Диагностические медицинские классификаторы

Б. М. Гавриков, Н. В. Пестрякова

Федеральный исследовательский центр "Информатика и управление" РАН, Москва, Россия

Аннотация. Для предварительной диагностики заболеваний человека по данным анализа периферической крови разработан метод статистической классификации, основанный на полиномиальной регрессии. Он реализован в качестве двух приложений. Один тип классификаторов позволяет оценить состояние пациента от здорового до максимальной степени поражения (онкология) в рамках отдельной системы организма. С помощью классификатора другого вида выявляется область локализации опухоли при онкологическом заболевании. Для различных вариантов классификации исследуется структура обучающего множества, проводится сравнительный анализ.

Ключевые слова: онкологическое заболевание, система организма, периферическая кровь, классификация, полиномиальная регрессия, обучающее множество.

DOI 10.14357/20718632240302 **EDN** BTQWIO

Введение

Используемый нами подход к медицинской диагностике основывается на концепции крупнейших гематологов, который основывается на том, что многие заболевания человека вносят изменения в состав его крови [1]. Проведенные ранее статистические исследования баз параметров крови служат подтверждением правомочности этого представления [2-8].

При оценке состояния здоровья человека (СЗЧ) гематологи предлагают использовать не менее пяти показателей периферической крови (из пальца) [1]. Организм представляется в виде совокупности систем органов: пищеварения, дыхания, урологической, эндокринной, гинекологической (у женщин), опорно-двигательного аппарата, печени и желчевыводящих путей, грудных желез (у женщин), центральной нервной системы (ЦНС) и органов чувствительности. Мужчины и женщины рассматриваются по отдельности, поскольку диапазоны вариации параметров крови среди множества людей существенно зависят от пола.

СЗЧ включает четыре градации: здоровые, начальные отклонения от нормы, выраженные изменения, тяжелое заболевание (онкология).

В серии работ мы показали, что статистический метод классификации может успешно применяться для оценивания СЗЧ по показателям периферической крови [2-8]. Для каждой системы организма (СО) проводилось исследование СЗЧ при помощи соответствующего классификатора, обученного на профильной выборке. Базы параметров крови были созданы с использованием верифицированных диагнозов [1]. Позднее мы разработали способ диагностики онкологических заболеваний человека различной локализации [9-13]. Обучение статистического классификатора проводилось на наборах показателей периферической крови пациентов с опухолями в ряде СО.

В настоящей работе для мужчин рассматриваются семь вышеприведенных СО. Описываются классификаторы двух видов. Первый относится к каждой отдельно взятой СО. Классы «1», «2», «3», «4» соответствует СЗЧ. Второй

определяет локализацию онкопатологии по семи СО – классы «1», «2», «3», «4», «5», «6», «7».

Для классификаторов этих двух типов проводится сравнительный анализ структуры обучающего множества. А именно, в рамках некой СО рассматривается его часть, относящаяся к онкобольным. В первом случае она представляет собой класс «4» для данной СО, а во втором – один из семи классов, соответствующих набору СО.

1. Метод классификации

Используется восемь показателей крови. Их общепринятые обозначения и размерность следующие: RBC [L^{-1}] – эритроциты, HGB [gL^{-1}] – гемоглобин, PLT [L^{-1}] – тромбоциты, WBC [L^{-1}] – лейкоциты, LIMPH [L^{-1}], [%] – лимфоциты, GRAN [L^{-1}], [%] – гранулоциты (GRAN=NEUT+EOS+BASO, где NEUT [L^{-1}], [%] – нейтрофилы, EOS [L^{-1}],[%] – эозинофилы, BASO [L^{-1}],[%] – базофилы).

Изложим по шагам постановку задачи для двух приложений. **Вариант 1** – классификатор, позволяющий определить класс здоровья, строится отдельно по каждой СО. **Вариант 2** предназначен для нахождения СО, в которой локализована онкологическая опухоль.

Шаг 1. Вариант 1. Рассматриваем определенную СО человека. Вводим вектор $\mathbf{v}\in\mathbf{R}^N$, i-я компонента которого – отнормированная на отрезок [0,1] величина i-го показателя крови, причем N=8.

Нормировка на отрезок [0,1] проводится следующим образом. По обучающей выборке фиксированной СО, включающей все четыре градации СЗЧ, для каждого i-го показателя крови находим минимальное и максимальное значение $v_i^{\min}$, $v_i^{\max}$, причем $i = 1,..., N$.

$$\begin{aligned} v_i^{\min} &= \min_j \{v_i^j\}, j = 1, \ldots, J, \\ v_i^{\max} &= \max_j \{v_i^j\}, j = 1, \ldots, J, \end{aligned} \tag{1}$$

где J – объем выборки по данной СО. Затем выполняем следующее преобразование:

$$v_i \rightarrow (v_i - v_i^{min})/(v_i^{max} - v_i^{\min}). \tag{2}$$

Отождествляем k-й элемент множества классов СЗЧ с базисным вектором $\mathbf{e}_k = (0\ldots1\ldots0)$ (здесь 1 находится на k-м месте, $1\le k\le K$, причем $K = 4$) из $\mathbf{R}^K$. Обозначаем $Y = \{\mathbf{e}_1,\ldots,\mathbf{e}_K\}$.

Пусть $p_k(\mathbf{v})$ – вероятность того, что набор отнормированных показателей крови соответствует k–му элементу СЗЧ, где $1\le k\le K$. Искомый класс СЗЧ будет иметь порядковый номер r, получивший максимальное значение вероятности:

$$p_r(\mathbf{v}) = \max_k \{p_k(\mathbf{v})\},\ 1\le k\le K. \tag{3}$$

Шаг 1. Вариант 2. Рассматриваем K определенных перенумерованных СО человека, $1\le k\le K$, причем $K = 7$. Вводим вектор $\mathbf{v}\in\mathbf{R}^N$, i-я компонента которого – отнормированная на отрезок [0,1] величина i-го показателя крови онкобольных, где $N = 8$.

Нормирование проводим следующим образом. Рассмотрим обучающие выборки параметров крови онкобольных всех семи исследуемых СО. Для каждого i-го показателя крови находим (1) минимальное и максимальное значение $v_i^{\min}$, $v_i^{\max}$, где $i = 1,..., N$. В (1) J – суммарный объем выборок по совокупности рассматриваемых СО. Затем выполняем преобразование (2).

Отождествляем k-й элемент множества СО с базисным вектором из $\mathbf{R}^K$: $\mathbf{e}_k = (0\ldots1\ldots0)$, причем 1 находится на k-м месте, $1\le k\le K$, K=7. Обозначим $Y = \{\mathbf{e}_1,\ldots,\mathbf{e}_K\}$.

Пусть существует $p_k(\mathbf{v})$ – вероятность того, что набор отнормированных показателей крови онкобольных соответствует k–му элементу СО, где $1\le k\le K$. Искомый элемент СО будет иметь порядковый номер r, получивший максимальное значение вероятности (3).

Шаг 2. Вариант 1, 2. Приближенные значения $p_1(\mathbf{v}),\ldots,p_K(\mathbf{v})$ представляются в виде конечных многочленов от координат $\mathbf{v} = (v_1,\ldots,v_N)$ и определяются выбором базисных мономов:

$$p_k(\mathbf{v}) \cong c_0^{(k)} + \sum_{i=1}^{N} c_i^{(k)} v_i + \sum_{i,j=1}^{N} c_{i,j}^{(k)} v_i v_j + \ldots, \tag{4}$$
$$1 \le k \le K.$$

Запишем упорядоченные базисные мономы из (4) в виде вектора размерности L:

$$\mathbf{x}(\mathbf{v}) = (1, v_1, \ldots, v_N, \ldots)^{\mathrm{T}}. \tag{5}$$

Тогда (4) имеет следующий вид:

$$\mathbf{p}(\mathbf{v}) = (p_1(\mathbf{v}),\ldots,p_K(\mathbf{v}))^{\mathrm{T}} \cong A^{\mathrm{T}}\mathbf{x}(\mathbf{v}), \tag{6}$$

где A – матрица размера $L\times K$, столбцами которой являются векторы $\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}$. Каждый такой

вектор составлен из коэффициентов при мономах соответствующей строки (4) (с совпадающим верхним индексом), упорядоченных так же, как в векторе $\mathbf{x}(\mathbf{v})$ – (5).

