

Диссертацией является рукопись.

Работа выполнена в Донецком национальном техническом университете
Министерства образования и науки Украины.

Научный руководитель:

доктор технических наук, профессор
Скобцов Юрий Александрович
Донецкий национальный технический
университет МОН Украины, г. Донецк,
заведующий кафедры автоматизированных
систем управления.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Предвидение явлений в медицине является наиболее актуальной задачей в профессиональной деятельности врача. Совершенствование известных и создание новых методов диагностики и прогнозирования позволяет повышать качество оказания медицинской помощи населению.

В течение нескольких десятилетий внимание ученых всего мира привлекает одна из драматичных и все еще невыясненных проблем медицины – синдром внезапной смерти грудного ребенка (СВСГР). Интерес к этой проблеме не ослабевает, так как число жертв СВСГР не имеет тенденции к снижению. СВСГР является одной из ведущих причин смерти младенцев, а в некоторых зарубежных странах выходит на лидирующие позиции в общей структуре детской смертности.

Данная проблема актуальна как в развивающихся, так и в развитых странах, где число случаев СВСГР составляет: в Австрии 0,54, Германии – 0,78, Италии – 1,0, Голландии – 0,14, России – 0,43, Словакии – 0,14, в Украине смертность от СВСГР составляет 0,7 на 1000 детей рожденных живыми. Во многих странах проводятся программы по профилактике СВСГР, а в США в 1983 г. создан Национальный институт СВСГР.

В связи с этим актуальной научной задачей является создание экспертной системы (ЭС) для определения степени риска СВСГР, использующей современные методы извлечения знаний. Применение экспертной системы позволит определять высокую степень риска СВСГР на различных сроках беременности и сразу после рождения ребенка, что позволит раньше начинать профилактические меры, направленные на снижение степени риска СВСГР.

Связь работы с научными программами, планами, темами. Диссертационная работа выполнена в рамках госбюджетных тем Донецкого национального технического университета Д-11-04 «Разработка и

исследования нейросетевых и эволюционных методов построения систем принятия решения» (№ государственной регистрации 0104U002394), Н-3-07 «Разработка научных основ построения компьютерных систем технической медицинской диагностики» (98-15 от 04.04.07) и Д-11-07 «Разработка и исследование эволюционных и нейросетевых моделей, методов и алгоритмов в системах диагностики и прогнозирования» (№ государственной регистрации 0107U001482).

Цель работы и задачи исследования. Цель диссертационной работы - разработка экспертной системы определения степени риска СВСГР на различных сроках беременности и сразу после рождения ребенка.

Для разработки экспертной системы определения степени риска СВСГР необходимо решить следующие задачи:

- разработка структуры экспертной системы;
- разработка методов кодирования данных;
- разработка методов выделения информативных данных;
- разработка методов интеллектуального анализа данных;

Объект исследования: процесс проектирования и разработки методов извлечения знаний для медицинских экспертных систем.

Предмет исследования: экспертная система определения степени риска синдрома внезапной смерти грудного ребенка.

Методы исследования, используемые в работе, базируются на методах искусственного интеллекта, основных положениях теории вероятности и математической статистики.

Научная новизна полученных результатов.

1. Впервые разработана ЭС предназначенная для определения степени риска СВСГР, что позволило определять степень риска еще на во время беременности и сразу после рождения ребенка.

2. Разработана фитнес-функция для генетического алгоритма, реализующего выбор информативных факторов риска, что позволило

достигнуть высокой эффективности выделения информативных факторов риска в медицине за счет регулирования соотношения количества факторов и ошибки классификации.

3. Усовершенствован древовидный способ кодирования хромосомы для булевой функции, что позволило получать классификационные деревья для определения степени риска СВСГР.

4. Усовершенствован способ кодирования хромосомы в виде дерева, реализующего булеву функцию в дизъюнктивной нормальной форме, что позволило получать классификационные правила для определения степени риска СВСГР.

5. Впервые предложена троичная логика в булевых вычислениях классификационного дерева, что позволило выполнять классификацию в условиях неопределенных значениях некоторых входных параметров.

Практическое значение работы.

1. Разработана и внедрена экспертная система определения степени риска СВСГР в городской больнице №3 города Донецка, что позволило определять степень риска во время беременности и начинать профилактические меры еще до рождения ребенка.

2. В госбюджетной теме Д-11-04 «Разработка и исследования нейросетевых и эволюционных методов построения систем принятия решения» применены алгоритмы и программное обеспечение системы прогнозирования атеротромбогенного инсульта на базе нейронных сетей. В госбюджетной теме Н-3-07 «Разработка научных основ построения компьютерных систем технической медицинской диагностики» применены алгоритмы и программное обеспечение системы определения высокой степени риска СВСГР на основе генетического программирования (ГП). В госбюджетной теме Д-11-07 «Разработка и исследование эволюционных и нейросетевых моделей, методов и алгоритмов в системах диагностики и прогнозирования» применены алгоритмы и программное обеспечение

системы определения степени риска СВСГР на основе ГП в условиях неопределенных некоторых входных состояний.

3. Разработанные алгоритмы и программное обеспечение применяются в учебном процессе при выполнении курсовых работ, дипломных проектов и магистерских работ студентами специальностей «Информационные управляющие системы и технологии» и «Специализированные компьютерные системы» Донецкого национального технического университета.

Публикации. По теме диссертации опубликовано 17 научных работ, из них 7 – в ведущих специальных изданиях, входящих в перечень ВАК Украины, 2 патента на полезную модель, остальные в сборниках конференций.

Апробация результатов диссертации. Результаты диссертационной работы докладывались на международной научной конференции «Нейросетевые технологии и их применение» (ДГМА, Краматорск, 2002г.); 1-й и 3-й международных конференциях «Advanced Computer Systems and Networks: Design and Application» (НУ«ЛП», Львов, 2003г., 2007г.); II научной – конференции «ДОНБАС -2020: НАУКА І ТЕХНІКА – ВИРОБНИЦТВУ» (ДонНТУ, Донецк 2004г.); на V международной научно-практической конференции «Комп'ютерні системи в автоматизації виробничих процесів» (ХНУ, Хмельницький, 2007); IX международной научно-практической конференции «Інформаційні технології в освіті та управлінні» (ХНТУ, Новая Каховка, 2007), международной научной конференции «Компьютерные науки и информационные технологии» (Саратов, 2007), международных научно-практических конференциях «ИНФОТЕХ-2007», «ИНФОТЕХ-2009» (СевНТУ, Севастополь, 2007, 2009), IX международной научно технической конференции «Искусственный интеллект. Интеллектуальные системы – ИИ-2008» (ИПИИ, Кадивели, 2008), международной научной конференции «Інтелектуальні системи прийняття

рішень і проблеми обчислювального інтелекту (ISDMCI'2008)» (ХНТУ, Євпаторія, 2008), II Всеукраїнська научно-практична конференція «Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці» (Луганськ, 2008).

Структура и объем диссертации. Диссертация состоит из введения, 5 разделов, заключения, списка литературы из 93 источников и 10 приложений. Она содержит 62 иллюстрации, 48 таблиц. Общий объем диссертации составляет 217 страниц, в том числе 166 страниц основного текста.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность диссертационной работы. Сформулированы цель и задачи исследований. Изложены основные результаты, полученные в работе, показана их новизна, а также отмечены их научная и практическая значимость.

Первый раздел работы посвящен обзору научной литературы, определению направлений исследований. Рассмотрены современные научные подходы, направления и методы прогнозирования в медицине. Показано, что задачи диагностики и прогнозирования это комплексный процесс и рассмотрены его этапы.

Выполнен анализ способов представления обучающих данных, приведена классификация входной информации и способы кодирования в зависимости от типа данных. Проведен анализ методов выбора информативной информации из общего набора всех параметров.

Проанализированы современные методы построения экспертных систем. Показано, что большинство из них имеют существенные недостатки, так как не решают поставленную задачу или делают это не оптимальным образом. Обоснован новый метод построения экспертной системы с использованием генетического программирования.

На основании проведенного анализа сформулирована цель, поставлены

задачи и выбраны направления исследований для разработки экспертной системы медицинской диагностики (на примере определения высокой степени риска СВСГР).

Во втором разделе разработана структура и описаны функции экспертной системы.

Для определения степени риска СВСГР экспертная система должна выполнять следующие функции:

- ввод общей информации о пациенте, течении текущей беременности и исходах предыдущих;
- обработка информации: представление ее в виде, пригодном для дальнейших действий (обучение или эксплуатация системы);
- оценка значимости каждого фактора в отдельности для всего общего набора входной информации;
- выделение информативных факторов риска из общего набора входной информации;
- обучение экспертной системы, система должна позволять получать новые знания при накоплении информации;
- прогнозирование высокой степени риска СВСГР, на различных этапах беременности и после рождения ребенка;
- хранение исходных данных, информации, получаемой в процессе обработки, а также результатов прогнозирования СВСГР.

