья проявляется значениями признаков из соответствующих интервалов;

5) полученный запрос экспортируется в решатель Matlab. В решателе осуществляется формирование диагностической модели и расчет апостериорных вероятностей с помощью метода Байеса.

Таким образом, разрабатываемая информационная система состоит из следующих блоков: базы данных MS Access, решателя MS Excel, решателя MatLab, пользовательского интерфейса. База данных предназначена для хранения, накопления и обработки результатов обследования и диагнозов пациентов с сахарным диабетом 1 и 2 типа, сведений о пациентах и медперсонале. Результаты обследования пациентов экспортируются в решатель MS Excel, в котором производится статистическая обработка данных, целью которой является группировка, определение законов распределения и оценка корреляции признаков. Из решателя MS Excel

данные загружаются в базу данных MS Access для формирования запроса, в строках которого отражаются формы сахарного диабета, в столбцах – интервалы значений признаков, в ячейках – количество пациентов, у которых класс состояния здоровья проявляется значениями признаков из соответствующих интервалов. Полученный запрос экспортируются в решатель Matlab. Решатель предназначен для формирования вероятностной диагностической модели и расчета апостериорных вероятностей с помощью метода Байеса. Работа со всеми элементами системы осуществляется через пользовательский интерфейс.

2.4 База данных пациентов с заболеванием сахарным диабетом

База данных служит для накопления и обработки сведений о результатах обследований и диагнозов из историй болезней пациентов с сахарным диабетом 1 и 2 типа (рисунок 2.6).

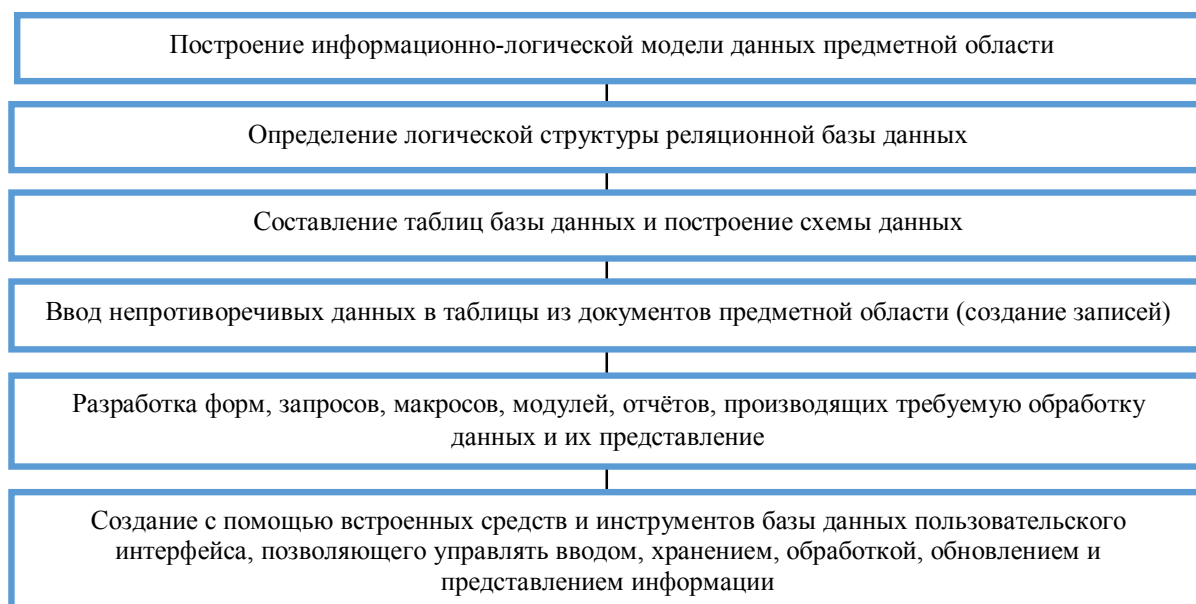


Рисунок 2.6 — Этапы проектирования и создания базы данных

На первом этапе осуществляется выявление атрибутов и подбор данных, которые будут храниться в базе данных.

На втором этапе определяются сущности базы данных, и представление каждой из них таблицей базы данных. Также устанавливается первичный ключ каждой таблицы.

В разрабатываемой базе данных пять стержневых сущностей:

- 1) «Пациенты» (код пациента, фамилия, имя, отчество, пол, дата рождения). Первичный ключ – код пациента (рисунок 2.7);
- 2) «Обследования» (код обследования, HbA1C %, ППГ ммоль/л, ТГ ммоль/л, Ац, дата обследования, код специалиста). Первичный ключ – код обследования (рисунок 2.8);
- 3) «Диагнозы» (код диагноза, код пациента, код болезни, дата установления диагноза, код специалиста, класс состояния здоровья). Первичный ключ – код пациента (рисунок 2.9);
- 4) «Специалисты» (код специалиста, фамилия, имя, отчество, код специальности, код категории). Первичный ключ – код специалиста (рисунок 2.10).
- 5) «Признаки» (код признака, код признака по МКБ, название признака по МКБ, обозначение признака). Первичный ключ – код признака (рисунок 2.11).

В разрабатываемой БД четыре характеристические сущности:

- 1) «Болезни» (код болезни, код болезни по МКБ, название болезни). Первичный ключ – код болезни (рисунок 2.12);
- 2) «Специальности» (код специальности, специальность). Первичный ключ – код специальности (рисунок 2.13);
- 3) «Категории» (код категории, категория). Первичный ключ – код категории (рисунок 2.14);
- 4) «Услуги» (код услуги, шифр услуги, название услуг). Первичный ключ – код услуги (рисунок 2.15).

Код пациента	Фамилия	Имя	Отчество	Пол	Дата рождения
Имя поля	Тип данных	Описание			
Код пациента	Числовой	Код пациента — уникальный код пациента [Числовое поле]			
Фамилия	Текстовый	Фамилия — фамилия пациента [Текстовое поле] (15 знаков);			
Имя	Текстовый	Имя — имя пациента [Текстовое поле] (15 знаков);			
Отчество	Текстовый	Отчество — отчество пациента [Текстовое поле] (15 знаков);			
Пол	Текстовый	Пол — пол пациента [Текстовое поле] (1 знак);			
Дата рождения	Дата/время	Дата рождения — дата рождения пациента [Поле типа дата] (10 знаков).			

Рисунок 2.7– Таблица «Пациенты»

Код обследования	Код пациента	НbA1C,%	ППГ, ммоль/л	ТГ, ммоль/л	Ац	Дата обследования	Код специалиста
Имя поля	Тип данных	Описание (необязательно)					
Код обследования	Счетчик	уникальный код обследования пациента					
Код пациента	Числовой	код пациента из таблицы ПАЦИЕНТЫ					
НbA1C,%	Числовой						
ППГ, ммоль/л	Числовой						
ТГ, ммоль/л	Числовой						
Ац	Числовой						
Дата обследования	Дата и время	дата обследования пациента					
Код специалиста	Числовой	код специалиста из таблицы СПЕЦИАЛИСТЫ					

Рисунок 2.8 – Таблица «Обследования»

Код диагноза	Код пациента	Код болезни	Дата установления диагноза	Код специалиста	Класс состояния здоров
Имя поля	Тип данных	Описание			
Код диагноза	Числовой	Код диагноза — уникальный код диагноза [Числовое поле];			
Код пациента	Числовой	Код пациента — код пациента из таблицы ПАЦИЕНТЫ [Числовое поле];			
Код болезни	Числовой	Код болезни — код болезни из таблицы БОЛЕЗНИ [Числовое поле];			
Дата установления диагноза	Числовой	Дата установления диагноза — дата составления заключения о болезни пациента [Поле типа дата];			
Код специалиста	Числовой	Код специалиста — код специалиста из таблицы СПЕЦИАЛИСТЫ [Числовое поле]			
Класс состояния здоровья	Текстовый				

Рисунок 2.9 – Таблица «Диагнозы»

Код специалиста	Фамилия	Имя	Отчество	Код специальности	Код категории
Имя поля	Тип данных	Описание			
Код специалиста	Числовой	Код специалиста — уникальный код специалиста [Числовое поле];			
Фамилия	Текстовый	Фамилия — фамилия специалиста [Текстовое поле] (15 знаков);			
Имя	Текстовый	Имя — имя специалиста [Текстовое поле] (15 знаков);			
Отчество	Текстовый	Отчество — отчество специалиста [Текстовое поле] (15 знаков);			
Код специальности	Числовой	Код специальности — код специальности из таблицы СПЕЦИАЛЬНОСТИ [Числовое поле];			
Код категории	Числовой	Код категории — код категории специалиста из таблицы КАТЕГОРИИ [Числовое поле]			

Рисунок 2.10 – Таблица «Специалисты»

Код признака	Код признака по МКБ	Название признака по МКБ	Обозначение признака
Имя поля	Тип данных	Описание (необязательно)	
Код признака	Числовой	уникальный код диагностического признака, выявленного при обследовании пациента [Числовое поле];	
Код признака по МКБ	Короткий текст	код признака по МКБ-10 [Текстовое поле] (6 знаков);	
Название признака по МКБ	Короткий текст	название признака по МКБ-10 [Текстовое поле] (20 знаков);	
Обозначение признака	Короткий текст	условное обозначение признака, содержащее порядковый номер признака и порядковый н	

Рисунок 2.11 – Таблица «Признаки»

Код болезни	Код болезни по МКБ	Название болезни
<input type="text"/>	<input type="text"/>	<input type="text"/>
Имя поля	Тип данных	Описание
Код болезни	Числовой	Код болезни — уникальный код болезни [Числовое поле];
Код болезни по МКБ	Текстовый	Код болезни по МКБ — код болезни по МКБ-10 [Текстовое поле] (6 знаков);
Название болезни	Текстовый	Название болезни по МКБ — название болезни по МКБ-10 [Текстовое поле] (20 знаков);

Рисунок 2.12 – Таблица «Болезни»

Код специальности		Специальность
Имя поля	Тип данных	Описание
Код специальности	Текстовый	Код специальности — уникальный код специальности [Числовое поле];
Специальность	Текстовый	Специальность — название специальности из Номенклатуры специальностей специалистов

Рисунок 2.13 – Таблица «Специальности»

Код категории		Категория
		

Категории		
Имя поля	Тип данных	Описание
Код категории	Числовой	Код категории — уникальный код категории специалиста [Числовое поле];
Категория	Текстовый	Категория — квалификационная категория специалиста [Текстовое поле] (6 знаков)

Рисунок 2.14 – Таблица «Категории»

Код услуги	Шифр услуги	Название услуги
<input type="text"/>	<input type="text"/>	<input type="text"/>
Имя поля	Тип данных	Описание (необязательно)
Код услуги	Числовой	Уникальный код услуги [Числовое поле];
Шифр услуги	Короткий текст	Буквенно-цифровой шифр услуги по «Номенклатуре медицинских услуг» [Текстовое поле];
Название услуги	Короткий текст	Название медицинской услуги по «Номенклатуре медицинских услуг» [Текстовое поле]

Рисунок 2.15 – Таблица «Услуги»

Обеспечение целостности взаимосвязанных данных при вводе и корректировке данных в таблицах, упрощение конструирования многотабличных форм, запросов, отчетов возможно благодаря разработке взаимосвязей между таблицами в схеме данных. Конечная схема данных представлена на рисунке 2.16.

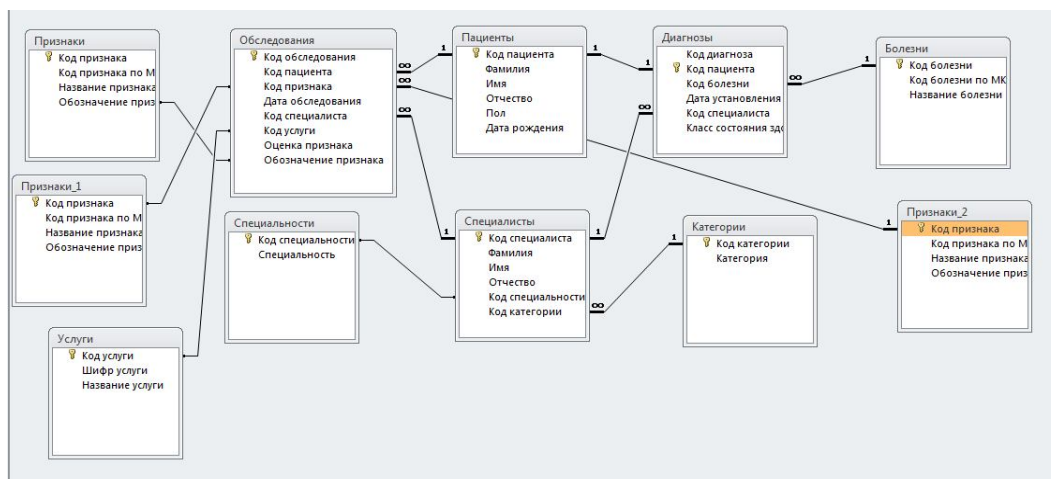


Рисунок 2.16– Схема данных

На основе сформированных таблиц созданы одноименные формы (рисунок 2.17). Благодаря формам обеспечивается ускорение ввода данных в таблицы. Формы можно использовать для ввода, изменения или отображения данных, для управления доступом к данным: с их помощью можно определять, какие поля или строки данных будут отображаться.

Посредством форм произведено заполнение таблиц данными из историй болезней о пациентах, о результатах обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа, о врачах, установивших заключительный диагноз (рисунок 2.18).

Пациенты	
Код пациента:	16823
Фамилия:	Соломянов
Имя:	Владимир
Отчество:	Вячеславович
Пол:	М
Дата рождения:	05.09.1964

Рисунок 2.17 – Форма «Пациенты»

	Код пациента	Фамилия	Имя	Отчество	Пол	Дата рождения
+	16136	Ишмухаметов	Сания	Сулейманова	Ж	24.08.1956
+	16227	Шарапова	Нина	Григорьевна	Ж	04.06.1953
+	16338	Ежкова	Ольга	Павловна	Ж	28.09.1959
+	16343	Романцова	Елена	Александровна	Ж	23.03.1964
+	16344	Мирясова	Валентина	Сергеевна	Ж	26.01.1970
+	16430	Белов	Валентина	Геннадьевич	М	03.01.1957
+	16438	Шишкова	Нина	Ивановна	Ж	05.01.1957
+	16456	Егорова	Татьяна	Владимировна	Ж	02.07.1961
+	16473	Поляков	Валерий	Иванович	М	15.09.1950
+	16496	Горохов	Сергей	Вячеславович	М	15.07.1996
+	16530	Акинина	Елена	Анатольевна	Ж	06.12.1984
+	16583	Кулькова	Ирина	Викторовна	Ж	02.06.1963
+	16601	Блохина	Любовь	Николаевна	Ж	01.08.1949
+	16634	Кожевников	Евгений	Валентинович	М	13.08.1967
+	16652	Осипочева	Марина	Николаевна	Ж	06.05.1992

Рисунок 2.18 – Заполненная таблица «Пациенты»

Операции обработки данных (сортировка, фильтрация, объединение данных из разных источников, преобразование данных) осуществляются с помощью запросов (таблица 2.2).

Таблица 2.2 – Запросы данных

Название запроса	Результат выполнения запроса
1 «Заболевания»	Таблица, содержащая сведения о классах состояния здоровья каждого пациента
2 «Результаты лабораторного обследования»	Таблица, содержащая сведения об установленных диагнозах и значениях диагностических признаков каждого пациента. Полученная таблица экспортируется в решатель MS Excel с целью статистической обработки исходных данных (рисунок 2.19)
3 «Число больных»	Таблица, содержащая сведения о суммарном числе пациентов по каждому классу состояния здоровья
4 «Результаты обследования»	Таблица, содержащая сведения о числе пациентов по каждому классу состояния здоровья с обнаруженным значением каждого признака состояния здоровья
5 «Результаты и диагнозы»	Таблица, содержащая сведения о суммарном числе пациентов по каждому классу состояния здоровья с обнаруженным значением каждого признака состояния здоровья. Полученная таблица используется для построения диагностической модели (рисунок 2.20).

Код пациента	Название болезни	HbA1C, %	ППГ, ммоль	ТГ, ммоль/л	Ац
16343	Сахарный диабет 1 типа	9,45	10,70	0,66	0
16496	Сахарный диабет 1 типа	13,45	13,40	1,65	0
16530	Сахарный диабет 1 типа	8,67	10,50	2,94	1
16652	Сахарный диабет 1 типа	9,06	10,30	1,45	0
16783	Сахарный диабет 1 типа	6,80	13,80	1,40	0
16823	Сахарный диабет 1 типа	10,50	18,30	0,82	0
17313	Сахарный диабет 1 типа	7,91	9,30	0,98	0
17404	Сахарный диабет 1 типа	10,49	14,10	0,81	0
17418	Сахарный диабет 1 типа	10,71	20,90	1,43	0
17563	Сахарный диабет 1 типа	8,62	16,40	1,06	0
17669	Сахарный диабет 1 типа	10,62	12,90	2,07	0
17703	Сахарный диабет 1 типа	13,16	13,90	0,56	0
17794	Сахарный диабет 1 типа	7,09	13,30	0,99	0
17882	Сахарный диабет 1 типа	8,21	16,10	2,19	0
17896	Сахарный диабет 1 типа	12,40	24,10	2,77	1
18079	Сахарный диабет 1 типа	7,47	11,70	1,49	0
18429	Сахарный диабет 1 типа	8,20	12,50	0,88	0
18659	Сахарный диабет 1 типа	8,70	8,90	2,11	0
18973	Сахарный диабет 1 типа	6,40	11,00	1,21	0
19114	Сахарный диабет 1 типа	7,02	9,40	1,20	0
19231	Сахарный диабет 1 типа	7,66	10,20	0,92	0
19529	Сахарный диабет 1 типа	11,09	8,50	2,60	0
19612	Сахарный диабет 1 типа	10,11	17,10	1,62	0
19653	Сахарный диабет 1 типа	9,88	15,70	0,50	0

Рисунок 2.19 – Таблица с результатами запроса «Результаты лабораторного обследования»

Название болезни	Count	u11	u12	u13	u14	u15	u16	u17
Обследование (здоров)	30	18	13					
Сахарный диабет 1 типа	27			4	7	6	6	1
Сахарный диабет 2 типа	28		1	8	9	5	4	1

Рисунок 2.20 – Таблица с результатами запроса «Результаты и диагнозы»

Автоматизации часто выполняемых задач достигается за счет применения макросов. В базе данных макросы используются для автоматизации открытия объектов (таблиц, форм, запросов, отчетов), а также экспорта данных таблицы «Результаты лабораторного обследования» из Access в книгу Excel (рисунок 2.21–2.23).

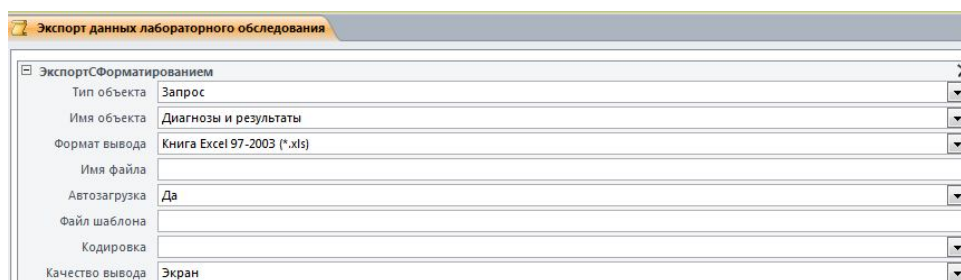


Рисунок 2.21 – Макрос экспорта данных

При выполнении макроса на экспорт появляется окно для сохранения файла в формате xls (рисунок 2.22):

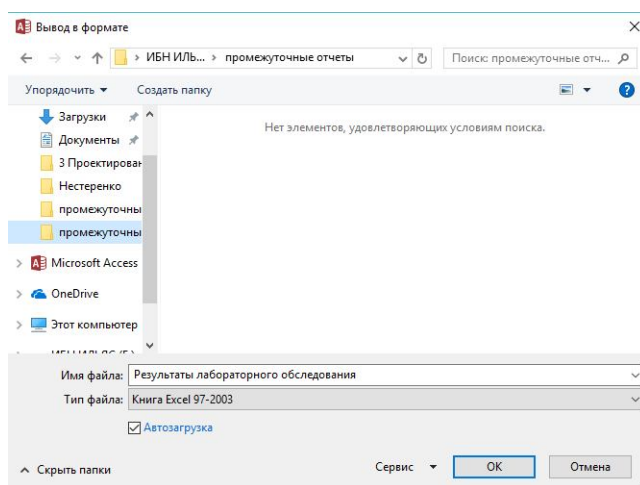


Рисунок 2.22 – Экспорт данных

Оставляем автозагрузку по умолчанию. Открывается среда Excel с экспортированным файлом (рисунок 2.23):

	А	В	С	Д	Е	Ф
	Код пациента	Название болезни	НbA1C, %	ППг, ммоль/л	ТТ, ммоль/л	Ас
1	16343	Сахарный диабет 1 типа	9,45	10,70	0,66	0
2	16496	Сахарный диабет 1 типа	13,45	13,40	1,65	0
3	16530	Сахарный диабет 1 типа	8,67	10,50	2,94	1
4	16652	Сахарный диабет 1 типа	9,06	10,30	1,45	0
5	16783	Сахарный диабет 1 типа	6,80	13,80	1,40	0
6	16823	Сахарный диабет 1 типа	10,50	18,30	0,82	0
7	17313	Сахарный диабет 1 типа	7,91	9,30	0,98	0
8	17404	Сахарный диабет 1 типа	10,49	14,10	0,81	0
9	17418	Сахарный диабет 1 типа	10,71	20,90	1,43	0
10	17563	Сахарный диабет 1 типа	8,62	16,40	1,06	0
11	17669	Сахарный диабет 1 типа	10,62	12,90	2,07	0
12	17703	Сахарный диабет 1 типа	13,16	13,90	0,56	0
13	17794	Сахарный диабет 1 типа	7,09	13,30	0,99	0
14	17882	Сахарный диабет 1 типа	8,21	16,10	2,19	0
15	17896	Сахарный диабет 1 типа	12,40	24,10	2,77	1
16	18079	Сахарный диабет 1 типа	7,47	11,70	1,49	0
17	18429	Сахарный диабет 1 типа	8,20	12,50	0,88	0
18	18659	Сахарный диабет 1 типа	8,70	8,90	2,11	0
19	18973	Сахарный диабет 1 типа	6,40	11,00	1,21	0
20	19114	Сахарный диабет 1 типа	7,02	9,40	1,20	0
21	19231	Сахарный диабет 1 типа	7,66	10,20	0,92	0
22	19529	Сахарный диабет 1 типа	11,09	8,50	2,60	0
23						

Рисунок 2.23– Экспортированные данные таблицы «Результаты лабораторного обследования»

Вывод структурированной информации, хранимой в базе данных, на экран, в печать или файл осуществляется с помощью отчётов. База данных пациентов с сахарным диабетом 1 и 2 типа содержит отчеты, полученные на основе таблиц и запросов (рисунок 2.24).

