



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 24 за 2023 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

**Б.М. Гавриков, М.Б. Гавриков,
И.М. Лебеденко, Н.В. Пестрякова**

Нахождение области
локализации онкопатологии
по параметрам крови
больного

Статья доступна по лицензии
Creative Commons Attribution 4.0 International



Рекомендуемая форма библиографической ссылки: Нахождение области локализации онкопатологии по параметрам крови больного / Б.М. Гавриков [и др.] // Препринты ИПМ им. М.В.Келдыша. 2023. № 24. 14 с. <https://doi.org/10.20948/prepr-2023-24>
<https://library.keldysh.ru/preprint.asp?id=2023-24>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

**Б.М. Гавриков, М.Б. Гавриков, И.М. Лебедеико,
Н.В. Пестрякова**

**Нахождение области локализации
онкопатологии по параметрам крови
больного**

Москва — 2023

Б.М. Гавриков, М.Б. Гавриков, И.М. Лебеденко, Н.В. Пестрякова

Нахождение области локализации онкопатологии по параметрам крови больного

Описывается математический подход, позволяющий по данным анализа крови онкопациента установить местонахождение опухоли. Используется разработанный авторами статистический классификатор, основанный на полиномиальной регрессии и имеющий вероятностные оценки. Обучение классификатора осуществлялось на базах периферической крови онкологических больных по пяти различным системам организма.

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

Boris Mikhailovich Gavrikov, Mikhail Borisovich Gavrikov, Irina Matveevna Lebedenko, Nadejda Vladimirovna Pestryakova

Finding the area of localization of oncopathology according to the parameters of the patient's blood

A mathematical approach is described that allows, according to the blood test of a cancer patient, to locate the tumor. A statistical classifier developed by the authors is used, based on polynomial regression and having probabilistic estimates. The classifier was trained on the bases of peripheral blood of oncological patients in five different body systems.

Key words: cancer, body system, peripheral blood, classification, polynomial regression, training set

Оглавление

Введение.....	3
Метод классификации.....	4
Расстояние между «своими» и «чужими» элементами.....	6
Отклонение от центра масс своих и чужих элементов.....	8
Распределение числа своих и чужих элементов при удалении от центра масс.....	10
Заключение.....	12
Библиографический список.....	13

Введение

В основе описываемого в данной работе исследования лежит концепция крупнейших гематологов, согласно которой многие заболевания человека вносят изменения в состав его крови. При оценке состояния здоровья человека (СЗЧ) гематологи предлагают использовать не менее пяти показателей периферической крови (из пальца) [1].

СЗЧ включает четыре градации: 1) практически здоровые, 2) начальные отклонения состояния здоровья, 3) выраженные изменения, 4) тяжелое заболевание (онкология).

Авторы ранее показали, что статистический метод классификации может успешно применяться для оценивания СЗЧ по данным лабораторного анализа периферической крови [2-8]. В рамках используемой модели организм представляется как совокупность систем организма (СО) – пищеварения, дыхания, урологической, эндокринной, гинекологической (для женщин), опорно-двигательного аппарата, печени и желчевыводящих путей, грудных желез (для женщин), центральной нервной системы и органов чувствительности.

Для каждой СО оценивание СЗЧ проводилось при помощи профильного классификатора, обученного на соответствующей выборке. Выборки включают наборы из восьми параметров периферической крови. Они были созданы на основе верифицированных диагнозов, поставленных большому числу пациентов.

Для мужчин и женщин построены и обучены различные классификаторы, поскольку диапазоны вариации показателей крови существенно зависят от пола.

Онкология, относящаяся к четвертой градации СЗЧ, является одним из наиболее опасных и коварных заболеваний. При диагностике не исключены ошибки относительно местонахождения онкологической опухоли. Аналогичная проблема может возникнуть, если классификатор СЗЧ для некоторого анализа крови выдал на выходе именно этот класс здоровья для более чем одной СО.

Авторы задались вопросом: можно ли по анализу крови онкологического больного определить область локализации патологии?

Очевидного ответа на него нет. Для практикующих врачей изучение этой проблемы является весьма трудоемким и математически сложным, поскольку требуется собрать базы параметров крови пациентов с онкологическими опухолями различной локализации, а затем провести их сравнительный анализ.

