

СТАТИСТИЧНЕ ОБРОБЛЕННЯ ЦИТОЛОГІЧНИХ ЗОБРАЖЕНЬ

Описано запропонований алгоритм статистичного оброблення цитологічних зображень. Досліджено закон розподілу ядерно-цитоплазматичного відношення (ЯЦВ) та проведено ідентифікацію різних ступенів дисплазії на основі кількісної оцінки ЯЦВ. Розроблений алгоритм реалізований в інформаційно-аналітичній системі морфометричного аналізу гістологічних і цитологічних зображень.

The offered algorithm of the statistical processing of cytological images is described. Distributing of nucleocytoplasmic ratio (NCR) is investigated and authentication of different degrees of dysplasia is conducted on the basis of quantitative estimation of NCR. The developed algorithm is realized in information-analytical system of morphometrical analysis of histological and cytological images.

Ключові слова: цитологічне зображення, морфометрія, варіаційний ряд, статистичний розподіл.

Вступ

Біомедичні зображення – класи зображень, які отримані в різних діапазонах електромагнітного випромінювання живої природи. У даній роботі ми будемо розглядати клас зображень у діапазоні видимого світла окремих клітин людини, тобто цитологічні зображення. Крім якісної оцінки зображень при постановці діагнозу використовують кількісні методи (морфометричні). Морфометричні методи є сукупністю прийомів, що дозволяють дати кількісну оцінку параметрів клітинних і тканинних структур, на гістологічних або цитологічних препаратах (або їх фотографіях) [1]. Шляхом використання цих методів визначають такі параметри, як, наприклад, діаметр, висоту, товщину, площу перетину, кількість об'єктів на одиниці площі, їх форму та ін. При морфометрії об'єктів на гістологічних препаратах необхідно враховувати, що оцінювані параметри відносяться не до тканинних компонентів, а до їх перерізів на зрізах.

Ручна морфометрія базується на проведенні підрахунків візуально, безпосередньо під мікроскопом або на мікрофотографіях з використанням лінійок, сіток (у тому числі у вигляді окулярних вставок) та інших пристосувань. Методи напівавтоматичного і автоматичного аналізу зображення із застосуванням комп'ютерів набули широкого поширення завдяки своїй високій продуктивності. Вони дозволяють швидко кількісно оцінити велике число ознак на препараті, що вивчається, і за їх сукупністю ідентифікувати різні клітини [2].

В роботі досліджено дисплазію – гістологічне спрощення тканини регресивного характеру, яке пов'язане із зниженням процесу диференціювання. Вона характеризується заміщенням частини шару епітелію ектоцервіксу клітинами з різним ступенем атипії. Дисплазії є важливим передраковим процесом і цитологічний метод займає особливе місце при їх виявленні як для профілактичного скринінгового так і для діагностичного досліджень. Серед патологічних процесів шийки матки дисплазія зустрічається з найбільшою частотою і діагностується в 38,5% жінок, причому 67% з них знаходяться в активному репродуктивному віці до 30 років. В зв'язку з цим рання діагностика і активне лікування псевдоерозії шийки матки є важливою ланкою в профілактиці злоякісних новоутворень у жінок.

Постановка задачі

Нехай дано п'ять груп цитологічних зображень пацієнтів $A = \{x_i\}$, $B = \{y_i\}$, $C = \{z_i\}$, $D = \{u_i\}$, $E = \{w_i\}$, $i = \overline{1, n}$, де A – норма, B – дисплазія III А, C – дисплазія III Б, D – дисплазія III С, E – дисплазія IV. У кожній групі для кожної клітини обчислено морфометричний показник – ядерно-цитоплазматичне відношення $k_{яцв} = \frac{S_{я}}{S_{ц}}$, де $S_{я}$ – площа ядра, $S_{ц}$ – площа цитоплазми клітини.

На основі цих показників необхідно провести статистичний аналіз $k_{яцв}$ (побудувати варіаційний ряд, гістограму, статистичний розподіл, обчислити середнє вибіркове, середнє квадратичне відхилення, моду M_o , емпіричну функцію розподілу $F^*(x)$, медіану m_e , розмах варіації R , коефіцієнт варіації Cv), визначити належність емпіричного розподілу певному розподілу (нормальному, показниковому, біноміальному і т.д.). Використовуючи ядерно-цитоплазматичне відношення як основний інформативний морфометричний показник при постановці діагнозу необхідно дослідити залежності ядерно-цитоплазматичного відношення $k_{яцв}$ від типу дисплазії епітелію.