Приближенное вычисление A производится при обучении на конечной последовательности: $[\mathbf{v}^{(1)},\mathbf{y}^{(1)}]$, …, $[\mathbf{v}^{(J)},\mathbf{y}^{(J)}]$. Здесь используются следующие обозначения.

Вариант 1: $\mathbf{v}^{(j)}$ – набор параметров крови, соответствующий элементу СЗЧ с номером k ($1\le k\le K$, K=4).

Вариант 2: $\mathbf{v}^{(j)}$ – набор параметров крови онкобольных, соответствующий элементу СО с номером k ($1\le k\le K$, K=7).

Вариант 1, 2: $\mathbf{y}^{(j)} = (0\ldots1\ldots0)$ – базисный вектор размерности 4 или 7 соответственно, где 1 стоит на k-м месте, $1\le j\le J$).

Итак,

$$A \cong \Big(\frac{1}{J}\sum_{j=1}^{J}\mathbf{x}^{(j)}(\mathbf{x}^{(j)})^{\mathrm{T}}\Big)^{-1}\Big(\frac{1}{J}\sum_{j=1}^{J}\mathbf{x}^{(j)}(\mathbf{y}^{(j)})^{\mathrm{T}}\Big). \quad (7)$$

Поскольку проблема обращения заполненной матрицы большой размерности до сих пор не решена [14], правую часть (7) получаем посредством рекуррентной процедуры [15].

Использовались различные модификации вектора $\mathbf{x}(\mathbf{v})$ – (5). Ниже указаны типы классификатора (**Вариант 1** или **Вариант 2**), а также в первом случае СО, для которых при $\mathbf{x}(\mathbf{v})$ данного вида были получены наилучшие результаты, и более сложные модификации не использовались. Точность классификации приведена, если она меньше 100%.

Далее выражения в фигурных скобках соответствуют цепочкам элементов вектора, вычисляемым по всем показателям крови из имеющегося набора.

1) **Вариант 1.** *Печень и желчевыводящие пути. ЦНС и органы чувствительности.*

$$\mathbf{x} = (1,\{v_i\},\{v_i^3\},\{v_i^4\},\{v_i v_j\}),\ 1\le i\le 8,\ i\le j\le 8. \quad (8)$$

Длина полинома 61. Имеются мономы степенного вида первого, второго, третьего и четвертого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

2) **Вариант 1.** *Опорно-двигательный аппарат.*

$$\mathbf{x} = (1,\{v_i\},\{v_i^3\},\{v_i^4\},\{v_i^5\},\{v_i^6\},\{v_i v_j\}),\ 1\le i\le 8,\ i\le j\le 8. \quad (9)$$

Длина полинома 77. Имеются мономы степенного вида первого, второго, третьего, четвертого, пятого и шестого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

3) **Вариант 1.** *Органы дыхания.*

$$\mathbf{x} = (1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\}),\ 1\le i\le 8,\ i\le j\le 8,\ j\le k\le 8. \quad (10)$$

Длина полинома 165. Имеются мономы первого, второго и третьего порядка. Перекрестные произведения используются для мономов второго и третьего порядка.

4) **Вариант 1.** *Пищеварительная система (98,2%). Урологическая система (97,3%). Эндокринная система (95%).*

$$\mathbf{x} = (1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\},\{v_i v_j v_k v_l\},\{v_i v_j v_k v_l v_m\}),\ 1\le i\le 8,\ i\le j\le 8,\ j\le k\le 8,\ k\le l\le 8,\ l\le m\le 8. \quad (11)$$

Длина полинома 1287. Имеются мономы первого, второго, третьего, четвертого и пятого порядка. Перекрестные произведения используются для мономов второго, третьего, четвертого и пятого порядка.

5) **Вариант 2.** *(95%).*

$$\mathbf{x} = (1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\},\{v_i v_j v_k v_l\},\{v_i v_j v_k v_l v_m\},\ \{v_i v_j v_k v_l v_m v_n\},\ \{v_i v_j v_k v_l v_m v_n v_p\}),\ 1\le i\le 8,\ i\le j\le 8,\ j\le k\le 8,\ k\le l\le 8,\ l\le m\le 8,\ m\le n\le 8,\ n\le p\le 8. \quad (12)$$

Длина полинома 6435. Имеются мономы первого, второго, третьего, четвертого, пятого, шестого и седьмого порядка. Перекрестные произведения используются для мономов второго, третьего, четвертого, пятого, шестого и седьмого порядка.

Итак, исследуется база периферической крови мужчин по семи СО. Построены и обучены классификаторы двух типов.

Вариант 1 для каждой из этих семи СО по отдельности имеет четыре класса, соответствующие градациям СЗЧ: «1» – C^1, «2» – C^2, «3» – C^3, «4» – C^4.

Вариант 2 относится к базам крови онкологических больных по семи СО для мужчин: пищеварительная – C^1, органы дыхания – C^2, опорно-двигательный аппарат – C^3, урологическая система – C^4, эндокринная – C^5, печени и

желчевыводящих путей – C^6, центральной нервной системы и органов чувствительности – C^7.

Далее **Вариант 1** (классы здоровья) и **Вариант 2** (локализация онкологической опухоли) будут рассмотрены в рамках проводимого сравнительного анализа структуры обучающих множеств: $C^1UC^2UC^3UC^4$ и $C^1UC^2UC^3UC^4UC^5UC^6UC^7$ соответственно.

2. Расстояния между «своими» и «чужими» элементами

Для каждого из рассматриваемых классов C^1, C^2, C^3, C^4 (**Вариант 1**) или C^1, C^2, C^3, C^4, C^5, C^6, C^7 (**Вариант 2**) в отдельности найдем минимальное, максимальное и среднее расстояние между своими векторами (принадлежащими данному классу). Для множества векторов k-го класса определяем их следующим образом.

Минимальное расстояние:

$$U\_k_{\min} = \min_{V^k} \{\|\mathbf{v}^k - \mathbf{u}^k\|\},\ \mathbf{v}^k \in V^k,\ \mathbf{u}^k \in V^k,\ \mathbf{v}^k \neq \mathbf{u}^k \quad (13)$$

Максимальное расстояние:

$$U\_k_{\max} = \max_{V^k} \{\|\mathbf{v}^k - \mathbf{u}^k\|\},\ \mathbf{v}^k \in V^k,\ \mathbf{u}^k \in V^k \quad (14)$$

где $\mathbf{v}^k$ и $\mathbf{u}^k$ – пары различных векторов, принадлежащих множеству элементов k-го класса V^k.

Среднее расстояние определим с приведением алгоритма нахождения этой величины:

$$U\_k_{\text{ср}} = \sum_{j=1}^{J_k} \sum_{j1=j+1}^{J_k} \| \mathbf{w}^{k,j} - \mathbf{w}^{k,j1} \| / (J_k\ (J_k - 1)/2),$$
$$\mathbf{w}^{k,j} \in V^k, j = 1, \ldots, J_k \quad (15)$$

где $\{\mathbf{w}^{k,j}, j = 1, \ldots, J_k\} = V^k$ – представление совокупности элементов k-го класса в виде множества перенумерованных векторов.

