

Лекція 14. Комп'ютерна діагностика рака молочної залози

В даний час добре відомо, що організм має потенційну спроможність адекватно відповідати на зміни зовнішнього і внутрішнього середовища в цілому або деякими своїми внутрішніми системами (нервовою, ендокринною, імунною системою тощо). Ця спроможність називається реактивністю організму [1]. Існує множина реакцій, оскільки вони виявляються в різноманітних системах, органах і клітинах організму, більш того, ми можемо говорити про цитологічну реактивність організму, припускаючи під цим зміни, що виникають у клітинах у відповідь на різноманітні чинники. Проблема цитологічної реактивності під час утворення злоякісної або доброякісної пухлини в організмі є досить складною і мало вивченою. У 1962 р. Н. Nieburgs et al. [2], а потім й інші дослідники, повідомили про характерні зміни в клітинах буккального епітелію пацієнта з пухлиною, локалізованої поза порожниною рота, і назвали це змінами, асоційованими з малігнізацією (MAC - malignancy associated changes) [3,4]. Ці зміни характеризуються зростанням розмірів епітеліоцитів, розривами ядерної мембрани, зростанням розмірів зон, пов'язаних із хроматином, оточеним світлими областями. Ці автори зареєстрували схожі зміни при наявності передпухлинних процесів в організмі. У великій кількості робіт, що з'явилися за минулий час (детальна бібліографія приведена в роботі [1]), дослідники вивчали якісні показники, що характеризують вплив пухлини на різноманітні структурні елементи організму, віддалені від пухлини, але лише деякі автори повідомили про використання кількісних методів оцінок стана клітин, присвячених визначенню оптичної щільності і розмірів ядер.

Матеріал і метод. Для дослідження були взяті групи жінок із діагнозами "рак молочної залози" (2-й і 3-й стадії) і "фіброаденоматоз" у віці від 25 до 53 років (усього 104 пацієнта). Після полоскання і зняття поверхневого прошарку клітин слизової оболонки порожнини рота були отримані клітини з різноманітної глибини шиповатого прошарку. Клітини були висушені при кімнатній температурі і зафіксовані в суміші Нікіфорова, а потім проведена реакція Фельгена з холодним гідролізом у 5 н. HCl протягом 15 хв. при температурі $t=21-22$ °C. Оптична щільність ядра реєструвалася цитоспектрофотометром МЦФУ-2МТ (ЛОМО) за допомогою методу сканування з довжиною хвилі 575 нм і діаметром зонда 0.05 мкм. У кожному препараті досліджувалося від 10 до 30 клітин. Утримання ДНК-фуксина в ядрах епітеліоцитів визначався як добуток оптичної щільності і площини. Таким чином, при дослідженні інтерфазних ядер були отримані сканограми розподілу ДНК, які є

матрицею $R = \|r_{ij}\|_{i=1, m}^{j=1, n}$, де r_{ij} характеризує вміст ДНК в осередку сітки з номером (i, j) , m - число рядків і n - число стовпчиків у матриці R . Число $m \times n$ називається розміром реєстраційного поля сканограми. При формуванні навчальних вибірок для забезпечення їхньої статистичної однорідності цей розмір варіювався в досить вузьких межах від 56 до 81.

Морфо- і денситометричні показники інтерфазного ядра. Сканограма, або цифровий портрет клітини, являє собою матрицю $R = \|r_{ij}\|_{i=1, m}^{j=1, n}$, де r_{ij} - значення поточної оптичної щільності хроматину інтерфазного ядра клітини, виражене в умовних одиницях; n, m - число точок сканограм по вертикалі і горизонталі, відповідно.

На підставі цієї цитоспектрофотометричної інформації обчислюються такі морфо- і денситометричні показники, що характеризують структурні і текстурні особливості хроматину [5-8]:

1. Площа ядра.

$$x_1 - \text{число елементів матриці } R, \text{ таких що } r_{ij} \geq 0.08.$$

2. Площа конденсованого хроматину.

$$x_2 - \text{число елементів матриці } R, \text{ таких що } r_{ij} \geq 0.35.$$

3. Площа деконденсованого хроматину.

$$x_3 - \text{число елементів матриці } R, \text{ таких що } 0.08 \leq r_{ij} < 0.35.$$

4. Площа сильно деконденсованого хроматину.

$$x_4 - \text{число елементів матриці } R, \text{ таких що } 0.08 \leq r_{ij} < 0.15$$

5. Питома площа конденсованого хроматину.

$$x_5 = \frac{x_2}{x_1}$$

6. Питома площа деконденсованого хроматину.

$$x_6 = \frac{x_3}{x_1}.$$

7. Інтегральна щільність.

$$x_7 = \sum_{i=1}^m \sum_{\substack{j=1 \\ r_{ij} \geq 0.08}}^n r_{ij},$$

де нерівність $r_{ij} \geq 0.08$ означає, що підсумовування провадиться лише по тим індексам i і j , для яких $r_{ij} \geq 0.08$.

8. Середня щільність.

$$x_8 = \frac{x_1}{nm - p},$$

де p - число точок (i, j) , у яких $r_{ij} < 0.08$.

9. Усереднена сума перепадів.

$$x_9 = \frac{1}{q} \sum_{k=1}^q v_k,$$

де q - число точок (i, j) , таких що $\min(r_{ij}, r_{i+1,j}, r_{i,j+1}, r_{i+1,j+1}) \geq 0.08$;

$$v_k = \max(r_{ij}, r_{i+1,j}, r_{i,j+1}, r_{i+1,j+1}) - \min(r_{ij}, r_{i+1,j}, r_{i,j+1}, r_{i+1,j+1}), \quad k = \overline{1, q}.$$

(Підсумовування провадиться по вищевказаних елементах).

10. Спільний показник кластерності.

$$x_{10} = \frac{1}{q} \sum_{k=1}^q v_k^2.$$

11. Дисперсійний коефіцієнт.

$$x_{11} = \left(\frac{\sum_{k=1}^q (v_k - x_9)^2}{q - 1} \right)^{1/2}.$$

12. Показник варіації перепадів.

$$x_{12} = x_9 + x_{11}.$$

13. Показник рельєфності.

$$x_{13} = \frac{\sum_{i=2}^m \sum_{j=1}^n |r_{ij} - r_{i-1,j}|}{(2mn - m + n - q)},$$

де q - число точок (i, j) , таких що $\max(r_{ij}, r_{i-1,j}) < 0.08$.

14. Текстурний коефіцієнт.

$$x_{14} = \frac{x_{13}}{\varepsilon}, \quad \varepsilon = \frac{\sum_{i=1}^m \sum_{j=1}^n (r_{ij} - x_7)}{mn - p}.$$

15. Коефіцієнт взаєморозташування.

$$x_{15} = \frac{a}{x_8^2 b},$$

$$\text{де } a = \sum_{i=1}^m \sum_{j=1}^n \left(\sum_{k=l=j+1}^m \sum_{l=j+1}^n \frac{r_{ij}r_{kl}}{(k-i)^2 + (l-j)^2} + \sum_{k=i+1}^m \sum_{l=1}^n \frac{r_{ij}r_{kl}}{(k-i)^2 + (l-j)^2} \right),$$

$$b = \sum_{i=1}^m \sum_{j=1}^n \left(\sum_{k=l=j+1}^m \sum_{l=j+1}^n \frac{1}{(k-i)^2 + (l-j)^2} + \sum_{k=i+1}^m \sum_{l=1}^n \frac{1}{(k-i)^2 + (l-j)^2} \right),$$

підсумовування провадиться лише по тим елементам, для котрих

$$\min(r_{ij}, r_{kl}) > 0.875 \max_{i=1,2,\dots,n; j=1,2,\dots,m; r_{ij} \geq 0.08} r_{ij}.$$

Алгоритм розпізнавання образів. Розглянемо алгоритми автоматизованої діагностики раку молочної залози і міру близькості, на якій вони засновані.

