

приводит к большому разбросу результатов и заметной нелинейности калибровочной кривой.

Заключение. Предварительные результаты использования макета прецизионного фотограмметрического автомата «Зенит» для решения задач фото- и астрометрии подтверждают высокие точностные параметры устройства. Структура автомата, подключенного непосредственно к ЭВМ, при наличии развитого математического обеспечения, хранилища звездного каталога и автоматического отождествления звезд снимка с каталогом [3] позволяет полностью автоматизировать фото- и астрометрические измерения. Считывание коэффициента пропускания в каждой точке дает возможность решать задачи более тонкого анализа изображений: построение эквиденсит протяженных объектов, оценивание параметров характеристической кривой фотоматериала, обработка снимков планет с компенсацией искажений, вызванных рассеянием и т. д.

ЛИТЕРАТУРА

1. Бурый Л. В., Коронкевич В. П., Нестерихин Ю. Е., Нестеров А. А., Пушной Б. М., Ткач С. Е., Щербаченко А. М. Прецизионный фотограмметрический автомат.— «Автометрия», 1974, № 4, с. 83—89.
2. Киричук В. С., Пушной Б. М., Чейдо Г. П. Построение процедур считывания изображений с выпуклыми границами.— «Автометрия», 1977, № 3, с. 65—74.
3. Воронцова Л. А., Чейдо Г. П. Алгоритм автоматического отождествления звезд снимка с каталогом.— «Автометрия», 1974, № 4, с. 103—111.

Поступила в редакцию 27 декабря 1976 г.

УДК 513.06 : 621.391 : 681.3.01

**Л. А. АНДРИАНОВ, В. С. КИРИЧУК, В. Г. КОСЫХ,
Г. П. ЧЕЙДО**

(Новосибирск)

АВТОМАТИЧЕСКИЙ АНАЛИЗ ЦИТОГРАММ

Введение. Цитологический диагноз опухолевого заболевания представляется довольно сложной проблемой, поскольку требует выделения редко встречающихся форм клеток среди достаточно большой и неоднородной популяции [1]. Автоматизация цитофотометрических исследований в значительной степени ускоряет и повышает точность определения состава цитологических препаратов [2, 3] в тех случаях, когда классификация может быть построена на основе количественных признаков, поддающихся алгоритмическому анализу (признаки формулируются на основе параметров, измеряемых конкретным прибором).

Данная работа посвящена описанию анализа клеток системой «Зенит». Приведенные результаты отражают начальный этап поиска, направленного на разработку современных методов онкодиагностики на основе автоматического анализа цитологических препаратов.

Для анализа микрообъектов требуется высокая разрешающая способность устройства. Используя оптическую систему, переносящую растр ЭЛТ на препарат с 80-кратным уменьшением, удается получить диаметр светового пятна в плоскости препарата около 4,5 мкм. Измерения, произведенные в точках, находящихся на расстоянии, равном

диаметру пятна, не коррелированы. Если же измерения производить в точках, находящихся на расстоянии $1/3$ диаметра друг от друга, то коэффициент корреляции соседних измерений не превысит 0,5 для любого пятна, обладающего центральной симметрией. Ввиду того что распределение интенсивности света в считывающем пятне хорошо известно [4], реальное разрешение системы может быть улучшено с помощью последующей математической обработки полученной информации.

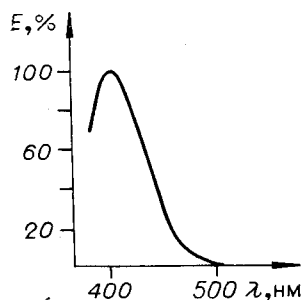


Рис. 1.

При работе с окрашиваемыми препаратами следует иметь в виду, что спектр излучения ЭЛТ, используемый в «Зените», смещен в сине-фиолетовую область (рис. 1).

Поэтому красители должны подбираться таким образом, чтобы препарат имел максимальный контраст в указанной области. Вполне приемлемым красителем, как показывает опыт, является акрединовый оранжевый в разведении 1 : 10 000 или 1 : 1000.