Алгоритм статистичної обробки цитологічних зображень

Після визначення морфометричних показників (площі ядра, площі цитоплазми, площі клітини, ядерно-цитоплазматичного відношення) отримуємо вибірки для кожного показника x_j , $j = \overline{1, 4}$, тобто

G_A^1, \dots, G_A^4 – група пацієнтів А,

G_B^1, \dots, G_B^4 – група пацієнтів В,

G_C^1, \dots, G_C^4 – група пацієнтів С,

G_D^1, \dots, G_D^4 – група пацієнтів D,

G_E^1, \dots, G_E^4 – група пацієнтів E.

При визначенні обсягу вибірки використовуємо відповідну таблицю кількості одиниць спостережень в вибірках при вивченні мірних ознак для робіт середньої точності [3].

Вибірki опрацьовано за таким алгоритмом [4]:

1. Будуємо варіаційний ряд і статистичний розподіл вибірки біомедичного дослідження. Оскільки ЯЦВ – неперервна випадкова величина, тоді статистичний розподіл вибірки задається відповідністю між інтервалами і частотами, які потрапляють у ці інтервали, тобто інтервальним статистичним розподілом вибірки:

$$\begin{array}{c|cccc} [x_i, x_{i+1}) & [x_1, x_2) & [x_2, x_3) & \dots & [x_k, x_{k+1}) \\ \hline n_i & n_1 & n_2 & \dots & n_k \end{array}$$

2. Отримуємо візуалізацію досліджуваного процесу за допомогою гістограми та полігону частот, і кумуляти при нарощуванні частот або збільшенні величини ознаки.

3. Обчислюємо наступні числові характеристики вибірки. Оскільки для знаходження числових характеристик вибірових даних необхідно мати варіанти x_i , то за них взято середини інтервалів інтервального статистичного розподілу вибірки. Обчислюємо середню арифметичну \bar{x}_g , моду за формулою

$M_o = x_m + \frac{n_m - n_{m-1}}{2n_m - n_{m-1} - n_{m+1}} h_m$, де $[x_m, x_{m+1})$ – модальний інтервал, n_m – число варіант з цього

інтервалу. Знаходимо значення емпіричної функції розподілу

$$F^*(x) = W(X < x) = \frac{n_x}{n}, \quad (1)$$

де n_x – сума частот тих варіант, які менші від x , n – обсяг вибірки. Знаходимо медіану за формулою

$$M_e = x_m + \frac{0,5 - F^*(x_m)}{F^*(x_{m+1}) - F^*(x_m)} (x_{m+1} - x_m), \quad (2)$$

де $[x_m, x_{m+1})$ – медіанний частинний інтервал ($1 \leq m \leq k$), для якого виконуються нерівності $F^*(x_m) < 0,5$, $F^*(x_{m+1}) > 0,5$. Знаходимо кватилі – значення варіант, що ділять варіаційний ряд разом з медіаною на 4 частини.

4. Обчислюємо числові характеристики вибірки для оцінювання ступеня розсіювання (малий, середній, сильний) варіант навколо середньої. Обчислюємо розмах варіації, дисперсію σ_g^2 , середнє квадратичне відхилення σ_g , коефіцієнт варіації $CV = \frac{\sigma_g}{\bar{x}_g} \cdot 100\%$.

5. Обчислюємо помилки репрезентативності вибірки за формулою $m = \pm \frac{\sigma}{\sqrt{n}}$ і робимо висновки за розміром помилки, наскільки середня величина вибіркової сукупності відрізняється від середньої генеральної сукупності.

З таблиці значень функції Лапласа вибираємо довірчий коефіцієнт t для заданого ступеня надійності p з рівності $2\Phi(t) = p$.

Знаходимо межі середньої арифметичної генеральної: $M_{gen} = \bar{x} \pm tm$.

