

УДК 004.932 + 612.1

Г.В. БОНДАРЬ (д-р мед. наук, проф.)²,
А.Б. СТУПИН (д-р техн. наук, проф.)¹,
В.П. ШЕВЧЕНКО (д-р физ.-мат. наук, проф.)¹,
А.А. ПОЛИВАНОВ¹,
В.С. ОВЕРКО (ст. науч. сотр.)¹
¹Донецкий национальный университет,
²Донецкий областной противоопухолевый центр

ПРИМЕНЕНИЕ КЛАСТЕРНЫХ МЕТОДОВ ДЛЯ ИДЕНТИФИКАЦИИ ФОРМЕННЫХ ЭЛЕМЕНТОВ В «ЖИВОЙ» КАПЛЕ КРОВИ

В статье изложены основные положения, касающиеся построения нейросетевых алгоритмов, применяющихся для решения данной проблемы. Дано их подробное описание, которое позволит другим исследователям в полной мере использовать полученные результаты. Авторы успешно аргументируют свой подход к решению поставленной задачи с помощью наглядного представления результатов расчетов.

Ключевые слова: *нейронная сеть, обработка изображений, кластеризация, кровь, форменные элементы крови*

При исследовании влияния гидродинамически-активных добавок на движение крови нужно учитывать то обстоятельство, что кровь является неньютоновской жидкостью. Принято считать, что основным источником, вызывающим отклонение гидродинамических свойств крови от свойств обычных ньютоновских жидкостей, является присутствие в крови различных форменных элементов (эритроциты, кристаллы холестерина, кристаллы солей фосфорной и мочево́й кислот и др.) [1]. Эритроциты имеют способность к агрегации и дезагрегации при различных режимах течения крови в сосудах. Это свойство эритроцитов, а также неупорядоченное движение кристаллов холестерина, солей фосфорной кислоты и др. могут привести к изменению осредненных гидродинамических характеристик потока крови в сосудах и, как следствие, к изменению гидравлических сопротивлений последних.

Поэтому распознавание и идентификация форменных элементов в «живой» капле крови является важным дополнением к гидродинамическим исследованиям крови (без и с гидродинамически-активными добавками).

Известен метод функциональной оценки форменных элементов крови, проводимый с помощью достаточно мощного (увеличение в 2400 раз) темнопольного микроскопа [2]. Однако, на результаты указанных исследований значительное влияние оказывает субъективность восприятия оператора, в связи с чем предлагается использовать методы компьютерной идентификации и распознавания форменных элементов «живой капли» крови. Данные методы могут дать не только качественные, но и количественные характеристики. В частности: определить плотность, цветность, количество и размеры кристаллов холестерина; форму и число агрегаций эритроцитов; количество и размеры кристаллов мочево́й и фосфорной кислот и др.

В работе [3] такая идентификация осуществлялась с помощью нейросетевого метода. Его суть состоит в следующем: перед началом работы программы в базу данных заносятся картинки, которые представляют собой эталонные образцы искомым включений (кристаллы холестерина, мутагены и т.д.) – создается так называемое обучающее множество. Далее то изображение, которое нужно проанализировать, делится на фрагменты, каждый из которых сравнивается с любым из эталонных образцов по формуле вероятностных нейронных сетей. При этом сопоставляются числовые характеристики текущего фрагмента и элемента обучающего множества, в качестве которых выступают усредненные значения интенсивностей красного, зеленого и синего цветов (RGB-палитра).

Однако, данному подходу присущи определенные недостатки. В частности, крайне сложно изначально подобрать все те эталонные образцы, которые полностью бы исчерпывали включения, встречающиеся в плазме крови. Кроме того, при делении изображения на равные части во фрагмент могут попадать элементы сразу нескольких включений, что затрудняет идентификацию.

Решить эти вопросы можно, если применить подход «распознавание образов без учителя» (в то время, как вероятностные нейронные сети относятся к «обучению с учителем»). Этот подход называется кластерным анализом [4].

Кластерный анализ представляет собой совокупность подходов, методов и алгоритмов, предназначенных для нахождения некоторого разбиения исследуемой совокупности объектов на подмножества относительно сходных, похожих между собой объектов. При этом исходным допущением для выделения таких подмножеств, получивших специальное название кластеров, служит предположение о том, что объекты, относимые к одному кластеру, должны иметь большее сходство между собой, чем с объектами из других кластеров.