Роздивимося такі критерії для перевірки гіпотези H про рівність функцій розподілу $F_G(u)$ і $F_{G'}(u)$ генеральних сукупностей G і G' , відповідно [9]. Нехай $x = (x_1, \dots, x_n) \in G$ і $x' = (x'_1, \dots, x'_m) \in G'$, $x_{(1)} \leq \dots \leq x_{(n)}$, $x'_{(1)} \leq \dots \leq x'_{(m)}$ - порядкові статистики. Припустимо, що $F_G(u) = F_{G'}(u)$. Тоді в силу формули [10]

$$p(A_{ij}) = p(x'_k \in (x_{(i)}, x_{(j)})) = p_{ij} = \frac{j-i}{n+1}. \quad (1)$$

Якщо ми маємо вибірку $x' = (x'_{(1)}, \dots, x'_{(m)})$, ми можемо знайти частоту h_{ij} випадкової події A_{ij} і довірчі межі $p_{ij}^{(1)}, p_{ij}^{(2)}$ для можливості p_{ij} , що відповідають заданому рівню значимості β , такому що $B = \{p_{ij} \in (p_{ij}^{(1)}, p_{ij}^{(2)})\}$, $p(B) = 1 - \beta$. Ці межі були обчислені по формулах [11]:

$$p_{ij}^{(1)} = \frac{h_{ij}m + g^2/2 - g\sqrt{h_{ij}(1-h_{ij})m + g^2/4}}{m + g^2},$$

$$p_{ij}^{(2)} = \frac{h_{ij}m + g^2/2 + g\sqrt{h_{ij}(1-h_{ij})m + g^2/4}}{m + g^2}, \quad (2)$$

де g задовольняє умові $\Phi(g) = 1 - \beta/2$, $\Phi(u)$ - функція нормованого нормального розподілу (якщо m мало, то відповідно до правила "3 σ " $g=3$).

Варто зауважити, що довірчі межі (2) не є точними довірчими межами, які відповідають заданому рівню значущості навіть для моделі Бернуллі, але вони є асимптотичними довірчими межами.

Позначимо через N усі довірчі інтервали $I_{ij} = (p_{ij}^{(1)}, p_{ij}^{(2)})$, ($N = n(n-1)/2$) і L - число тих інтервалів I_{ij} , що містять ймовірності p_{ij} . Покладемо $h = \rho(F^*, F^{*'}) = \rho(x, x') = \frac{L}{N}$. Оскільки h - частота випадкової події $B = \{p_{ij} \in I_{ij}\}$, що має ймовірність $p(B) = 1 - \beta$, то, враховуючи, що $h_{ij} = h$, $m = N$ і $g = 3$ у формулі (2), ми одержимо довірчий інтервал $I = (p^{(1)}, p^{(2)})$ для ймовірності $p(B)$, що має довірчий рівень, рівний 0.95. Критерій для перевірки гіпотези H із рівнем значущості, приблизно рівним 0.05, може бути сформульований у такий спосіб: якщо довірчий $I = (p^{(1)}, p^{(2)})$ інтервал містить ймовірність $p(B) = 1 - \beta$, то гіпотеза H приймається, у протилежному випадку вона відхиляється. Статистика h називається p -статистикою (статистика Петуніна). Вона є мірою близькості $\rho(x, x')$ між вибірками x і x' .

Розглянемо тепер алгоритм диференціальної діагностики раку молочної залози і фіброаденоматоза [12]:

1) На першій стадії ми формуємо дві групи пацієнтів G_1 і G_2 ; перша група містить пацієнтів, що хворіють на рак молочної залози (РМЖ), а друга складається з пацієнтів, що хворіють на фіброаденоматоз (ФАМ) (діагнози пацієнтів у кожній групі повинні бути установлені точно!).

Це - так звані навчальні вибірки. На основі цих груп ми будемо діагностувати захворювання.

2). На другій стадії застосовується квадратичний тест на основі p -статистик (статистик Петуніна). Обчислюються відстані (міри близькості) між показниками сканограм пацієнта і показниками пацієнтів, що відповідають групам G_1 і G_2 . Це робиться таким чином. припустимо, що пацієнт Q належить першій групі G_1 : $Q \in G_1, G_1 = (Q_1, \dots, Q_n)$, $Q = Q_i$ ($i = 1, 2, \dots, n$). Ми виключаємо пацієнта $Q = Q_i$ з групи G_1 і одержуємо групу $G_1^{(i)} = \{Q_1, \dots, Q_{i-1}, Q_{i+1}, \dots, Q_n\}$.

Нехай

$$X_{C_1}^{(k)} = (x_{1k}^{(1)}, x_{2k}^{(1)}, \dots, x_{15k}^{(1)})$$

$$X_{C_2}^{(k)} = (x_{1k}^{(2)}, x_{2k}^{(2)}, \dots, x_{15k}^{(2)})$$

$$X_{C_{j_k}}^{(k)} = \left(x_{1k}^{(j_k)}, x_{2k}^{(j_k)}, \dots, x_{15k}^{(j_k)} \right)$$

$(k = 1, 2, \dots, n; 10 \leq j_k \leq 30)$ - вектор показників клітин пацієнта Q_k . Тут

$x_{C_i}^{(k)} = \left(x_{1k}^{(i)}, x_{2k}^{(i)}, \dots, x_{15k}^{(i)} \right)$ - вектор показників клітини C_i пацієнта $Q_k, k = 1, 2, \dots, n$.

Потім ми формуємо навчальні вибірки для кожного показника $x_i, i = 1, 2, \dots, 15$; перша навчальна вибірка для показника x_1 :

$$X_1^{(1)} = \left(x_{11}^{(1)}, x_{11}^{(2)}, \dots, x_{11}^{(j_1)} \right) \text{ (від першого пацієнта)}$$

$$X_2^{(1)} = \left(x_{12}^{(1)}, x_{12}^{(2)}, \dots, x_{12}^{(j_2)} \right) \text{ (від другого пацієнта)}$$

$$X_n^{(1)} = \left(x_{1n}^{(1)}, x_{1n}^{(2)}, \dots, x_{1n}^{(j_n)} \right) \text{ (від } n\text{-го пацієнта);}$$

друга навчальна вибірка (для показника x_2) :

$$X_1^{(2)} = \left(x_{21}^{(1)}, x_{21}^{(2)}, \dots, x_{21}^{(j_1)} \right) \text{ (від першого пацієнта)}$$

$$X_2^{(2)} = \left(x_{22}^{(1)}, x_{22}^{(2)}, \dots, x_{22}^{(j_2)} \right) \text{ (від другого пацієнта)}$$

$$X_n^{(2)} = \left(x_{2n}^{(1)}, x_{2n}^{(2)}, \dots, x_{2n}^{(j_n)} \right) \text{ (від } n\text{-го пацієнта);}$$

і так далі. В результаті ми одержимо останню навчальну вибірку, тобто вибірку для 15-го показника $X_1^{(15)}, X_2^{(15)}, \dots, X_n^{(15)}$ (n - число пацієнтів у групі G_1). Тепер ми можемо обчислити значення p -статистик для вибірок $X_i^{(15)}, X_i^{(15)}, \dots, X_i^{(15)}$ i -го пацієнта і відповідних вибірок для інших пацієнтів із номером $k (\neq i)$ (і фіксоване!):

$$\rho_{ik}^{(1)} = \rho\left(X_i^{(1)}, X_k^{(1)}\right), \rho_{ik}^{(2)} = \rho\left(X_i^{(2)}, X_k^{(2)}\right), \dots, \rho_{ik}^{(15)} = \rho\left(X_i^{(15)}, X_k^{(15)}\right)$$

і знайти значення усереднених p -статистик

$$\rho_i^{(1)} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n \rho\left(X_i^{(1)}, X_k^{(1)}\right),$$

$$\rho_i^{(2)} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n \rho\left(X_i^{(2)}, X_k^{(2)}\right),$$

.....

$$\rho_i^{(15)} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n \rho(X_i^{(15)}, X_k^{(15)})$$

(i фіксовано!), що подають міру близькості між Q_i (точніше, його показниками) і групою $G_1^{(i)}$ ($i = 1, 2, \dots, n$).

