

Ю. В. Марчук, М. А. Кожевников

Пакет средств информационно-программной поддержки раннего выявления риска развития ретинопатии у недоношенных детей

В статье приведены основные результаты исследований, проведенных на базе научно-практического центра «Бонум». Была поставлена задача разработать систему поддержки прогнозирования степени риска развития ретинопатии. В связи с низкой эффективностью частных подходов к решению этой задачи предлагается вариант адаптации методов многомерного анализа в рамках пакета средств поддержки прогнозирования, повышающий его точность.

В настоящее время актуальна проблема моделирования (в том числе математического) при решении теоретических и практических задач поддержки прогнозирования в медицине с применением информационно-интеллектуальных систем.

В медико-биологических науках математическое моделирование играет в большинстве случаев подчиненную роль в связи со сложностью объектов, процессов и явлений, вариабельностью их характеристик, наличием индивидуальных особенностей объектов исследования [1, 3, 7, 8, 11].

Теоретическим основам математического моделирования в медицине и его применению для решения актуальных практических задач, а также проблемам в области разработки медицинских информационных систем поддержки деятельности врачей и НИОКР в части интеллектуально-информационных систем посвящено большое количество работ отечественных и зарубежных авторов [2, 4, 5, 9, 10, 12].

От медицинских учреждений научно-практического типа все чаще поступает социальный заказ на выполнение работ по моделированию и реализации систем ин-

формационно-интеллектуальной поддержки прогнозирования.

Одним из тяжелых многофакторных заболеваний, возникающих у новорожденных с экстремально низкой массой тела (менее 2000 г), является ретинопатия¹ недоношенных. Несмотря на многочисленные исследования по выявлению факторов риска, влияющих на развитие ретинопатии, по разработке правил составления прогнозов и способов диагностики, необходимо создать пакет средств поддержки (алгоритмы, способы, программы) прогнозирования степени риска возникновения этого заболевания на раннем, доклиническом этапе развития патологии.

Материалы и методы

Объектом исследования являлись недоношенные дети с массой тела менее 2000 г и сроком гестации² менее 32 недель при рождении, поступившие на 1–2-е сутки жизни в реанимационно-консультативное отделение. Диагноз определялся на основе данных офтальмоскопии и результатов исследования на ретинальной камере. В соответствии с поставленным диагнозом новорожденные были подразделены на 3 груп-

¹ Ретинопатия — тяжелое заболевание глаз, проявляющееся в виде отслойки сетчатки с последующим развитием слепоты.

² Срок гестации — отрезок времени, в который происходит внутриутробное развитие плода до момента рождения.

пы: I — группа риска без ретинопатии, II — дети с ретинопатией I–II степени, III — дети с ретинопатией III–V степени.

Дети наблюдались и лечились в Областном центре ретинопатии недоношенных, расположенном на базе государственного учреждения здравоохранения Свердловской области — Детской клинической больницы восстановительного лечения Научно-практического центра (ГУЗ СО ДКБВЛ НПЦ) «Бонум».

Результаты и их обсуждение

Исследование было затруднено тем, что данные, характеризующие состояние новорожденных, находились в нескольких несогласованных источниках, при этом частично эти данные дублировались. Часть данных отсутствовала в предполагаемом месте их хранения, поэтому необходимо было обнаружить их в дублирующем хранилище. Отмечено, что взятые в отдельности признаки, характеризующие состояние пациента на доклиническом этапе развития патологии, не обладают достаточной информативностью. При использовании их в совокупности построить механизм поддержки прогнозирования можно лишь на основе применения специальной последовательности методов многомерного анализа.

Для разработки механизма поддержки прогнозирования был предложен алгоритм построения адаптированного средства (которое может включать как методические рекомендации, альбом алгоритмов, так и программное средство) определения уровня риска развития ретинопатии и степени ее тяжести у недоношенных детей (рис. 1).

Для получения необходимой информации о клинико-анамнестических, лабораторных и инструментальных данных из различных источников, а также для обеспечения их целостности, полноты и непротиворечивости разработан алгоритм синхронизации баз данных (рис. 2).

В ходе исследований выявлена характерная особенность объектов медицинской природы (понимается объект, описываемый

переменными, имеющими медицинское происхождение) — значительная внутригрупповая дисперсия, которая вносит существенный вклад в общую ошибку прогнозистического правила.

Для увеличения степени межгруппового различия и повышения концентрации объектов внутри группы разработан алгоритм усиления межгрупповых различий и повышения концентрации объектов внутри группы (рис. 3).

