

Автоматическое получение «упрощенной лейкоцитарной формулы» методами проточной цитофлуориметрии и кластерного анализа

А. В. Орехов, В. И. Шишкин, Г. В. Кудрявцева, Г. В. Павилайнен, В. В. Шишкин, Н. С. Люджевич

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: Орехов А. В., Шишкин В. И., Кудрявцева Г. В., Павилайнен Г. В., Шишкин В. В., Люджевич Н. С. Автоматическое получение «упрощенной лейкоцитарной формулы» методами проточной цитофлуориметрии и кластерного анализа // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2023. Т. 19. Вып. 4. С. 469–483. <https://doi.org/10.21638/11701/spbu10.2023.404>

Лейкоцитарная формула — это процентное отношение основных групп белых клеток крови (лейкоцитов). По морфологическим признакам среди лейкоцитов можно выделить три субпопуляции: лимфоциты, моноциты и гранулоциты, последние разделяют на нейтрофильные, эозинофильные и базофильные клетки. Автоматическая типологизация белых клеток крови является нерешенной задачей, поэтому при их цитометрическом исследовании подсчет количества клеток в различных субпопуляциях лейкоцитов фактически производится вручную, что, в свою очередь, обуславливает субъективность эксперимента и большие значения погрешностей при вычислениях. Для решения проблемы автоматической типологизации белых клеток крови предпринимались попытки использования методов кластерного анализа. В вычислительных экспериментах было показано, что применение стандартных алгоритмов, таких как метод К-средних, EM-алгоритм, DBSCAN и т. п., не эффективно. За последние годы опубликовано большое количество исследовательских работ, в которых описываются специализированные алгоритмы кластеризации для обнаружения и определения субпопуляций белых клеток крови, некоторые из них нашли практическое применение, но по-прежнему актуальными остаются проблемы, связанные с наличием значительного количества шумов и с различной плотностью распределения данных при кластеризации лейкоцитов методами проточной цитометрии. Рассматривается подход к построению стратегии автоматического выделения лимфоцитов, моноцитов и гранулоцитов при помощи модифицированного агломеративного центроидного метода кластеризации. Обсуждаются результаты вычислительных экспериментов.

Ключевые слова: лейкоцитарная формула, проточная цитометрия, кластерный анализ, марковский момент, метод наименьших квадратов.

1. Основные принципы проточной цитофлуориметрии. Проточный цитофлуориметр — прибор для измерения оптических свойств клеток. Методы проточной цитофлуориметрии позволяют измерять светорассеяние клеток под разными углами и их флуоресценцию в различных диапазонах спектра. Анализ собранных данных дает возможность выделить группы клеток, обладающих теми или иными свойствами, определить их относительное и абсолютное содержание в образце [1, 2].

Работу проточной системы цитофлуориметра можно описать следующим образом. Суспензия клеток из пробирки с образцом забирается специальным приспособ-

лением, клетки узкой струей впрыскиваются в центр ламинарного потока проточной жидкости, который протекает через проточную ячейку по тонкому каналу из оптически прозрачного материала. В одной из точек канала сфокусирован лазерный луч, в момент пересечения клеткой этого участка рассеянный ею свет и флуоресценция регистрируются оптической системой цитофлуориметра. Прибор может быть оснащен несколькими лазерами. Рассеянный клетками свет и флуоресценция с помощью системы линз, светоделителей и фильтров спектрально распределяются по системе детекторов: FSC-детектор, SSC-детектор и набор FL-детекторов [1, 2].

Детектор прямого светорассеяния (FSC-детектор) располагается по ходу лазерного луча и собирает излучение, рассеянное в пределах малых углов ($2\text{--}16^\circ$). При прохождении клетки через лазерный луч часть излучения рассеивается и попадает на периферическую часть детектора. Зарегистрированный сигнал преобразуется в электрический импульс, величина которого пропорциональна количеству попавшего на детектор света. Появление сигнала FSC свидетельствует о прохождении какого-либо объекта через лазерный луч и применяется для подсчета количества клеток. Объекты значительного размера вызывают большие величины малоуглового светорассеяния, как следствие численное значение FSC-сигнала позволяет оценить величину объекта [2].

Детектор бокового светорассеяния (SSC-детектор) располагается под углом 90° относительно направления лазерного луча и собирает излучение, рассеянное в пределах больших углов ($75\text{--}105^\circ$). Рассеяние света под такими углами является следствием многократного преломления луча лазера при прохождении через клетку. Регистрация этого излучения позволяет судить о сложности внутреннего строения клетки (наличие гранул, везикул и т. п.). Объекты с более сложной внутренней структурой вызывают большее светорассеяние, поэтому величина SSC-сигнала дает возможность оценить степень гранулярности клетки [2].

Комбинация FSC- и SSC-параметров позволяет судить о морфологии клетки в целом и даже без анализа флуоресценции достаточно информативна при изучении популяций клеток в образце.

Система для регистрации флуоресценции состоит из набора зеркал, светофильтров и FL-детекторов флуоресценции, каждый из которых регистрирует излучение в строго определенном диапазоне длин волн. Анализ флуоресценции клеток дает возможность получить наибольшее количество информации об изучаемом образце, поскольку даже клетки с идентичной морфологией могут отличаться по составу белков и других макромолекул. Одновременное использование нескольких красителей способствует выделению популяции клеток с различным сочетанием исследуемых признаков. Такой подход очень важен при анализе содержания в крови различных субпопуляций лейкоцитов, каждая из которых отличается уникальным сочетанием поверхностных белков — кластеров дифференциации (сокращенно CD).

Результатом измерения является массив данных, где в виде таблицы записаны числовые значения параметров клеток со всех рабочих детекторов для каждого из зарегистрированных событий (в терминологии проточной цитофлуориметрии единственный сигнал от клетки принято называть событием). Для облегчения обработки полученных данных они могут быть представлены в графическом виде на экране консоли, подключенной к проточному цитофлуориметру.

Наиболее распространенная форма графического представления данных в проточной цитометрии — точечная диаграмма, элементы которой — точки, соответствующие упорядоченным парам параметров некоторого события [2].

2. Проблема автоматической типологизации лейкоцитов. Лейкоцитарной формулой называется процентное соотношение основных субпопуляций лейкоцитов: лимфоцитов, моноцитов, нейтрофилов, эозинофилов и базофилов [3].

При анализе крови методами проточной цитометрии используется «упрощенная лейкоцитарная формула», в которой нейтрофилы, эозинофилы и базофилы объединяются в одну общую субпопуляцию — гранулоцитов. Разделение лейкоцитов на три группы при цитометрическом исследовании обусловлено технической спецификой проточных цитофлуориметров и морфологическими особенностями белых клеток крови. Лимфоциты и моноциты имеют простое несегментированное ядро и небольшую зернистость цитоплазмы, а гранулоциты, напротив, клетки с более сложным строением и специфической зернистостью цитоплазмы [4].