Замінюючи пацієнта Q_i пацієнтом \bar{Q}_i із групи G_2 (пригадаємо, що G_2 складається з пацієнтів із фіброаденоматозом), ми одержимо аналогічні усереднені p -статистики для групи G_2 : $\bar{\rho}_i^{(1)}, \bar{\rho}_i^{(2)}, \dots, \bar{\rho}_i^{(15)}$ ($i = 1, 2, \dots, m$; $m = \text{card } G_2$)

$$\bar{\rho}_i^{(t)} = \frac{1}{n} \sum_{k=1}^n \rho(\bar{X}_i^{(t)}, X_k^{(t)}), \quad (i = 1, 2, \dots, m; t = 1, 2, \dots, 15)$$

де $\bar{X}_i^{(t)}$ - показник пацієнта, що \bar{Q}_i відповідає. Якщо ми замінимо групу G_1 на G_2 і зробимо аналогічні обчислення, ми одержимо усереднені p -статистики $d_i^{(1)}, d_i^{(2)}, \dots, d_i^{(15)}$ ($i = 1, 2, \dots, n$) і $\bar{d}_j^{(1)}, \bar{d}_j^{(2)}, \dots, \bar{d}_j^{(15)}$ ($j = 1, 2, \dots, m$):

$$d_i^{(t)} = \frac{1}{m-1} \sum_{k=1, k \neq i}^m \rho(\bar{X}_i^{(t)}, \bar{X}_k^{(t)}), \quad (i = 1, 2, \dots, m; t = 1, 2, \dots, 15),$$

$$\bar{d}_j^{(t)} = \frac{1}{m} \sum_{k=1}^m \rho(X_j^{(t)}, \bar{X}_k^{(t)}), \quad (j = 1, 2, \dots, m; t = 1, 2, \dots, 15).$$

3). На третій стадії квадратичного тест ми робимо об'єднання цих усереднених p -статистик $(\rho_i^{(t)}, \rho_i^{(s)}), (\bar{\rho}_j^{(t)}, \bar{\rho}_j^{(s)})$ ($i=1, 2, \dots, n$; $j=1, 2, \dots, m$; $t, s=1, 2, \dots, 15$) і аналогічно одержуємо точки $(d_i^{(t)}, d_i^{(s)}), (\bar{d}_j^{(t)}, \bar{d}_j^{(s)})$ ($i=1, 2, \dots, n$; $j=1, 2, \dots, m$; $t, s=1, 2, \dots, 15$). Потім ми будуємо так звані довірчі еліпси E_{ts} , що містять усереднені p -статистики $(\rho_i^{(t)}, \rho_i^{(s)})$ ($i=1, 2, \dots, n$) для групи G_1 , тобто еліпси мінімальної площі, що містять точки $(\rho_i^{(t)}, \rho_i^{(s)})$ ($i=1, 2, \dots, n$; $t, s=1, 2, \dots, 15$). Точніше, ми будуємо еліпси E_{ts} за допомогою алгоритму, що дає наближений розв'язок цієї проблеми [13].

Потім ми будуємо довірчі еліпси \bar{E}_{ts} для усереднених p -статистик $(\bar{\rho}_j^{(t)}, \bar{\rho}_j^{(s)})$ ($j=1, 2, \dots, m$; $t, s=1, 2, \dots, 15$) і аналогічні еліпси E_{ts}^* і \bar{E}_{ts}^* , використовуючи точки $(d_i^{(t)}, d_i^{(s)}), (\bar{d}_j^{(t)}, \bar{d}_j^{(s)})$ ($i=1, 2, \dots, n$; $j=1, 2, \dots, m$; $t, s=1, 2, \dots, 15$), відповідно.

Крім того, для опису так званого лінійного тесту ми розглядаємо дискримінантні функції Фішера $f_{ts}(u, v)$ ($f_{ts}^*(u, v)$), що відокремлюють множину точок $M_{ts}^p = \left\{ \left(\rho_i^{(t)}, \rho_i^{(s)} \right), i = 1, 2, \dots, n \right\}$ від множини точок $\bar{M}_{ts}^p = \left\{ \left(\bar{\rho}_j^{(t)}, \bar{\rho}_j^{(s)} \right), j = 1, 2, \dots, m \right\}$ і множину точок $M_{ts}^d = \left\{ \left(d_i^{(t)}, \rho_i^{(s)} \right), i = 1, 2, \dots, n \right\}$ від множини точок $\bar{M}_{ts}^d = \left\{ \left(\bar{d}_j^{(t)}, \bar{d}_j^{(s)} \right), j = 1, 2, \dots, m \right\}$. Функція $f_{ts}(u, v)$ побудована таким чином, що пряма лінія $l_{ts}^p = \{(u, v): f_{ts}(u, v) = 0\}$ є перпендикулярною до відрізка, що з'єднує центри множин M_{ts}^p і \bar{M}_{ts}^p і проходить через середину цього відрізка; аналогічно побудована $f_{ts}^*(u, v)$. Крім того, центр множини M_{ts}^p належить нижній напівплощині π_{ts} , а центр множини \bar{M}_{ts}^p належить верхній напівплощині λ_{ts} (аналогічно $\pi_{ts}^*, \lambda_{ts}^*$). Отже, ми маємо 15 показників і 210 пар еліпсів (E_{ts}, \bar{E}_{ts}) і $(E_{ts}^*, \bar{E}_{ts}^*)$ ($t < s; t, s = 1, 2, \dots, 15$) і, крім того, 210 пар напівплощин $(\pi_{ts}, \lambda_{ts}), (\pi_{ts}^*, \lambda_{ts}^*)$ ($t < s; t, s = 1, 2, \dots, 15$).

Нехай Q - пацієнт, що страждає на рак молочної залози (гіпотеза H_1) або фіброаденоматозом (гіпотеза H_2). Використовуючи алгоритми, згадані вище, ми можемо обчислити для цього пацієнта усереднені p -статистики $\rho_Q^{(t)}, d_Q^{(t)}$ ($t=1, 2, \dots, 15$):

$$\rho_Q^{(t)} = \frac{1}{n} \sum_{k=1}^n \rho(X_Q^{(t)}, X_k^{(t)}), \quad d_Q^{(t)} = \frac{1}{m} \sum_{k=1}^m \rho(X_Q^{(t)}, \bar{X}_k^{(t)}),$$

де $X_Q^{(t)}$ - відповідний показник пацієнта Q , і сформувати $(\rho_Q^{(t)}, \rho_Q^{(s)})$ точки $(d_Q^{(t)}, d_Q^{(s)})$, ($t < s, t, s = 1, 2, \dots, 15$). Розглянемо такі випадкові події

$$\begin{aligned} A_1 &= \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in E_{ts} \right\}, \quad A_2 = \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in \bar{E}_{ts} \right\}, \quad A_3 = \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in E_{ts} - \bar{E}_{ts} \right\}, \\ A_4 &= \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in \bar{E}_{ts} - E_{ts} \right\}, \quad A_1^* = \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in E_{ts}^* \right\}, \quad A_2^* = \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in \bar{E}_{ts}^* \right\}, \\ A_3^* &= \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in E_{ts}^* - \bar{E}_{ts}^* \right\}, \quad A_4^* = \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in \bar{E}_{ts}^* - E_{ts}^* \right\}, \quad B_1 = \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in \pi_{ts} \right\}, \\ B_2 &= \left\{ \left(\rho_Q^{(t)}, \rho_Q^{(s)} \right) \in \lambda_{ts} \right\}, \quad B_1^* = \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in \pi_{ts}^* \right\}, \quad B_2^* = \left\{ \left(d_Q^{(t)}, d_Q^{(s)} \right) \in \lambda_{ts}^* \right\}, \quad t < s, \\ C_1 &= A_3 \text{ Y } A_4^*, \quad C_2 = A_4 \text{ Y } A_3^*, \quad C_3 = A_1 \text{ Y } A_2^*, \end{aligned}$$