В настоящей работе рассматривается подход, позволивший решить поставленную задачу. Для этого обучение статистического классификатора проводили на наборах параметров периферической крови четвертого класса СЗЧ (онкологических больных), относящихся к пяти различным СО.

Проводить классификацию по всем СО нет необходимости; допустимо ограничиться только наиболее проблемными.

Вполне естественными представляются еще два вопроса, возникшие у авторов. Как именно соотносятся между собой фрагменты обучающей выборки, относящиеся к каждому из классов? Как можно визуализировать векторные объекты восьмимерного пространства (по числу параметров крови)? Ответом на них является описанный оригинальный способ исследования структуры обучающей выборки, а также полученные результаты.

Метод классификации

Общепринятые обозначения и размерность используемых восьми показателей крови следующие: RBC [L⁻¹] – эритроциты, HGB [gL⁻¹] – гемоглобин, PLT [L⁻¹] – тромбоциты, WBC [L⁻¹] – лейкоциты, LIMPН [L⁻¹], [%] – лимфоциты, GRAN [L⁻¹], [%] – гранулоциты (GRAN=NEUT+EOS+BASO, где NEUT[L⁻¹], [%] – нейтрофилы, EOS[L⁻¹], [%] – эозинофилы, BASO[L⁻¹], [%] – базофилы).

Рассматриваем K определенных перенумерованных СО человека, $1 \leq k \leq K$, $K=5$. Вводим вектор $\mathbf{v} \in \mathbb{R}^N$, i -я компонента которого – отнормированная на отрезок, не выходящий за пределы $[0,1]$, величина i -го показателя крови онкобольных, где $N=8$.

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

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

$$p_r(\mathbf{v}) = \max_k \{p_k(\mathbf{v})\}, 1 \leq k \leq K. \quad (1)$$

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

$$p_k(\mathbf{v}) \cong c_0^{(k)} + \sum_{i=1}^N c_i^{(k)} v_i + \sum_{i,j=1}^N c_{i,j}^{(k)} v_i v_j + \dots, 1 \leq k \leq K. \quad (2)$$

Представим упорядоченные базисные мономы из (2) в виде вектора размерности L :

$$\mathbf{x}(\mathbf{v}) = (1, v_1, \dots, v_N, \dots)^T.$$

Тогда (2) можно записать в векторном виде:

$$\mathbf{p}(\mathbf{v}) = (p_1(\mathbf{v}), \dots, p_K(\mathbf{v}))^T \cong A^T \mathbf{x}(\mathbf{v}), \quad (3)$$

где A – матрица размера $L \times K$, столбцами которой являются векторы $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(K)}$. Каждый такой вектор составлен из коэффициентов при мономах соответствующей строки (2) (с совпадающим верхним индексом), упорядоченных так же, как в векторе $\mathbf{x}(\mathbf{v})$.

Значение A вычисляется приближенно в процессе обучения с использованием последовательности данных: $[\mathbf{v}^{(1)}, \mathbf{y}^{(1)}], \dots, [\mathbf{v}^{(J)}, \mathbf{y}^{(J)}]$. Здесь $\mathbf{v}^{(j)}$ – набор параметров крови, соответствующий элементу СО с номером k ($1 \leq k \leq K$), $\mathbf{y}^{(j)} = (0 \dots 1 \dots 0)$ – его базисный вектор, где 1 стоит на k -м месте, $1 \leq j \leq J$):

$$A \cong \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{x}^{(j)})^T \right)^{-1} \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{y}^{(j)})^T \right). \quad (4)$$

Поскольку проблема обращения заполненной матрицы большой размерности до сих пор не решена [9], правую часть (4) получаем посредством рекуррентной процедуры [10].

В данной работе рассмотрено пять СО для мужчин: пищеварительная система (C^1), органы дыхания (C^2), опорно-двигательный аппарат (C^3), урологическая система (C^4), эндокринная система (C^5).

Нормирование проводим следующим образом. Рассмотрим совокупность пяти обучающих выборок исследуемых СО. Для каждого i -го показателя крови находим минимальное и максимальное значение v_i^{\min}, v_i^{\max} , где $i = 1, \dots, N$.

$$v_i^{\min} = \min_j \{v_i^j\}, j=1, \dots, J,$$

$$v_i^{\max} = \max_j \{v_i^j\}, j=1, \dots, J,$$

где J – общий объем выборки по всем рассматриваемым СО.

