

УДК 004.048:004.622

©2011. Т. А. Васяева, С. В. Хмелевой, Д. Е. Иванов

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ ВЫСОКОГО РИСКА ПАТОЛОГИЧЕСКОЙ ПОТЕРИ КРОВИ ПРИ РОДАХ

В данной работе предлагается решение актуальной задачи определения риска возникновения патологического акушерского кровотечения. Применение генетического программирования (ГП) позволяет выполнять классификацию на патологическое и допустимое кровотечение. Предлагается модификация способа кодирования особей для генетического программирования. Выполнено тестирование на реальных медицинских данных.

Ключевые слова: генетическое программирование, классификация, кодирование.

Введение. Разработка и внедрение медицинских компьютерных аналитических систем является одной из наиболее актуальных задач предметной области информационных технологий. В данной работе предлагается решение актуальной задачи определения риска возникновения патологического кровотечения при родах [1]. Для сравнения ежедневно в мире от осложнений, связанных с беременностью и родами, умирает около 1500 женщин. Причем 80% смертей обусловлено прямыми причинами: кровотечения, инфекции, гестозы, патологические роды. Лидирует в этом списке смертность в результате массивной кровопотери. При этом если даже удастся предотвратить гибель женщины, то следствием перенесенного критического состояния является развитие какой-либо тяжелой хронической патологии – вплоть до инвалидности. Таким образом, данная проблема носит не только медицинский, но и выраженный социальный характер. По оценкам экспертов большинство этих случаев можно было предотвратить. Для практики важно определять риск возможной потери крови при родах более чем 0,5% от массы тела, с целью принятия соответствующих мер.

1. Постановка задачи. При решении задачи прогнозирования необходимо определить функцию вида (1)

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

где Y – исследуемая величина, зависящая от факторов $X = (x_1, x_2, \dots, x_n)$.

В нашем же случае важно определить риск потери крови при родах более чем 0,5% от массы тела, с целью принятия соответствующих мер. По сути, необходимо выполнить разбиение на два класса: патологическое кровотечение и допустимое. То есть наша целевая переменная принимает дискретные значения, и в контексте нашей задачи рассматривается классификация.

Итак, Y – кровотечение патологическое или допустимое, x_1, x_2, \dots, x_n – факторы риска развития патологической потери крови при родах, функция (1) определяет тип кровотечения при определенных значениях факторов x_1, x_2, \dots, x_n . Значение функции равно единице означает высокую степень риска патологического акушер-