Алгоритм анализа препаратов и результаты. При просмотре цитологических препаратов специалистом-цитологом диагноз формулируется на основе многих признаков, реальная диагностическая ценность которых зависит от специфики каждого отдельного препарата. Алгоритмизация процесса постановки диагноза с использованием автоматизированной системы обработки препаратов должна предшествовать работе по сопоставлению параметров клеток, определяемых системой «Зенит», с признаками цитологического материала, которые используются при анализе препарата цитологом.

Альтернативный вариант, когда распознающий (диагностический) алгоритм использует критерии, не формулируемые на языке, привычном для цитологов, в настоящей работе не рассматривается.

Значительное число признаков, на которых сосредоточено внимание цитолога, связано со свойствами ядра. Настоящая работа в основном и посвящена сопоставлению характеристик ядер, получаемых с помощью системы типа «Зенит»; с характеристиками ядер, использующимися в обоснованиях цитологической диагностики.

Сопоставление величин ядер для нормальных и измененных клеток может быть проведено системой с гораздо большей точностью, чем это делает человек. Получение гистограмм площадей ядер во многих случаях позволяет (при достаточном объеме выборки) отличать повышение пролиферативной активности от присутствия совокупности клеток с укрупненными ядрами. При усиленном размножении клеток характер распределения ядерных площадей связан с временными соотношениями митотического цикла. Используя модельные представления, можно провести проверку гипотезы о наличии популяции с измененным характерным размером.

Значительное число признаков, используемых в цитологических описаниях, может быть оценено автоматизированной системой, исходя из значений периметра анализируемого объекта — ядра или клетки (круглые, овальные сегментированные ядра; округлые, полигональные, веретенообразные клетки и т. д.).

Использование данных об интегральной оптической плотности ядер при условии, что контрастирующая окраска их пропорциональна количеству хроматина, дает возможность составить представление о количестве окрашиваемого материала в ядре и может быть применена при модельном отыскании числа выделяемых клеточных популяций различной плоидности.

Структура хроматина, часто служащая цитологам отправной точкой для выделения измененных клеток, определяется как функция рас-

предела оптической плотности внутри отдельных ядер; число ядер с выраженной неравномерностью распределения, мера которой может быть приведена в соответствие с цитологическими критериями, дает представление о наличии в популяции клеток с измененной картиной окраски ядерного материала.

Наличие в препарате групп клеток, связанных в те или иные структуры, также часто служит важным диагностическим признаком. Автомат «Зенит», ведущий измерение в фиксированной системе координат, позволяет выделить группы близко лежащих ядер, учесть их взаимную ориентацию, а значит, и построить критерий, определяющий вероятность принадлежности рассматриваемой группы к единому тканевому образованию.

Ядерное плазменное отношение существенным образом зависит от функционального состояния клетки. Используя имеющиеся данные о цикличности функциональных изменений, можно на основе соотношения площадей и оптических плотностей ядер и клеток построить модельный критерий, реагирующий на функциональные сдвиги в популяции, например на наличие секретирующих клеток.

В наших исследованиях определялись следующие параметры ядер клетки:

- координаты центра,
- площадь,
- максимальный и минимальный радиус,
- ориентация максимального и минимального радиуса,
- распределение интегрального пропускания в зависимости от выбранного порога,
- радиус инерции ядра,
- периметр ядра.

Алгоритм определения всех перечисленных параметров работает следующим образом:

1) осуществляется поиск ядер клеток на заданном поле, координаты найденных ядер запоминаются (размер анализируемого поля 230×230 мкм²);

2) производится считывание найденных ядер и в память ЭВМ поступает информация о коэффициенте пропускания всех внутренних точек ядра, на основе которой вычисляются оценки всех перечисленных параметров;

3) меняется поле и вся процедура обработки повторяется.

Производительность такой системы — полный анализ ядра клетки за 5—6 с при 300—500 измерениях коэффициента пропускания на одно ядро*.

Принципиально важный вопрос при определении всех приведенных выше параметров — выделение ядра на фоне клетки. На рис. 2 приведены линии равных значений коэффициента пропускания, снятые системой с одной клетки. Расстояние между уровнями $\Delta t = 0,03$. Можно заметить, что на границе ядра и клетки происходит резкое изменение коэффициента пропускания. На рис. 3, 4 показана динамика изменения площади ядра (найденная алгоритмом) в зависимости от выбранного порогового уровня (т. е. вычислена площадь сечения ядра заданным уровнем пропускания). Шаг

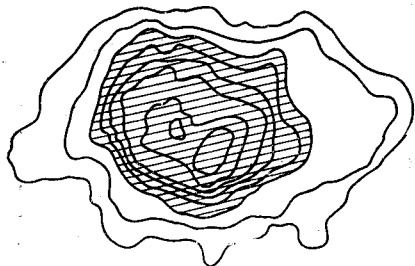


Рис. 2.