Благодаря реализации алгоритма усиления можно повысить чувствительность и специфичность механизма поддержки прогнозирования степени риска развития ретинопатии недоношенных. Несмотря на использование этого алгоритма, часть объектов попадает в пересечение множеств объектов, принадлежащих к разным группам (пограничная область, где невозможно однозначно определить принадлежность объекта к одной из групп, на границе которых он находится). Эта область названа нами областью спорных ситуаций (рис. 4). В центре окружностей расположены групповые центроиды, характеризующие типовые объекты своей группы. Окружность имеет радиус, равный расстоянию от наиболее удаленного объекта группы из обучающей выборки до своего центроида.

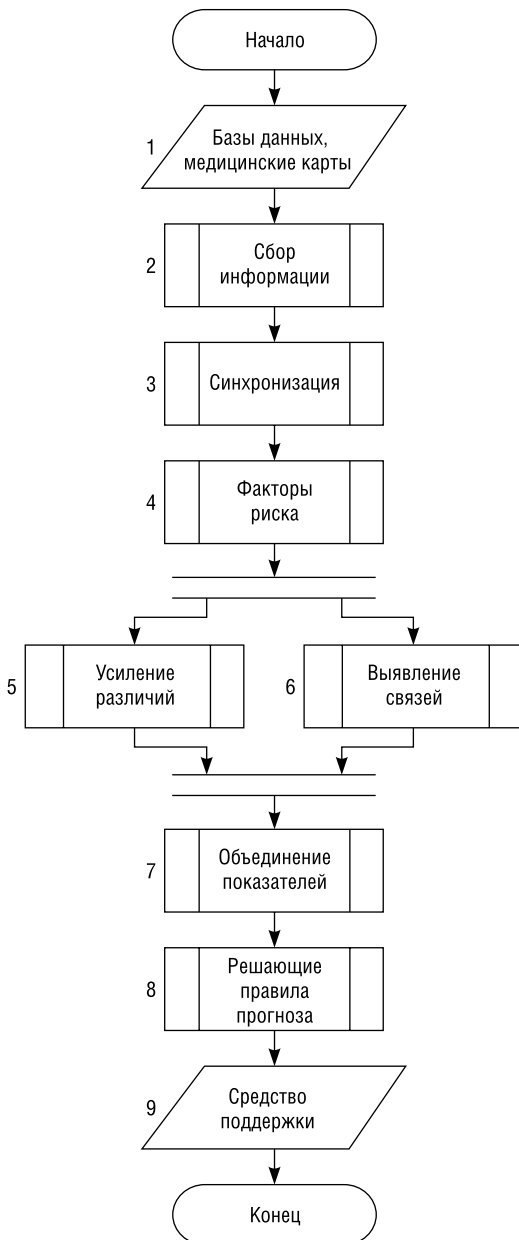
Авторы предлагают алгоритм выявления многосторонних спорных ситуаций в ответах решающего правила при определении принадлежности объекта (рис. 5), а также алгоритм их разрешения и получения однозначного ответа.

Для выявления приоритетного варианта ответа при неоднозначном ответе решающего правила предлагается использовать операционные характеристики прогностического правила для каждого варианта прогноза (чувствительность — Se , специфичность — Sp).

Для решения поставленной задачи используются множества значений чувствительности $\mathbf{Se} = \{Se_1, Se_2, \dots, Se_n\}$ и специфичности $\mathbf{Sp} = \{Sp_1, Sp_2, \dots, Sp_n\}$ решающего правила при определении каждого из выделенных заболеваний, где n — число альтернатив при работе решающего правила, также исполь-

зуется множество величин, характеризующих степень удаленности объекта от соответствующих центроидов групп $\mathbf{R} = \{R_1, R_2, \dots, R_n\}$ в абстрактном пространстве канонических линейных дискриминантных функций. Опре-

деление приоритетной альтернативы осуществляется с помощью двух вспомогательных множеств: множества рейтингов патологий $\mathbf{T} = \{T_1, T_2, \dots, T_n\}$ и антирейтингов патологий $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$.



- 1 — сведения, регистрируемые в базах данных специализированных медицинских учреждений;
- 2 — формирование запросов на предоставление информации о различных сторонах состояния объекта исследования;
- 3 — синхронизация собираемых сведений с помощью специализированного информационно-программного средства с обеспечением полноты и непротиворечивости информации;
- 4 — выявление обособленных факторов риска развития тяжелой патологии зрительного анализатора;
- 5 — усиление совокупной информативности каждой группы переменных (параметров) с формированием абстрактных группировочных переменных и правил идентификации их значений (применяется кластерный, факторный и дискриминантный анализы);
- 6 — итерационный процесс построения модели, отражающей связь факторов риска, параметров состояния и развития патологического процесса в организме;
- 7 — агломерация разрозненных групп переменных и абстрактных параметров с целью получения целостного массива обучающей информации;
- 8 — определение решающих правил прогноза степени риска возникновения и тяжести протекания ретинопатии у недоношенных (применяется дискриминантный анализ);
- 9 — ключевой механизм поддержки раннего выявления и прогнозирования степени тяжести патологического процесса в зрительном анализаторе.