Пусть исследуемая совокупность данных представляет собой конечное множество элементов $A = \{a_1, a_2, \dots, a_n\}$, которое называется множеством объектов кластеризации [6]. В рассмотрение

также вводится конечное множество признаков или атрибутов $P = \{p_1, p_2, \dots, p_q\}$, каждый из которых количественно представляет некоторое свойство или характеристику элементов рассматриваемой проблемной области. При этом натуральное n определяет общее количество объектов данных, а натуральное q – общее количество измеримых признаков объектов. Далее предполагается, что для каждого из объектов кластеризации измерены все признаки множества P в некоторой количественной шкале. Тем самым каждому из элементов $a_i \in A$ поставлен в соответствие вектор $x_i = (x_1^i, x_2^i, \dots, x_q^i)$, где x_j^i – количественное значение признака $p_j \in P$.

Задача нечеткого кластерного анализа формулируется так: на основе исходных данных x определить такое нечеткое разбиение на заданное число нечетких кластеров C , которое доставляет экстремум целевой функции f . В качестве целевой функции рассматривается сумма квадратов отклонений координат объектов кластеризации x от центров искоемых нечетких кластеров v_k .

Перенесем этот понятийный аппарат на реалии нашей задачи. Множество A – это совокупность фрагментов, на которые делится анализируемое изображение. Множество P в нашем случае состоит из трех признаков – интенсивности красного, зеленого и синего цветов для каждого из фрагментов. Количественное значение этих признаков x_i равняется среднему значению всех пикселей, составляющих фрагмент изображения. В свою очередь, подмножества, на которые предстоит разбить совокупность фрагментов A – это искомые форменные элементы (ими могут являться эритроциты, кристаллы холестерина, мочевой кислота и др.).

В прикладном пакете Matlab, в котором разрабатывалась программа обработки изображения, задача нечеткой кластеризации решается двумя функциями – fcm (метод нечетких средних) и subclust (метод субтрактивной нечеткой кластеризации). Они принципиально различаются. Функция fcm возвращает координаты центров кластеров и степень принадлежности каждого объекта из множества A к этим кластерам. Важно отметить, что степень принадлежности указывается нечетко – не 0/1 (не принадлежит/принадлежит), а, скажем, 0.4 (принадлежит в определенной степени). Именно этим обуславливается концептуальная взаимосвязь между кластерным анализом и теорией нечетких множеств.

Метод fcm применительно к нашей задаче имеет существенный недостаток – он требует указывать точное число искоемых кластеров. Но мы не можем знать это априори, так как на изображении может быть три вида включений, пять или же одно. В этом случае необходимо

воспользоваться методом *subclust*, который позволяет найти оптимальное число кластеров для данной совокупности объектов. Однако, эта функция не возвращает степени принадлежности (только координаты центров кластеров), а значит, нельзя определить, к какому классу относится данный фрагмент изображения.

Мы предлагаем комбинированное использование этих процедур – сперва *subclust* определяет оптимальное количество кластеров для наших исходных данных, а затем этот параметр передается программе *fcm*, которая возвращает нам массив со значениями функций принадлежности. Стоит напомнить, что для объектов степень принадлежности указывается нечетко – например, в случае наличия трех кластеров объект a_1 может принадлежать подмножеству C_1 со степенью 0.5, подмножеству C_2 со степенью 0.3 и подмножеству C_3 со степенью 0.2 (их сумма должна равняться единице). Тогда можно применить процесс дефаззификации (метод максимального значения функции принадлежности) и считать, что данный объект принадлежит подмножеству C_1 и никакому другому.

Еще один важный момент, который необходимо осветить – это количество объектов нечеткой кластеризации n (применительно к нашей задаче – количество фрагментов, на которое дробится изображение). В работе [3] любое изображение (вне зависимости от своего типа) делилось на n равных частей. Но тогда не исключена ситуация, при которой у двух фрагментов будут примерно одинаковые усредненные характеристики по RGB-интенсивностям, хотя разброс значений будет совершенно разным (скажем, 0..200 против 90..110). В таком случае программа отнесет к одному кластеру различные по своей природе фрагменты, чего нужно избежать.

