

ционном способом разбиения ряда на несколько возрастных групп. Процесс дифференцировки и созревания миелоидных клеток имеет непрерывный и монотонный характер. Поэтому разбиение клеточного ряда на ступени созревания, по меньшей мере, неадекватно и является лишь первым приближением при реализации популяционного подхода в анализе клеточных множеств. Использование же непрерывного количественного критерия зрелости более соответствует характеру биологических процессов созревания клеток.

Этот критерий может найти применение в диагностике ряда гематологических заболеваний, связанных с нарушениями процессов дифференцировки гемопоэтических клеток.

ЛИТЕРАТУРА

1. Кассирский И. А. Клиническая гематология.— М.: Медицина, 1970.
2. Teick J. M., Young I. T. Transformation of nuclear morphology during cellular maturation // J. Histochem. Cytochem.— 1979.— 27, N 1.— P. 193.
3. Gray S. B. Local properties of binary imager in two dimensions // IEEE Trans. Computers.— 1971.— C-20, N 5.— P. 551.
4. Киричук В. С. Алгоритмы анализа клеточных ядер // Автоматизированный анализ клеточных популяций.— Новосибирск: ИАиЭ СО АН СССР, 1978.
5. Косых В. П. Структура, аппаратное и прикладное программное обеспечение автоматизированного комплекса для анализа микроизображений: Автореф. дис. ... канд. техн. наук.— Новосибирск, 1985.
6. Serra J. Image Analysis and Mathematical Morphology.— L.; N. Y.: Academ. Press, 1982.
7. Косых В. П., Пустовских А. И., Тарасов Е. В., Яковенко Н. С. Морфологический процессор // Автометрия.— 1984.— № 4.

Поступила в редакцию 12 мая 1988 г.

УДК 681.3 : 681.7.08 : 576.3

И. А. КИРИЧУК, В. П. КОСЫХ, А. И. ПЕТУНИН

(Красноярск — Новосибирск)

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ ЛЕЙКОЦИТОВ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ КОМПЛЕКСА «ЗЕНИТ-К»

Созданный в ИАиЭ СО АН СССР комплекс для автоматизированного анализа микроизображений [1] в различных его модификациях в течение ряда лет использовался при исследовании клеточных популяций. В настоящей статье приводятся результаты применения комплекса для морфометрического анализа клеток крови, который играет важную роль в диагностике гематологических, ряда опухолевых и инфекционных заболеваний. В клинической и исследовательской практике с этой целью наиболее часто используется так называемый дифференциальный счет лейкоцитов, который представляет собой исследование 100—200 лейкоцитов на специально окрашенном препарате с целью разделения клеток на несколько отличных по морфологии классов [2]. Результаты такого анализа подвержены значительным вариациям: во-первых, в силу случайных причин, поскольку объем выборки весьма мал, во-вторых, при визуальных исследованиях фиксируются лишь качественные характеристики, в-третьих, на результат анализа большое влияние оказывают опыт и интуиция исследователя. Другой недостаток «ручной» обработки — низкая скорость анализа. Вполне закономерно поэтому желание цитологов получить более объективные и эффективные методы анализа клеток крови.

В результате исследований, выполненных в основном в прошлом десятилетии, за рубежом разработан ряд установок для автоматического

дифференциального счета лейкоцитов [3, 4]. Отсутствие аналогичной отечественной аппаратуры явилось основной причиной для проведения работы, результаты которой изложены ниже. Цель данного этапа заключалась в создании алгоритмов предобработки изображений клеток крови и поиске таких параметров, в пространстве которых клетки различных классов образуют разделимые кластеры. Задача обеспечения производительности, достаточной для проведения рутинных анализов, здесь не ставилась.