Затем выполняем преобразование:

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

Использовалась следующая модификация $\mathbf{x}(\mathbf{v})$:

$$\begin{aligned} \mathbf{x} = & (1, \{v_i\}, \{v_i v_j\}, \{v_i v_j v_k\}, \{v_i v_j v_k v_l\}, \{v_i v_j v_k v_l v_m\}, \\ & \{v_i v_j v_k v_l v_m v_n\}, \{v_i v_j v_k v_l v_m v_n v_p\}), \end{aligned} \quad (5)$$

$$1 \leq i \leq 8, i \leq j \leq 8, j \leq k \leq 8, k \leq l \leq 8, l \leq m \leq 8, m \leq n \leq 8, n \leq p \leq 8.$$

В (5) выражения в фигурных скобках соответствуют цепочкам элементов вектора, вычисляемым по всем показателям крови из имеющегося набора. 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.

Исследование проведено на классификаторе из пяти классов. Обучающее множество $C^1UC^2UC^3UC^4UC^5$ имеет следующий объем (количество наборов

крови), дифференцированный по классам: $|C^1| = |C^3| = |C^5| = 33$, $|C^2| = 21$, $|C^4| = 32$.

Классификатор обеспечил точность более 94% на обучающем множестве из 152 элементов (остается 10 ошибок).

Проведем анализ структуры обучающего множества $C^1UC^2UC^3UC^4UC^5$ [4, 8, 11, 12].

Расстояние между «своими» и «чужими» элементами

Для каждого из пяти рассматриваемых классов C^1 , C^2 , C^3 , C^4 , C^5 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:

$$U_{k_{\min}} = \min_{V^k} \{ \|\mathbf{v}^k - \mathbf{u}^k\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^k \in V^k, \mathbf{v}^k \neq \mathbf{u}^k. \quad (6)$$

Максимальное расстояние:

$$U_{k_{\max}} = \max_{V^k} \{ \|\mathbf{v}^k - \mathbf{u}^k\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^k \in V^k, \quad (7)$$

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:

$$U_{k_{\text{cp}}} = \sum_{j=1}^{J_k} \sum_{j_1=j+1}^{J_k} \|\mathbf{w}^{k,j} - \mathbf{w}^{k,j_1}\| / (J_k (J_k - 1) / 2), \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k, \quad (8)$$

где $\{\mathbf{w}^{k,j}, j = 1, \dots, J_k\} = V^k$ – 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. Для обучающего множества, содержащего элементы всех пяти классов, $V = \{C^1UC^2UC^3UC^4UC^5\}$.

Минимальное расстояние:

$$U_{kZ_{\min}} = \min_V \{ \|\mathbf{v}^k - \mathbf{u}^{-k}\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^{-k} \in V^{-k}. \quad (9)$$

Максимальное расстояние:

$$U_{kz_{\max}} = \max_V \{ \|\mathbf{v}^k - \mathbf{u}^{-k}\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^{-k} \in V^{-k}, \quad (10)$$