$C_4 = A_2 Y A_1^*$, $C_5 = B_1 Y B_2^*$, $C_6 = B_2 Y B_1^*$ і позначимо через $h_i = h(C_i)$, $i=1,2,\dots,6$, частоту події C_i при 210 тестах (експериментах), коли $t,s=1,2,\dots,15$; $t < s$. Використовуючи формулу (2), ми можемо одержати асимптотичні довірчі межі, що відповідають заданому рівню значимості для можливості $p_i = p(C_i)$ на основі частоти $h_i = h(C_i)$ (у зв'язку з цим, ми повинні взяти $h_{ij} = h_j$, $m = 210$). Ці межі, очевидно, будуть занадто широкими. Ми назвемо частоту h_1 показником РМЖ (скорочено - РМЖ), а h_3 - тотальним показником РМЖ (скорочено - ТРМЖ), оскільки ці показники є мірою близькості між сканограмами інтерфазних ядер клітин досліджуваного пацієнта Q і відповідними сканограмами пацієнтів, що страждають на рак молочної залози. На підставі подібних аргументів ми позначимо частоти h_2 і h_4 як показник ФАМ (ФАМ) і тотальний показник ФАМ (ТФАМ), відповідно. Аналогічно ми позначимо частоти h_5 і h_6 як лінійний показник РМЖ (ЛРМЖ) і лінійний показник ФАМ (ЛФАМ), відповідно.

Розглянемо тепер так званий порядковий тест. Перший етап цього тесту аналогічний попереднім тестам. Нехай $x_{C_i}^{(k)} = (x_{1k}^{(i)}, x_{2k}^{(i)}, \dots, x_{15k}^{(i)})$ ($i = 1,2,\dots, j_k; k = 1,2,\dots, n$) - вектор показників клітини C_i пацієнта Q_k з групи G_1 , а $Y_{D_i}^{(k)} = (y_{1k}^{(i)}, y_{2k}^{(i)}, \dots, y_{15k}^{(i)})$, ($i = 1,2,\dots, l_k; k = 1,2,\dots, m$) - вектор відповідних показників пацієнта $\bar{Q}_k \in G_2$. На другому етапі обчислюється усереднений вектор показників

$$\bar{X}^{(k)} = \frac{1}{j_k} \sum_{t=1}^{j_k} X_{C_t}^{(k)} = (\bar{x}_{1k}, \bar{x}_{2k}, \dots, \bar{x}_{15k})$$

для кожного пацієнта $\bar{c}_t^{\min}, \bar{c}_t^{\max}$; аналогічно, для кожного пацієнта $\bar{Q}_k \in G_2$ усереднений вектор показників $\bar{Y}^{(k)}$ має вигляд

$$\bar{Y}^{(k)} = \frac{1}{l_k} \sum_{t=1}^{l_k} Y_{D_t}^{(k)} = (\bar{y}_{1k}, \bar{y}_{2k}, \dots, \bar{y}_{15k}).$$

Покладемо $X_{tk}^{\min} = \min(x_{tk}^{(1)}, x_{tk}^{(2)}, \dots, x_{tk}^{(j_k)}), k = 1,2,\dots, n; t = 1,2,\dots, 15;$

$$X_{tk}^{\max} = \max(x_{tk}^{(1)}, x_{tk}^{(2)}, \dots, x_{tk}^{(j_k)}), k = 1,2,\dots, n; t = 1,2,\dots, 15;$$

$$Y_{tk}^{\min} = \min(y_{tk}^{(1)}, y_{tk}^{(2)}, \dots, y_{tk}^{(l_k)}), k = 1,2,\dots, m; t = 1,2,\dots, 15;$$

$$Y_{tk}^{\max} = \max(y_{tk}^{(1)}, y_{tk}^{(2)}, \dots, y_{tk}^{(l_k)}), k = 1,2,\dots, m; t = 1,2,\dots, 15;$$

$$\begin{aligned}
a_t^{\min} &= \min(x_{t1}^{\min}, x_{t2}^{\min}, \dots, x_{tm}^{\min}), \quad a_t^{\max} = \max(x_{t1}^{\min}, x_{t2}^{\min}, \dots, x_{tm}^{\min}), \\
b_t^{\min} &= \min(x_{t1}^{\max}, x_{t2}^{\max}, \dots, x_{tm}^{\max}), \quad b_t^{\max} = \max(x_{t1}^{\max}, x_{t2}^{\max}, \dots, x_{tm}^{\max}), \\
\bar{a}_t^{\min} &= \min(y_{tk}^{\min}, k = 1, 2, \dots, m); \quad \bar{a}_t^{\max} = \max(y_{tk}^{\min}, k = 1, 2, \dots, m); \\
\bar{b}_t^{\min} &= \min(y_{tk}^{\max}, k = 1, 2, \dots, m); \quad \bar{b}_t^{\max} = \max(y_{tk}^{\max}, k = 1, 2, \dots, m); \\
c_t^{\min} &= \min(\bar{x}_{tk}, k = 1, 2, \dots, n), \quad c_t^{\max} = \max(\bar{x}_{tk}, k = 1, 2, \dots, n), \\
I_t &= \begin{cases} 1, & \text{if } \bar{x}_t \notin \gamma_t, \\ 0, & \text{if } \bar{x}_t \in \gamma_t, \end{cases} \quad \bar{c}_t^{\max} = \max(y_{tk}, k = 1, 2, \dots, m), \quad t = 1, 2, \dots, 15;
\end{aligned}$$

Тоді a_t^{\min}, a_t^{\max} будуть мінімальною і максимальною порядковими статистиками відповідно, як і $b_t^{\min}, b_t^{\max}, \bar{a}_t^{\min}, \bar{a}_t^{\max}, \bar{b}_t^{\min}, \bar{b}_t^{\max}, c_t^{\min}, c_t^{\max}, \bar{c}_t^{\min}, \bar{c}_t^{\max}$. За допомогою цих порядкових статистик визначаються довірчі інтервали $\alpha_t = (a_t^{\min}, a_t^{\max})$, $\beta_t = (b_t^{\min}, b_t^{\max})$, $\bar{\alpha}_t = (\bar{a}_t^{\min}, \bar{a}_t^{\max})$, $\bar{\beta}_t = (\bar{b}_t^{\min}, \bar{b}_t^{\max})$, $\gamma_t = (c_t^{\min}, c_t^{\max})$, $\bar{\gamma}_t = (\bar{c}_t^{\min}, \bar{c}_t^{\max})$.

Нехай Q - пацієнт, що обстежується, і $X_{C_i} = (x_1^{(i)}, x_2^{(i)}, \dots, x_{15}^{(i)})$, $i=1, 2, \dots, j$ - вектор показників цього пацієнта. На третьому етапі порядкового тесту ми обчислюємо вектори усереднених показників пацієнтів Q_j :

$$\bar{X} = \frac{1}{j} \sum_{t=1}^j X_{C_t} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_5)$$

і показники $x_t^{\min} = \min(x_t^{(i)}, i = 1, 2, \dots, j)$, $x_t^{\max} = \max(x_t^{(i)}, i = 1, 2, \dots, j)$, $t=1, 2, \dots, 15$,

а потім визначаємо індикатори виходу за межі $I_t^{\min}, I_t, I_t^{\max}$:

$$I_t^{\min} = \begin{cases} 1, & \text{if } x_t^{\min} \notin \alpha_t, \\ 0, & \text{if } x_t^{\min} \in \alpha_t, \end{cases} \quad I_t = \begin{cases} 1, & \text{if } \bar{x}_t \notin \gamma_t, \\ 0, & \text{if } \bar{x}_t \in \gamma_t, \end{cases} \quad I_t^{\max} = \begin{cases} 1, & \text{if } x_t^{\max} \notin \beta_t, \\ 0, & \text{if } x_t^{\max} \in \beta_t, \end{cases}$$

$t=1, 2, \dots, 15$. Аналогічно визначаються індикатори $\bar{I}_t^{\min}, \bar{I}_t, \bar{I}_t^{\max}$. Далі, ми оцінюємо показники

$$\alpha_1 = \sum_{t=1}^{15} (I_t^{\min} + I_t + I_t^{\max}), \quad \alpha_2 = \sum_{t=1}^{15} (\bar{I}_t^{\min} + \bar{I}_t + \bar{I}_t^{\max}).$$

Ці показники також є мірами близькості між сканограмами інтерфазних ядер клітин пацієнта Q , що обстежується, і відповідними сканограмами пацієнтів, які хворіють на рак молочної залози і фіброаденоматоз, відповідно.