Результаты лабораторного исследования					20 мая 2017 г. 20:39:31
Код пациента	Название болезни	HbA1C,%	ППГ, ммоль/л	ТТ, ммоль/л	Ац
16343	Сахарный диабет 1 типа	9,45	10,70	0,66	0
16496	Сахарный диабет 1 типа	13,45	13,40	1,65	0
16530	Сахарный диабет 1 типа	8,67	10,50	2,94	1
16652	Сахарный диабет 1 типа	9,06	10,30	1,45	0
16783	Сахарный диабет 1 типа	6,80	13,80	1,40	0
16823	Сахарный диабет 1 типа	10,50	18,30	0,82	0
17313	Сахарный диабет 1 типа	7,91	9,30	0,98	0
17404	Сахарный диабет 1 типа	10,49	14,10	0,81	0
17418	Сахарный диабет 1 типа	10,71	20,90	1,43	0
17563	Сахарный диабет 1 типа	8,62	16,40	1,06	0
17669	Сахарный диабет 1 типа	10,62	12,90	2,07	0

Рисунок 2.24 – Отчет «Результаты лабораторного обследования»

Таким образом, база данных служит для накопления и обработки сведений, полученных из историй болезней. В реляционные таблицы базы данных занесены сведения пациентов с сахарным диабетом 1 и 2 типа, полученные из историй болезней за период с 1 июля по 31 августа 2016 года на базе эндокринологического отделения ПОКБ им. Н.Н. Бурденко. Применение базы данных, содержащей результаты обследований и достоверные диагнозы пациентов, позволяет повысить достоверность диагноза и ограничить требования к опыту специалистов, за счёт принятия решений на основе опыта постановки диагноза высококвалифицированных специалистов, накопленного в базе данных.

Ввод, изменение и отображение данных реляционных таблиц осуществляется посредством форм. Применение форм позволяет упростить и ускорить работу пользователя с базой данных. Автоматизация работы с базой данных достигается использованием макросов. За счет операций экспорта и импорта осуществляется обмен данных между решателем и базой

данных. Таблица «Результаты лабораторного обследования», получаемая в результате выполнения одноименного запроса, с помощью макроса ЭКСПОРТ С ФОРМАТИРОВАНИЕМ экспортируются в решатель MS Excel, в котором осуществляется статистическая обработка данных (раздел 3). После выполнения статистической обработки данных в решателе таблица импортируются в базу данных, и на основе обработанных данных формируются запрос «Результаты и диагнозы». В результате выполнения данного запроса формируется таблица с исходными данными для построения диагностической модели.

Ограничение требований к квалификации специалистов в области компьютерных технологий достигается за счёт применения «дружественного» к специалисту графического интерфейса пользователя.

2.5 Пользовательский интерфейс базы данных

Организации эффективной работы пользователя с базой данных обеспечивается целостным приложением, в котором все компоненты базы данных сгруппированы по функциональному назначению. При этом обеспечивается удобный графический интерфейс пользователя. Приложение дает возможность пользователю работать с базой данных, затрачивая при этом меньше усилий.

При создании приложения пользователя главную роль играют формы, так как именно они являются основным диалоговым средством работы пользователя с базой данных. Для объединения объектов в едином диалоговом приложении создана главная кнопочная форма. По существу, кнопочная форма представляет собой панель управления приложением. Кнопки такой формы обеспечивают вызов других кнопочных форм, а также отдельных объектов базы данных – таблиц, отчетов, запросов, форм и т.д. Здесь же находятся кнопки для возврата к кнопочным формам более высокого уровня и для выхода из базы данных.

Дополнительным элементом приложения пользователя является организация вызова главной кнопочной формы при открытии базы данных, что позволяет пользователю сразу же начать работу в среде приложения.

Главная кнопочная форма создается с помощью диспетчера кнопочных форм. Для создания главной кнопочной формы необходимо выбрать команду «Сервис» «Служебные программы» «Диспетчер кнопочных форм». В окне «Диспетчер кнопочных форм» следует подтвердить создание кнопочной формы. В следующем окне появится окно со строкой «Главная кнопочная форма». В этом окне формируется список кнопочных форм разных уровней.

Для создания кнопочной формы нужно нажать кнопку «Создать». В окне «Создание» в поле «Имя страницы кнопочной формы» нужно ввести имя кнопочной формы – «База данных пациентов с сахарным диабетом 1 и 2 типа». После нажатия кнопки «ОК» в окне диспетчера кнопочных форм появится строка с именем созданной кнопочной формы. Затем необходимо создать страницы главной кнопочной формы (рисунок 2.25).

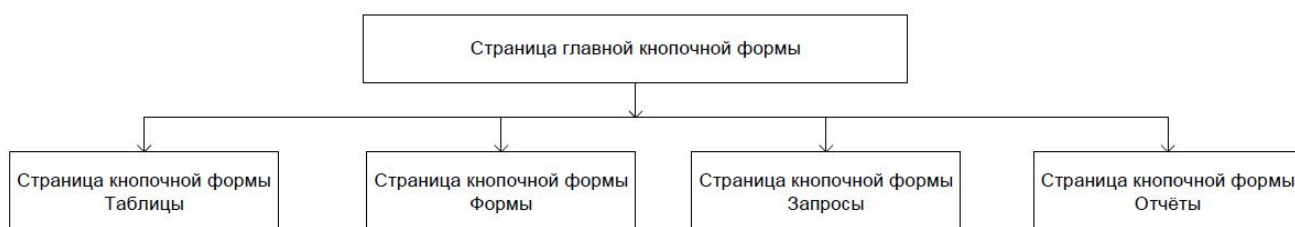


Рисунок 2.25 – Структура главной кнопочной формы

Названия страниц служат названиями кнопок, которые вызывают соответствующую страницу с хранимой в ней информацией (рисунок 2.26 – 2.27).



Рисунок 2.26 – Пользовательский интерфейс базы данных, основное меню



Рисунок 2.27 – Пользовательский интерфейс базы данных, страница
«Таблицы»

Таким образом, создан интуитивно понятный для пользователя интерфейс, в полной мере удовлетворяющий его потребности. Благодаря интерфейсу работа пользователя с базой данных становится простой в освоении. Простота освоения базы данных пользователем

достигается за счет применения меню базы данных (главной кнопочной формы), позволяющей экономить время на поиск нужных для пользователя объектов и легко ориентироваться в базе данных. Интерфейс позволяет сэкономить время пользователя на выполнение задач, решаемых с помощью базы данных (ввод, обработка, просмотр данных). Субъективная удовлетворенность потребителя достигается благодаря приятному визуальному дизайну интерфейса (рисунок 2.26 – 2.27).

2.6 Основные результаты и выводы

1 Информационная система распознавания сахарного диабета относится к системам базового уровня, поскольку предназначена для информационного обеспечения принятия решений в профессиональной деятельности врачей. Согласно функциональной классификации ИС разрабатываемая информационная система относится к классу медико-технологических, а именно к классу информационных экспертных систем. Основными составными частями экспертной системы являются: база данных, решатель, пользовательский интерфейс. Разрабатываемая информационная экспертная система предназначена для поддержки принятия решений специалистам в области эндокринной патологии по дифференциальной диагностике двух форм сахарного диабета, а также для обучения студентов медицинских вузов. Данная экспертная система реализуется на персональном компьютере, по степени интеграции относится к гибридным, поскольку представляет программный комплекс из следующих прикладных программ: Microsoft Access, Microsoft Excel, MatLab.

2 Анализ существующих экспертных систем распознавания сахарного диабета позволил установить, что в анализируемых экспертных системах базы знаний не содержат в своей структуре данных результатов обследования и диагнозов пациентов. В экспертных системах для задач

распознавания используются субъективные жалобы, которые имеют более низкую информативность по сравнению с результатами лабораторного обследования. 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 сахарного диабета пациентов, а также решатель для формирования вероятностной диагностической модели является актуальной задачей.

3 Предлагаемая информационная система состоит из базы данных MS Access, решателя MS Excel, решателя MatLab, пользовательского интерфейса (рисунок 2.5). База данных предназначена для хранения, накопления и обработки результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа, сведений о пациентах и медперсонале. Результаты обследования пациентов экспортируются в решатель MS Excel, в котором производится статистическая обработка данных, целью которой является группировка, определение распределения и его параметров, оценка корреляции признаков. Из решателя MS Excel данные загружаются в базу данных MS Access для формирования запроса, результатом выполнения которого является таблица. В строках таблицы отражаются формы сахарного диабета, в столбцах – интервалы значений признаков, в ячейках – количество пациентов, у которых класс состояния здоровья проявляется значениями признаков из соответствующих интервалов. Полученная таблица экспортируется в решатель Matlab. Решатель предназначен для формирования диагностической модели и расчета апостериорных вероятностей с помощью метода Байеса. Работа со всеми элементами системы осуществляется через пользовательский интерфейс.

4 База данных служит для накопления и обработки сведений, полученных из историй болезней. В реляционные таблицы базы данных занесены сведения пациентов с сахарным диабетом 1 и 2 типа, полученные

из историй болезней за период с 1 июля по 31 августа 2016 года на базе эндокринологического отделения ПОКБ им. Н.Н. Бурденко. Применение базы данных, содержащей результаты обследований и достоверные диагнозы пациентов, позволяет повысить достоверность диагноза и ограничить требования к опыту специалистов, за счёт принятия решений на основе опыта постановки диагноза высококвалифицированных специалистов, накопленного в базе данных. Ввод, изменение и отображение данных реляционных таблиц осуществляется посредством форм. Применение форм позволяет упростить и ускорить работу пользователя с базой данных. Автоматизация работы с базой данных достигается использованием макросов. За счет операций экспорта и импорта осуществляется обмен данных между решателем и базой данных. Таблица «Результаты лабораторного обследования», получаемая в результате выполнения одноименного запроса, с помощью макроса ЭкспортсФорматированием экспортируются в решатель MS Excel, в котором осуществляется статистических обработка данных (раздел 3). После выполнения статистической обработки данных в решателе таблица импортируются в базу данных, и на основе обработанных данных формируются запрос «Результаты и диагнозы». В результате выполнения данного запроса формируется таблица с исходными данными для построения диагностической модели.

5 Разработан интуитивно понятный для пользователя интерфейс, в полной мере удовлетворяющий его потребности. Благодаря интерфейсу работа пользователя с базой данных становится простой в освоении. Простота освоения базы данных пользователем достигается за счет применения меню базы данных (главной кнопочной формы), позволяющей экономить время на поиск нужных для пользователя объектов и легко ориентироваться в базе данных. Интерфейс позволяет сэкономить время пользователя на выполнение задач, решаемых с помощью базы данных (ввод, обработка, просмотр данных). Субъективная удовлетворенность потребителя

достигается благодаря приятному визуальному дизайну интерфейса (рисунок 2.26 – 2.27).

6 Повышение достоверности распознавания состояния здоровья методом Байеса возможно благодаря наличию информации о зависимости и распределениях вероятностей значений диагностических признаков. Для получения сведений о зависимости и распределениях вероятностей значений диагностических признаков в среде MS Excel выполняется статистическая обработка данных лабораторного обследования пациентов, которая включает в себя экспорт данных из MS Access в MS Excel, группировку, построение гистограммы распределения частот, расчет общих статистических характеристик, определение законов распределения, оценку корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков (раздел 3).

3 ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ ОБСЛЕДОВАНИЯ В EXCEL

3.1 Построение вариационных рядов значений признаков

Анализ литературных источников показал, что рассматриваемые диагностические признаки подчиняются нормальному распределению. Одной из целей статистической обработки является подтверждение нормального распределения признаков, поскольку математический аппарат метода Байеса может быть применен к диагностическим признакам распределенных нормально.

Статистическая обработка данных лабораторного обследования пациентов является одним из этапов формирования вероятностной диагностической модели и включает в себя экспорт данных из MS Access в MS Excel, группировку, построение гистограммы распределения частот, расчет общих статистических характеристик, определение распределения, оценку корреляции диагностических признаков, вычисление вероятностей интервалов значений признаков, экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.1).

Экспорт данных результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа из базы данных в решатель осуществляется с помощью макроса Экспорт с форматированием (рисунок 2.23). Данные пациентов с сахарным диабетом 1 и 2 типа получены из историй болезней за период с 1 июля по 31 августа 2016 года на базе эндокринологического отделения ПОКБ им. Н.Н. Бурденко.

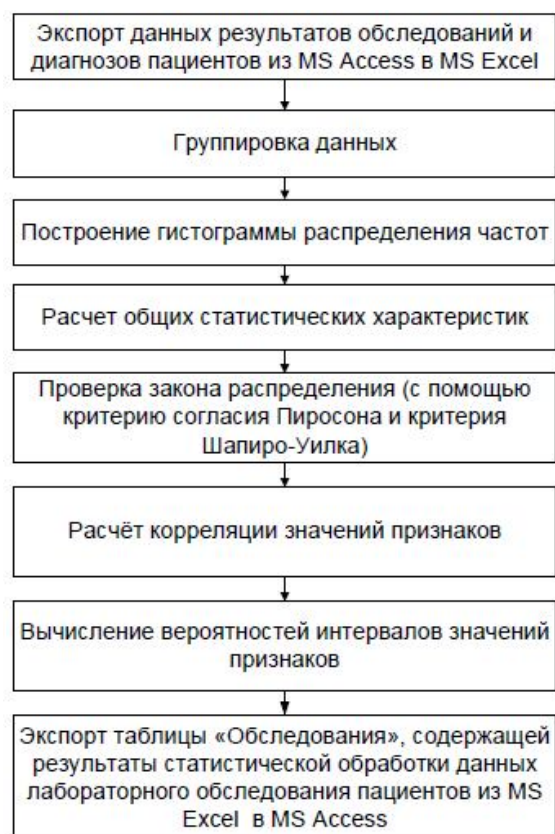


Рисунок 3.1 – Этапы методики статистической обработки данных в решателе MS Excel

В данной работе для группы здоровых пациентов используются сгенерированные данные на основе известных для каждого показателя распределений и средних значений показателей (таблица 3.1).

Таблица 3.1 – Средние значения показателей для здоровых пациентов

Показатель	Гликированный гемоглобин (HbA1C) [21]	Постприандиальная гликемия [20]	Триглицериды
Здоровые пациенты	4,58 %	4,74 ммоль/л	1,18 ммоль/л

На основании данных средних значений для здоровых пациентов в среде Excel, с использованием пакета «Анализ данных», были сгенерированы выборки со следующими параметрами (таблица 3.2):

Таблица 3.2 – Параметры генерации выборок для здоровых пациентов

Параметр генерации	HbA1C, %	ППГ, ммоль/л	ТГ, ммоль/л
Число переменных	1		
Число случайных чисел	30		
Распределение	нормальное		
Среднее	4,58	4,74	1,18
Стандартное отклонение	0,44	0,13	0,84

После генерации результаты были округлены до двух знаков после запятой. Результат генерации с последующим округлением для здоровых пациентов показан на рисунке 3.2:

	А	В	С
1	HbA1C здоровые	ППГ здоровые	ТГ здоровые
2	4,91	4,68	0,88
3	4,18	4,64	1,24
4	4,37	4,70	0,41
5	5,30	4,71	0,58
6	5,07	4,88	0,09
7	4,83	5,05	2,34
8	4,15	4,76	2,09
9	4,79	4,80	1,32
10	4,80	4,78	0,50
11	4,29	4,72	0,81
12	4,83	4,45	0,85
13	4,23	4,78	0,29
14	4,34	4,85	1,84
15	4,12	4,76	1,08
16	4,95	4,91	0,06
17	4,37	4,81	1,21
18	4,64	4,68	2,39
19	5,01	4,86	0,00
20	4,49	4,63	1,51
21	4,61	4,90	1,77
22	4,34	4,89	1,41
23	4,38	4,84	0,27
24	4,63	4,88	0,32
25	4,26	4,76	0,68
26	4,86	4,58	2,34
27	4,63	4,76	0,59

Рисунок 3.2 – Генерация выборок для здоровых пациентов

Статистическому анализу подвергаются следующие тестовые выборки:

1.1) выборка здоровых пациентов, HbA1C (гликированный гемоглобин);

- 1.2) выборка здоровых пациентов, ППГ (постпрандиальная гликемия);
- 1.3) выборка здоровых пациентов, ТГ (триглицериды);
- 2.1) выборка пациентов с сахарным диабетом 1 типа;
- 2.2) выборка пациентов с сахарным диабетом 1 типа, ППГ;
- 2.3) выборка пациентов с сахарным диабетом 1 типа, ТГ;
- 3.1) выборка пациентов с сахарным диабетом 2 типа, HbA1C;
- 3.2) выборка пациентов с сахарным диабетом 2 типа, ППГ;
- 3.3) выборка пациентов с сахарным диабетом 2 типа, ТГ.

Каждая выборка пациентов размещается на отдельном листе книги MS Excel (рисунок 3.3).

	А	В	С	Д
1	Hb, % СД1			
2	10,50			
3	6,80			
4	9,06			
5	8,67			
6	7,47			
7	9,88			
8	11,09			
9	9,40			
10	7,02			
11	6,40			
12	8,20			
13	12,40			
14	7,09			
15	10,62			
16	8,62			
17	10,49			

Рисунок 3.3 – Выборка пациентов с сахарным диабетом 1 типа
(гликированный гемоглобин, %)

Вторым этапом является систематизация собранных данных. Процесс систематизации результатов наблюдений, объединения их в относительно однородные группы по некоторому признаку называется группировкой данных.

Вариационным рядом (рядом распределения) называют двойной ряд чисел, отражающий соотношение ранжированных значений признака с частотой их встречаемости в данной выборке. Значение ряда позволяет

определить закономерность варьирования (закон распределения) изучаемого признака.

В зависимости от того, в каком диапазоне и как варьирует признак – дискретно или непрерывно, – статистическая совокупность может распределяться в безынтервальный или интервальный вариационные ряды. 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.

Таким образом, если $\lambda = 1$ или $\lambda \approx 1$, то строится безынтервальный ряд, если $\lambda \neq 1$, то строится интервальный ряд [32].

При построении интервального вариационного ряда следует поступать так, чтобы минимальная варианта совокупности попадала примерно в середину первого классического интервала. Этому требованию удовлетворяет формула:

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

где x_n – нижняя граница первого классического интервала;

x_{min} – минимальная варианта совокупности;

λ – величина классového интервала.

Наметив классовые интервалы, распределяют по ним все варианты совокупности, т.е. определяют частоты каждого класса.

Последовательно прибавляя к x_n λ , ряд разбивается на классовые интервалы. Путем уменьшения верхних границ на величину, равную точности, принятой при измерении признака, достигается необходимое разграничение классов.

Наиболее точно центральную величину классového интервала можно получить по формуле [33]:

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

где x_i – центральная величина классového интервала;

x_n – нижняя граница первого классového интервала;

x_k – конечная точка интервала.

Рассмотрим обработку данных для пациентов с сахарным диабетом 1 типа по значению HbA1C (остальные выборки обрабатываются совершенно аналогично).

Перенесем на лист «СД 1, HbA1C» соответствующие значения признака в столбец А (ячейки A2:A28). 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 классového интервала по формуле (3.1). В среде Excel она примет вид «=(В4–В2)/(1+3,32*LOG10(СЧЁТ(А:А)))».

Определим нижнюю границу первого классového интервала по формуле (3.3). Для этого в ячейку В8 запишем формулу «=В2–0,5*В6». В столбцах С и D произведем разбивку ряда на классовые интервалы (рисунок 3.3):

	C	D
1	Xd	Xu
2	=B\$8	=C3-0,01
3	=C2+B\$6	=C4-0,01
4	=C3+B\$6	=C5-0,01
5	=C4+B\$6	=C6-0,01
6	=C5+B\$6	=C7-0,01
7	=C6+B\$6	=C8-0,01
8	=C7+B\$6	=C9-0,01
9	=C8+B\$6	

Рисунок 3.3 – Разбивка ряда на классовые интервалы в среде Excel

Рассчитаем центральные величины классовых интервалов по формуле (3.4), как это показано на рисунке 3.4:

	G
1	X_i
2	$=(C2+D2)/2$
3	$=(C3+D3)/2$
4	$=(C4+D4)/2$
5	$=(C5+D5)/2$
6	$=(C6+D6)/2$
7	$=(C7+D7)/2$
8	$=(C8+D8)/2$

Рисунок 3.4 – Расчет центральных величин классовых интервалов в Excel

Для того чтобы более наглядно представить закономерность варьирования количественных признаков, вариационные ряды принято изображать в виде графиков. По оси абсцисс откладывают срединные значения классов, по оси ординат – частоты. Частоты – это абсолютные числа, показывающие, сколько раз в совокупности встречается данное значение признака. Сумма всех частот должна быть равна численности единиц всей совокупности. Высота перпендикуляров, восставляемых по оси абсцисс, соответствует частотам классов.

Расчёт частот в среде Excel приведен на рисунке 3.5.

	E
1	fi
2	=СЧЁТЕСЛИМН(A:A;">="&C2;A:A;"<="&D2)
3	=СЧЁТЕСЛИМН(A:A;">="&C3;A:A;"<="&D3)
4	=СЧЁТЕСЛИМН(A:A;">="&C4;A:A;"<="&D4)
5	=СЧЁТЕСЛИМН(A:A;">="&C5;A:A;"<="&D5)
6	=СЧЁТЕСЛИМН(A:A;">="&C6;A:A;"<="&D6)
7	=СЧЁТЕСЛИМН(A:A;">="&C7;A:A;"<="&D7)
8	=СЧЁТЕСЛИМН(A:A;">="&C8;A:A;"<="&D8)

Рисунок 3.5 – Расчёт частот в Excel

Для построения гистограммы распределения частот по оси абсцисс откладывают границы классовых интервалов, по оси ординат – частоты интервалов [34].

Гистограмма распределения частот значений уровня HbA1C у пациентов с сахарным диабетом 1 типа приведена на рисунке 3.6.



Рисунок 3.6 – Гистограмма распределения частот у пациентов с сахарным диабетом 1 типа

Таким образом, исходные данные для формирования выборок значений диагностических признаков пациентов с сахарным диабетом 1 и 2 типа

содержатся в таблице «Результаты лабораторного обследования» (рисунок 2.33), экспортируемой из MS Access в MS Excel. Модельные выборки здоровых пациентов сгенерированы в среде Excel в соответствии с параметрами нормального распределения, полученными из литературных источников. Методика обработки включает в себя экспорт данных из MS Access в MS Excel, группировку, 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 Вычисление характеристик распределений

Для полного описания варьирующих объектов используют статистические обобщающие характеристики. К ним относят, прежде всего, средние величины и показатели вариации (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. Мода не зависит от крайних значений вариантов и может применяться для характеристики центра в рядах распределения с неопределенными границами

Продолжение таблицы 3.3

Асимметрия	Величина, характеризующая асимметрию распределения данной случайной величины. Величина асимметрии может быть положительной (для правосторонней асимметрии) и отрицательной (для левосторонней асимметрии). Применение данного показателя дает возможность определить не только величину асимметрии, но и проверить ее наличие в генеральной совокупности.
Эксцесс	Для симметричных распределений может быть рассчитан показатель эксцесса, который показывает, насколько резкий скачок имеет изучаемое явление [33]

В среде Excel существуют готовые функции нахождения указанных выше характеристик. На рисунке 3.7 показан расчет с использованием данных возможностей.

	Н
1	Объем выборки
2	=СЧЁТ(А:А)
3	Выборочное среднее
4	=СРЗНАЧ(А2:А31)
5	Дисперсия
6	=ДИСП(А:А)
7	СКО
8	=СТАНДОТКЛОН(А:А)
9	Медиана
10	=МЕДИАНА(А:А)
11	Мода
12	=МОДА(А:А)
13	Эксцесс
14	=ЭКСЦЕСС(А:А)
15	Ассиметрия
16	=СКОС(А:А)

Рисунок 3.7 – 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.

