

УДК 616.71+533.9.07

**ПІДТРИМКА ДІАГНОСТИЧНИХ РІШЕНЬ НА ОСНОВІ КЛАСТЕРНОГО  
АНАЛІЗУ ЗОБРАЖЕНЬ ГАЗОРОЗРЯДНОГО СВІТІННЯ КРОВІ**

Н.В. Глухова\*, Л.А. Пісоцька\*\*

\*ДВНЗ «Національний гірничий університет»

\*\*ДВНЗ «Дніпропетровська медична академія МОЗ України»

e-mail: GINaVi@ukr.net

**Вступ.** Сучасна діагностика захворювань крові людини ґрунтується переважно на застосуванні стандартних дорогих гематологічних методів дослідження. Вони включають комплекс цитохімічних, імунофенотипових, патоморфологічних, гістологічних, вірусологічних, молекулярно-біологічних і серологічних методів [1, 2]. На їх основі визначають кількість лейкоцитів, лімфоцитів та інших формених або ферментних елементів крові, на підставі чого ставлять діагноз. Вказані методи оцінки кількісних параметрів складу крові дозволяють виявляти патологію крові на стадії клінічних проявлень.

**Мета та постановка завдання.** В основу досліджень авторів поставлено завдання удосконалення способу ранньої діагностики та ризику захворювань крові, зокрема лейкозів, на основі автоматизованої обробки зображень газорозрядного випромінювання мазків крові. Дослідження характеристик окремих крапель або мазків крові проводиться шляхом фіксації випромінювання від зразка у полі високовольтного розряду. Реєстрація структури газорозрядного випромінювання навколо дослідного об'єкту в електромагнітному полі відбувається за рахунок можливості візуалізації газорозрядного випромінювання у зоні контакту зразка з фотоматеріалом. При розташуванні краплі або мазка у зоні дії змінного електромагнітного поля з високою напруженістю навколо досліджуваного об'єкта виникає характерне світіння (ефект Кірліан) [3, 4].

**Основна частина.** Дослідження характеристик біологічних рідин здійснювалося шляхом фіксації газорозрядного випромінювання від зразка у полі високовольтного розряду. При цьому виконується фіксація структури газорозрядного світіння навколо дослідного зразка в електромагнітному полі з можливістю візуалізації зображення випромінювання у зоні контакту біологічного об'єкту з фотоматеріалом. Були проведені експериментальні дослідження газорозрядних випромінювань зразків крові в умовах медичного закладу у хворих на лейкомію в порівнянні з контролем.

У газовому розряді навколо зразка крові здорової людини має місце більш велика гіллястість і поширеність розряду (площа засвітлення), ніж у хворого. При лейкозах порушується структура крові взагалі, що відбивається на зображеннях газорозрядного випромінювання у вигляді пошкодження окремих фрагментів картини світіння.

З метою зменшення вкладу суб'єктивного фактору при реалізації діагностичних процедур запропоновано спосіб автоматизованої програмної обробки зображень газорозрядного випромінювання. У результаті сканування фотографій газорозрядного випромінювання дослідних зразків отримуємо напівтонове растрове зображення. Градації сірого кольору, які характеризують яскравість випромінювання окремих пікселів, зберігаються у вигляді двовимірною масиву. Для вихідних зображень структури газорозрядного випромінювання будують гістограму яскравості пікселів [5].

З метою підвищення достовірності дослідження зразків біологічних об'єктів за рахунок виключення випадкових впливаючих факторів у ході вимірювального експерименту виконуються серії досліджень. Наступний етап обробки полягає в отриманні усереднених оцінок кількості пікселів у кожному піддіапазоні яскравості для вибірки в цілому. Спеціально проведені дослідження для великої кількості типових зразків дозволили встановити, що розподіл значень кількості пікселів за піддіапазонами не

відповідає нормальному. Теоретично доказано [6], що гіпотеза нормального розподілу результатів експериментальних досліджень, що є підґрунтям усіх класичних результатів, не може бути обґрунтовано перевірена у межах однієї вибірки даних, а, крім того, фактично не може бути застосована для багатовимірного випадку.

З точки зору робастних методів обробки експериментальних даних, у таких випадках найкращою стійкою оцінкою найбільш ймовірного значення величини є значення медіани. З урахуванням вказаного теоретичного обґрунтування у якості кількісних характеристик оцінки параметрів газорозрядного випромінювання було обрано значення медіан кількості пікселів у піддіапазонах. Саме значення медіан є вихідними даними для кластерного аналізу зображень газорозрядного випромінювання. Математичний апарат кластерного аналізу був обраний, оскільки він являє собою специфічну методологію для реалізації класифікації неоднорідних статистичних сукупностей.