Рис. 1. Алгоритм построения адаптированного средства информационно-программной поддержки прогнозирования уровня риска развития ретинопатии и степени ее тяжести у недоношенных детей

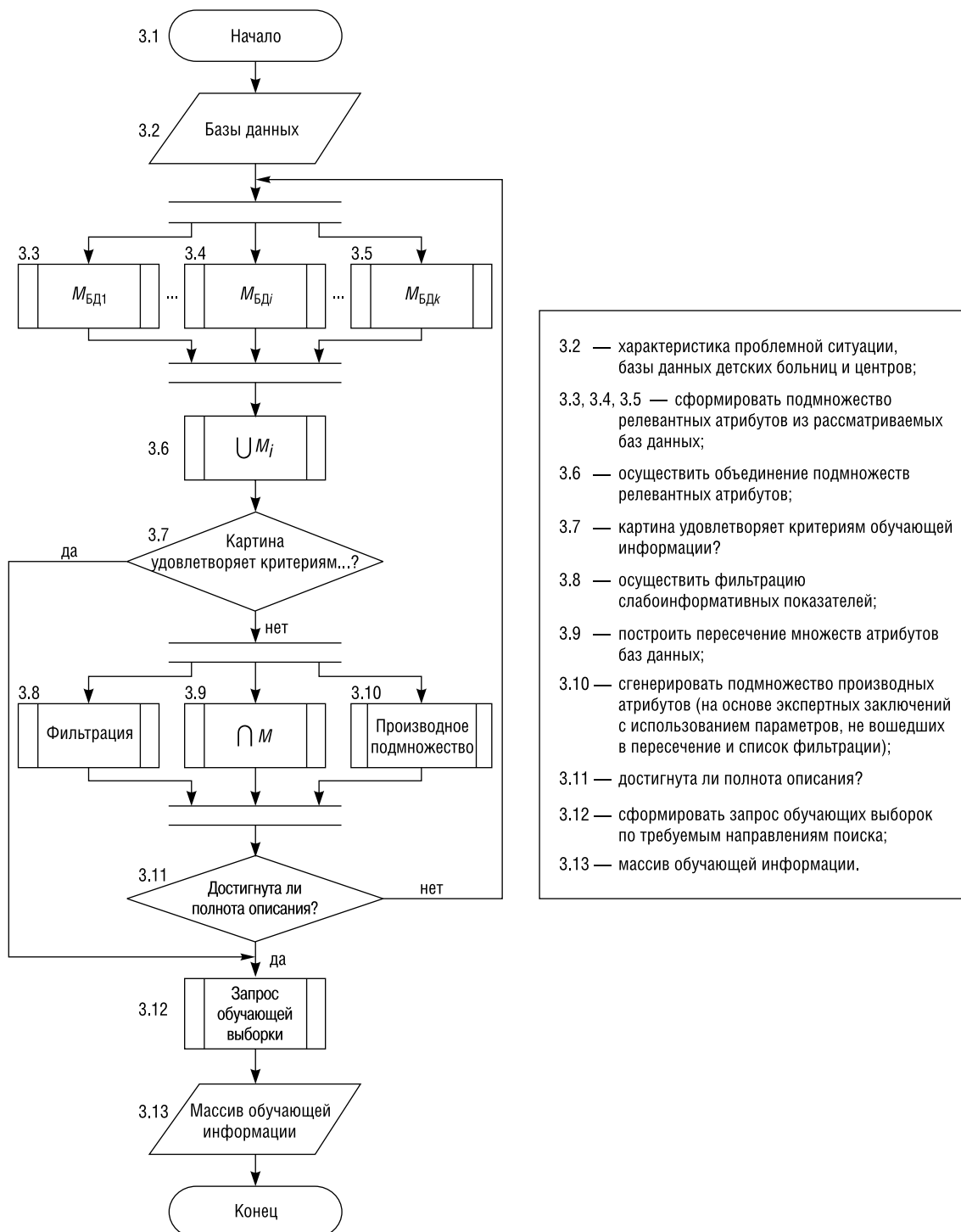
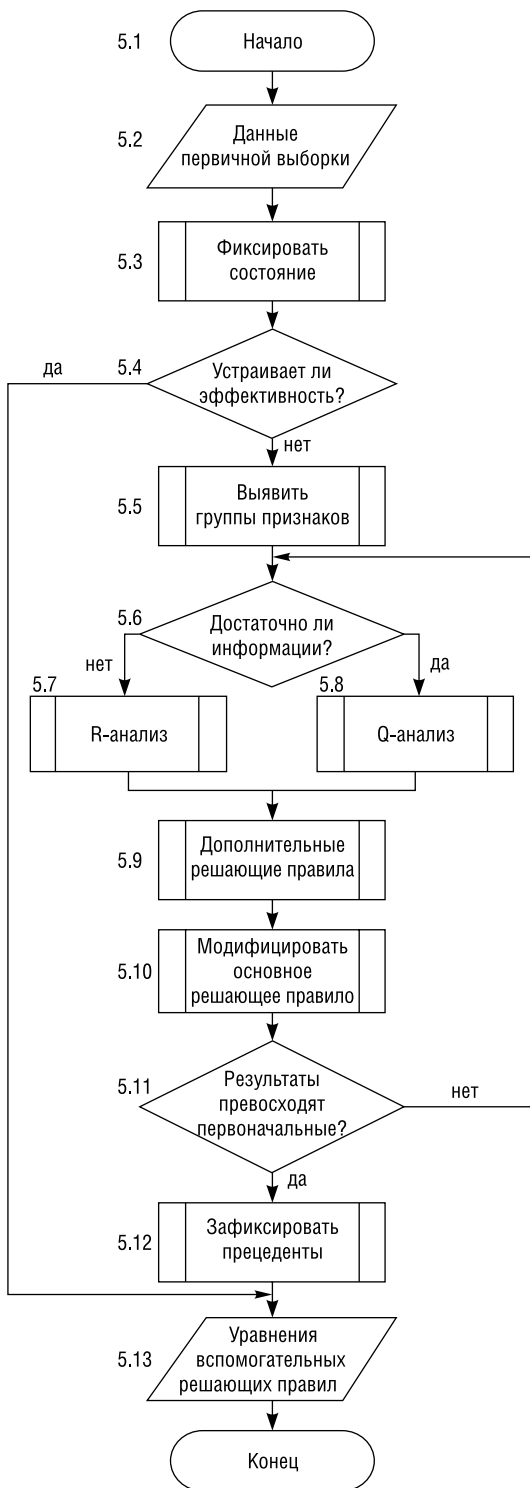