Таблица 3.4 – Общие статистические характеристики для диагностических признаков сахарного диабета

Диагностические признаки	Здоровое состояние			Сахарный диабет 1 типа			Сахарный диабет 2 типа		
	HbA1C, %	ППГ, ммоль/л	ТГ, ммоль/л	HbA1C, %	ППГ, ммоль/л	ТГ, ммоль/л	HbA1C, %	ППГ, ммоль/л	ТГ, ммоль/л
Объем выборки	30			27			28		
Выборочное среднее	4,62	4,77	1,24	9,19	13,51	1,32	8,23	11,41	3,65
Дисперсия	0,1	0,01	0,8	3,64	13,72	0,53	2,58	17,82	3,45
Стандартное отклонение	0,31	0,12	0,9	1,91	3,7	0,73	1,61	4,22	1,86
Медиана	4,63	4,77	1,15	8,7	13,3	1,2	8,08	10,6	3,47
Мода	–	–	–	7,47	13,4	–	–	13	4,57
Экссесс	0,85	1,34	-0,81	-0,16	1,39	-0,17	-0,24	-0,77	0,52
Ассиметрия	0,09	-0,38	0,49	0,68	1,05	0,76	0,59	0,49	0,85

Проверка распределения на нормальность может осуществляться по нескольким критериям (таблица 3.5).

Таблица 3.5 – 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 [32]	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. [35]. Необходимо, чтобы частоты значений признака в крайних классах были ≥ 5 [32]
Критерий Шапиро - Уилка	Считается наиболее мощным критерием, если объем выборки небольшой ($8 \leq n \leq 50$) [32]	Малые выборки $n \leq 8$ при обнаружении отклонений от нормального распределения не дают достоверных результатов [36]
Критерий Колмогорова – Смирнова	Обладает большей мощностью для непрерывных переменных [32]	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 [35]

В данной работе использовано два из них: критерий согласия Пирсона и критерий Шапиро – Уилка, наиболее подходящий для малых выборок.

Критерий согласия Пирсона (χ^2) применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению $F(x)$. Критерий применим для любых видов функции $F(x)$, даже при неизвестных значениях их параметров. В этом заключается его универсальность.

Использование критерия χ^2 предусматривает разбиение размаха варьирования выборки на интервалы и определения числа наблюдений (частоты) для каждого из интервалов. Для удобства оценок параметров распределения интервалы выбирают одинаковой длины.

Недостатком критерия согласия Пирсона является потеря части первоначальной информации, связанная с необходимостью группировки результатов наблюдений в интервалы и объединения отдельных интервалов с малым числом наблюдений. В связи с этим рекомендуется дополнять проверку соответствия распределений по критерию χ^2 другими критериями. Особенно это необходимо при сравнительно малом объеме выборки [34].

Проверка по данному критерию в среде Excel показана на рисунке 3.8 – 3.9.

	I	J	K	L	M	N	O
1	Проверка на нормальность по критерию Пирсона						
2	$F(x_{i_min})$	$F(x_{i_max})$	p_i	$n \cdot p_i$	$p_i - n \cdot p_i$	$(p_i - n \cdot p_i)^2 / n \cdot p_i$	
3	=НОРМРАСП(C2;\$H\$4;\$H\$8;1)	=НОРМРАСП(D2;\$H\$4;\$H\$8;1)	=J3-I3	=K3*\$H\$2	=E2-L3	=M3^2/L3	
4	=НОРМРАСП(C3;\$H\$4;\$H\$8;1)	=НОРМРАСП(D3;\$H\$4;\$H\$8;1)	=J4-I4	=K4*\$H\$2	=E3-L4	=M4^2/L4	
5	=НОРМРАСП(C4;\$H\$4;\$H\$8;1)	=НОРМРАСП(D4;\$H\$4;\$H\$8;1)	=J5-I5	=K5*\$H\$2	=E4-L5	=M5^2/L5	
6	=НОРМРАСП(C5;\$H\$4;\$H\$8;1)	=НОРМРАСП(D5;\$H\$4;\$H\$8;1)	=J6-I6	=K6*\$H\$2	=E5-L6	=M6^2/L6	
7	=НОРМРАСП(C6;\$H\$4;\$H\$8;1)	=НОРМРАСП(D6;\$H\$4;\$H\$8;1)	=J7-I7	=K7*\$H\$2	=E6-L7	=M7^2/L7	
8	=НОРМРАСП(C7;\$H\$4;\$H\$8;1)	=НОРМРАСП(D7;\$H\$4;\$H\$8;1)	=J8-I8	=K8*\$H\$2	=E7-L8	=M8^2/L8	
9	=НОРМРАСП(C8;\$H\$4;\$H\$8;1)	=НОРМРАСП(D8;\$H\$4;\$H\$8;1)	=J9-I9	=K9*\$H\$2	=E8-L9	=M9^2/L9	
10	=НОРМРАСП(C9;\$H\$4;\$H\$8;1)					Хи-квадрат расчетное	=СУММ(N3:N9)
11						Хи-квадрат критическое	=ХИ2ОБР(0,05;5)
12	т.к. $\chi^2_{кр} > \chi^2_{расчет}$, гипотеза нормальности НЕ ОТКЛОНЯЕТСЯ						

Рисунок 3.8 – Проверка выборки на нормальность по критерию Пирсона
(режим формул)

	I	J	K	L	M	N	O
1	Проверка на нормальность по критерию Пирсона						
2	$F(x_{i_min})$	$F(x_{i_max})$	p_i	$n \cdot p_i$	$p_i - n \cdot p_i$	$(p_i - n \cdot p_i)^2 / n \cdot p_i$	
3	0,04	0,13	0,09	2,38	-0,38	0,06	
4	0,13	0,31	0,18	4,86	4,14	3,53	
5	0,31	0,55	0,25	6,64	-0,64	0,06	
6	0,56	0,78	0,23	6,09	-1,09	0,20	
7	0,78	0,92	0,14	3,75	-1,75	0,82	
8	0,92	0,98	0,06	1,55	-0,55	0,19	
9	0,98	1,00	0,02	0,43	1,57	5,78	
10	1,00					Хи-квадрат расчетное	10,63
11						Хи-квадрат критическое	11,07
12	т.к. $\chi^2_{\text{расчет}} > \chi^2_{\text{критическое}}$, гипотеза нормальности НЕ ОТКЛОНЯЕТСЯ						

Рисунок 3.9 – Проверка выборки на нормальность по критерию Пирсона (результаты расчетов)

Здесь теоретическая (ожидаемая) частота попадания элементов выборки в i -ый интервал группировки для принятой гипотезы о нормальном распределении генеральной совокупности рассчитывается с использованием функции НОРМРАСП. Критическое значение в данном случае имеет распределение с пятью степенями свободы, определяется при помощи функции ХИ2ОБР.

Во всех выборках расчетное значение критерия оказалось меньше критического, следовательно, гипотеза, состоящая в том, что генеральная совокупность подчиняется нормальному закону распределения, принимается.

Критерий Шапиро – Уилка применим при объеме выборки от 8 до 50. Малые выборки с $n < 8$ при обнаружении отклонений от нормального распределения не дают достоверных результатов.

Критерий основан на регрессионном анализе порядковых статистик по их ожидаемым значениям. Это критерий типа дисперсионного анализа для полной выборки. Статистика критерия – отношение квадрата суммы линейной разности выборочных порядковых статистик к обычной оценке дисперсии [35].

Процедуру проверки нулевой гипотезы, сводящейся к предположению, что выборка принадлежит нормально распределенной величине, начинают с построения ранжированного ряда значений в порядке убывания.

По выборочным данным вычисляют сумму квадратов центральных отклонений C^2

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

где C^2 – сумма квадратов центральных отклонений;

X_i – значение из ранжированного ряда;

$\bar{X}$ – среднее выборки.

Далее находят величину k . Для четных выборок она равна половине её объема, в противном случае:

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

где k – искомый индекс;

n – объем выборки.

Вспомогательная величина B вычисляется по формуле (3.7) с использованием табличных коэффициентов a_{n-i+1} .

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

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

Затем вычисляют величину W по формуле (3.9):

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

Значение W служит статистикой для проверки гипотезы о нормальности распределения. Если $W < W_{st}$ то с уровнем значимости α распределение считается отличным от нормального. 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 [36].

В среде Excel расчет данного критерия выглядит следующим образом (рисунок 3.10 – 3.11):

	l	j	k	l	m
14	Проверка на нормальность по критерию Шапиро-Уилка				
15	$(X_i - \bar{X}_{ср})^2$	$\alpha/(n-i+1)$	X_{n-i}	$X_{n-i} - X_i$	$(X_{n-i} - X_i) * \alpha$
16	=A2-\$H\$4)^2	0,4366	6,4	=K16-K42	=J16*L16
17	=A3-\$H\$4)^2	0,3018	6,8	=K17-K41	=J17*L17
18	=A4-\$H\$4)^2	0,2522	7,02	=K18-K40	=J18*L18
19	=A5-\$H\$4)^2	0,2152	7,09	=K19-K39	=J19*L19
20	=A6-\$H\$4)^2	0,1848	7,47	=K20-K38	=J20*L20
21	=A7-\$H\$4)^2	0,1584	7,47	=K21-K37	=J21*L21
22	=A8-\$H\$4)^2	0,1346	7,66	=K22-K36	=J22*L22
23	=A9-\$H\$4)^2	0,1128	7,69	=K23-K35	=J23*L23
24	=A10-\$H\$4)^2	0,0923	7,91	=K24-K34	=J24*L24
25	=A11-\$H\$4)^2	0,0728	8,2	=K25-K33	=J25*L25
26	=A12-\$H\$4)^2	0,054	8,21	=K26-K32	=J26*L26
27	=A13-\$H\$4)^2	0,0358	8,62	=K27-K31	=J27*L27
28	=A14-\$H\$4)^2	0,0158	8,67	=K28-K30	=J28*L28
29	=A15-\$H\$4)^2		8,7	=K29-K29	=J29*L29
30	=A16-\$H\$4)^2		9,06		
31	=A17-\$H\$4)^2	k	9,4		B
32	=A18-\$H\$4)^2	=ОЧЁТ(J16:J28)	9,45		=СУММ(M16:M29)
33	=A19-\$H\$4)^2	Уровень значимости	9,88		B^2
34	=A20-\$H\$4)^2	0,05	10,11		=M32^2
35	=A21-\$H\$4)^2	Wst(для n=27)	10,49		
36	=A22-\$H\$4)^2	0,923	10,5		
37	=A23-\$H\$4)^2	W	10,62	C^2	
38	=A24-\$H\$4)^2	=M34/L38	10,71	=СУММ(J16:J43)	
39	=A25-\$H\$4)^2		11,09	т.к. W>Wst, гипотеза нормальности НЕ ОТКЛОНЯЕТСЯ	
40	=A26-\$H\$4)^2		12,4		
41	=A27-\$H\$4)^2		13,16		
42	=A28-\$H\$4)^2		13,45		
43					

Рисунок 3.10 – Проверка на нормальность по критерию Шапиро-Уилка в Excel (режим формул)

	I	J	K	L	M	N
14	Проверка на нормальность по критерию Шапиро-Уилка					
15	$(X_i - \bar{X}_{ср})^2$	$a(n-i+1)$	X_{n-i}	$X_{n-i} - \bar{X}_i$	$(X_{n-i} - \bar{X}_i) * a$	
16	1,71	0,44	6,40	-7,05	-3,08	
17	5,73	0,30	6,80	-6,36	-1,92	
18	0,02	0,25	7,02	-5,38	-1,36	
19	0,27	0,22	7,09	-4,00	-0,86	
20	2,97	0,18	7,47	-3,24	-0,60	
21	0,47	0,16	7,47	-3,15	-0,50	
22	3,60	0,13	7,66	-2,84	-0,38	
23	0,04	0,11	7,69	-2,80	-0,32	
24	4,72	0,09	7,91	-2,20	-0,20	
25	7,80	0,07	8,20	-1,68	-0,12	
26	0,99	0,05	8,21	-1,24	-0,07	
27	10,28	0,04	8,62	-0,78	-0,03	
28	4,43	0,02	8,67	-0,39	-0,01	
29	2,03		8,70	0,00	0,00	
30	0,33		9,06			
31	1,68	k	9,40		B	
32	18,12	13,00	9,45		-9,44	
33	0,07	уровень значимост	9,88		B^2	
34	1,65	0,05	10,11		89,06	
35	2,30	Wst(для n=27)	10,49			
36	0,84	0,92	10,50			
37	2,35	W	10,62	C^2		
38	2,26	0,94	10,71	94,57		
39	0,24		11,09	т.к. W>Wst, гипотеза норм. НЕ ОТКЛОНЯЕТСЯ		
40	2,97		12,40			

Рисунок 3.11 – Проверка на нормальность признака HbA1C по критерию Шапиро-Уилка в Excel

Во всех выборках расчетное значение критерия оказалось больше критического, следовательно, гипотеза, состоящая в том, что генеральная совокупность подчиняется нормальному закону распределения, принимается.

Таким образом, рассматриваемые в работе диагностические признаки (постпрандиальная гликемия, уровень гликированного гемоглобина и триглицеридов) для здоровых пациентов и пациентов с сахарным диабетом 1 и 2 типа подчиняются нормальному закону распределения. Данный вывод, получен благодаря статистическому анализу исследуемых выборок в среде Excel, а именно проверки закона распределения с помощью критерия согласия Пирсона и критерия Шапиро-Уилка (рисунок 3.8 – 3.11). 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

диагностических признаков обладает признак – постпрадиальная гликемия, величина стандартного отклонения при сахарном диабете 1 типа составляет 3,7 ммоль/л, при сахарном диабете 2 типа – 4,22 ммоль/л.

3.3 Расчёт корреляции значений признаков

Оценка степени взаимной корреляции исследуемых признаков. При корреляционной связи изменение результативного признака y не всецело зависит от факторного признака x , а лишь частично, так как возможно влияние прочих факторов ε :

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

Характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе.

Для определения тесноты корреляционной связи применяется коэффициент корреляции. Коэффициент корреляции изменяется от -1 до +1 и показывает тесноту и направление корреляционной связи. Наиболее удобной формулой для расчета коэффициента корреляции является (метод Пирсона):

$$r_{XY} = \frac{\text{cov}_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}, \quad (3.11)$$

где средние значения выборок

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t; \quad (3.12)$$

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t. \quad (3.13)$$

Коэффициент парной корреляции показывает силу и направление влияния факторного показателя на результативный. Чем ближе к 1, тем связь сильнее, чем ближе к 0, тем связь слабее. Если коэффициент корреляции имеет положительное значение, то связь прямая, если отрицательное – обратная.

В среде Excel существует готовая функция КОРРЕЛ для вычисления данного показателя. Расчеты для пар признаков Нб и ППГ, Нб и ТГ, ТГ и ППГ для здоровых пациентов приведены на рисунках 3.12– 3.13:

А	
1	Коэффициент корреляции признаков у здоровых людей
2	Нб и ППГ
3	=КОРРЕЛ('Нб здоровые'!\$A\$2:\$A\$31;'ППГ здоровые'!\$A\$2:\$A\$31)

В	
1	
2	Нб и ТГ
3	=КОРРЕЛ('Нб здоровые'!\$A\$2:\$A\$31;'ТГ здоровые'!\$A\$2:\$A\$31)

С	
1	
2	ТГ и ППГ
3	=КОРРЕЛ('ТГ здоровые'!\$A\$2:\$A\$31;'ППГ здоровые'!\$A\$2:\$A\$31)

Рисунок 3.12 – Формулы в Excel для расчета коэффициента парной корреляции

	А	В	С
1	Коэффициент корреляции признаков у здоровых людей		
2	Нб и ППГ	Нб и ТГ	ТГ и ППГ
3	0,075	0,004	-0,054
4			

Рисунок 3.13 – Результаты расчета коэффициента корреляции признаков у здоровых пациентов

Для пациентов с сахарным диабетом 1 и 2 типа вычисления проводятся аналогичным способом. В результате получены коэффициенты парной корреляции, представленные в таблице 3.5.

Таблица 3.5 – Коэффициенты парной корреляции признаков

Коэффициенты парной корреляции признаков у здоровых пациентов		
Нб и ППГ	Нб и ТГ	ТГ и ППГ
0,075	0,004	-0,054
Коэффициенты парной корреляции признаков у пациентов с сахарным диабетом 1 типа		
Нб и ППГ	Нб и ТГ	ТГ и ППГ
0,428	0,211	0,034
Коэффициенты парной корреляции признаков у пациентов с сахарным диабетом 2 типа		
Нб и ППГ	Нб и ТГ	ТГ и ППГ
0,409	-0,05	-0,248

Таким образом, между рассматриваемыми диагностическими признаками наблюдается слабая теснота связи (коэффициент парной корреляции находится в пределах от 0,1 до 0,3), за исключением пары признаков Нб и ППГ у пациентов с сахарным диабетом. Между Нб и ППГ отмечается умеренная теснота связи (коэффициент парной корреляции находится в пределах от 0,3 до 0,5). Следовательно, связь между диагностическими признаками незначительная и её допустимо не учитывать при формировании диагностической модели.

3.4 Вычисление вероятностей интервалов значений признаков

Построение диагностической модели возможно после разделения значений количественных признаков на интервалы по формулам (3.1) – (3.4). Результаты разбиения для признака гликированный гемоглобин представлены на рисунке 3.14.

	A	B	C	D	E	F
1	НbA1C, %	Минимальное значение	Xd	Xu	Интервалы НbA1C	Обозначение НbA1C
2	9,45	=МИН(A:A)	=B\$8	=C3-0,01	НbA1C 3,49-4,74	u11
3	13,45	Максимальное значение	=C2+B\$6	=C4-0,01	НbA1C 4,75-6,00	u12
4	8,67	=МАКС(A:A)	=C3+B\$6	=C5-0,01	НbA1C 6,01-7,26	u13
5	9,06	λ	=C4+B\$6	=C6-0,01	НbA1C 7,27-8,52	u14
6	6,8	= $(B4-B2)/(1+3,32*\text{LOG10}(\text{СЧЁТ}(A:A)))$	=C5+B\$6	=C7-0,01	НbA1C 8,53-9,78	u15
7	10,5	Xn	=C6+B\$6	=C8-0,01	НbA1C 9,79-11,04	u16
8	7,91	=B2-0,5*B6	=C7+B\$6	=C9-0,01	НbA1C 11,05-12,30	u17
9	10,49		=C8+B\$6	=C10-0,01	НbA1C 12,31-13,56	u18
10	10,71		=C9+B\$6			

Рисунок 3.14 – Интервалы значений признака гликированный гемоглобин, %

Каждому значению количественного признака должен соответствовать определенный интервал. Определение принадлежности конкретного значения к определенному интервалу осуществляется с помощью логической функции ЕСЛИ. Формула расчета выглядит следующим образом:

=ЕСЛИ(И(С2>='Интервалы'!\$C\$2;С2<='Интервалы'!\$D\$2);"u11";ЕСЛИ(И(С2>='Интервалы'!\$C\$3;С2<='Интервалы'!\$D\$3);"u12";ЕСЛИ(И(С2>='Интервалы'!\$C\$4;С2<='Интервалы'!\$D\$4);"u13";ЕСЛИ(И(С2>='Интервалы'!\$C\$5;С2<='Интервалы'!\$D\$5);"u14";ЕСЛИ(И(С2>='Интервалы'!\$C\$6;С2<='Интервалы'!\$D\$6);"u15";ЕСЛИ(И(С2>='Интервалы'!\$C\$7;С2<='Интервалы'!\$D\$7);"u16";ЕСЛИ(И(С2>='Интервалы'!\$C\$8;С2<='Интервалы'!\$D\$8);"u17";ЕСЛИ(И(С2>='Интервалы'!\$C\$9;С2<='Интервалы'!\$D\$9)).

Результаты определения принадлежности значения признака к определенному интервалу приведены на рисунке 3.15.

	A	B	C	D	E	F	G	H	I	J
1	Код пациента	Название болезни	HbA1C,%	значения	ППГ, ммоль/л	значение	ТГ, ммоль/л	значение	Ац	Ац, обозначение
2	16343	Сахарный диабет 1 типа	9,45	u15	10,70	u23	0,66	u32	0	u41
3	16496	Сахарный диабет 1 типа	13,45	u18	13,40	u24	1,65	u32	0	u41
4	16530	Сахарный диабет 1 типа	8,67	u15	10,50	u23	2,94	u34	1	u42
5	16652	Сахарный диабет 1 типа	9,06	u15	10,30	u23	1,45	u32	0	u41
6	16783	Сахарный диабет 1 типа	6,80	u13	13,80	u25	1,40	u32	0	u41
7	16823	Сахарный диабет 1 типа	10,50	u16	18,30	u26	0,82	u32	0	u41
8	17313	Сахарный диабет 1 типа	7,91	u14	9,30	u23	0,98	u32	0	u41
9	17404	Сахарный диабет 1 типа	10,49	u16	14,10	u25	0,81	u32	0	u41
10	17418	Сахарный диабет 1 типа	10,71	u16	20,90	u27	1,43	u32	0	u41
11	17563	Сахарный диабет 1 типа	8,62	u15	16,40	u26	1,06	u32	0	u41
12	17669	Сахарный диабет 1 типа	10,62	u16	12,90	u24	2,07	u33	0	u41
13	17703	Сахарный диабет 1 типа	13,16	u18	13,90	u25	0,56	u31	0	u41
14	17794	Сахарный диабет 1 типа	7,09	u13	13,30	u24	0,99	u32	0	u41
15	17882	Сахарный диабет 1 типа	8,21	u14	16,10	u25	2,19	u33	0	u41
16	17896	Сахарный диабет 1 типа	12,40	u18	24,10	u28	2,77	u33	1	u42
17	18079	Сахарный диабет 1 типа	7,47	u14	11,70	u24	1,49	u32	0	u41
18	18429	Сахарный диабет 1 типа	8,20	u14	12,50	u24	0,88	u32	0	u41
19	18659	Сахарный диабет 1 типа	8,70	u15	8,90	u23	2,11	u33	0	u41
20	18973	Сахарный диабет 1 типа	6,40	u13	11,00	u24	1,21	u32	0	u41
21	19114	Сахарный диабет 1 типа	7,02	u13	9,40	u23	1,20	u32	0	u41
22	19231	Сахарный диабет 1 типа	7,66	u14	10,20	u23	0,92	u32	0	u41
23	19529	Сахарный диабет 1 типа	11,09	u17	8,50	u23	2,60	u33	0	u41
24	19612	Сахарный диабет 1 типа	10,11	u16	17,10	u26	1,62	u32	0	u41
25	19653	Сахарный диабет 1 типа	9,88	u16	15,70	u25	0,50	u31	0	u41
26	19686	Сахарный диабет 1 типа	7,47	u14	13,40	u24	0,32	u31	1	u42
27	20093	Сахарный диабет 1 типа	9,40	u15	15,30	u25	0,34	u31	0	u41

Рисунок 3.15 – Принадлежность значений признака к интервалам

Последний этап статистической обработки представляет собой экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.16).