Основна задача, яка вирішується за допомогою кластерного аналізу, полягає у розподілу усієї множини вихідних об'єктів у однорідні групи. На сьогодні розроблено цілий спектр методів кластерного аналізу [7]. Але для використання методів у прикладних галузях необхідно розв'язання нетривіального наукового завдання – виділення з загального обсягу параметрів, властивостей, характеристик дослідних об'єктів саме таких, що забезпечать ефективне диференціювання вихідної множини об'єктів на однорідні підмножини.

Нехай  $X = \{x_1, x_2, \dots, x_n\}$  – множина з  $n$  біологічних об'єктів (зразків крові), зображення газорозрядного випромінювання яких характеризуються  $m$  ознаками. Тоді кожний біологічний об'єкт виступає як точка у  $m$ -вимірному просторі ознак. В загальному випадку вихідні дані для кластерного аналізу зображень випромінювання аналітично представляємо у вигляді матриці  $m \times n$ :

$$X_{m \times n} = \begin{pmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_1^m & x_2^m & \dots & x_n^m \end{pmatrix}. \quad (1)$$

Матриця вигляду (1) має назву матриця «об'єкт – властивість», елементи якої  $x_i^t$  виступають кількісною оцінкою  $t$ -ї ознаки  $i$ -го об'єкту. У кожному  $i$ -му стовпчику  $X_i = (x_i^1, x_i^2, \dots, x_i^m)$  записані параметри досліджуваного об'єкту  $x_i$ . Геометричний сенс кожного об'єкта – це точка  $m$ -вимірного простору ознак  $I^m(X)$ .

З урахуванням розглянутих позначень проблема кластерного аналізу зображень газорозрядного випромінювання полягає у розбитті однорідних класів вихідної експериментальної множини зображень  $X = \{x_1, x_2, \dots, x_n\}$ , представлених у вигляді «об'єкт - властивість». Приклад реальних даних для опису запропонованої методики наведений у табл.1. Кількість оброблених вибірок реальних даних за об'ємом значно перевищує представлений у табл. 1 обсяг матеріалу, а тому виходить за рамки можливостей відображення у публікації.

Для нечіткої кластеризації використаний алгоритм Беждека-Данна (Fuzzy ISODATA algorithm, FCM algorithm) [8]. Кластерний аналіз полягає у розв'язанні оптимізаційної задачі:

$$Q(P) \rightarrow \min.$$

Таблиця 1.

Значення медіан у піддіапазонах  $g_i$  гістограми яскравості пікселів зображень газорозрядного випромінювання мазків крові

№ пп	1	2	3	4	5	6	7	8
	контроль	контроль	лейкоз	лейкоз	контроль	контроль	лейкоз	лейкоз
$g_1$	0	0	0	0	0	0	0	0
$g_2$	106,5	308,5	2376,5	2097,5	303	217	1068,5	1691,5
$g_3$	2726,5	2271	9605	9587,5	2162	2606	8974	10161
$g_4$	5213	3824,5	6050,5	5084	3748	3713	4903	6340,5
$g_5$	3918,5	3915,5	5465	4586,5	2963,5	2734	3983,5	4746
$g_6$	4636	4164,5	5609	4178,5	4976	3579,5	3659	4394,5
$g_7$	5999,5	5055,5	6301	4909,5	6426,5	6068,5	4476,5	5093
$g_8$	5856	4918,5	6908	5635	5966	5579,5	4781,5	4876
$g_9$	7421,5	7699,5	10158	8074	7022	7720,5	6199	7055
$g_{10}$	37291	39687	64668	62996	46961	36353	58237	63404
$g_{11}$	44278	57632	95357	54863	48141	45885	67464	62230
$g_{12}$	2	0	2,5	1,5	4,5	2,5	0	1,5

Якщо вихідні дані представлено у вигляді матриці «об'єкт – властивість»  $X_{m \times n} = [x_i^t]$ ,  $i = 1 \dots n$ ,  $t = 1 \dots m$ , то функціонал якості розподілу на класи визначається залежністю:

$$Q(P) = \sum_{l=1}^c \sum_{i=1}^n \sum_{l=1}^n g(p(x_i), \mu_{li}) d(x_i, \tau^l),$$

де  $g$  – певний функціонал;  $p(x_i)$  – апіорна вага для кожної точки  $x_i$  при виконанні умови  $\sum_{i=1}^n p(x_i) = 1$ ;  $\mu_{li}$  – функція приналежності елементу  $x_i \in X$  деякому нечіткому кластеру  $A^l \in \{A^1, \dots, A^c\}$ ;  $d(x_i, \tau^l)$  – функція, що визначає відстань від елементу  $x_i$ ,  $x_i \in X$ ,  $i = 1, \dots, n$  до деякого елемента  $\tau^l$ ,  $\tau^l \in X$ ,  $l = 1, \dots, c$ . Обмеження виступають у якості умов:

$$\mu_{li} \geq 0, \sum_{l=1}^c \mu_{li} = 1, i = 1, \dots, n; l = 1, \dots, c.$$

