

УДК 616.71+533.9.07

ДІАГНОСТИКА ЗАХВОРЮВАНЬ КРОВІ НА ОСНОВІ КЛАСТЕРНОГО АНАЛІЗУ ЗОБРАЖЕНЬ ГАЗОРОЗРЯДНОГО СВІТІННЯ

Н. В. Глухова, Л. А. Пісоцька*

*ДВНЗ “Національний гірничий університет”, м. Дніпропетровськ
ДВНЗ “Дніпропетровська медична академія МОЗ України”**

Розглянуто проблеми діагностики захворювань крові та ризику їх розвитку, зроблено аналіз недоліків існуючих методів. Запропоновано новий підхід до реалізації скринінгових досліджень, який засновано на отриманні зображень газорозрядного випромінювання мазків або крапель крові в електромагнітному полі високої напруги на рентгенівській плівці. З метою підвищення достовірності діагностичних рішень шляхом усунення суб'єктивних факторів запропоновано автоматизовану комп'ютерну обробку зображень світіння крові на базі побудови гістограм яскравості пікселів та виконання процедур нечіткого кластерного аналізу за методом «fuzzy c-means». Обґрунтований вибір методології обробки зображень випромінювання біологічних об'єктів дозволив отримати задовільні результати у галузі скринінгової діагностики захворювань крові, раннього визначення ризику їх виникнення. Практична цінність запропонованої діагностичної методики підтверджується результатами обробки експериментальних даних з використанням кластерного аналізу.

Ключові слова: газорозрядне випромінювання крові, автоматизована комп'ютерна обробка.

ДИАГНОСТИКА ЗАБОЛЕВАНИЙ КРОВИ НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА ИЗОБРАЖЕНИЙ ГАЗОРОЗРЯДНОГО СВЕЧЕНИЯ

Н. В. Глухова, Л. А. Писоцкая*

*ГВУЗ “Национальный горный университет”, г. Днепропетровск
ГВУЗ “Днепропетровская медицинская академия МОЗ Украины”**

Рассмотрены проблемы диагностики заболеваний крови и раннего выявления риска их возникновения, выполнен анализ недостатков существующих методов. Предложен новый подход к реализации скрининговых исследований, который основан на получении изображений газоразрядного излучения мазков или капель крови на рентгеновской пленке в электромагнитном поле высокой напряженности. С целью повышения достоверности диагностических решений путем устранения субъективных факторов предложена автоматизированная компьютерная обработка изображений свечения крови на базе построения гистограм яркости пикселей и выполнения процедур нечеткого кластерного анализа по методу “fuzzy c-means”. Обоснованный выбор методологии обработки изображений излучения биологических объектов позволил получить удовлетворительные результаты в области скрининговой диагностики заболеваний крови и риска их возникновения. Практическая ценность предложенной диагностической методики подтверждается результатами обработки экспериментальных данных с использованием кластерного анализа.

Ключевые слова: газоразрядное свечение крови, автоматизированная компьютерная обработка.

DIAGNOSIS OF DISEASES OF THE BLOOD BASED ON CLUSTER ANALYSIS BASED ON GAS-DISCHARGE RADIATION

N. V. Hlukhova, L. A. Pisotska*

*SHEI “National Mining University”, Dnipropetrovsk
SHEI “Dnipropetrovsk Medical Academy of MPH of Ukraine”**

The problems of diagnosis of diseases of blood and the risk of their development, the analysis of the shortcomings of existing methods. A new approach to the implementation of screening studies, which is based on image capture gas-discharge radiation smears or drops of blood in the electromagnetic field of high voltage on x-ray film. In order to improve the reliability of diagnostic decisions by eliminating subjective factors proposed automated computer processing of images glow blood-based histogram brightness of pixels and performing procedures fuzzy cluster analysis method «fuzzy c-means». Informed choice of methodology radiation imaging of biological objects yielded satisfactory results in the field of diagnostic

© Н. В. Глухова, Л. А. Пісоцька

screening for blood diseases, early definition of risk incurred. The practical value of the proposed diagnostic technique is confirmed by the results of experimental data using cluster analysis.

Key words: radiation smears of drops of blood, automated computer processing.

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]. 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 проявів.

Застосовується також спосіб діагностики, запропонований в [2]. Суть його полягає у визначенні в плазмі крові людини вільної та зв'язаної води, на основі чого кількісно оцінюється значення коефіцієнта гідратації плазми крові. Але такий підхід також вимагає наявності дорогого устаткування, для експлуатації якого необхідний персонал спеціальної кваліфікації, що робить його неможливим для скринінгових обстежень.