Код пациен	Название болезни	Оценка при	Обозначен
16136	Сахарный диабет 2 типа	6,11	u13
16136	Сахарный диабет 2 типа	6,50	u22
16136	Сахарный диабет 2 типа	3,72	u34
16136	Сахарный диабет 2 типа	0,00	u41
16227	Сахарный диабет 2 типа	6,58	u13
16227	Сахарный диабет 2 типа	6,50	u22
16227	Сахарный диабет 2 типа	2,31	u33
16227	Сахарный диабет 2 типа	0,00	u41
16338	Сахарный диабет 2 типа	8,28	u14
16338	Сахарный диабет 2 типа	8,90	u23
16338	Сахарный диабет 2 типа	2,66	u33
16338	Сахарный диабет 2 типа	0,00	u41
16343	Сахарный диабет 1 типа	9,45	u15

Рисунок 3.16 - Экспортированные данные статистической обработки из MS Excel в базу данных

Таким образом, диапазон значений вероятностных диагностических признаков сахарного диабета разделен на интервалы. Диапазон значений признаков гликированный гемоглобин, уровень постпрандиальной гликемии и уровень триглицеридов разделен на 8 интервалов, признак ацетонурия разбит на 2 интервала, общее количество интервалов для формирования вероятностной диагностической модели составляет 26. В среде MS Excel с помощью логической функции ЕСЛИ осуществлен подсчет числа больных с диагнозом E_i , у которых значения признака попадают в определенный интервал, а также выполнена проверка несовместности событий попадания значений признака в разные интервалы. В результате получены необходимые статистические данные результатов обследований и диагнозов пациентов для формирования вероятностной диагностической модели.

3.5 Основные результаты и выводы

1 Исходные данные для формирования выборок значений диагностических признаков пациентов с сахарным диабетом 1 и 2 типа содержатся в таблице «Результаты лабораторного обследования» (рисунок 2.33), экспортируемой из MS Access в MS Excel. Модельные выборки здоровых пациентов сгенерированы в среде Excel в соответствии с параметрами нормального распределения, полученными из литературных источников. Методика обработки включает в себя экспорт данных из MS Access в MS Excel, группировку, 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 Рассматриваемые в работе диагностические признаки (постпращаальная гликемия, уровень гликированного гемоглобина и триглицеридов) для здоровых пациентов и пациентов с сахарным диабетом 1 и 2 типа подчиняются нормальному закону распределения. Данный вывод, получен благодаря статистическому анализу исследуемых выборок в среде Excel, а именно проверки распределения вероятностей с помощью критерия согласия Пирсона и критерия Шапиро-Уилка (рисунок 3.8 – 3.11). 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 диагностических признаков обладает признак – постпращаальная гликемия, величина стандартного отклонения при сахарном диабете 1 типа составляет 3,7 ммоль/л, при сахарном диабете 2 типа – 4,22 ммоль/л.

3 Расчет коэффициента парной корреляции позволил установить слабую и умеренную тесноту связи между рассматриваемыми диагностическими признаками (пары признаков: HbA1C и ППГ, HbA1C и ТГ, ТГ и ППГ) для выборок здоровых пациентов и пациентов с сахарным диабетом 1 и 2 типа. Для данных признаков коэффициент парной корреляции лежит в пределах от 0 до 0,4, что свидетельствует о незначительной корреляционной связи. Таким образом, слабая и умеренная теснота корреляционной связи между диагностическими признаками позволяет принять гипотезу о независимости значений диагностических признаков при разработке диагностической модели.

4 Диапазон значений вероятностных диагностических признаков сахарного диабета разделен на интервалы. Диапазон значений признаков гликированный гемоглобин, уровень постпрандиальной гликемии и уровень триглицеридов разделен на 8 интервалов, признак ацетонурия разбит на 2 интервала, общее количество интервалов для формирования вероятностной диагностической модели составляет 26. В среде MS Excel с помощью логической функции ЕСЛИ осуществлен подсчет числа больных с диагнозом E_i , у которых значения признака попадают в определенный интервал, а также выполнена проверка несовместности событий попадания значений признака в разные интервалы.

5 Последний этап статистической обработки представляет собой экспорт данных статистической обработки из MS Excel в MS Access (рисунок 3.16).

4 МЕТОДИКА РАСПОЗНАВАНИЯ САХАРНОГО ДИАБЕТА

4.1 Методы распознавания болезней

В процессе развития теории распознавания появляется большое количество методов, позволяющих решать поставленные задачи, так что классификация этих методов является условной и неоднозначной. Различные авторы (Ю.Л. Барабаш, В.И. Васильев [38], А.Л. Горелик, В.А. Скрипкин [39], Р. Дуда, П. Харт [40] и др.) дают различную типологию методов распознавания образов. Одни авторы различают параметрические, непараметрические и эвристические методы, другие – выделяют группы методов, исходя из исторически сложившихся школ и направлений в данной области. Например, в работе [40], в которой дан критический обзор методов распознавания, используется следующая типология методов распознавания образов:

- 1) вероятностные методы;
- 2) методы, основанные на предположениях о классе решающих функций;
- 3) логические методы;
- 4) лингвистические (структурные) методы;

Вероятностные методы распознавания заимствованы из классической теории статистических решений, в которой объекты исследования рассматриваются как реализации многомерной случайной величины, распределенной в пространстве признаков по какому-либо закону. Они базируются на байесовской схеме принятия решений, апеллирующей к априорным вероятностям принадлежности объектов к тому или иному распознаваемому классу и условным плотностям распределения значений признаков. Данные методы сводятся к определению отношения правдоподобия в многомерном пространстве признаков.

Группа методов, основанных на оценке плотностей распределения значений признаков, имеет прямое отношение к методам дискриминантного анализа. Байесовский подход к принятию решений относится к наиболее разработанным в современной статистике так называемым параметрическим методам, для которых считается известным аналитическое выражение закона распределения (в данном случае нормальный закон) и требуется оценить лишь небольшое количество параметров (векторы средних значений и ковариационные матрицы).

Основными трудностями применения вероятностных методов считаются необходимость запоминания всей обучающей выборки для вычисления оценок локальных плотностей распределения вероятностей и высокая чувствительность к непредставительности обучающей выборки.

В методах, основанных на предположениях о классе решающих функций считается известным общий вид решающей функции и задан функционал ее качества. На основании этого функционала по обучающей последовательности находят наилучшее приближение решающей функции. Самыми распространенными являются представления решающих функций в виде линейных и обобщенных нелинейных полиномов. Функционал качества решающего правила обычно связывают с ошибкой классификации.

Основным достоинством методов, основанных на предположениях о классе решающих функций, является ясность математической постановки задачи распознавания, как задачи поиска экстремума. Многообразие методов этой группы объясняется широким спектром используемых функционалов качества решающего правила и алгоритмов поиска экстремума. Однако достижению практических целей в данном случае не сопутствует извлечение новых знаний о природе распознаваемых объектов. Возможность извлечения этих знаний, в частности знаний о механизмах взаимодействия атрибутов (признаков), здесь принципиально ограничена заданной

структурой такого взаимодействия, зафиксированной в выбранной форме решающих функций.

Логические методы распознавания образов базируются на аппарате алгебры логики и позволяют оперировать информацией, заключенной не только в отдельных признаках, но и в сочетаниях значений признаков. В этих методах значения какого-либо признака рассматриваются как элементарные события.

В самом общем виде логические методы можно охарактеризовать как разновидность поиска по обучающей выборке логических закономерностей и формирование некоторой системы логических решающих правил, каждое из которых имеет собственный вес. Группа логических методов разнообразна и включает методы различной сложности и глубины анализа. Для дихотомических (булевых) признаков популярными являются так называемые древообразные классификаторы, метод тупиковых тестов, алгоритм «Кора» и др. Логические методы распознавания образов, является достаточно трудоемким в вычислительном отношении, поскольку при отборе конъюнкций необходим полный перебор. Поэтому при применении логических методов предъявляются высокие требования к эффективной организации вычислительного процесса, и эти методы хорошо работают при сравнительно небольших размерностях пространства признаков и только на мощных компьютерах.

Лингвистические методы распознавания образов основаны на использовании специальных грамматик, порождающих языки, с помощью которых может описываться совокупность свойств распознаваемых объектов.

Таким образом, каждый объект представляет собой совокупность непроеизводных элементов, «соединенных» между собой теми или иными способами или, другими словами, «предложением» некоторого «языка».

Путем синтаксического анализа (грамматического разбора) «предложения» определяется его синтаксическая «правильность» или, что эквивалентно, может ли некоторая фиксированная грамматика, описывающая класс, породить имеющееся описание объекта.

Однако задача восстановления (определения) грамматик по некоторому множеству высказываний (предложений - описаний объектов), порождающих данный язык, является трудно формализуемой [40].

Каждый из существующих методов классификации не является универсальным, то есть может быть применим только для решения определенного класса задач. Кроме того, каждому методу присущи свои достоинства и недостатки (таблица 4.1).

Таблица 4.1 – Области и ограничения применения методов распознавания состояния здоровья [40]

Классификационные группировки методов	Область применения	Ограничения (недостатки)
Вероятностные методы	Задачи с известным распределением, как правило, нормальным, необходимость набора большой статистики	Необходимость перебора всей обучающей выборки при распознавании, высокая чувствительность к непредставительности обучающей выборки и артефактам
Методы, основанные на предположениях о классе решающих функций	Классы должны быть хорошо разделяемыми, система признаков – ортонормированной	Должен быть заранее известен вид решающей функции. Невозможность учета новых знаний о корреляциях между признаками
Логические методы	Задачи небольшой размерности пространства признаков	При отборе логических решающих правил (конъюнкций) необходим полный перебор. Высокая вычислительная трудоемкость
Лингвистические (структурные) методы	Задачи небольшой размерности пространства признаков	Задача восстановления (определения) грамматики по некоторому множеству высказываний (описаний объектов), является трудно формализуемой. Нерешенность теоретических проблем

Одним из перспективных направлений решения задач дифференциальной диагностики является использование вероятностных методов (таблица 4.2). Это объясняется тем, что для многих заболеваний принципиально могут совпадать симптомокомплексы или диапазоны возможных значений отдельных диагностических признаков, различаясь только частотами, с которыми при тех или иных заболеваниях встречаются отдельные признаки, их сочетания или количественные значения.

Таблица 4.2 – Области и ограничения применения вероятностных методов распознавания состояния здоровья

Метод	Область применения	Ограничения
Метод Байеса	Применение критерия Байеса целесообразно в случае, когда система распознавания многократно осуществляет распознавание неизвестных объектов или явлений в условиях неизменного признакового пространства, стабильном описании классов.	Недостатками метода Байеса являются большой объем предварительной информации, «угнетение» редко встречающихся классов и другие. Метод Байеса целесообразно использовать, когда имеются необходимые статистические данные.
Метод последовательного анализа (метод Вальда)	Применяется для дифференциальной диагностики (распознавания двух состояний). Число обследований заранее не устанавливается, их проводится столько, сколько необходимо для принятия решения с определенной степенью риска. В методе последовательного анализа рассматриваемые отношения вероятностей признаков (отношения правдоподобия) составляются не сразу, а в последовательном порядке; поэтому, как правило, требуется меньшее число обследований [41].	Ограничением метода является необходимость дихотомического рассмотрения при решении многоальтернативных задач.

Среди вероятностных методов распознавания метод Байеса, занимает особое место благодаря простоте и эффективности. Разумеется, метод Байеса

имеет недостатки: большой объем предварительной информации, «угнетение» редко встречающихся диагнозов и др. Однако в случаях, когда объем статистических данных позволяет применить метод Байеса, его целесообразно использовать как один из наиболее надежных и эффективных методов [41].

Методикой распознавания состояния здоровья, основанной на методе Байеса, предусматривается сбор статистических данных о заболеваемости и признаках сахарного диабета, проектирование базы данных результатов обследований и диагнозов пациентов, статистическая обработка данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.1)

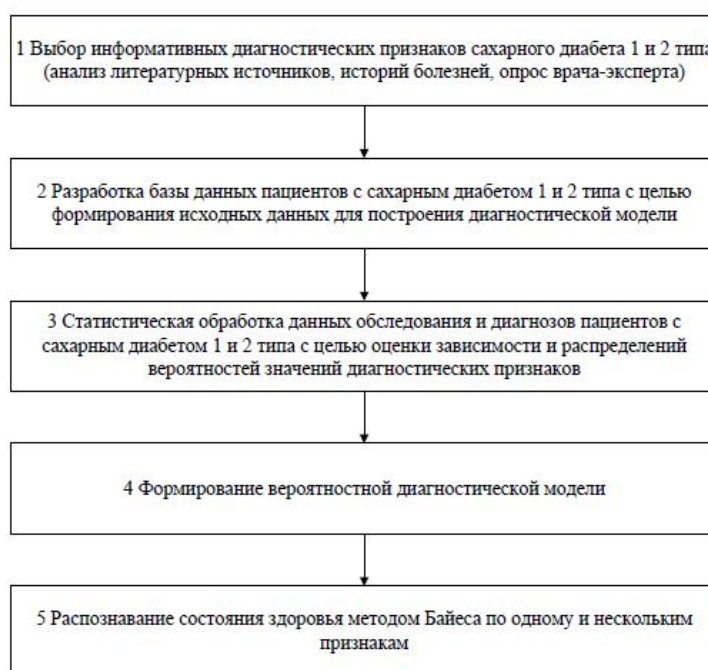


Рисунок 4.1 – Методика распознавания сахарного диабета

Таким образом, одним из перспективных направлений решения задач распознавания состояния здоровья является использование вероятностных методов. Основное преимущество вероятностных методов распознавания

состоит в возможности одновременного учета признаков различной физической природы, так как они характеризуются безразмерными величинами – вероятностями их появления при различных состояниях системы. Среди методов диагностики метод, основанный на обобщенной формуле Байеса, занимает особое место благодаря простоте и эффективности. Разумеется, метод Байеса имеет недостатки: большой объем предварительной информации, «угнетение» редко встречающихся диагнозов и др. Однако в случаях, когда объем статистических данных позволяет применить метод Байеса, его целесообразно использовать как один из наиболее надежных и эффективных методов [41].

Методикой распознавания состояния здоровья, основанной на методе Байеса, предусматривается сбор статистических данных о заболеваемости и признаках сахарного диабета, проектирование базы данных результатов обследований и диагнозов пациентов, статистическая обработка данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.1)

4.2 Диагностические модели сахарного диабета

Задача распознавания состояния здоровья по вероятностным диагностическим признакам решается на основе вероятностных диагностических моделей статистическими методами.

Классы состояния здоровья пациентов могут толковаться как 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. Область (диапазон) значений вероятностного диагностического признака разделяется на интервалы. События попадания значения диагностического признака 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.

Оценки вероятностей классов состояния здоровья и вероятностей принятия признаками значений из выбранных интервалов вычисляются по статистическим данным результатов обследований и диагнозов пациентов. По данным историй болезней пациентов определяется плотность распределения вероятностей значений диагностического признака. 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 [37].

Таким образом, вероятностная диагностическая модель представляет собой таблицу, которая содержит результаты вычислений статистических оценок вероятностей диагнозов и условных вероятностей интервалов значений признаков. Табличная форма модели позволяет без затруднений реализовать математический аппарат распознавания методом Байеса на языке MatLab.

4.3 Методика формирования диагностической модели

Формирование вероятностной диагностической модели включает в себя: сбор статистических данных результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа из историй болезней, накопление и обработку собранных данных в базе данных (раздел 2), оценку зависимости признаков и определение распределений вероятностей значений диагностических признаков в решателе MS Excel (раздел 3), импортирование статистических данных из решателя MS Excel в MatLab, вычисление условных вероятностей интервалов значений признаков и вероятностей диагнозов (рисунок 4.2).

Этап вычисления условных вероятностей интервалов значений признаков и вероятностей диагнозов осуществляется в решателе MatLab

(таблица 4.3), этапы построения диагностической модели предшествующие данному описаны в предыдущих разделах.

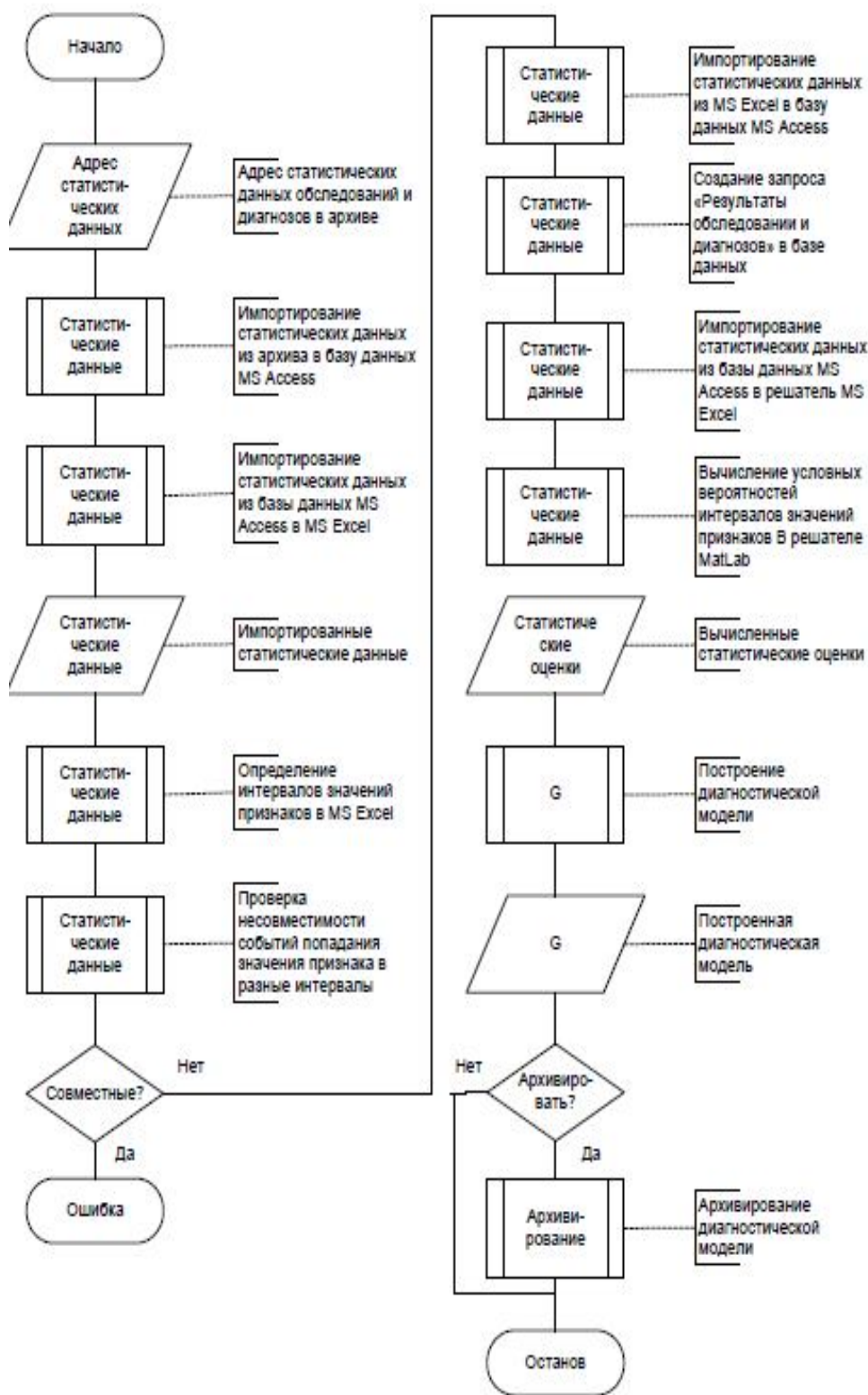


Рисунок 4.2 – Схема методики построения диагностической модели

Таблица 4.3 – Текст программного комплекса импортирования данных и формирования вероятностной диагностической модели

Операторы	Комментарии
Кнопка «Загрузка»	
function pushbutton1_Callback(hObject, eventdata, handles)	% обработка события Callback кнопки «Загрузка»
global S	% объявляет переменные глобальными
[FName, PName] = uigetfile('*.*)')	% диалоговое окно открытия файла с содержимым текущего каталога (рисунок 4.3)
if ~ isequal(FName, 0)	% проверка был ли выбран файл
FullName = strcat(PName, FName)	% если файл выбран, происходит считывание данных из файла
S=importdata(FullName) end	% импортирование загруженных данных в матрицу S
S(isnan(S))=0 %	% замена значений "NaN" на "0"
set(handles.uitable1, 'Visible','on'); set(handles.uitable1, 'Data',S);	% вывод матрицы в таблицу (рисунок 4.4)
Кнопка "Оценка условных вероятностей интервалов значений признаков, $P(u_{ij}/E_i)$ " (рисунок 4.5)	
function pushbutton2_Callback(hObject, eventdata, handles)	% обработка события Callback кнопки "Оценка условных вероятностей интервалов значений признаков, $P(u_{ij}/E_i)$ "
global S n k	% объявляет переменные глобальными
n=size(S,1)	% количество строк матрицы S
k=size(S,2);	% количество столбцов матрицы S
for i =1:n	%начало цикла от 1 до конца вектора n
for j =2:k	%начало цикла от 1 до конца вектора k
S(i,j)=S(i,j)/S(i,1)	%расчет условных вероятностей интервалов значений признаков $p(u_{ij}/E_i)$
end	%окончание цикла
set(handles.uitable1, 'Visible','on'); set(handles.uitable1, 'Data',S);	%вывод матрицы матрицы S в таблицу
Кнопка «Оценка вероятности диагноза $p(E_i)$ » (рисунок 4.6)	
function pushbutton3_Callback(hObject, eventdata, handles)	% обработка события Callback кнопки «Оценка вероятности диагноза $p(E_i)$ »
global S n k	% объявляет переменные глобальными
d=sum(S(:,1))	%расчет суммы пациентов, знаменателя для формулы $p(E_i)=M_i/M$
for i =1:n;	%начало цикла
S(i,1) = S(i,1)/d	%расчет оценки вероятности диагноза, $p(E_i)=M_i/M$
end	%окончание цикла
set(handles.uitable1, 'Visible','on'); set(handles.uitable1, 'Data',S);	%вывод матрицы матрицы S в таблицу

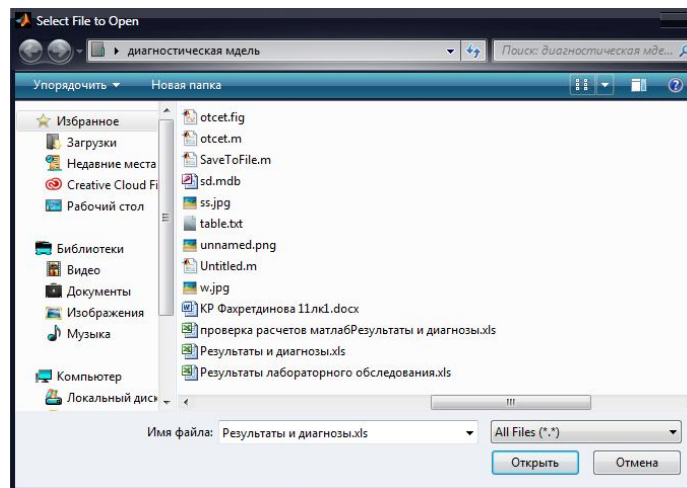


Рисунок 4.3 – Диалоговое окно открытия файла

	u11	u12	u13	u14	u15	u16	u17
Обследование (здоров)	30	18	13	0	0	0	0
Сахарный диабет 1 типа	27	0	0	4	7	6	6
Сахарный диабет 2 типа	28	0	1	8	9	5	4

Рисунок 4.4 – Таблица, импортированных данных

	u11	u12	u13	u14	u15	u16	u17
Обследование (здоров)	30	0.6000	0.4333	0	0	0	0
Сахарный диабет 1 типа	27	0	0	0.1481	0.2593	0.2222	0.2222
Сахарный диабет 2 типа	28	0	0.0357	0.2857	0.3214	0.1786	0.1429

Рисунок 4.5 – Результаты вычисления условных вероятностей интервалов значений признаков, $P(u_{ij}/E_i)$

	u11	u12	u13	u14	u15	u16	u17
Обследование (здоров)	0.3529	0.6000	0.4333	0	0	0	0
Сахарный диабет 1 типа	0.3176	0	0	0.1481	0.2593	0.2222	0.2222
Сахарный диабет 2 типа	0.3294	0	0.0357	0.2857	0.3214	0.1786	0.1429

Рисунок 4.6 – Вероятностная диагностическая модель

Таким образом, формирование вероятностной диагностической модели включает в себя: сбор статистических данных результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа из историй болезней, накопление и обработку собранных данных в базе данных (раздел 2), оценку зависимости признаков и определение распределений вероятностей значений диагностических признаков в решателе MS Excel (раздел 3), импортирование статистических данных из решателя MS Excel в MatLab, вычисление

условных вероятностей интервалов значений признаков и вероятностей диагнозов (рисунок 4.3).