* Данные относятся к системе с ЭВМ «Минск-22», в настоящее время замененной в 5—6 раз более производительной М-4030.

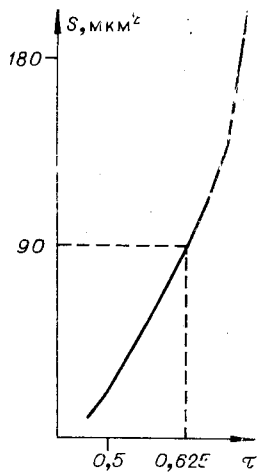


Рис. 3.

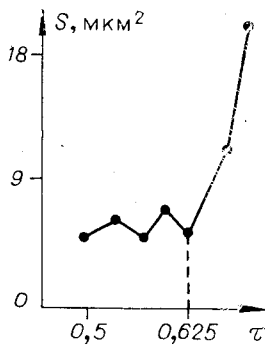


Рис. 4.

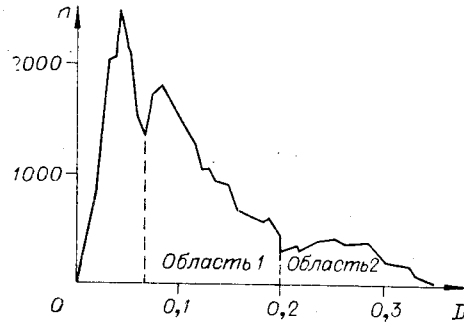


Рис. 5.

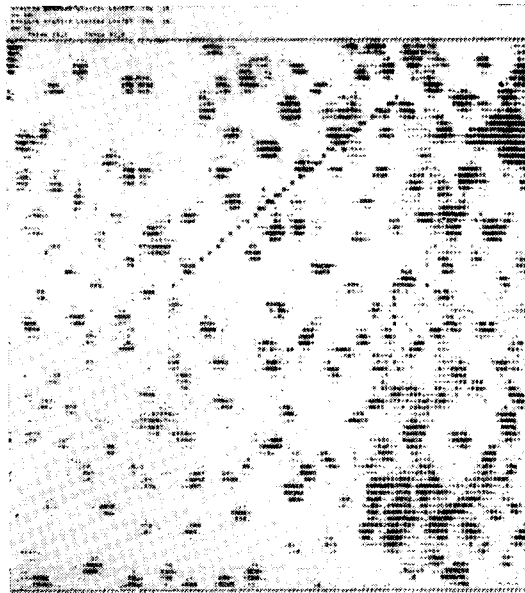


Рис. 6.

по уровню пропускания равен 0,03. На рис. 4 видно резкое возрастание площади при изменении порогового уровня от 0,625 до 0,66. Следовательно, границей клетки и ядра здесь можно считать уровень пропускания, равный 0,625. Это подтверждается и гистограммой распределения оптических плотностей той же клетки (рис. 5); гистограмма построена по результатам считывания поля размерами 38×38 мкм², содержащего анализируемую клетку. Первый пик гистограммы соответствует оптической плотности стеклянной подложки. Область 1 представляет оптическую плотность тела клетки, область 2 относится к ядру. Переход от клетки к ядру происходит на уровне оптической плотности $D=0,2$ ($\tau=0,625$). На поиске этого скачка градиента площадей при варьировании порогового уровня и основан алгоритм выделения ядер.

Рассмотрим более подробно обработку одного поля препарата. Содержание этого поля приведено на рис. 6. На рис. 7 и 8 представлены его фрагменты, дополнительно увеличенные в 2 и 4 раза. Изображение отдельной клетки приведено на рис. 9. Шаг считывания по обоим осям равен 0,15 мкм.