Рис. 2. Алгоритм синхронизации баз данных с целью формирования обучающего множества прецедентов



- 5.2 — данные первичной обучающей выборки, априорная информация, дополнительные организационные сведения;
- 5.3 — фиксация текущего состояния обучающей выборки, результатов основного РП и его операционных характеристик;
- 5.4 — устраивает ли эффективность имеющегося механизма прогнозирования?
- 5.5 — выявить самостоятельные группы параметров (на основании результатов опроса экспертов);
- 5.6 — достаточно ли экспертной информации?
- 5.7 — задействовать факторный анализ для выявления вспомогательных переменных;
- 5.8 — применить кластерный анализ для формирования абстрактных признаков, не имеющих предметной интерпретации, но вносящих вклад в межгрупповую дисперсию;
- 5.9 — построить дополнительные решающие правила на основании условных классификаций по абстрактным группировочным признакам (применяется дискриминантный анализ);
- 5.10 — модифицировать основное решающее правило с учетом абстрактных и вспомогательных переменных, с применением средства разрешения многосторонних споров;
- 5.11 — результаты модификации превосходят первоначальные? (сравнение с позиции операционных характеристик);
- 5.12 — зафиксировать результаты условной классификации и абстрактные признаки;
- 5.13 — уравнения вспомогательных решающих правил, модифицированное основное решающее правило РП, операционные характеристики РП

Рис. 3. Алгоритм усиления межгрупповых различий и повышения внутригрупповой концентрации объектов

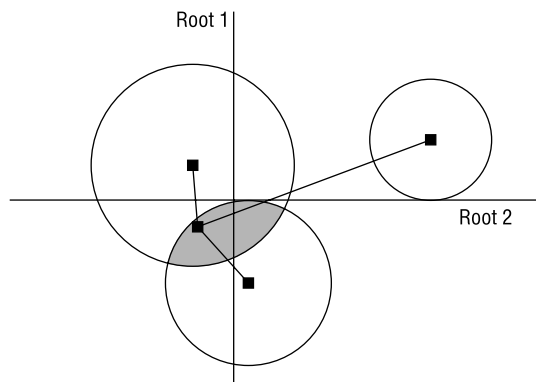


Рис. 4. Иллюстрация множества ответов, попадающих в область спорных ситуаций (выделено серым) в пространстве классифицирующих функций Root 1 и Root 2

Предварительно строятся вспомогательные множества $\overline{\text{Se}} = \{\max\{Se_i\}, med1\{Se_i\}, \dots, \min\{Se_i\}\}$ и $\overline{\text{Sp}} = \{\max\{Sp_j\}, med1\{Sp_j\}, \dots, \min\{Sp_j\}\}$. В построенных множествах все элементы упорядочены по убыванию их величины. Также построим множество $\overline{\text{R}} = \{\min\{R_k\}, med1\{R_k\}, \dots, \max\{R_k\}\}$, здесь упорядочение ведется по возрастанию значений удаленности объекта от центров классов. В обозначенных выше множествах функция $med1, med2 \dots med_{n-2}$ является функцией отыскания следующего по величине элемента исходных множеств $\text{Se}, \text{Sp}, \text{R}$.

