
КОМП'ЮТЕРНИЙ ЦИТОГЕНЕТИЧНИЙ МЕТОД ДЛЯ ДІАГНОСТИКИ РАКУ

Київ – 2012

Епідеміологічні показники

- Рак молочної залози - це найпоширеніше онкологічне захворювання серед жінок. Щороку в Україні виявляють близько 16 тисяч випадків захворювання і близько 8 тисяч жінок гинуть від цієї недуги.
- Кожний новий випадок захворювання реєструється кожні 35 хвилин, майже кожну годину вмирає від раку молочної залози одна жінка.

Причини смертності

- Внаслідок пізнього звернення до лікаря кожна п'ята жінка вмирає впродовж першого року після встановлення діагнозу.
- За питомою вагою за давнених форм раку молочної залози Україна знаходиться на одному з перших місць (близько 25% жінок, що звертаються до клінік, мають III–IV стадію захворювання).

Причини смертності

- На базі однієї з київських клінік було організовано обстеження для жінок, які могли безоплатно зробити знімок і проконсультуватися у мамолога. При цьому за півроку обстеження пройшло лише 700 жінок.
- В Україні за рік лише 40 жінок із 1000 проходять мамографію.

Переваги існуючих методів

- У Західній Європі для постановки діагнозу використовується Triple Assessment – комбінація клінічного обстеження, мамографії і тонкогольчастої аспірації. Загальна точність цього стандарту 99,6% (Br J Cancer 2009; 100: 1873–1878).

Недоліки клінічного обстеження і мамографії

- Деякі дослідники вважають, що близько 12% випадків раку молочної залози є результатом рентгенівського опромінення, отриманого жінками при проходженні мамографії з метою діагностики.
- Виключення тонкогольчастої аспірації знижує точність і чутливість діагностики: клінічне обстеження - 40-69 і 88-95%, мамографія - 77-95% і 45-90% відповідно.
- При розмірах до 5-10 мм пухлину важко виявити як клінічно, так і за допомогою мамографа.

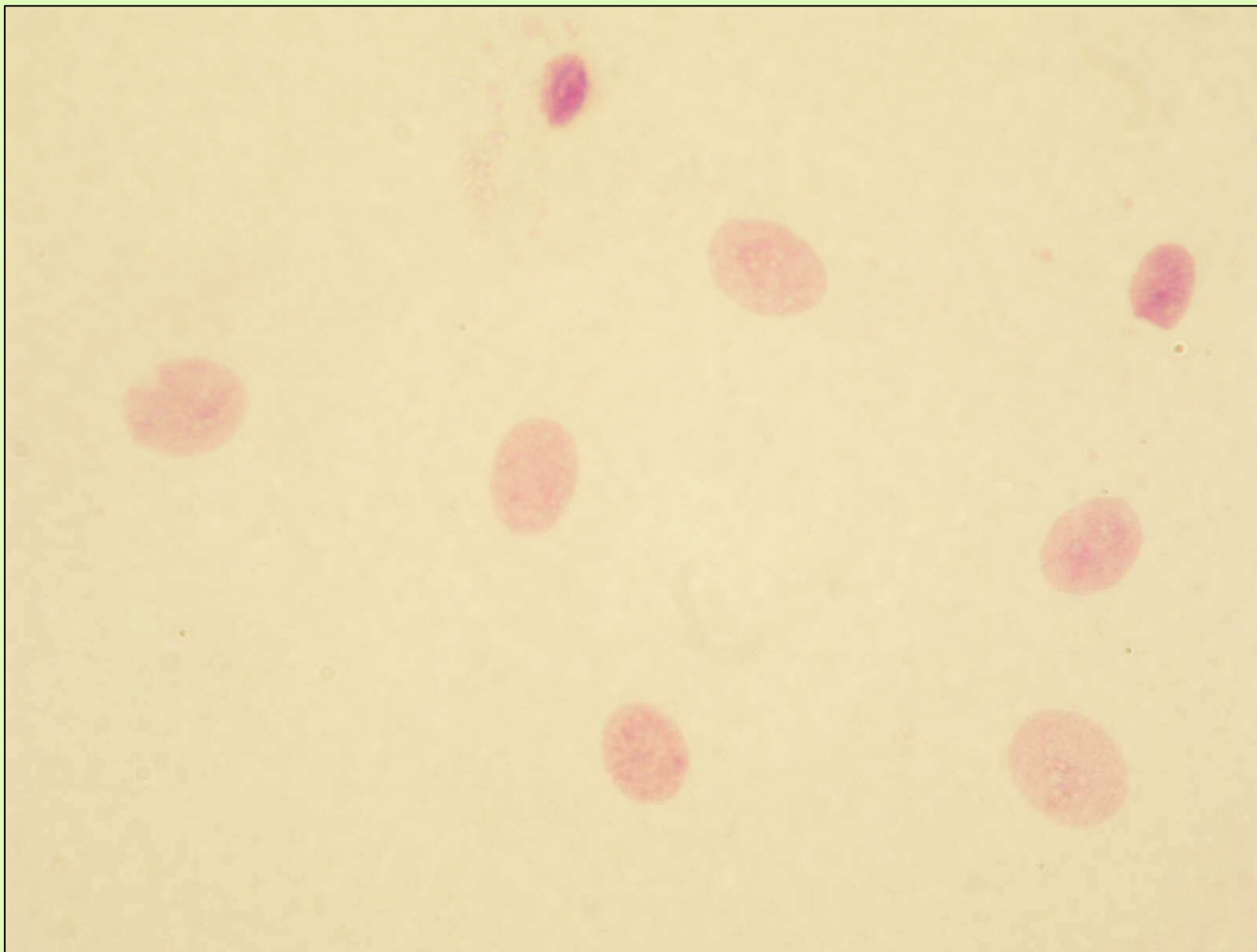
Недоліки тонкогольчастої аспірації

- При використанні тонкогольчастої аспірації виникає небезпека метастазів.
- Вірогідність методу тонкогольчастої аспірації за даними різних авторів складає від 60 до 98%.