Аналогично, получим соответствующие значения для пар свой – чужой по каждому из классов. Чужой вектор – не принадлежащий рассматриваемому классу. Обучающее множество, содержащее элементы всех четырех классов (**Вариант 1**) суть $V = \{V^1 U V^2 U V^3 U V^4\}$; для семи классов (**Вариант 2**) $V = \{V^1 U V^2 U V^3 U V^4 U V^5 U V^6 U V^7\}$.

Минимальное расстояние:

$$U\_kz_{\min} = \min_{V} \{\|\mathbf{v}^k - \mathbf{u}^{\neg k}\|\},\ \mathbf{v}^k \in V^k,\ \mathbf{u}^{\neg k} \in V^{\neg k} \quad (16)$$

Максимальное расстояние:

$$U\_kz_{\max} = \max_{V} \{\|\mathbf{v}^k - \mathbf{u}^{\neg k}\|\},\ \mathbf{v}^k \in V^k,\ \mathbf{u}^{\neg k} \in V^{\neg k} \quad (17)$$

где $\mathbf{v}^k$ и $\mathbf{u}^{\neg k}$ – пары векторов, из которых $\mathbf{v}^k$ принадлежит множеству элементов k-го класса V^k, а $\mathbf{u}^{\neg k}$ принадлежит множеству чужих элементов $V^{\neg k}$ классов, отличных от k-го: $V^{\neg k} = V \backslash V^k$.

Среднее расстояние:

$$U\_kz_{\text{ср}} = \sum_{j=1}^{J_k} \sum_{j1=1}^{J_{\neg k}} \| \mathbf{w}^{k,j} - \mathbf{w}^{\neg k,j1} \| / (J_k\ J_{\neg k}),$$
$$\mathbf{w}^{k,j} \in V^k, j = 1, \ldots, J_k, \mathbf{w}^{\neg k,j1} \in V^{\neg k}, j1 = 1, \ldots, J_{\neg k}, \quad (18)$$

где $\{\mathbf{w}^{k,j}, j = 1, \ldots, J_k\} = V^k$ – представление совокупности своих элементов k-го класса в виде множества перенумерованных векторов, аналогично для множества чужих элементов классов, отличных от k-го: $\{\mathbf{w}^{\neg k,j1}, j1 = 1, \ldots, J_{\neg k}\} = V^{\neg k}$, $V^{\neg k} = V \backslash V^k$.

Далее в метрике L_2 представлены (Рис. 1, а – 14, а) минимальное, среднее и максимальное расстояние (значения ординат для точек 1, 2, 3 по оси абсцисс) между своими векторами (Ряд 1), аналогичные величины для пар свой–чужой (Ряд 3).

Вариант 1. Для каждой из СО здесь и ниже в разделах 3 - 5 используются обозначения: «свой» – класс C^4 ($k = 4$), «чужой» – совокупность $C^1UC^2UC^3$. Эти части обучающего множества имеют следующий объем.

Пищеварительная система: $|C^4| = 33$, $|C^1UC^2UC^3| = 71$ (Рис. 1а).

Органы дыхания: $|C^4| = 21$, $|C^1UC^2UC^3| = 50$ (Рис. 2а).

Опорно-двигательный аппарат: $|C^4| = 33$, $|C^1UC^2UC^3| = 38$ (Рис. 3а).

Урологическая система: $|C^4| = 32$, $|C^1UC^2UC^3| = 73$ (Рис. 4а).

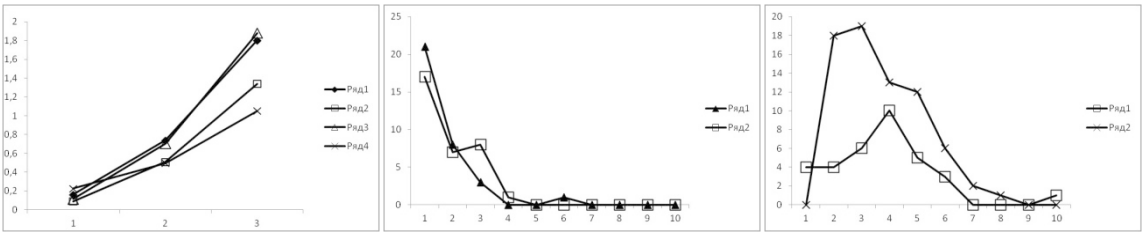
Эндокринная система: $|C^4| = 34$, $|C^1UC^2UC^3| = 82$ (Рис. 5а).

ЦНС и органы чувствительности: $|C^4| = 26$, $|C^1UC^2UC^3| = 30$ (Рис. 6а).

Печень и желчевыводящие пути: $|C^4| = 31$, $|C^1UC^2UC^3| = 38$ (Рис. 7 а).

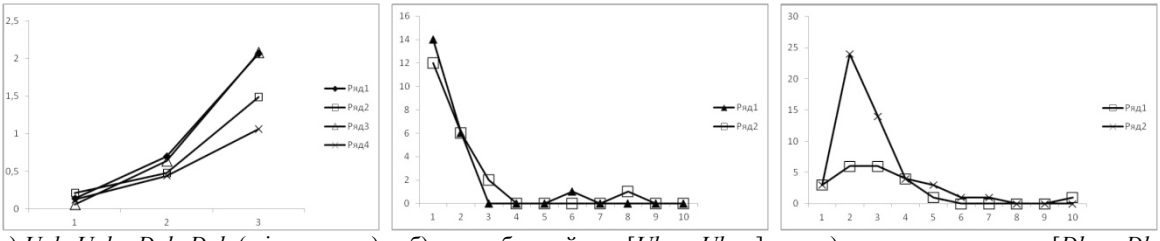
Для всех семи СО **Варианта 1** (Рис. 1, а – 7,а) Ряд 1 превышает Ряд 3 по средним значениям. Если для минимальных значений Ряд 1 превышает Ряд 3, то для максимальных – Ряд 3 больше, чем Ряд 1 (Рис. 1, 2, 4, 5, 7, а). Если для минимальных величин Ряд 3 больше, чем Ряд 1,

то для максимальных – Ряд 1 превышает Ряд 3 (Рис. 3, а и 6, а).



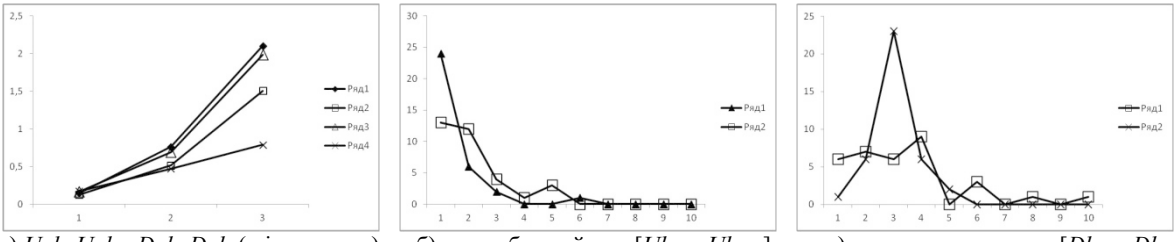
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 1. Вариант 1. Пищеварительная система