В решателе MatLab осуществлена реализация программного комплекса построения вероятностной диагностической модели. Программный комплекс осуществляет расчет статистических оценок вероятностей диагнозов, вероятностей и условных вероятностей интервалов значений признаков. По результатам вычислений статистических оценок вероятностей диагнозов и условных вероятностей значений признаков формируется вероятностная диагностическая модель состояния здоровья в табличной форме (рисунок 4.6).

На основе сформированной диагностической модели осуществляется распознавание состояния здоровья методом Байеса по одному и нескольким диагностическим признакам.

4.4 Методика распознавания сахарного диабета методом Байеса

4.4.1 Методика распознавания состояния здоровья методом Байеса по одному признаку

Соответствие состояния здоровья классу E_i и принятие диагностическим признаком значения из интервала u_{jl} отображаются вероятностной диагностической моделью (рисунок 4.6) как зависимые события E_i и u_{jl} с известными вероятностями. При обследовании пациента обнаруживается значение диагностического признака из интервала u_{jl} . Для постановки диагноза требуется вычислить вероятности классов состояния здоровья после обследования пациента, т.е. условные вероятности классов состояния здоровья $p(E_i/u_{jl})$. Решение задачи методом Байеса основывается на предположении, что классами состояния здоровья образуется полная группа несовместных событий.

Искомые вероятности классов состояния здоровья вычисляются по формуле Байеса

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

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

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

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

Выводы из формулы Байеса:

1) сумма апостериорных вероятностей классов состояния здоровья равна единице. Действительно, суммирование апостериорных вероятностей (4.8) приводит к следующему результату:

$$\sum_{i=1}^N p(E_i / u_{jl}) = 1, \quad (4.9)$$

поскольку после суммирования в правой части формулы (4.8) числитель равен знаменателю;

2) если состояние здоровья E_k проявляется детерминирующим или патогномичным признаком, то после обнаружения этого признака при обследовании пациента апостериорная вероятность состояния здоровья E_k равняется единице, а апостериорные вероятности остальных классов состояния здоровья равны нулю.

Решающее правило, в соответствии с которым принимается решение о диагнозе, основывается на сопоставлении апостериорных вероятностей классов состояния здоровья, вычисляемых по формуле Байеса. Диагноз относится к классу состояния здоровья с максимальным значением

апостериорной вероятности. Обычно решающее правило дополняется пороговым значением апостериорной вероятности класса состояния здоровья:

$$\sum_{i=1}^N p(E_i / u_{jl}) \geq P, \quad (4.10)$$

где P – выбранное пороговое значение апостериорной вероятности для принятия решения о диагнозе E_i . Если условие (4.10) не выполняется, то решение о диагнозе не принимается (отказ от распознавания). Необходимо дополнительное обследование пациента. Методика распознавания состояния здоровья методом Байеса по одному диагностическому признаку представлена на рисунке 4.7.

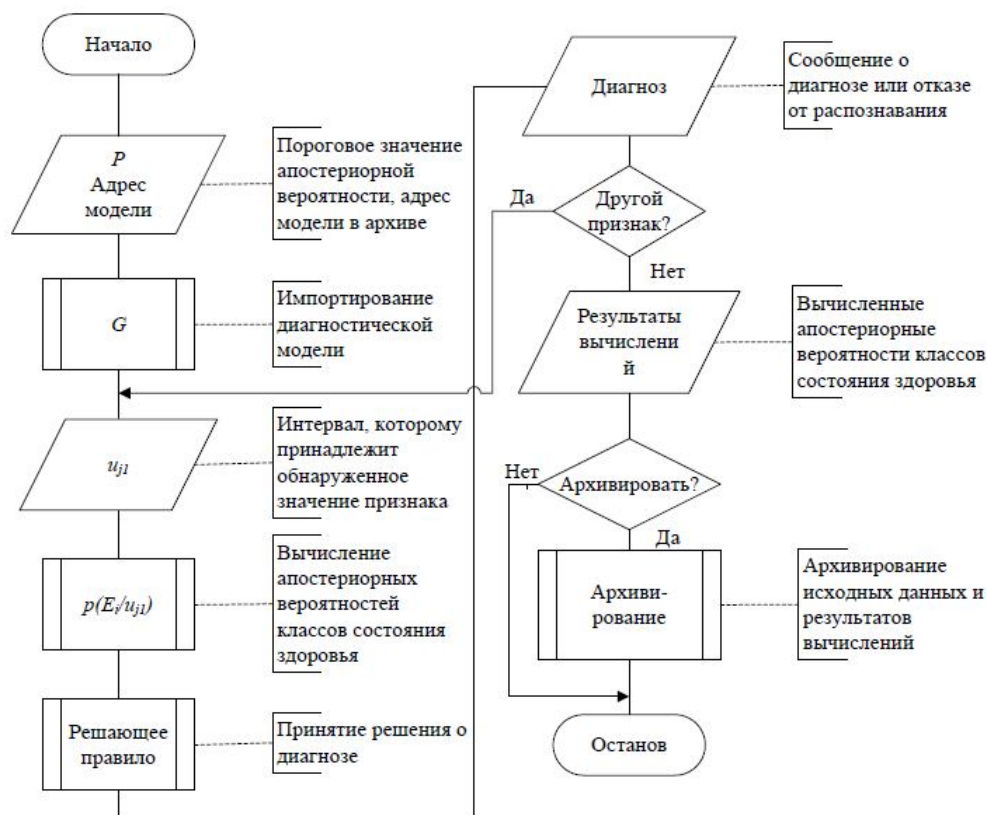


Рисунок 4.7 – Схема методики распознавания состояния здоровья по значению одного признака

4.4.2 Методика распознавания состояния здоровья методом Байеса по нескольким признакам

Комплексным обследованием пациента выявляются значения нескольких диагностических признаков. Интервалы, к которым относятся обнаруженные значения признаков, составляют множество

$$U_{JL} = \{u_{jl}\} \subset U_{KL}, j = 1, 2, 3, \dots, J; l \in \{1, 2, 3, \dots, L_j\}, \quad (4.11)$$

где U_{JL} – множество интервалов значений признаков;

u_{jl} – интервал значений признака;

U_{KL} – множество интервалов значений признаков в вероятностной диагностической модели;

J – число обнаруженных признаков;

L_j – число интервалов значений признака.

Вероятности классов состояния здоровья после обнаружения при обследовании пациента значений нескольких диагностических признаков вычисляются по формуле Байеса в обобщенной форме

$$p(E_i / U_{JL}) = \frac{p(E_i) p(U_{JL} / E_i)}{\sum_{i=1}^N p(E_i) p(U_{JL} / E_i)}, \quad (4.12)$$

где $p(E_i / U_{JL})$ – условная вероятность состояния здоровья E_i (после обнаружения значений признаков из интервалов U_{JL} ;

$p(E_i)$ – априорная вероятность состояния здоровья E_i ;

$p(U_{JL} / E_i)$ – условная вероятность обнаружения значений признаков из интервалов U_{JL} в состоянии здоровья E_i ;

$p(U_{JL})$ – вероятность принятия диагностическими признаками значений из интервалов U_{JL} .

В соответствии с решающим правилом метода Байеса диагноз состояния здоровья со значениями признаков из интервалов U_{JL} относится к классу состояния здоровья с наибольшей апостериорной вероятностью. Решающее правило дополняется введением порогового значения для апостериорной вероятности:

$$\sum_{i=1}^N p(E_i / U_{JL}) \geq P, \quad (4.13)$$

где P – выбранное пороговое значение апостериорной вероятности для принятия решения о диагнозе E_i . Если условие (4.13) не выполняется, то решение о диагнозе откладывается до получения результатов дополнительного обследования пациента [37].

Методика распознавания состояния здоровья методом Байеса по сочетанию значений диагностических признаков представлена на рисунке 4.8.

Таким образом, методика распознавания сахарного диабета методом Байеса по одному признаку состоит из этапов: импортирования диагностической модели, определения интервала, которому принадлежит обнаруженное значение признака, вычисления апостериорных вероятностей классов состояния здоровья по формуле Байеса, принятия решения о диагнозе (рисунок 4.7). Однако, для принятия достоверного диагноза одного диагностического признака недостаточно, требуется комплексное обследования пациента. В связи с этим помимо распознавания по одному признаку в работе применяется методика распознавания по сочетанию признаков (рисунок 4.8).

Взаимодействие программного комплекса с пользователем осуществляется посредством управляющей программы с использованием графического интерфейса GUI MatLab в интерактивном режиме.

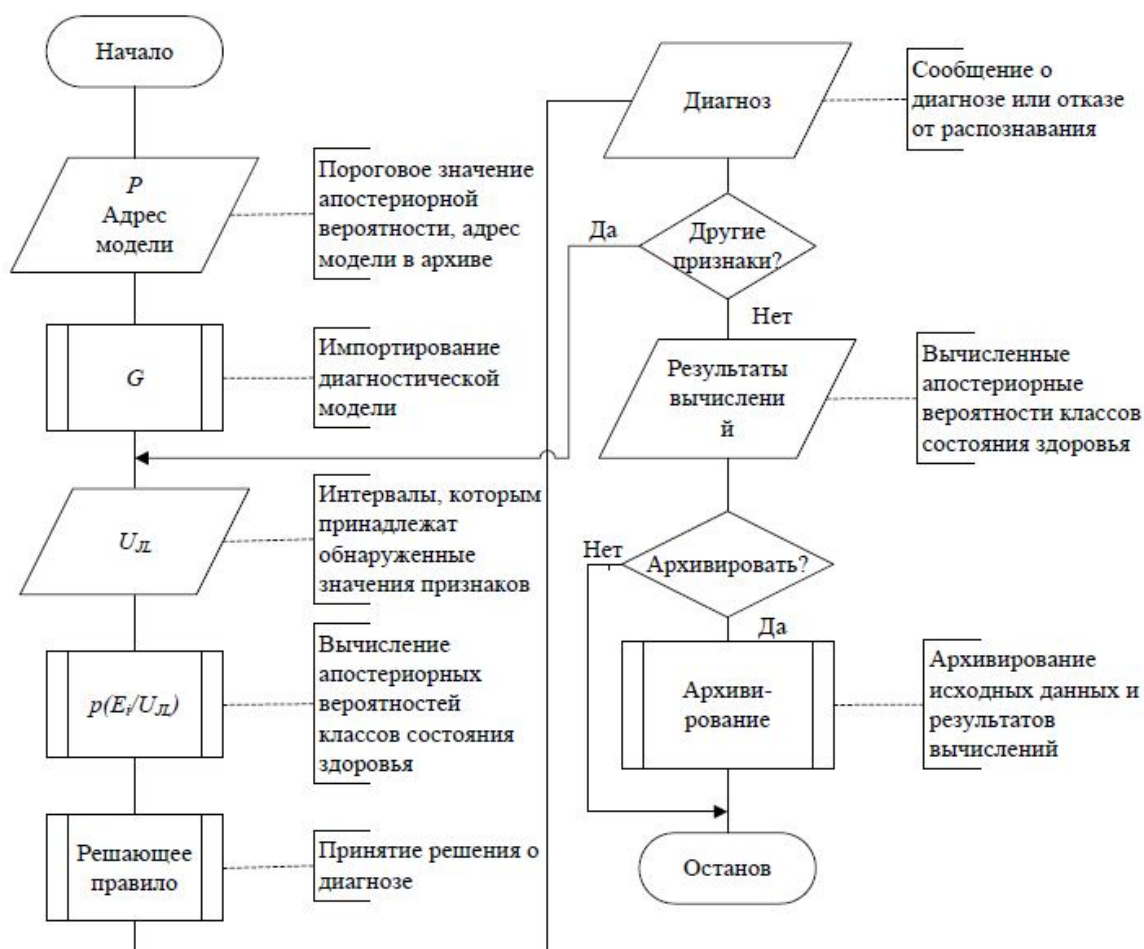


Рисунок 4.8 – Схема методики распознавания состояния здоровья по сочетанию значений диагностических признаков

4.5 Интерфейс пользователя

Максимально удобная работа пользователя с программным комплексом распознавания сахарного диабета 1 и 2 типа возможна благодаря «дружественному» пользовательскому интерфейсу. Применение интерфейса позволяет освоить навыки работы с программой даже начинающему пользователю персонального компьютера. Работа с программой не требует от пользователя знаний в области математики и теории распознавания.

Графический интерфейс пользователя создан с помощью специализированной среды GUIDE MatLab.

Как и любой процесс проектирования, процесс построения графического интерфейса пользователя можно разбить на этапы (рисунок 4.9).

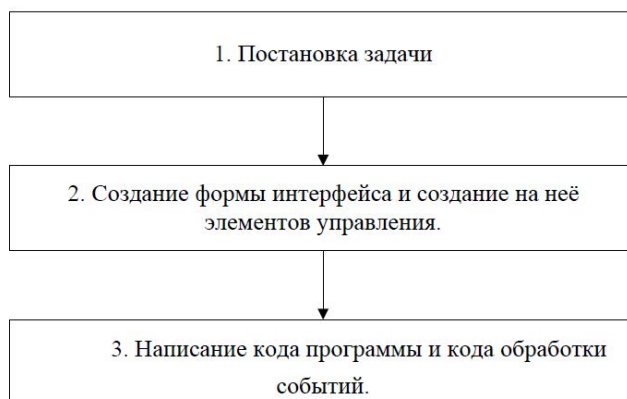


Рисунок 4.9 – Этапы построения графического интерфейса пользователя

На первом этапе проводится анализ поставленной задачи и определяется количество и состав элементов управления необходимых для решения задачи. Пользовательский интерфейс программного комплекса распознавания сахарного диабета включает следующие элементы управления.

I. Кнопки:

- 1) «Загрузка»;
- 2) «Оценка условных вероятностей интервалов значений признаков, $P(u_j|E_i)$ »;
- 3) «Оценка вероятности диагноза $p(E_i)$ »;
- 4) «Метод Байеса по признаку HbA1C, %»;
- 5) «Метод Байеса по признаку ТГ, ммоль/л»;
- 6) «Метод Байеса по признаку ППГ, ммоль/л»;
- 7) «Метод Байеса по признаку Ац»;
- 8) «Метод Байеса по нескольким признакам»;
- 9) «Сохранение»;

8) «Справка».

II. Таблицы для вывода диагностической модели, значений рассчитанных апостериорных вероятностей.

III. Текстовые области для вывода результатов работы.

На втором этапе создаётся форма графического интерфейса, на которой создаются и размещаются элементы управления. Здесь же описываются их свойства (рисунок 4.10).

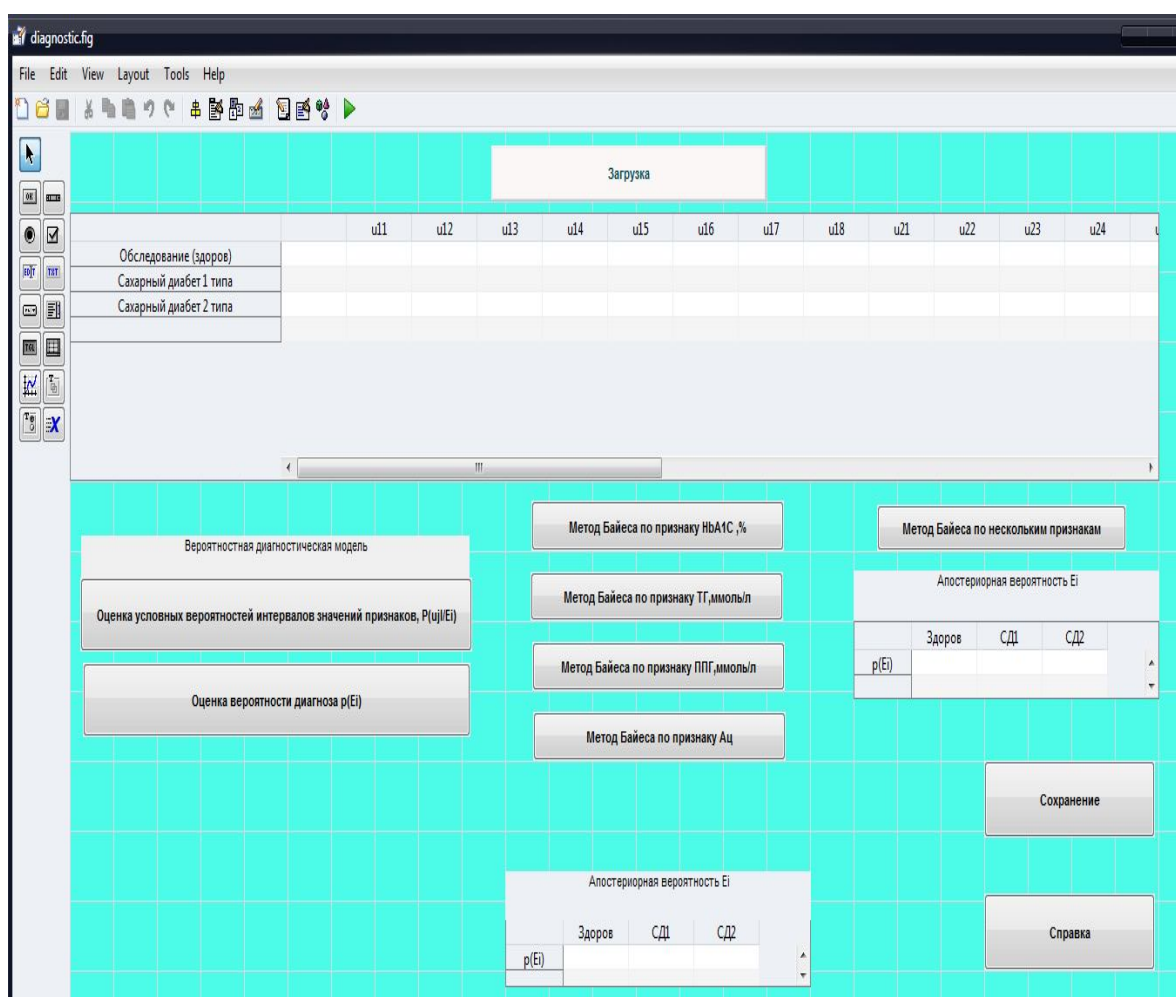


Рисунок 4.10 – Форма графического интерфейса

Задавать расположение и выравнивать элементы на форме, описывать их свойства можно вручную, но для удобства и быстроты применяется редактор выравнивания объектов (The Alignment Tool) и редактор свойств (The Property Editor).

Существует два способа создания формы и элементов управления, а также задания или изменения их свойств:

1) использование команды `WORKSPACE` (то есть использование команды операционной среды `MATLAB`).

2) использование средств панели инструментов – совокупности средств для быстрого создания GUI (`The Control Panel`).

При построении элементов управления первым способом удобно использовать скрипт-файл, в котором последовательно с помощью команд `WORKSPACE` описывается создание элементов управления и устанавливаются их свойства.

На практике чаще применяют второй способ создания графического интерфейса с элементами управления. Это объясняется тем, что при использовании панели управления с её редакторами свойств, событий, выравнивания очень удобно работать, и создавать GUI значительно быстрее, чем в первом случае.

На третьем этапе создания графического интерфейса пользователя (GUI) пишется код основной программы вычисления и код для обработки событий (ПРИЛОЖЕНИЕ Б).

Код основной программы вычисления, пишется на языке программирования операционной среды `Matlab`, в виде `m`-файла. Созданные `m`-файлы закрепляются за событием какого-нибудь элемента управления или формы.

Конечным результатом является программа с графическим интерфейсом пользователя (рисунок 4.11).