Скорость автоматического дифференциального счета лейкоцитов с использованием комплекса «Зенит-К» ограничивается в основном на этапах поиска и сегментации клеток. При поиске приходится целиком просматривать каждое поле зрения, хотя и с большим шагом (4 мкм), что при низкой концентрации лейкоцитов на мазке приводит к непроизводительным затратам времени. Низкая концентрация лейкоцитов характерна для традиционных методик приготовления мазка. Серьезный недостаток мазков — низкое качество поверхности предметных стекол, требующее перефокусировки при каждой смене поля зрения, — тоже приводит к значительным затратам времени на этапе поиска. В среднем на поиск лейкоцита затрачивается около 4 с, что совершенно неприемлемо при рутинном анализе.

Наиболее трудоемкий этап сегментации — отделение лейкоцитов от соприкасающихся с ними эритроцитов. Сложность этого этапа обусловлена отсутствием в данном варианте денситометра цветоделения, в связи с чем разделение соприкасающихся клеток опирается только на анализ формы границ, который требует приблизительно 15 с на клетку.

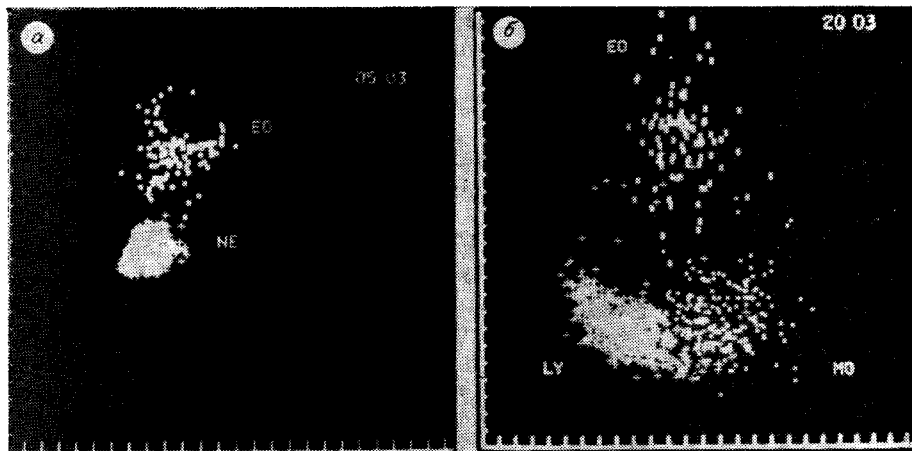
Тем не менее разработка классификатора с использованием имеющегося оборудования послужила решению ряда методических задач. Во-первых, при создании и эксплуатации классификатора определены требования к качеству мазков, пригодных для автоматического анализа, и выявлены основные времяемкие этапы, для сокращения которых необходимо развитие аппаратной поддержки. Во-вторых, реализован полный цикл анализа, разработаны алгоритмы всех его этапов, что может послужить основой для создания специализированного устройства для рутинных анализов. В-третьих, результаты автоматической классификации могут использоваться при отборе нетрадиционных морфометрических параметров, а также при проверке соответствия морфометрических параметров визуальным характеристикам клеток. Наконец, автоматическая классификация и методы кластерного анализа могут служить основой для выделения субпопуляций клеток, не выявляемых традиционными методами.

Вид классификатора. В настоящей работе были поставлены задачи построения классификатора для четырех нормальных классов белой крови: нейтрофильных гранулоцитов (NE), эозинофильных гранулоцитов (EO), лимфоцитов (LY), моноцитов (MO) — и оценивания ошибок классификации. Базофильные гранулоциты не рассматривались из-за малочисленности полученной выборки. С формальной точки зрения задача классификации клеток сводится к построению разделяющих поверхностей для распределений исследуемых клеточных классов в пространстве измеряемых параметров и определению решающего правила, относящего произвольный экспериментальный вектор параметров X к распределению соответствующего класса.

