

УДК 004:932.2:616-006.06

О.М. БЕРЕЗЬКИЙ, К.М. БЕРЕЗЬКА, Г.М. МЕЛЬНИК, Т.М. ДОЛИНЮК
Тернопільський національний економічний університет

СТАТИСТИЧНИЙ АНАЛІЗ КОЕФІЦІЄНТА СТРУКТУРНОЇ АТИПІЇ ПАТОЛОГІЧНИХ СТАНІВ МОЛОЧНОЇ ЗАЛОЗИ

Для діагностики раку використовуються гістологічні зображення. Ці зображення відображають ділянки людських тканин. Хід злоякісних процесів характеризується зміною структури тканин. Для аналізу структурних властивостей тканин та оцінки структурних залежностей між мікрооб'єктами використовується коефіцієнт структурної атипії. Запропоновано алгоритм статистичної обробки гістологічних зображень. Проведено статистичний аналіз коефіцієнта структурної атипії (S) на основі гістологічних зображень молочної залози. Досліджено закон розподілу S , ідентифіковано різні типи патології молочної залози на основі кількісної оцінки S . Алгоритм був реалізований в гібридній інтелектуальній системі аналізу біомедичних зображень.

Ключові слова: гістологічне зображення, атипія, патологія, варіаційний ряд, статистичний розподіл.

OLEH BEREZSKY, KATERYNA BEREZKA, GRIGORIY MELNYK, TARAS DOLYNUK
Ternopil National Economic University

STATISTICAL ANALYSIS OF STRUCTURAL ATYPIA FACTOR OF BREAST PATHOLOGICAL CONDITIONS

For the diagnosis of cancer histological images are used. These images reflect sections of human tissues. The course of malignant processes is characterized by a change in the structure of tissues. To analyze the structural properties of tissues and to evaluate the structural dependencies between microobjects, the structural atypia factor is used. The algorithm of statistical processing of histological images is proposed. The statistical analysis of the structural atypia factor (S) based on the histological images of the mammary gland was performed. The distribution law of S was investigated, various types of breast pathology were identified on the basis of quantitative evaluation of S . The algorithm was implemented in the hybrid biomedical images analysis intelligence system.

Keywords: histology, mammary gland, structural atypia factor, statistical analysis.

Вступ

Для діагностування ракових захворювань використовуються зображення зрізів зафіксованих тканин людини – гістологічні зображення (ГЗ). Протікання злоякісних процесів характеризується зміною структури тканин. ГЗ належать до класу структурних текстур, тобто складаються з обмеженої кількості непохідних елементів-клітин, що знаходяться у відношенні просторового порядку – мають структурні залежності. Отже, аналіз структурних властивостей ГЗ складається із операцій виділення непохідних елементів – мікрооб'єктів та оцінки структурних залежностей між ними. Для виділення мікрооб'єктів проводиться сегментація зображення.

Для тестування розроблених алгоритмів сегментації в мережі доступні набори сегментованих анованих зображень, наприклад The Berkeley Segmentation Dataset and Benchmark. Проте вони не містять гістологічних зображень. Існуючі ж бази біомедичних зображень, наприклад Breast Cancer Dataset, визначають тільки клас зображень і не містять анотацій і ручної сегментації. Велику тестову вибірку зображень та її ручну експертну сегментацію можна отримати тільки в рамках співпраці із медичними навчальними закладами та установами. Крім того, якщо зображень тканин при хворобах можна вибрати достатню для експериментів кількість, то ановані зображення здорових тканин містяться тільки у підручниках та інших навчальних матеріалах.

Аналіз ГЗ – тема, що досліджується протягом багатьох років. Оглядові роботи [1, 2] підсумовують як розроблені алгоритми і методи всіх етапів аналізу ГЗ так і структури і функції комп'ютерних систем аналізу. Із зростанням доступності обчислювальних ресурсів для аналізу ГЗ почали використовуватись штучні нейронні мережі [3]. Автори використали методи глибинного навчання ШНМ для класифікації тканин молочної залози. При застосуванні експертних систем в медичній діагностиці критичним є пояснення системою зроблених висновків

Розглянемо структуру тканин ГЗ. Молочна залоза складається з дольок, розділених прошарками сполучної тканини. Долька складається з протоки, що утворює численні трубчасті розгалуження, що закінчуються кінцевими відділами альвеолами. Схематично дані об'єкти представляються групою клітин, розташованих у вигляді колоподібних структур (рис. 1).