Для каждой из n альтернатив (групп, патологий) осуществляются проверки ряда условий, выполнение или невыполнение которых влечет за собой увеличение соответствующих элементов множеств рейтингов T и антирейтингов S .

Алгоритм заключается в отыскании порядкового номера элемента множеств $\overline{\text{Se}}, \overline{\text{Sp}}$ и $\overline{\text{R}}$, соответствующих рассматриваемому объекту (предполагается, что удаленность объекта от центров групп уже определена и известен порядок на множестве элементов каждой из операционных характеристик). Рейтинг/антирейтинг высчитывается при помощи сопоставления величин специфичности/чувствительности (номера ее позиции во множестве $\overline{\text{Se}}$ или $\overline{\text{Sp}}$ соответственно) решаю-

щего правила при определении рассматриваемой группы и удаленности объекта от центра этой группы (номер позиции этого значения во множестве $\overline{\text{R}}$).

Введем функции, которые извлекают порядковый номер элемента для множеств $\overline{\text{Se}}, \overline{\text{Sp}}$ и $\overline{\text{R}}$:

$$\begin{aligned} F_{\overline{\text{R}}} &: \overline{\text{R}} \longrightarrow i_{\text{R}}; \\ F_{\overline{\text{Se}}} &: \overline{\text{Se}} \longrightarrow i_{\text{Se}}; \\ F_{\overline{\text{Sp}}} &: \overline{\text{Sp}} \longrightarrow i_{\text{Sp}}. \end{aligned}$$

В качестве аргументов функции получают название рассматриваемой группы. В качестве результата возвращают порядковый номер позиции, соответствующий рассматриваемой группе.

Таким образом, для каждой группы (патологии, диагноза и т.д.) можно вычислить ее рейтинг и антирейтинг, осуществив следующие действия.

Для каждой существующей группы ($j = 1; j++; j = n$)

$$\begin{aligned} S_j &+ F_{\overline{\text{Se}}}(j) \cdot F_{\overline{\text{R}}}(j), \\ T_j &+ F_{\overline{\text{Sp}}}(j) \cdot F_{\overline{\text{R}}}(j). \end{aligned}$$

Цикл по n альтернативам окончен, два множества S и T сформированы.

Рассмотрим производное множество: $\alpha = \left\{ \frac{S_1}{T_1}, \frac{S_2}{T_2}, \dots, \frac{S_n}{T_n} \right\}$. При условии, что T_i не равно 0 (в противном случае патология не имеет никакого приоритета при диагностике), чем больше значение T_i (рейтинга i -й альтернативы) и ниже S_i (антирейтинга альтернативы i), тем меньше значение отношения $\frac{S_i}{T_i}$ и тем больше приоритет у i -й альтернативы в сложившейся ситуации (расстояния всегда будут различными).

Если $\alpha_i = \min\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, тогда i -я патология имеет высший приоритет в сложившейся ситуации.

Если $\alpha_i = \alpha_j = \min\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, тогда необходимо руководствоваться следующим правилом:

Пусть $\text{Rmax} = \{R_{\max 1}, R_{\max 2}, \dots, R_{\max n}\}$ — множество внешних радиусов на одном уровне распознавания. Назовем коэффициентом