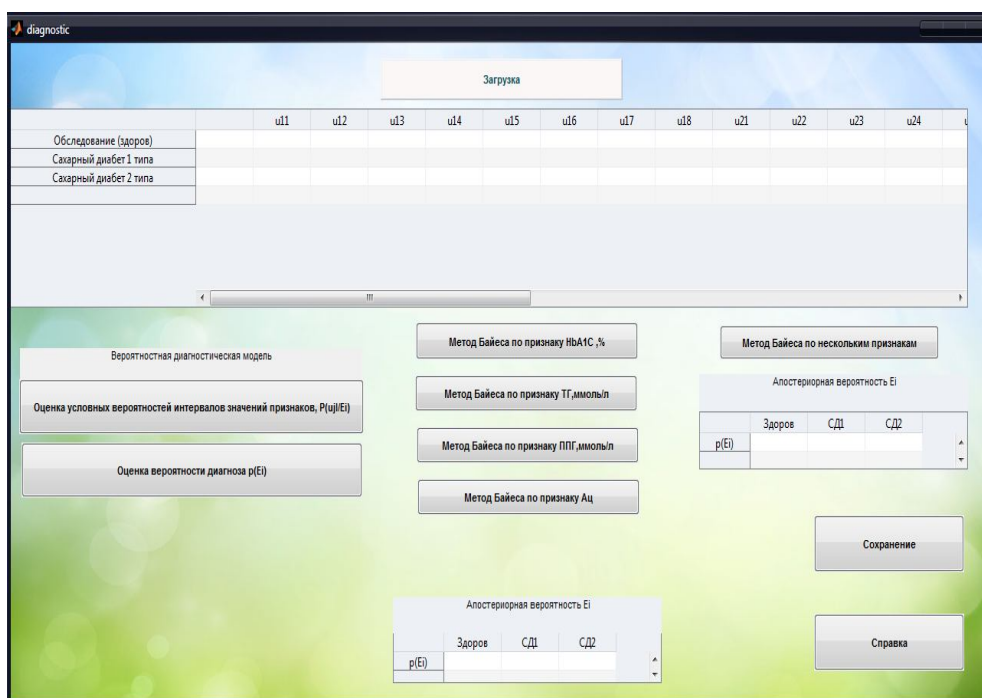
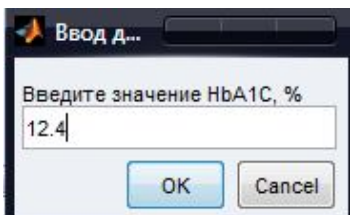
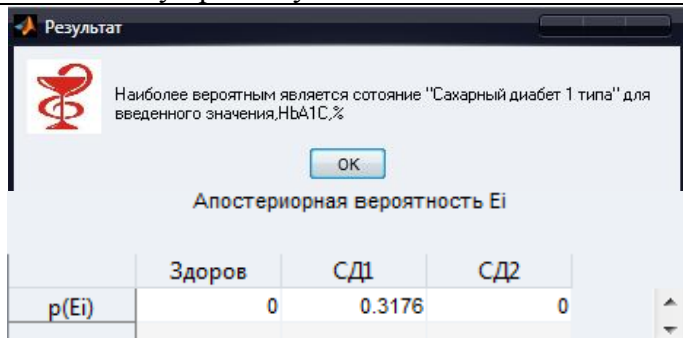
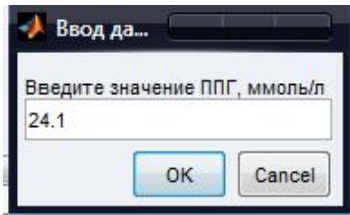
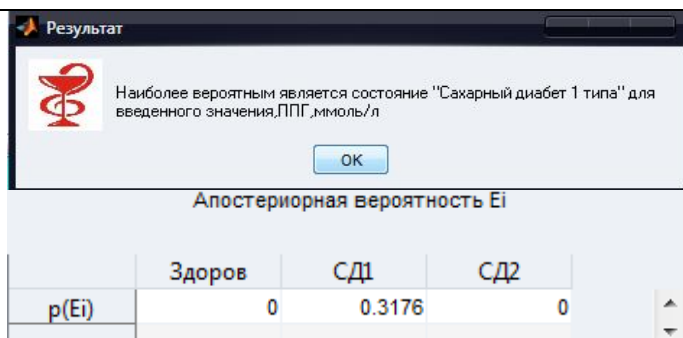
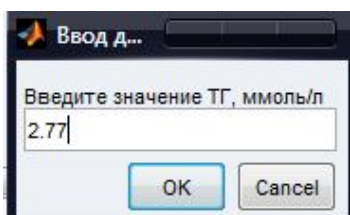
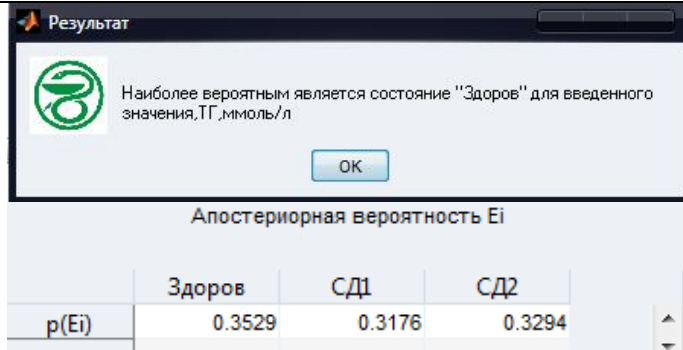
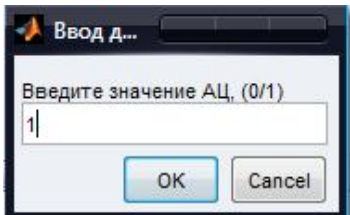
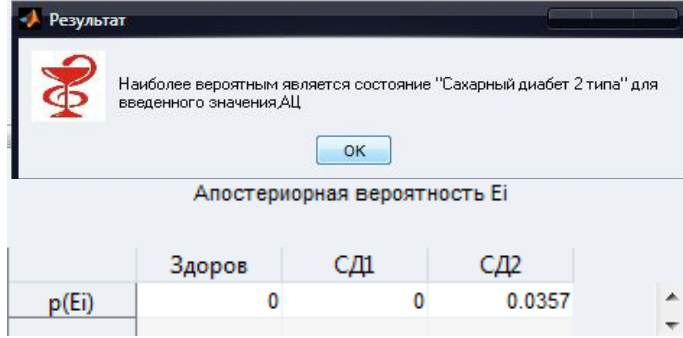


Рисунок 4.11 – Интерфейс программы распознавания сахарного диабета

Создание графического интерфейса является последним этапом разработки программы распознавания сахарного диабета.

Проверка работоспособности программы распознавания сахарного диабета проводится путем тестирования программы по значениям диагностических признаков пациентов с установленными диагнозами (таблица 4.3). Значения диагностических признаков, используемые для тестирования, содержатся в базе данных пациентов с сахарным диабетом в таблице «Результаты лабораторного обследования» (рисунок 2.19).

Таблица 4.3 – Проверка работоспособности программного комплекса распознавания сахарного диабета по одному признаку и сочетанию признаков

Окно ввода значения диагностического признака	Окно вывода результата о наиболее вероятном состоянии здоровья пациента
Проверка распознавания у пациента с сахарным диабетом 1 типа по одному диагностическому признаку	
	
	
	
	

Продолжение таблицы 4.3

Проверка распознавания у пациента с сахарным диабетом 1 типа по сочетанию диагностических признаков									
<div><div>Ввод да...</div><div>Введите значение HbA1C, % 12.4</div><div>Введите значение ППГ, ммоль/л 24.1</div><div>Введите значение ТГ, ммоль/л 2.77</div><div>Введите значение АЦ, 0/1 1</div><div>OKCancel</div></div>	<div><div>Результат</div><div>Наиболее вероятным является состояние "Сахарный диабет 1 типа"</div><div>OK</div></div>								
Проверка распознавания у пациента с сахарным диабетом 2 типа по одному диагностическому признаку									
<div><div>Ввод д...</div><div>Введите значение HbA1C, % 6.9</div><div>OKCancel</div></div>	<div><div>Результат</div><div>Наиболее вероятным является состояние "Сахарный диабет 2 типа" для введенного значения HbA1C, %</div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><th></th><th>Здоров</th><th>CD1</th><th>CD2</th></tr><tr><th>$p(E_i)$</th><td>0</td><td>0.3176</td><td>0.3294</td></tr></table></div>		Здоров	CD1	CD2	$p(E_i)$	0	0.3176	0.3294
	Здоров	CD1	CD2						
$p(E_i)$	0	0.3176	0.3294						
<div><div>Ввод да...</div><div>Введите значение ППГ, ммоль/л 16.7</div><div>OKCancel</div></div>	<div><div>Результат</div><div>Наиболее вероятным является состояние "Сахарный диабет 2 типа" для введенного значения ППГ, ммоль/л</div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><th></th><th>Здоров</th><th>CD1</th><th>CD2</th></tr><tr><th>$p(E_i)$</th><td>0</td><td>0.3176</td><td>0.3294</td></tr></table></div>		Здоров	CD1	CD2	$p(E_i)$	0	0.3176	0.3294
	Здоров	CD1	CD2						
$p(E_i)$	0	0.3176	0.3294						
<div><div>Ввод д...</div><div>Введите значение ТГ, ммоль/л 1.52</div><div>OKCancel</div></div>	<div><div>Результат</div><div>Наиболее вероятным является состояние "Здоров" для введенного значения ТГ, ммоль/л</div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><th></th><th>Здоров</th><th>CD1</th><th>CD2</th></tr><tr><th>$p(E_i)$</th><td>0.4333</td><td>0</td><td>0.0357</td></tr></table></div>		Здоров	CD1	CD2	$p(E_i)$	0.4333	0	0.0357
	Здоров	CD1	CD2						
$p(E_i)$	0.4333	0	0.0357						

Продолжение таблицы 4.3

<div><div>Ввод д...</div><div>Введите значение АЦ, (0/1)</div><div><input type="text" value="0"/></div><div><div>OK</div><div>Cancel</div></div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров" для введенного значения АЦ</div></div><div><div>OK</div></div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>p(E_i)</td><td>0.3529</td><td>0.3176</td><td>0.3294</td></tr></table></div>		Здоров	СД1	СД2	p(E _i)	0.3529	0.3176	0.3294
	Здоров	СД1	СД2						
p(E _i)	0.3529	0.3176	0.3294						
Проверка распознавания у пациента с сахарным диабетом 2 типа по сочетанию диагностических признаков									
<div><div>Ввод да...</div><div>Введите значение HbA1C, %</div><div><input type="text" value="6.9"/></div><div>Введите значение ППГ, ммоль/л</div><div><input type="text" value="16.7"/></div><div>Введите значение ТГ, ммоль/л</div><div><input type="text" value="1.52"/></div><div>Введите значение АЦ, 0/1</div><div><input type="text" value="0"/></div><div><div>OK</div><div>Cancel</div></div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Сахарный диабет 2 типа"</div></div><div><div>OK</div></div></div>								
Проверка распознавания у здорового пациента по одному диагностическому признаку									
<div><div>Ввод д...</div><div>Введите значение HbA1C, %</div><div><input type="text" value="4.18"/></div><div><div>OK</div><div>Cancel</div></div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров" для введенного значения HbA1C, %</div></div><div><div>OK</div></div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>p(E_i)</td><td>0.3529</td><td>0</td><td>0</td></tr></table></div>		Здоров	СД1	СД2	p(E _i)	0.3529	0	0
	Здоров	СД1	СД2						
p(E _i)	0.3529	0	0						
<div><div>Ввод да...</div><div>Введите значение ППГ, ммоль/л</div><div><input type="text" value="4.74"/></div><div><div>OK</div><div>Cancel</div></div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров" для введенного значения ППГ, ммоль/л</div></div><div><div>OK</div></div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>p(E_i)</td><td>0.3529</td><td>0</td><td>0.3294</td></tr></table></div>		Здоров	СД1	СД2	p(E _i)	0.3529	0	0.3294
	Здоров	СД1	СД2						
p(E _i)	0.3529	0	0.3294						

Продолжение таблицы 4.3

<div><div>Ввод д...</div><div>Введите значение ТГ, ммоль/л</div><div>1.24</div><div>OKCancel</div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров" для введенного значения ТГ, ммоль/л</div></div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>$p(E_i)$</td><td>0.4333</td><td>0</td><td>0.0357</td></tr></table></div>		Здоров	СД1	СД2	$p(E_i)$	0.4333	0	0.0357
	Здоров	СД1	СД2						
$p(E_i)$	0.4333	0	0.0357						
<div><div>Ввод д...</div><div>Введите значение АЦ, (0/1)</div><div>0</div><div>OKCancel</div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров" для введенного значения ППГ, ммоль/л</div></div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>$p(E_i)$</td><td>0.3529</td><td>0.3176</td><td>0.3294</td></tr></table></div>		Здоров	СД1	СД2	$p(E_i)$	0.3529	0.3176	0.3294
	Здоров	СД1	СД2						
$p(E_i)$	0.3529	0.3176	0.3294						
Проверка распознавания у здорового пациента по сочетанию диагностических признаков									
<div><div>Ввод да...</div><div>Введите значение HbA1C, %</div><div>4.18</div><div>Введите значение ППГ, ммоль/л</div><div>4.74</div><div>Введите значение ТГ, ммоль/л</div><div>1.24</div><div>Введите значение АЦ, 0/1</div><div>0</div><div>OKCancel</div></div>	<div><div>Результат</div><div><div></div><div>Наиболее вероятным является состояние "Здоров"</div></div><div>OK</div><div>Апостериорная вероятность E_i</div><table><tr><td></td><td>Здоров</td><td>СД1</td><td>СД2</td></tr><tr><td>$p(E_i)$</td><td>1</td><td>0</td><td>0</td></tr></table></div>		Здоров	СД1	СД2	$p(E_i)$	1	0	0
	Здоров	СД1	СД2						
$p(E_i)$	1	0	0						

Таким образом, в среде GUI MatLab разработан графический пользовательский интерфейс программы распознавания сахарного диабета. Применение пользовательского интерфейса позволяет освоить работу с программой даже начинающему пользователю персонального компьютера. Работа с программой не требует от пользователя знаний математического аппарата, заложенного в ней. Дружественный визуальный интерфейс (рисунок 4.11) делают работу пользователя с программой удобной и простой.

Вывод результатов в диалоговое окно программы позволяет пользователю получить не только предполагаемый диагноз пациента в виде сообщения с наиболее вероятным диагнозом (таблица 4.3), а также получить количественную оценку состояния здоровья пациента, поскольку в диалоговое окно выводятся апостериорные вероятности заболеваний.

Оценка эффективности разработанного программно-методического комплекса осуществляется посредством вычислительного эксперимента.

4.6 Вычислительный эксперимент по распознаванию сахарного диабета

Оценка эффективности предложенной методики распознавания проводилась по результатам тестирования программно-методического комплекса распознавания сахарного диабета по сочетанию диагностических признаков. Тестирование осуществлялась по результатам обследования пациентов с сахарным диабетом 1 и 2 типа и по модельным выборкам для здоровых пациентов (рисунок 4.12).

Код пациента	Название болезни	HbA1C, %	ППГ, ммоль	ТГ, ммоль/л	Ац
19529	Сахарный диабет 1 типа	11,09	8,50	2,60	0
19612	Сахарный диабет 1 типа	10,11	17,10	1,62	0
19653	Сахарный диабет 1 типа	9,88	15,70	0,50	0
19686	Сахарный диабет 1 типа	7,47	13,40	0,32	1
20093	Сахарный диабет 1 типа	9,40	15,30	0,34	0
20225	Сахарный диабет 1 типа	7,69	13,00	0,75	0
16136	Сахарный диабет 2 типа	6,11	6,50	3,72	0
16227	Сахарный диабет 2 типа	6,58	6,50	2,31	0
16338	Сахарный диабет 2 типа	8,28	8,90	2,66	0
16344	Сахарный диабет 2 типа	6,62	9,50	1,73	0
16430	Сахарный диабет 2 типа	7,50	10,00	4,33	0

Рисунок 4.12 – Результаты обследования пациентов (фрагмент таблицы из базы данных)

Результаты тестирования представлены в таблице 4.4.

Таблица 4.4 – Тестирование программно-методического комплекса по сочетанию диагностических признаков

Название болезни	Количество верно установленных диагнозов	Количество неверно установленных диагнозов
Сахарный диабет 1 типа	24	3
Сахарный диабет 2 типа	23	5
Обследование (здоров)	28	2
Итого:	77	8

Оценка эффективности предложенной методики проводилась путем вычисления прогностической значимости результатов по формулам (4.14) – (4.18) (таблица 4.5):

$$\text{Ч} = (\text{ИП} / (\text{ИП} + \text{ЛО})) * 100\%; \quad (4.14)$$

$$\text{С} = (\text{ИО} / (\text{ИО} + \text{ЛП})) * 100\%; \quad (4.15)$$

$$\text{ПЦПТ} = (\text{ИП} / (\text{ИП} + \text{ЛП})) * 100\%; \quad (4.16)$$

$$\text{ПЦОТ} = (\text{ИО} / (\text{ИО} + \text{ЛО})) * 100\%; \quad (4.17)$$

$$\text{ПТ} = ((\text{ИП} + \text{ИО}) / \text{N}) * 100\%, \quad (4.18)$$

где N – число наблюдений;

ИП – истинно положительные результаты;

ИО – истинно отрицательные результаты;

ЛП – ложноположительные результаты;

ЛО – ложноотрицательные результаты;

Ч – чувствительность;

С – специфичность;

ПЦПТ – предсказующая ценность положительного теста;

ПЦОТ – предсказующая ценность отрицательного теста;

ПТ – предсказующая точность.

Таблица 4.5 – Оценка эффективности программно-методического комплекса распознавания сахарного диабета

Показатель	Значение, %
Чувствительность	85,45
Специфичность	93,33
Предсказуемая ценность положительного теста	95,92
Предсказуемая ценность отрицательного теста	77,77
Предсказуемая точность	88,23

Таким образом, анализ данных таблицы позволяет утверждать, что предложенная методика распознавания сахарного диабета является эффективной. В связи с чем, может быть внедрена в качестве программы для обучения молодых специалистов в области эндокринологии, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам.

Наличие расхождений в диагнозах, объясняется тем, что для принятия диагноза недостаточно рассматриваемых диагностических признаков. Повышение достоверности распознавания возможно благодаря увеличению количества диагностических признаков в вероятностной диагностической модели и за счет обучения информационной системы (пополнения базы данных пациентов с сахарным диабетом новыми сведениями).

4.6 Основные результаты и выводы

1 Одним из перспективных направлений решения задач распознавания состояния здоровья является использование вероятностных методов. Основное преимущество вероятностных методов распознавания состоит в возможности одновременного учета признаков различной физической природы, так как они характеризуются безразмерными величинами – вероятностями их появления при различных состояниях системы. Среди методов диагностики метод, основанный на обобщенной формуле Байеса, занимает особое место благодаря простоте и эффективности. Разумеется, метод Байеса имеет недостатки: большой

объем предварительной информации, «угнетение» редко встречающихся диагнозов и др. Однако в случаях, когда объем статистических данных позволяет применить метод Байеса, его целесообразно использовать как один из наиболее надежных и эффективных методов [41].

2 Методикой распознавания состояния здоровья, основанной на методе Байеса, предусматривается сбор статистических данных о заболеваемости и признаках сахарного диабета, проектирование базы данных результатов обследований и диагнозов пациентов, статистическая обработка данных обследований и диагнозов, формирование диагностической модели, применение метода Байеса для распознавания состояния здоровья (рисунок 4.1)

3 Вероятностная диагностическая модель представляет собой таблицу, которая содержит результаты вычислений статистических оценок вероятностей диагнозов и условных вероятностей интервалов значений признаков. Табличная форма модели позволяет без затруднений реализовать математический аппарат распознавания методом Байеса на языке MatLab.

4 Формирование вероятностной диагностической модели включает в себя: сбор статистических данных результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа из историй болезней, накопление и обработку собранных данных в базе данных (раздел 2), оценку зависимости признаков и определение распределений вероятностей значений диагностических признаков в решателе MS Excel (раздел 3), импортирование статистических данных из решателя MS Excel в MatLab, вычисление условных вероятностей интервалов значений признаков и вероятностей диагнозов (рисунок 4.3). В решателе MatLab осуществлена реализация программного комплекса построения вероятностной диагностической модели. Программный комплекс осуществляет расчет статистических оценок вероятностей диагнозов, вероятностей и условных вероятностей интервалов значений признаков. По результатам вычислений статистических оценок

вероятностей диагнозов и условных вероятностей значений признаков формируется вероятностная диагностическая модель состояния здоровья в табличной форме (рисунок 4.6).

5 Методика распознавания сахарного диабета методом Байеса по одному признаку состоит из этапов: импортирования диагностической модели, определения интервала, которому принадлежит обнаруженное значение признака, вычисления апостериорных вероятностей классов состояния здоровья по формуле Байеса, принятия решения о диагнозе (рисунок 4.7). Однако, для принятия достоверного диагноза одного диагностического признака недостаточно, требуется комплексное обследования пациента. В связи с этим помимо распознавания по одному признаку в работе применяется методика распознавания по сочетанию признаков (рисунок 4.8).

6 В среде GUI MatLab разработан графический пользовательский интерфейс программы распознавания сахарного диабета. Применение пользовательского интерфейса позволяет освоить работу с программой даже начинающему пользователю персонального компьютера. Работа с программой не требует от пользователя знаний математического аппарата, заложенного в ней. Дружественный визуальный интерфейс (рисунок 4.11) делают работу пользователя с программой удобной и простой. Вывод результатов в диалоговое окно программы позволяет пользователю получить не только предполагаемый диагноз пациента в виде сообщения с наиболее вероятным диагнозом (таблица 4.3), а также получить количественную оценку состояния здоровья пациента, поскольку в диалоговое окно выводятся апостериорные вероятности заболеваний.

7 Оценка эффективности предложенной методики распознавания сахарного диабета показала высокую предсказующую точность программно-методического комплекса. В связи с чем, информационная система может

быть внедрена в учебный процесс для обучения молодых специалистов в области эндокринологии, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам.

Наличие расхождений в диагнозах, объясняется тем, что для принятия диагноза недостаточно рассматриваемых диагностических признаков. Повышение достоверности распознавания возможно благодаря увеличению количества диагностических признаков в вероятностной диагностической модели и за счет обучения информационной системы (пополнения базы данных пациентов с сахарным диабетом новыми сведениями).

ЗАКЛЮЧЕНИЕ

1 Сахарный диабет является социально-значимым заболеванием, которое, получило статус неинфекционной эпидемии XXI века. Заболеваемость сахарным диабетом в Российской Федерации имеет тенденцию к увеличению (рисунок 1.2), при этом в Пензенской области она превышает среднюю величину по РФ (рисунок 1.4). К тому же для Пензенского региона характерны высокие показатели смертности при СД, в рейтинге субъектов область занимает пятое место (рисунок 1.4). Распознавание сахарного диабета трудоемко, поскольку является коллективным интеллектуально сложным процессом принятия решений в условиях неопределённости и ограничений по времени, достоверность которого зависит от квалификации и опыта специалистов.

2 Одним из путей упрощения задач распознавания заболеваний является применение информационных систем. Однако результаты исследований по разработке ИС 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. Поэтому в дипломной работе представлена задача разработки методики проектирования базы данных пациентов, методики статистической обработки результатов обследования пациентов, методики формирования вероятностной диагностической модели, методики распознавания сахарного диабета методом Байеса. Перечисленные методики необходимы для проектирования информационной системы распознавания сахарного диабета, которая бы позволила обеспечить снижение трудоёмкости и повышение достоверности постановки диагноза, ограничение требований к квалификации специалистов. Методики, представленные в дипломной работе могут применяться для задач распознавания не только сахарного диабета, но и других заболеваний.

3 Проведённый анализ литературы и историй болезней пациентов за 2016 год на базе эндокринологического отделения ПОКБ им. Н.Н. Бурденко с установленным диагнозом сахарного диабета, позволил обосновать выбор четырех наиболее информативных признаков сахарного диабета первого и второго типа: постпрандиальной гликемии, уровня гликозилированного гемоглобина и триглицеридов, наличия ацетонурии. Данные признаки заложены в диагностическую модель.

4 В дипломной работе предлагается информационная система распознавания сахарного диабета, составными частями которой являются: база данных, решатель, пользовательский интерфейс (рисунок 2.5). База данных предназначена для хранения, накопления и обработки результатов обследований и диагнозов пациентов с сахарным диабетом 1 и 2 типа. Применение базы данных, содержащей результаты обследований и достоверные диагнозы пациентов, позволяет повысить достоверность диагноза и ограничить требования к опыту специалистов, за счёт принятия решений на основе опыта постановки диагноза высококвалифицированных специалистов, накопленного в базе данных. Благодаря базе данных ИС относится к системам с обучением.

5 Методика статистической обработки данных лабораторных анализов по малым выборкам, основанная на сочетании критерия согласия Пирсона и критерия Шапиро-Уилка подтвердила гипотезу о нормальном распределении количественных признаков сахарного диабета (рисунок 3.8. – 3.11). В результате получены оригинальные данные о распределениях значений диагностических признаков сахарного диабета. Выявлено наличие слабой тесноты связи (коэффициент парной корреляции находится в пределах от 0,1 до 0,3), за исключением пары признаков Hb и ППГ у пациентов с сахарным диабетом. Между Hb и ППГ обнаружена умеренная теснота связи (коэффициент парной корреляции находится в пределах от 0,3 до 0,5). Следовательно, связь между диагностическими признаками незначительная и может не учитываться при разработке диагностической модели (таблица 3.5).

6 Решатель MatLab представляет собой программный комплекс формирования диагностической модели и распознавания сахарного диабета методом Байеса по одному и сочетанию диагностических признаков.

7 Дружественный визуальный интерфейс ИС (рисунок 2.26 – 2.27, 4.11) делают работу пользователя с программой удобной и простой. Пользовательский интерфейс ИС позволяет освоить работу с программой даже начинающему пользователю персонального компьютера. Работа с программой не требует от пользователя знаний математического аппарата, заложенного в ней.