Для визуализации «упрощенной лейкоцитарной формулы» применяются точечные диаграммы (рис. 1) в осях FSC и SSC [5]. Лимфоциты — это самые маленькие белые клетки крови с простым внутренним строением, моноциты крупнее и сложнее по внутреннему строению, а гранулоциты наиболее сложные и самые большие лейкоциты. Важно отметить еще одну группу точек, расположенную на рис. 1 внизу, левее всех остальных клеток, это так называемый дребрис (осколки клеток).

Нерешенной проблемой остается задача автоматической типологизации субпопуляций лейкоцитов, поэтому их «гейтирование» при цитометрическом исследовании фактически производится вручную [1, 2]. Согласно общепринятому соглашению, «гейт (от англ. “gate” — ворота) — это инструмент для выделения из всего массива полученных данных отдельных популяций клеток, удовлетворяющих определенным условиям» [2, с. 31]. Для создания гейтов используются графические инструменты, позволяющие оператору в интерактивном режиме выделять группы клеток на экране консоли в виде прямоугольного или полигонального гейта (рис. 2). Затем программными средствами определяется количество объектов в выделенной оператором ограниченной области [1, 2], что, в свою очередь, обуславливает субъективность эксперимента и большие значения погрешностей при вычислении количества клеток в различных субпопуляциях лейкоцитов.

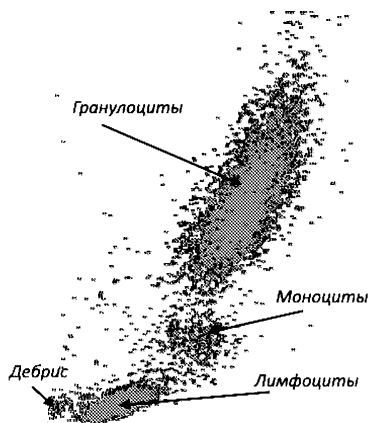


Рис. 1. Стандартное распределение лимфоцитов, моноцитов, гранулоцитов и дребриса в осях FSC и SSC

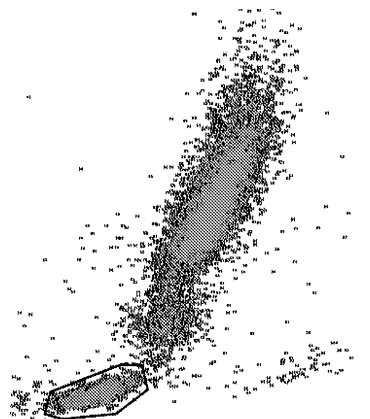


Рис. 2. Выделение кластера лимфоцитов при помощи полигонального гейта (замкнутая черная ломаная линия)

В связи с этим возникает следующая проблема.

Доказательная медицина — это процесс систематического пересмотра, оценки и использования результатов клинических исследований с целью оказания оптимальной медицинской помощи пациентам, который сочетает в себе определенные принципы и методы. Благодаря их действию инструкции и стратегии в медицине основываются на текущих подтверждающих данных об эффективности разных форм лечения и медицинских услуг в целом. В том числе эти принципы предъявляют повышенные требования к качеству лабораторных исследований [6].

В различных медицинских учреждениях и у нас в стране, и за рубежом проводился один и тот же эксперимент: нескольким исследователям предлагалось произвести гейтирование клеток из одной и той же пробы, и всегда получались неодинаковые результаты, окончательная максимальная погрешность достигала 30 % [7–10].

Свой вклад в погрешность результатов обработки анализов крови вносят два вида ошибок: случайные и системные.

Случайные ошибки сопутствуют любому измерению, как бы тщательно оно ни проводилось, и проявляются в некотором различии результатов измерения одного и того же элемента, выполненного данным методом. Они обусловлены в том числе точностью работы персонала лаборатории (неточное считывание результатов, ошибка утомления, неверный подбор класса точности инструментов и т. д.).

Системные ошибки зависят от применяемых приборов и реактивов, определяют точность приборов, происходят от неправильного или неточного выполнения операции, зависят от личных способностей оператора, его органов чувств, привычек и т. д.

Субъективный фактор имеет большое значение и при случайных, и при системных ошибках, поэтому для улучшения качества исследований средствами проточной цитометрии следует разработать машинные (формальные) методы гейтирования белых клеток крови. Возможным подходом к решению этой задачи может быть кластерный анализ результатов цитометрического исследования. Необходимо сразу отметить: в вычислительных экспериментах было показано, что применение стандартных алгоритмов кластеризации не позволяет получить желаемые результаты [11].

При численном моделировании процесса кластеризации лейкоцитов по FSC- и SSC-параметрам хуже всего проявил себя метод К-средних [12, 13]. Разделение гранулоцитов на два кластера является ошибочным в принципе. Чуть лучше результаты кластерного анализа лейкоцитов при помощи EM-алгоритма [14].

В целом иерархические агломеративные алгоритмы [15, 16] дают возможность получить удовлетворительные результаты, но не более того. Остаются открытыми проблемы завершения процесса кластеризации и определения предпочтительного числа кластеров. Не удастся достоверно отделять лимфоциты от мелкого дебриса [11].

Лучшие результаты были получены при использовании метода DBSCAN [17]. Однако при наличии значительного объема шумов или существенной разницы в плотности распределения лейкоцитов подбор параметров, необходимых для выполнения этого алгоритма, вызывает серьезные затруднения [11].

За последние годы опубликовано много исследовательских работ, в которых описываются специализированные алгоритмы кластеризации для обнаружения и определения популяций белых клеток крови, некоторые из них нашли практическое применение [18, 19].

Но по-прежнему актуальными остаются проблемы, связанные с наличием большого количества шумов и различной плотностью распределения данных при класте-

ризации лейкоцитов методами проточной цитометрии. Также не решена внутренняя задача кластерного анализа, связанная с определением предпочтительного числа кластеров и моментом остановки самого процесса.

На рис. 3 и 4 изображены аномальные распределения лейкоцитов в осях FSC и SSC: на рис. 3 хорошо видно большое количество шумов распределенных по периферии стандартных кластеров: дегрессива, лимфоцитов, моноцитов и гранулоцитов, а на рис. 4 присутствует еще один, так называемый «пятый кластер».