Процедура выбора вида классификатора, оптимального в смысле минимизации вероятности ошибки, описана, например, К. Фукунага [5]. Классификатором такого вида является решающее правило, опирающееся на байесовское отношение правдоподобия:

$$L(X) = \frac{P(X/\omega_i)}{P(X/\omega_j)} \geq \frac{P(\omega_j)}{P(\omega_i)} \rightarrow X \in \begin{cases} \omega_i, & i \neq j, \\ \omega_j, & \end{cases}$$

где $P(\omega_i)$, $P(\omega_j)$ — априорные вероятности событий ω_i , ω_j (в нашем



Разделимость лейкоцитарных классов:

а — по оси абсцисс — дисперсия оптической плотности цитоплазмы, по оси ординат — оптическая плотность цитоплазмы: □ — эозинофилы; + — нейтрофилы; *б* — по оси абсцисс — площадь ядра, по оси ординат — контраст ядро/цитоплазма: □ — эозинофилы (EO); □ — моноциты (MO); + — лимфоциты (LY)

случае принадлежность к классу ω_i, ω_j ; $P(X/\omega_i), P(X/\omega_j)$ — соответственно вычисляемые апостериорные вероятности. В случае нормальных распределений с математическими ожиданиями μ_i, μ_j и ковариационными матрицами V_i, V_j приведенное к логарифмической форме решающее правило приобретает следующий вид:

$$H(X) = \ln L(X) = \frac{1}{2} (X - \mu_i)^T V_i^{-1} (X - \mu_i) - \frac{1}{2} (X - \mu_j)^T V_j^{-1} (X - \mu_j) + \frac{1}{2} \ln \frac{|V_i|}{|V_j|} \leq \ln \frac{P(\omega_i)}{P(\omega_j)} \rightarrow X \in \left\{ \begin{array}{l} \omega_i \\ \omega_j \end{array} \right.$$

Исходя из предположений о нормальности распределений рассматриваемых классов клеток (NE, LY, EO, MO) и равенства априорных вероятностей для каждого класса программно было реализовано байесовское решающее правило в виде

$$H(X) = \min [(X - \hat{\mu}_i)^T \hat{V}_i^{-1} (X - \hat{\mu}_i) + \ln |\hat{V}_i|], \quad i = \overline{1, k}, \quad (1)$$

где k — число классов; $\hat{\mu}_i$ — оценка вектора математического ожидания i -го класса; \hat{V}_i — оценка ковариационной матрицы i -го класса. Процедура «обучения» классификатора вида (1) сводится, таким образом, к получению оценок $\hat{\mu}_i$ и \hat{V}_i по выборке клеток, классифицированных предварительно специалистом-цитологом.

Оценка ошибок классификации. Оценивание вероятностей ошибок классификации проводилось по методу «скользящего экзамена», заключающемуся в следующем. Из выборки, содержащей N объектов, исключается один объект. По оставшимся $N - 1$ оценивается вектор математических ожиданий и ковариационная матрица. Неиспользованный объект классифицируется по правилу, описанному выше. Процедура повторяется N раз с подсчетом неправильно классифицированных объектов для каждого класса.

Этот метод весьма экономичен в смысле использования имеющейся выборки как для синтеза классификатора, так и для оценки его качества. Метод «скользящего экзамена» в общем случае дает оценку верхней границы вероятности ошибки [5].

Отдельным и важным вопросом при разработке автоматического классификатора является оптимизация набора классификационных параметров. В данной работе классификационные параметры выбирались

Таблица 1

	NE	LY	MO	EO	N_1	NE	LY	MO	EO	P_1
NE	628	14	79	4	97	0,866	0,019	0,109	0,005	0,134
LY	4	362	17	3	24	0,010	0,938	0,044	0,008	0,062
MO	13	4	137	2	19	0,083	0,026	0,878	0,013	0,121
EO	0	0	0	100	0	0,000	0,000	0,000	1,000	0,000
N_2	17	18	96	9						

Примечание. Число параметров 4.

