



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

Препринты ИПМ • Препринт № 72 за 2022 г.



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

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

Статистический подход для
диагностики онкологических
заболеваний по параметрам
крови

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



Рекомендуемая форма библиографической ссылки: 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.
<https://doi.org/10.20948/prepr-2022-72>
<https://library.keldysh.ru/preprint.asp?id=2022-72>

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

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

**Статистический подход для диагностики
онкологических заболеваний
по параметрам крови**

Москва — 2022

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

Статистический подход для диагностики онкологических заболеваний по параметрам крови

Предложен математический метод для предварительной диагностики онкологических заболеваний человека различной локализации по параметрам крови. Авторами разработан статистический классификатор, основанный на полиномиально-регрессионном подходе и имеющий вероятностные оценки. Обучение осуществлялось на базах данных анализа периферической крови онкобольных по различным системам организма.

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

Boris Mikhailovich Gavrikov, Mikhail Borisovich Gavrikov, Nadejda Vladimirovna Pestryakova

Statistical approach for diagnosing oncological diseases by blood parameters

A mathematical method is proposed for the preliminary diagnosis of human oncological diseases of various localization by blood parameters. The authors have developed a statistical classifier based on a polynomial-regression approach and having probabilistic estimates. Training was carried out on databases of analysis of peripheral blood of cancer patients in various 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.....	11
Библиографический список.....	11

Введение

В основе математической модели, рассматриваемой в данной работе, 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 [1]. В рамках предлагаемой модели организм представляется как совокупность систем органов.

Авторы ранее показали, что статистический метод классификации может успешно применяться для оценивания СЗЧ по показателям периферической крови [2-8].

СЗЧ включает четыре градации – от практически здорового состояния до максимальной степени поражения организма:

- 1 класс – здоровые;
- 2 класс – начальные отклонения состояния здоровья;
- 3 класс – выраженные отклонения;
- 4 класс – тяжелое заболевание (онкология).

По каждой системе организма (СО) – пищеварения, дыхания и пр. – проводилось самостоятельное исследование СЗЧ при помощи отдельного классификатора, обученного на профильной выборке. Эти базы, включающие восемь параметров периферической крови, были созданы на основе диагнозов, поставленных врачами высокой классификации.

Для мужчин и женщин строили различные классификаторы, поскольку диапазоны вариации показателей крови среди множества людей существенно зависят от пола.

В настоящей работе рассматривается подход, позволяющий определить область локализации онкологии. Для этого обучение статистического классификатора предлагается проводить на наборах параметров периферической крови четвертого класса (онкологических больных), относящихся к нескольким различным СО.

Нет необходимости проводить классификацию по всем СО, вполне допустимо ограничиться только проблемными. Был построен и обучен классификатор для четырех СО. Аналогичная уточняющая процедура проведена также для возможных комбинаций, включающих по три либо по две СО из имеющегося набора.

В данной постановке не возникает вопрос о «чистоте эксперимента», хотя этот анализ практически всегда имеет место при численном моделировании на основе измеренных данных. Дело в том, что наборы параметров крови онкобольных получены одним и тем же способом и имеют достаточный (сопоставимый) объем. Результаты, относящиеся к структуре наборов крови, на основе которых строились обучающие выборки, были подробно описаны [3, 4, 7, 8].

Отсутствуют также вырожденные случаи по отдельным признакам. Вследствие этого проводимое здесь моделирование предварительной

диагностики является даже более обоснованным, по сравнению с оцениванием состояния здоровья [1-3].

Метод классификации

Общепринятые обозначения и размерность используемых восьми показателей крови: 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).

Рассматриваем K определенных перенумерованных СО человека, $1 \leq k \leq K$. 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:

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

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:

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

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 базы 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 ($1 \leq k \leq K$), $\mathbf{y}^{(i)} = (0 \dots 1 \dots 0)$ – 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 рассмотрены четыре СО для мужчин: пищеварительная система (C^1), органы дыхания (C^2), опорно-двигательный аппарат (C^3), урологическая система (C^4).