8 Оценка эффективности предложенной методики распознавания сахарного диабета показала высокую предсказующую точность (88,23%) программно-методического комплекса (таблица 4.5). В связи с чем, информационная система может быть внедрена в учебный процесс для обучения молодых специалистов в области эндокринологии, а также в лечебно-профилактические учреждения в качестве «помощника» начинающим специалистам. Наличие расхождений в диагнозах, объясняется тем, что для принятия диагноза недостаточно рассматриваемых диагностических признаков. Повышение достоверности распознавания возможно благодаря увеличению количества диагностических признаков в вероятностной диагностической модели и за счет обучения информационной системы (пополнения базы данных пациентов с сахарным диабетом новыми сведениями).

9 Результаты в полной мере соответствуют поставленным задачам дипломной работы. Перспективами применения данных методик, алгоритмов и программных комплексов является создание медицинских информационных систем, позволяющих осуществить дифференциальную диагностику болезней других органов и систем; дополнение существующих диагностических моделей другими диагностическими признаками.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Атлас диабета IDF /6–е издание, 2013 г. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.diabetesatlas.org/> (дата обращения: 07.02.2017);

2 Гаспарян С. А. Классификация медицинских информационных систем / Врач и информационные технологии. 2005. №3 [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://cyberleninka.ru/article/n/klassifikatsiya-meditsinskih-informatsionnyh-sistem-1> (дата обращения: 15.02.2017);

3 Систематизация информационных медицинских систем [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://refwy.ru/249010262.html>;

4 Козлов, А.Н. Интеллектуальные информационные системы: учебник /А.Н. Козлов. Мин-во с-х. РФ, ФГБОУ ВПО Пермская ГСХА. – Пермь : 2013. – 278 с. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://pgsha.ru:8008/books/study/%CA%E7%E2%20%C0.%CD.%20%C8%ED%F2%E5%EB%E5%EA%F2%F3%E0%EB%FC%ED%FB%E5%20%E8%ED%F4%EE%F0%EC%E0%F6%E8%EE%ED%ED%FB%E5%20%F1%E8%F1%F2%E5%EC%FB.pdf>;

5 Стандарт организации «Информационные системы в здравоохранении». Общие требования СТО МОСЗ 91500.16.0002–2004. – Москва, 2004;

6 CNewsPublichealth: Крупнейшие поставщики решений для здравоохранения 2015 [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://www.cnews.ru/reviews/publichealth2016/review_table/40902a632035a82afdflaa227e7c59613adf3bbc/;

7 Рейтинг: Крупнейшие поставщики ИТ в здравоохранении России [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.cnews.ru/reviews/free/publichealth2012/rating/rating2.shtml>;

8 Иванов В.Г. Лабораторные методы определения глюкозы: методические рекомендации для студентов и врачей-лаборантов / В.Г. Иванов, Т.О. Толстолуцкая. – Ижевск, 2002. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://sialotest80.narod.ru/glukoz1.htm>;

9 Хамнуева Л.Ю., Андреева Л.С., Шагун О.В. Диагностика, лечение и профилактика сахарного диабета. Современные представления и рекомендации: руководство для врачей. – Иркутск: ИГМУ, 2011. – 138 с. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://www.ismu.baikal.ru/src/downloads/1deb9b62_sd_ruk-vo_dlya_vracheiy.pdf;

10 Рошин Д.О. Потери от сахарного диабета и пути решения проблемы их оценки: диссертация на соискание ученой степени кандидата медицинских наук [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://www.mednet.ru/images/stories/files/replay/Roschin_text2.pdf;

11 Стандарт специализированной медицинской помощи при инсулинзависимом сахарном диабете от 8 декабря 2013 года;

12 Дедов И.И., Мельниченко Г.А., Фадеев В.Ф. Эндокринология. – М.: ГЭОТАР – Медиа, 2007. – 432 с.: ил.;

13 Калашникова М.Ф., Сунцов Ю.И., Белоусов Д.Ю., Кантемирова М.А. Анализ эпидемиологических показателей сахарного диабета 2 типа среди взрослого населения города Москвы. – Сахарный диабет. 2014;(3):5–16;

14 Титович Е.В., Кураева Т.Л., Иванова О.Н., Степанова С.М., Петеркова В.А., Дедов И.И. Прогнозирование сахарного диабета 1 типа в семьях больных (проспективное 16–летнее наблюдение). Акцент на будущее. Сахарный диабет. 2014; 3: 83–89. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://endojournals.ru>;

15 Беляева Ю.Б., Рахматуллов Ф.К., Вакина Т.Н. Сахарный диабет в практике терапевта: учебн. пособие. – Пенза: Изд-во ПГУ, 2010. – 131 с.;

16 Дедов И.И., Балаболкин М.И., Марова Е.И. Болезни органов эндокринной системы: руководство для врачей. – М.: Медицина, 2000. – 568 с.: ил.;

17 Диабет, общие сведения [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.pseudology.org/health/DiabetBook.pdf>;

18 Дедов И.И., Шестакова М.В., Викулова О.К. Государственный регистр сахарного диабета в Российской Федерации: статус 2014 г. и перспективы развития Сахарный диабет. 2015;18(3):5-23 [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://elibrary.ru/download/22477395.pdf>;

19. Литвицкий П.В. Патофизиология: учеб. пособие / П. Ф. Литвицкий. – 4-е изд., испр. и доп. – М. : ГЭОТАР – Медиа, 2009. – 496 с. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://vmede.org/sait/?page=8&id=Patofiziologija_litvickii_2009&menu=Patofiziologija_litvickii_2009;

20 Дробышев В. А., Карева Н. П., Туниковская О. В., Чернова Н.Н. Применение общесистемной магнитотерапии в комплексном лечении сахарного диабета 2 типа / Медицина и образование в Сибири. 2011. №6. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://cyberleninka.ru/article/n/primenenie-obschesistemnoy-magnitoterapii-v-kompleksnom-lechenii-saharnogo-diabeta-2-tipa> (дата обращения: 04.03.2017);

21 Фотина И.А. Сравнительный анализ биохимических показателей сыворотки крови и слюны у здоровых и больных сахарным диабетом 2 типа: автореферат диссертации на соискание ученой степени кандидата биологических наук [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.unn.ru/pages/disser/972.pdf>;

22 Поддубный Д. А., Ребров А. П. Кардиоваскулярный риск у больных анкилозирующим спондилитом: роль системного воспаления и дисфункции эндотелия / РФК. 2008. №5. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://cyberleninka.ru/article/n/kardiovaskulyarnyy-risk-u-bolnyh-ankiloziruyuschim-spondilitom-rol-sistemnogo-vospaleniya-i-disfunktsii-endoteliya> (дата обращения: 08.04.2017);

23 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]. – Электрон., текстовые дан. – Режим доступа: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=LAW&n=210955&fld=134&dst=100001,0&rnd=0.3568275418827398#0>;

24 Никонова Т.В. Современные аспекты патогенеза сахарного диабета 1 типа / Сахарный диабет, №3, 2006;

25 Жариков О. Г., Литвин А. А., Ковалёв В. А. Экспертные системы в медицине [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://mednovosti.by/journal.aspx?article=4013>. – Загл. с экрана;

26 Гусев А. В. Обзор рынка комплексных медицинских информационных систем // Врач и информационные технологии. – 2009. – № 6. – С. 4-17 [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://www.kmis.ru/site.nsf/pages/2009_obzor_kmis.htm .– Загл. с экрана;

27 DXplain – система поддержки принятия решения [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.mghlcs.org/projects/dxplain>;

28 Экспертные системы. Классификация экспертных систем. Разработка простейшей экспертной системы [Электронный ресурс].– Электрон., текстовые дан. – Режим доступа: <http://ref.by/refs/67/14877/1.html>;

29 EasyDiagnosis [Электронный ресурс] .– Электрон., текстовые дан. – Режим доступа: <http://www.easydiagnosis.com/cgi-bin/expert/diagnose.cgi>;

30 AImedica – интернет-услуга по компьютерной диагностике заболеваний [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://aimedica.ru/about/aimedica>;

31 Тараник М. А., Копаница Г. Д. Анализ задач и методов построения интеллектуальных медицинских систем // Врач и информационные технологии. 2014. №3. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: : <http://cyberleninka.ru/article/n/analiz-zadach-i-metodov-postroeniya-intellektualnyh-meditsinskih-sistem> (дата обращения: 30.03.2017);

32 Сиделев С. И. Математические методы в биологии и экологии: введение в элементарную биометрию: учебное пособие / С.И. Сиделев. – Ярославль : ЯрГУ, 2012. – 140 с. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.lib.uniyar.ac.ru/edocs/iuni/20110318.pdf>;

33 Лакин Г. Ф. Биометрия: учеб. пособие для биолог. спец. вузов / Г.Ф.Лакин – М. : Высш. шк., 2009. – 352 с.;

34 Лабораторная работа «Проверка гипотезы о нормальном распределении выборки по критерию Пирсона» [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://www.studfiles.ru/preview/4029502/>;

35 ГОСТ Р ИСО 54–79–2002 Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: <http://docs.cntd.ru/document/gost-r-iso-5479-2002>;

36 Гиниятуллин К.Г. Статистическая обработка результатов научных исследований : краткий конспект лекций / К.Г. Гиниятуллин. – Казань: 2014. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://dspace.kpfu.ru/xmlui/bitstream/handle/net/21919/02_10_kl-000830.pdf;

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

38 Васильев В.И. Распознающие системы: справочник. – К.: Наукова думка, 1983. – 230 с.;

39 Горелик А.Л., Скрипкин В.А. Методы распознавания. Изд.2. – М.: Высшая школа, 1984. – 219 с.;

40 Дуда Р., Харт П. Распознавание образов и анализ сцен: Пер.с англ. – М.: Мир, 1978. – 510 с;

39 Пасынкова Ирина Вячеславовна. Разработка моделей и алгоритмов дифференциальной диагностики патологии коленного сустава : Дис. ... канд. техн. наук : 05.13.01 : Воронеж, 2005 141 с.

41 Дмитриенко А.Г., Блинов А.В., Волков Д.В., Волков В.С. Техническая Диагностика. Оценка состояния и прогнозирование остаточного ресурса технически сложных объектов : учебн. пособие, Пенза, 2013. – 62 с. [Электронный ресурс]. – Электрон., текстовые дан. – Режим доступа: http://dep_pribor.pnzgu.ru/files/dep_pribor.pnzgu.ru/uchebnoe_posobie_iyul_2013.pdf.

ПРИЛОЖЕНИЕ А РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ВЫБОРОК

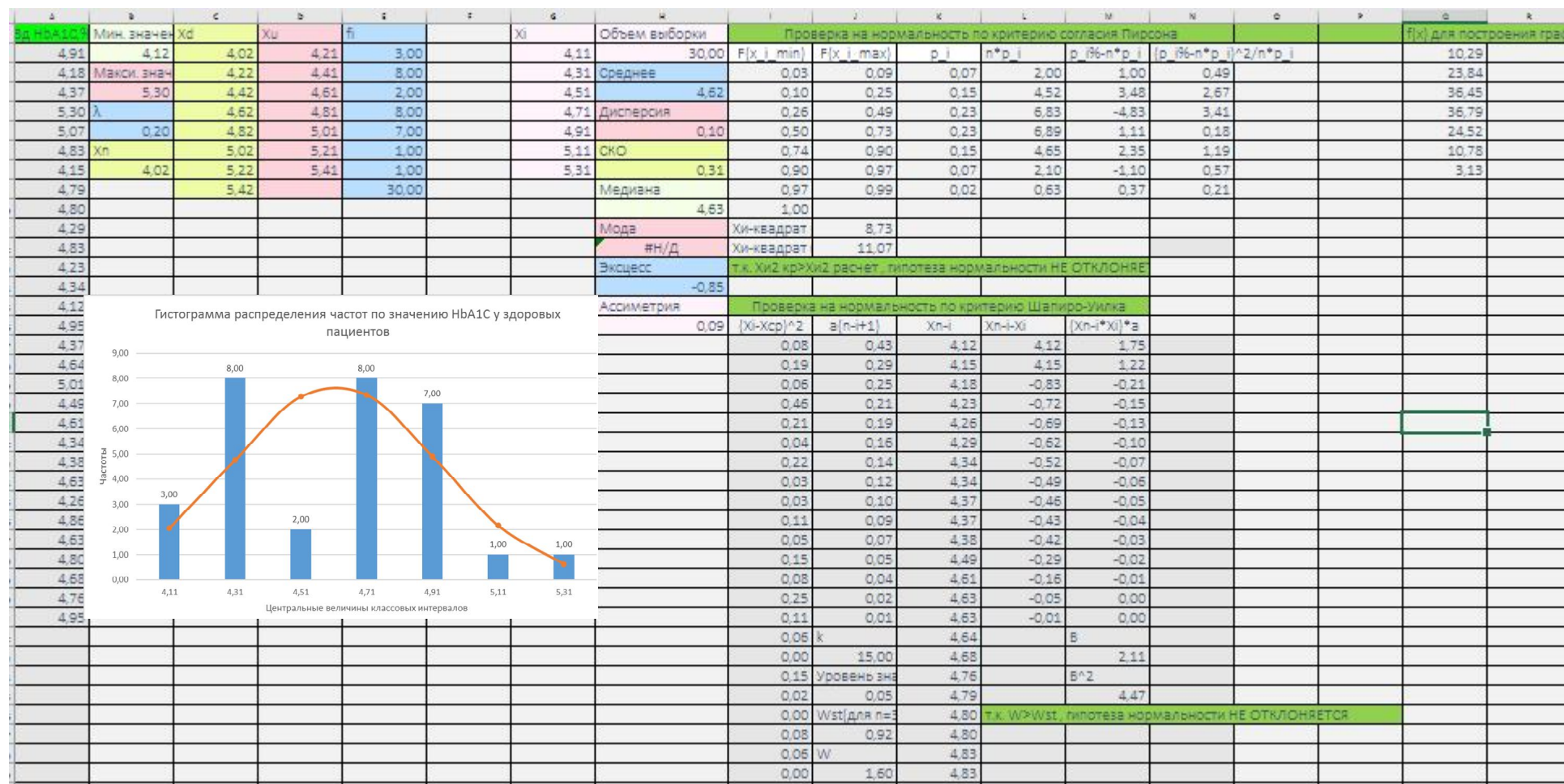


Рисунок А1 – Обработка модельной (генерированной) выборки по значению HbA1C у здоровых пациентов

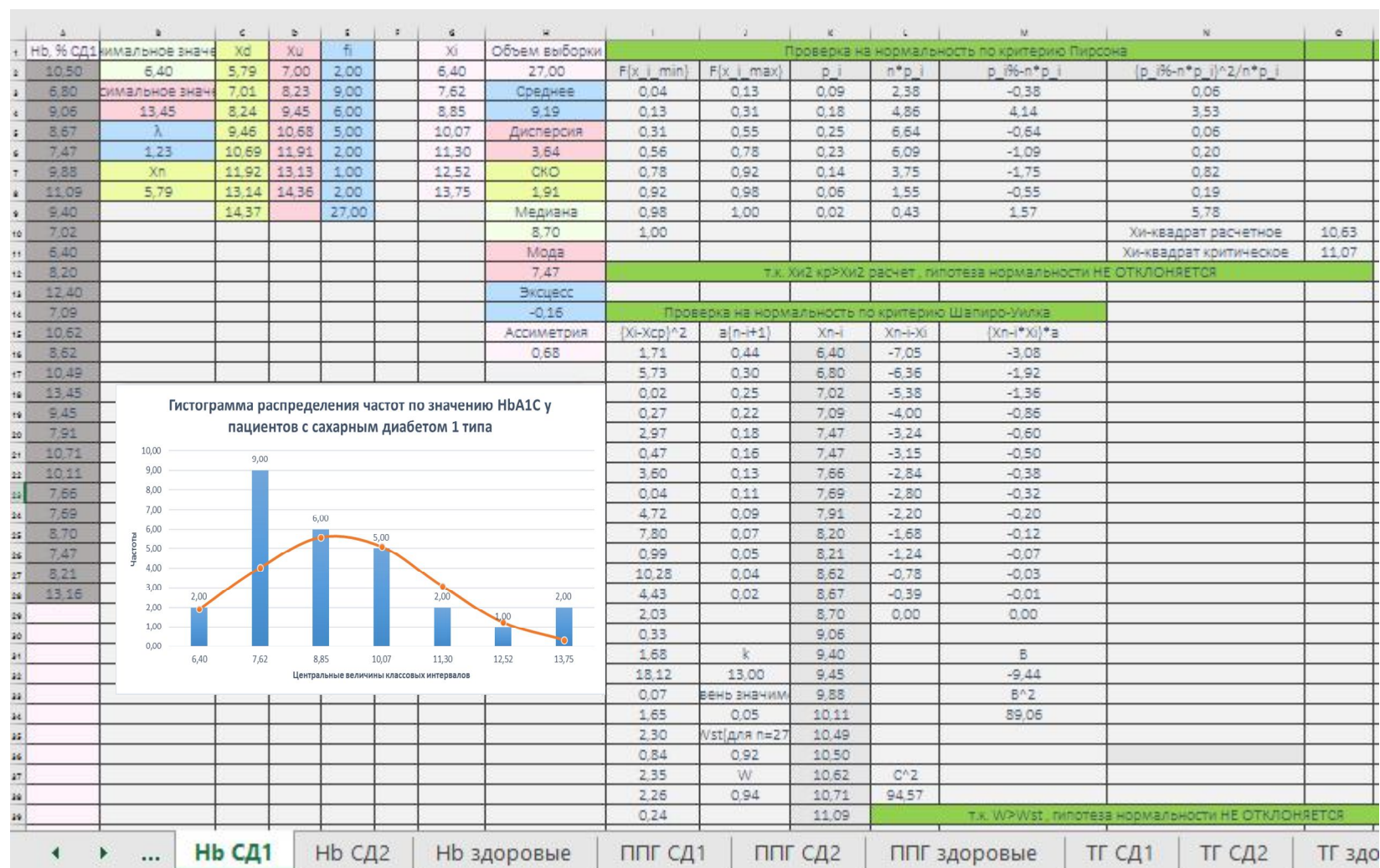


Рисунок А2 – Обработка выборки из историй болезней по значению HbA1C у пациентов с сахарным диабетом 1 типа

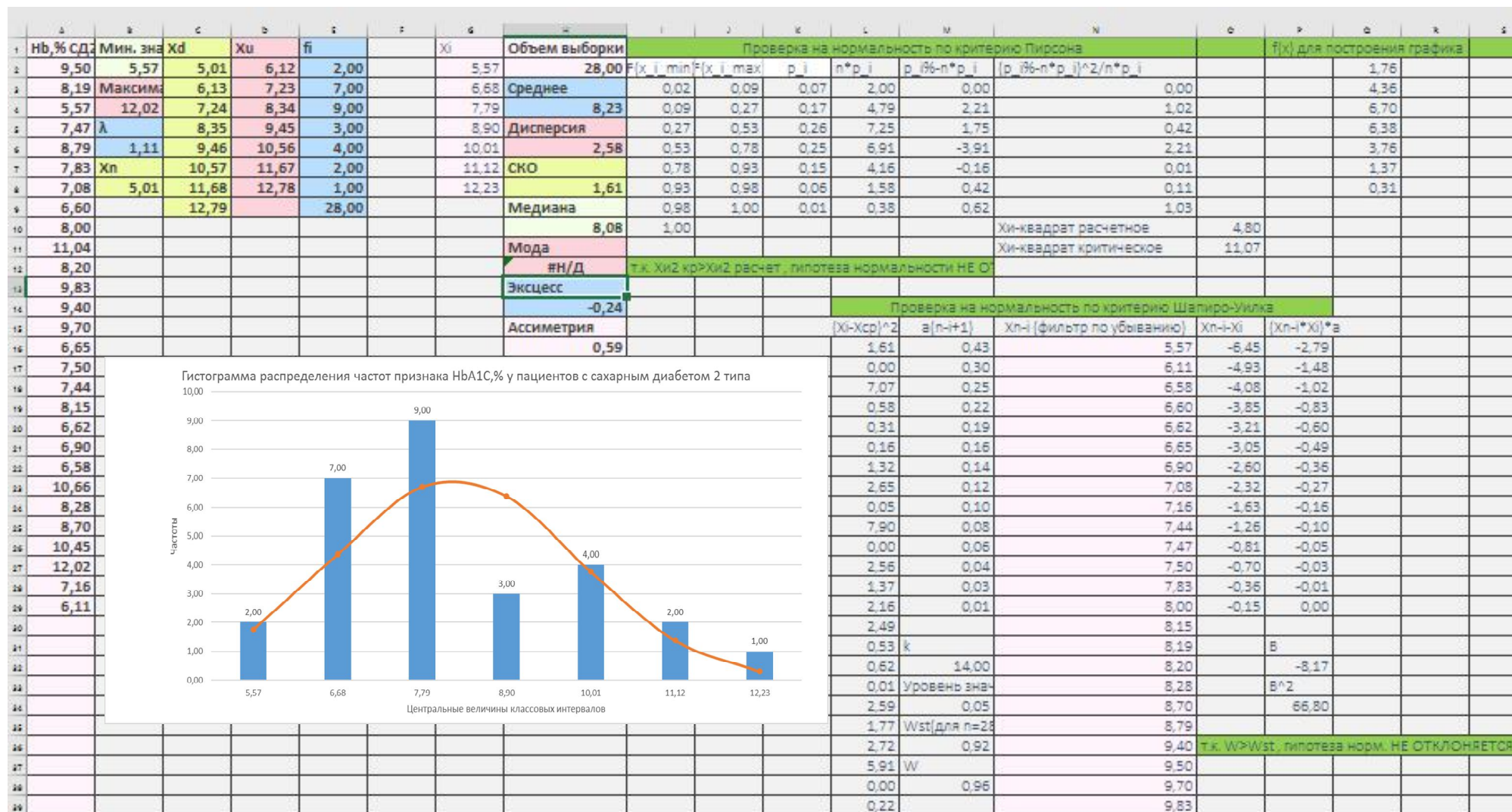


Рисунок А3 – Обработка выборки из историй болезней по значению HbA1C у пациентов с сахарным диабетом 2 типа