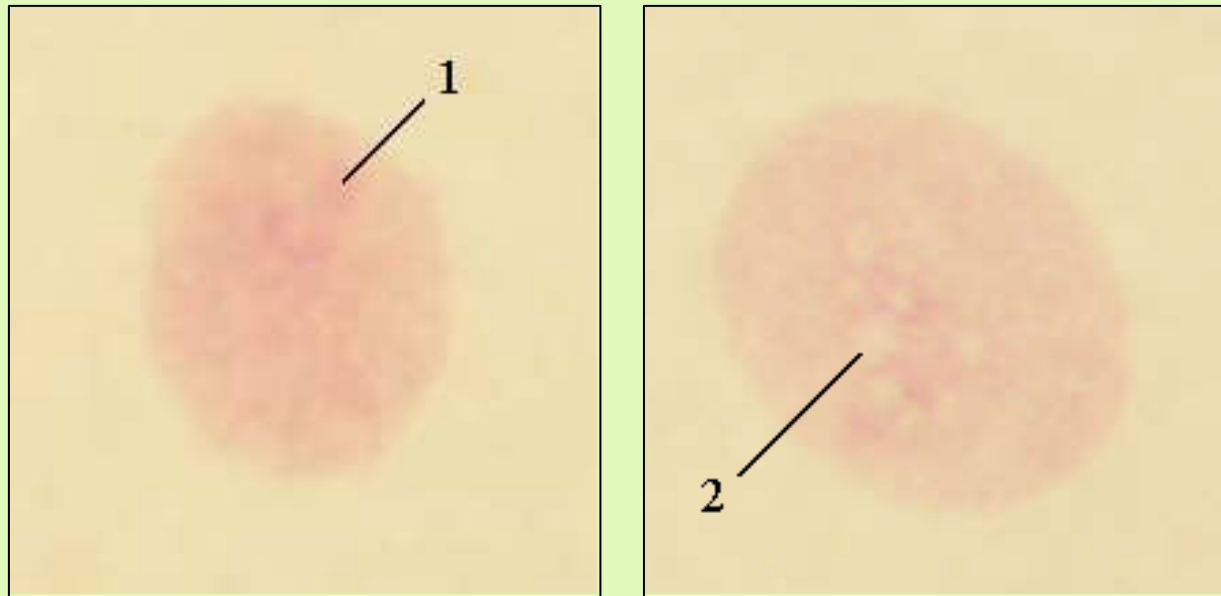
Загальна схема діагностики

- 1) підготовка біологічного матеріалу;
 - 2) отримання зображень;
 - 3) попередня обробка зображень (виділення ядер);
 - 4) обчислення параметрів ядер;
 - 5) застосування до параметрів методу класифікації.
-

Попередня обробка зображень



Розподіл хроматину в ядрах



Розподіл хроматину в ядрах букального епітелію:

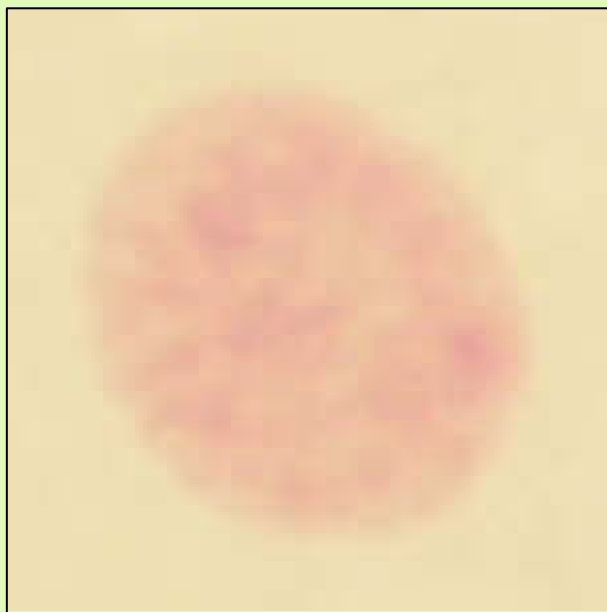
1 – конденсований хроматин, 2 – деконденсований хроматин.

Виділення ядер: правило 3σ

$$y_j \in G_1 \iff y_j < \bar{x}_\phi - 3\sigma_\phi$$

$$\bar{x}_\phi = \frac{1}{M} \sum_{i=1}^M x_i$$

$$\sigma_\phi = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (x_i - \bar{x}_\phi)^2}$$



Адаптивна порогова обробка зображень

Нехай N – кількість точок зображення, L – кількість градацій сірого кольору, C_1 – клас з інтенсивностями $[1, \dots, k]$, C_2 – клас з $[k+1, \dots, L]$

$$T = k \left| \sigma_{C_1 C_2}^2 \right. \max \quad \sigma_{C_1 C_2}^2 = \omega_1 \omega_2 (\mu_2 - \mu_1)^2$$

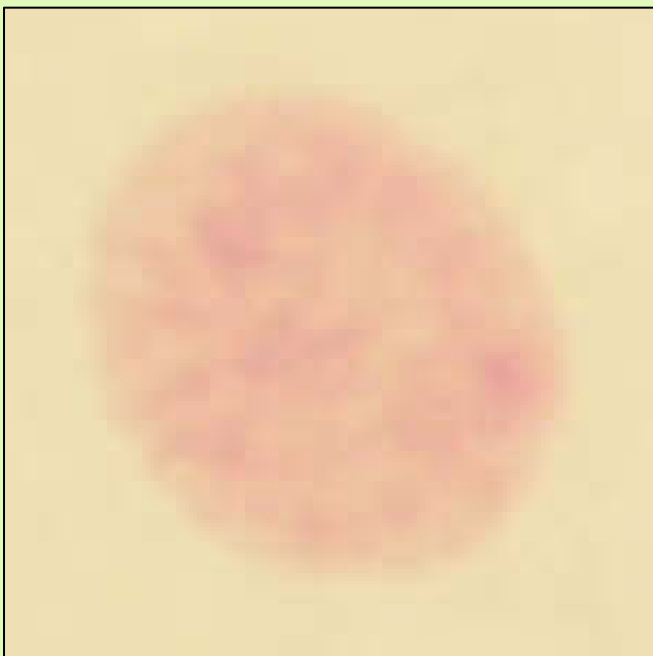
де ω_1 – ймовірність потрапляння в C_1

ω_2 – ймовірність потрапляння в C_2

μ_1 – значення середньої яскравості в C_1

μ_2 – значення середньої яскравості в C_2

Адаптивна порогова обробка зображень



Обчислення параметрів ядер

- 1) Площа ядра
- 2) Мінімальне значення
- 3) Максимальне значення
- 4) Середнє значення
- 5) Середньоквадратичне відхилення
- 6) Коефіцієнт асиметрії
- 7) Коефіцієнт ексцесу
- 8) Ентропія
- 9) Енергія
- 10) Медіана
- 11) Верхній квартиль
- 12) Нижній квартиль
- 13) 5%-процентіль
- 14) 95%-процентіль