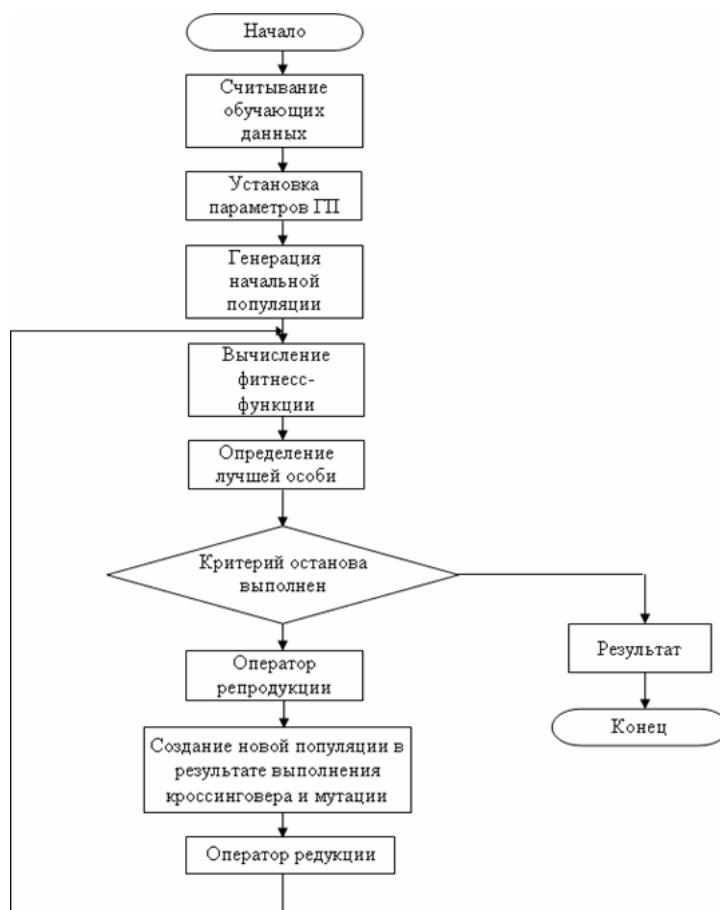


Рис. 1. Обобщенный алгоритм работы ГП

ского кровотока, а нулю – соответственно низкую. Задача решается методом, который предполагает обучение с учителем. При обучении с учителем необходимо сформировать обучающую выборку. В нашем случае обучающая выборка представляет собой реальные медицинские данные, полученные при обследовании беременных женщин.

Так как решается задача классификации на два класса, то для ее решения можно использовать математический аппарат булевых функций в виде древообразных структур [2]. При этом листьями дерева являются факторы риска, которые приведены к бинарным значениям, функциональными узлами – логические функции (И, ИЛИ, НЕ), а корень дерева определяет значение булевой функции, где единица соответствует патологическому кровотечению, а ноль – допустимому. После этапа обучения полученное дерево, реализующее определенную булеву функцию, будет выполнять классификацию на допустимую или патологическую потерю крови при родах.

В качестве ошибки классификации будем использовать долю правильной клас-

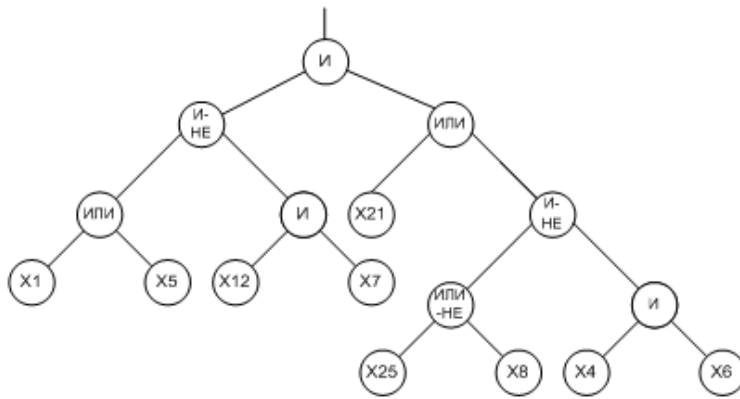


Рис. 2. Пример структуры дерева для классификации акушерских кровотечений

сификации (2):

$$E = \frac{1}{M} \cdot \left(M - \sum_{i=1}^M (|F_i - Y_i|) \right), \quad (2)$$

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

2. Разработка алгоритма. Исходные данные должны быть предварительно обработаны, основное назначение предобработки – преобразовать входное обучающее множество в булевозначный вектор. Далее определяются терминальное и функциональное множества, фитнес-функция, значения параметров эволюционного алгоритма. Решением задачи будет булева функция, представленная в виде дерева, которое можно использовать для классификации акушерских кровотечений. Общий алгоритм представлен на рис. 1.

В контексте нашей задачи ГП используется для получения классификационного дерева (пример структуры показан рис. 2). Дереву на рис. 2 соответствует булева функция (3):

$$F(x_1, x_2, \dots, x_{176}) = \overline{(x_1 \vee x_5) \cdot (x_{12} \cdot x_7)} \cdot (x_{21} \vee \overline{(x_{25} \vee x_8)} \cdot (x_4 \cdot x_6)). \quad (3)$$

Терминальное множество в данном случае составляют факторы риска развития патологического акушерского кровотечения, которые после предобработки представляют собой булевозначный вектор.

Основными логическими операциями являются: ИЛИ, И, НЕ. Все эти операции всегда имеют один выход, но могут иметь разное количество входов. Так как первые две функции могут иметь два и более входа, а последняя только один, то для более удобной программной реализации выполним замену операции НЕ на следующие две: ИЛИ-НЕ и И-НЕ. Такая замена позволит унифицировать количество входов для всех операций (их будет всегда 2). Таким образом, функциональное множество составляют 4 логические функции ИЛИ, И, ИЛИ-НЕ и И-НЕ. К тому же, такой

Таблица 1. Параметры эволюционного алгоритма

Параметр	Допустимые значения
Мощность популяции	Задается
Максимальная глубина особи	Задается
Метод генерации начальной популяции	– растущая $p_g = 1$ – полная $p_c = 1$ – смешенная $p_g + p_c = 0,5 + 0,5 = 1$
Вероятность функционального узла	0,5
Вероятность терминального узла	0,5
Вероятность кроссинговера	Задается
Вероятность мутации	Задается
Отбор родителей	– рулетка

логический базис позволяет избавиться от так называемых интронов (бесполезных участков), в данном случае – от двойного отрицания ($\text{НЕ}(\text{НЕ}(X))$).