Калібрування навчальних вибірок. Діагностика онкологічних захворювань, запропонована нами, ґрунтується на застосуванні тестів і алгоритмах статистичної і геометричної теорії розпізнавання образів.

На першій стадії сформуємо дві групи сканограм пацієнтів $A = \{X_i\}_{i=\overline{1,N}}$ і $B = \{Y_j\}_{j=\overline{1,M}}$, діагнози яких точно ідентифіковані. Надалі для певності будемо вважати, що група A містить сканограми хворих із злоякісною пухлиною (рак молочної залози), а група B - із доброякісним новоутворенням (фіброаденоматоз). Після процедур реєстрації і визначення морфо- і денситометричних показників ми одержуємо навчальні вибірки для кожного показника x_k ($k = 1, 2, \dots, 15$): $G_A^{(1)}, G_A^{(2)}, \dots, G_A^{(15)}$ для пацієнтів групи A і $G_B^{(1)}, G_B^{(2)}, \dots, G_B^{(15)}$ для пацієнтів групи B .

Розглянемо проблему визначення числа навчальних вибірок у групах A і B , що забезпечує достатньо високий рівень достовірності діагностики.

Спочатку природно припустити, що число вибірок у групах A і B повинно бути однаковим. Для прийняття або відхилення цієї гіпотези ми застосували *процедуру калібрування навчальних вибірок (method 1-hold-out)*, що складається з таких етапів:

1. Виключимо пацієнта $X_i, i = \overline{1, N}$ (або $Y_j, j = \overline{1, M}$) із множини $A \cup B$.
2. На основі множини вибірок $\{A \cup B\} \setminus X_i$ (або $\{A \cup B\} \setminus Y_j$) побудуємо тести, що використовують пари еліпсів $(E_{ts}, \bar{E}_{ts}), (E_{ts}^*, \bar{E}_{ts}^*)$ і напівплощин $(\pi_{ts}, \lambda_{ts}), (\pi_{ts}^*, \lambda_{ts}^*)$.
3. Обчислимо значення статистик $h_k = h(C_k)$ ($k = \overline{1, 6}$) для пацієнта $X_i, i = \overline{1, N}$ (або $Y_j, j = \overline{1, M}$).
4. Повернемо пацієнта $X_i, i = \overline{1, N}$ (або $Y_j, j = \overline{1, M}$) у множину $A \cup B$ і повторимо процедуру для іншого пацієнта.

Введемо в розгляд такі критерії діагностики:

- 1) квадратичний: $h_3 > h_4 \Rightarrow \text{РМЖ}; h_3 \leq h_4 \Rightarrow \text{ФАМ};$
- 2) лінійний: $h_5 > h_6 \Rightarrow \text{РМЖ}; h_5 \leq h_6 \Rightarrow \text{ФАМ}.$

Введемо такі позначення: D_1 - діагноз "РМЖ", D_2 - діагноз "ФАМ", v_{11} - частота події D_1 серед вибірок РМЖ, v_{21} - частота події D_2 серед вибірок РМЖ, v_{12} - частота події D_1 серед вибірок ФАМ, v_{22} - частота події D_2 серед вибірок ФАМ.

Аналіз результатів, отриманих при калібруванні вибірок із груп A і B однакового обсягу (табл.1) дозволяє зробити такі висновки.

1. У переважній більшості випадків для групи B (навчальна вибірка пацієнтів із ФАМ) спостерігається перевищення (домінування) статистики h_4 (тотальний ФАМ) над h_3 (тотальний РМЖ), а також статистики h_2 (ФАМ) над h_1 (РМЖ) (ми будемо називати цей феномен *ефектом стійкого домінування*), а для групи A такого ефекту не спостерігається.

2. Для лінійного критерію події D_1 і D_2 практично рівноймовірні, як для групи A (навчальна вибірка пацієнтів із РМЖ), так і для групи B (навчальна вибірка пацієнтів із ФАМ), тому цей критерій непридатний для діагностики РМЖ і ФАМ.

Таблиця 1.

Частота випадкової події D_k ($k = 1,2$) при калібруванні навчальних вибірок (25 РМЖ і 25 ФАМ)				
Частоти	v_{11}	v_{21}	v_{22}	v_{12}
Критерії				
Квадратичний	0.28	0.72	0.80	0.20
Лінійний	0.56	0.44	0.48	0.52
Комбінований	0.72	0.28	0.80	0.20

3. Квадратичний критерій для групи B дає значно кращі результати (у 80% випадків ставиться правильний діагноз - подія D_2 , а в 20% неправильний діагноз - подія D_1), а для групи A спостерігається протилежний ефект - у 28% випадків ставиться правильний діагноз, в у 72% - неправильний діагноз, тому цей критерій так само непридатний для диференціальної діагностики.

Таким чином, використання навчальних вибірок A і B однакового обсягу і лише самих квадратичного і лінійного критеріїв не дозволяє одержати прийнятні для клінічної медицини результати. У зв'язку з цим було здійснене калібрування навчальних вибірок для випадку, коли група A (25 сканограм пацієнтів із РМЖ) майже вдвічі перевищувала по обсязі групу B (12 сканограм пацієнтів із ФАМ); такий вибір обсягів повинний був забезпечити

перевищення (домінування) статистики h_3 (тотальний РМЖ) над h_4 (тотальний ФАМ), а також h_1 (РМЖ) над h_2 (ФАМ) (табл. 2).

Таблиця 2.

Частота випадкової події D_k ($k = 1, 2$) при калібруванні навчальних вибірок (25 РМЖ і 12 ФАМ)

Критерії	Частоти	v_{11}	v_{21}	v_{22}	v_{12}
Квадратичний		0.92	0.08	0.42	0.56
Лінійний		0.68	0.32	0.58	0.42
Комбінований		0.92	0.08	0.58	0.42

Аналізуючи результати калібрування, ми прийшли до таких висновків:

1. У переважній більшості випадків для групи A спостерігається перевищення (домінування) статистики h_3 (тотальний РМЖ) над h_4 (тотальний ФАМ), а також h_1 (РМЖ) над h_2 (ФАМ) (ефект стійкого домінування); а для групи B такого ефекту не спостерігається.

2. Для лінійного критерію події D_1 і D_2 практично рівноймовірні, тому він непридатний для диференціальної діагностики РМЖ і ФАМ.

3. Квадратичний критерій для групи A дають гарні результати (у 92% випадків ставиться правильний діагноз - подія D_1 , а в 8% неправильний - подія D_2), а для групи B правильний діагноз ставиться в 56% випадків, а неправильний - у 44%; тому цей критерій непридатний для диференціальної діагностики.

Відзначимо, що описаний вище ефект домінування статистики h_3 над h_4 для групи A виявляється лише у випадку, коли розміри поля реєстрації сканограми варіюють у досить вузьких межах (у даному випадку від 56 до 81); якщо цей принцип порушується, те утворюється статистично неоднорідна вибірка, і ефект домінування стає слабо вираженим.