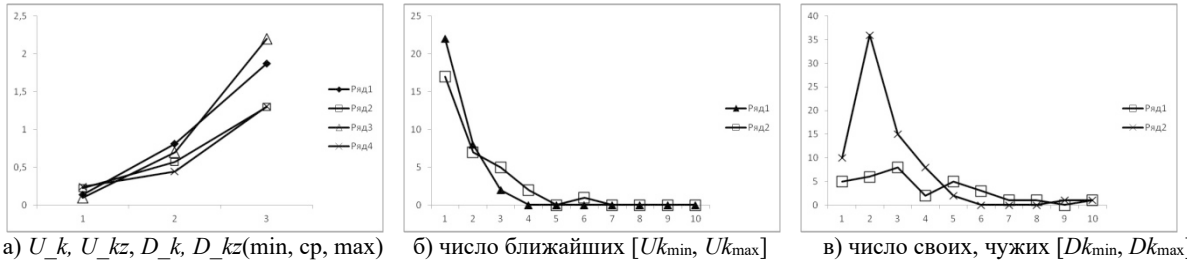
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 2. Вариант 1. Органы дыхания



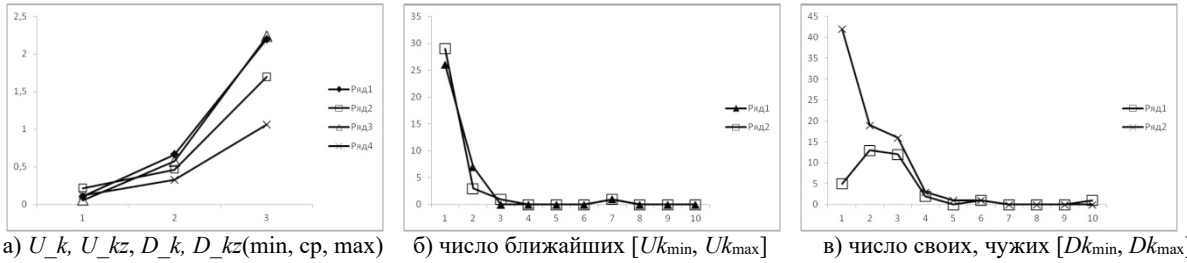
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 3. Вариант 1. Опорно-двигательный аппарат



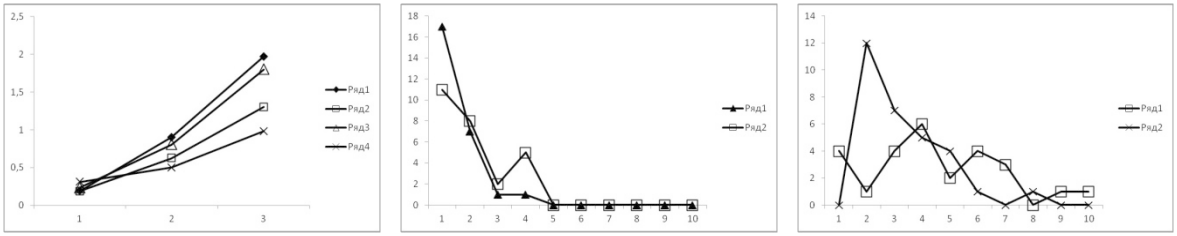
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 4. Вариант 1. Урологическая система.



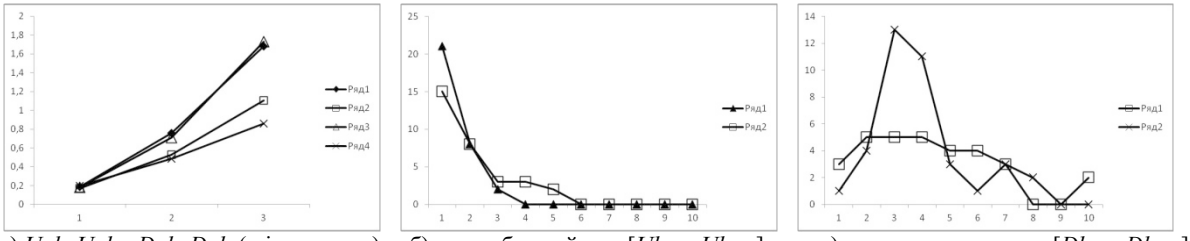
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 5. Вариант 1. Эндокринная система



а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 6. Вариант 1. ЦНС и органы чувствительности



а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 7. Вариант 1. Печень и желчевыводящие пути

Вариант 2. Здесь и далее в разделах 3 - 5 используются следующие обозначения: «свой» – один из рассматриваемых семи классов, «чужой» – совокупность остальных классов. Объем этих частей обучающего множества следующий.

C^1 – пищеварительная система:
$|C^1| = 33$, $|C^2UC^3C^4UC^5UC^6UC^7| = 174$ (Рис. 8, а).

C^2 – органы дыхания:
$|C^2| = 21$, $|C^1UC^3C^4UC^5UC^6UC^7| = 186$ (Рис. 9, а).

C^3 – опорно-двигательный аппарат:
$|C^3| = 33$, $|C^1UC^2C^4UC^5UC^6UC^7| = 174$ (Рис. 10, а).

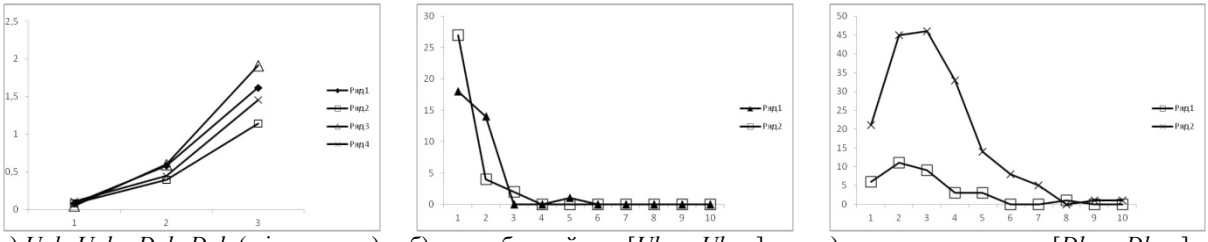
C^4– урологическая система:
$|C^4| = 31$, $|C^1UC^2UC^3C^5UC^6UC^7| = 177$ (Рис. 11, а).

C^5 – эндокринная система:
$|C^5| = 33$, $|C^1UC^2UC^3C^4UC^6UC^7| = 174$ (Рис. 12, а).

C^6 – ЦНС и органы чувствительности:
$|C^6| = 25$, $|C^1UC^2UC^3C^4UC^5UC^7| = 182$ (Рис. 13, а).

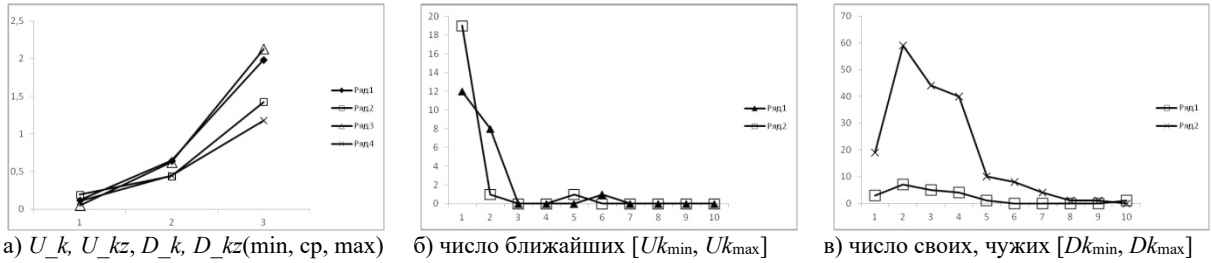
C^7 – печень и желчевыводящие пути:
$|C^7| = 31$, $|C^1UC^2UC^3C^4UC^5UC^6| = 176$ (Рис. 14, а).

Для всех семи классов **Варианта 2** (Рис. 8, а – 14, а) Ряд 1 превышает Ряд 3 по минимальным значениям (причем, обе малы); первая величина меньше второй для максимальных величин. Для средних – имеются оба варианта различия.