Ациноси (кінцеві галузенья протоків), епітелій, міоепітелій та стінки проток складаються із впорядковано розміщених клітин.

При злоякісному процесі проходить зміна структури тканин. Однією з тих змін є атипія – це порушення розташування клітин, характерного для нормального стану клітин [4]. Ракові клітини розташовуються хаотично. Для визначення ступеня атипії тканин введено коефіцієнт структурної атипії [5, 6]. Він характеризує рівень відхилення розташування клітин від нормального стану. Від того наскільки порушене розташування клітин залежить діагноз хворого. Цей показник ступеня атипії введено завдяки припущенню, що структуру тканини в нормальному стані можна описати з допомогою плоских кристаліграфічних груп [6], а структуру патологічних тканин за допомогою плоских кристаліграфічних груп зі спотвореннями.

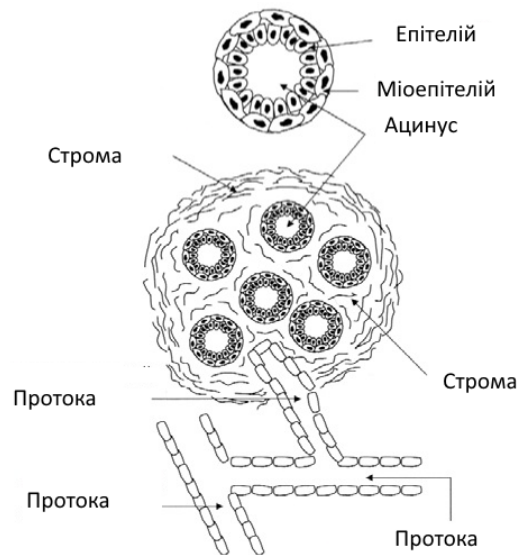


Рис. 1. Будова долек молочної залози на гістологічному зображенні

Отже, проведений аналіз показав, що існує проблема дослідження нових ознак мікрооб'єктів та їх структурних залежностей котрі відображають процеси в тканинах.

Постановка задачі

Нехай дано п'ять груп гістологічних зображень пацієнтів $A = \{x_i\}$, $B = \{y_i\}$, $C = \{z_i\}$, $D = \{u_i\}$, $E = \{w_i\}$, $i = \overline{1, n}$, де A – непроліферативна мастопатія, B – проліферативна мастопатія, C – фіброкістозна мастопатія, D – фіброаденома, E – листовидна фіброаденома. У кожній групі для кожної клітини обчислено коефіцієнт структурної атипії.

На основі цього показника необхідно провести статистичний аналіз S (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 коефіцієнт структурної атипії при постановці діагнозу необхідно дослідити залежність коефіцієнта структурної атипії від типу патології.

Коефіцієнт структурної атипії

Існують такі геометричні породжуючі перетворення паралельний перенос, центральна симетрія, осьова симетрія, ковзне відображення, поворот, їх комбінації [4]. Кожне геометричне перетворення в

афінному просторі задається в матричному вигляді так: $T = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ m & n & 1 \end{pmatrix}$, де a, b, c і d здійснюють

відповідно зсув, поворот, відображення, локальне масштабування; m і n виконують зміщення. Нехай множина геометричних породжуючих перетворень $S = \{T_1, T_2, \dots, T_n\}$ відповідає аксіомам абстрактної групи.

Величина абсолютної похибки Δ для перетворень T_1, \dots, T_{n-1}, T_n , враховуючи похибки від коефіцієнтів a, b, c, d, m, n рівна:

$$\Delta = \sum_{j=1}^3 \sum_{i=1}^3 |t_{ij}^* - t_{ij}|,$$

де t_{ij} – точне значення коефіцієнта матриці еталонного перетворення T^e ;

t_{ij}^* – значення коефіцієнта матриці реального перетворення T .

Обчислене значення похибки (спотворення) справедливе тільки для одного рапорту. В загальному випадку зображення складається із n рапортів і тоді коефіцієнт структурної атипії (загальна похибка спотворення) рівний:

$$S = \frac{\sum_{i=1}^n |\Delta_i|}{n}.$$

де n – кількість рапортів на зображенні.