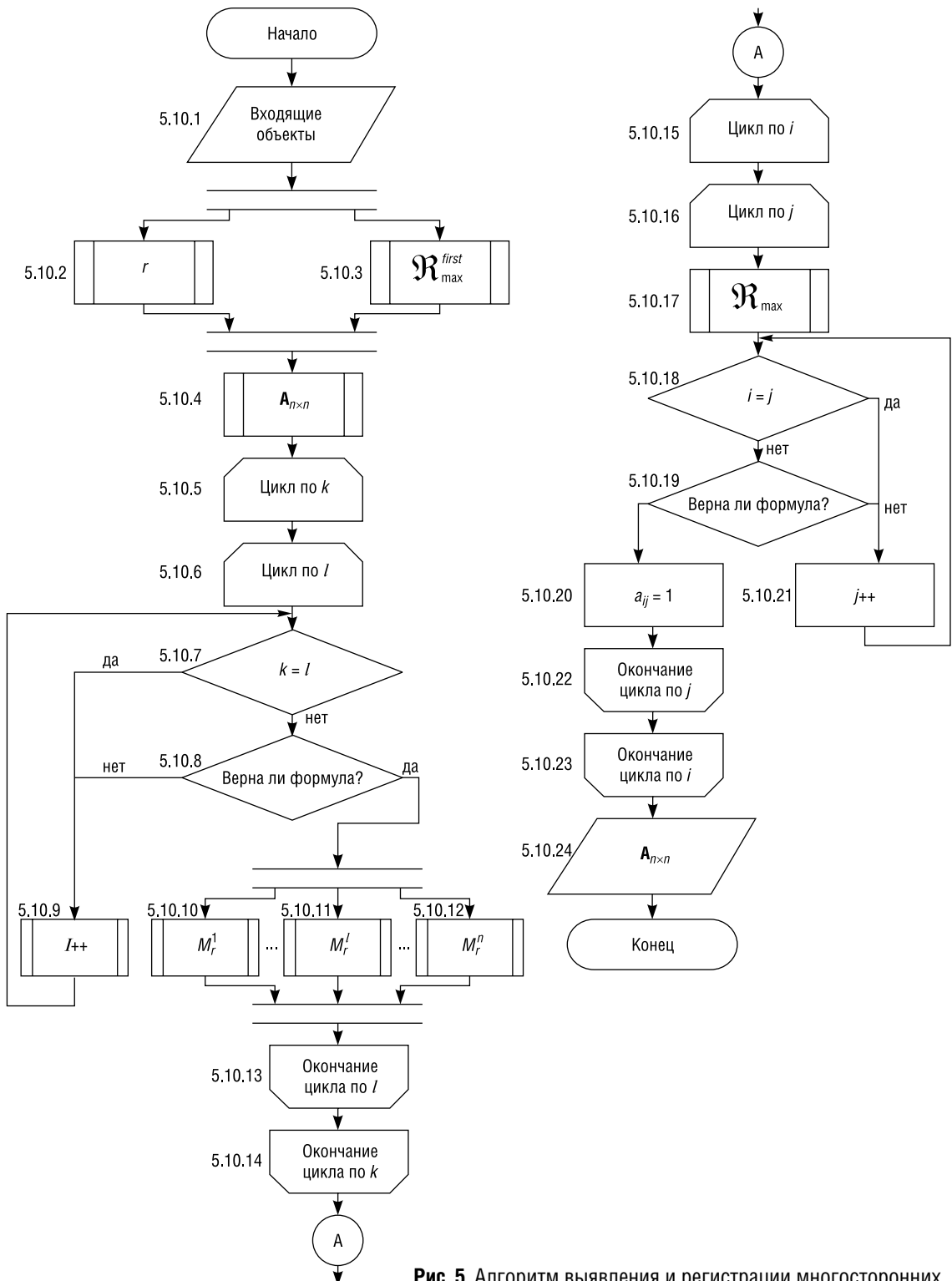


Рис. 5. Алгоритм выявления и регистрации многосторонних

- 5.10.1 — координаты центроидов групп, координаты исследуемого объекта, координаты объектов, принадлежащих соответствующим группам;
- 5.10.2, 5.10.3 — для каждой группы рассчитать расстояние от ее центроида до наиболее удаленного объекта, принадлежащего этой группе; построить множество $\mathcal{R}_{\max}^{first}$ — множество максимальных радиусов каждой из рассматриваемых групп. Рассчитать удаленность всех объектов обучающей выборки от центроидов групп, к которым они принадлежат, и построить множества $r_{1,2,3}$ — множества всех расстояний от объектов до центроидов своих групп;
- 5.10.4 — создать матрицу **A** размерностью $n \times n$, где n — число рассматриваемых групп (все элементы матрицы нулевые);
- 5.10.5 — $k = 1, \dots, n$, где n — число рассматриваемых групп;
- 5.10.6 — $l = 1, \dots, n$, где n — число рассматриваемых групп. Начинается поиск объектов, попадающих в первичное пересечение областей рассматриваемых групп;
- 5.10.7 — проверка попадания механизма перебора на главную диагональ матрицы **A**;
- 5.10.8 — $r_i \leq R_{\max i}^{first}$ & $r_j \leq R_{\max j}^{first}$;
- 5.10.9 — переход к следующей итерации без каких-либо действий;
- 5.10.10, 5.10.11, 5.10.12 — пополнить массивы M_r^1 , M_r^l и M_r^n расстояний от соответствующих центроидов до объектов, попавших в пересечение областей соседних рассматриваемых групп. Значения расстояний до объектов каждой группы поместить в отдельный массив;
- 5.10.13, 5.10.14 — цикл заканчивается после перебора всех групп, рассматриваемых в соответствующем пространстве канонических линейных дискриминантных функций;
- 5.10.15 — $i = 1, \dots, n$, где n — число рассматриваемых групп;
- 5.10.16 — $j = 1, \dots, n$, где n — число рассматриваемых групп. Начинается поиск объектов, попадающих во вторичное пересечение областей рассматриваемых групп;
- 5.10.17 — построить трехмерный массив \mathcal{R}_{\max} , состоящий из максимальных значений расстояний от центроида соответствующей группы до объекта, попавшего в область спорных вопросов для пары рассматриваемых групп (номер строки — одна группа, номер столбца — другая группа). В ячейке хранятся два значения: максимальное значение расстояния от объекта до центроида одной и второй групп соответственно;
- 5.10.18 — проверка попадания объекта на главную диагональ матрицы **A**;
- 5.10.19 — $r_i \leq R_{\max 1}^{ij}$ & $r_j \leq R_{\max 2}^{ij}$;
- 5.10.20 — присвоить элементу a_{ij} матрицы **A** значение 1 (номер строки — одна группа, номер столбца — другая группа);
- 5.10.21 — переход к следующей итерации без каких-либо действий;
- 5.10.22, 5.10.23 — цикл заканчивается после перебора всех групп, рассматриваемых в соответствующем пространстве канонических линейных дискриминантных функций;
- 5.10.24 — матрица регистрации неоднозначных результатов.