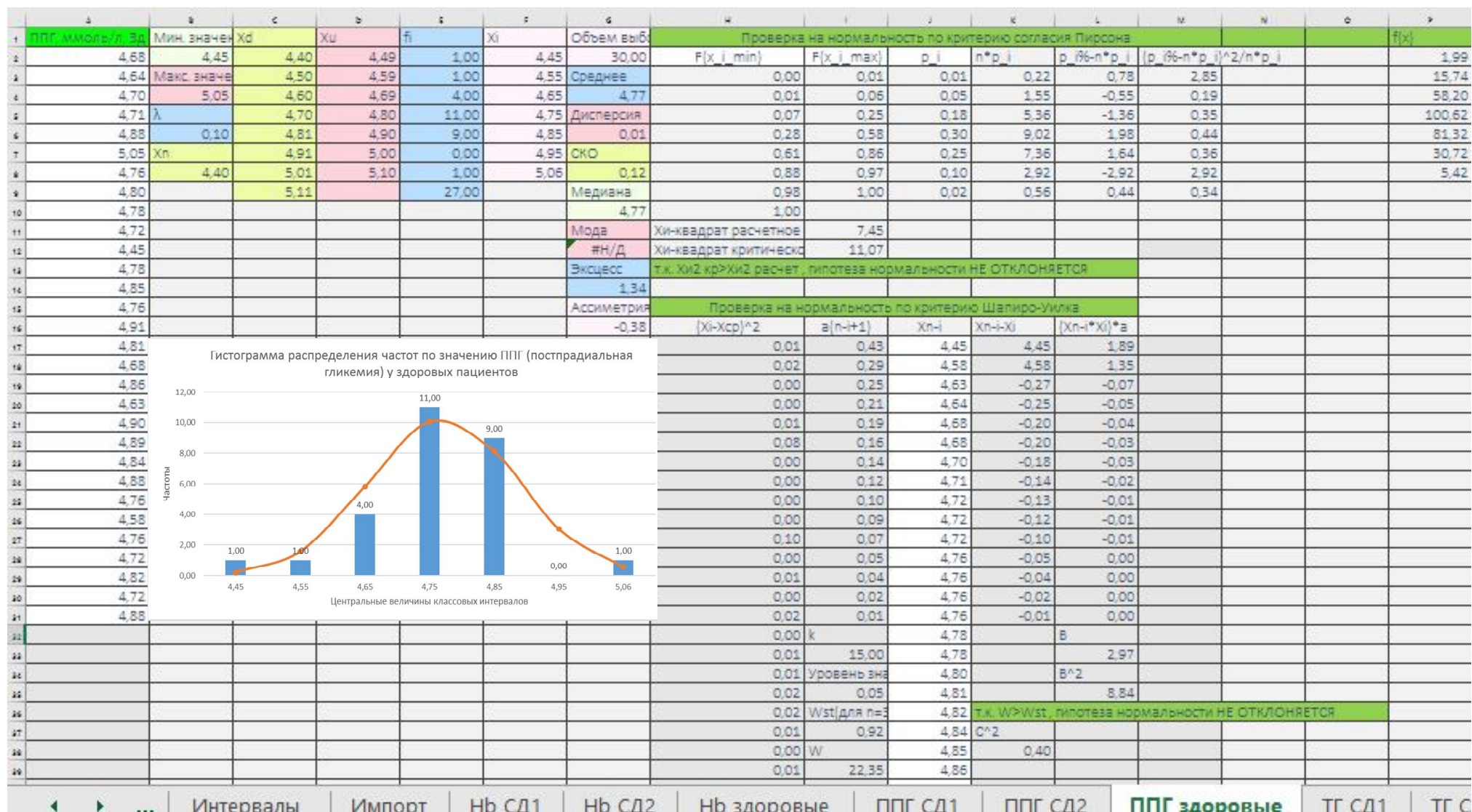


Рисунок А4 – Обработка модельной (генерированной) выборки по значению ППГ (постпрадиальная гликемия) у здоровых пациентов



Рисунок А5 – Обработка выборки из историй болезней по значению ППГ (постпрадиальная гликемия) у пациентов с сахарным диабетом 1 типа



Рисунок А6 – Обработка выборки из историй болезней по значению ППГ (постпрадиальная гликемия) у пациентов с сахарным диабетом 2 типа

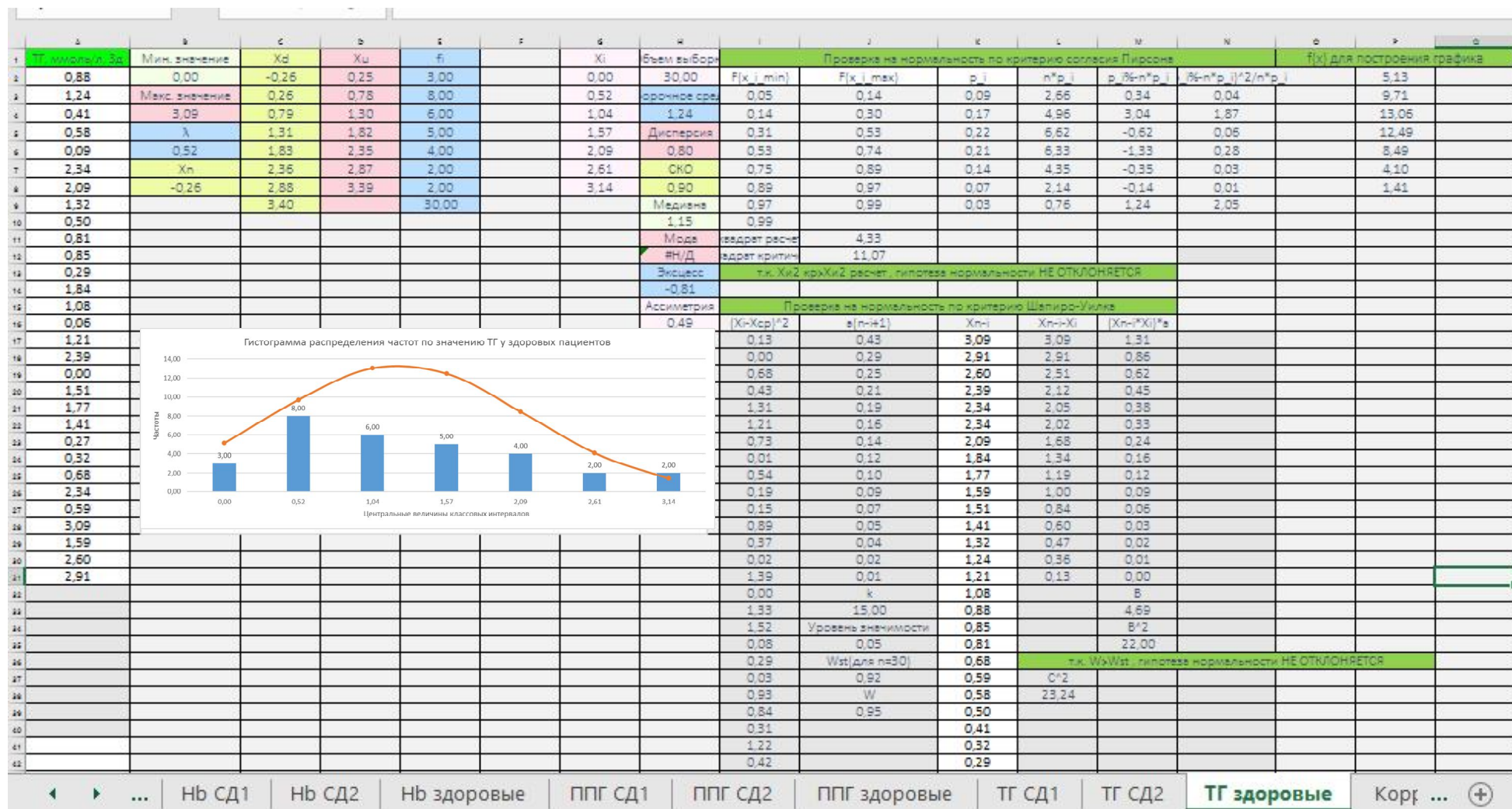


Рисунок А7 – Обработка модельной (генерированной) выборки по значению ТГ (триглицериды) у здоровых пациентов

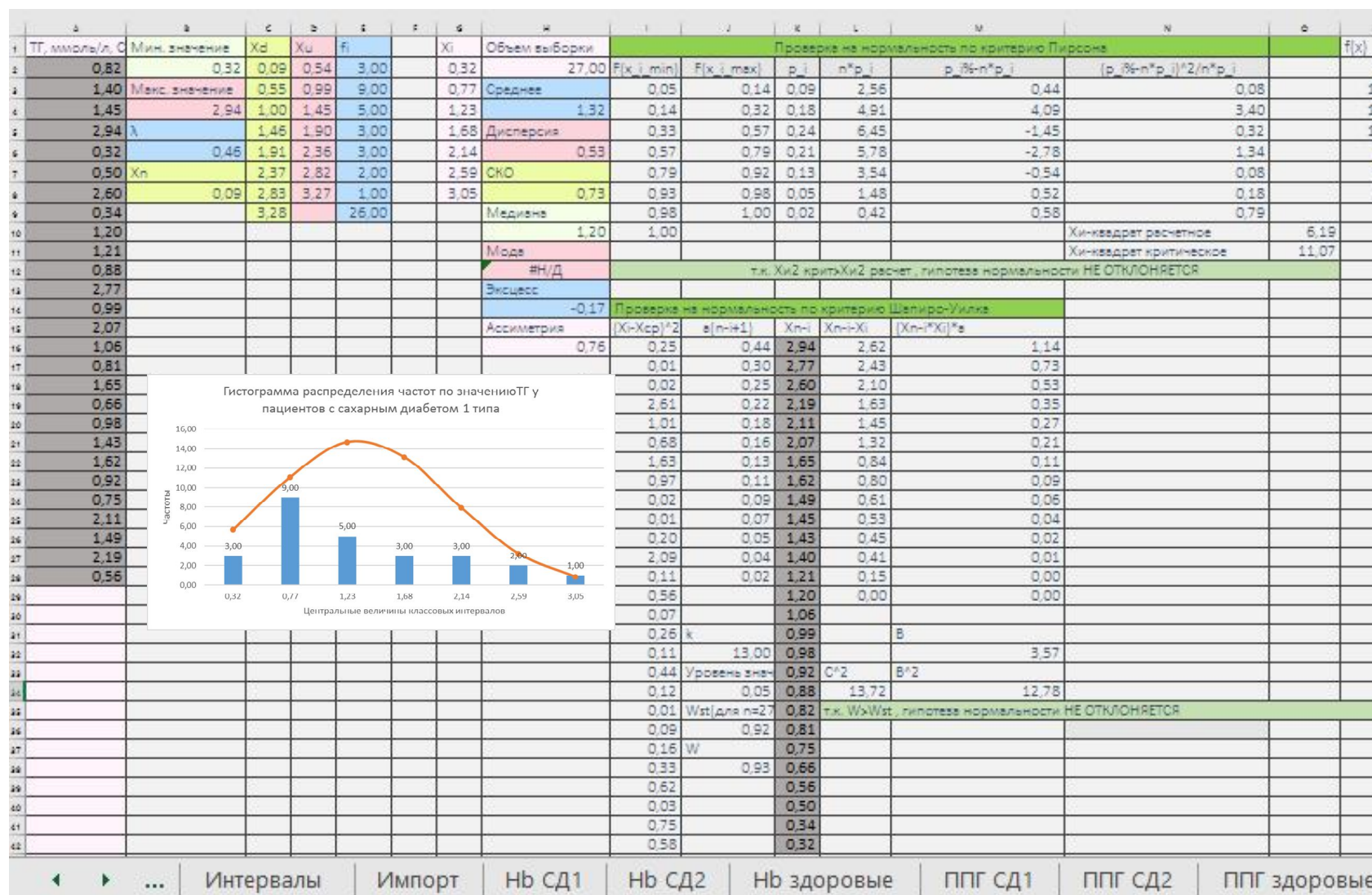


Рисунок А8 – Обработка выборки из историй болезней по значению ТГ (триглицериды) у пациентов с сахарным диабетом 1 типа

ПРИЛОЖЕНИЕ Б

ПРОГРАММНЫЙ КОМПЛЕКС ФОРМИРОВАНИЯ ДИАГНОСТИЧЕСКОЙ МОДЕЛИ И РАСПОЗНОВАНИЯ САХАРНОГО ДИАБЕТА МЕТОДОМ БАЙЕСА

Программный код загрузки исходных данных формирования
диагностической модели

```
%кнопка "Загрузка"
function pushbutton1_Callback(hObject, eventdata, handles)
%объявляет переменные глобальными
global S myicon
myicon=imread('zm.jpg');
%диалоговое окно открытия файла с содержимым текущего каталога
[FName, PName] = uigetfile('*.*)
if ~ isequal(FName, 0)
%если файл выбран, происходит считывание данных из файла
FullName = strcat(PName, FName);
%импортирование данных
S=importdata(FullName)
end
%Замена значений "NaN" на "0"
S(isnan(S))=0
%вывод матрицы в таблицу
set(handles.uitable1, 'Visible','on');
set(handles.uitable1, 'Data',S);
```

Программный код формирования диагностической модели

```
%кнопка "Оценка условных вероятностей интервалов значений
признаков, P(uj1/Ei)"
function pushbutton2_Callback(hObject, eventdata, handles)
%объявляет переменные глобальными
global S n k
% количество строк
n=size(S,1);
%количество столбцов
k=size(S,2);
%расчет условных вероятностей интервалов значений признаков
p(uj1/Ei)
for i =1:n
for j =2:k
S(i,j)=S(i,j)/S(i,1)
end
end
%вывод матрицы в таблицу
```

```

set(handles.uitable1, 'Visible','on');
set(handles.uitable1, 'Data',S);
%кнопка "Оценка вероятности диагноза p(Ei)"
function pushbutton3_Callback(hObject, eventdata, handles)
%объявляет переменные глобальными
global S n k
%расчет суммы пациентов, знаменателя для формулы  $p(E_i)=M_i/M$ 
d=sum(S(:,1));
%расчет оценки вероятности диагноза,  $p(E_i)=M_i/M$ 
for i =1:n;
S(i,1) = S(i,1)/d;
end
%вывод матрицы в таблицу
set(handles.uitable1, 'Visible','on');%вывод матрицы в таблицу
set(handles.uitable1, 'Data',S);

```

Программный код распознавания методом Байеса по одному признаку

```

%кнопка "Метод Байеса по признаку HbA1C,%"
function pushbutton4_Callback(hObject, eventdata, handles)
%объявляет переменные глобальными
global S n k myicon custom myicon1
%диалоговое окно ввода данных
w = inputdlg({'Введите значение HbA1C,%'})
a = str2num(w{1})
%определение принадлежности введенного значения HbA1C к
определенному интервалу
%интервал u11
if a>=3.49 & a<=4.74
S1=S(:,2)
end
%интервал u12
if a>=4.75 & a<=6
S1=S(:,3)
end
%интервал u13
if a>=6.01 & a<=7.26
S1=S(:,4)
end
%интервал u14
if a>=7.27 & a<=8.52
S1=S(:,5)
end
%интервал u15
if a>=8.53 & a<=9.78
S1=S(:,6)
end
%интервал u16
if a>=9.79 & a<=11.04
S1=S(:,7)
end

```

```

%интервал u17
if a>=11,05 & a<=12.3
S1=S(:,8)
end
%интервал u18
if a>=12.31 & a<=13.56
S1=S(:,9)
end
%допустимые значения HbA1C, в случае, если
HbA1C>=13.57, выводится сообщение об ошибке
if a>=13.57
errordlg({'Недопустимое значение'}, 'Ошибка')
end
%допустимые значения HbA1C, в случае, если HbA1C<=3.48, выводится
сообщение об ошибке
if a<=3.48
h=errordlg({'Недопустимое значение'}, 'Ошибка')
end
%расчет числителя  $p(E_i) \cdot p(u_{j1}/E_i)$  для состояния "Здоров"
D1= S(1,1)*S1(1,:);
%расчет числителя  $p(E_i) \cdot p(u_{j1}/E_i)$  для состояния "СД1"
D2= S(2,1)*S1(2,:);
%расчет числителя  $p(E_i) \cdot p(u_{j1}/E_i)$  для состояния "СД2"
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 "СД1"
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 "СД2"
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.uitable2, 'Visible','on');
set(handles.uitable2, 'Data',U11);

```

```

%принятие решение о диагнозе
% если апостериорная вероятность здорового состояния больше
апостериорных вероятностей СД1 и СД2, диагноз "Здоров"
if F1>F2 & F1>F3
e=msgbox ('Наиболее вероятным является состояние "Здоров" для
введенного значения,HbA1C,%', 'Результат', 'custom',myicon)
end
% если апостериорная вероятность СД1 больше апостериорных
вероятностей СД2 и здорового состояния, диагноз "СД1"
if F2>F1 & F2>F3
myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние "Сахарный
диабет 1 типа" для введенного
значения,HbA1C,%', 'Результат', 'custom',myicon1)
end
% если апостериорная вероятность СД2 больше апостериорных
вероятностей СД1 и здорового состояния, диагноз "СД2"
if F3>F2 & F3>F1
myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние "Сахарный
диабет 2 типа" для введенного
значения,HbA1C,%', 'Результат', 'custom',myicon1)
end

```

Программный код распознавания методом Байеса по одному признаку для значений ППГ, ТГ, Ац аналогичен.

Программный код распознавания методом Байеса по сочетанию признаков

```

function pushbutton5_Callback(hObject, eventdata,
handles)%кнопка "Метод Байеса по нескольким признакам"
global S n k myicon custom myicon1; %объявление глобальных
переменных
w = inputdlg({'Введите значение HbA1C,%', 'Введите значение ППГ,
ммоль/л', 'Введите значение ТГ, ммоль/л', 'Введите значение
АЦ, 0/1'}, 'Ввод данных')% диалоговое окно ввода данных
a = str2num(w{1})% введенное значение HbA1C
%определение принадлежности введенного значения HbA1C к
определенному
%интервалу
if a>=3.49 & a<=4.74 %интервал u11
S1=S(:,2)
end
% интервал u12
if a>=4.75 & a<=6
S1=S(:,3)
end
end

```

```

% интервал u13
if a>=6.01 & a<=7.26
S1=S(:,4)
end
% интервал u14
if a>=7.27 & a<=8.52
S1=S(:,5)
end
%интервал u15
if a>=8.53 & a<=9.78
S1=S(:,6)
end
% интервал u16
if a>=9.79 & a<=11.04
S1=S(:,7)
end
% интервал u17
if a>=11,05 & a<=12.3
S1=S(:,8)
end
%интервал u18
if a>=12.31 & a<=13.56
S1=S(:,9)
end
a1 = str2num(w{2})%введенное значение ППГ,ммоль/л
%определение принадлежности введенного значения ППГ к
определенному интервалу
%интервал u21
if a1>=2.97 & a1<=5.64
S2=S(:,10)
end
% интервал u22
if a1>=5.65 & a1<=8.31
S2=S(:,11)
end
%интервал u23
if a1>=8.32 & a1<=10.98
S2=S(:,12)
end
% интервал u24
if a1>=10.99 & a1<=13.65
S2=S(:,13)
end
%интервал u25
if a1>=13.66 & a1<=16.32
S2=S(:,14)
end
% интервал u26
if a1>=16.33 & a1<19
S2=S(:,15)
end
% интервал u27
if a1>=19.01 & a1<=21.67

```

```

S2=S(:,16)
end
%интервал u28
if a1>=21.68 & a1<=24.34
S2=S(:,17)
end
a2 = str2num(w{3})%введенное значение ТГ,ммоль/л
%определение принадлежности введенного значения ТГ к
определенному интервалу
%интервал u31
if a2>=0 & a2<=0.57
S3=S(:,18)
end
%интервал u32
if a2>=0.58 & a2<=1.73
S3=S(:,19)
end
%интервал u33
if a2>=1.74 & a2<=2.88
S3=S(:,20)
end
%интервал u34
if a2>=2.89 & a2<=4.04
S3=S(:,21)
end
%интервал u35
if a2>=4.05 & a2<=5.19
S3=S(:,22)
end
%интервал u36
if a2>=5.2 & a2<6.35
S3=S(:,23)
end
%интервал u37
if a2>=6.36 & a2<=7.5
S3=S(:,24)
end
%интервал u38
if a2>=7.51 & a2<=8.66
S3=S(:,25)
end
a3 = str2num(w{4})%введенное значение Ац,
%определение принадлежности введенного значения Ац к
определенному интервалу
%интервал u41
if a3==0
S4=S(:,26)
end
%интервал u42
if a3==1
S4=S(:,27)
end
%случай, если значение HbA1C отсутствует

```

```

if a==999
S1(:,1)=[1;1;1]
end
%случай, если значение ППГ отсутствует
if a1==999
S2(:,1)=[1;1;1]
end
%случай, если значение ТГ отсутствует
if a2==999
S3(:,1)=[1;1;1]
end
%случай, если значение АЦ отсутствует
if a3==999
S4(:,1)=[1;1;1]
end
%Расчет числителя обобщенной формулы Байеса  $p(E_i/U_{j1})$ 

X1=S(1,1)*S1(1,1)*S2(1,1)*S3(1,1)*S4(1,1)
X2=S(2,1)*S1(2,1)*S2(2,1)*S3(2,1)*S4(2,1)
X3=S(3,1)*S1(3,1)*S2(3,1)*S3(3,1)*S4(3,1)
Y1=0
%Расчет знаменателя обобщенной формулы Байеса  $p(E_i/U_{j1})$ 
%Расчет знаменателя для состояния "Здоров"
for i=1:n
y1=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
Y1 = Y1+y1
End
%апостериорная вероятность здорового состояния
W1=X1/Y1
%Расчет знаменателя для состояния "СД1"
Y2=0
for i=1:n
y2=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
Y2 = Y2+y2
End
%апостериорная вероятность СД1
W2=X2/Y2
%Расчет знаменателя для состояния "СД2"
Y3=0
for i=1:n
y3=S(i,1)*S1(i,1)*S2(i,1)*S3(i,1)
Y3 = Y3+y3
End
%апостериорная вероятность СД2
W3=X3/Y3
% рассчитанные апостериорные вероятности
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
% если апостериорная вероятность здорового состояния больше
апостериорных вероятностей СД1 и СД2, диагноз "Здоров"
if W1>W2 & W1>W3

```

```

% диалоговое окно вывода диагноза
e=msgbox ('Наиболее вероятным является состояние "Здоров" ',
,'Результат','custom',myicon);
end
% если апостериорная вероятность СД1 больше апостериорных
вероятностей СД2 и здорового состояния, диагноз "СД1"
if W2>W1 & W2>W3
myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние "Сахарный
диабет 1 типа" ', 'Результат','custom',myicon1
end
% если апостериорная вероятность СД2 больше апостериорных
вероятностей СД1 и здорового состояния, диагноз "СД2"
if W3>W2 & W3>W1 myicon1=imread('zmkr.jpg');
e=msgbox ('Наиболее вероятным является состояние "Сахарный
диабет 2 типа" ', 'Результат','custom',myicon1);
end

```

Программный код кнопки «Справка»

```

%кнопка "Справка"
function pushbutton9_Callback(hObject, eventdata, handles)
%вывод сообщения с краткой информацией о программном комплексе
%распознавания
myicon1=imread('zmkr.jpg')
msgbox({'Программный комплекс распознавания сахарного диабета
предназначается для внедрения в диагностических центрах и в
учебный процесс медицинского института.'
'Распознавание состояния здоровья осуществляется статистическим
методом, а именно методом Байеса.'
'Апостериорные вероятности классов состояния здоровья
вычисляются на основе вероятностной диагностической модели,
содержащей априорные вероятности классов состояния здоровья и
условные плотности распределений вероятностей значений
диагностических признаков.'
'Программный комплекс осуществляет формирование вероятностной
диагностической модели, на основе данных таблицы «Результаты
лабораторного обследования», загружаемой из книги MS Excel. Для
формирования вероятностной диагностической модели необходимо
воспользоваться кнопкой «Загрузка» для импортирования данных, а
затем кнопками: "Оценка условных вероятностей интервалов
значений признаков,  $P(u_{jl}/E_i)$ " и «Оценка вероятности диагноза
 $p(E_i)$ ».'
'Распознавание состояния здоровья возможно как по одному
признаку (кнопки: «Метод Байеса по признаку HbA1C», «Метод
Байеса по признаку ТГ, ммоль/л», «Метод Байеса по признаку
ППГ, ммоль/л», «Метод Байеса по признаку Ац») так и по сочетанию
диагностических признаков (кнопка «Метод Байеса по нескольким
признакам»' })

```