p -статистика

Нехай $x = (x_1, \dots, x_n) \in G_1$ та $y = (y_1, \dots, y_n) \in G_2$

$$x_{(1)} \leq \dots \leq x_{(n)}, \quad y_{(1)} \leq \dots \leq y_{(n)}, \quad F_{G_1}(u) = F_{G_2}(u)$$

$$P_{ij}^{(n)} = p(A_{ij}^{(k,n)}) = p\left(y_k \in (x_{(i)}, x_{(j)})\right) = \frac{j-i}{n+1}$$

***p*-статистика, продовження**

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

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

де $h_{ij}^{(n)}$ - частота появи події $A_{ij}^{(k,n)}$ в n випробуваннях,

$$g = 3.$$

ρ -статистика, продовження

$$h^{(N)} = \rho(x, y) = \frac{L}{N},$$

де L – кількість подій $B_{ij}^{(n)} = \{p_{ij}^{(n)} \in I_{ij}^{(n)}\}$,

N – кількість інтервалів $I_{ij}^{(n)} = (p_{ij}^{(n,1)}, p_{ij}^{(n,2)})$,

$$N = n(n-1)/2$$

Довірчий інтервал для p -статистики

$$I^{(N)} = \left(p^{(N,1)}, p^{(N,2)} \right),$$

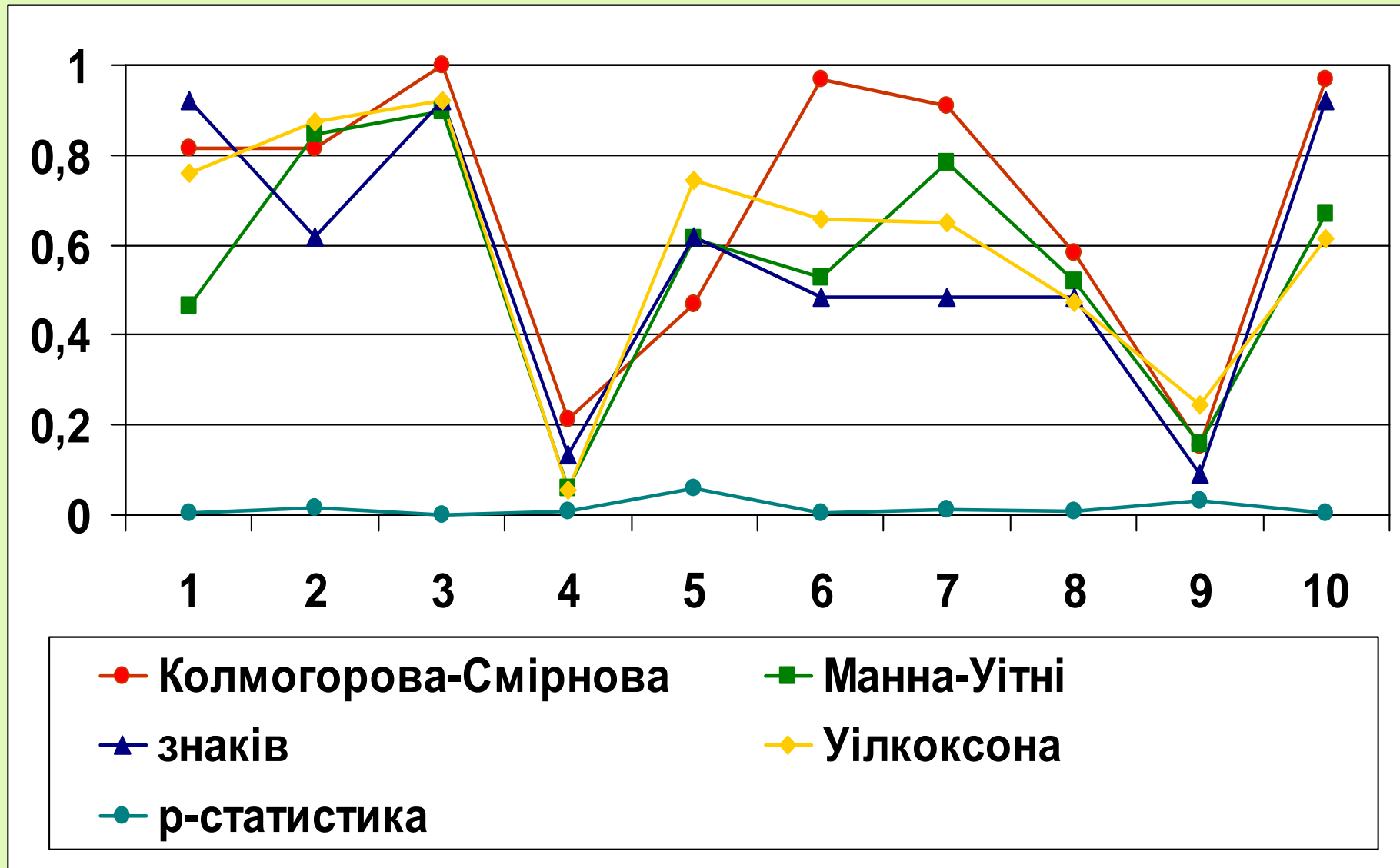
$$p^{(N,1)} = \frac{h^{(N)} N + g^2 / 2 - g \sqrt{h^{(N)} (1 - h^{(N)}) N + g^2 / 4}}{N + g^2},$$

$$p^{(N,2)} = \frac{h^{(N)} N + g^2 / 2 + g \sqrt{h^{(N)} (1 - h^{(N)}) N + g^2 / 4}}{N + g^2}.$$