Нормирование проводим следующим образом. Рассмотрим все четыре обучающие выборки 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 – объем выборки по данной совокупности СО.

Затем выполняем следующее преобразование:

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

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

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

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

В (5) выражения в фигурных скобках соответствуют цепочкам элементов вектора, вычисляемым по всем показателям крови из имеющегося набора. Длина полинома 3003. Имеются мономы первого, второго, третьего, четвертого, пятого и шестого порядка. Перекрестные произведения используются для мономов второго, третьего, четвертого, пятого и шестого порядка.

Исследование проведено на классификаторе из четырех классов. Обучающее множество $C^1UC^2UC^3UC^4$ имеет следующий объем (количество наборов крови), дифференцированный по классам: $|C^1| = |C^3| = 33$, $|C^2| = 21$, $|C^4| = 32$. Классификатор обеспечил точность 93,3% на обучающем множестве из 119 элементов (сохраняется восемь ошибок).

Также рассматривались классификаторы, основанные на объединениях по три класса из четырех имеющихся. На таких комбинациях была получена более

высокая точность уже на полиномах пятого порядка с длиной полинома 1287, без комбинации элементов в последней паре фигурных скобок (5).

Множества из двух классов классифицируются безошибочно, причем в ряде случаев достаточно использовать модификацию вектора $\mathbf{x}(\mathbf{v})$ четвертого порядка с длиной полинома 495.

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

Расстояние между «своими» и «чужими» элементами

Для каждого из четырех рассматриваемых классов C^1, C^2, C^3, C^4 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 четырех классов, $V = \{C^1UC^2UC^3UC^4\}$.

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: $V^{-k} = V \setminus V^k$.