Показано, что экспертную систему целесообразно реализовать, используя модульный принцип. Использование модульного построения позволяет получить максимальную гибкость настройки, высокую универсальность, а также возможность расширения функциональности системы в процессе эксплуатации. Разработанная структура экспертной системы представлена на рисунке 1.

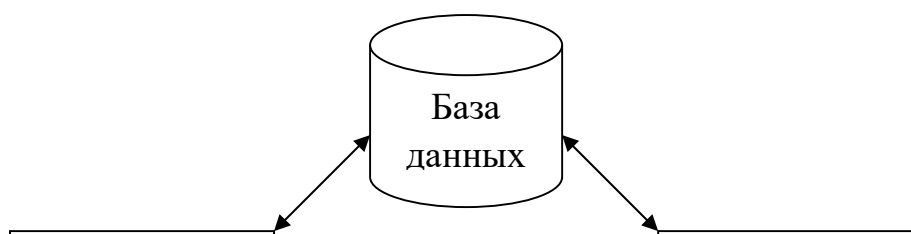


Рисунок 1. Структура экспертной системы.

Выполнена постановка задачи к формированию знаний для ЭС. Рассмотрено, что при решении задач предсказания необходимо определить функцию вида:

$$Y = f(x_1, x_2, \dots, x_n), \quad (1)$$

где Y - исследуемая величина, зависящая от факторов x_1, x_2, \dots, x_n . Искомая функция может быть найдена явно или нет.

Предсказание в зависимости от того какой вид имеет целевая переменная может быть выполнено на основе классификации или регрессии. Мы рассматриваем задачу классификации.

1. Пусть необходимо выполнить классификацию на l классов, и функция для классификации имеет вид формула (2).

$$Y = \begin{cases} K_1, F \in [0; N_1) \\ K_2, F \in [N_1; N_2) \\ \dots \\ K_l, F \in [N_{l-1}; N_l] \end{cases} \quad 2)$$

$$F = \sum_{i=1}^n x_i * w_i \quad 3)$$

где $K=(K_1, K_2, \dots, K_l)$ - возможные результаты классификации; F - некоторая функция представленная формулой (3); $X=(x_1, x_2, \dots, x_n)$ - факторы риска; $W=(w_1, w_2, \dots, w_n)$ - весовые коэффициенты факторов риска; $N=(N_1, N_2, \dots, N_l)$ - некоторые пороговые значения. Разработка метода классификации заключается в нахождении весов W , и пороговых значений N , а непосредственно классификация в суммировании весов присутствующих у пациента факторов риска.

2. Пусть нет необходимости найти функцию (1) явно. Необходимо разработать математический аппарат, который предскажет поведение этой функции. Примером реализации может быть нейронная сеть.

3. Пусть необходимо получить классификационные правила для ЭС с двумя возможными результатами классификации. Тогда система правил может быть представлена, например, так:

ЕСЛИ *условие 1=да* или *условие 2=да* или ... или *условие m=да* ТО $Y=K_1$, ИНАЧЕ $Y=K_2$.

Перечисленные условия представляют собой классификационные правила, которые можно математически представить следующим образом:

ЕСЛИ $f_1(x_1, x_2, \dots, x_n)=1$ или $f_2(x_1, x_2, \dots, x_n)=1$ или ... или $f_m(x_1, x_2, \dots, x_n)=1$ ТО $Y=K_1$, ИНАЧЕ $Y=K_2$.

Тогда необходимо найти функции вида:

$$f_j(x_1, x_2, \dots, x_n)=1 \quad (4)$$

Функции (4) должны быть представлены булевым выражением и

состоять из логических функций: И, ИЛИ, НЕ. Исключим ИЛИ из функций (4), теперь вся конструкция ЕСЛИ будет представлена в дизъюнктивной нормальной форме (ДНФ) и иметь вид:

$$F(x_1, x_2, \dots, x_n) = \bigvee_{j=1}^m f_j(x_1, x_2, \dots, x_n) \quad (5)$$

Разработка классификационных правил заключается в получении булевых функций (4), и вся система правил имеет следующий вид:

$$\text{ЕСЛИ } F(x_1, x_2, \dots, x_n) = 1 \text{ ТО } Y = K_1, \text{ ИНАЧЕ } Y = K_2 \quad (6)$$

Выполнена постановка задачи к предварительной обработке данных. Предположим, существует набор данных, который содержит m примеров. Каждый пример состоит из n определяющих параметров и параметр – результат формулы (7) и (8).

$$S = \{X_{i,j}, Y_i\}, i \in [1, m] \quad (7)$$

где

$$X_j = \{x_1, x_2, \dots, x_n\} \quad (8)$$

Т.е. каждый i -й пример набора данных S представлен набором факторов X и результатом Y . При этом факторы $\{x_1, x_2, \dots, x_n\}$ набора данных S могут включать в себя как подмножество факторов содержащих полезную информацию о принадлежности к Y , так и подмножество факторов частично или полностью неинформативных. При этом определенную величину шума могут содержать и информативные факторы.

Критерием оценки качества может быть оценка качества всего набора данных S или некоторая мера качества (значимости) каждого отдельного фактора.

При необходимости оценить качество всего набора критерием оценки может быть выбрана точность классификации (например, погрешность), обозначим критерий оценки $E(X)$. Тогда задача сводится к нахождению такого множества X , при котором достигается допустимая точность классификации при минимальном количестве входных параметров:

$$E(X) \rightarrow \min \text{ и } n \rightarrow \min \quad (9)$$

где n , количество факторов.

При вычислении меры значимости каждого фактора, необходимо найти такие веса $W = \{w_1, w_2, \dots, w_n\}$, при котором выполняется условие (10).

$$E(X * W) \rightarrow \min \quad (10)$$

где $E(X * W)$ - критерий оценки классификации (например, погрешность).

Выполнена реализация методов предобработки входных данных, включая кодирование нечисловой информации и нормировку для числовой. Выполнена предварительная обработка данных для различных типов ЭС. Закодированы нечисловые характеристики, учитывая их тип. Для числовой информации выполнена нормировка.

Выполнена разработка и реализация методов выбора информативных факторов для понижения размерности обучающей выборки.

С помощью корреляционного анализа оценена значимость каждого входного параметра отдельно.

Выбран информативный набор факторов риска с помощью нейронных сетей (НС) и генетических алгоритмов (ГА).

На первом этапе подбирается архитектура НС минимальной сложности, которая позволяет с достаточной точностью определять высокую степень риска СВСГР.

Под сложностью сети будем предполагать вычислительную сложность алгоритма обучения. Под вычислительной сложностью алгоритма обучения будем понимать количество операций за один шаг обучения. Известно, что для алгоритма обратного распространения количество операций связано линейной зависимостью с синаптическими весам W (включая пороги).

Учитывая, что количество входов не меняется (определено факторами риска), вычислительную сложность определяет количество скрытых слоев и количество нейронов на каждом из них.

Экспериментально выбрана архитектура сети, которая содержит один скрытый слой. Количество входов обусловлено максимальным количеством факторов риска (после кодирования получили 99). Выбраны активационные функции – гиперболический тангенс на скрытом слое и линейная на выходном. Количество нейронов в сети: на скрытом слое 6 нейронов, на выходном – один, входов – 99.

Далее с помощью ГА определяется оптимальный набор входных параметров

Схема работы генетического алгоритма такова: каждый возможный вариант набора входных переменных представляется в виде битовой маски. Ноль в соответствующей позиции означает, что эта входная переменная не включена во входной набор, единица – что включена. Таким образом, особь представляет собой строку битов – по одному на каждую возможную входную переменную (рис. 2).

Факторы риска

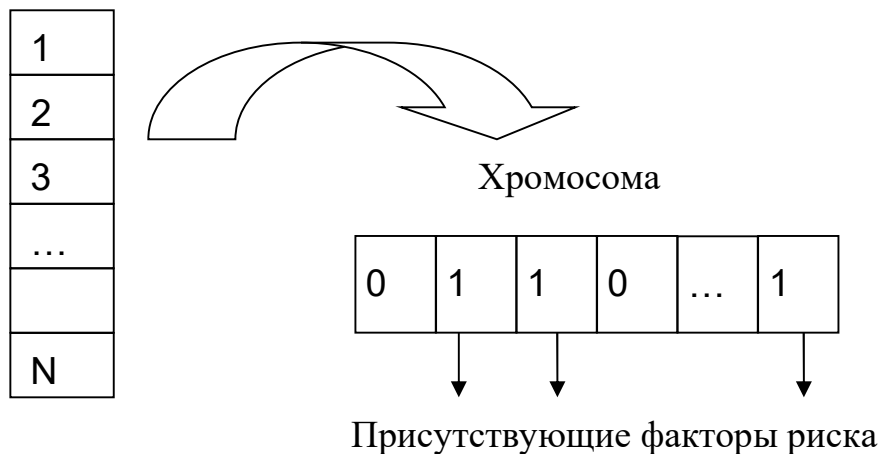


Рисунок 2. Кодирование хромосомы.