Підводячи результати проведених досліджень, ми повинні констатувати, що для вибірок A і B із співвідношенням обсягу 2:1 (точніше кажучи, $\left[\frac{1}{2} \text{card } A\right] \approx \text{card } B$, де $[x]$ означає цілу частину числа x) використання самих лише квадратичних і лінійних критеріїв не приводить до задовільних результатів. Проте описаний вище ефект стійкого домінування частот, що спостерігається при навчальних вибірках однакового обсягу

($card A \approx card B$) і вибірках з обсягами $\left[\frac{1}{2} card A \right] \approx card B$ дозволяє сформулювати так званий фільтруючий критерій, головна ідея якого полягає в наступному.

Фільтруючий критерій. Розглянемо спочатку результати калібрування для навчальних вибірок з однаковим обсягом. Як було показано, у цьому випадку для групи B (пацієнти з ФАМ) виникає ефект стійкого домінування статистики h_4 (тотальний ФАМ) над h_3 (тотальний РМЖ), а для групи A (пацієнти з РМЖ) цей ефект відсутній. Нехай Q - пацієнт, діагноз якого ми намагаємося поставити. Обчислимо для нього на підставі цих груп (навчальних вибірок) A і B значення статистик $h_3(Q)$ і $h_4(Q)$. Припустимо, що $h_3(Q) \geq h_4(Q)$. З якою гіпотезою D_1 і D_2 найкраще узгоджуються ці експериментальні дані? Оскільки для групи B спостерігається ефект стійкого домінування $h_3(Q) \leq h_4(Q)$, то можливість події "пацієнт P має те ж захворювання, що і пацієнти групи B (тобто ФАМ)" буде малою, отже в цього пацієнта більш ймовірно захворювання РМЖ, виходить, гіпотеза D_2 краще узгоджується з експериментальними даними. Якщо ж для даного пацієнта $h_3(Q) < h_4(Q)$, то ми не можемо прийняти на підставі двох груп A і B однакового обсягу ніякого рішення про захворювання пацієнта (тобто не можемо поставити діагноз), оскільки з такими даними можуть бути як пацієнти з РМЖ, так і з ФАМ.

Ми побудували так званий A -фільтр, що дозволяє в деяких випадках діагностувати РМЖ у пацієнтів, що обстежуються. Будемо говорити, що сканограми обстежуваного пацієнта проходять A -фільтр, якщо $h_3(Q) \geq h_4(Q)$ і не проходять A -фільтр, якщо $h_3(Q) < h_4(Q)$.

Перейдемо тепер до опису B -фільтра. Для цього використовується інша пара навчальних вибірок A і B з обсягами $\left[\frac{1}{2} card A \right] \approx card B$, тобто коли обсяг A вдвічі перевищує обсяг B . Дані калібрування показують, що для цього випадку спостерігається обернений ефект стійкого домінування $h_3(Q) > h_4(Q)$ для групи A . Отже, значення статистик $h_3^*(Q)$ і $h_4^*(Q)$, які отримані для обстежуваного пацієнта за допомогою цієї пари навчальних вибірок A і B , що задовольняють нерівності $h_3^*(Q) \leq h_4^*(Q)$, свідчать із більшою ймовірністю про наявність у пацієнта ФАМ, ніж РМЖ; якщо ж $h_3^*(Q) > h_4^*(Q)$, то

діагноз не ставиться (неприйняття рішення). Таким чином, ми описали другу частину фільтруючого критерію, який називається B -фільтром, що дозволяє в деяких випадках діагностувати ФАМ у що обстежаться пацієнтів. Будемо говорити, що сканограм обстежуваного пацієнта проходять B -фільтр, якщо $h_3^*(Q) \leq h_4^*(Q)$, і не проходять B -фільтр, якщо $h_3^*(Q) > h_4^*(Q)$.

Тоді схема квадратичного фільтруючого критерію може бути описана в такий спосіб.

1. Формуються дві пари навчальних вибірок A і B з обсягами $card A \approx card B$ і

$\left[\frac{1}{2} card A \right] \approx card B$. Перша пара використовується для побудови A -фільтра, а друга - B -фільтра.

2. Проводиться описаний вище процес фільтрації сканограм обстежуваного пацієнта через A -фільтр і B -фільтр. Якщо ці сканограми проходять A -фільтр, то ставиться діагноз РМЖ, а якщо ж вони проходять B -фільтр - діагноз ФАМ; у протилежному випадку, коли вони не проходять ні через один фільтр, діагноз не ставиться (неприйняття рішення).

Квадратичний фільтруючий критерій має деякий інтерес для клінічної медицини, проте задовільним його не можна визнати, оскільки в цього критерію достатньо велика можливість неприйняття рішення.

У зв'язку з цим виникає ідея використання комбінованого фільтруючого критерію (лінійного, квадратичного і порядкового), до опису якого ми зараз переходимо.

Комбінований фільтруючий критерій також складається з двох фільтрів: A -фільтра і B -фільтра. Як і раніше, позначимо через $h_i, i = \overline{1,8}$ визначені раніше статистики обстежуваного пацієнта P , що отримані за допомогою навчальних вибірок A і B приблизно однакового обсягу ($card A \approx card B$), а через $h_i^*, i = \overline{1,8}$ відповідних статистики, що отримані за допомогою навчальних вибірок $A: card=25$ і $B: card=12$. Будемо вважати, що сканограми пацієнта P проходять A -фільтр, якщо виконується хоча б одне з таких нерівностей:

$$h_3 > h_4; \alpha_1 \leq \alpha_2$$

(тобто справедливо висловлення $h_3 > h_4 \vee \alpha_1 \leq \alpha_2$), і проходять B -фільтр, якщо має місце одне з нерівностей:

$$h_3^* \leq h_4^*; \alpha_1^* > \alpha_2^*$$

(тобто справедливо висловлення $h_3^* \leq h_4^* \vee \alpha_1^* > \alpha_2^*$).

Комбінований фільтруючий критерій формується в такий спосіб: якщо сканограми обстежуваного пацієнта проходять A -фільтр, то ставиться діагноз РМЖ, якщо B -фільтр – ФАМ, а в усіх інших випадках діагноз не ставиться (неприйняття рішення). Використання процедури калібрування навчаючих вибірок для оцінки якості цього критерію некоректно, оскільки при цьому утворюються ті ж результати, що і для квадратичного фільтруючого критерію; крім того, лише один порядковий критерій у цьому випадку дає безпомилковий діагноз.

У зв'язку з цим для експериментальної перевірки якості запропонованих критеріїв були вибрані 17 хворих раком молочної залози (РМЖ) і 7 хворих фіброаденоматозом. Усі ці пацієнти не входили в навчальні вибірки A і B . A -фільтр був побудований на підставі сканограм 24 хворих РМЖ (група A) і 25 хворих ФАМ (група B), а B -фільтр - на підставі сканограм 25 хворих РМЖ (група A) і 12 хворих ФАМ (група B). Результати тестування по обох фільтрах приведені в табл.3.

Аналіз експериментальних результатів показує, що у випадку використання комбінованого фільтруючого критерію ми можемо одержати три можливості: 1) ми діагностуємо ФАМ для обстежуваного пацієнта; 2) ми діагностуємо РМЖ; і 3) ми не діагностуємо ніякого захворювання (неприйняття рішення). Якщо ми діагностуємо ФАМ за умови, що пацієнт хворий на РМЖ, то виникає похибка першого роду. Якщо ми діагностуємо РМЖ за умови, що пацієнт хворий на ФАМ, то виникає похибка другого роду. На основі результатів експерименту ми можемо констатувати, що ймовірність (точніше, частота) ФАМ-діагнозу для РМЖ-пацієнта приблизно дорівнює 0.06, а ймовірність РМЖ-діагнозу для ФАМ-пацієнта дорівнює 0.00. Крім того, ймовірність діагностувати захворювання (прийняття рішення) дорівнює 94% для РМЖ-пацієнтів і 43% для ФАМ-пацієнтів. Таким чином, ми можемо діагностувати рак молочної залози з високою ймовірністю після першого аналізу, а для більш точного діагностування фіброаденоматозу ми повинні повторити аналіз.

Таблиця 3.