Чтобы разрешить эту ситуацию, нами был разработан рекурсивный метод неравномерной фрагментации. На первом шаге целиком анализируется все изображение, и если его среднеквадратичное отклонение по RGB-интенсивностям не превышает определенной величины ε , то процесс завершается и это изображение будет составлять единственный объект нашего множества кластеризации. В противном случае изображение делится на 4 равных части, каждая из которых рассматривается аналогично, и т.д. Процесс лимитируется только площадью дробления.

Рассмотрим действие этих алгоритмов на практике.

а) идентификация кристаллов фосфорной кислоты

Кристаллы солей фосфорной кислоты в изображении, полученном от темнопольного микроскопа, выделяются черным цветом (рис. 1).

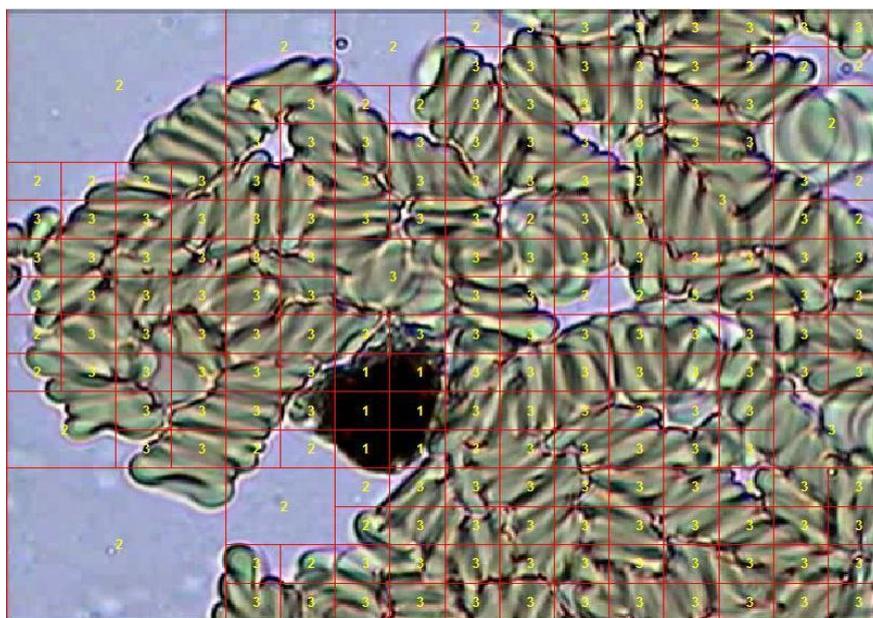


Рисунок 1 – Идентификация кристаллов солей фосфорной кислоты

Кристаллы солей фосфорной кислоты отображаются фрагментами с цифрой “1”, ко второму классу относится фон, а к третьему – эритроциты. Программа подсчитывает площадь этих включений – в данном случае класс 1 занимает 2,4% от площади изображения, класс 2 – 26,5%, а класс 3 (эритроциты) – 71,1%. Диаграмма распределения числовых характеристик фрагментов изображения представлена на рис. 2.