Алгоритм следит за некоторым набором таких хромосом, оценивая каждую из них по контрольной ошибке (ошибка обучения НС). По значениям ошибки производится отбор лучших вариантов хромосом, которые

комбинируются друг с другом с помощью искусственных генетических операторов: скрещивания и мутации.

Разработана фитнес-функция для генетического алгоритма, при которой достигается высокая эффективность выделения информативных факторов риска в медицине. Данная фитнес функция позволяет выбирать соотношение точности классификации и количества факторов риска.

$$F = \left(\frac{X_i}{X_n} \right) * Q_1 + \left(\frac{E_i - E_n}{E_n} \right) * Q_2 \quad (11)$$

где X_i – количество факторов для i – ой хромосомы, X_n – максимальное количество факторов, E_i – ошибка обучения для i – ой хромосомы, E_n – ошибка обучения при использовании максимального количества факторов, Q_1 и Q_2 – мера влияния на фитнес-функцию. Мера влияния каждого слагаемого может корректироваться. Диапазон допустимых значений – $Q_1, Q_2 \in (0,1)$, при выполнении условия (12):

$$Q_1 + Q_2 = 1 \quad (12)$$

Выполнена реализация данного метода. Проведены экспериментальные исследования по определению оптимального набора входных параметров, такой набор определен для дальнейших исследований.

В третьем разделе разработаны методы извлечения знаний для ЭС, в соответствии с постановкой задачи в разделе 2.

1. Согласно постановке задачи, функция классификации может быть описана формулами (2) и (3). Из чего следует:

- необходимо задать количество и интерпретации классифицируемых вариантов, $K = \{K_1, K_2, \dots, K_l\}$ – возможные результаты классификации (степень риска СВСГР), где l – количество классов;
- рассчитать весовые коэффициенты $W = \{w_1, w_2, \dots, w_n\}$ для факторов риска $X = \{x_1, x_2, \dots, x_n\}$, где n – количество факторов;
- определить некоторые пороговые значения $N = \{N_1, N_2, \dots, N_l\}$, где l – количество прогнозируемых вариантов;

Получена классификационная таблица с помощью корреляционного анализа, которая представляет собой шкалу оценивания степени риска, а классификация разработанным методом сводится к подсчету баллов для всех присутствующих факторов риска, формула (3). Предложенные пороговые значения и число градаций степени риска, в зависимости от срока беременности, приведены в таблице 1.

Таблица 1.

Определение степени риска СВСГР.

Срок беременности	Степень риска			
	Очень низкий риск	Низкий риск	Высокий риск	Очень высокий риск
До 24 недель	До 14	14—27	27—55	Более 55
После 24 недель	До 15	15—31	31—61	Более 61
После родов	До 20	20—40	40—80	Более 80

2. Для реализации ЭС, когда нет необходимости находить функцию (1) явно, предложено использовать НС. Возможны следующие архитектуры НС (рис. 3-4).

Так же можно использовать аппарат генетического программирования (ГП), который позволяет построить эволюционным алгоритмом на основе обучающей выборки программу, которая выполняет классификацию на два класса. Обобщенный алгоритм работы ГП представлен на рисунке 5. Предлагается использовать древовидную структуру хромосомы. Пример древовидной структуры показан на рисунке 6. Хромосома представляет собой дерево, которому соответствует логическая функция (13).

$$F(x_1, x_2, \dots, x_n) = \overline{(x_1 + x_5)} \cdot \overline{(x_{12} \cdot x_7)} \cdot (x_{21} + \overline{(x_{25} + x_8)} \cdot (x_4 \cdot x_6)) \quad (13)$$

Если в результате расчета этой функции (на выходе дерева) получаем 1, то высокая степень риска СВСГР, 0 – низкая.

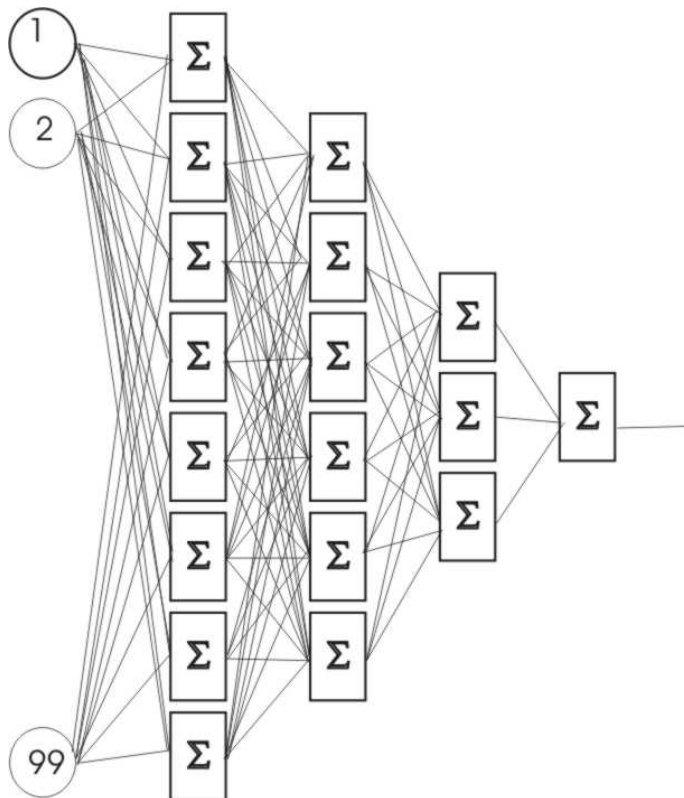


Рисунок 3. Архитектура нейронной сети. Выбранные активационные функции – гиперболический тангенс, сигмоид, гиперболический тангенс, линейная.

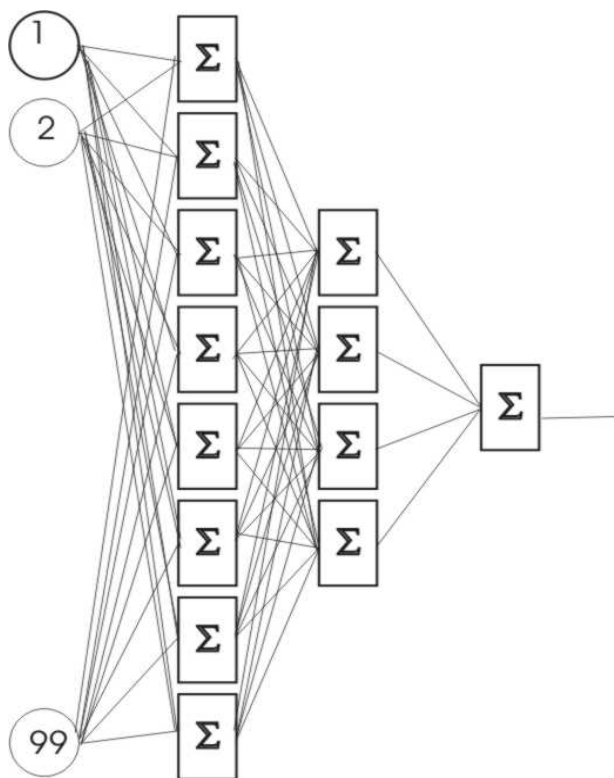


Рисунок 4. Архитектура нейронной сети. Выбранные активационные функции – гиперболический тангенс, сигмоид, линейная.

