

ІНТЕРВАЛЬНИЙ ТА КЛАСТЕРНИЙ АНАЛІЗ ОЗНАК ЗАХВОРЮВАНЬ НА ОСНОВІ МІКРОЕЛЕМЕНТНОГО ДОСЛІДЖЕННЯ ВОЛОСЯ

Дивак М.П.¹⁾, Довбака І.В.²⁾

Тернопільський національний економічний університет

¹⁾ д.т.н., професор; ²⁾ магістрант

І. Постановка проблеми

Одним із сучасних методів оцінювання ризику певних видів захворювань є мікроелементний аналіз волосся (МАВ) за допомогою абсорбційної атомної спектрометрії (АSА), індукційної спектрометрії з використанням концентрованої плазми (ІСАР), а також шляхом нейтронної активації (NAA) [1].

Такий метод дає можливість на основі встановлення вмісту певних мікроелементів та їх співвідношень встановити ризик захворювання, попередити його розвиток та допомогти у ефективному лікуванні та реабілітації [2,3]. У сучасній медичній діагностиці використовують методи кластерного аналізу ознак і на їх основі будують програмні системи підтримки прийняття рішень та класифікації типу захворювань. Разом з тим, на сьогодні відсутнє математичне забезпечення програмних систем, які встановлюють не тільки тип захворювань, але й ризик його виникнення та ступінь в залежності від вмісту мікроелементів та їх співвідношень в мікроелементному аналізі волосся (МАВ). Таке математичне забезпечення вимагає іншого підходу до кластерного аналізу ознак, оскільки ґрунтується на вибірці даних для різних пацієнтів і може відображати різну міру ризику та ступеня захворювання для однакових значень вмісту та співвідношень мікроелементів. У цьому випадку маємо неоднорідну вибірку даних первинних ознак і відповідно групи значень цих ознак формують певні інтервали ризику та можливого ступеня встановленого захворювання.

За цих умов математичне забезпечення програмної системи медичної діагностики на основі мікроелементного аналізу волосся доцільно побудувати із використанням кластерного та інтервального аналізу [4-7].

ІІ. Метод кластерного та інтервального аналізу ознак

Нехай для мікроелементного аналізу волосся використано $t=1, \dots, T$ ознак (кількість мікроелементів та їх відношень) і сформовано вибірку для N пацієнтів:

$$\{x_{1k}, \dots, x_{Tk}, k = 1, \dots, N\} \rightarrow \{[y_k], k = 1, \dots, N\}, \quad (1)$$

де $\{x_{1k}, \dots, x_{Tk}, k = 1, \dots, N\}$ $k = 1, \dots, N$ - набори первинних ознак для мікроелементного аналізу для k - того пацієнта; $[y_k], k = 1, \dots, N$ - нормовані до одиничного інтервалу інтервальні оцінки ступеня певного виду захворювання.

Оскільки первинні ознаки для мікроелементного аналізу є не однорідними, проведемо групування наборів цих ознак у вибірці (1) з метою виявлення подібних груп характеристик. Для цих цілей застосуємо метод «гірської» кластеризації, який не потребує складних обчислювальних процедур, а кількість кластерів регулюється лише за допомогою єдиного параметра – радіуса кластера [8]. З метою одержання якісних результатів кластеризації вибірку первинних ознак доцільно пронормувати.

На основі зазначеного методу кластеризації отримаємо P подібних груп із наборами центрів $\{\hat{x}_{1p}, \dots, \hat{x}_{Tp}, p = 1, \dots, P\}$ та заданим радіусом r . Відсортуємо у вибірці (1) P – груп даних за ознакою:

$$\sqrt{(\bar{x}_{1k} - \hat{x}_{1p})^2 + \dots + (\bar{x}_{Tk} - \hat{x}_{Tp})^2} \leq r, \quad p = 1, \dots, P \quad (2)$$

де $\{\bar{x}_{1k}, \dots, \bar{x}_{Tk}, k = 1, \dots, N\}$ - нормовані до «1» набори первинних ознак, r – деяка константа, задана емпірично для забезпечення якісних результатів кластеризації у межах $0,1 \leq r \leq 0,5$.