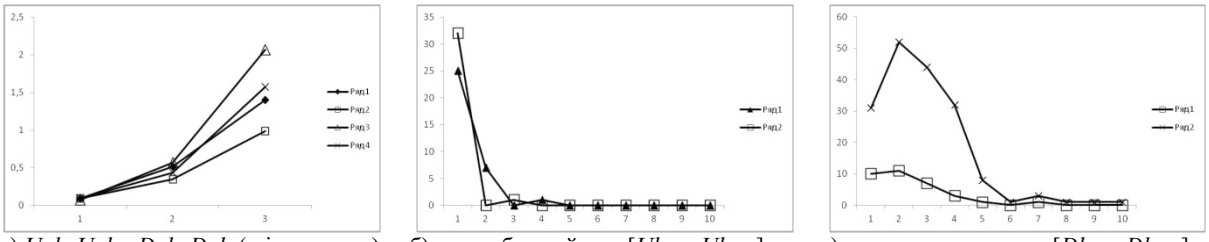
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 8. Вариант 2. С – пищеварительная система.



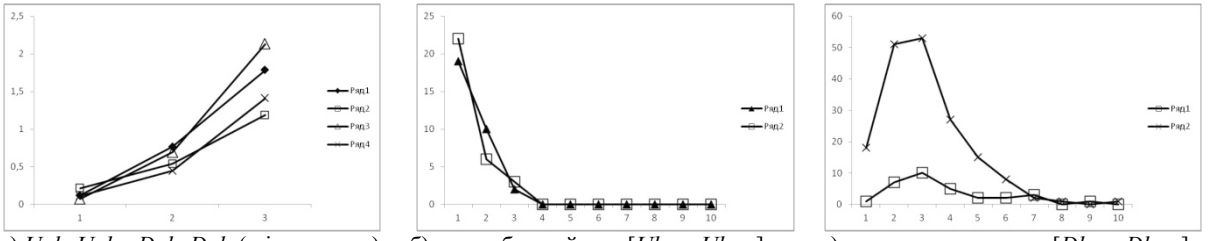
а) *U_k, U_kz*, *D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 9. Вариант 2. С – органы дыхания



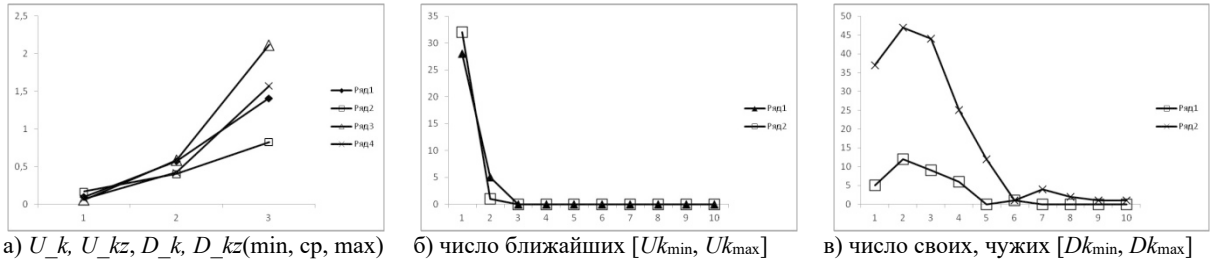
а) *U_k, U_kz, D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 10. Вариант 2. C^3 – опорно-двигательный аппарат



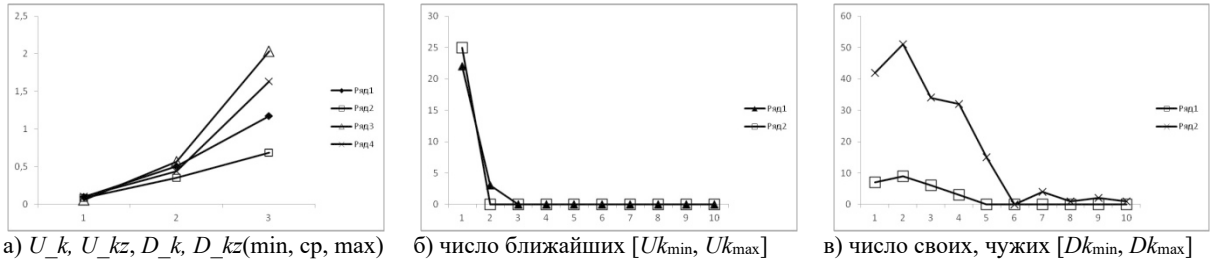
а) *U_k, U_kz, D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 11. Вариант 2. C^4 – урологическая система



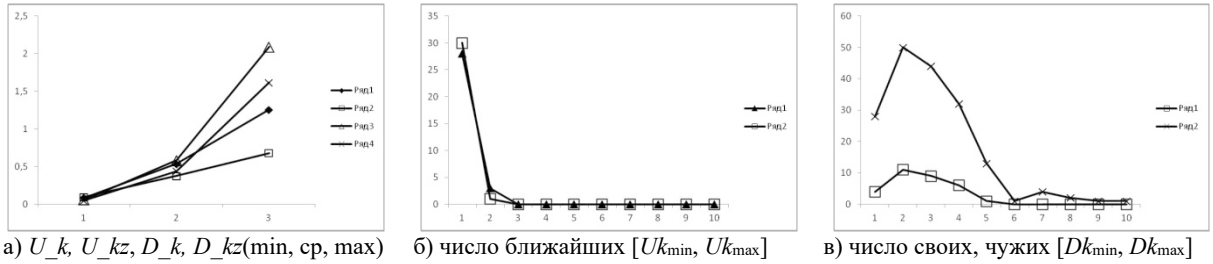
а) *U_k, U_kz, D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 12. Вариант 2. C^5 – эндокринная система



а) *U_k, U_kz, D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 13. Вариант 2. C^6 – ЦНС и органы чувствительности



а) *U_k, U_kz, D_k, D_kz*(min, ср, max) б) число ближайших [Uk_{min}, Uk_{max}] в) число своих, чужих [Dk_{min}, Dk_{max}]

Рис. 14. Вариант 2. C^7 – печень и желчевыводящие пути

3. Отклонение от центра масс своих и чужих элементов

Для каждого из рассматриваемых классов C^1, C^2, C^3, C^4 (**Вариант 1**) или C^1, C^2, C^3, C^4, C^5, C^6, C^7 (**Вариант 2**) в отдельности получим среднестатистический вектор длины 8, принадлежащий исходному векторному пространству $\mathbf{R}^8$. Иногда такой вектор называют центром масс.

Для центра масс k-го класса значение i–го параметра крови равно среднему арифметическому значений i-х параметров крови по всем J_k имеющимся в базе наборам показателей крови, относящихся к данному классу:

$$v_i^{k,cp} = \left(\sum_{j=1}^{J_k} v_i^{k,j}\right) / J_k , \qquad (19)$$

где $\mathbf{v}^{k,j}$ – перенумерованные элементы k-го класса: $\{\mathbf{v}^{k,j}=(v^{k,j}{}_1,\ldots,v^{k,j}{}_N), j = 1, \ldots, J_k\} = V^k$.

Для каждого из классов C^1, C^2, C^3, C^4 (**Вариант 1**) или C^1, C^2, C^3, C^4, C^5, C^6, C^7 (**Вариант 2**) найдем минимальное, максимальное и среднее расстояние между центром масс и своими векторами. Указанные величины для множества векторов k-го класса определяем следующим образом.