В качестве фитнес-функции используется доля пациентов с правильно определенным акушерским кровотечением (2), в которой переменные Y (желаемый результат классификации) и F (полученный результат классификации) принимают булевы значения 0 или 1. Единица соответствует высокой степени риска развития патологической потери крови при родах, а ноль – низкой степени риска развития патологической потери крови при родах.

Для получения классификационного дерева необходимо выполнить следующие этапы:

1. Установить параметры ГП, которые представлены в табл. 1.
2. Сгенерировать начальную популяцию. Популяция состоит из некоторого количества хромосом (особей). Каждая особь соответствует определенному дереву, которое представляет собой решение, например, как на рис. 2. Особь (дерево) на начальном этапе генерируется случайным образом и состоит из функциональных и терминальных узлов.
3. По значению фитнес-функции (2) оценить все особи, входящие в популяцию.
4. Выполнить генетические операторы.
5. Проверить условие останова. Если критерий останова выполнен, то перейти на шаг 6, иначе на шаг 3.
6. Выбрать лучшую особь (с максимальным значением фитнес-функции), которая будет результатом работы ГП, а соответствующее ей дерево – решением поставленной задачи (классификационным деревом, выполняющим классификацию на патологическое акушерское кровотечение и допустимое).
7. В процессе поиска решений можно использовать различные значения параметров ГП.

3. Генерация начальной популяции. На данном этапе происходит генерация начальной популяции в соответствии с заданными параметрами. Популяция состо-

ит из набора деревьев, сгенерированных случайным образом. Генерация каждого дерева происходит рекурсивно, начиная с генерации функционального узла и его аргументов. Для каждого дочернего узла (аргумента) случайным образом определяется, будет данный узел функциональным или терминальным, далее в соответствии с типом узла выбирается его значение из соответствующего терминального или функционального множества. Процесс выполняется по левой ветви до тех пор, пока не будет выбран дочерним терминальный узел. Затем генерируются правые ветви.

Предусмотрены следующие методы создания деревьев [2]:

1. **Полный.** При генерации случайного дерева каждая ветвь имеет одинаковую (максимальную) глубину.

2. **Растущий.** При генерации случайного дерева каждая ветвь может иметь различную глубину, но не более чем заданная максимальная.

3. **Комбинированный.** Половина деревьев всей популяции генерируется полным методом, вторая половина – растущим.

4. Применение генетических операций.

Отбор родителей [2-3]. Важнейшим фактором, который влияет на эффективность ГП, является оператор отбора. Слепое следование принципу "выживает сильнейший" может привести к сужению пространства области поиска решения и попаданию найденного решения в область локального экстремума фитнес-функции. С другой стороны, достаточно слабый оператор отбора может привести к медленному росту качества популяции, что приведет к замедлению поиска решения. Кроме того, популяция при таком подходе может не только не улучшаться, но и ухудшаться. Использован отбор пропорционально значению фитнес-функции, реализованный методом рулетки.

Кроссинговер. Для древообразной формы представления используются следующие три основных типа оператора кроссинговера (ОК) [2, 4-5]:

- узловой кроссинговер;
- кроссинговер поддеревьев;
- смешанный.

Узловой оператор кроссинговера выполняется следующим образом:

– выбираются два дерева и узлы в этих деревьях. Так как узлы в деревьях могут быть разного типа, сначала необходимо убедиться, что выбранные узлы у родителей являются взаимозаменяемыми. Если тип узла во втором родителе отличен от типа узла первого родителя, то во втором родителе случайным образом выбирается другой узел, после чего опять выполняется проверка на предмет совместимости;

- производится обмен выбранных узлов.

Пример узлового ОК показан на рис. 3.