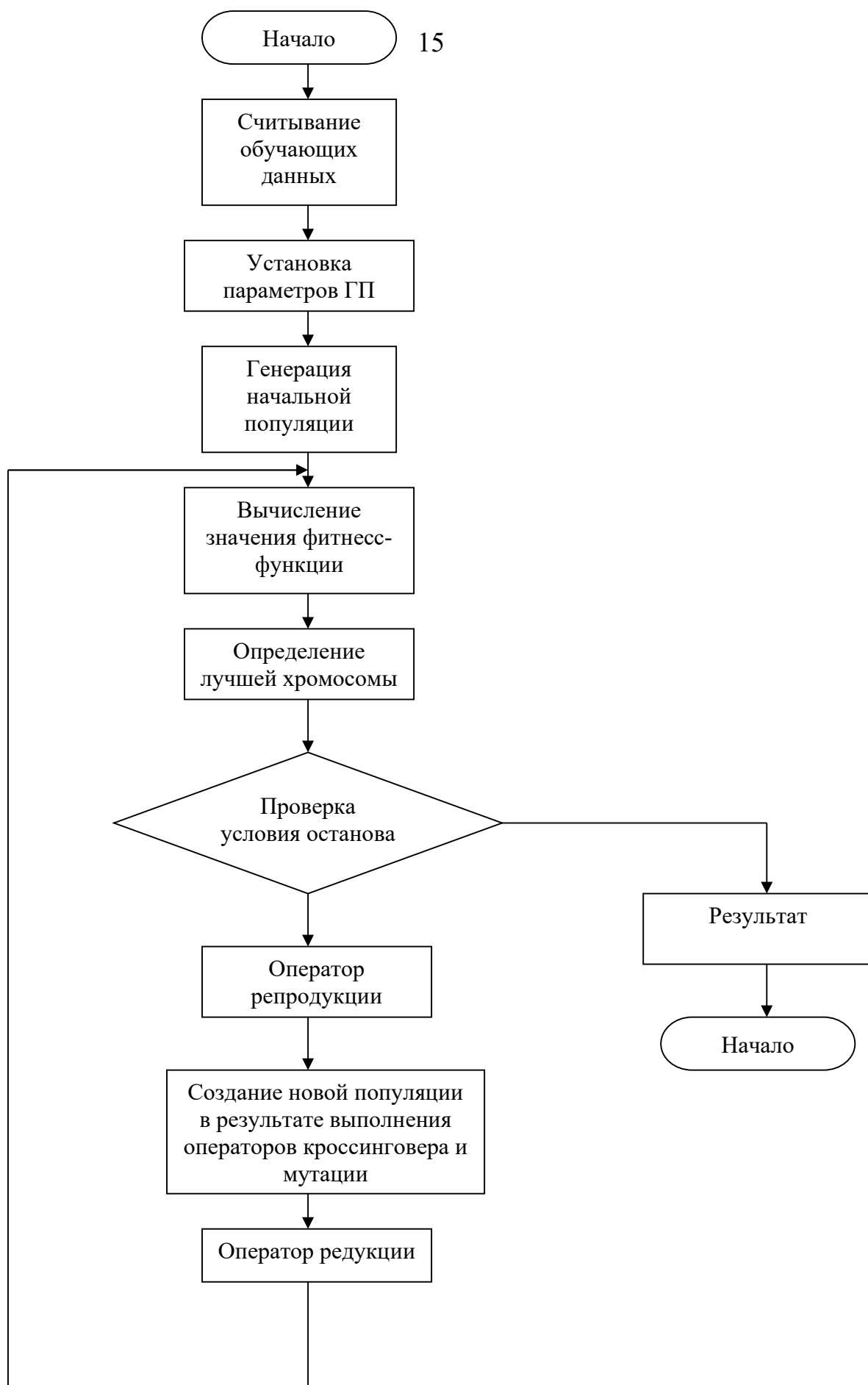


Рисунок 5. Обобщенный алгоритм получения дерева с помощью ГП.

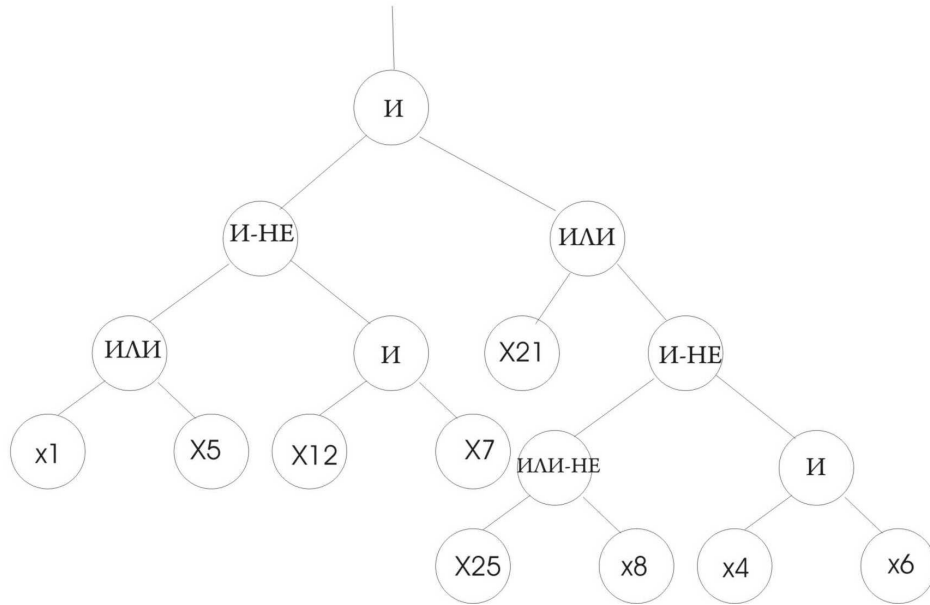


Рисунок 6. Пример дерева прогнозирования.

Терминальное множество в данном случае состоит из выделенных ранее информативных факторов риска, которые после предобработки представляют собой булевозначный вектор. Функциональное множество состоит из логических операций: AND, OR, NOT. Так как первые две операции могут иметь два и более входов и один выход, а последняя операция всегда имеет один вход и один выход, то для более удобной программной реализации заменим операцию NOT на AND-NOT и OR-NOT. Такая замена выполнена с целью унифицировать количество входов для всех операций (их будет всегда 2). Таким образом, функциональное множество состоит из 4 логических операций AND, OR, AND-NOT и OR-NOT. Так же подобная замена позволяет избавиться от так называемых интронов (бесполезных участков кода), например двойное отрицание (NOT(NOT(X))). В качестве фитнес-функции рассматривается доля пациентов с правильно поставленным диагнозом:

$$E = \frac{1}{M} * (M - \sum_{i=1}^M (|F_i - Y_i|)) \quad (14)$$

где M – количество обучающих примеров, F – полученный результат классификации, Y – желаемый результат классификации.

3. Рассмотренный подход, как и НС не позволяет логически объяснять полученный результат. Поэтому предлагается новый метод кодирования особей для генетического программирования, который позволяет интерпретировать булеву функцию в виде классификационных правил. Как и ранее особь представляет собой дерево, только теперь это дерево представляет собой булеву функцию в дизъюнктивной нормальной форме. На рисунке 7 представлен пример дерева соответствующего функции представленной в дизъюнктивной нормальной форме.

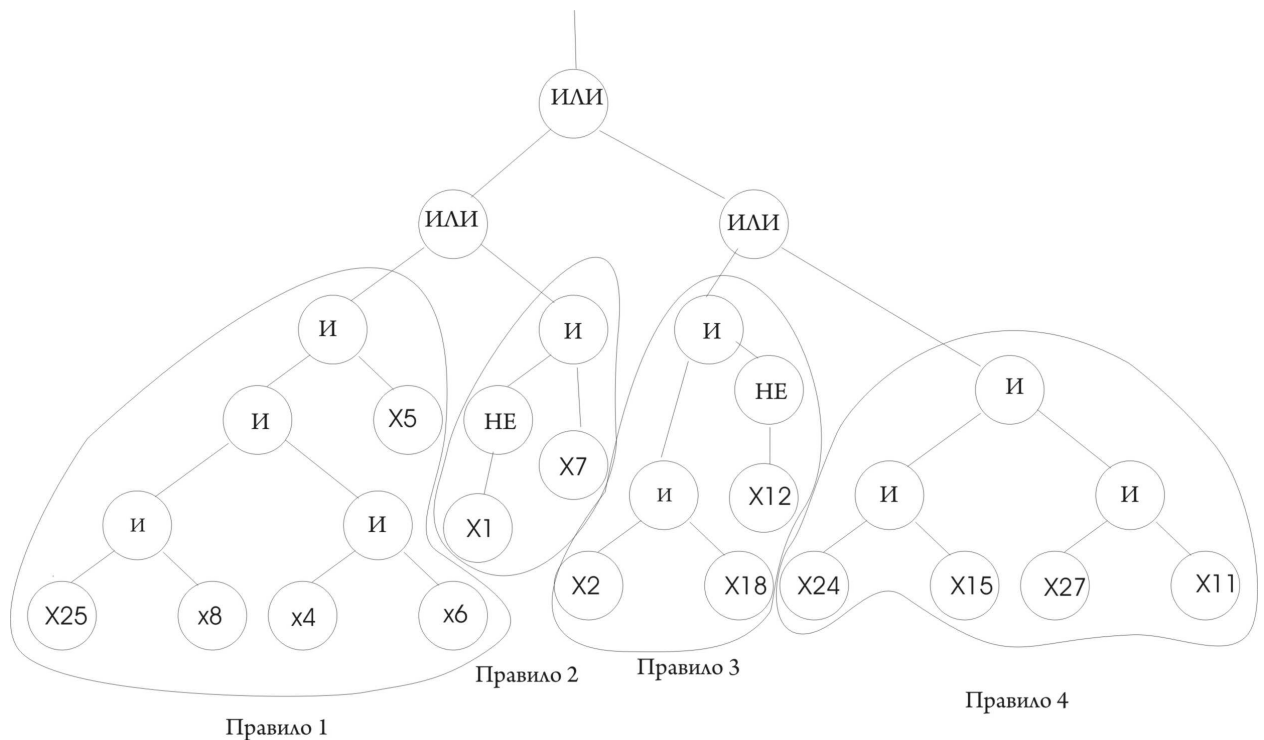


Рисунок 7. Пример дерева, которое представляет булеву функцию в дизъюнктивной нормальной форме.