Алгоритм статистичної обробки гістологічних зображень

Опишемо розроблений алгоритм статистичної обробки гістологічних зображень.

Алгоритм складається з таких кроків [7]:

1. Будуємо варіаційний ряд і статистичний розподіл вибірки гістологічного дослідження. Оскільки коефіцієнт структурної атипії – неперервна випадкова величина, тоді статистичний розподіл вибірки задаємо інтервальним розподілом:

$[x_i, x_{i+1})$	$[x_1, x_2)$	$[x_2, x_3)$...	$[x_k, x_{k+1})$
n_i	n_1	n_2	...	n_k

2. Візуалізуємо досліджуваній процес за допомогою гістограми та полігону частот, і кумуляти при нарощуванні частот або збільшенні величини ознаки.

3. Обчислюємо такі числові характеристики вибірки: 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 $F^*(x) = W(X < x) = \frac{n_x}{n}$, де n_x – сума частот тих варіант, які менші від x , n – обсяг вибірки. Знаходимо медіану за формулою

$M_e = x_m + \frac{0,5 - F^*(x_m)}{F^*(x_{m+1}) - F^*(x_m)} (x_{m+1} - x_m)$, 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: 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: $M_{gen} = \bar{x} \pm tm$.

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

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

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:

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

де $\Phi(x)$ – функція Лапласа.

Робимо висновки стосовно узагальнених показників значень варіантів морфометричних ознак, їх варіабельності, розподілу, меж середньої арифметичної генеральної.

Досліджуємо незміщеність, спроможність і ефективність коефіцієнта структурної атипії.

Наведемо визначення незміщеності, спроможності і ефективності.

Визначення 1. Незміщеною називають статистичну оцінку Q^* , математичне сподівання якої дорівнює оцінюваному параметру Q за будь-якого обсягу вибірки, тобто $M(Q^*) = Q$.

Зміщеною називають оцінку, математичне сподівання якої не дорівнює оцінюваному параметру.

Визначення 2. Ефективною називають статистичну оцінку, яка за заданого обсягу вибірки n має найменшу можливу дисперсію.

При розгляді вибірок великого обсягу (n велике) до статистичних оцінок ставиться вимога спроможності.

Визначення 3. Спроможною називають статистичну оцінку, яка при $n \rightarrow \infty$ прямує за ймовірністю до оцінюваного параметру. Наприклад, якщо дисперсія незміщеної оцінки при $n \rightarrow \infty$ прямує до нуля, то така оцінка виявляється і спроможною [8].

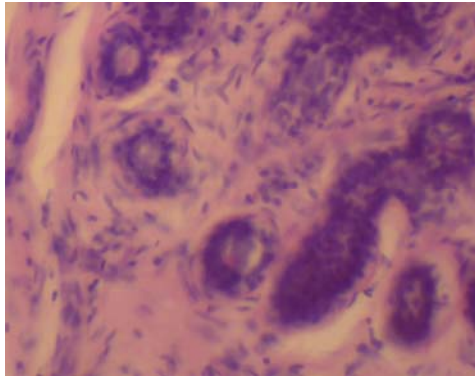
Результати експериментальних досліджень

В рамках держбюджетної теми «Гібридна інтелектуальна інформаційна технологія діагностування передракових станів молочної залози на основі аналізу зображень» нами було проведено експериментальне дослідження коефіцієнта структурної атипії, для чого використано 500 ГЗ тканин епітеліального шару по

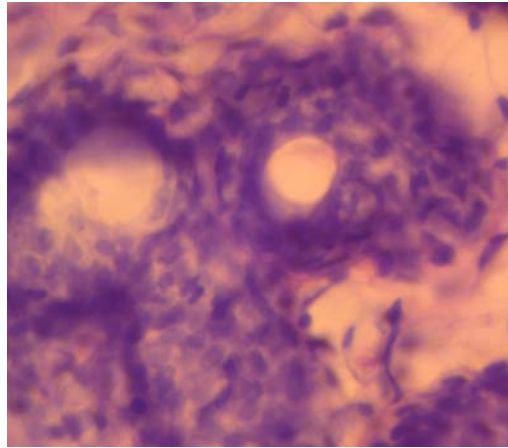
100 зразків для класів непроліферативної, проліферативної та фіброкістозної мастопатії, фіброаденоми та листовидної фіброаденоми. Кожен клас відображає певний діагноз [11].