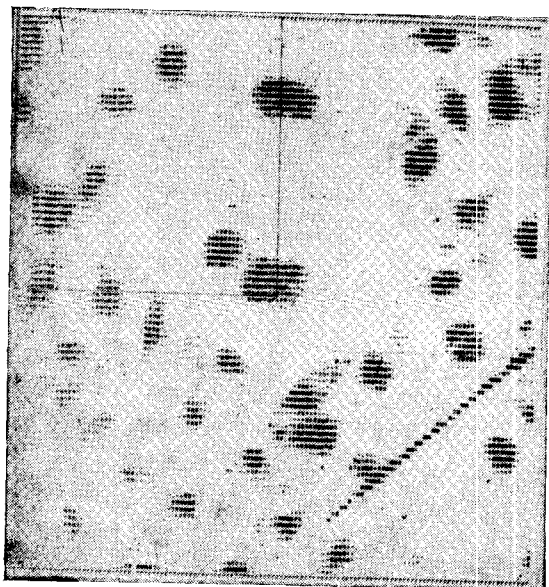


Рис. 7.

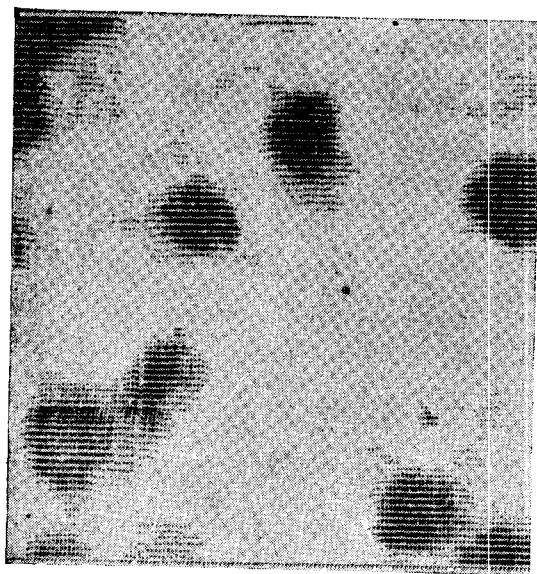


Рис. 8.

средней оптической плотности ядра. Распределение унимодально и близко к гауссовскому. Заметим, что средняя оптическая плотность ядер для второго пика в распределении (см. рис. 5) равна $D=0,273$ с $\sigma=0,005$, в то время как для первого пика среднее значение оптической плотности равно $0,252$ с $\sigma=0,003$, т. е. отличия в средних значениях оптической плотности для первого и второго пиков значимы. Точность измерения оптической плотности ядра составляет 4—5%.

Координаты центров ядер представляют интерес при анализе их взаимного расположения и ориентации и определяются как

Алгоритм выделил на всем поле 46 ядер и для каждого из них определил следующие параметры:

Площадь. Гистограмма распределения площадей ядер представлена на рис. 10. Ядра имеют площади от 63 до 92 мкм². На гистограмме даже при таком малом объеме выборки четко вырисовываются два пика: первый — со средней площадью 68 мкм², второй — со средней площадью 82 мкм². Анализ точности измерения площадей показал, что относительная погрешность измерения равна примерно 5%. На рис. 4, как отмечалось выше, приведена типичная гистограмма площадей в зависимости от выбранного уровня пропускания.

Интегральное пропускание (оптическая плотность) ядра. На рис. 11 приведена гистограмма интегральной оптической плотности внутренних точек ядра; на ней также вырисовываются два пика: первый — размытый пик со средним значением интегральной плотности $20,3 D$ мкм², второй — узкий пик со средним значением интегральной плотности $29 D$ мкм². (Заметим, что в интегральную плотность внутренних точек ядра входят в качестве аддитивных добавок оптические плотности клетки и стекла, проинтегрированные по площади ядра.)

На рис. 12 приведена гистограмма распределения

центр тяжести ядра:

$$\bar{x} = \frac{\sum_{\Omega} x_i \tau(x_i, y_i)}{\sum_{\Omega} \tau(x_i, y_i)}; \quad \bar{y} = \frac{\sum_{\Omega} y_i \tau(x_i, y_i)}{\sum_{\Omega} \tau(x_i, y_i)}, \quad (1)$$

где область суммирования Ω — внутренняя область ядра. Оценки координат имеют погрешность $\sigma = 1 \div 1,5$ мкм.