Дерево представлено 4-мя правилами. Такое представление особи значительно упрощает интерпретацию результата. В данном примере расшифровка будет следующей:

ЕСЛИ выполняется правило 1 ИЛИ выполняется правило 2 ИЛИ выполняется правило 3 ИЛИ выполняется правило 4, ТО результат 1, ИНАЧЕ результат 2.

Где правила 1-4 соответствуют булевым выражениям (15)-(18), а

выполнение какого-либо правила означает, что значение соответствующего ему выражения равно 1.

$$f_1(x_1, x_2, \dots, x_n) = x_{25}x_8 \cdot x_4x_6 \cdot x_5 \quad (15)$$

$$f_2(x_1, x_2, \dots, x_n) = \bar{x}_1x_7 \quad (16)$$

$$f_3(x_1, x_2, \dots, x_n) = x_2x_{18}\bar{x}_{12} \quad (17)$$

$$f_4(x_1, x_2, \dots, x_n) = x_{24}x_{15} \cdot x_{27}x_{11} \quad (18)$$

Соответственно все дерево можно представить выражением (19).

$$F(x_1, x_2, \dots, x_n) = x_{25}x_8 \cdot x_4x_6 \cdot x_5 + \bar{x}_1x_7 + x_2x_{18}\bar{x}_{12} + x_{24}x_{15} \cdot x_{27}x_{11} \quad (19)$$

Выполнение равенства (20) означает выполнение конструкции ЕСЛИ, т.е. результат 1, не выполнение – ИНАЧЕ т.е. результат 2. Применительно к нашей задаче выполнение равенства (20) будет означать высокую степень риска СВСГР, не выполнение – низкую.

$$F(x_1, x_2, \dots, x_n) = 1 \quad (20)$$

где m – количество классификационных правил, n – число терминальных переменных.

При работе с медицинскими данными, достаточно часто возникает ситуация, когда некоторые параметры неизвестны. Причем в большинстве случаев у разных пациентов отсутствуют данные о разных факторах риска, формирование обучающей выборки выполняется с существенной потерей данных. С целью минимизировать потерю данных при обучении и получить возможность работы при неизвестных значениях некоторых факторов риска предлагается использовать троичную логику в булевых вычислениях.

В таблицах 2-4 приведены таблицы истинности для следующих логических функций: И, ИЛИ и НЕ.

В четвертом разделе разработана информационная система, которая представляет собой программный комплекс, предназначенный для надежного

хранения информации, ее обработки и выполнения необходимых вычислений. На рисунке 8 представлена схема обработки информации. Для реализации базы данных построена логическая модель данных. На основании этого спроектирована и внедрена база данных системы, позволяющая вводить, хранить информацию и использовать ее при функционировании экспертной системы определения степени риска СВСГР.

Таблица 2

N ₁	N ₂	И
0	0	0
0	1	0
1	0	0
1	1	1
0	*	0
1	*	*
*	0	0
*	1	*
*	*	*

Таблица 3

N ₁	N ₂	ИЛИ
0	0	0
0	1	1
1	0	1
1	1	1
0	*	*
1	*	1
*	0	*
*	1	1
*	*	*

Таблица 4

N ₁	НЕ
0	1
1	0
*	*

Разработана структура программного обеспечения. Ядром экспертной системы является подсистема интеллектуального анализа данных. Подсистема интеллектуального анализа данных реализована в виде нескольких модулей:

- модуль определения степени риска СВСГР на основе классификационной таблицы. Модуль выполняет считывание из БД информации о наличии факторов риска СВСГР у пациента, выполняет суммирование весовых коэффициентов присутствующих факторов риска. Весовые коэффициенты получены с помощью корреляционного анализа. Блок-схема укрупненного алгоритма работы модуля представлена на рисунке 9;

- модуль определения высокой степени риска СВСГР с использованием НС. Выполняется моделирование работы НС, на входы которой подаются закодированные факторы риска, на выходе получаем закодированный результат, интерпретация которого может быть такой: высокая или низкая степень риска СВСГР;
- модуль определения степени риска СВСГР, который использует классификационное дерево, полученное на основе ГП. Выполняется расчет значения бинарного дерева. Единичное значение на выходе интерпретируется как высокая степень риска СВСГР, нулевое – низкая;
- модуль определения степени риска СВСГР, который использует классификационные правила, полученные на основе ГП.
- модуль определения степени риска СВСГР, который использует классификационные правила с возможностью работы в условиях некоторых неопределенных состояний. Классификационные правила получены с использованием ГП.

На рисунке 10 представлена структурная схема взаимодействия модулей экспертной системы. На рисунке 11. представлена укрупненная блок-схема алгоритма, реализующего определение степени риска СВСГР.

Все программное обеспечение разработано под операционную систему (ОС) Windows XP или другую ОС семейства Windows. Основные модули экспертной системы реализованы с использованием среды Borland C++ Builder 6.

В пятом разделе проведены экспериментальные исследования функционирования системы на реальных медицинских данных.

При тестировании получены следующие результаты: классификационная таблица - правильные результаты составляют 92,5%.

Проведены эксперименты по выбору НС. В результате выбрана НС, которая состоит из входного слоя (46 входов соответствуют выбранным 46

значимым факторам риска); скрытого слоя состоящего из 6 нейронов, с функцией активации – гиперболический тангенс; выходного слоя состоящего из 1 нейрона, с линейной функцией активации. Правильные результаты классификации составляют - 97,85%.

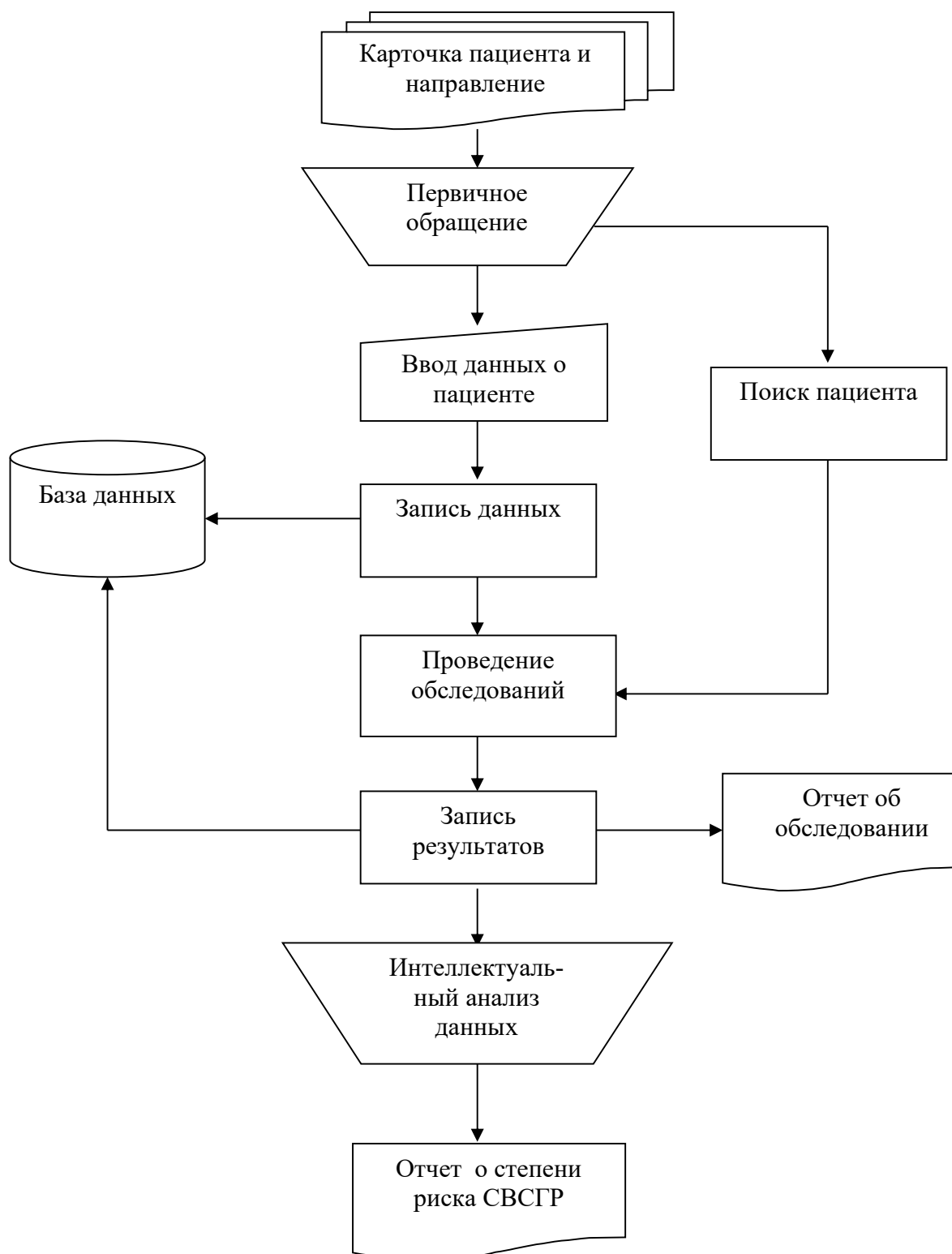


Рисунок 8. Схема обработки информации.