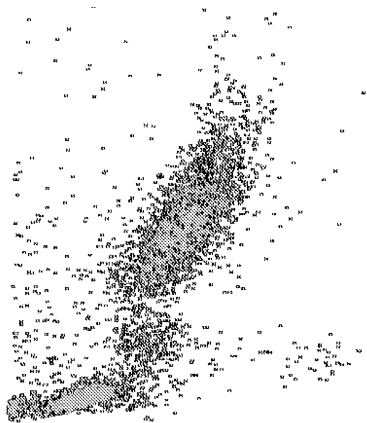


Рис. 3. Распределение лейкоцитов с большим количеством шумов

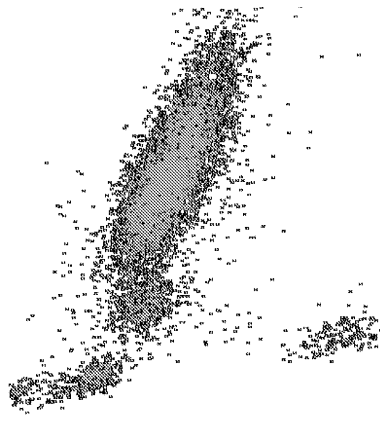


Рис. 4. Аномальное распределение лейкоцитов с «пятым кластером»

Замечание. В 2018 г. при исследовании методов автоматической типологизации белых клеток крови в осях FSC и SSC было обнаружено появление еще одной группы лейкоцитов, которая получила название «пятого кластера» [20].

3. Модифицированный центроидный метод кластеризации. Для устранения шумов и преодоления проблем, связанных с различной плотностью распределения лейкоцитов в шкалах FSC и SSC, рассмотрим модифицированный агломеративный центроидный алгоритм кластеризации с марковским моментом остановки.

Формально этот метод можно описать следующим образом. Произвольный алгоритм кластеризации является отображением

$$A: \begin{cases} X \longrightarrow \mathbb{N}, \\ \bar{x}_i \longmapsto k, \end{cases}$$

которое ставит в соответствие любому элементу \bar{x}_i из выборки X единственное натуральное число k , являющееся номером кластера, которому принадлежит \bar{x}_i . Процесс кластеризации разбивает выборку X на попарно дизъюнктные подмножества X_h , называемые кластерами. Следовательно, отображение A задает на X отношение эквивалентности; в качестве независимых представителей классов эквивалентности выбирают элементы, называемые центроидами. В n -мерном евклидовом пространстве \mathbb{E}^n координаты центроидов равны среднему арифметическому соответствующих координат всех элементов (векторов), входящих в кластер (класс эквивалентности). Если отождествить каждый вектор из \mathbb{E}^n с материальной точкой единичной массы, то центроиды можно рассматривать как центры масс [21].

«Расстояние» между центроидами в предлагаемой модификации центроидного метода вычисляется по формуле

$$\widehat{\rho}(\widehat{X}_k, \widehat{X}_l) = \left| \rho(\widehat{X}_k, \widehat{X}_l) - w \cdot n_k \cdot n_l \right|, \quad (1)$$

где \widehat{X}_k и \widehat{X}_l — центроиды кластеров k и l ; $\rho(\widehat{X}_k, \widehat{X}_l)$ — евклидово расстояние между центроидами этих кластеров; n_k и n_l — количество элементов в соответствующих кластерах; w — параметр, называемый «коэффициентом притяжения» и используемый для тонкой настройки модифицированного алгоритма в интерактивном режиме [5].

Очевидно, что «расстояние», задаваемое (1), не является метрикой, так как оно не рефлексивно.

Если выборочная совокупность X содержит m элементов (векторов), то в начале любого агломеративного метода кластеризации полагают, что X разбита на m классов эквивалентности (кластеров), содержащих по одному элементу; при этом понятно, что кластеры, состоящие из единственного элемента, и их центроиды совпадают: $X_h = \widehat{X}_h$ для $\forall h \mid 1 \leq h \leq m$.

Итерации алгоритма \mathcal{A} , реализующего модифицированный центроидный метод, можно описать следующим образом. Первым шагом 1-й итерации \mathcal{A}_1 алгоритма \mathcal{A} , реализующего модифицированный центроидный метод, является построение диагональной матрицы расстояний между \widehat{X}_h , затем определяется ее минимальный элемент: $F_1 = \min(\widehat{\rho}(\widehat{X}_h, \widehat{X}_l))$, где $1 \leq h, l \leq m$; F_1 — минимальное расстояние при \mathcal{A}_1 .

После чего X_h и X_l , для которых $\widehat{\rho}$ минимально, объединяются в один класс эквивалентности, который обозначим как X_1 , а его центроид — как \widehat{X}_1 . Кластеры X_h и X_l (при \mathcal{A}_1 элементы \bar{x}_h и \bar{x}_l) заменяются на центроид \widehat{X}_1 . Таким образом, после \mathcal{A}_1 выборочная совокупность X оказывается разбитой на $m - 1$ элемент.

Не умаляя общности, будем считать, что в начале g -й итерации \mathcal{A}_g модифицированного алгоритма \mathcal{A} выборочная совокупность X разбита на p кластеров. Первым шагом \mathcal{A}_g является построение диагональной матрицы расстояний, затем так же, как и при \mathcal{A}_1 , находится минимальный элемент этой матрицы: $F_g = \min(\widehat{\rho}(\widehat{X}_h, \widehat{X}_l))$, где $1 \leq h, l \leq p$; F_g — минимальное расстояние при \mathcal{A}_g . Элементы X_h и X_l , для которых расстояние $\widehat{\rho}$ минимально, объединяются в кластер. Обозначим его как X_g . Его центроид \widehat{X}_g имеет координаты, равные среднему арифметическому соответствующих координат всех векторов из X_h или X_l , объединенных в X_g . В конце итерации \mathcal{A}_g элементы X_h и X_l заменяются на \widehat{X}_g . Таким образом, после завершения \mathcal{A}_g выборочная совокупность X оказывается разбитой на $p - 1$ элемент [21].

4. Параболический аппроксимационно-оценочный критерий. Если нет правила завершения процесса кластеризации, то после $m - 1$ итерации модифицированного центроидного алгоритма выборочная совокупность X будет объединена в один кластер, что является абсурдным результатом. Для определения формального признака завершения процесса кластеризации, реализующего модифицированный центроидный метод в \mathbb{E}^n , будем использовать «кортеж минимальных расстояний» F_1, F_2, \dots, F_{m-1} .

В начале алгоритма при формировании кластеров объединяются близкие точки и числовые значения F_j растут очень медленно. Допуская вольность речи, можно сказать, что это возрастание является «почти линейным», но при слиянии сформировавшихся кластеров должен произойти резкий скачок числового значения F_j , который, по здравому смыслу, совпадает с моментом завершения процесса кластеризации [21]. Такой скачок лучше аппроксимировать не прямой, а параболой.