6. Робимо перевірку приналежності вибірових даних нормальному розподілу із визначеними параметрами за критерієм узгодженості Пірсона (χ^2):

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - n_i^0)^2}{n_i^0}, \quad (3)$$

де n_i – частота варіанти x_i , $n_i^0 = np_i$ – теоретична частота варіанти x_i , n – обсяг вибірки, p_i – імовірність варіанти x_i , яка обчислюється за формулою:

$$p_i = \Phi\left(\frac{x_{i+1} - \bar{x}}{\sigma}\right) - \Phi\left(\frac{x_i - \bar{x}}{\sigma}\right), \quad (4)$$

де $\Phi(x)$ – функція Лапласа.

7. Здійснюємо висновки стосовно узагальнених показників значень варіантів морфометричних

ознак, їх варіабельності, розподілу, меж середньої арифметичної генеральної.

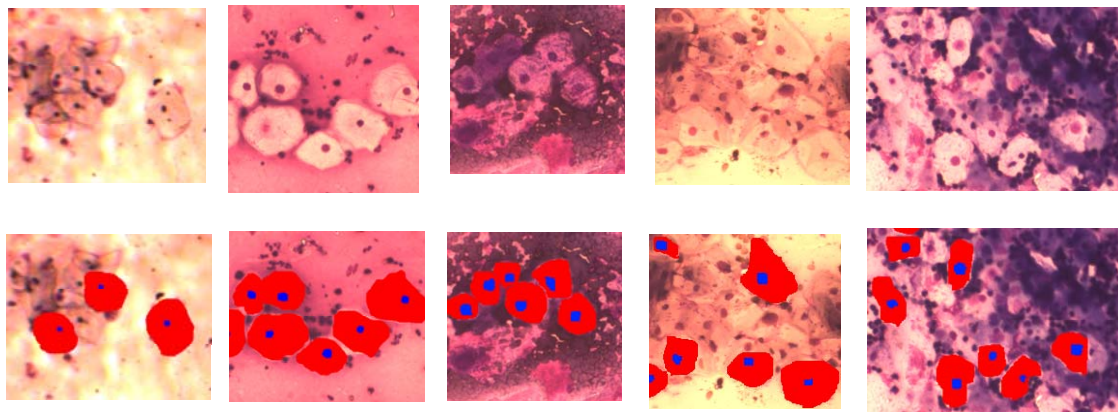
Результати експериментальних досліджень

В рамках держбюджетної теми "Інформаційно-аналітична система для дослідження та діагностування пухлинних (ракових) клітин людини на основі аналізу їх зображень" розроблено програмне забезпечення МорфоСист [5], яке було використане для дослідження морфометричних показників багаточарового плоского незроговілого епітелію при різних видах дисплазії епітелію шийки матки у жінок репродуктивного віку.

Для морфометричного аналізу застосовані цитологічні препарати, які виготовлені за методикою Романовського. Морфометрію клітин проведено за методом Автанділова Г.Г. [1]. Аналіз препаратів здійснено за допомогою системи автоматизованої мікроскопії на базі комп'ютера з процесором AMD Sempron 1500 МГц та цифрової відеокамери MAA Group IEEE-1394 (1.3 Мпкс) і програмного забезпечення МорфоСист.

Проведено цитологічне дослідження багаточарового плоского незроговілого епітелію шийки матки при різних ступенях дисплазій, зображення яких показано на рис. 1.

Досліджувані клітини візуалізувалися на екрані монітору у псевдокольорах.



Норма

Дисплазія IIIA

Дисплазія IIIB

Дисплазія IIIV

Дисплазія IV

Рис. 1. Зображення різних типів дисплазії

Результати експериментальних досліджень приведені в таблицях 1 і 2.

Таблиця 1

Результати першого експерименту

	Площа ядра	Площа клітини	Площа цитоплазми	$k_{яцв}$
Середнє вибіркоче	384,33	15764	15380	0,030041
Мінімальні значення	215	6632	6224	0,013391
Максимальні значення	611	33998	33387	0,065553
Середнє квадратичне відхилення	29,012	1924,6	1914,3	0,0042853

Таблиця 2

Результати другого експерименту

	Площа ядра	Площа клітини	Площа цитоплазми	$k_{яцв}$
Середнє вибіркоче	36,2	2191,4	2155,3	0,017135
Мінімальні значення	26	1417	1378	0,01029
Максимальні значення	56	2751	2721	0,028302
Середнє квадратичне відхилення	2,0782	77,87	77,413	0,0010594

Крім цього досліджено залежність ядерно-цитоплазматичного відношення $k_{яцв}$ від типу дисплазії епітелію (рис. 2).