Минимальный и максимальный радиусы и их ориентация определяются сравнением длин восьми радиусов, исходящих из центра клетки и ориентированных под углами, кратными $\pi/4$. Точность определения радиусов характеризуется $\sigma = 1,5$ мкм.

Радиус инерции ядра определяется по формуле

$$R = \left\{ \frac{\sum_{\Omega} \tau(x_i, y_i) [(x_i - \bar{x})^2 + (y_i - \bar{y})^2]}{2 \sum_{\Omega} \tau(x_i, y_i)} \right\}^{1/2}. \quad (2)$$

Основываясь на величине радиуса инерции, легко построить критерии разделения ядер, поскольку радиус инерции очень чувствителен к перекрытию и соприкосновению ядер.

Периметр ядра вычисляется как длина линии, соединяющей все граничные точки объекта. Его отношение к площади ядра может служить мерой изрезанности границ ядра.

Так как алгоритм рассчитан на обработку больших выборок, требования к его быстрдействию вынудили нас ориентировать его на анализ неперекрывающихся клеток и ядер. Поэтому препарат должен содержать достаточно большое количество отдельно лежащих клеток и ядер. Идентификация клеток, входящих в многослойные клеточные скопления, представляет значительные трудности; такие скопления могут быть идентифицированы алгоритмом как один объект и поэтому исключаются из анализа. В то же время неперекрывающиеся ядра в клеточных пластах или группах плотно лежащих клеток будут проанализированы, а число таких групп с учетом «тесноты» расположения ядер и их вза-

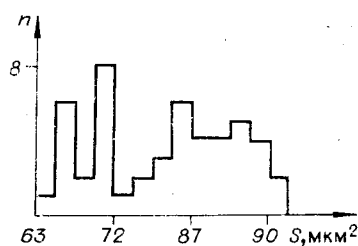
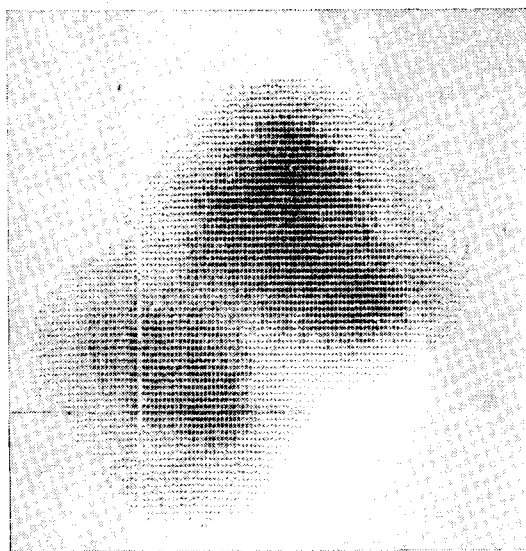


Рис. 10.

Рис. 9.

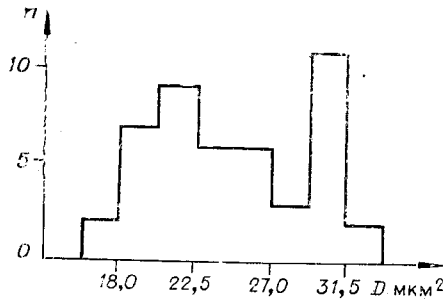


Рис. 11.

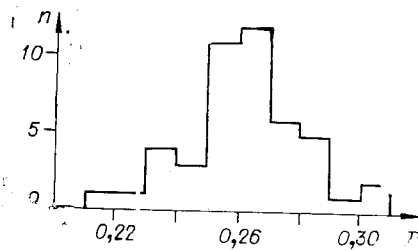


Рис. 12.

имной ориентации может использоваться как диагностический признак. **Диагностические критерии.** Полученные при анализе препарата оценки параметров — основа для формулирования диагноза. При этом используются диагностические критерии, в основу построения которых может быть положено представление об опухолевой клетке, выраженное в терминах определяемых признаков (параметров), либо модельное представление о соотношении признаков для нормальной и опухолевой популяций клеток. К сожалению, в настоящее время мы не располагаем достаточным количеством клинических данных для построения математической модели опухолевого процесса; возможно, что без применения автоматизированных методов такие данные получить вообще не удастся.

Процесс установления диагноза можно разбить на следующие этапы.