Аналитически момент, когда монотонный рост произвольной числовой последовательности y_t изменяется от линейного вида к нелинейному, можно определить при помощи квадратичной формы $\delta^2(4_0)$, параболического аппроксимационно-оценочного критерия, построенного по четырем узлам аппроксимации [21, 22]:

$$\delta^2 = \delta^2(4_0) = \frac{1}{245}(19y_1^2 - 11y_2^2 + 41y_3^2 + 12y_1y_2 - 64y_1y_3 - 46y_2y_3). \quad (2)$$

Параболический аппроксимационно-оценочный критерий строится в виде разности квадратичных погрешностей: линейной аппроксимации и неполной параболической аппроксимации одной и той же последовательности y_t . Линейная аппроксимация производится по методу наименьших квадратов в классе функций $ax + b$, неполная параболическая аппроксимация — в классе функций $cx^2 + d$. Квадратичная форма параболического аппроксимационно-оценочного критерия $\delta^2(4_0)$ строится по четырем узлам аппроксимации: $(0, y_0); (1, y_1); (2, y_2); (3, y_3)$. Так как подстрочный индекс однозначно определяет натуральный аргумент, для обозначения узла аппроксимации вместо упорядоченной пары (i, y_i) можно использовать соответствующий элемент последовательности y_i и называть его натуральным узлом аппроксимации.

При построении квадратичных форм аппроксимационно-оценочных критериев значения последовательности y_t рассматриваются в узлах y_0, y_1, \dots, y_{k-1} , при этом всегда $y_0 = 0$, что значительно облегчает вычисления. Для выполнения этого условия на любом шаге аппроксимации применяется преобразование

$$y_0 = y_j - y_j, \quad y_1 = y_{j+1} - y_j, \dots, y_{k-1} = y_{j+k-1} - y_j. \quad (3)$$

Можно сказать, что вблизи элемента y_k характер возрастания числовой последовательности y_t изменился с линейного на параболический, если для узлов y_0, y_1, \dots, y_{k-1} линейная аппроксимация не хуже неполной параболической, т. е. справедливо неравенство $\delta^2 \leq 0$, а для набора точек, y_1, y_2, \dots, y_k , сдвинутых на один шаг дискретности, неполная параболическая аппроксимация стала точнее линейной, т. е. выполняется неравенство $\delta^2 > 0$.

Рассмотрим пример применения параболического аппроксимационно-оценочного критерия $\delta^2(4_0)$.

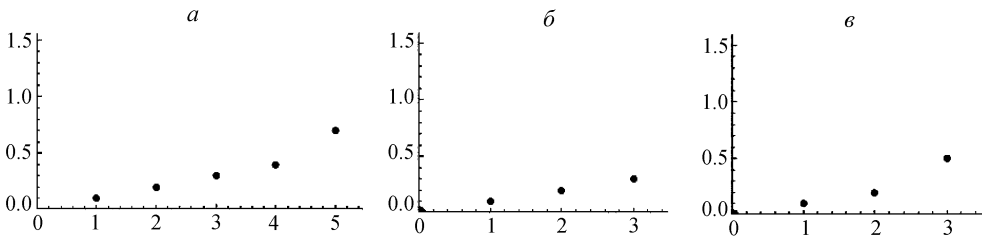


Рис. 5. Пример применения параболического аппроксимационно-оценочного критерия $\delta^2(4_0)$

На оси абсцисс отложены номера компонент кортежа, на оси ординат — их числовые значения. a — общий вид кортежа y_1, y_2, y_3, y_4, y_5 ; $б$ — вид кортежа y_1, y_2, y_3, y_4 после преобразования (3); $в$ — вид кортежа y_2, y_3, y_4, y_5 после преобразования (3).

Пусть $y_1 = 0.1, y_2 = 0.2, y_3 = 0.3, y_4 = 0.4, y_5 = 0.7$ (рис. 5, a). Используя преобразование (3), вычислим $\delta^2(4_0)$ для y_1, y_2, y_3, y_4 (рис. 5, $б$) и получим, что

$\delta^2(4_0) = -0.00408163$, затем при помощи (3) преобразуем узлы y_2, y_3, y_4, y_5 (рис. 5, 6) и имеем $\delta^2(4_0) = 0.00995918$.

Таким образом, для y_1, y_2, y_3, y_4 возрастание числовых значений y_t является линейным, а для узлов y_2, y_3, y_4, y_5 оно стало параболическим.

Кроме «кортежа минимальных расстояний» $\{F_1, F_2, \dots, F_{m-1}\}$ будем рассматривать множество, которое назовем «кортежем примитивных минимальных расстояний»: $\{D_1, D_2, \dots, D_{m/2}\}$, его элементами являются числовые значения, которые вычисляются следующим образом: строится матрица евклидовых расстояний между элементами выборки X , определяется минимальный элемент этой матрицы, его значение присваивается элементу D_1 , и элементы, соответствующие минимальному расстоянию, удаляются из X . Процедура повторяется, новое минимальное значение матрицы расстояний присваивается D_2 , и элементы, соответствующие новому минимальному расстоянию, удаляются из X , и так происходит до тех пор, пока не будут удалены все элементы из X . Если X содержит нечетное количество элементов, то $D_{m/2} = D_{m/2-1}$.

Для определения момента остановки агломеративного алгоритма кластеризации и предпочтительного числа кластеров опишем кластеризацию как квазидетерминированный дискретный случайный процесс $\xi = \{\xi_t, t \in T\}$ [22], который можно рассматривать как функцию двух переменных $\xi = \xi(t, \omega)$, где t — натуральный аргумент, ω — случайное событие. Если зафиксировать t , то получим случайную величину ξ_t ; если же зафиксировать случайное событие ω_0 , то функцию от натурального аргумента t , которая называется «траекторией» случайного процесса $\xi = \{\xi_t, t \in T\}$ и является случайной последовательностью $\xi_t(\omega_0)$ [23]. В этом случае случайным событием $\omega \in \Omega$ будет извлечение выборки X из \mathbb{E}^n .

Пусть τ — момент наступления некоторого события в случайном процессе ξ . Если для $\forall t_0 \in T$ можно однозначно сказать, наступило событие τ или нет при условии, что известны значения ξ_t только в прошлом (слева от t_0), то тогда τ — марковский момент относительно случайного процесса $\xi = \{\xi_t, t \in T\}$. А если наступление τ в конечный момент времени — достоверное событие, то τ — марковский момент остановки случайного процесса ξ [23–25].

В евклидовом пространстве \mathbb{E}^n при кластерном анализе выборочных данных модифицированным центроидным алгоритмом одной из основных характеристик процесса будет множество минимальных расстояний. Естественно считать его значения случайной величиной $\xi_t: \Omega \rightarrow \mathbb{R}$, полагая, что t — номер итерации модифицированного алгоритма кластеризации \mathcal{A} . Для любого фиксированного случайного события $\omega_0 \in \Omega$ соответствующая траектория $\xi_t(\omega_0) = F_t$ — монотонно возрастающая случайная последовательность (если $F_i < F_{i-1}$, по теореме Миллигана [26] для центроидного метода такое возможно, F_i присваивается значение F_{i-1}).