Порівняння p -статистики

Критерій	Частота похибки при порівнянні вибірки з $N(0,1)$ з вибірками з генеральних сукупностей				
	$N(0,1)$	$N(0,2)$	$N(1,1)$	$U(0,1)$	$U(-1,1)$
Колмогорова-Смірнова	0	0,3	0	0	0,7
знаків	0	1	0	0,2	1
Уїлкоксона	0	0,9	0	0	1
Манна-Уїтні	0	1	0	0	1
p -статистика	0	0	0	0	0

$N(0;1)$ з $N(0;1)$



Усреднена ρ -статистика

Нехай $x = (x_1, \dots, x_n) \in G_1$, $y^{(m)} = (y_1^{(m)}, \dots, y_n^{(m)}) \in G_2$

$$x_{(1)} \leq \dots \leq x_{(n)}, \quad y_{(1)}^{(1)} \leq \dots \leq y_{(n)}^{(1)}, \quad y_{(1)}^{(m)} \leq \dots \leq y_{(n)}^{(m)}$$

$$h = \frac{1}{M} \sum_{m=1}^M h^{(n,m)},$$

де $h^{(n,m)} = \rho(x, y^{(m)})$.

Довірчий інтервал для p -статистики

$$I^{(N)} = \left(p^{(N,1)}, p^{(N,2)} \right),$$

$$p^{(N,1)} = \frac{h^{(N)} N + g^2 / 2 - g \sqrt{h^{(N)} (1 - h^{(N)}) N + g^2 / 4}}{N + g^2},$$

$$p^{(N,2)} = \frac{h^{(N)} N + g^2 / 2 + g \sqrt{h^{(N)} (1 - h^{(N)}) N + g^2 / 4}}{N + g^2}.$$

Математична статистична модель

Нехай $G_1 = \{X_i, i = 1, \dots, N\}$ та $G_2 = \{Y_j, j = 1, \dots, M\}$ – генеральні сукупності, кожен елемент яких характеризується k параметрами:

$$X_i = (x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)}) \quad Y_j = (y_1^{(j)}, y_2^{(j)}, \dots, y_k^{(j)})$$

Для кожного фіксованого $l, l = \overline{1, k}$, величини $x_l^{(i)}$ є однаково розподіленими, $y_l^{(j)}$ є також однаково розподіленими.

Необхідно встановити, до якої генеральної сукупності належить $z = (z_1, z_2, \dots, z_k)$

Розв'язання задачі класифікації

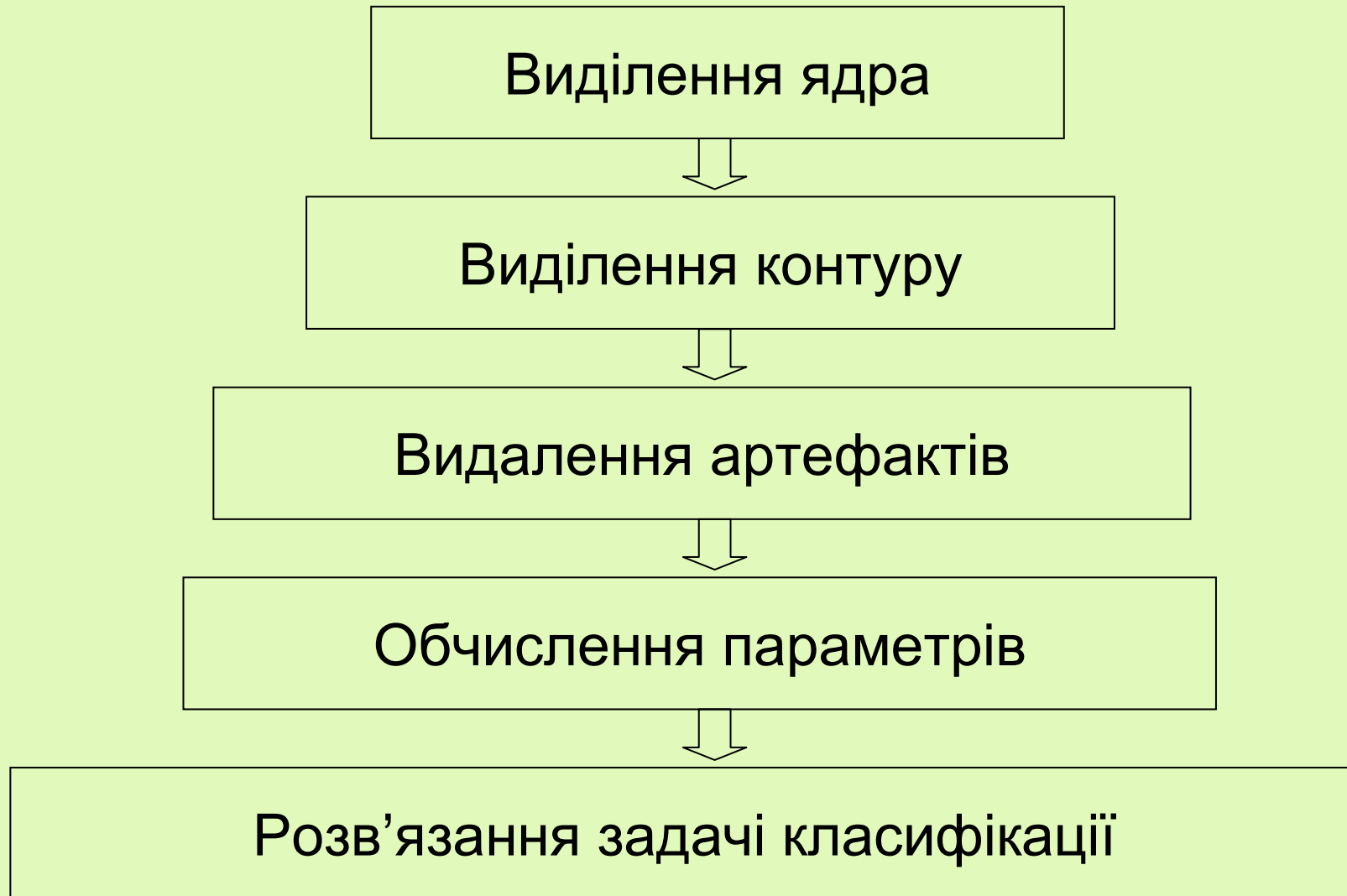
$$z \in G_1 \Leftrightarrow h^{(1)} > h^{(2)}, \quad z \in G_2 \Leftrightarrow h^{(1)} < h^{(2)},$$

$$h^{(1)} = \frac{1}{k} \sum_{l=1}^k h_l^{(1)}, \quad h^{(2)} = \frac{1}{k} \sum_{l=1}^k h_l^{(2)},$$