Для морфометричного аналізу застосовані цитологічні препарати, які виготовлені за методикою Романовського. Аналіз препаратів здійснено за допомогою гібридній інтелектуальній системі аналізу біомедичних зображень на базі комп'ютера з процесором Intel Pentium 3,2 ГГц та цифрової відеокамери 6 Мпкс.

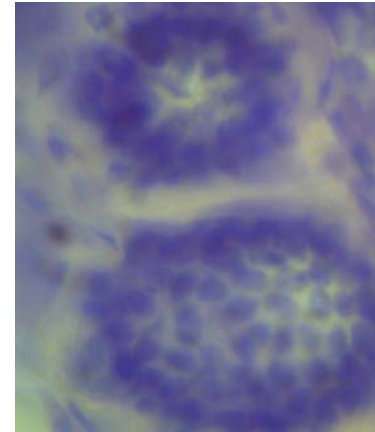
Проведено гістологічне дослідження молочної залози при різних ступенях мастопатії, зображення яких показано на рис. 2.



Непроліферативна мастопатія



Проліферативна мастопатія



Фіброаденома

Рис. 2. Зображення різних типів мастопатії

Результати експериментальних досліджень наведені в табл. 1.

Таблиця 1

Основні статистичні показники

Статистичний показник	Непроліферативна мастопатія	Проліферативна мастопатія	Фіброкістозна мастопатія	Фіброаденома	Листовидна фіброаденома
Медіана	0,8551	0,9016	0,87	0,9136	0,9324
Мода	0,9206	0,947	0,7901	0,8276	0,9229
Вибіркова середня	0,8252	0,8640	0,8779	0,9213	0,9289
Середнє квадратичне відхилення	0,3007	0,3249	0,3257	0,257	0,3416
Помилка середнього квадратичного відхилення	0,0301	0,0325	0,0326	0,0257	0,0342
Розмах варіації	1,4733	1,6049	1,4860	1,3346	1,7089
Максимальне значення	1,5852	1,6253	1,6027	1,5476	1,8708
Мінімальне значення	0,1119	0,0204	0,1167	0,2130	0,1619
Коефіцієнт варіації	0,3645	0,376	0,371	0,279	0,3678
Асиметрія	-0,0519	-0,237	0,0907	-0,1666	0,1462
Екссес	-0,169	-0,162	-0,476	0,155	0,071

Досліджено залежність коефіцієнта структурної атипії S від типу патології молочної залози (рис. 3).

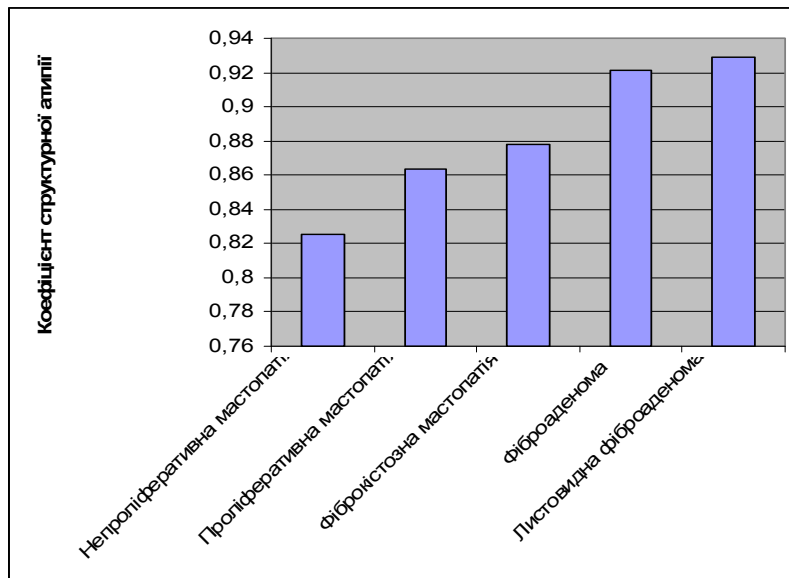


Рис. 3. Графік залежності коефіцієнта структурної атипії S від типу патології молочної залози

Незміщеність оцінки коефіцієнта структурної атипії характеризувалася відносною і абсолютною похибкою, ефективність цієї оцінки визначалася за допомогою середньоквадратичного відхилення. Спроможність оцінки забезпечувалася шляхом зміни розміру вибірки при обчисленні числових характеристик коефіцієнта структурної атипії. Значення розміру вибірки вибирались від 5 до 100 коефіцієнтів структурної атипії з кроком 5.