Частота випадкової події D_k ($k = 1, 2$) при тестуванні сканограм пацієнтів по A - і B -фільтрам

Частоти	v_{11}	v_{21}	v_{22}	v_{12}
Критерії				

Квадратичний	0.29	0.06	0.22	0.11
Лінійний	0.35	0.06	0.67	0.11
Комбінований	0.94	0.06	0.43	0.00

Повторні аналізи. Для зменшення кількості невизначених відповідей проводилися два типи повторних аналізів:

1) *Послідовний аналіз.* Якщо після першого аналізу рішення не прийняте, те повторювалося взяття зіскобу, вимір сканограм, обчислення показників і процедура фільтрації сканограм через A -фільтр і B -фільтр.

2) *Множинний аналіз.* Щоб збільшити точність відповіді (тобто, зменшити ймовірність помилково-позитивного і помилково-негативного діагнозів) і зменшити ймовірність непевних відповідей, одночасно береться декілька зіскобов буккального епітелію, і з кожним зіскобом здійснюється описана вище процедура комп'ютерної діагностики.

Проаналізуємо проблему виключення неприйняття рішення і постановки певного діагнозу шляхом послідовного аналізу. Надалі будемо думати, що обсяг групи обстежуваних пацієнтів достатньо велика, так що частота h і відповідна ймовірність p події (правильного діагнозу, помилкового діагнозу і неприйняття рішення) практично однакові. Припустимо, що після першого аналізу можливість (частота) помилкового діагнозу дорівнює α , можливість неприйняття вирішення дорівнює β , а можливість правильного діагнозу дорівнює γ , так що $\alpha + \beta + \gamma = 1$. Нехай n - обсяг групи обстежуваних пацієнтів. Позначимо через i_k число помилкових діагнозів, m_k - число випадків неприйняття рішення і l_k - число правильних діагнозів, отриманих при k -м повторенні аналізу. Припустимо, що при кожному повторенні аналізу ймовірності α, β, γ не змінюються. Тоді $i_1 = \alpha n$, $m_1 = \beta n$, $l_1 = \gamma n$, $i_2 = \alpha\beta n$, $m_2 = \beta^2 n$, $l_2 = \gamma\beta n$, $i_3 = \alpha\beta^2 n$, $m_3 = \beta^3 n$, $l_3 = \gamma\beta^2 n$, ..., $i_k = \alpha\beta^{k-1} n$, $m_k = \beta^k n$, $l_k = \gamma\beta^{k-1} n$, оскільки послідовний аналіз застосовується лише для пацієнтів із невизначеним діагнозом (неприйняття рішення). Після N повторень аналізу формула для обчислення числа помилкових діагнозів $i(N)$ має вигляд

$$i(N) = \sum_{k=1}^N i_k = \alpha n \sum_{k=1}^N \beta^{k-1} = \alpha n \frac{1 - \beta^{N-1}}{1 - \beta} \approx \frac{\alpha n}{1 - \beta},$$

оскільки при великому N величина β^{N-1} є достатньо малою, то для кількості випадків неприйняття рішення маємо

$$m(N) = m_N = \beta^N n \approx 0.$$

Таким чином, при великих N можливість помилкового діагнозу α_N визначається формулою

$$\alpha_N = \frac{i(N)}{n} = \alpha \frac{1 - \beta^{N-1}}{1 - \beta} < \frac{\alpha}{1 - \beta},$$

а формула для обчислення ймовірності неприйняття рішення β_N має вигляд

$$\beta_N = \frac{m(N)}{n} = \beta^N,$$

отже, при $\beta < \frac{1}{2}$ ймовірність помилкового діагнозу зростає не більш, ніж вдвічі, а ймовірність неприйняття рішення є нехтувано малою величиною (оскільки β^N дуже швидко прямує до нуля при зростанні N , то при $N = 5$, як правило, $\beta^N \approx 0$). Отже, можливість правильного діагнозу дорівнює $\gamma_N = 1 - \alpha_N - \beta_N$.

Наприклад, відповідно до результатів комп'ютерної діагностики для осіб, що страждають на рак молочної залози, ми маємо $\alpha = 0.06$, $\beta = 0$, $\gamma = 0.94$, так що для послідовного аналізу теоретична оцінка ймовірності є постійною і не залежить від N . Проте для осіб, хворих на фіброаденоматоз $\alpha = 0$, $\beta = 0.57$, $\gamma = 0.43$, і

$$\alpha_2 = 0, \beta_2 = 0.3249, \gamma_2 = 0.6751,$$

$$\alpha_3 = 0, \beta_3 = 0.1852, \gamma_3 = 0.8148,$$

$$\alpha_4 = 0, \beta_4 = 0.1056, \gamma_4 = 0.8944,$$

$$\alpha_5 = 0, \beta_5 = 0.0602, \gamma_5 = 0.9398,$$

$$\alpha_6 = 0, \beta_6 = 0.0343, \gamma_6 = 0.9657.$$

Оцінка числа повторення аналізів, що залежить від значення довірчого рівня β^* для ймовірності неприйняття рішення визначається формулою:

$$N = \left\lceil \frac{\ln \beta^*}{\ln \beta} \right\rceil + 1.$$

Розглянемо тепер множинний аналіз. Варто підкреслити, що в цьому випадку, на відміну від послідовного аналізу, повторні аналізи не залежать від прийняття або неприйняття рішення про захворювання пацієнта, і здійснюються для усіх пацієнтів

одночасно. Позначимо через A_1 подію, що полягає в постановці помилкового діагнозу при першому аналізі (терміни "перший", "другий" і так далі означають лише порядковий номер у серії зіскобов для множинного аналізу), A_2 - у неприйнятті рішення (невизначений діагноз) і A_3 - у постановці правильного діагнозу. Відповідно до уведеного вище позначеннями, можливості цих подій визначаються в такий спосіб $p(A_1) = \alpha$, $p(A_2) = \beta$, $p(A_3) = \gamma$. В результаті множинного аналізу можуть відбутися такі 9 подій: $A_i^{(1)} A_j^{(2)}$, $i, j = 1, 2, 3$, де добуток подій означає, що вони відбуваються одночасно, верхній індекс означає порядковий номер аналізу.

Припустимо, що пацієнт P хворіє РМЖ, тоді подію A_1 означає помилковий діагноз "ФАМ", а A_3 - правильний діагноз "РМЖ". Після серії з двох аналізів помилковий діагноз "ФАМ" ставиться в таких випадках: $A_1^{(1)} A_1^{(2)}$, $A_1^{(1)} A_2^{(2)}$, $A_2^{(1)} A_1^{(2)}$, тому ймовірність помилкового діагнозу визначається в такий спосіб:

$$\begin{aligned} p_1 &= p\left(A_1^{(1)} A_1^{(2)} + A_1^{(1)} A_2^{(2)} + A_2^{(1)} A_1^{(2)}\right) = \\ &= p\left(A_1^{(1)} A_1^{(2)}\right) + p\left(A_1^{(1)} A_2^{(2)}\right) + p\left(A_2^{(1)} A_1^{(2)}\right) = \\ &= \left(p\left(A_1^{(1)}\right)\right)^2 + 2p\left(A_1^{(1)}\right)p\left(A_2^{(1)}\right) = \alpha^2 + 2\alpha\beta, \end{aligned}$$

за умови, що результати повторних аналізів є незалежними подіями (як і раніше, ми припускаємо, що при повторних аналізах частоти α, β, γ постійні).