спорных ситуаций в ответах решающего правила

пропорциональной удаленности двух альтернатив i и j (КПУ_{ij}) отношение двух внешних радиусов пересекающихся альтернатив

$$\text{КПУ}_{ij} = \frac{R \max_i}{R \max_j}.$$

Если $\frac{R_i}{R_j} > \frac{R \max_i}{R \max_j}$ — предпочтение отдается j -й патологии.

Если $\frac{R_i}{R_j} < \frac{R \max_i}{R \max_j}$ — предпочтение отдается i -й патологии.

Если $\frac{R_i}{R_j} = \frac{R \max_i}{R \max_j}$ — предпочтение отдается наиболее тяжелому варианту патологии (сочетание патологий тяжелее монопатологии).

Приведенные выше решения проблемы разрозненности, слабоинформативности входных данных, а также низкого качества работы решающих правил поддержки прогнозирования, построенных по типовым схемам, реализованы в виде программного продукта «Система поддержки прогнозирования степени риска развития ретинопатии недоношенных GimRN 1.0». Принципиальная схема работы программы представлена на рис. 6.

Под множеством гибких решающих правил подразумеваются специализированные

правила, способные функционировать при отсутствии некоторого подмножества значений признаков.

Для проверки качества работы построенной системы поддержки прогнозирования проведено исследование ее операционных характеристик и их сравнение с характеристиками типовых решений, встречающихся в литературе. Как видно на рис. 7, показатели качества работы построенного решающего механизма превосходят показатели типовых решений.

Построенная модель прогнозирования ретинопатии недоношенных с использованием последовательности методов многомерного анализа данных и информационной системы апробирована на базе областного центра ретинопатии недоношенных научно-практического центра «Бонум». Результаты показали преимущество предлагаемого решения над имеющимися аналогичными схемами и средствами поддержки.

Аппарат многомерного анализа клинических, анамнестических и лабораторно-инструментальных данных не позволяет обеспечить информационно-программную поддержку надлежащего качества без предварительной адаптации. Благодаря применению специализированной последовательности