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

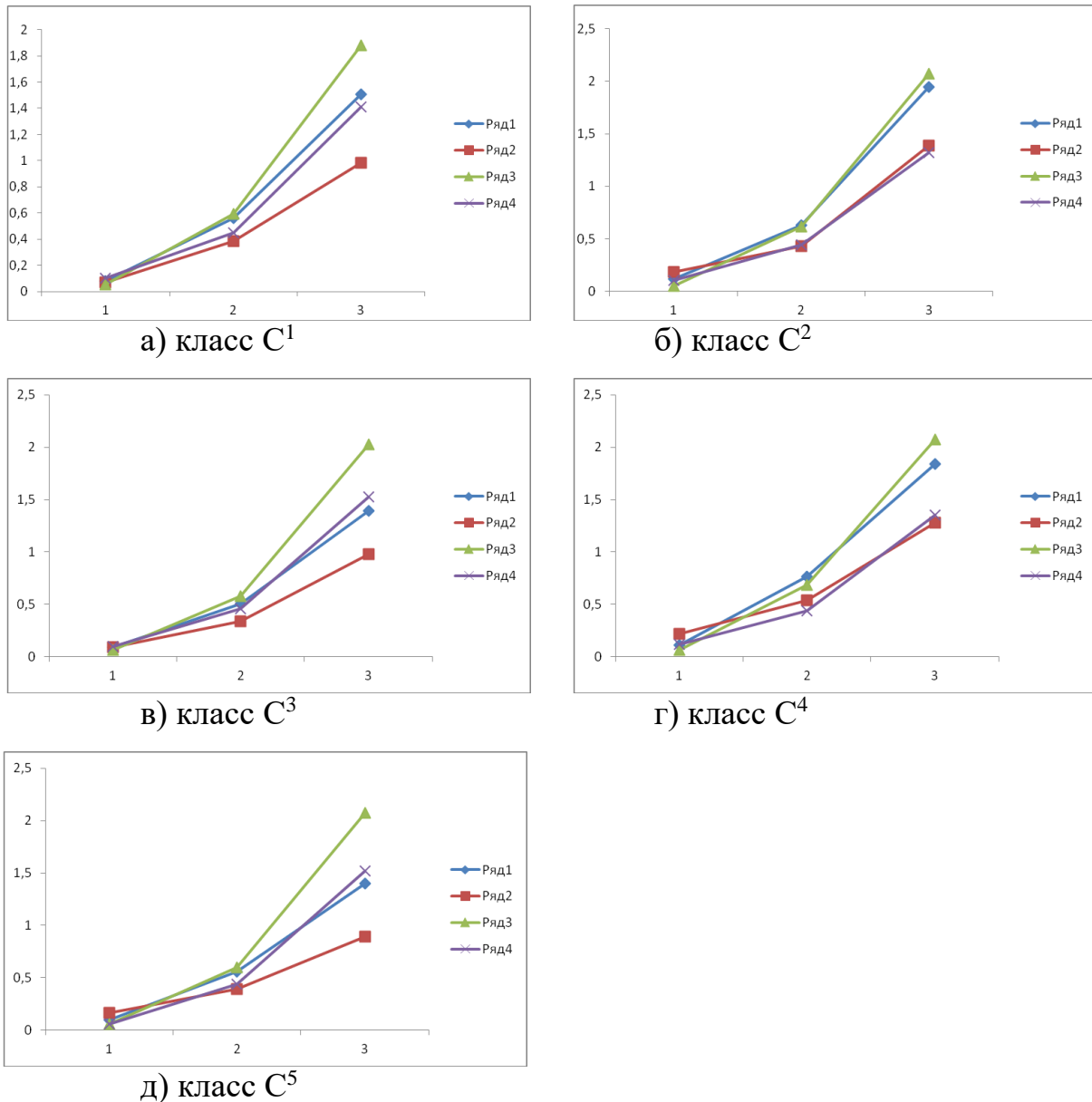


Рис. 1. Минимальное, максимальное и среднее расстояние между парами векторов: свой–свой, свой–чужой, центр масс–свой, центр масс–чужой.

Среднее расстояние:

$$U_{kzcp} = \sum_{j=1}^{J_k} \sum_{j1=1}^{J_{-k}} \| \mathbf{w}^{k,j} - \mathbf{w}^{-k,j1} \| / (J_k J_{-k}), \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k, \quad (11)$$

$$\mathbf{w}^{-k,j1} \in V^{-k}, j1 = 1, \dots, J_{-k},$$

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

Продемонстрируем, какие значения принимают перечисленные величины. Расстояние между векторами определяем в метрике L_2 .

Для классов C^1, C^2, C^3, C^4 и C^5 соответственно представлены (рис. 1а, б, в, г, д) 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).

Для всех пяти классов (рис. 1) Ряд 1 немного превышает Ряд 3 по минимальным значениям (причем оба малы). Первая величина меньше второй для максимальных значений (как существенно, так и незначительно). Для средних имеются оба варианта различия (небольшого).

Аналогичное исследование приведено для наборов из двух и трех классов [12]; в [11] рассматривалась совокупность из четырех классов. Сопоставление полученных результатов показывает их сходство.