1. Подготовка препарата для исследования на автоматизированной системе.
2. Получение оценок выбранного множества параметров.
3. Обработка полученных данных программой выделения признаков.
4. Сопоставление признаков с принятыми диагностическими критериями.

На заключительном этапе, используя полученные признаки и построенное на их основе решающее правило, по которому клетки относятся к нормальным и патологическим, делается вывод о концентрации в препарате патологических клеток [5].

Рассмотрим задачу определения объема выборки (числа клеток) для надежного выявления (с заданной вероятностью) наличия или отсутствия в анализируемой популяции патологических клеток.

Пусть некоторое множество состоит из элементов двух видов: нормальных (N) и патологических (P). Необходимо по случайной выборке объема n из этого множества определить при помощи некоторого решающего правила наличие или отсутствие в исходном множестве элементов P . Распознавание элементов N и P производится решающим правилом H с ошибками двух видов:

α — вероятность принять N за P (ложная тревога),

β — вероятность принять P за N (пропуск).

Следовательно, если исходное множество объема n содержит n_1 элементов P и выбор осуществляется случайным образом, то вероятность выбора элемента P равна $p = n_1/n$, а вероятности отнесения наугад выбранного элемента правилом H к N и P следующие:

$$\gamma_N = (1-p)(1-\alpha) + p\beta; \quad \gamma_P = p(1-\beta) + (1-p)\alpha. \quad (3)$$

Для решения поставленной задачи необходимо по выборке объема n проверить гипотезу $p=0$ при альтернативе $p>0$. Поскольку выборка

случайна, то число элементов P в выборке объема W подчиняется биномиальному закону распределения

$$p(m) = C_W^m (1-p)^{W-m} p^m$$

со средним значением, равным pW , и дисперсией $p(1-p)W$. Однако из-за ошибок распознавания число обнаруженных элементов P состоит как из правильно идентифицированных элементов P , так и из элементов N , ошибочно принятых за P . Общее число K обнаруженных элементов P подчиняется биномиальному закону распределения:

$$P_I(K) = C_W^K (1-\gamma_p)^{W-K} \gamma_p^K. \quad (4)$$

Аналогично при $p=0$, т. е. при отсутствии в исходном множестве элементов P ,

$$P_{II}(K) = C_W^K (1-\alpha)^{W-K} \alpha^K. \quad (5)$$

При достаточно малых α и β и больших W (100 и более) биномиальное распределение может быть аппроксимировано распределением Пуассона [6]. В этом случае конечный расчет вероятностей существенно упрощается:

$$\begin{aligned} P_I(K) &= \frac{\lambda_1^K}{K!} e^{-\lambda_1}, \quad \text{где } \lambda_1 = \alpha W + (1-\alpha-\beta)pW = \alpha W + \Delta\lambda; \\ P_{II}(K) &= \frac{\lambda_2^K}{K!} e^{-\lambda_2}, \quad \text{где } \lambda_2 = \alpha W. \end{aligned} \quad (6)$$

Следовательно, задача выявления наличия или отсутствия элементов P в исходном множестве сводится к проверке гипотезы $H_0: \Delta\lambda=0$ при альтернативе $H_1: \Delta\lambda>0$ для распределения Пуассона. Равномерно наиболее мощный критерий [7] отвергает гипотезу H_0 , если число K обнаруженных элементов P достаточно велико, и принимает ее в противном случае.

Гипотеза принимается, если $K \leq I_p$ (I_p — квантиль распределения Пуассона, определяемая из уравнения $\sum_{i=I_p+1}^{\infty} \frac{\lambda_2^i}{i!} e^{-\lambda_2} \leq \alpha_I$), и отвергается в противном случае. Здесь α_I — ошибка первого рода, равная вероятности сделать вывод о наличии в исходном множестве элементов P при их отсутствии (вероятность ложной тревоги). Ошибка второго рода α_{II} — вероятность вывода об отсутствии элементов P при их наличии (вероятность пропуска):

$$\alpha_{II} = \sum_{i=0}^{I_p} \frac{\lambda_1^i}{i!} e^{-\lambda_1}.$$

Следовательно, задаваясь вероятностью ложной тревоги α_I и вероятностью пропуска α_{II} , получаем систему уравнений (7), связывающую объем выборки W с минимальной концентрацией, достаточной для постановки диагноза:

$$\begin{aligned} \sum_{i=I_p+1}^{\infty} \frac{\lambda_2^i}{i!} e^{-\lambda_2} &= \alpha_I, \quad \lambda_2 = \alpha W; \\ \sum_{i=0}^{I_p} \frac{\lambda_1^i}{i!} e^{-\lambda_1} &= \alpha_{II}, \quad \lambda_1 = \lambda_2 + \Delta\lambda, \quad \Delta\lambda = pW(1-\alpha-\beta). \end{aligned} \quad (7)$$