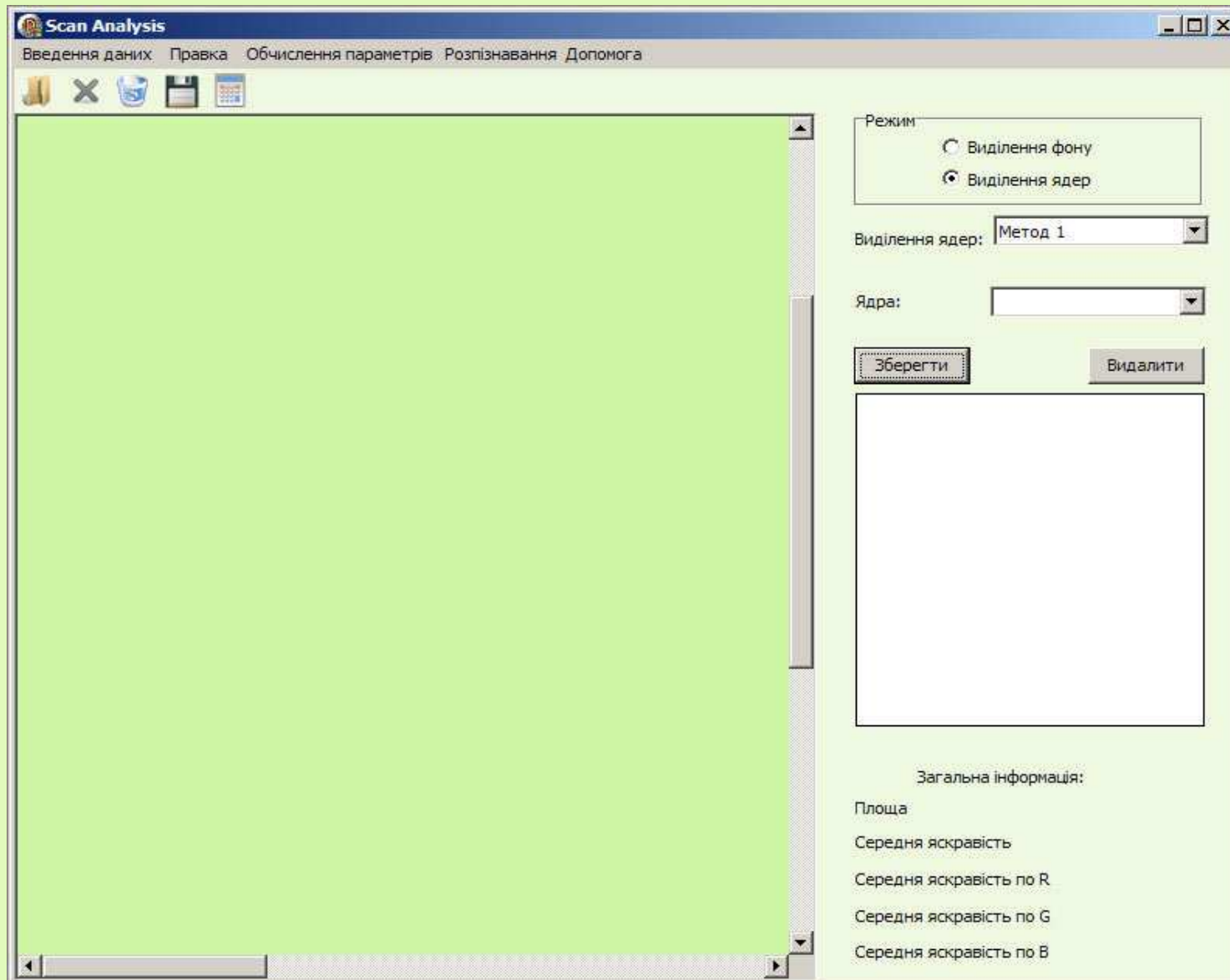
де $h_l^{(1)}$ – усереднена р-статистика між вибіркою z_l та множиною вибірок $x_l^{(i)}$, $i = \overline{1, N}$

$h_l^{(2)}$ – усереднена р-статистика між вибіркою z_l та множиною $y_l^{(j)}$, $j = \overline{1, M}$