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 відбитого випромінювання від поверхні зразка у полі високовольтного розряду. Реєстрація структури газорозрядного випромінювання навколо дослідного об'єкта 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).

Під час взаємодії краплі рідини з електромагнітним полем через неї протікає електричний струм, за рахунок протікання якого через газовий прошарок виникає іонізація оточуючого газу, що супроводжується світінням. Шляхом поширення спектра електромагнітного впливу у режимі збудження газорозрядного випромінювання крапель або мазків крові отримуємо зображення, на основі кореляції параметрів

якого зі структурними змінами крові при наявності патологічних процесів здійснюється діагностика захворювань крові [3, 4].

Основна частина. Дослідження характеристик біологічних рідин здійснювалося шляхом фіксації газорозрядного випромінювання від поверхні 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'єкта з фотоматеріалом.

Були проведені експериментальні дослідження газорозрядних випромінювань зразків крові в умовах медичного закладу у хворих на лейкоемію в порівнянні з контролем.

При якісному аналізі експерт враховує наступні основні ознаки газорозрядного випромінювання: внутрішнє кільце, з якого виходять радіально спрямовані стримери, які утворюють середнє кільце та тонкі люмінесценції, що характеризують ширину зовнішньої засвітки; структуру випромінювання у зоні контакту біологічного об'єкта з фотоматеріалом.

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

З метою зменшення внеска суб'єктивного фактора при реалізації діагностичних процедур запропоновано спосіб автоматизованої програмної обробки зображень газорозрядного випромінювання. У результаті сканування фотографій газорозрядного випромінювання дослідних зразків отримуємо півтонове растрове зображення. Градації сірого кольору, які характеризують яскравість випромінювання окремих пікселів, зберігаються у вигляді двовимірного масиву. Для вихідних зображень структури газорозрядного випромінювання будують гістограму яскравості пікселів. Гістограма являє собою графік, що складається з 256 стовпчиків, які відповідають градаціям яскравості кольору для півтонового растрового зображення структури випромінювання [5].

З метою підвищення достовірності дослідження зразків біологічних об'єктів за рахунок виключення випадкових факторів впливу у ході вимірюваль-

ного експерименту виконуються серії досліджень. У результаті для кожного зразка з певними характеристиками отримуємо серію фотографій, що містить корисну інформативну складову специфічних характеристик зразка, а також стохастичну компоненту, обумовлену впливом випадкових зовнішніх факторів. Для отримання остаточного результату цифрової обробки зображень для вибірки в цілому необхідно обчислити точкові характеристики з можливістю їх усереднення у межах вихідної реалізації. Загальний інтервал яскравостей поділяємо на піддіапазони, для кожного з яких підраховується кількість пікселів.

Наступний етап обробки полягає в отриманні усереднених оцінок кількості пікселів у кожному піддіапазоні 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 доведено [6], що гіпотеза нормального розподілу результатів експериментальних досліджень, яка 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 розподілі усієї множини вихідних об'єктів у однорідні групи. На сьогодні розроблено цілий спектр методів кластерного аналізу [7]. Але для використання методів у прикладних галузях необхідне 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. У 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^l виступають кількісною оцінкою l -ї ознаки i -го об'єкта. У кожному i -му стовпчику $X_i = (x_i^1, x_i^2, \dots, x_i^m)$ записані параметри досліджуваного об'єкта. Геометричний сенс кожного об'єкта – це точка m -вимірного простору ознак $I^m(X)$.

Вихідні дані можуть бути представлені також на базі опису взаємних відстаней між об'єктами вихідної множини:

$$d_{n \times n} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}, \quad (2)$$

де d_{ij} – відстань (метрика) об'єктів x_i та x_j між собою. Іншою формою представлення є використання протилежної за змістом матриці, що характеризує близькість об'єктів. Перехід від матриці відстаней до матриці близькості виявляється елементарним, тому у спеціальній літературі вводиться одне загальне позначення функції близькості або віддаленості, а замість відповідних матриць використовується так звана матриця “об'єкт – об'єкт”.

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 таблиці 1. Кількість оброблених вибірок реальних даних за об'ємом значно перевищує представлений у таблиці 1 обсяг матеріалу, а тому виходить за рамки можливостей відображення у публікації.

Таблиця 1. Значення медіан у піддіапазонах g_i гістограми яскравості пікселів зображень газорозрядного випромінювання мазків крові

№ за п.	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

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

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

Якщо вихідні дані представлено у вигляді матриці “об’єкт – властивість” $X_{m \times n} = [x_i^l]$, $i=1, \dots, n$, $l=1, \dots, m$, то функціонал якості розподілу на класи визначається залежністю:

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

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

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

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

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

де γ – показник нечіткості класифікації.