В кроссинговере поддеревьев родители обмениваются не самими узлами, а соответствующими им поддеревьями. Такой ОК выполняется следующим образом:

– выбираются родители и узлы в них. Далее необходимо убедиться, что выбранные узлы принадлежат одному типу (функциональному или терминальному). Если

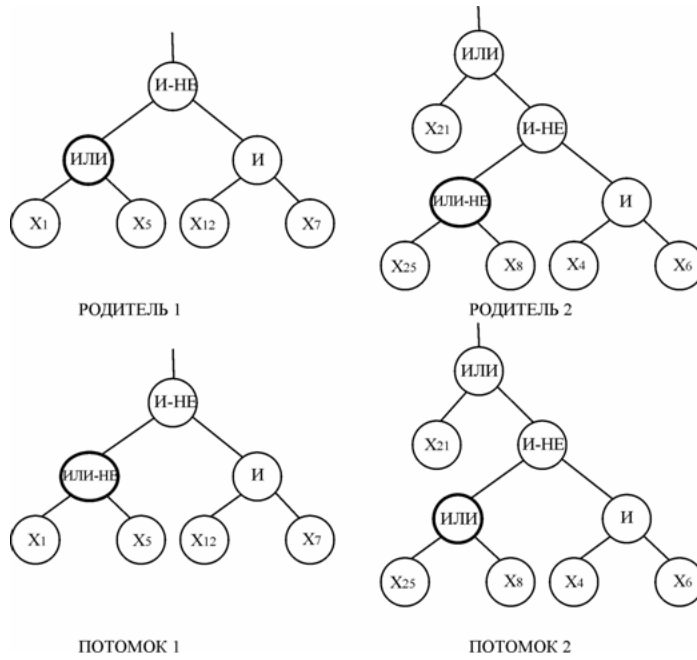


Рис. 3. Пример выполнения узлового ОК

это условие не выполняется, то, как и в предыдущем случае, во втором дереве выбирается другой узел с последующей проверкой на взаимозаменяемость;

– производится обмен поддеревьев, которые соответствуют этим выбранным узлам. Вычисляется глубина дерева для каждого потомка. Если ожидаемый размер не превышает заданный порог, то потомки запоминаются.

Обмен поддеревьями возможен и в одном родителе. Пример ОК поддеревьев показан на рис. 4.

При смешанном операторе кроссинговера для одних особей выполняется узловой оператор кроссинговера, а для других – кроссинговер поддеревьев.

Мутация. Для деревьев применимы следующие операторы мутации (ОМ) [2, 4-5]:

- узловая мутация;
- усекающая мутация;
- растущая мутация.

Узловой оператор мутации выполняется следующим образом:

- выбирается случайным образом узел, подлежащий мутации;
- определяется тип выбранного узла (функциональный или терминальный);
- случайно выбирается из множества, соответствующего типу выбранного узла,

значение узла, отличное от рассматриваемого;

- заменяется исходное значение узла на выбранное.

Пример узлового оператора мутации показан на рис. 5.

Усекающий оператор мутации производится так:

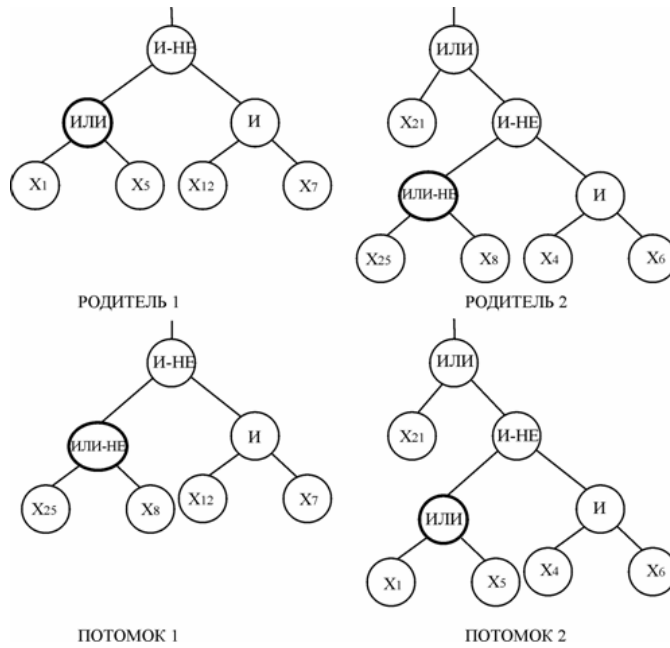


Рис. 4. Пример выполнения ОК поддеревьев

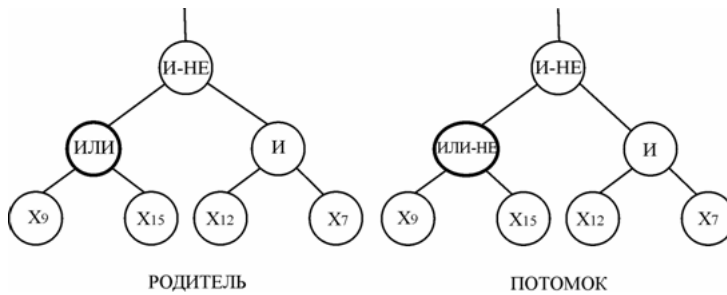


Рис. 5. Пример выполнения узлового ОМ

- случайным образом выбирается узел;
- случайным образом выбирается символ из заданного терминального множества;

- обрезаются ветвь выбранного узла мутации;
- вместо обрезанной ветви ставится терминальный символ.

Пример усекающего оператора мутации показан на рис. 6.

Растущий оператор мутации выполняется следующим образом:

- случайным образом определяется узел мутации;
- если узел не является терминальным, то необходимо отсечь ветви, которые исходят из него, иначе выбрать другой узел;
- вычислить глубину остатка дерева;
- вместо отсеченной ветви дерева вырастить случайным образом новую ветвь

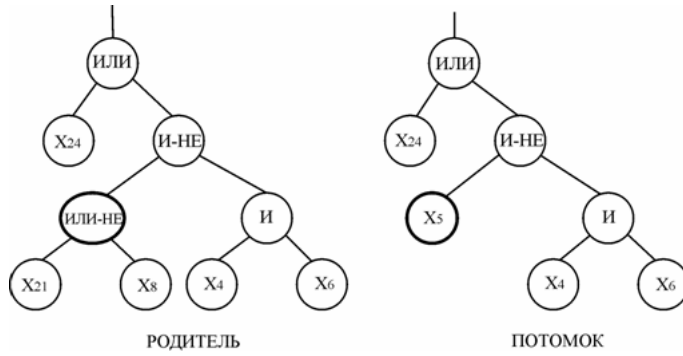


Рис. 6. Пример выполнения усекающего ОМ

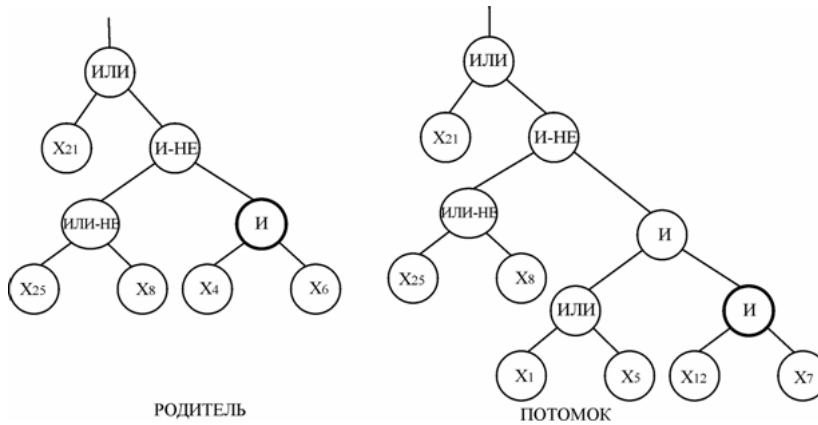


Рис. 7. Пример выполнения растущего ОМ

так, чтобы размер нового построенного дерева не превышал заданный порог.

Пример растущего оператора мутации показан на рис. 7. Это очень мощный оператор, обладающий большими возможностями.

Редукция [2-3]. Оператор редукции выполняется с целью сохранения размера популяции. Типы оператора практически совпадают с типами оператора отбора родителей. Но процедуры редукции и отбора родителей разнесены по действию во времени и имеют разный смысл. Возможно выполнение следующих вариантов редукции:

- элитная стратегия;
- чистая замена;
- равномерная случайная замена (с указанием количества заменяемых особей в процентах).