Отклонение от центра масс своих и чужих элементов

Для каждого из пяти рассматриваемых классов C^1, C^2, C^3, C^4, C^5 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:

$$v_i^{k,cp} = \left(\sum_{j=1}^{J_k} v_i^{k,j} \right) / J_k, \quad (12)$$

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

Для каждого из классов C^1, C^2, C^3, C^4 и C^5 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:

$$D_{k_{\min}} = \min_{V^k} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^k\| \}, \mathbf{u}^k \in V^k. \quad (13)$$

Максимальное расстояние:

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

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:

$$D_{k_{cp}} = \sum_{j=1}^{J_k} \|\mathbf{w}^{k,j} - \mathbf{v}^{k,cp}\| / J_k, \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k, \quad (15)$$

где $\{\mathbf{w}^{k,j}, j = 1, \dots, J_k\} = V^k$ – 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:

$$D_{kz_{\min}} = \min_{V^{-k}} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^{-k}\| \}, \mathbf{u}^{-k} \in V^{-k}. \quad (16)$$

Максимальное расстояние:

$$D_{kz_{\max}} = \max_{V^{-k}} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^{-k}\| \}, \mathbf{u}^{-k} \in V^{-k}, \quad (17)$$

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

Среднее расстояние:

$$D_{kz_{cp}} = \sum_{j=1}^{J_{-k}} \|\mathbf{v}^{k,cp} - \mathbf{w}^{-k,j}\| / J_{-k}, \mathbf{w}^{-k,j} \in V^{-k}, j = 1, \dots, J_{-k}, \quad (18)$$

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

Для каждого класса соответственно (рис. 1) из объединения $C^1UC^2UC^3UC^4UC^5$ представлено 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).

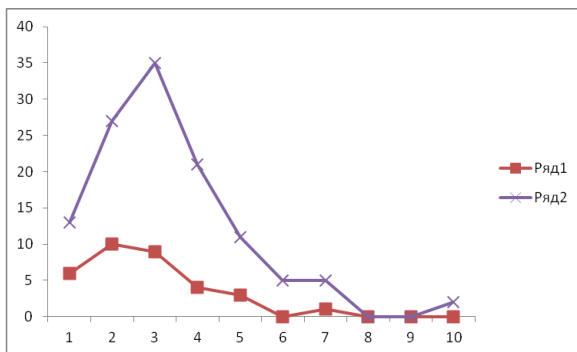
Для классов C^1 и C^3 (рис. 1а, в) Ряд 4 всюду превышает Ряд 2. Аналогичная картина наблюдается для класса C^5 , исключая минимальные значения (рис. 1д). При рассмотрении этих трех классов для минимальных и средних значений разница небольшая, а для максимальных – существенная. Для C^2 имеется такое же соотношение по средним значениям и противоположное – по минимальным и максимальным, причем везде Ряд 2 и Ряд 4 отличаются мало (рис. 1б). В классе C^4 (рис. 1г) Ряд 2 немного превышает Ряд 4 для минимальных и средних значений, а для максимальных – ситуация обратная, но различие незначительное.

Для рассмотренных ранее наборов из двух и трех классов [12], а также из четырех классов [11] получены аналогичные результаты.

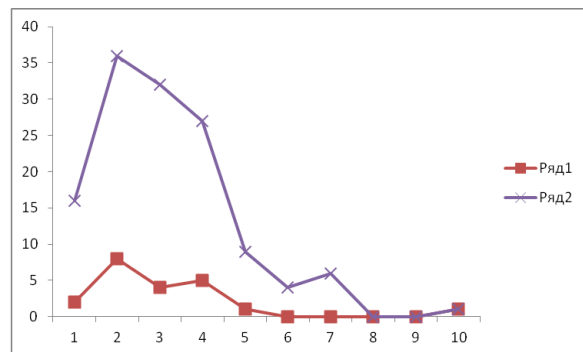
Распределение числа своих и чужих элементов при удалении от центра масс

Диапазон расстояний между центром масс k -го класса CO и векторами этого же класса («своими», $\mathbf{v}^k \in V^k$) по рассматриваемой базе, согласно формулам (13), (14), находится на отрезке $[D_{k_{\min}}, D_{k_{\max}}]$. Диапазон расстояний между центром масс k -го класса CO и векторами всех других классов («чужими», $\mathbf{z}^k \in \{V \setminus V^k\}$), согласно формулам (16), (17), – на отрезке $[D_{kz_{\min}}, D_{kz_{\max}}]$. Пусть

$$\begin{aligned} Dk_{\min} &= \min(D_{k_{\min}}, D_{kz_{\min}}), \\ Dk_{\max} &= \max(D_{k_{\max}}, D_{kz_{\max}}). \end{aligned} \quad (19)$$