Критерій якості кластеризації використовується з урахуванням залежностей:

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

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

Алгоритм, що оптимізує функціонал Дж. Беждека–Дж. Данна, називається методом нечітких c -середніх (fuzzy c -means), та являє собою параметричне сімейство за умри 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:

$$P^* = \arg \min_P \left\{ \begin{array}{l} Q_l(P) : P = (A^1, \dots, A^c), A^l = \mu_{1l}, \dots, \mu_{nl}, 0 \leq \mu_{ij} \leq 1; \\ \sum_{i=1}^n \mu_{ij} = 1, \sum_{i=1}^n \mu_{ij} > 0, i=1, \dots, n; j=1, \dots, c; \end{array} \right.$$

де c – кількість нечітких кластерів у шуканому розбитті P^* ; γ – 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:

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

Отримана геометрична інтерпретація задачі показана на рисунку 1.

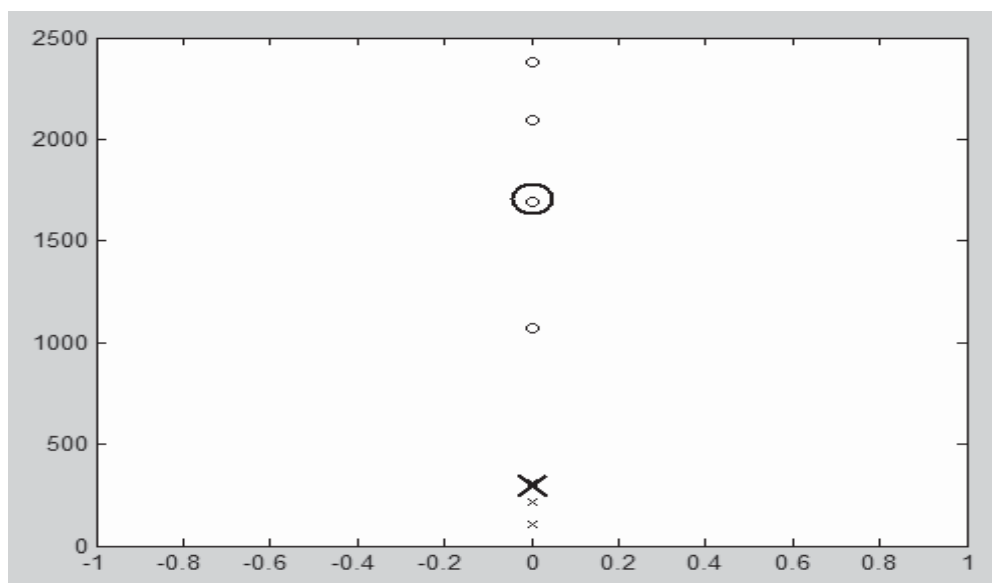


Рис. 1. Геометрична інтерпретація розв'язання задачі кластеризації зображень газорозрядного випромінювання.

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, є 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.

Обґрунтований вибір методології обробки зображень випромінювання біологічних об'єктів дозволив отримати задовільні результати у галузі скринінгової

діагностики ризику захворювань крові. Практична цінність запропонованої методики діагностики підтверджується на основі обробки експериментальних даних з використанням кластерного аналізу. Огляд отриманих результатів показує, що алгоритм нечіткої кластеризації "fuzzy c-means" забезпечив повністю вірне розбиття вихідних даних на групи «контроль» № 1, 2, 5, 6, що відповідають стану системи кровотворення здорової людини, та інший клас з № 3, 4, 7, 8, у який попали усі результати пацієнтів з підтвердженою низкою класичних методів діагностики патологією системи кровотворення.

Література

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. Кирлиан-фотография мазков крови при анемиях / Н. В. Глухова, Л. А. Песоцкая, Т. О. Опрятная // Академический журнал Западной Сибири Academic Journal of West Siberia – 2013 – Т. 9, № 5 (48). – С. 23–24.
5. Глухова Н. В. Анализ изображений кирлиановского свечения капель воды / Н. В. Глухова, Л. А. Песоцкая, В. Н. Лапицкий // Научный вестник Национального горничного университета. – №1. – 2013. – С. 91–96.
6. Мандель И. Д. Кластерный анализ / И. Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с.
7. Fern X. Z. Clustering ensembles for high dimensional data clustering / X. Z. Fern, C. E. Brodley // In Proc. International Conference on Machine Learning. – 2003. – P. 186–193.
8. Coordinate Descent and Clustering / J. C. Bezdek, R. J. Hathaway, R. E. Howard, C. A. Wilson // Control and Cybernetics. – Vol. 15. – P. 195–204.