Критерий останова. В качестве критерия останова можно использовать указание определенного числа итераций или указание определенного числа повторения лучшего результата.

Апробация. Для реализации поставленной задачи написана программа в сре-

Таблица 2. Результаты экспериментов по выбору вероятности ОМ

Вероятность ОМ	Доля правильной классификации					
	Обучающая выборка			Проверочная выборка		
	Узловой	Усекающий	Растущий	Узловой	Усекающий	Растущий
0,025	0,93571	0,95	0,96428	0,86428	0,87142	0,95714
0,05	0,87142	0,95	0,96428	0,89285	0,98571	0,97857
0,075	0,92857	0,96428	0,97142	0,88571	0,97857	0,95
0,1	0,95	0,95	0,96428	0,97857	0,97857	0,97857
0,125	0,94285	0,97142	0,97142	0,88571	0,97857	0,94285
0,15	0,96428	0,96428	0,98571	0,97857	0,97857	0,94285
0,175	0,97142	0,96428	0,96428	0,98571	0,97857	0,97857
0,2	0,96428	0,96428	0,96428	0,97857	0,97857	0,97857
0,225	0,97142	0,96428	0,97142	0,87857	0,97857	0,98571
0,25	0,97142	0,97142	0,96428	0,97142	0,93571	0,97142
0,275	0,96428	0,96428	0,97142	0,97857	0,97857	0,98571
0,3	0,97857	0,96428	0,99285	0,92857	0,98571	0,97142
0,325	0,96428	0,96428	0,98571	0,97857	0,98571	0,99285
0,35	0,97142	0,97857	0,99285	0,98571	0,9	0,97142
0,375	0,96428	0,97857	0,97142	0,97857	0,99285	0,97857
0,4	0,96428	0,96428	0,97142	0,97857	0,97142	0,97142
0,425	0,97857	0,96428	0,98571	0,97142	0,98571	0,91428
0,45	0,97142	0,96428	0,97142	0,91428	0,86428	0,95
0,475	0,97142	0,96428	0,99285	0,98571	0,97857	0,97142
0,5	0,96428	0,96428	0,97142	0,97857	0,97857	0,98571

де C++ Builder 6, которая выполняет рассмотренный алгоритм. Проведены эксперименты с целью определения значений параметров ГП для построения дерева классификации акушерских кровотечений. Например, для каждого значения вероятности ОМ выполнялось несколько экспериментов и выбиралось лучшее значение. Результаты экспериментов представлены в табл. 2.

На рис. 8 и 9 показаны зависимости доли правильной классификации от вероятности и типа ОМ.

Очевидно, что, начиная с вероятности ОМ 0,275, результаты классификации достаточно высокие для усекающего и растущего ОМ. Процент правильной классификации составляет более 96% на обучающей выборке и на проверочной, за некоторым исключением. Поэтому выберем усекающий ОМ и вероятность, равную 0,3. Выбор усекающего ОМ обусловлен лучшей классификацией на проверочной выборке, а также тем, что при данном типе ОМ деревья получаются меньшего размера, что упрощает работу с ними при их интерпретации.

Аналогичным образом подбирались и другие значения ГП. В результате некоторых экспериментов получено дерево, которое на обучающей выборке дает 100% правильной классификации, на проверочной 98,57%. В табл. 3 приведены значения параметров ГП, при которых оно получено.

5. Заключение. Разработанный аппарат ГП создан и протестирован на приме-

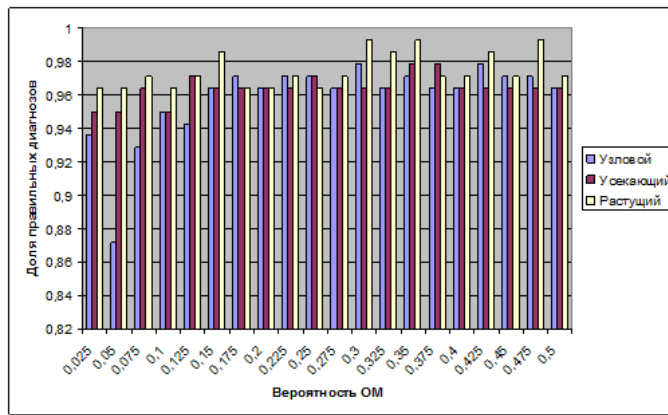


Рис. 8. Зависимость корректности акушерских кровотечений от вероятности и типа ОМ на обучающей выборке