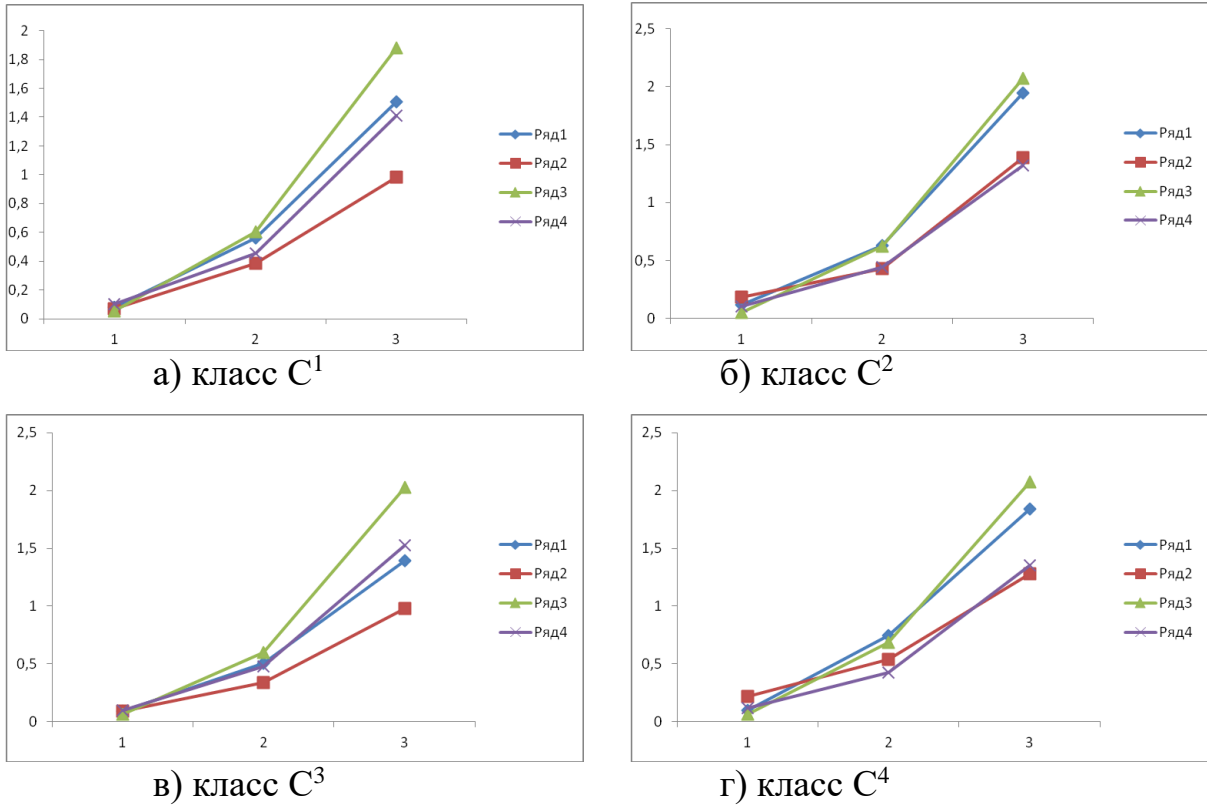


Рис. 1. Минимальное, максимальное и среднее расстояние между парами векторов: свой–свой, свой–чужой, центр масс–свой, центр масс–чужой.

Среднее расстояние:

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

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

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: $\{\mathbf{w}^{-k,j_1}, j_1 = 1, \dots, J_{-k}\} = V^{-k}, V^{-k} = V \setminus V^k$.

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

Для классов C^1, C^2, C^3 и C^4 соответственно представлены (рис.1а, б, в, г) 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 четырех классов (рис. 1) Ряд 1 немного превышает Ряд 3 по минимальным значениям (причем обе малы); кроме того, первая величина меньше второй для максимальных значений. Для средних имеются оба варианта различия, но оно небольшое.

Отклонение от центра масс своих и чужих элементов

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

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

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:

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

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 $C^1UC^2UC^3UC^4$ 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 класса C^1 (рис. 1а) Ряд 4 всюду превышает Ряд 2, причем для минимальных и средних значений разница небольшая, а для максимальных – существенная. Аналогичная картина наблюдается для класса C^3 , исключая практически равные минимальные значения (рис. 1в). Для C^2 соотношения противоположные (при малом различии) для минимальных и максимальных значений, а средние – почти равны (рис. 1б). В классе C^4 (рис. 1г) Ряд 2 немного

превышает Ряд 4 для минимальных и средних значений, а для максимальных – ситуация противоположная при незначительном различии.

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 (13), (14), 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 (16), (17), – на отрезке $[D_{kz_{\min}}, D_{kz_{\max}}]$. Пусть