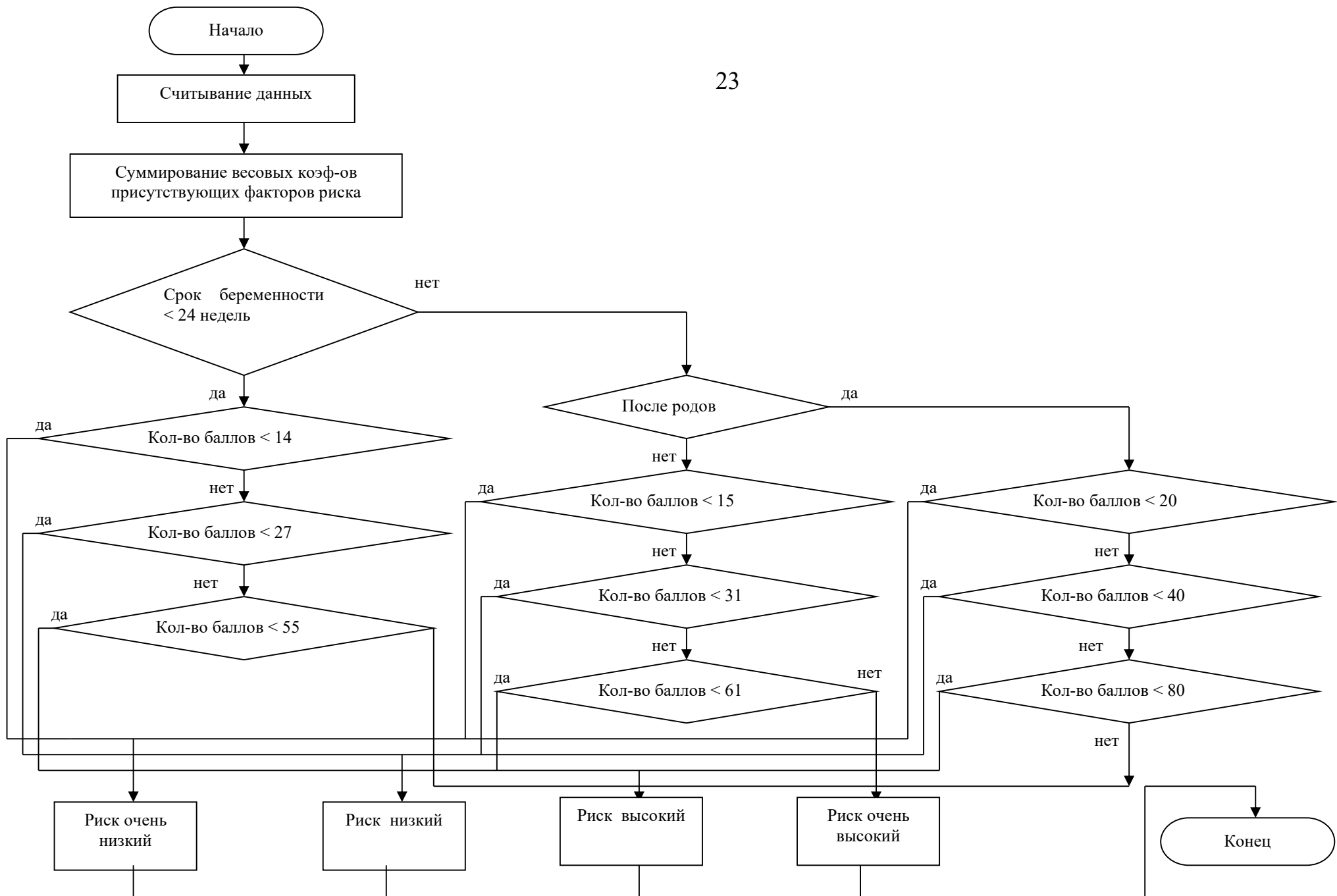


Рисунок 9. Блок-схема укрупненного алгоритма работы модуля определения степени риска СВСГР на основе прогностической таблицы

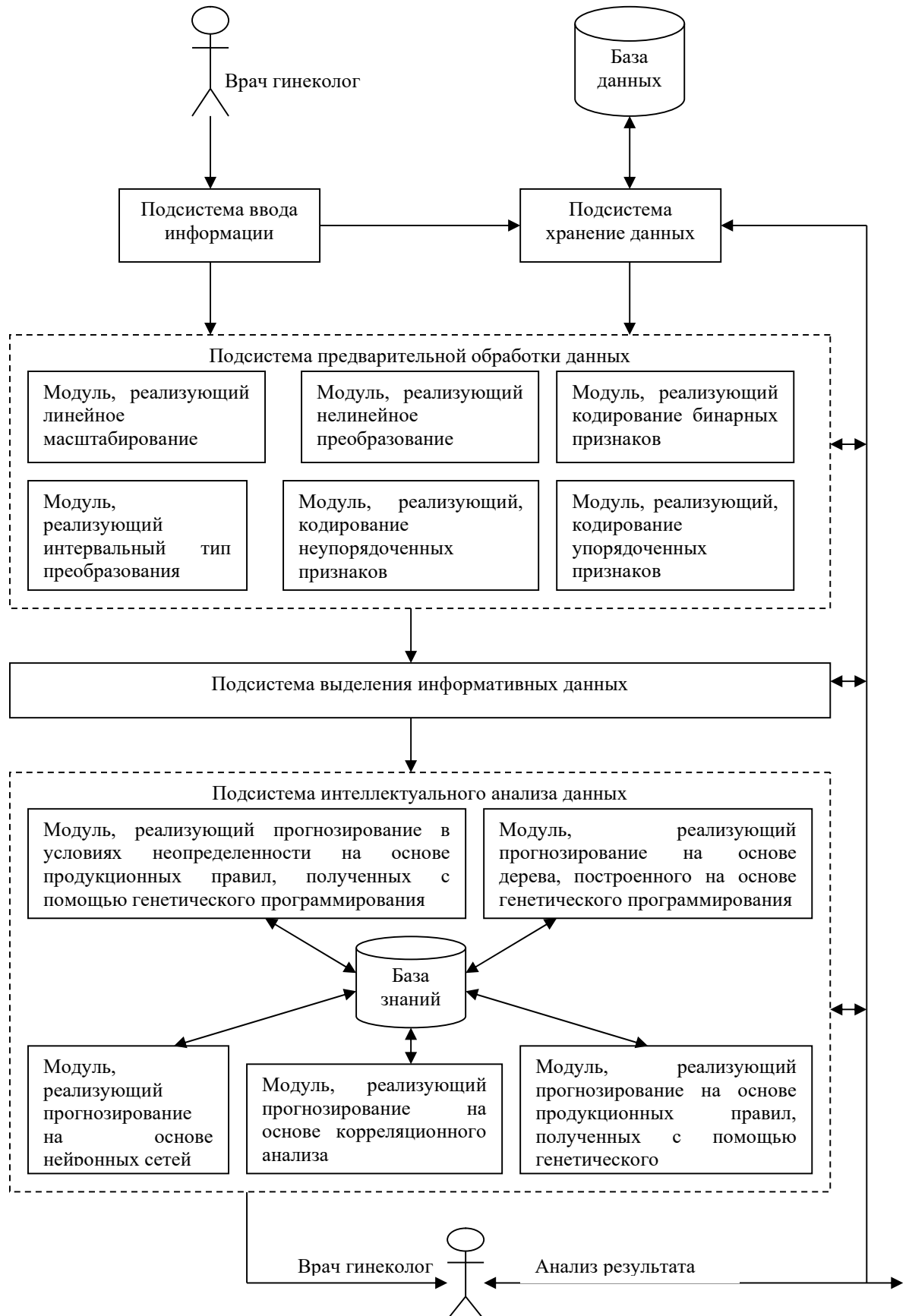


Рисунок 10. Структурная схема взаимодействия модулей экспертной системы.