На рис. 13 представлена зависимость величины $\Delta\lambda$ от λ_2 . Исходя из этой зависимости, нетрудно определить необходимый объем выборки

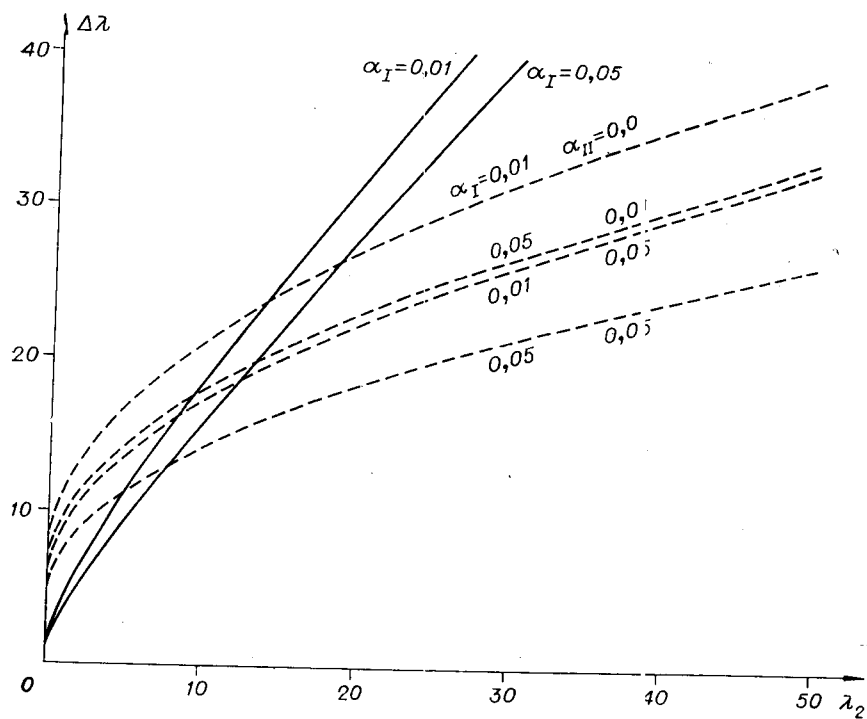


Рис. 13.

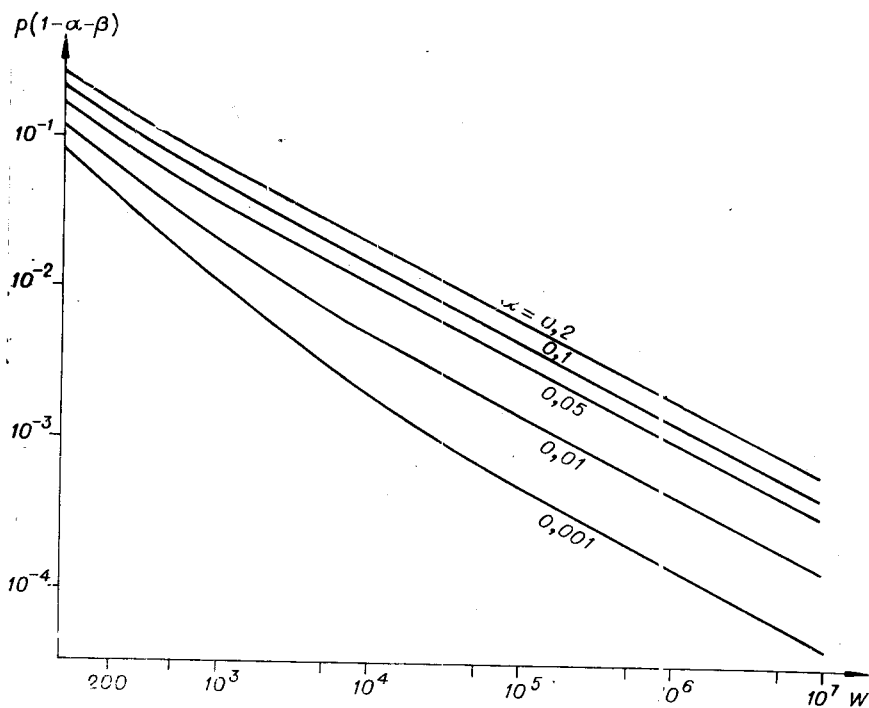


Рис. 14.