Минимальное расстояние:

$$D\_k_{\min} = \min_{V^k} \{\|\mathbf{v}^{k,cp} - \mathbf{u}^k\|\}, \mathbf{u}^k \in V^k \qquad (20)$$

Максимальное расстояние:

$$D\_k_{\max} = \max_{V^k} \{\|\mathbf{v}^{k,cp} - \mathbf{u}^k\|\}, \mathbf{u}^k \in V^k \qquad (21)$$

где $\mathbf{u}^k$ – вектор, принадлежащий множеству элементов k-го класса V^k, $\mathbf{v}^{k,cp}$ – среднестатистический вектор этого класса**.**

Среднее расстояние определим более детально с приведением алгоритма нахождения этой величины:

$$D\_k_{\text{ср}} = \sum_{j=1}^{J_k} \| \mathbf{w}^{k,j} - \mathbf{v}^{k,cp} \| / J_k , \mathbf{w}^{k,j} \in V^k,$$
$$j = 1, \ldots, J_k \qquad (22)$$

где $\{\mathbf{w}^{k,j}, j = 1, \ldots, J_k\} = V^k$ – представление совокупности элементов k-го класса в виде множества перенумерованных векторов.

Аналогично, получим соответствующие значения по каждому из классов между центром масс и чужими векторами. Эти результаты зависят от количества классов, входящих в обучающее множество.

Минимальное расстояние:

$$D\_kz_{\min} = \min_{V^{\neg k}} \{\|\mathbf{v}^{k,cp} - \mathbf{u}^{\neg k}\|\}, \mathbf{u}^{\neg k} \in V^{\neg k} \qquad (23)$$

Максимальное расстояние:

$$D\_kz_{\max} = \max_{V^{\neg k}} \{\|\mathbf{v}^{k,cp} - \mathbf{u}^{\neg k}\|\}, \mathbf{u}^{\neg k} \in V^{\neg k} \qquad (24)$$

где $\mathbf{u}^{\neg k}$ –вектор, принадлежащий множеству чужих элементов $V^{\neg k}$ классов, отличных от k-го: $V^{\neg k} = V \backslash V^k$, $\mathbf{v}^{k,cp}$ – среднестатистический вектор k-го класса**.**

Среднее расстояние:

$$D\_kz_{\text{ср}} = \sum_{j=1}^{J_{\neg k}} \| \mathbf{v}^{k,cp} - \mathbf{w}^{\neg k,j} \| / J_{\neg k} ,$$
$$\mathbf{w}^{\neg k,j} \in V^{\neg k}, j = 1, \ldots, J_{\neg k} \qquad (25)$$

где $\{\mathbf{w}^{\neg k,j}, j = 1, \ldots, J_{\neg k}\} = V^{\neg k}$, $V^{\neg k} = V \backslash V^k$ – представление совокупности чужих элементов классов, отличных от k-го, в виде множества перенумерованных векторов**.**

В метрике L_2 представлено (Рис. 1,а –14,а) минимальное, среднее и максимальное расстояние (значения ординат для точек 1, 2, 3 по оси абсцисс) между центром масс и своими векторами (Ряд 2), аналогично между парами центр масс – чужой вектор (Ряд 4).

Вариант 1. Для шести из семи СО Ряд 2 превышает Ряд 4 по средним и максимальным значениям (Рис. 1, а -7, а). Исключением является *урологическая система,* где для максимальной величины соотношение противоположное, но крайне незначительное (Рис. 4, а). Минимальные значения малы; по ним имеются оба вида различия (Рис. 1, а -7, а).

Вариант 2. Ряд 4 превышает Ряд 2 по средним и максимальным значениям (Рис. 8, в - 14, в). Имеются исключения, где соотношение противоположное по одной из величин: для максимальной – класс *C^2, органы дыхания* (Рис. 9, а), для средней – *C^4, урологическая система* (Рис. 11, а). Минимальные значения малы; по ним имеются оба типа различия (Рис. 8, а -14, а).

4. Ближайшие элементы. Распределения числа своих и чужих элементов

Перенумеруем все элементы k-го класса (**Вариант 1, 2**) по рассматриваемой базе: $\{\mathbf{w}^{k,j}, j = 1, \ldots, J_k\} = V^k$. Для каждого такого вектора $\mathbf{w}^{k,j}$ найдем расстояние до ближайшего элемента этого же класса (своего):

$$U\_k^j = \min_{V^k \setminus \mathbf{w}^{k,j}} \{\|\mathbf{w}^{k,j} - \mathbf{u}^k\|\},\ \mathbf{w}^{k,j} \in V^k, j = 1, \ldots, J_k,$$

$$\mathbf{u}^k \in \{V^k \setminus \mathbf{w}^{k,j}\} \qquad (26)$$

Заметим, что один и тот же вектор $\mathbf{u}^k$ может оказаться ближайшим более чем для одного элемента, например, для $\mathbf{w}^{k,j1}$ и $\mathbf{w}^{k,j2}$, где $j1 \neq j2$. В то же время, для элемента $\mathbf{w}^{k,j}$ ближайшими могут быть одновременно несколько различных равноудаленных от него векторов, например $\mathbf{v}^k \in \{V^k \setminus \mathbf{w}^{k,j}\}$, $\mathbf{u}^k \in \{V^k \setminus \mathbf{w}^{k,j}\}$, $U\_k^j = \|\mathbf{w}^{k,j} - \mathbf{v}^k\| = \|\mathbf{w}^{k,j} - \mathbf{u}^k\|$, $\mathbf{v}^k \neq \mathbf{u}^k$.

Аналогично, для каждого перенумерованного элемента k-го класса $\mathbf{w}^{k,j}$ найдем расстояние до ближайшего вектора, не принадлежащего этому классу (чужого):

$$U\_kz^j = \min_{V^{\neg k}} \{\|\mathbf{w}^{k,j} - \mathbf{u}^{\neg k}\|\},\ \mathbf{w}^{k,j} \in V^k,$$

$$j = 1, \ldots, J_k,\ \mathbf{u}^{\neg k} \in V^{\neg k} \qquad (27)$$

Заметим, что один и тот же вектор $\mathbf{u}^{\neg k}$ может оказаться ближайшим более чем для одного элемента, например, для $\mathbf{w}^{kj1}$ и $\mathbf{w}^{kj2}$, где $j1 \neq j2$. В то же время для элемента $\mathbf{w}^{kj}$ ближайшими могут быть одновременно несколько различных равноудаленных от него векторов, например $\mathbf{v}^{\neg k} \in V^{\neg k}$, $\mathbf{u}^{\neg k} \in V^{\neg k}$, $U\_kz^j = \|\mathbf{w}^{kj} - \mathbf{v}^{\neg k}\| = \|\mathbf{w}^{kj} - \mathbf{u}^{\neg k}\|$, $\mathbf{v}^{\neg k} \neq \mathbf{u}^{\neg k}$.

Диапазон расстояний между своими элементами k-го класса по рассматриваемой базе, согласно формулам (13), (14), находится на отрезке $[U\_k_{\min}, U\_k_{\max}]$. Диапазон расстояний между элементами k-го класса и векторами всех других классов (чужими), согласно формулам (16), (17) находится на отрезке $[U\_kz_{\min}, U\_kz_{\max}]$. Пусть

$$Uk_{\min} = \min(U\_k_{\min}, U\_kz_{\min})$$
$$Uk_{\max} = \max(U\_k_{\max}, U\_kz_{\max}) \qquad (28)$$

Делим отрезок $[Uk_{\min}, Uk_{\max}]$ (ось абсцисс на Рис. 1,б -14,б) на десять равных по длине частей – один отрезок и девять полуинтервалов: $[Uk_{\min}, Uk_{\min} + u]$, $(Uk_{\min} + u, Uk_{\min} + 2u]$, … , $(Uk_{\min} + 9u\ Uk_{\min} + 10u]$, где $u = (Uk_{\max} - Uk_{\min})/10$.