Функціонал Дж. Беждека – Дж. Данна представляє собою зважений критерій якості розбиття [8]:

$$Q(P) = \sum_{l=1}^c \sum_{i=1}^n \mu_{li}^\gamma d(x_i, \tau^l),$$

де  $\gamma$  – показник нечіткості класифікації.

Критерій якості кластеризації використовується з урахуванням залежностей:

$$g(p(x_i), \mu_{li}) = \frac{\mu_{li}^\gamma}{n}, \quad 1 \leq \gamma \leq \infty,$$

$$\tau^l = \sum_{i=1}^n \mu_{li}^\gamma \cdot x_i / \sum_{i=1}^n \mu_{li}^\gamma.$$

Алгоритм, що оптимізує функціонал Дж. Беждека – Дж. Данна називається методом нечітких  $c$ -середніх (fuzzy  $c$ -means), та являє собою параметричне сімейство за  $\gamma$  при фіксованій кількості кластерів.

У якості метрики – функції відстані у функціоналі  $Q(P)$  використовується квадрат евклідової норми у  $m$ -вимірному просторі ознак  $I^m(X)$ :

$$d(x_i, \tau^l) = \|x_i - \tau^l\|^2.$$

Кластерний аналіз реалізується шляхом розв'язання наступної задачі класифікації:

$$P^* = \arg \min_P \begin{cases} Q_l(P): P = (A^1, \dots, A^c), A^l = (\mu_{l1}, \dots, \mu_{ln}), 0 \leq \mu_{li} \leq 1; \\ \sum_{l=1}^c \mu_{li} = 1, \sum_{i=1}^n \mu_{li} > 0, i = 1, \dots, n; l = 1, \dots, c; \end{cases}$$

де  $c$  – кількість нечітких кластерів у шуканому розбитті  $P^*$ ;  $\gamma$  – показник нечіткості класифікації.

Результати використання розглянутого алгоритму кластерного аналізу до вихідних даних у вигляді кількісної оцінки медіан для зображень газорозрядного випромінювання (табл. 1) мають наступний вигляд:

$$\begin{aligned} \text{class1} &= 3 \quad 4 \quad 7 \quad 8; \\ \text{class2} &= 1 \quad 2 \quad 5 \quad 6. \end{aligned}$$

**Висновки.** Розробка нових сучасних методів експрес-діагностики захворювань крові, що не потребують коштовного обладнання та мінімізують вклад суб'єктивних факторів при інтерпретації результатів експериментальних досліджень, виявляється актуальною проблемою. Розглянутий метод отримання зображень газорозрядного випромінювання зразків крові людини у комбінації з новітніми методами цифрової обробки та аналізу даних відповідають вимогам, що висуваються сучасною медициною.

#### Список літератури

1. Руководство по гематологии: в 3 т. Т.1. Под ред. А.И. Воробьева. М.: Ньюдиамед; 2002, 280 с.
2. Литвинов А.В. Способ дифференциальной диагностики развернутой и терминальной стадий хронического лейкоза // Патент РФ №2098821, МПК G01N 33/48, 1997.
3. Пісоцька Л.А., Глухова Н.В. Спосіб діагностики захворювань крові // Пат. 82295 Україна: МПК G01N 33/49 / Заявник та патентовласник Пісоцька Л.А. Заявл. 21.02.2013; опубл. 25.07.2013.
4. Глухова Н.В., Песоцкая Л.А., Опрятная Т.О. Кирлиан-фотографиямазков крови при анемиях // Академический журнал Западной Сибири Academic Journal of West Siberia - г.Тюмень, Россия. - 2013, Том 9, № 5 (48). –С.23-24.
5. Глухова Н.В., Песоцкая Л.А., Лапицкий В.Н. Анализ изображений кирлиановского свечения капель воды // Науковий вісник Національного гірничого університету. - №1, 2013, с. 91 - 96.
6. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика. 1988. – 176 с.
7. Fern, X.Z., Brodley, C.E. Clustering ensembles for high dimensional data clustering // InProc. International Conference on Machine Learning - 2003. - p. 186 - 193.
8. Bezdek J.C., Hathaway R.J., Howard R.E., Wilson C.A. Coordinate Descent and Clustering // Control and Cybernetics. – Vol. 15. – pp. 195 -204.