Построим статистический критерий завершения процесса кластеризации как марковский момент остановки τ , для этого рассмотрим бинарную задачу проверки статистических гипотез H_0 и H_1 , где нулевая гипотеза H_0 — случайная последовательность $\xi_t(\omega_0)$ возрастает линейно, а альтернативная гипотеза H_1 — случайная последовательность $\xi_t(\omega_0)$ увеличивается нелинейно (параболически). Для проверки статистической гипотезы необходимо построить критерий как строгое математическое правило, позволяющее ее принять или отвергнуть, в качестве такого правила будем использовать параболический аппроксимационно-оценочный критерий $\delta^2(4_0)$.

Тогда, по определению, марковским моментом остановки процесса кластеризации будет статистика

$$\tau = \min\{t \in T \mid \delta_t^2 > 0\},$$

где δ_t^2 задается формулой (2). То есть марковским моментом остановки процесса кластеризации является минимальное значение t , при котором отвергается нулевая гипотеза — H_0 (последовательность минимальных расстояний возрастает линейно) и принимается альтернативная гипотеза — H_1 (последовательность минимальных расстояний возрастает параболически) [21, 22].

Для того чтобы окончательно сформулировать условие завершения описанного выше агломеративного процесса кластеризации, осталось рассмотреть «проблему чувствительности» аппроксимационно-оценочного критерия δ^2 , которую можно связать с понятием «устойчивой кластеризации».

Решим «обратную задачу», пусть известны значения последовательности y_n в узлах y_0, y_1, y_2 и требуется определить, при каком y_3 характер увеличения последовательности y_n изменился с линейного на параболический. Приравняем к нулю квадратичную форму (2) и, заменив y_3 на x , решим квадратное уравнение

$$41x^2 - (64y_1 + 46y_2)x + (19y_1^2 + 12y_2y_1 - 11y_2^2) = 0,$$

для которого

$$x_{1,2} = \frac{32y_1 + 23y_2 \pm 7\sqrt{5}(y_1 + 2y_2)}{41}.$$

Учитывая, что $0 \leq y_1 \leq y_2 \leq y_3$, окончательно находим, что

$$y_3 = \frac{32y_1 + 23y_2 + 7\sqrt{5}(y_1 + 2y_2)}{41}. \quad (4)$$

Заметим, что если $y_j = y_{j+1} = y_{j+2}$, то тогда не только $y_0 = 0$, но и $y_1 = y_2 = 0$. Согласно (2), для любого $y_{j+3} > y_{j+2}$, даже если $y_3 = y_{j+3} - y_j > 0$ сколь угодно мало, квадратичная форма $\delta^2 > 0$.

Если ввести преобразование $y_i = F_i + q \cdot i$, то получим множество $\{y_1, y_2, \dots, y_k\}$, которое назовем «множеством тренда», а q — «коэффициентом тренда». При применении критерия δ^2 не к набору $\{F_1, F_2, \dots, F_n, \dots\}$, а к множеству $\{y_1, y_2, \dots, y_n, \dots\}$ результат кластеризации качественно меняется.

Выполнение процесса кластеризации завершается при помощи параболического аппроксимационно-оценочного критерия, который оценивает скачки монотонно возрастающей последовательности минимальных расстояний. Величина значимого скачка, достаточного для остановки процесса кластеризации, зависит от чувствительности критерия остановки, которая задается неотрицательным коэффициентом тренда q . Чем больше значение q , тем меньше чувствительность критерия остановки процесса кластеризации. Максимальной чувствительностью критерий остановки обладает при $q = 0$, в этом случае при кластеризации получится наибольшее число кластеров. Увеличивая q , можно уменьшить чувствительность критерия остановки так, что процесс будет продолжаться до тех пор, пока все m векторов не объединятся в один кластер. Действительно, если узлы аппроксимации y_0, y_1, y_2 изменяются как арифметическая прогрессия с разностью q , то формула (4) принимает вид

$$y_3 = \frac{1}{41} (78 + 35\sqrt{5}) q \simeq 3.811q,$$

и значения узлов аппроксимации в этом случае равны $0, q, 2q, 3,811q$. А это означает, что при увеличении коэффициента тренда q чувствительность критерия уменьшается, и для достижения критического значения необходима большая величина изменения минимального расстояния.

5. Результаты вычислительных экспериментов. Для численных экспериментов с модифицированным алгоритмом использовался программный код, написанный на языке программирования Python 3.7, с подключением библиотек NumPy, SciPy и с использованием оболочки PyCharm, разработанной компанией JetBrains на основе IntelliJ IDEA, и среды свободного доступа Colaboratory для Jupiter Notebook от компании Google.

При численных экспериментах применялась многоступенчатая кластеризация: сначала удалялись шумы, для этого вычислялось среднеквадратичное отклонение d случайной величины D_j и определялись изолированные точки выборки X как точки, окрестности которых радиуса $r = d/n$ содержат не более чем l клеток из X . После этого все изолированные точки удалялись из множества X . Было установлено, что $n = 0.5, l = 20$ являются оптимальными параметрами для устранения шума в большинстве наборов экспериментальных данных.

На рис. 6, *а* изображена точечная диаграмма до очистки от шумов. После очистки данных от шума (рис. 6, *б*) лейкоциты группируются в три этапа. При этом в качестве управляющих параметров используются q — коэффициент тренда и w — коэффициент притяжения.

На первом этапе выделялась группа лимфоцитов (как кластер, координаты центроида которого находятся ближе всего к статистическим оценкам соответствующего математического ожидания) и эта группа клеток удалялась из выборки. На этом этапе кластеризации $w = 0.0000001, q = 2$.

На втором этапе производилась кластеризация оставшихся данных, после чего лимфоциты возвращались в выборку X , при этом $w = 0.0000005, q = 18$.

На последнем, третьем, этапе автоматической типологизации лейкоцитов модифицированным методом выделялись крупные кластеры, координаты центроидов которых ближе всего к статистическим оценкам координат скоплений моноцитов и гранулоцитов, после чего оставшиеся мелкие кластеры удалялись из выборки так же, как и шумы (рис. 6, *в*).



Рис. 6. Точечные диаграммы

а — распределение лейкоцитов до очистки от шумов; *б* — распределение лейкоцитов после очистки от шумов; *в* — окончательный результат кластеризации лейкоцитов.

На рис. 7 представлены результаты применения модифицированного алгоритма к девяти различным наборам данных [5]. Можно видеть, что в этих наборах предложенный метод правдоподобно выделяет основные группы белых клеток крови.