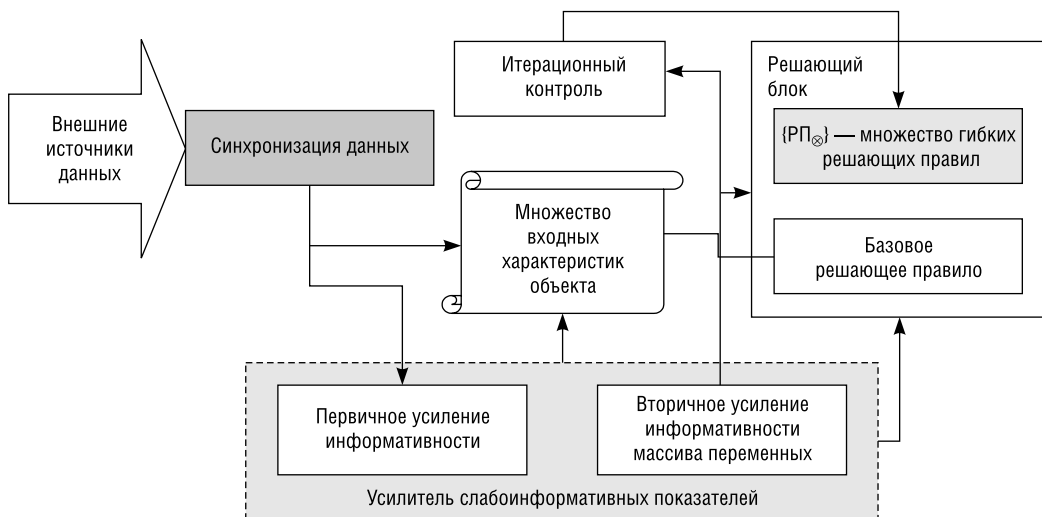


Рис. 6. Схема работы программы GimRN 1.0

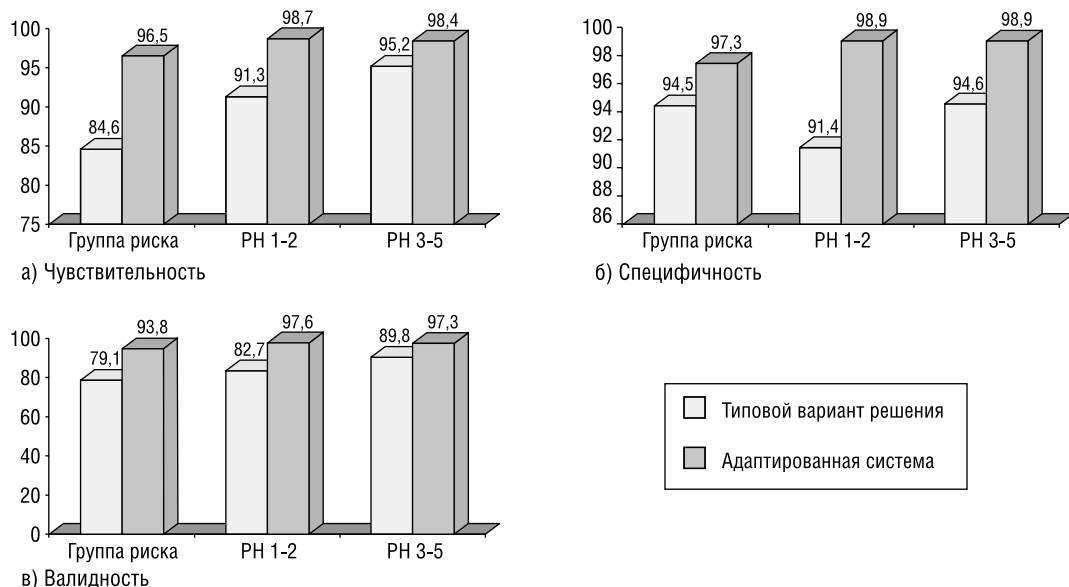


Рис. 7. Показатели качества решения задачи прогнозирования

на основе типового метода и адаптированной последовательности методов многомерного анализа

методов многомерного анализа можно увеличить значения операционных характеристик механизма поддержки прогнозирования до требуемого уровня.

Использование значений операционных характеристик в рамках алгоритма анализа компетентности прогностического правила позволяет добиться пустого пересечения множеств объектов различных классов в задаче прогнозирования, когда геометрическая интерпретация не дает однозначно интерпретируемой картины.

Список литературы

1. Бейли Н. Математика в биологии и медицине / Н. Бейли. М.: Мир, 1970.
2. Беллман Р. Математические методы в медицине / Р. Беллман. М.: Мир, 1987.
3. Гридин В.Н. Построение интеллектуальных диагностических систем в медицинских приложениях [Текст] / В.Н. Гридин, О.Б. Тарасова // Информационные технологии. 2007. №7. С. 54–58.
4. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях / В. Дюк, В. Эмануэль. СПб.: Питер, 2003.
5. Казанцев В.С. Математические методы и новые информационные технологии в решении

медицинских задач: лекции / В.С. Казанцев. Екатеринбург: НПЦ Уралмедсоцэкономпроблем, 2002.

6. Кельмансон И.А. Принципы доказательной педиатрии. СПб.: ООО «Издательство Фолиант», 2004.

7. Кудрин А.Н., Пономарева Г.Т. Применение математики в экспериментальной и клинической медицине / А.Н. Кудрин, Г.Т. Пономарева. М.: Медицина, 1967.

8. Лапко А.В. Статистические методы моделирования и принятия решений в развивающихся медико-биологических системах / А.В. Лапко. Новосибирск, 1991.

9. Мазуров В.Д. О плохо формализуемых задачах анализа сложных систем / В.Д. Мазуров // Математическое моделирование процессов в медицинских и биологических системах. Свердловск: УНЦ АН СССР, 1982. С. 3–7.

10. Охтилев М.Ю. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов [Текст] / М.Ю. Охтилев, Б.В. Соколов, Р.М. Юсупов. М.: Наука, 2006.

11. Смит Дж. Математические идеи в биологии / Дж. Смит. М.: Мир, 1970.

12. Штейн Л.Б. Опыт прогнозирования с помощью ЭВМ / Л.Б. Штейн. Л.: Ленинградский институт, 1984.