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

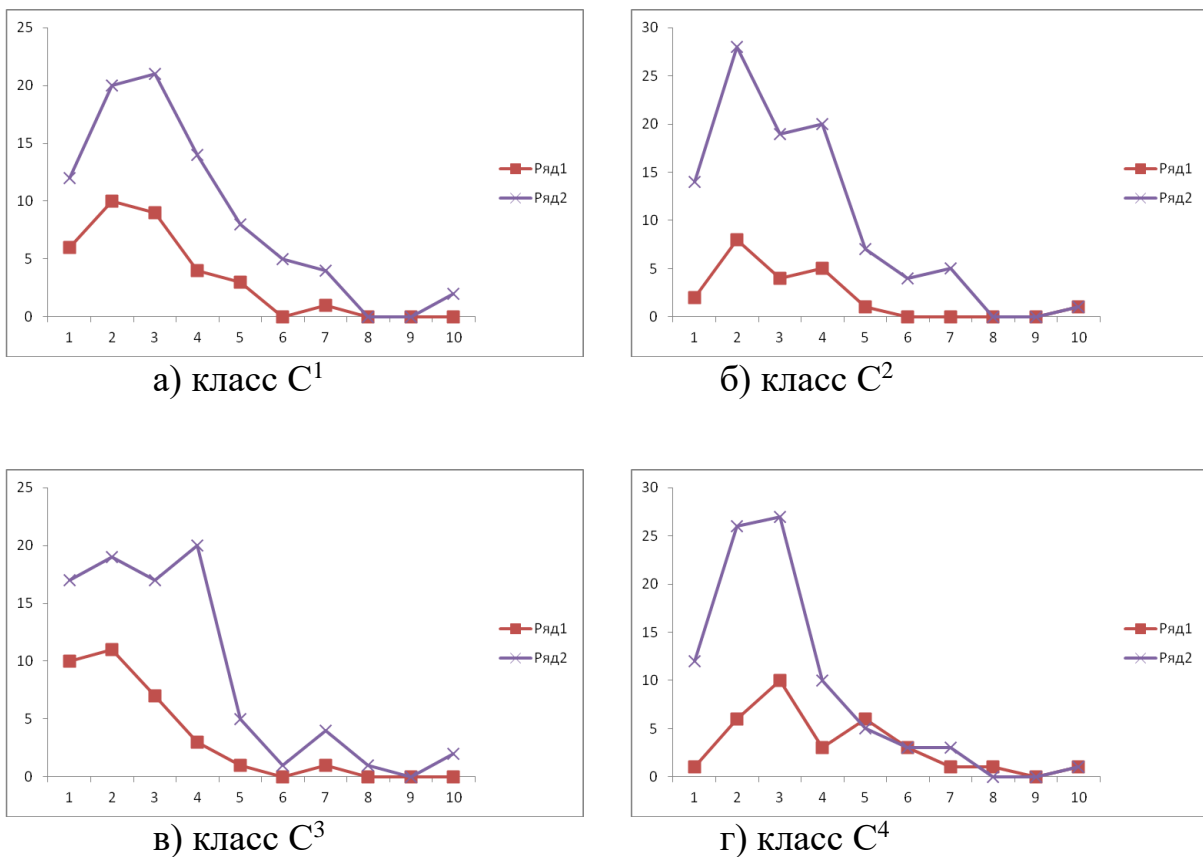


Рис. 2. Распределение числа своих и чужих элементов при удалении от центра масс ($C^1 \cup C^2 \cup C^3 \cup C^4$)

Делим отрезок $[Dk_{\min}, Dk_{\max}]$ (оси абсцисс на рис. 2а, б, в, г) на десять равных по длине частей – один отрезок и девять полуинтервалов: $[Dk_{\min}, Dk_{\min} + d]$, $(Dk_{\min} + d, Dk_{\min} + 2d]$, \dots , $(Dk_{\min} + 9d, Dk_{\min} + 10d]$, 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 $C^1UC^2UC^3UC^4$ 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 четырьмя классами. Эта же аналогия наблюдается в отношении распределений своих–чужих элементов. Отметим, что для классов C^1, C^3 получены результаты (рис. 2а, в), 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 и C^3 (рис. 2а, в).

Заключение

Разработан классификатор, позволяющий уточнить область локализации онкологии по показателям периферической крови с использованием статистического метода распознавания, основанного на полиномиальной регрессии.

Проведено исследование структуры обучающего множества, состоящего из четырех классов (C^1, C^2, C^3, C^4), соответствующих следующим СО для мужчин: пищеварительная система, органы дыхания, опорно-двигательный аппарат, урологическая система. Аналогичная процедура проведена для комбинаций, составленных из трех классов из имеющегося набора, а также из пар элементов.

Для каждого из четырех классов C^1, C^2, C^3, C^4 в отдельности 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 классов C^1, C^2, C^3, C^4 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 на отрезке их нахождения при удалении от центра масс.

Проведенный анализ позволил выявить сходство и различие в структуре множеств C^1, C^2, C^3, C^4 .

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.