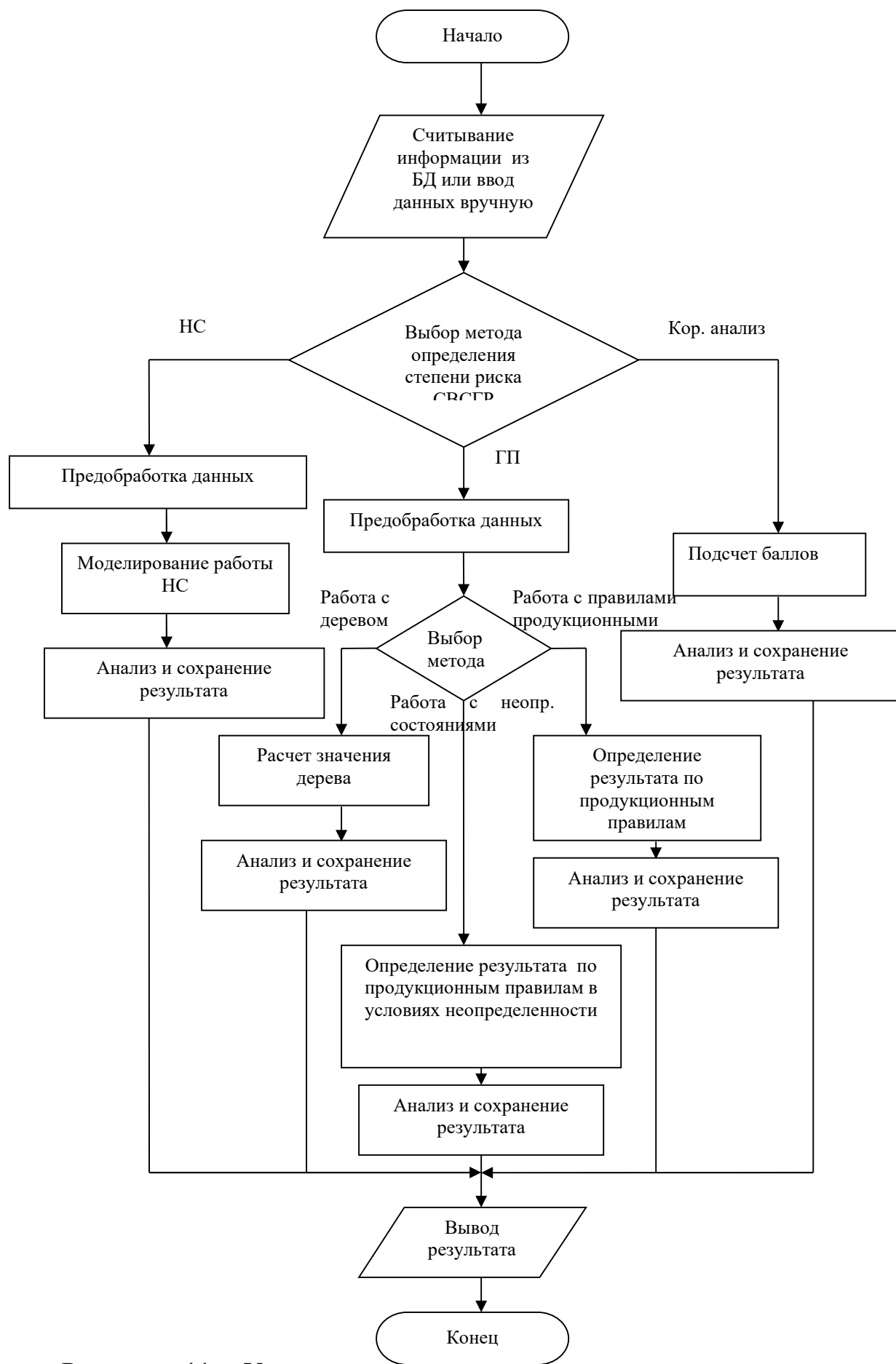


Рисунок 11. Укрупненная блок-схема алгоритма, реализующего функционирование ЭС определения степени риска СВSGP.

Проведены эксперименты по выбору параметров ГП, для получения классификационного дерева. Получено дерево, для определения степени риска СВСГР, которое на обучающей выборке дает 100% правильной классификации, на проверочной 98,57%.

Проведены эксперименты по выбору параметров ГП, для получения классификационных правил. Получены классификационные правила, которые на обучающей выборке дают 100% правильной классификации, на проверочной 97,14%.

Проведены эксперименты на полученном наборе правил при некоторых неопределенных входных состояниях в проверочной выборке. Получено более 90% правильной классификации на проверочной выборке с неопределенными состояниями до 5 %.

В заключении изложены наиболее важные научные и практические результаты, которые получены в диссертационной работе.

В приложениях приведены материалы, содержащие результаты экспериментов, проведенных в работе, а также акты внедрения основных результатов диссертационной работы.

ВЫВОДЫ

В диссертационной работе решена актуальная научная задача создания экспертной системы определения степени риска СВСГР. Разработанная экспертная система позволяет определять степень риска СВСГР на различных сроках беременности и сразу после рождения ребенка. При проведении исследований получены следующие основные результаты:

1. Впервые разработана экспертная система, предназначенная для определения степени риска СВСГР. Создана информационная база экспертной системы. Предложена структурная схема взаимодействия программных модулей. Разработаны алгоритмы и основные программные модули экспертной системы.

2. Разработана фитнес-функция для генетического алгоритма, реализующего выбор информативных факторов риска, что позволило достигнуть высокой эффективности выделения факторов риска в медицине за счет регулирования соотношения количества факторов и ошибки классификации. Экспериментально выбран оптимальный набор параметров для определения степени риска СВСГР.

3. Усовершенствован древовидный способ кодирования хромосомы для булевой функции, что позволило получать классификационные деревья для определения степени риска СВСГР. Определен набор параметров ГП, при котором достигается лучший результат и получено дерево для определения высокой степени риска СВСГР.

4. Усовершенствован способ кодирования хромосомы в виде дерева, реализующего булеву функцию в дизъюнктивной нормальной форме, что позволило получать классификационные правила для определения степени риска СВСГР. Экспериментально выбраны наиболее эффективные параметры ГП и получены классификационные правила для определения степени риска СВСГР.

5. Впервые предложена троичная логика в булевых вычислениях классификационного дерева, что позволило выполнять классификацию в условиях неопределенных значениях некоторых входных параметров. При неизвестных факторов риска до 5% в обучающих данных, удается получать до 95% правильной классификации на проверочной выборке.

6. Проведены испытания экспертной системы определения степени риска СВСГР в условиях городской больницы №3 города Донецка. Применение разработанной экспертной системы позволяет на различных сроках беременности определять степень риска СВСГР, в случае высокой степени риска применяются профилактические меры, направленные на снижение степени риска СВСГР.

7. Результаты работы использованы в городской больнице №3 города Донецка в госбюджетных темах Д-11-04, Н-3-07, Д-11-07, а также в учебном процессе кафедры «Автоматизированные системы управления» Донецкого национального технического университета.

СПИСОК ТРУДОВ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Skobtsov Y.A., Vasyaeva T.A., Rodin Y.V. Neural network prediction of aterotrombogenic strocke. // Advanced Computer Systems and Networks: Design and Application. Proceedings of 1-st International Conference ACSN-2003. 24-26 September, 2003, Lviv, Ukraine
2. Скобцов Ю.О., Васяева Т.О. Нейромережевий підхід до прогнозу результату хвороби Годжкина. // Наукові праці Донецького державного технічного університету. Серія: Обчислювальна техніка та автоматизація. Випуск 64.—Донецьк: РВА ДонДТУ.- 2003.
3. Чайка В.К., Яковлева Е.Б., Тутов С.М. Васяева Т.О. Спосіб прогнозування розвитку синдрому раптової смерті грудних дітей. // Патент України, МПК7: А61В17/42. – Заява № 4200500814 від 31.01.2005. – Опубл. 15.08.2005. – Бюл. № 8.
4. Чайка В.К., Яковлева Е.Б., Тутов С.М. Васяева Т.О. Спосіб прогнозування розвитку синдрому раптової смерті грудних дітей. // Патент України, МПК7: А61В17/42. – Заява № 4200500807 від 31.01.2005. – Опубл.19.09.2005.—Бюл. №9.
5. Васяева Т.А. Визначення групи ризику щодо синдрому раптової смерті немовляти. // Наукові праці Донецького державного технічного університету. Серія: Обчислювальна техніка та автоматизація, випуск 88. - Донецьк: ДонНТУ, 2005.
6. Васяева Т.А., Скобцов Ю.А. Выделение информативных факторов риска синдрома внезапной смерти грудных детей для формирования

- обучающей выборки при построении экспертной системы. // Вісник Хмельницького національного університету №2 Том 1 (90) Технічні науки. Хмельницький 2007.
7. Скобцов Ю.А., Васяева Т.А. Подготовка данных при разработке медицинских экспертных систем. // Вестник Херсонского национального технического университета №4(27) -2007г
 8. Скобцов Ю.А., Скобцов В.Ю., Васяева Т.А. Применение искусственного интеллекта для вычисления информативности признаков в медицинских задачах. //КОМПЬЮТЕРНЫЕ НАУКИ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ Тезисы докладов Международной научной конференции, посвященной памяти профессора А.М.Богомолова 2-4 июля 2007г., Саратов 110-112с.
 9. Васяева Т.А. Предобработка входной информации для построения и обучения экспертной системы прогнозирования синдрома внезапной смерти грудных детей.// Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація. Випуск 13(121). –Донецьк: ДонНТУ.- 2007. 118-125с.
 - 10.Скобцов Ю.А., Васяева Т.А., Лобачева М.В. Формирование знаний для медицинских экспертных систем на основе генетического программирования.// Информационные технологии и информационная безопасность в науке, технике и образовании "ИНФОТЕХ - 2007" Материалы международной научно-практической конференции Часть 2. г. Севастополь, 10-16 сентября 2007г. 38-42с
 - 11.Skobtsov Y.A., Vasyaeva T.A. Diagnosis of SIDS using Genetic Programming. // Advanced Computer Systems and Networks: Design and Application. Proceedings of 3-st International Conference ACSN-2007. 20-22 September, 2007, Lviv, Ukraine 92-93с.
 - 12.Скобцов Ю.А., Васяева Т.А. Извлечение знаний на основе генетических алгоритмов и генетического программирования. //