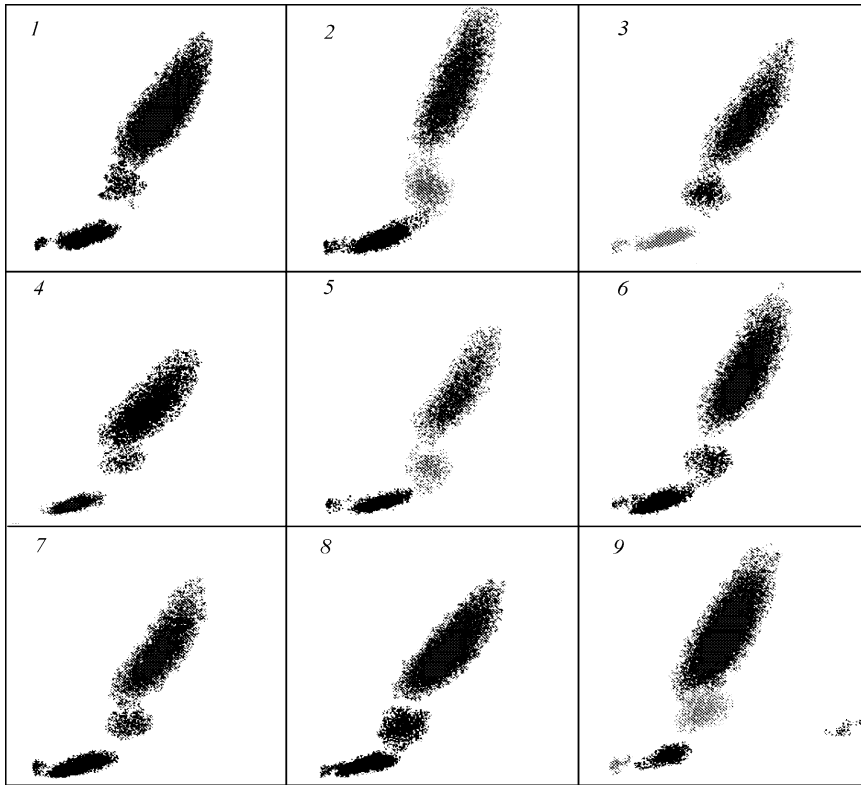


Рис. 7. Результаты применения модифицированного центроидного алгоритма кластеризации лейкоцитов к девяти различным наборам данных

В таблице сравниваются результаты по вычислению доли лимфоцитов «вручную» и автоматически при помощи центроидного алгоритма кластеризации [5].

Таблица. Сравнение результатов по вычислению доли лимфоцитов в интерактивном режиме гейтирования лаборантом и автоматически при помощи модифицированного центроидного алгоритма кластеризации

| Номер набора | Результаты гейтирования, % | Результаты кластеризации, % |
|--------------|----------------------------|-----------------------------|
| 1 | 16.8 | 18.6 |
| 2 | 23.4 | 27.2 |
| 3 | 19.0 | 18.6 |
| 4 | 22.2 | 24.3 |
| 5 | 27.4 | 26.2 |
| 6 | 21.9 | 24.0 |
| 7 | 29.1 | 27.9 |
| 8 | 17.8 | 18.9 |
| 9 | 3.6 | 4.7 |

6. Заключение. Лейкоцитарная формула — важный гематологический показатель, имеющий клиническое и диагностическое значение. Методы получения этой формулы при помощи «ручного выделения» субпопуляций лейкоцитов, моноцитов и гранулоцитов страдают определенной долей субъективности. Автоматическая типологизация белых клеток крови является сложной задачей, но предполагает получение объективных экспериментальных данных. Знание об ожидаемой структуре данных дает возможность использовать дополнительные инструменты для увеличения эффективности кластеризации. Предложенный модифицированный алгоритм агломеративной кластеризации со взвешенной функцией расстояния демонстрирует перспективные результаты. Основной проблемой описанного метода остается большое количество параметров, значения которых выбираются эвристически: радиус окрестностей и предельное число точек, используемых для очистки данных от шумов, коэффициент притяжения в применяемой метрике, а также коэффициент тренда, влияющий на чувствительность аппроксимационно-оценочного критерия.

Дальнейшие исследования предполагают сведение к минимуму указанных недостатков и совершенствование программного кода так, чтобы стало возможным автоматическое гейтирование лейкоцитов не только в осях FSC и SSC, но и по наборам поверхностных белков — кластеров дифференциации.

Литература

1. Зурочка А. В., Хайдуков С. В., Кудрявцев И. В., Черешнев В. А. Проточная цитометрия в медицине и биологии. 2-е изд. Екатеринбург: Урал. отд. РАН, 2014. 574 с.
2. Балалаева И. В. Проточная цитофлуориметрия: учеб.-метод. пособие. Нижний Новгород: Нижегородский государственный университет, 2014. 75 с.
3. Агаджанян Н. А., Смирнов В. М. Нормальная физиология: учебник для студентов медицинских вузов. М.: Медицинское информационное агентство, 2009. 520 с.
4. Хайтов Р. М., Игнатъева Г. А., Сидорович И. Г. Иммунология: учебник. М.: Медицина, 2000. 432 с.
5. Orekhov A. V., Shishkin V. I., Lyudkevich N. S. Clusterization of white blood cells on the modified upgmc method // Stability and Control Processes. Proceedings of the 4th International Conference dedicated to the memory of professor Vladimir Zubov. Cham: Springer, 2022. P. 559–566.
6. Основы доказательной медицины: учеб. пособие для системы послевузовского и дополнительного профессионального образования врачей / под общ. ред. Р. Г. Оганова. М.: Силиция-Полиграф, 2010. 136 с.
7. Pedersen N. W., Chandran P. A., Qian Y., Rebhahn J., Petersen N. V., Hoff M. D., White S., Lee A. J., Stanton R. H. Ch., Jakobsen K., Mosmann T., Gouttefangeas C., Chan C., Scheuermann R. H., Hadrup S. R. Automated analysis of flow cytometry data to reduce inter-lab variation in the detection of major histocompatibility complex multimer-binding T cells // Front Immunol. 2017. Vol. 8. P. 858.
8. Daneau G., Buyze J., Wade D., Diaw P. A., Dieye T. N., Sopheak T., Florence E., Lynen L., Kestens L. CD4 results with a bias larger than hundred cells per microliter can have a significant impact on the clinical decision during treatment initiation of HIV patients // Cytometry B Clin Cytom. 2017. Vol. 92. N 6. P. 476–484. <https://doi.org/10.1002/cyto.b.21366>
9. Qian Y., Kim H., Purawat Sh., Wang J., Stanton R., Lee A., Xu W., Altintas I., Sinkovits R., Scheuermann R. H. FlowGate: towards extensible and scalable web-based flow cytometry data analysis // XSEDE '15: Proceedings of the 2015 XSEDE Conference. Scientific advancements enabled by enhanced cyberinfrastructure. July 2015. Art. N 5. P. 1–8. <https://doi.org/10.1145/2792745.2792750>
10. Omana-Zapata I., Mutschmann C., Schmitz J., Gibson S., Judge K., Indig M. A., Lu B., Tauffman D., Sanfilippo A. M., Shallenberger W., Graminske Sh., McLean R., Hsen R. I., d'Empaire N., Dean K., O'Gorman M. Accurate and reproducible enumeration of T-, B-, and NK lymphocytes using the BD FACSLyric 10-color system: A multisite clinical evaluation // PLoS One. 2019. Vol. 14. N 1. Art. N e0211207. <https://doi.org/10.1371/journal.pone.0211207>
11. Лепский А. И. Сравнительный анализ алгоритмов кластеризации лейкоцитов по FS и SS параметрам при цитофлуориметрическом исследовании крови // Информационные технологии. 2020. Т. 26. № 1. С. 56–61.