а) класс C^1



б) класс C^2

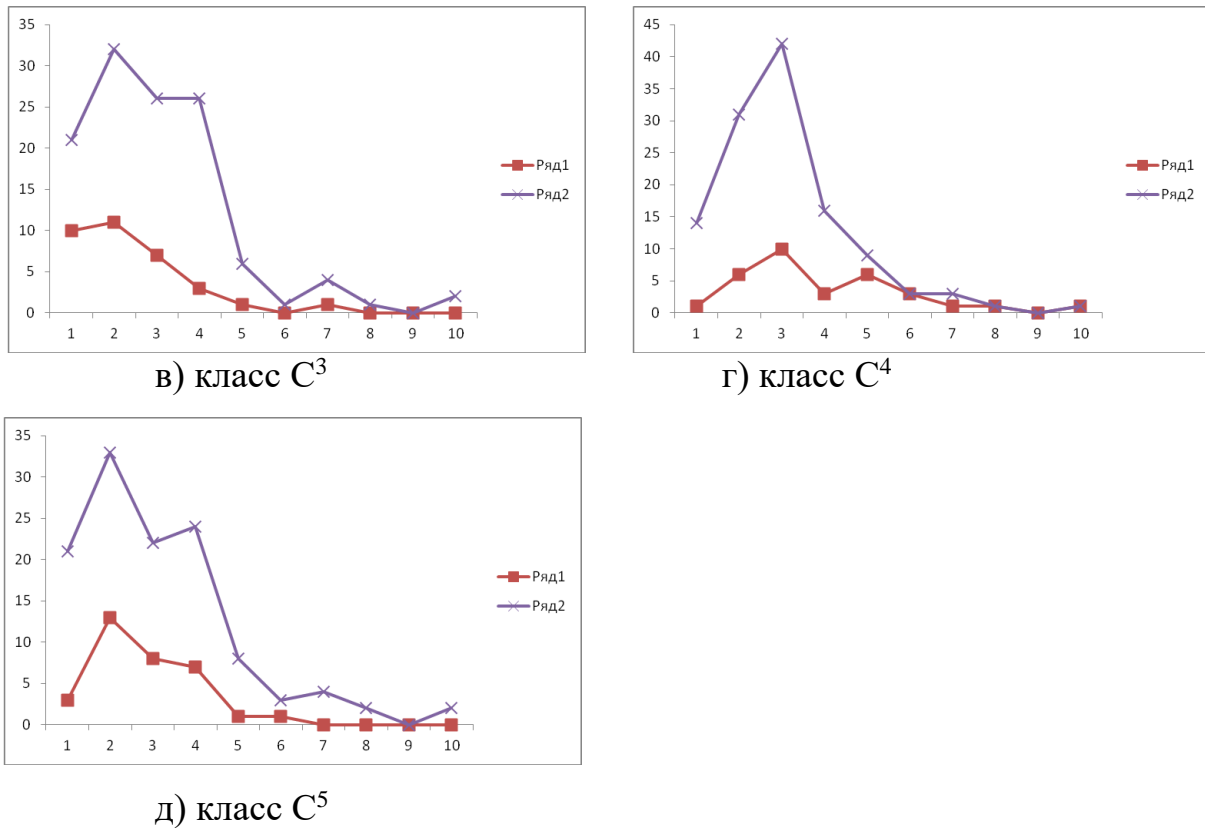


Рис. 2. Распределение числа своих и чужих элементов при удалении от центра масс ($C^1UC^2UC^3UC^4UC^5$)

Делим отрезок $[Dk_{\min}, Dk_{\max}]$ (оси абсцисс на рис. 2а, б, в, г, д) 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 (чужих) векторов на отрезке $[Dk_{\min}, Dk_{\max}]$.

Для каждого класса (рис. 2) соответственно из объединения $C^1UC^2UC^3UC^4UC^5$ представлено распределение количества своих (Ряд 1) и чужих (Ряд 2) элементов на отрезке $[Dk_{\min}, Dk_{\max}]$.