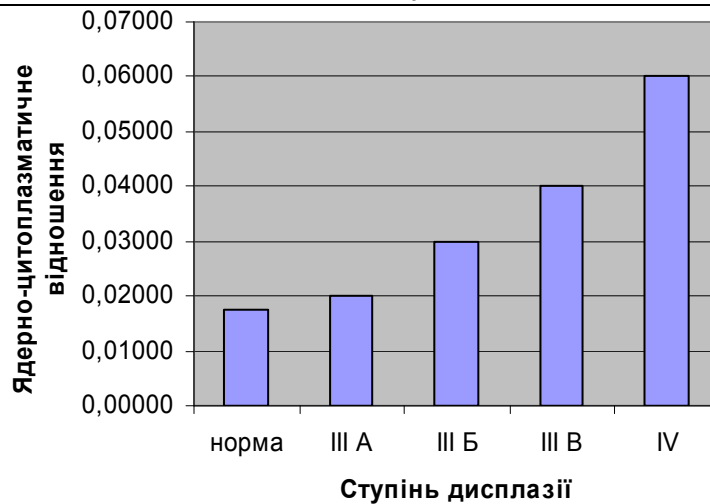


Рис. 2. Графік залежності ядерно-цитоплазматичного відношення $k_{яцв}$ від типу дисплазії епітелію

Висновок

Розроблена інформаційно-аналітична система дає можливість здійснювати морфометричні вимірювання гістологічних і цитологічних зображень. Статистичний аналіз показав, що закон розподілу ядерно-цитоплазматичного відношення є нормальний і дав змогу провести ідентифікацію різних ступенів дисплазії на основі кількісної оцінки $k_{яцв}$.

Література

1. Автандилов Г.Г. Основы количественной паталогической анатомии / Автандилов Г.Г. – М. : Медицина, 2002. – 240 с.
2. Егорова О. Компьютерная микроскопия / Егорова О., Пантелеев В., Клыкова Е. – М. : Техносфера, 2005. – 300 с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика / Гмурман В.Е. – М. : Высшая школа, 1972. – 368 с.
4. Березький О.М. Алгоритми статистичної обробки біомедичних зображень / О.М. Березький, К.М. Березька, Г.М. Мельник // Матеріали міжнар. конф. „Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту” (ISDMCI’2009), м. Євпаторія, 18–22 травня 2009 р. – Херсон : ХНТУ, 2009. – Т. 2. – С. 227–230.
5. Березький О.М. Інформаційно-аналітична система дослідження та діагностування пухлинних клітин на основі аналізу їх зображень / О.М. Березький, Ю.М. Батько, Г.М. Мельник // Вісник Хмельницького національного університету. – Хмельницький, 2008. – № 3. – Т. 1. – С. 120–130.

Надійшла 15.9.2012 р.
Рецензент: к.т.н. Коваль В.С.

УДК 621.391:004.73

К.В. КОЛЕСНИКОВ, С.И. КАУНЕНКО
Черкасский государственный технологический университет

ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВОГО КОНТЕЙНЕРА ДЛЯ ИДЕНТИФИКАЦИИ ПО ОТПЕЧАТКАМ ПАЛЬЦЕВ

Рассматривается алгоритм использования нейросетевого контейнера при биометрической идентификации по рисунку пальца с заполнением личной информации. Данный алгоритм позволяет защитить биометрический шаблон и усилить программную биометрическую защиту для минимизации возможности подмены шаблона с целью взлома.

This algorithm allows to protect the biometric template and strengthen the protection of biometric software to minimize the possibility of substitution pattern in order to break-in.

Ключевые слова: идентификация по отпечаткам пальцев, биометрия, нейросетевой контейнер.

Введение

В настоящее время активно используются биометрические средства аутентификации при доступе к информации, находящейся на ПК, flash-накопителях, при удаленной идентификации и при доступе к ключу формирования ЭЦП [1,2]. Подобные средства имеют разные биометрические датчики ввода информации,