12. *Steinhaus H.* Sur la division des corps materiels en parties // *Bull. Acad. Polon. Sci. C1. III.* 1956. Vol. IV. P. 801–804.
13. *Lloyd S.* Least squares quantization in PCM // *IEEE Transactions on Information Theory.* 1982. Vol. 28. Iss. 2. P. 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
14. *Dempster A. P., Laird N. M., Rubin D. B.* Maximum Likelihood from incomplete data via the EM algorithm // *Journal of the Royal Statistical Society. Series B.* 1977. Vol. 39. Iss. 1. P. 1–38.
15. *Everitt B. S.* Cluster analysis. Chichester: John Wiley & Sons Ltd, 2011. 330 p.
16. *Hartigan J. A.* Clustering algorithms. New York; London; Sydney; Toronto: John Wiley & Sons Inc. Press, 1975. 351 p.
17. *Ester M., Kriegel H.-P., Sander J., Xu X.* A density-based algorithm for discovering clusters in large spatial databases with noise // *Proceedings of the 2nd International Conference on knowledge discovery and data mining (KDD-96)* / eds E. Simoudis, J. Han, U. M. Fayyad. Portland: AAAI Press, 1996. P. 226–231.
18. *Weber L. M., Robinson M. D.* Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data // *Cytometry. Pt A.* 2016. Vol. 89A. Iss. 12. P. 1084–1096.
19. *Zhang C., Xiao X., Li X., Chen Y.-J., Zhen W., Chang J., Zheng Ch., Liu Z.* White blood cell segmentation by color-space-based K-means clustering // *Sensors.* 2014. Vol. 14. Iss. 9. P. 16128–16147.
20. *Виль М. Ю.* Анализ статистической связи между клиническими факторами и появлением аномальной субпопуляции лейкоцитов // *Процессы управления и устойчивость.* 2020. Т. 7. № 1. С. 143–147.
21. *Орехов А. В.* Марковский момент остановки агломеративного процесса кластеризации в евклидовом пространстве // *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления.* 2019. Т. 15. Вып. 1. С. 76–92. <https://doi.org/10.21638/11702/spbu.2019.106>
22. *Orekhov A. V.* Quasi-deterministic processes with monotonic trajectories and unsupervised machine learning // *Mathematics.* 2021. Vol. 9. Art. N 2301. <https://doi.org/10.3390/math9182301>
23. *Буллинский А. В., Ширяев А. Н.* Теория случайных процессов. М.: Физматлит, 2003. 400 с.
24. *Wald A.* Sequential analysis. New York: John Wiley & Sons. Inc. Press, 1947. 212 p.
25. *Sirjaev A. N.* Statistical sequential analysis: Optimal stopping rules. Providence, Rhode Island: American Mathematical Society, 1973. 174 p.
26. *Milligan G. W.* Ultrametric hierarchical clustering algorithms // *Psychometrika.* 1979. Vol. 44. Iss. 3. P. 343–346.

Статья поступила в редакцию 27 мая 2023 г.

Статья принята к печати 12 октября 2023 г.

Контактная информация:

Орехов Андрей Владимирович — ст. преп.; a.orehov@mail.ru

Шишкин Виктор Иванович — д-р мед. наук, проф.; v.shishkin@spbu.ru

Кудрявцева Галина Васильевна — д-р биол. наук, проф.; g.v.kudryavtseva@spbu.ru

Павилайнен Галина Вольдемаровна — канд. физ.-мат. наук, доц.; g_v_pavilaynen@mail.ru

Шишкин Виктор Викторович — д-р экон. наук, доц.; viktor.shishkin@spbu.ru

Людкевич Николай Сергеевич — аспирант; nikolai.lyudkevich@gmail.com

Automatically obtaining by methods of flow cytometry and cluster analysis simplified leukocyte formula

A. V. Orekhov, V. I. Shishkin, G. V. Kudriavtseva, G. V. Pavilaynen, V. V. Shishkin, N. S. Lyudkevich

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Orekhov A. V., Shishkin V. I., Kudriavtseva G. V., Pavilaynen G. V., Shishkin V. V., Lyudkevich N. S. Automatically obtaining by methods of flow cytometry and cluster

analysis simplified leukocyte formula. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2023, vol. 19, iss. 4, pp. 469–483. <https://doi.org/10.21638/11701/spbu10.2023.404> (In Russian)

The leukocyte formula is the percentage of different groups of white blood cells. According to morphological features, three subpopulations can be distinguished among leukocytes: lymphocytes, monocytes and granulocytes. Granulocytes are divided into neutrophilic, eosinophilic, and basophilic cells. Automatic typologization of white blood cells is an unsolved problem, since at present, during cytometric research, the counting of the number of cells in various subpopulations of leukocytes is actually done manually, which in turn causes the subjectivity of the experiment and large values of errors in calculations. To solve this problem, attempts have been made repeatedly to use cluster analysis methods. In computational experiments, it was shown that the use of standard algorithms, such as the agglomerative methods, EM algorithm, DBSCAN, etc., does not allow to obtain the desired results. In recent years, a large number of research papers have been published describing specialized clustering algorithms for detecting and determining populations of white blood cells, some of them have found practical application, but the problems associated with the presence of a large amount of noise and different data density distribution during leukocyte clustering by flow cytometry methods remain relevant. The article considers an approach to constructing a strategy for automatic allocation of the main leukocyte subpopulations using a modified agglomerative centroid clustering method and discusses the results of computational experiments. The results of calculating the proportion of lymphocytes are compared “manually” and automatically using a modified centroid algorithm.

Keywords: leukocyte formula, flow cytometry, cluster analysis, Markov moment, least squares method.