из совокупности с концентрацией P патологических элементов, достаточный для обнаружения (с заданными вероятностями ложной тревоги и пропуска) элементов P (рис. 14). На рис. 13 представлена также зависимость I_p от $\Delta\lambda$ для λ_2 , равной 0,05 и 0,01 (сплошная линия).

При $\lambda_2 \geq 50$ распределение Пуассона можно заменить нормальным распределением. При этом система (7) преобразуется к виду

$$\lambda_2 + \Delta_2 \sqrt{\lambda_2} \leq \lambda_1 - \Delta_1 \sqrt{\lambda_1},$$

где Δ_1 и $\Delta_2 - \alpha_1$ - и α_1 -квантили нормального распределения. Разрешая уравнение относительно $\Delta\lambda$, получаем

$$\Delta\lambda = 0,5 \left\{ \Delta_1^2 + 2\Delta_2 \sqrt{\lambda_2} + \Delta_1 \sqrt{\Delta_1^2 + 4\Delta_2 \sqrt{\lambda_2} + 4\lambda_2} \right\}$$

или

$$P = \left\{ \Delta_1^2 + 2\Delta_2 \sqrt{\lambda W} + \Delta_1 \sqrt{\Delta_1^2 + 4\Delta_2 \sqrt{\alpha W} + 4\alpha W} \right\} / 2W (1 - \alpha - \beta).$$

При равенстве ошибок первого и второго рода ($\Delta_1 = \Delta_2 = \Delta$)

$$P = \frac{2\Delta}{1 - \alpha - \beta} \sqrt{\frac{\alpha}{W}} + \frac{\Delta^2}{(1 - \alpha - \beta) W}. \quad (8)$$

Заключение. Обработка серии цитологических препаратов показала, что система «Зенит» позволяет измерять обширное множество разнообразных параметров клетки и ядра с точностью, достаточной для определения признаков, характеризующих состав и функциональное состояние клеток в препарате.

Большим преимуществом системы является то, что результаты измерений поступают непосредственно в ЭВМ, где производится как оценивание параметров, так и определение признаков, характеризующих исследуемый материал. Такое построение системы допускает последовательное опробование нескольких вариантов алгоритмов оценивания параметров, формирования набора признаков и выработки диагностического критерия после того, как массив результатов измерений препарата однажды занесен в память ЭВМ. Возможно создание архива первичных данных, относящихся к отдельным классам препаратов.

Высокая производительность системы «Зенит» позволяет строить критерии, основанные на анализе больших выборок; при этом возможно достоверное определение малых концентраций измененных элементов. Такая возможность есть одно из необходимых условий ранней диагностики опухолевых заболеваний.

ЛИТЕРАТУРА

1. Moore D. H. Optimization of cancer detection.—“J. Histochemistry and Cytochemistry”, 1974, vol. 22, N 7, p. 663—667.
2. Автоматический анализ цитологических препаратов. Под ред. А. Я. Хесина. Рига, «Зинатне», 1975.
3. Ornstein L., Ansley H. R. Spectral matching of classical cytochemistry to automated cytology.—“J. Histochemistry and Cytochemistry”, 1974, vol. 22, N 7, p. 453—469.
4. Киричук В. С., Косых В. П., Перетягин Г. И. Восстановление распределения интенсивности светового потока в считывающем луче сканирующей системы.—«Автометрия», 1977, № 3, с. 57—65.
5. Либенсон М. И., Хесин А. Я., Янсон Б. А. Автоматизация распознавания телевизионных изображений. М., «Энергия», 1975.
6. Феллер В. Введение в теорию вероятностей и ее приложения. М., «Мир», 1967.
7. Леманн Э. Проверка статистических гипотез. М., «Наука», 1964.

Поступила в редакцию 27 декабря 1976 г.