Таблица 2

	NE	LY	MO	EO	N_1	NE	LY	MO	EO	P_1
NE	697	12	6	10	28	0,961	0,017	0,083	0,014	0,039
LY	5	372	7	2	14	0,013	0,964	0,018	0,005	0,036
MO	1	2	153	0	3	0,006	0,012	0,980	0,000	0,019
EO	0	0	1	99	1	0,000	0,000	0,010	0,990	0,000
N_2	6	14	14	12						

Примечание. Число параметров 7.

Таблица 3

	NE	LY	MO	EO	N_1	NE	LY	MO	EO	P_1
NE	704	12	5	4	21	0,971	0,017	0,007	0,005	0,029
LY	2	378	5	1	8	0,005	0,979	0,013	0,003	0,020
MO	1	2	153	0	3	0,006	0,013	0,980	0,000	0,019
EO	0	0	1	99	1	0,000	0,000	0,010	0,990	0,010
N_2	3	14	11	5						

Примечание. Число параметров 10.

на основе анализа двумерных распределений различных лейкоцитарных классов. Рисунок демонстрирует некоторые такие распределения для четырех основных классов.

Хорошо видно, что по ряду параметров наблюдается выраженное разделение рассматриваемых лейкоцитарных классов. Результатом этого анализа явился выбор десяти параметров, наилучших в смысле разделимости классов.

Результаты классификации для разного числа параметров приведены в табл. 1—3. Здесь в строках левой части таблицы указано число клеток данного типа, отнесенных классификатором в различные классы. Последняя колонка указывает количество пропусков, а последняя строка — количество ложных тревог в каждом классе. В правой части таблицы приведены соответствующие оценки вероятностей для каждого класса. Анализируя результаты классификации, следует признать, что для традиционного дифференциального счета они вполне удовлетворительны (средняя вероятность пропуска не превышала 0,02). Существенно выше вероятность ошибки второго рода («ложной тревоги»). Приведенные к реальным соотношениям лейкоцитов в крови, они достигают 0,2—0,3 для редких классов (эозинофилы и моноциты).

Разработанная процедура была использована также для классификации трех типов клеток нейтрофильного ряда: миелобластов (MB), миелоцитов (MC), сегментоядерных гранулоцитов (SG). Результаты классификации приведены в табл. 4. Как и следовало ожидать, наибольшие ошибки возникают при классификации миелоцитов, являющихся промежуточной формой. Тем не менее результаты достаточно удовлетворительны.

Таблица 4

	SG	MC	MB	N_1	SG	MC	MB	P_1
SG	369	9	1	10	0,974	0,024	0,002	0,026
MC	16	182	13	29	0,076	0,863	0,062	0,137
MB	0	2	68	2	0,000	0,029	0,971	0,028
N_2	16	11	14					

Примечание. Число параметров 9.

В заключение можно отметить, что в случае четырех нормальных классов лейкоцитов использование весьма ограниченного набора параметров (не более десяти) и байесовского решающего правила обеспечивает вполне удовлетворительную классификацию. Используемые классификационные параметры представлены лишь двумя основными группами: денситометрическими и геометрическими. В целях снижения ошибок классификации следует, по-видимому, при разработке классификатора учитывать реальные соотношения лейкоцитов в норме, например, в виде функций штрафа [5].

ЛИТЕРАТУРА

1. Косых В. П. Структура, аппаратное и прикладное программное обеспечение автоматизированного комплекса для анализа микроизображений: Автореф. дис. ... канд. техн. наук.— Новосибирск, 1985.
2. Кассирский И. А. Клиническая гематология.— М.: Медицина, 1970.
3. Green J. E. A practical application of computer recognition research. The abbot ADC-500 differential classifier // J. Hystochem. Cytochem.— 1979.— 27, N 1.— P. 174.
4. Тучко Р. Н., Anbalagan S., Lin M. C., Ornstein L. Automated leucocytes classification using cytochemically stained smears // Ibid.— 1976.— 22, N 1.— P. 178.
5. Фукунага К. Введение в статистическую теорию распознавания образов.— М.: Наука, 1979.

Поступила в редакцию 12 мая 1988 г.