Определим для каждого j какой из десяти этих частей принадлежит величина U_k^j. Затем сосчитаем, какое количество перенумерованных векторов $\mathbf{w}^{k,j}$ попало в каждый такой участок, то есть имеет ближайший вектор из этого же класса, расстояние до которого находится на этой части отрезка $[Uk_{\min}, Uk_{\max}]$. Итак, мы получим распределение числа своих векторов $\mathbf{w}^{k,j}$ по расстоянию до ближайшего своего вектора (Ряд 1) на отрезке $[Uk_{\min}, Uk_{\max}]$ (Рис. 1, б -14, б).

Аналогично, определим для каждого j к какой из десяти частей отрезка $[Uk_{\min}, Uk_{\max}]$ относится величина U_kz^j. Посчитаем сколько векторов $\mathbf{w}^{k,j}$ имеет ближайший вектор, не принадлежащий этому классу, расстояние до которого находится на этой части отрезка $[Uk_{\min}, Uk_{\max}]$. Получим распределение числа своих векторов $\mathbf{w}^{k,j}$ по расстоянию до ближайшего чужого вектора (Ряд 2) на отрезке $[Uk_{\min}, Uk_{\max}]$ (Рис. 1, б -14, б).

Вариант 1. Ряд 1 и Ряд 2 в целом убывают на отрезке $[Uk_{\min}, Uk_{\max}]$, причем до его конца достигают нуля (Рис. 1, б -7, б). В начальной точке имеется максимум, значение которого для расстояний между своими элементами (Ряд 1) больше, чем между своими и чужими (Ряд 2). Исключением является *Эндокринная система* (Рис. 5, б), но отклонение от общей закономерности компенсируется с учетом данных в соседней точке.

Вариант 2. Ряд 1 и Ряд 2 в целом убывают на отрезке $[Uk_{\min}, Uk_{\max}]$ так, что до его конца принимают нулевое значение (Рис. 8, б -14, б). В начальной точке имеется максимум, причем Ряд 2 превышает Ряд 1.

5. Распределение числа своих и чужих элементов при удалении от центра масс

Диапазон расстояний между центром масс k-го класса СО и векторами этого же класса («своими», $\mathbf{v}^k \in V^k$) по рассматриваемой базе, согласно формулам (20), (21), находится на отрезке $[D\_k_{\min}, D\_k_{\max}]$. Диапазон расстояний между центром масс k-го класса СО и векторами всех других классов («чужими», $\mathbf{z}^k \in \{V \setminus V^k\}$), согласно формулам (23), (24), – на отрезке $[D\_kz_{\min}, D\_kz_{\max}]$. Пусть

$$Dk_{\min} = \min(D\_k_{\min}, D\_kz_{\min})$$
$$Dk_{\max} = \max(D\_k_{\max}, D\_kz_{\max}) \quad (29)$$

Делим отрезок $[Dk_{\min}, Dk_{\max}]$ (оси абсцисс на Рис. 1, в – 14, в) на десять равных по длине частей – один отрезок и девять полуинтервалов: $[Dk_{\min}, Dk_{\min} + d]$, $(Dk_{\min} + d, Dk_{\min} + 2d]$, … , $(Dk_{\min} + 9d, Dk_{\min} + 10d]$, где $d = (Dk_{\max} - Dk_{\min})/10$. Определим, какое количество своих векторов попало в каждый такой участок (аналогично для чужих векторов). Затем рассмотрим распределение числа своих (Ряд 1) и чужих (Ряд 2) векторов на отрезке $[Dk_{\min}, Dk_{\max}]$.

Вариант 1. Ряд 1 и Ряд 2 в целом нарастают (либо в начале имеется максимум), затем убывают при удалении от центра масс. До конца отрезка $[Dk_{\min}, Dk_{\max}]$ присутствуют свои элементы, а чужие – не всегда (Рис. 1, в -7, в).

Вариант 2. Ряд 1 и Ряд 2 в целом нарастают, затем убывают при удалении от центра масс. В правой части отрезка $[Dk_{\min}, Dk_{\max}]$ свои элементы либо отсутствуют, либо преимущественно преобладают чужие (Рис. 8, в -14, в).

Заключение

Для предварительной медицинской диагностики по показателям периферической крови разработаны два типа классификаторов **(Вариант 1, 2)** с использованием статистического метода распознавания, основанного на полиномиальной регрессии.

Представлены семь СО для мужчин: пищеварения, дыхания, урологическая, эндокринная, опорно-двигательного аппарата, печени и желчевыводящих путей, ЦНС и органов чувствительности.

Вариант 1 служит для оценивания состояния здоровья пациента по четырем градациям – от практически здорового до максимальной степени поражения организма (онкология) В этом случае каждая СО рассматривается по отдельности. **Вариант 2** позволяет уточнить, в какой СО локализована онкологическая опухоль.

Мы предложили и формализовали подход к изучению обучающих множеств, а также продемонстрировали его на медицинских диагностических классификаторах.

Проведено исследование структуры обучающего множества, состоящего из четырех или семи классов соответственно. Для **Варианта 1** – это данные анализа крови пациентов, разделенные по четырем градациям СЗЧ отдельно для каждой СО. Для **Варианта 2** – наборы крови онкобольных, имеющих патологию в одной из семи СО.

Объектом изучения является совокупность наборов показателей крови онкологических больных, имеющих патологию в некоторой заданной СО. Она представляет собой класс «4» в обучающем множестве классификатора данной СО **(Вариант 1)**, а также один из семи классов, соответствующих рассматриваемым СО **(Вариант 2)**.

Используется следующая терминология.

Вариант 1. Для каждой из СО обозначено: «свой» – класс C^4, «чужой» – $C^1UC^2UC^3$

Вариант 2. «Свой» – класс, соответствующий данной СО, «чужой» – совокупность остальных СО.

Найдено минимальное, максимальное и среднее расстояние между своими векторами (принадлежащими данному классу). Также получены соответствующие значения для пар свой–чужой (элемент, не относящийся к рассматриваемому классу).

Для каждого из перенумерованных своих элементов определено минимальное расстояние до своих (чужих) векторов. Получено распределение числа векторов по степени близости к своим и чужим элементам.

Вычислили среднестатистический вектор, принадлежащий исходному векторному пространству $\mathbf{R}^8$ (центр масс). Найдено минимальное, максимальное и среднее расстояние между центром масс и своими (чужими) векторами.

Получено распределение количества своих и чужих элементов при удалении от центра масс.

Проведенный анализ позволил выявить сходство и различие в структуре обучающих множеств **Варианта 1** и **Варианта 2**.

Показано, что необходимость использования более сложных полиномов и соответственно увеличение его длины для **Варианта 2** по сравнению с **Вариантом 1** коррелируется с особенностями распределения числа векторов по расстоянию до ближайших своих и чужих элементов.