Відносна і абсолютна похибки обчислювалися відповідно за формулами:

$$\Delta a^* = a - a^*$$

$$\delta(a^*) = \frac{\Delta a^*}{|a^*|}$$

де a – точне значення, a^* – наближене значення деякої величини. За точне значення коефіцієнта структурної атипії було вибрано вибіркovo середню, за наближені значення – вибіркovo середні, що обчислювалися для цього коефіцієнта з кроком 5.

Графік залежності абсолютної похибки від величини вибірки для непроліферативної та фіброкістозної мастопатії наведено на рис. 4.

Висновки

Статистичний аналіз показав, що закон розподілу коефіцієнта структурної атипії є нормальний і дав змогу провести ідентифікацію різних ступенів мастопатії на основі кількісної оцінки S. Проведене дослідження дає можливість здійснювати класифікацію патологічних станів молочної залози на основі ознак зміни структури гістологічного зображення.

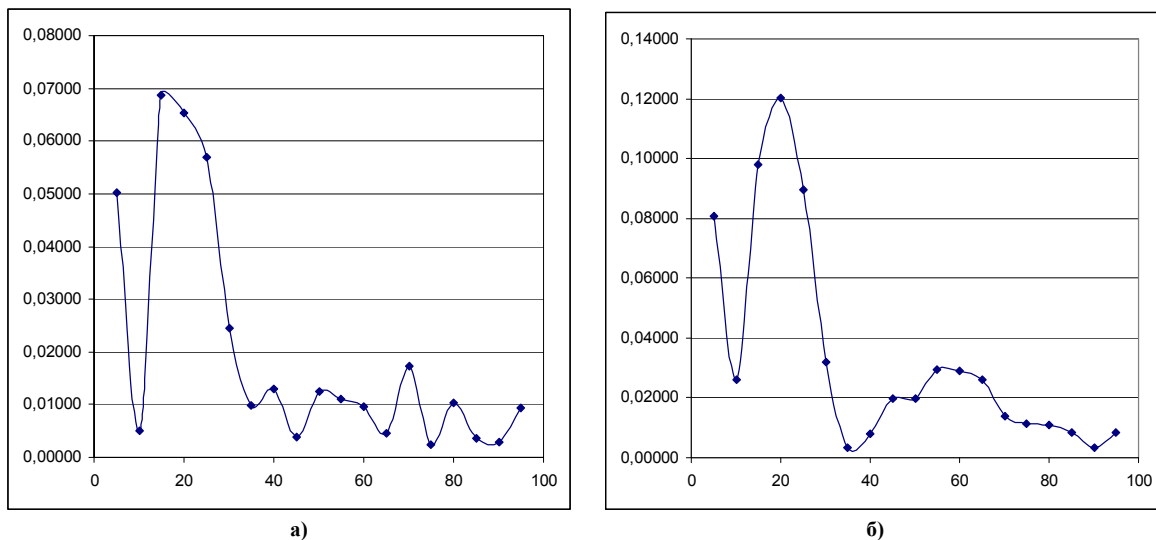


Рис. 4. Залежність абсолютної похибки від величини вибірки: а) непроліферативна, б) фіброкістозна мастопатія