References

1. Zurochka A. V., Khaidukov S. V., Kudryavtsev I. V., Chereshev V. A. *Protochnaya tsitometriya v meditsine i biologii* [Flow cytometry in medicine and biology]. 2nd ed. Ekaterinburg, Ural. dept. RAS Publ., 2014, 574 p. (In Russian)
2. Balalaeva I. V. *Protochnaya tsitofluorimetriya*. Ucheb.-metod. posobiye [Flow cytometry. Educational method. allowance]. Nizhny Novgorod, Nizhny Novgorod State University Press, 2014, 75 p. (In Russian)
3. Agadzhanian N. A., Smirnov V. M. *Normal'naya fiziologiya*. Uchebnik dlya studentov meditsinskikh vuzov [Normal physiology. A textbook for medical students]. Moscow, Medical Information Agency Press, 2009, 520 p. (In Russian)
4. Khaitov R. M., Ignatieva G. A., Sidorovich I. G. *Immunologiya*. Uchebnik [Immunology. Textbook]. Moscow, Medicine Publ., 2000, 432 p. (In Russian)
5. Orekhov A. V., Shishkin V. I., Lyudkevich N. S. Clusterization of white blood cells on the modified upgmc method. *Stability and Control Processes. Proceedings of the 4th International Conference dedicated to the memory of professor Vladimir Zubov*. Cham, Springer Publ., 2022, pp. 559–566.
6. *Osnovy dokazatel'noy meditsiny*. Uchebnoe posobiye dlya sistemy poslevuzovskogo i dopolnitel'nogo professional'nogo obrazovaniya vrachey [Fundamentals of evidence-based medicine. Textbook manual for the system of postgraduate and additional professional education of doctors]. Ed. by R. G. Oganova. Moscow, Silicea-Poligraf Publ., 2010, 136 p. (In Russian)
7. Pedersen N. W., Chandran P. A., Qian Y., Rebhahn J., Petersen N. V., Hoff M. D., White S., Lee A. J., Stanton R. H. Ch., Jakobsen K., Mosmann T., Gouttefangeas C., Chan C., Scheuermann R. H., Hadrup S. R. Automated analysis of flow cytometry data to reduce inter-lab variation in the detection of major histocompatibility complex multimer-binding T cells. *Front Immunol*, 2017, vol. 8, p. 858.
8. Daneau G., Buyze J., Wade D., Diaw P. A., Dieye T. N., Sopheak T., Florence E., Lynen L., Kestens L. CD4 results with a bias larger than hundred cells per microliter can have a significant impact on the clinical decision during treatment initiation of HIV patients. *Cytometry B Clin Cytom*, 2017, vol. 92, no. 6, pp. 476–484. <https://doi.org/10.1002/cyto.b.21366>
9. Qian Y., Kim H., Purawat Sh., Wang J., Stanton R., Lee A., Xu W., Altintas I., Sinkovits R., Scheuermann R. H. FlowGate: Towards extensible and scalable web-based flow cytometry data analysis.

XSEDE '15: Proceedings of the 2015 XSEDE Conference. Scientific advancements enabled by enhanced cyberinfrastructure, July 2015, art. no. 5, pp. 1–8. <https://doi.org/10.1145/2792745.2792750>

10. Omana-Zapata I., Mutschmann C., Schmitz J., Gibson S., Judge K., Indig M. A., Lu B., Tauffman D., Sanfilippo A. M., Shallenberger W., Graminske Sh., McLean R., Hsen R. I., d'Empaire N., Dean K., O'Gorman M. Accurate and reproducible enumeration of T-, B-, and NK lymphocytes using the BD FACSLyric 10-color system: A multisite clinical evaluation. *PLoS One*, 2019, vol. 14, no. 1, art. no. e0211207. <https://doi.org/10.1371/journal.pone.0211207>

11. Lepsky A. I. Sravnitel'nyy analiz algoritmov klasterizatsii leykotsitov po FS i SS parametram pri tsitofluorimetricheskom issledovanii krovi [Comparative analysis of leukocyte clustering algorithms according to FS and SS parameters in cytofluorometric blood tests]. *Information technologies*, 2020, vol. 26, no. 1, pp. 56–61. (In Russian)

12. Steinhaus H. Sur la division des corps materiels en parties. *Bull. Acad. Polon. Sci. C1. III*, 1956, vol. IV, pp. 801–804.

13. Lloyd S. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 1982, vol. 28, iss. 2, pp. 129–137. <https://doi.org/10.1109/TIT.1982.1056489>

14. Dempster A. P., Laird N. M., Rubin D. B. Maximum Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 1977, vol. 39, iss. 1, pp. 1–38.

15. Everitt B. S. *Cluster analysis*. Chichester, John Wiley & Sons Ltd Publ., 2011, 330 p.

16. Hartigan J. A. *Clustering algorithms*. New York, London, Sydney, Toronto, John Wiley & Sons Inc. Press, 1975, 351 p.

17. Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on knowledge discovery and data mining (KDD-96)*. Eds E. Simoudis, J. Han, U. M. Fayyad. Portland, AAAI Press, 1996, pp. 226–231.

18. Weber L. M., Robinson M. D. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry, Pt A*, 2016, vol. 89A, iss. 12, pp. 1084–1096.

19. Zhang C., Xiao X., Li X., Chen Y.-J., Zhen W., Chang J., Zheng Ch., Liu Z. White blood cell segmentation by color-space-based K-means clustering. *Sensors*, 2014, vol. 14, iss. 9, pp. 16128–16147.

20. Vil M. Yu. Analiz statisticheskoy svyazi mezhdru klinicheskimi faktorami i poyavleniyem anomal'noy subpopulyatsii leykotsitov [Analysis of the statistical relationship between clinical factors and the appearance of an abnormal subpopulation of leukocytes]. *Management Processes and Sustainability*, 2020, vol. 7, no. 1, pp. 143–147. (In Russian)

21. Orekhov A. V. Markovskii moment ostanovki aglomerativnogo protsesssa klasterizatsii v evklidovom prostranstve [Markov moment for the agglomerative method of clustering in Euclidean space]. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 1, pp. 76–92. <https://doi.org/10.21638/11702/spbu10.2019.106> (In Russian)

22. Orekhov A. V. Quasi-deterministic processes with monotonic trajectories and unsupervised machine learning. *Mathematics*, 2021, vol. 9, art. no. 2301. <https://doi.org/10.3390/math9182301>

23. Bulinsky A. V., Shiryaev A. N. *Teoriya sluchaynykh protsessov [Theory of random processes]*. Moscow, Fizmatlit Publ., 2003, 400 p. (In Russian)

24. Wald A. *Sequential analysis*. New York, John Wiley & Sons. Inc. Press, 1947, 212 p.

25. Sirjaev A. N. *Statistical sequential analysis: Optimal stopping rules*. Providence, Rhode Island, American Mathematical Society Publ., 1973, 174 p.

26. Milligan G. W. Ultrametric hierarchical clustering algorithms. *Psychometrika*, 1979, vol. 44, iss. 3, pp. 343–346.

Received: May 27, 2023.

Accepted: October 12, 2023.

Authors' information:

Andrey V. Orekhov — Senior Lecturer; a.orehov@mail.ru

Viktor I. Shishkin — Dr. Sci. in Medicine, Professor; v.shishkin@spbu.ru

Galina V. Kudryavtseva — Dr. Sci. in Biology, Professor; g.v.kudryavtseva@spbu.ru

Galina V. Pavilaynen — PhD in Physics and Mathematics, Associate Professor;
g_v_pavilaynen@mail.ru

Viktor V. Shishkin — Dr. Sci. in Economics, Associate Professor; viktor.shishkin@spbu.ru

Nikolay S. Lyudkevich — Postgraduate Student; nikolai.lyudkevich@gmail.com