В характере полученных распределений числа своих (чужих) элементов имеется структурное сходство между всеми пятью классами. Эта же аналогия наблюдается в отношении распределений своих–чужих элементов. Отметим, что для классов C^1, C^3, C^5 получены результаты (рис. 2а, в, д), отличающиеся от C^2, C^4 (рис. 2б, г). Для C^2, C^4 (рис. 2б, г) свои элементы имеются до конца отрезка $[Dk_{\min}, Dk_{\max}]$, в отличие от C^1, C^3 и C^5 (рис. 2а, в, д).

Схожие результаты были получены для множеств, состоящих из двух, трех, а также четырех классов [11, 12].

Более того, можно убедиться в сходстве структуры обучающих множеств [4, 8, 11, 12] классификаторов двух типов: 1) вычисляющих СО – описанных в данной работе и в [11, 12], 2) находящих СЗЧ [2-8].

Заключение

Предложен классификатор, с помощью которого по показателям периферической крови онкологического пациента можно конкретизировать область локализации патологии. Разработанный подход использует статистический метод распознавания, основанный на полиномиальной регрессии.

Полиномиальный вектор имеет 6435 элементов, представляющих собой мономы до седьмого порядка. Он построен и обучен для пяти классов (C^1 , C^2 , C^3 , C^4 , C^5), соответствующих следующим СО для мужчин: пищеварительная, урологическая и эндокринная системы, органы дыхания, опорно-двигательный аппарат.

На обучающем множестве из 152 наборов крови точность классификатора составляет более 94% (остается 10 ошибок).

Проведено исследование структуры обучающего множества.

Для каждого из пяти классов C^1 , C^2 , C^3 , C^4 , C^5 в отдельности найдено минимальное, максимальное и среднее расстояние между своими (принадлежащими данному классу) векторами – Ряд 1.

Также получены соответствующие значения для пар свой–чужой (элемент, не относящийся к рассматриваемому классу) – Ряд 3.

Для всех пяти классов Ряд 1 немного превышает Ряд 3 по минимальным значениям (оба малы). Ряд 1 меньше, чем Ряд 3 для максимальных значений (как существенно, так и незначительно). Для средних величин имеются оба варианта различия (небольшого).

Ранее для наборов из двух, трех и четырех классов были получены схожие результаты.

Для классов C^1 , C^2 , C^3 , C^4 , C^5 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 – Ряд 2 (Ряд 4).

Для классов C^1 и C^3 Ряд 4 всюду превышает Ряд 2. Аналогичная картина наблюдается для класса C^5 , исключая минимальные значения. При рассмотрении этих трех классов для минимальных и средних значений разница небольшая, а для максимальных – существенная. Для C^2 имеется такое же соотношение по средним значениям и противоположное – по минимальным и максимальным, причем везде Ряд 2 и Ряд 4 отличаются мало. В классе C^4 Ряд 2 немного превышает Ряд 4 для минимальных и средних значений, а для максимальных – ситуация обратная, но различие незначительное.

Для рассмотренных ранее наборов из двух, трех и четырех классов получены аналогичные результаты.

Построено распределение количества своих и чужих элементов на отрезке их нахождения при удалении от центра масс.

В характере распределения числа своих (чужих) элементов имеется структурное сходство между всеми пятью классами. Это же наблюдается в отношении распределений своих–чужих элементов.

Аналогичные результаты были получены ранее для множеств, состоящих из двух, трех, а также четырех классов.

Более того, наблюдается сходство структур соответствующих обучающих множеств классификаторов двух типов: с одной стороны, вычисляющих СО, а с другой стороны – определяющих СЗЧ.

Библиографический список

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. 18 с. 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. 40 с. DOI:10.20948/prepr-2018-255

9. Гавриков М.Б., Локуцкий О.В. Начала численного анализа. — М.: Янус, 1995.

10. Schürmann J. Pattern Classification. — New York: John Wiley&Sons, Inc., 1996.

11. 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

12. Гавриков Б.М., Пестрякова Н.В. Статистический классификатор для диагностики онкологических заболеваний // Информационные технологии и вычислительные системы. 2023. № 1. С. 39-49.