- ВІСНИК Східноукраїнського національного університету ім.. Даля.-№ 12 (130), частина 2.-2008.- с.5-11.
- 13.Скобцов Ю.А., Васяева Т.А. Эволюционный подход к формированию знаний для медицинских экспертных систем с учетом неопределенности данных. // Искусственный интеллект. -№ 4.-2008.- С.631-637.
- 14.Скобцов Ю.А., Васяева Т.А. Разработка экспертных систем медицинской диагностики с явным представлением продукционных правил на основе генетического программирования с учетом неопределенности данных. // Материалы IX международной научно технической конференции «Искусственный интеллект. Интеллектуальные системы –ИИ-2008».-Донецк:ИПИИ «Наука і освіта».-2008.-Т.2.-С.97-101.
- 15.Скобцов Ю.А., Васяева Т.А. Разработка экспертных систем медицинской диагностики с явным представлением продукционных правил на основе ГП. // Тези міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI'2008)» Херсон: ХНТУ, 2008. – Т3, Ч1.-68-71с.
- 16.Васяева Т.А. Получение правил вывода на основе генетического программирования. // Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: Матеріали II Всеукраїнської науково-практичної конференції. 8-10 квітня 2008р.,м. Луганськ. – Луганськ: Альма-матер, 2008. 19-21с.
- 17.Скобцов Ю.А., Васяева Т.А. Методы разработки экспертных систем медицинской диагностики. // Информационные технологии и информационная безопасность в науке, технике и образовании "ИНФОТЕХ - 2009" Материалы международной научно-практической конференции г. Севастополь, 7-12 сентября 2009г. 237-240с.

АННОТАЦИЯ

Васяева Т.А. Нейросетевые и эволюционные методы разработки экспертных систем медицинской диагностики (на примере синдрома внезапной смерти грудного ребенка) – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – «Информационные технологии». – Донецкий национальный технический университет, Донецк, 2009.

В диссертационной работе решена актуальная научная задача создания экспертной системы определения степени риска СВСГР.

Впервые разработана ЭС предназначенная для определения степени риска СВСГР, что позволило определять степень риска еще на этапе беременности и сразу после рождения ребенка.

Разработана фитнес-функция для генетического алгоритма, реализующего выбор информативных факторов риска, что позволило достигнуть высокой эффективности выделения информативных факторов риска в медицине за счет регулирования соотношения количества факторов и ошибки классификации.

Усовершенствован древовидный способ кодирования хромосомы для булевой функции, что позволило получать классификационные деревья для определения степени риска СВСГР.

Усовершенствован способ кодирования хромосомы в виде дерева, реализующего булеву функцию в дизъюнктивной нормальной форме, что позволило получать классификационные правила для определения степени риска СВСГР.

Впервые предложена троичная логика в булевых вычислениях классификационного дерева, что позволило выполнять классификацию в условиях неопределенных значениях некоторых входных параметров.

Проведены испытания экспертной системы определения степени риска

СВСГР в условиях городской больницы №3 города Донецка. Применение разработанной экспертной системы позволяет на различных сроках беременности определять степень риска СВСГР, в случае высокой степени риска применяются профилактические меры, направленные на снижение степени риска СВСГР.

Результаты работы использованы в городской больнице №3 города Донецка, в госбюджетной теме Д-11-04 «Разработка и исследования нейросетевых и эволюционных методов построения систем принятия решения» применены алгоритмы и программное обеспечение системы прогнозирования атеротромбогенного инсульта на базе нейронных сетей. В госбюджетной теме Н-3-07 «Разработка научных основ построения компьютерных систем технической медицинской диагностики» применены алгоритмы и программное обеспечение системы определения степени риска СВСГР на основе ГП. В госбюджетной теме Д-11-07 «Разработка и исследование эволюционных и нейросетевых моделей, методов и алгоритмов в системах диагностики и прогнозирования» применены алгоритмы и программное обеспечение системы определения степени риска СВСГР на основе ГП в условиях неопределенных некоторых входных состояний, а также в учебном процессе кафедры «Автоматизированные системы управления» Донецкого национального технического университета.

Ключевые слова: экспертная система, нейронные сети, генетические алгоритмы, генетическое программирование.

АНОТАЦІЯ

Васяєва Т.О. Нейромережеві та еволюційні методи розробки експертних систем медичної діагностики (на прикладі синдрому раптової смерті немовляти)– Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за фахом 05.13.06 – «Інформаційні технології». – Донецький національний технічний університет, Донецьк, 2009.

У дисертаційній роботі вирішено актуальне наукове завдання створення експертної системи (ЕС) визначення ступеню ризику синдрому раптової смерті немовляти (СРСН).

Вперше розроблено ЕС, що призначена для визначення ступеню ризику СРСН, що дозволило визначати ступень ризику на різних етапах вагітності і починати профілактичні заходи ще до народження дитини.

Розроблено фітнес-функцію для генетичного алгоритму, що реалізовує вибір інформативних факторів ризику, що дозволило досягти високої ефективності виділення інформативних чинників ризику в медицині за рахунок регулювання співвідношення кількості чинників і помилки класифікації.

Вдосконалений деревовидний спосіб кодування хромосоми для булевої функції, що дозволило отримувати класифікаційні дерева для визначення ступеню ризику СВСГР.

Вдосконалений спосіб кодування хромосоми у вигляді дерева, що реалізовує булеву функцію в диз'юнктивній нормальній формі, що дозволило отримувати класифікаційні правила для визначення ступеню ризику СВСГР.

Вперше запропонована троїчна логіка в булевих обчисленнях класифікаційного дерева, що дозволило виконувати класифікацію при невизначених значеннях деяких вхідних параметрів.

Проведені випробування ЕС визначення ступеню ризику СРСН в умовах міської лікарні №3 міста Донецька. Впровадження розробленої експертної системи дозволяє на різних термінах вагітності визначати ступень ризику СРСН, при встановленні високого ступеню ризику застосовуються профілактичні заходи, направлені на зниження ступеню ризику СРСН.

Ключові слова: експертна система, нейроні мережі, генетичні

алгоритми, генетичне програмування.

SUMMARY

Vasyaeva T.O. Neural network and evolutionary methods for expert system for medical diagnostics development. (on example of syndrome of sudden infant death) – Manuscript.

Candidate's of technical sciences thesis on specialty 05.13.06 – “Information technologies” – Donetsk National Technical University, Donetsk, 2009.

In the thesis an important scientific problem of an expert system for finding out SIDS risk range development is solved. The expert system developed allows increasing preciseness of finding out SIDS risk range by using methods, algorithms, software, and modern computer technologies.

For the first time the expert system for predicting a high SIDS risk range has been developed. That allowed finding out the risk range on an earlier stage, and beginning actions to decrease the risk range before an infant had been born.

A fitness function for the GA has been designed. That allowed reaching a high performance of extractions of valuable factors for medical parameters.

The predicting method based on genetic programming has been developed further, that allowed getting a tree that was able to detect a high SIDS risk range.

The predicting method based on genetic programming has been developed further, that allowed getting production rules for a high SIDS risk range prediction when some risk factors are unknown.

Testing of the expert system has been conducted in the Donetsk city hospital #3.

It is established, that using of this expert system increases accuracy of finding out of a SIDS risk range and supports decisions concerning preventive actions.

The results of this thesis have been used in the Donetsk city hospital #3, in

the state scientific research Д-11-04 «Development and researches of neural network and evolutionary methods for decisions support systems» the algorithms and the software of the system of prediction aterotrombogenic strocke based on neural networks were used. In the state scientific research H-3-07 “Development of scientific ground for computer system of medical and technical diagnostics building” the algorithms and the software of the system of predicting a high SIDS risk range were used. In the state scientific research Д-11-07 “Development and research of evolutionary and neural network models, methods and algorithms in the system of diagnostics and prediction” the algorithms and the software of the system of predicting a high SIDS risk range under vagueness of some input states were used. Also these methods were used in studying process of the chair “Automated control system” of the Donetsk National Technical University.

Keywords: expert system, neural networks, genetic algorithms, genetic programming