Література

1. Егорова О. В. Компьютерная микроскопия / О. В. Егорова, Е. И. Клыкова, В. Г. Пантелеев. – М. : Техносфера, 2005. – 304 с.
2. M.N. Gurcan Histopathological Image Analysis: A Review / M.N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, B. Yener // IEEE reviews in biomedical engineering. – 2009. – Vol. 2. – P. 147–171.
3. Classification of breast cancer histology images using Convolutional Neural Networks / T. Araujo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy // PLoS ONE. – 2017. – Vol. 12(6).
- 3.1. 4.Ошибки в клинической онкологии / под ред. В.И. Чиссова, А.Х. Трахтенберга. – М. : ГЭОТАР-Медиа, 2009. – 768 с.
4. Березький О. М. Статистична оцінка коефіцієнта структурної атипії на основі аналізу гістологічних зображень / О. М. Березький // Збірник наукових праць інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України. – 2013. – Вип. 67. – С. 113–121.
5. Мельник Г.М. Інформаційна технологія аналізу структурних текстур для опрацювання зображень ауто- та ксеногенних тканин / Г.М. Мельник // Вісник Хмельницького національного університету. – 2014. – № 6(219). – С. 73–79.
6. Березький О. М. Методи і алгоритми аналізу та синтезу асиметричних зображень / О. М. Березький // Искусственный интеллект. – 2010. – № 4. – С. 162–172.
7. Гмурман В.Е. Теория вероятностей и математическая статистика / Гмурман В.Е. – М. : Высшая школа, 1972. – 368 с.
8. Березький О. М. Статистичне оброблення цитологічних зображень / О. М. Березький, К.М. Березька, С. Ю. Попіна // Вісник Хмельницького національного університету. – 2012. – № 5. – С. 161–164.
9. Березький О. М. Нечітка база знань інтелектуальної системи діагностування видів раку молочної залози / О. М. Березький, Г.М. Мельник, К. М. Березька // Вісник Хмельницького національного університету. Технічні науки. – 2013. – № 6. – С. 284–292.
10. Розроблення реляційної бази даних інтелектуальної системи автоматизованої мікроскопії / О.М. Березький, О.Й. Піцун, С.О. Вербовий, Т.В. Дацко // Науковий вісник національного лісотехнічного університету України : збірник науково-технічних праць. – Львів : РВВ НЛТУ України. – 2017. – Т. 27, № 5. – С. 125–129

References

1. Egorova O. V. Computer microscopy / O. V. Egorova, E. I. Klykova, V. G. Panteleev - Moscow: Technosphere, 2005. - 304 p.
2. Histopathological Image Analysis: A Review / M.N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, B. Yener. // IEEE reviews in biomedical engineering. – 2009. – Vol. 2. – P. 147-171.
3. Classification of breast cancer histology images using Convolutional Neural Networks / T. Araujo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, // PLoS ONE. – 2017. – Vol. 12(6)
4. Errors in clinical oncology // ed. V.I. Chissova, A.H. Trachtenberg. - M.: GEOTAR-MEDIA, 2009. - 768 p.
5. Berezsky O. M. Statistical estimation of the coefficient of structural atypia on the basis of analysis of histological images / O. M. Berezky // Collection of scientific works of the Institute of Modeling Problems in Power Engineering them. G. E. Puhov NAS of Ukraine. - 2013. - Issue 67. - P. 113-121.
6. Melnyk G.M. Information technology for the analysis of structural textures for the processing of images of autogenous and xenogenous tissues / G.M. Melnyk // Bulletin of the Khmelnytsky National University. - 2014 - No. 6 (219). - P. 73-79.
7. Berezsky O. M. Methods and algorithms for analysis and synthesis of asymmetric images / O. M. Berezsky // Artificial Intelligence. - 2010. - No. 4. - P. 162-172.
8. Gmurman V.E. Probability theory and mathematical statistics. - M. : Higher school, 1972. - 368 p.
9. Berezsky O. M. Statistical processing of cytological images / O. M. Berezsky, K. M. Berezka, S. Yu. Popina // Bulletin of the Khmelnytsky National University. - 2012. - No. 5. - P. 161-164.
10. Berezsky O. M. Fuzzy knowledge base of the intelligent system for diagnosing breast cancer types / O. M. Berezsky, G. M. Melnyk, K. M. Berezka // Bulletin of the Khmelnytsky National University. Technical sciences. - 2013. - № 6. - P. 284-292.
11. Relational database of intelligent automated microscopy system / O.M. Berezky, O.Yo. Pitsun, S.O. Verbovyi, T.V. Datsko // Scientific Bulletin of Ukrainian National Forestry Universitys. - Lviv: UNFU. - 2017. - Vol. 27, No. 5. - P.125-129.

Рецензія/Peer review : 06.10.2017 р.

Надрукована/Printed :01.12.2017 р.

Рецензент: д.т.н., доц. Яцків В.В.