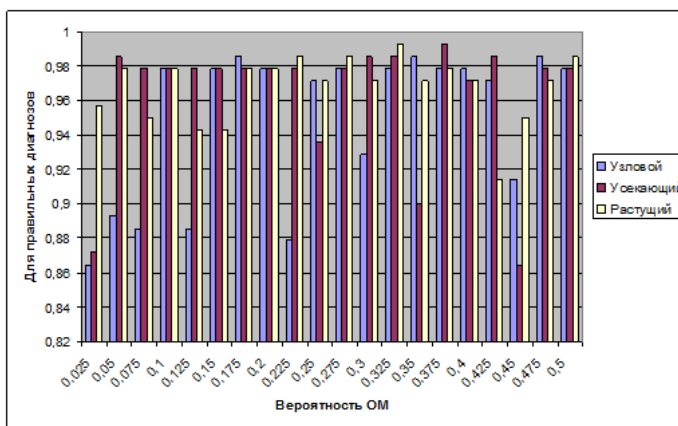


Рис. 9. Зависимость корректности классификации акушерских кровотечений от вероятности и типа ОМ на проверочной выборке

ре определения риска развития патологической потери крови при родах, но может быть использован и при решении других задач медицинской диагностики и прогнозирования.

1. Радзинский В.Е. Акушерский риск. Максимум информации минимум опасности для матери и младенца / В.Е. Радзинский, С.А. Князев, И.Н. Костин. – Изд.: Эксмо. – 2009. – С. 285.
2. Скобцов Ю.О. Основы эволюционных обчислень / Скобцов Ю.О. – Донецьк: ДонНТУ, 2009. – 316 с.
3. Курейчик В.М. Генетические алгоритмы. Состояние. Проблемы. Перспективы / Курейчик В.М. // Известия академии наук. Теория и системы управления. – 1999. – № 1. – С. 144-160.
4. Banzhaf W. Genetic Programming – an Introduction / Banzhaf W. et al. – Morgan Kaufman, Heidelberg; San-Francisco, 1998.
5. Koza J.R. Genetic Programming / Koza J.R. – Cambridge, MA MIT Press, 1998. – 609 с.

Таблица 3. Набор параметров ГП, при котором получено дерево для классификации

Параметр	Значения
Мощность популяции	150
Максимальная глубина особи	10
Отбор родителей (селекция)	Рулетка
Инициализация начальной популяции	Растущая
Редукция	Элитная стратегия
Вероятность кроссинговера	0,85
Вероятность мутации	0,30
Кроссинговер	Смешанный
Мутация	Усекающая
Количество итераций	200

T. A. Vasyaeva, S. V. Hmelevoy, D. E. Ivanov

Application of genetic programming to identify high-risk pathological blood loss at delivery.

A solution of the problem of determining the actual risk of pathological obstetric haemorrhage is proposed. The application of genetic programming allows performing the classification to the abnormal bleeding and valid one. A modification of the coding method of individuals for genetic programming is proposed. The testing on real medical data is performed.

Keywords: *genetic programming, classification, coding.*

T. O. Васяєва, С. В. Хмелевой, Д. Є. Іванов

Застосування генетичного програмування для визначення високого ризику патологічної втрати крові при пологах.

У даній роботі запропоновано вирішення актуальної задачі визначення ризику виникнення патологічної акушерської кровотечі. Застосування генетичного програмування дозволяє виконувати класифікацію на патологічну і допустиму кровотечу. Запропоновано модифікацію способу кодування особин для генетичного програмування. Виконано тестування на реальних медичних даних.

Ключові слова: *генетичне програмування, класифікація, кодування.*

Донецкий национальный технический ун-т
Ин-т прикл. математики и механики НАН Украины, Донецк

Получено 05.12.11

vasyaeva@gmail.com
hmelevoy@cs.donntu.edu.ua
ivanov@iamm.ac.donetsk.ua