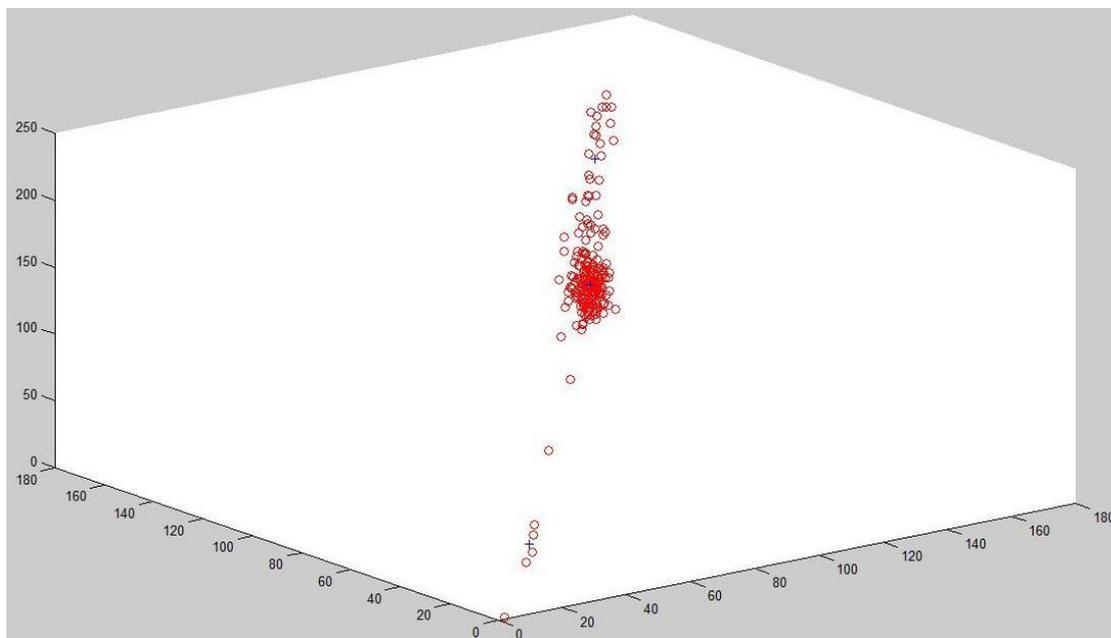


Рисунок 2 – Распределение RGB-интенсивностей фрагментов изображения

Синий плюс в нижней части рисунка обозначает центр самого малочисленного кластера – кластера фрагментов фосфорной кислоты (класс 1). Плюс в середине рисунка указывает местоположение центра кластера эритроцитов (класс 3), а плюс в верхней части изображения относится центру кластера для фоновых фрагментов (класс 2).

- b) идентификация кристаллов холестерина
- c)

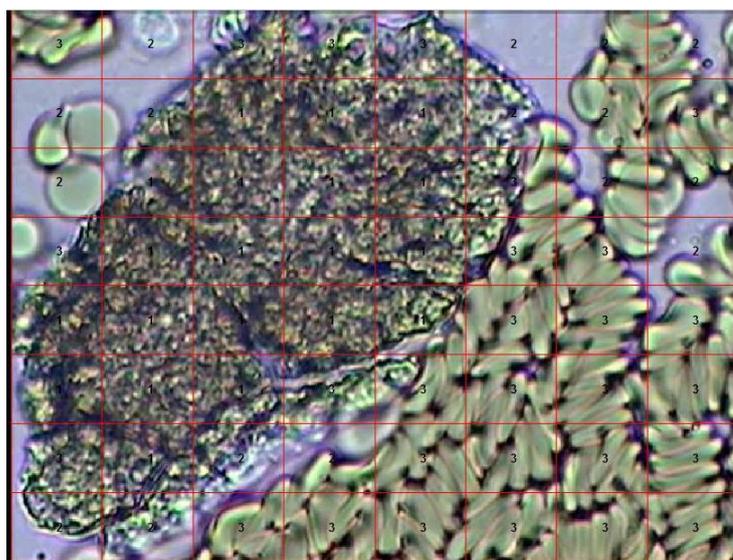


Рисунок 3 – Идентификация кристаллов холестерина

На рис. 3 фрагменты холестеринового включения отмечены цифрами “1”, их площадь составляет чуть менее 32 процентов. Также нами была разработана программа, которая очерчивает контуры включений более наглядно (рис. 4):



Рисунок 4- Очерченные контуры включений на изображении капли крови

d) идентификация мочевой кислоты

В качестве примера на рис. 5 представлена идентификация кристаллов мочевой кислоты, которые отображаются фрагментами с цифрой 1.