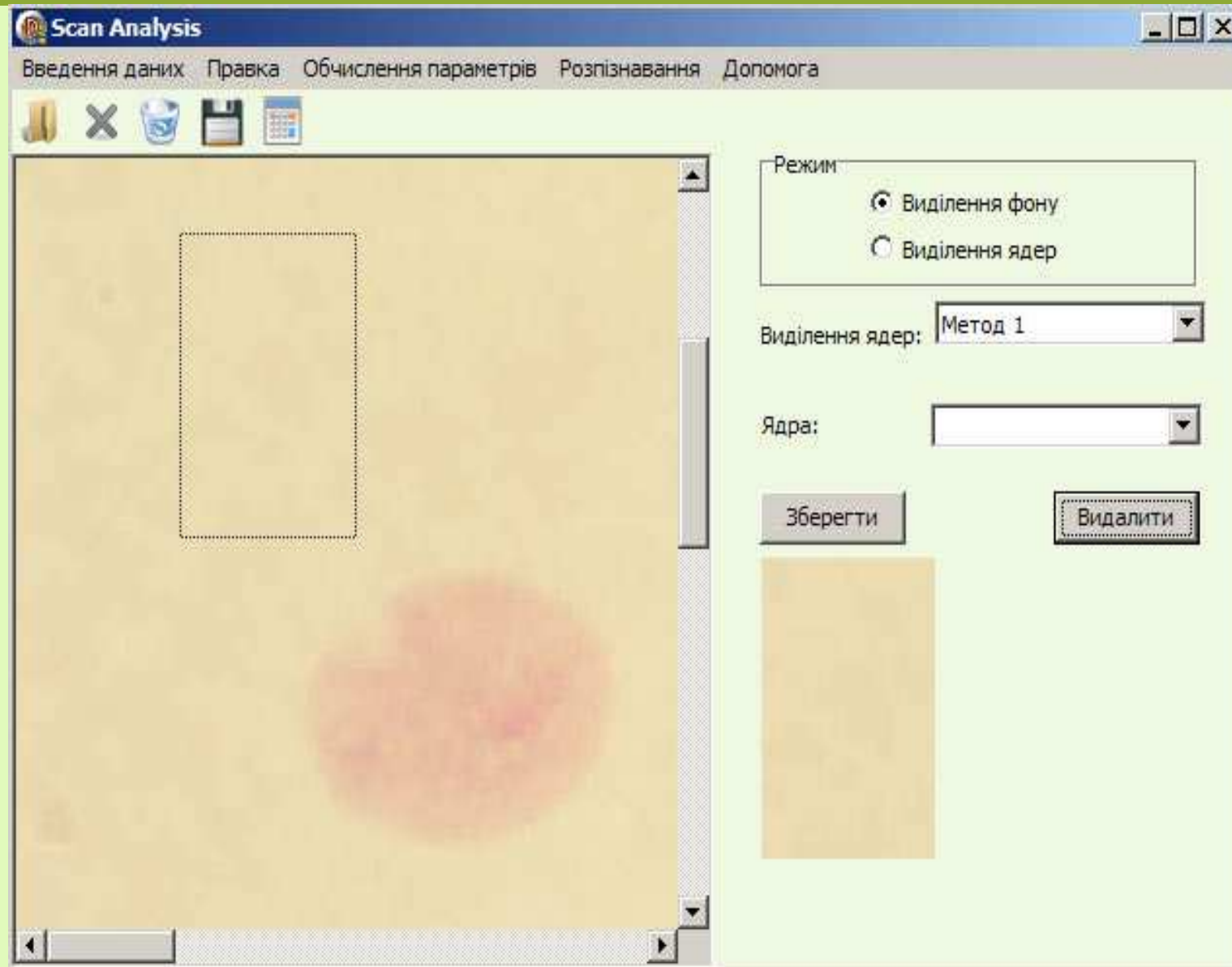
Програмний комплекс



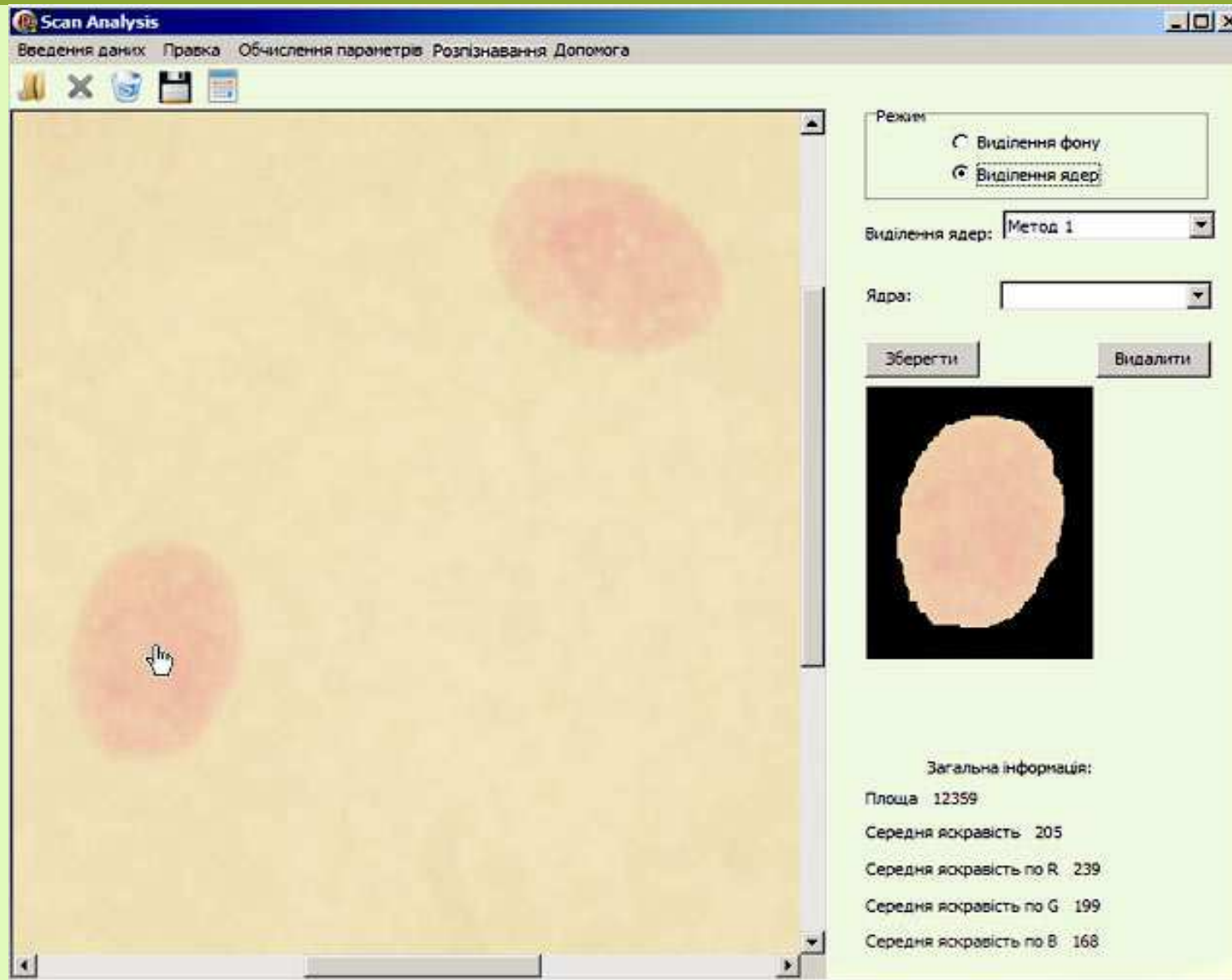
Програмний комплекс



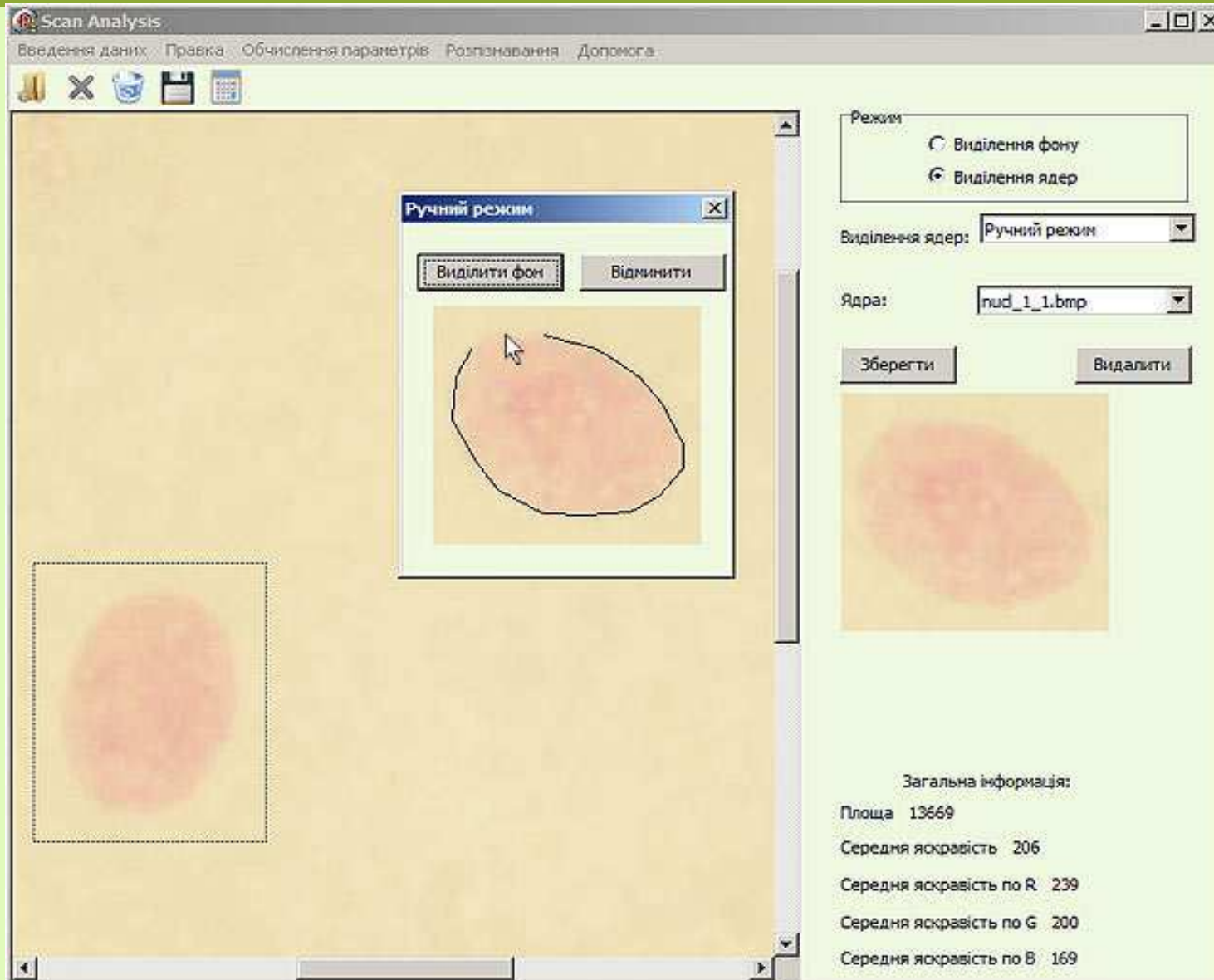
Програмний комплекс



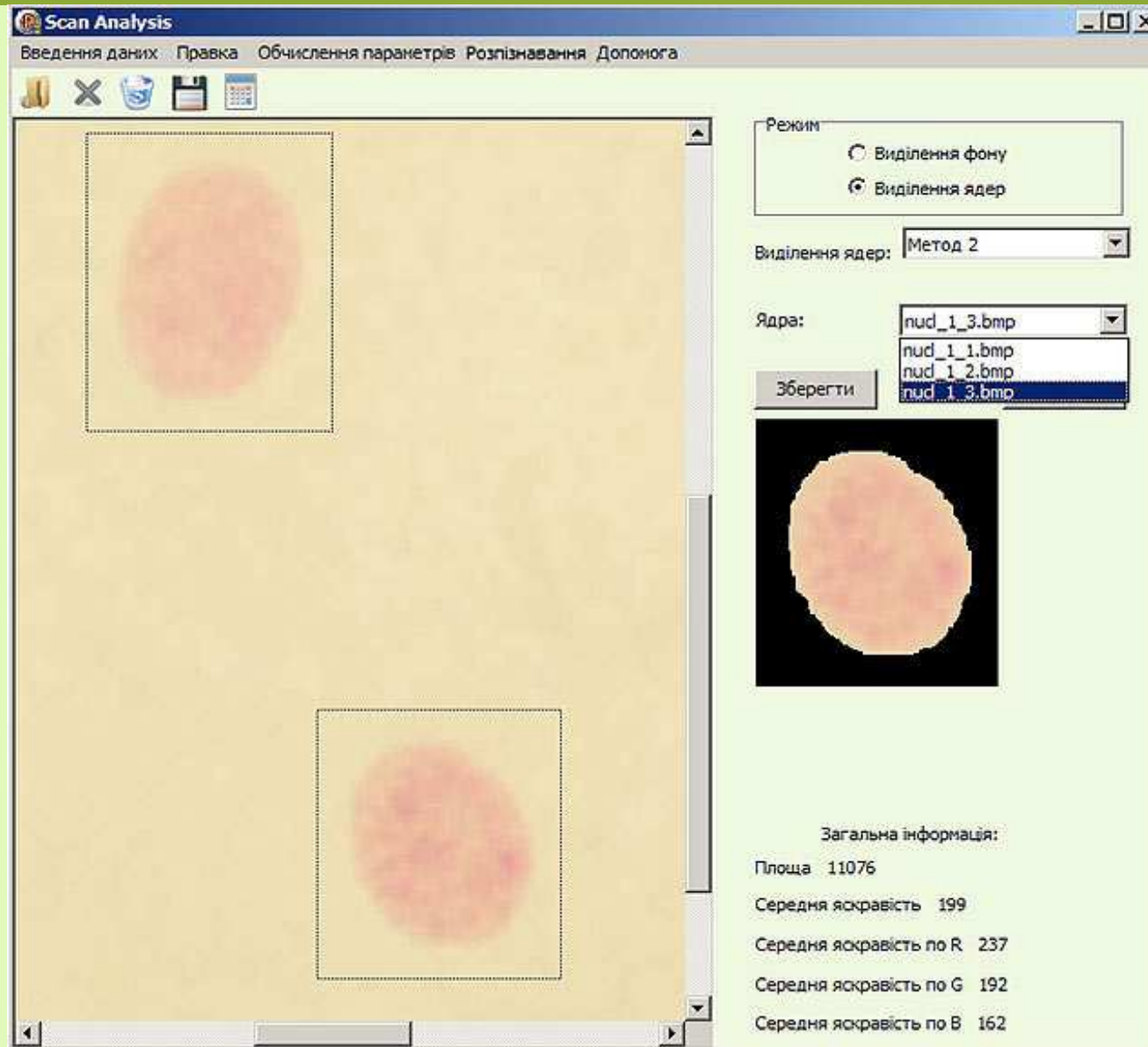
Програмний комплекс



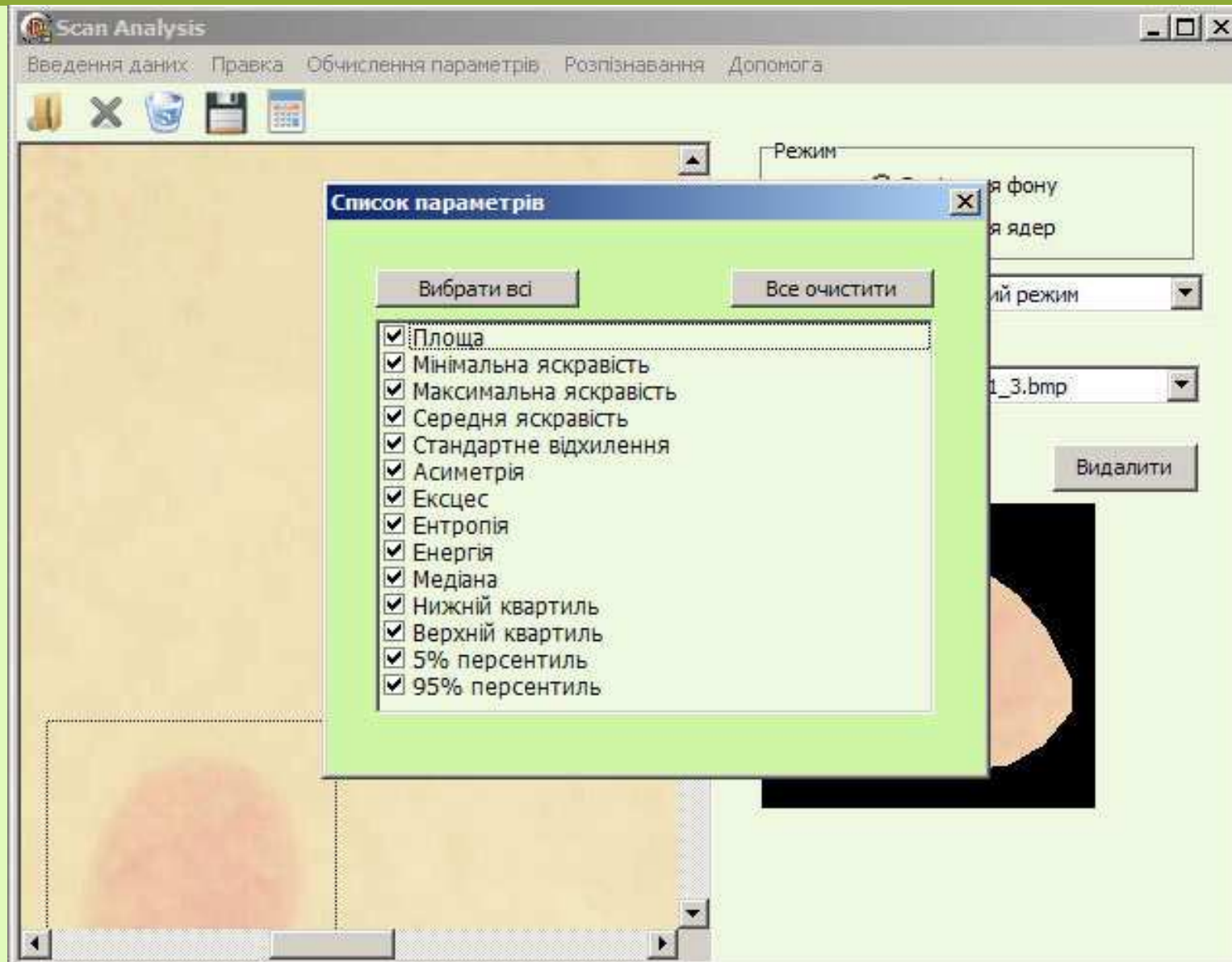
Програмний комплекс



Програмний комплекс



Програмний комплекс



Критерії діагностики

Критерій 1. Використовуються 14 вище вказаних показників та усереднена p-статистика. Похибка для діагностики раку дорівнює 7,9%, а фіброаденоматозу – 18,2%. Критерій має високий степінь точності і на відміну від більшості існуючих методів є абсолютно неінвазивним.

Критерій 2. Комп'ютерна цитогенетична діагностика раку молочної залози та фіброаденоматозу на підставі площі ядер. Похибка для ракових пацієнтів – 0%, для хворих на фіброаденоматоз – 18,18%.

Дякуємо за увагу!