Литература

1. Ставицкий Р. В., Лебедев Л. А., Лебедев А. Л., Смыслов А. Ю. Количественная оценка гомеостатической

активности здоровых и больных людей. - М.: ГАРТ. 2013. 131 с.
2. Гавриков Б. М., Лебеденко И. М., Пестрякова Н. В., Ставицкий Р. В. Об одном статистическом методе оценивания состояния здоровья человека // Труды ИСА РАН, 2016. Т. 66. № 2. С. 54-59.
3. Гавриков Б. М., Пестрякова Н. В. О построении признакового пространства в задаче обучения // Информационные технологии и вычислительные системы. 2018. №1. С. 22-29. DOI: 10.14357/20718632180104
4. Гавриков Б. М., Пестрякова Н. В., Ставицкий Р. В. О свойствах обучающих множеств // Информационные технологии и вычислительные системы. 2018. №4. С.97-107. DOI: 10.14357/207186321804010
5. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В. Статистический метод распознавания на основе нелинейной регрессии // Математическое моделирование. 2020. Т.32. №4. С.116-130. DOI: 0.20948/mm-2020-04-09
6. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В. О способности статистического классификатора к обобщениям// Информационные технологии и вычислительные системы. 2021. № 4. С. 38-50. DOI: 10.14357/20718632210404.
7. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В. О структуре базы обучения классификатора для оценивания состояния здоровья человека // Препринты ИПМ им. М. В. Келдыша. 2018. №126. 18с. DOI:10.20948/prepr-2018-126
8. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В., Ставицкий Р. В. Структура базы обучения статистического классификатора состояний систем организма человека // Препринты ИПМ им. М. В. Келдыша. 2018. №255. 40с. DOI:10.20948/prepr-2018-255
9. Гавриков Б. М., Пестрякова Н. В. Статистический классификатор для диагностики онкологических заболеваний // Информационные технологии и вычислительные системы. 2023. № 1. С. 39-49.
10. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В. Статистический подход для диагностики онкологических заболеваний по параметрам крови // Препринты ИПМ им. М. В. Келдыша. 2022. № 72. 12 с. DOI:10.20948/prepr-2022-72
11. Гавриков Б. М., Гавриков М. Б., Пестрякова Н. В. Прототип системы поддержки принятия решений в медицинской диагностике на основе статистического подхода // Препринты ИПМ им. М.В. Келдыша. – 2022. – № 76. – 23 с. DOI: 10.20948/prepr-2022-76.
12. Гавриков Б. М., Гавриков М. Б., Лебеденко И. М., Пестрякова Н. В. Нахождение области локализации онкопатологии по параметрам крови больного // Препринты ИПМ им. М. В. Келдыша. 2023. №24. 15 с. DOI: 10.20948/prepr-2023-24.
13. Гавриков Б. М., Гавриков М. Б., Лебеденко И. М., Пестрякова Н. В. Статистический классификатор для определения области локализации онкопатологии по анализу крови больного // Препринты ИПМ им. М.В. Келдыша. 2024. №22. 15 с. DOI: 10.20948/prepr-2024-22.
14. Гавриков М. Б., Локуциевский О. В. Начала численного анализа. – М.: Янус, 1995.
15. Schürmann J. Pattern Classification. – New York: John Wiley&Sons, Inc., 1996.

Гавриков Борис Михайлович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), Москва, Россия Аспирант. Область научных интересов: вычислительная математика, распознавание образов, медицинская физика. E-mail: bmgavrikov@gmail.com

Пестрякова Надежда Владимировна. Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), Москва, Россия. Ведущий научный сотрудник, доктор технических наук. Область научных интересов: вычислительная математика и физика, распознавание образов. E-mail: pestryakova@isa.ru

Diagnostic Medical Classifiers

B. M. Gavrikov, N. V. Pestryakova

Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

Abstract. For the preliminary diagnosis of human diseases based on peripheral blood analysis, a statistical classification approach based on polynomial regression has been developed. It is implemented as two applications. One type of classifier allows you to assess the patient’s condition from healthy to the maximum degree of damage (oncology) within a separate body system. Using a different type of classifier, the area of tumor localization in cancer is identified. For various classification options, the structure of the training set is examined and a comparative analysis is carried out.

Keywords: cancer, body system, peripheral blood, classification, polynomial regression, learning set.

DOI 10.14357/20718632240302 **EDN** BTQWIO

References

1. Stavitskii R.V., Lebedev L.A., Lebedev A.L., Smyslov A.IU. Kolichestvennaia otsenka gomeostaticheskoi aktivnosti zdorovykh i bolnykh liudei - M.: GART. 2013. 131 s.
2. Gavrikov B.M., Lebedenko I.M., Pestryakova N.V., Stavitskiy R.V. Ob odnom statisticheskom metode otsenivaniya sostoyaniya zdorov'ya cheloveka. // Trudy ISA RAN, 2016. T. 66. № 2. S. 54-59.
3. Gavrikov B.M., Pestryakova N.V. O postroyenii priznakovogo prostranstva v zadache obucheniya // Informatsionnyye tekhnologii i vychislitel'nyye sistemy. 2018. №1. S. 22-29. DOI: 10.14357/20718632180104
4. Gavrikov B.M., Pestryakova N.V., Stavitskiy R.V. O svoystvakh obuchayushchikh mnozhestv // Informatsionnyye tekhnologii i vychislitel'nyye sistemy. 2018. №4. S.97-107. DOI: 10.14357/207186321804010
5. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V. Statisticheskiy metod raspoznavaniya na osnove nelineynoy regressii. // Matematicheskoye modelirovaniye. 2020. T.32. №4. S.116-130. DOI: 0.20948/mm-2020-04-09
6. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V. O sposobnosti statisticheskogo klassifikatora k obobshcheniyam// Informatsionnyye tekhnologii i vychislitel'nyye sistemy. 2021. № 4. S. 38-50. DOI: 10.14357/20718632210404.
7. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V. O strukture bazy obucheniya klassifikatora dlya otsenivaniya sostoyaniya zdorov'ya cheloveka // Preprinty IPM im. M.V.Keldysha. 2018. №126. 18s. DOI:10.20948/prepr-2018-126
8. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V., Stavitskiy R.V. Struktura bazy obucheniya statisticheskogo klassifikatora sostoyaniy sistem organizma cheloveka // Preprinty IPM im. M.V.Keldysha. 2018. №255. 40s. DOI:10.20948/prepr-2018-255
9. Gavrikov B.M., Pestryakova N.V.Statisticheskiy klassifikator dlya diagnostiki onkologicheskikh zabolevaniy // Informatsionnyye tekhnologii i vychislitel'nyye sistemy. 2023. № 1. S. 39-49.
10. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V.Statisticheskiy podkhod k diagnostike onkologicheskikh zabolevaniy po parametram krovi // Preprinty IPM im. M.V.Keldysha. 2022. № 72. 12 s. DOI:10.20948/prepr-2022-72
11. Gavrikov B.M., Gavrikov M.B., Pestryakova N.V. Prototip sistemy podderzhki prinyatiya resheniy v meditsinskoy diagnostike na osnove statisticheskogo podkhoda // Preprinty IPM im. M.V.Keldysha. 2022. № 76. 23 s. DOI: 10.20948/prepr-2022-76.
12. B.M. Gavrikov, Gavrikov M.B., I.M. Lebedenko, N.V. Pestryakova. Nakhozhdeniye oblasti lokalizatsii onkopatologii po parametram krovi bol'nogo // Preprinty IPM im. M.V.Keldysha. 2023. №24. 15 s. DOI: 10.20948/prepr-2023-24.
13. B.M. Gavrikov, Gavrikov M.B., I.M. Lebedenko, N.V. Pestryakova. Statisticheskiy klassifikator dlya opredeleniya oblasti lokalizatsii onkopatologii po analizu krovi bol'nogo // Preprinty IPM im. M.V.Keldysha. 2024. №22. 15 s. DOI: 10.20948/prepr-2024-22.
14. Gavrikov M.B., Lokutsiyevskiy O.V. Nachala chislennogo analiza. — M.: Yanus, 1995.
15. Schürmann J. Pattern Slassification. — New York: John Wiley&Sons, Inc. 1996.

Gavrikov Boris M. Federal Research Center Computer Science and Control of Russian Academy of Sciences, Moscow, Russia. E-mail: bmgavrikov@gmail.com

Pestryakova Nadia V. Doctor of Technical Sciences, PhD in Physics and Mathematics. Federal Research Center Computer Science and Control of Russian Academy of Sciences, Moscow, Russia. E-mail: pestryakova@isa.ru