Неприйняття рішення виникає у випадках $A_2^{(1)} A_2^{(2)}$, $A_1^{(1)} A_3^{(2)}$, $A_3^{(1)} A_1^{(2)}$, і ймовірність p_2 цієї події обчислюється так:

$$\begin{aligned} p_2 &= p\left(A_2^{(1)} A_2^{(2)} + A_1^{(1)} A_3^{(2)} + A_3^{(1)} A_1^{(2)}\right) = \\ &= \left(p\left(A_2^{(1)}\right)\right)^2 + 2p\left(A_1^{(1)}\right)p\left(A_3^{(2)}\right) = \beta^2 + 2\alpha\gamma. \end{aligned}$$

Нарешті, правильний діагноз "РМЖ" ставиться у випадках $A_3^{(1)} A_3^{(2)}$, $A_3^{(1)} A_2^{(2)}$, $A_2^{(1)} A_3^{(2)}$, і ймовірність p_3 правильного діагнозу визначається формулою

$$p_3 = \gamma^2 + 2\beta\gamma.$$

Якщо

$$\alpha + 2\beta < 1, \quad 2\alpha\gamma < \beta(1 - \beta), \quad (3)$$

те $p_1 = \alpha^2 + 2\alpha\beta < \alpha$, $p_2 = \beta^2 + 2\alpha\gamma < \beta$, $p_3 = \gamma^2 + 2\beta\gamma > \gamma$. Отже, можливості помилкового діагнозу і неприйняття рішення зменшуються. Зауважимо, що нерівності (3) вірні завжди, якщо $0 < \delta < \beta < \frac{1}{2}$ і α практично дорівнює нулю. Подібна ситуація виникає і для пацієнтів, хворих на ФАМ.

Таким чином, відповідно до вищенаведених результатів, для пацієнтів, що страждають на РМЖ, $\alpha = 0.06$, $\beta = 0$, $\gamma = 0.94$, і при подвійному аналізі одержуємо такі теоретичні оцінки можливостей помилкового діагнозу, неприйняття рішення і правильного діагнозу:

$$p_1 = 0.0036, p_2 = 0.1128, p_3 = 0.8836.$$

Для осіб, хворих на ФАМ, $\alpha = 0$, $\beta = 0.57$, $\gamma = 0.43$ і

$$p_1 = 0, p_2 = 0.3249, p_3 = 0.6751.$$

Обчислення змін ймовірностей α, β, γ при множинному аналізі більш високого порядку здійснюються по вищевказаній схемі.

При повторних аналізах можливість помилкового діагнозу РМЖ або ФАМ незначно збільшується, а процент правильних діагнозів досягає 91 %.

Висновок. Проведені дослідження показали правомочність використання текстурних і денситометричних показників інтерфазних ядер епітеліоцитів буккального епітелію для виявлення злоякісних процесів, а також для диференціальної діагностики між доброякісними і злоякісними процесами в молочній залозі. Перевагами запропонованого методу є неінвазивність, можливість багатократного повторення аналізу і об'єктивність, яка обумовлена кількісним аналізом цитологічної інформації.

Для порівняння, зауважимо, що в Західній Європі [14-17] для постановки діагнозу РМЖ у клінічній практиці, усі пацієнти з підозрою на це захворювання повинні бути піддані таким дослідженням: клінічне обстеження, маммографія і/або ультразвукове дослідження, тонкоігольна аспірація. Цей комплекс діагностичних досліджень називається Triple Assessment. Точність індивідуального тесту пацієнта при цьому така: при клінічному обстеженні - до 84%; маммографії - до 78% (у залежності від віку пацієнта і розміру пухлини ця точність може зменшуватися); тонкоігольної аспірації - до 91%. Чутливість комплексу Triple Assessment, у випадку, коли хоча б один тест дає позитивний результат, лежить у межах від 95 до 100% , а точність - до 89%. Якщо три тести дають позитивні результати, то чутливість досягає 99%. У зв'язку з цим, варто зауважити, що кожне з обстежень, що входять у комплекс Triple Assessment, має значні хиби. Так, клінічне обстеження істотно залежить від

кваліфікації лікаря. Крім того, точність цього методу неможливо значно підвищити при багатократному повторенні обстеження. Діагностичні похибки при маммографії значно залежать від віку пацієнта, типу і розміру пухлини, а багатократне його повторення є небезпечним для здоров'я хворого. При використанні методу тонкоігольної аспірації не виключене засівання каналу проколювання пухлинними клітинами, що збільшує ризик рецидиву злоякісної пухлини молочної залози і викликає необхідність застосування більш радикальної операції у мінімально віддалені терміни. По тим же причинам повторні обстеження цим методом цілком виключені.

Розроблений нами комп'ютерний цитологічний метод діагностики раку молочної залози позбавлений перерахованих вище хиб і має більшу точність і чутливість при багатократних аналізах (при цьому максимально необхідне число аналізів на практиці не перевищує 4-5).

Література

1. Ганина К.П., Полищук Л.З., Бородай Н.В., Налескина Л.А., Исакова Л.М., Несина И.П., Бучинская Л.Г. Цитологическая реактивность онкологического больного. - Киев, Наукова думка. - 1995. - 151 с.
2. Nieburg H.E., Herman B.E., and Reisman H. Buccal cell changes in patients with malignant tumours // *Lab. Invest.* - 1962. -11, N 1.- P.80-88.
3. Palcic B., Calum MacAulay. Malignancy associated changes: Can they be employed clinically? // *Compendium on the Computerized Cytology and Histology Laboratory (Tutorials of Cytology, Chicago, Illinois, USA, 1994).* - PP.157-165.
4. Ogden G.R., Cowpe J.G., Green M.W. The effect of distant malignancy upon quantitative cytologic assessment of normal oral mucosa // *Cancer.* - 1990. - 65, N 3. - P.477-480.
5. Капанцян А.Л., Каралова Е.М., Магакян Ю.А. Коэффициент взаиморасположения частиц и его использование при количественном анализе вещества в клетках // *цитология.* - 1988. - 30, № 3. - С. 361-366.
6. Папаян Г.В., Магакян Ю.А., Агроскин Л.С., Каралова Е.М. Использование данных сканирующей цитометрии для анализа поглощающего вещества в клетках // *Цитология.* - 1982. - 24, № 11. - С. 1359-1366.
7. Бойко Ю.В., Литвинко П.Г., Петунин Ю.И., Ганина К.П. Показатель кластерности структуры хроматину интерфазных ядер // *Докл. АН УССР.* - 1990. - № 5. - С. 75-78.

8. Магакян Ю.А., Каралова Е.М. Цитофотометрия ДНК. - Ереван: Изд-во АН Армении, 1989. - 203 с.
9. Петунин Ю.И., Тимошенко Я.Г., Петунина М.Ю. Критерий для идентификации генеральной совокупности в случае конечного класса альтернативных гипотез // Докл. АН УССР. - Сер.А. - 1984, № 6. - С. 29-32.
10. Мадреимов И., Петунин Ю.И. Характеризация равномерного распределения с помощью порядковых статистик // Теор. вероятностей и мат. статистика. - 1982. - 27. - С. 96-102.
11. Ван дер Варден. Математическая статистика. - М.: Мир, 19__ . - __ с.
12. Andrushkiw R.I., Kljushin D.A., Petunin Yu.I. Nonlinear algorithm of pattern recognition for computer-aided diagnosis of breast cancer // Nonlinear analysis (Elsevier Science Ltd, Oxford). - 1997. - vol. 30. - P.5431-5436.
13. Петунин Ю.И., Рублев Б.В. Распознавание образов с помощью квадратичных дискриминантных функций // Вычисл. и прикл. математика. - 1996. - Вып. 80. - С. 90-105.
14. Ciatto S., Cariaggi P., Bulgaresi P., Confortini M., Bonardi R. Fine needle aspiration cytology of the breast: review of 9533 consecutive cases // The Breast. - 1993. - 2. - P. 87-90.
15. Ciatto S., Catarzi S., Morrone D., Rosselli Del Turco M. Fine needle aspiration cytology of nonpalpable breast lesions: US versus stereotaxic guidance // Radiology. - 1993. - 188. - P.195-199.
16. Dixon J.M., Anderson T.J., Lamb J., Nixon S.J., Forrest A.P.M. Fine needle aspiration cytology in relationship to clinical examination and mammography in the diagnosis of solid breast mass // Br. J. Surg. - 1984. - 71. - P. 593-596.
17. Layfield M., Glasgow B.J., Cramer H. Fine-needle aspiration in the management of breast masses // Pathol-Annu. - 1989. - 24 - P.23-62.