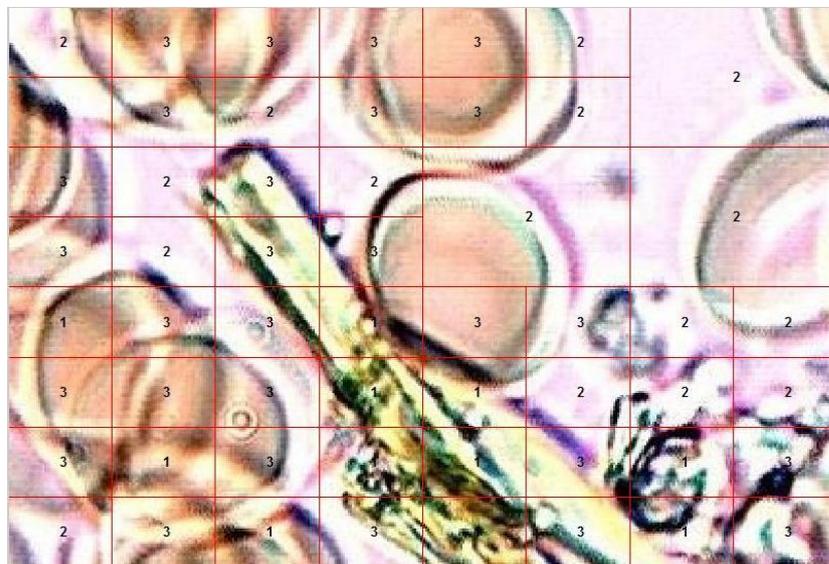


Рисунок 5 – Идентификация кристаллов мочевой кислоты

Таким образом, применение кластерных методов может быть эффективным для идентификации форменных элементов в изображении «живой капли» крови на темнопольном микроскопе.

Список литературы:

1. Отличия метода гемосканирования крови от клинического анализа крови [Электронный ресурс]. – Режим доступа: npanchenko.ru/diagnostika/gemoscanirovanie.html.
2. Грейндж Курт. Темнопольный микроскоп и его возможности в клинической практике / Курт Грейндж // Мат. журнала «Натуральная фармакология и косметология». – 2006. - №3.
3. Идентификация и распознавание объектов «blood-garbage» в крови человека / [Г.В.Бондарь, А.Б.Ступин, В.С.Оверко, А.А.Поливанов] // Труды Первого международного научно-практического семинара «Повховские научные чтения». – Донецк: ДонНУ, 2010. – С. 163-170.
4. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. – СПб.: БХВ-Петербург, 2005. – 736 с.

Надійшла до редакції 29.03.2013

Г.В. БОНДАРЬ, О.Б. СТУПИН, В.П. ШЕВЧЕНКО, О.О. ПОЛИВАНОВ, В.С. ОВЕРКО

ЗАСТОСУВАННЯ КЛАСТЕРНИХ МЕТОДІВ ДЛЯ ІДЕНТИФІКАЦІЇ ФОРМЕННИХ ЕЛЕМЕНТІВ В " ЖИВІЙ" КРАПЛІ КРОВІ

У статті викладені основні положення, що стосуються побудови нейромережесвих алгоритмів, що застосовуються для вирішення даної проблеми. Дано їх докладний опис, який дозволить іншим дослідникам в повній мірі використовувати отримані результати. Автори успішно аргументують свій підхід до вирішення поставленого завдання за допомогою наочного представлення результатів розрахунків.

Ключові слова: нейронна мережа, обробка зображень, кластеризація, кров, форменні елементи крові.

G. BONDAR, A. STUPIN, V. SHEVCHENKO, A. POLIVANOV, V. OVERKO

APPLICATION OF CLUSTER METHODS FOR IDENTIFICATION OF BLOOD CORPUSCLES IN A "LIVE" DROP OF THE BLOOD

For functional assessment of blood cells, the analysis employing dark-field microscope is applied. However, the results of these studies are strongly influenced by subjective perception of the researcher, and therefore we propose to use the techniques of computer identification and recognition of blood corpuscles in a "live drop" of blood. These methods can provide not only qualitative but also quantitative characteristics.

In most cases, the identification is performed using the neural network method. However, this approach has certain disadvantages. To solve this problem we suggest applying the cluster analysis. We have developed a recursive uneven fragmentation. First the entire image is analyzed, and if the standard deviation for RGB-intensity does not exceed a certain value, then the process is completed and the image will be the only object of our set of clustering. Otherwise, the image is divided into four equal parts, each of which is treated similarly, etc.

Keywords: neural network, image processing, clusterization, blood, blood corpuscles.

© Бондарь Г.В., Ступин А.Б., Шевченко В.П., Поливанов А.А., Оверко В.С., 2013