У підсумку, отримаємо P подібних груп первинних ознак вибірки даних (1):

$$\begin{aligned} \{\bar{x}_{1k}, \dots, \bar{x}_{Tk}, k = 1, \dots, N_1\} &\rightarrow \{[y_k], k = 1, \dots, N_1\}, \\ \{\bar{x}_{1k}, \dots, \bar{x}_{Tk}, k = N_{p-1}, \dots, N_p\} &\rightarrow \{[y_k], k = N_{p-1}, \dots, N_p\}, \dots, \\ \{\bar{x}_{1k}, \dots, \bar{x}_{Tk}, k = N_{p-1}, \dots, N_p\} &\rightarrow \{[y_k], k = N_{p-1}, \dots, N_p\} \end{aligned} \quad (3)$$

Тепер, користуючись отриманими наборами даних (3), сформуємо вибірку інтервальних даних нормованих ознак мікроелементного аналізу волосся у такий спосіб:

$$\{\hat{\bar{x}}_{11}, \dots, \hat{\bar{x}}_{T1}\} \rightarrow \left\{ \bigcup_{k=1}^{N_1} [y_k] \right\}, \{\hat{\bar{x}}_{1p}, \hat{\bar{x}}_{Tp}, p = 2, \dots, P\} \rightarrow \left\{ \bigcup_{k=N_{p-1}+1}^{N_p} [y_k] \right\}, p = 2, \dots, P. \quad (4)$$

На основі (4) отримуємо результуючі інтервали ступеня захворювань визначеного виду, що описують ознаки для мікроелементного аналізу:

$$\bigcup_{k=1}^{N_1} [y_k], \bigcup_{k=N_{p-1}+1}^{N_p} [y_k], p = 2, \dots, P \quad (5)$$

При цьому варто зазначити, що константу r необхідно інтеграційно підібрати таким чином, щоб результуючі інтервали (5) були неперервними:

$$\left(\bigcup_{k=1}^N [y_k] = [y_1^-; y_1^+] \cup [y_1^+; y_2^+] \cup \dots \cup [y_{p-1}^+; y_p^+] \right) \wedge \left(\bigcap_{k=1}^N [y_k] = \emptyset \right). \quad (6)$$

Із (6) формуємо вибірку вихідних експериментальних даних для задачі мікроелементного аналізу волосся:

$$\begin{pmatrix} \{\hat{\bar{x}}_{11}, \dots, \hat{\bar{x}}_{T1}\} \\ \{\hat{\bar{x}}_{12}, \dots, \hat{\bar{x}}_{T2}\} \\ \vdots \\ \{\hat{\bar{x}}_{1P}, \dots, \hat{\bar{x}}_{TP}\} \end{pmatrix} \rightarrow \begin{pmatrix} [y_1^-; y_1^+] \\ [y_2^-; y_2^+] \\ \vdots \\ [y_P^-; y_P^+] \end{pmatrix} \quad (7)$$

Побудовані інтервали (6) описують невизначеність ознак для мікроелементного аналізу, а це, в свою чергу, уможливило побудувати єдину математичну модель для класифікації ризику відповідного ступеня захворювання.

Далі у доповіді наведено приклад застосування розглянутого методу для класифікації ознак мікроелементного аналізу волосся для класифікації ступеня серцево-судинних захворювань. Наведено результати верифікації запропонованого методу кластерного та інтервального аналізу.

Висновки

1. Вперше запропоновано метод поєднання кластерного та інтервального аналізу для медичного діагностування на основі мікроелементного аналізу волосся людини.

2. Розглянуто приклад застосування запропонованого методу для встановлення ризику виникнення певного ступеня серцево-судинних захворювань. Проведено оцінку достовірності застосування методу.

Список використаних джерел

1. Скальный А.В. Микроэлементы для вашего здоровья/А.В.Скальный.-2-е изд. испр.и доп.-М.: Издательский дом «Оникс 21 век», 2004-320, (74)
2. Элементы в организме человека/ Электронный ресурс/ http://biomol.pl/biomol/ru_RU/html
3. Sławomir Puczkowski - Dlaczego przedmiotem badania analizy pierwiastkowej są włosy a nie krew?/електронний ресурс/ <https://www.youtube.com/watch?v=3bt0VSFPGjw>
4. Дивак М. П. Задачі математичного моделювання статичних систем із інтервальними даними / М. П. Дивак. - Тернопіль: Видавництво ТНЕУ «Економічна думка», 2011. – 216 с.
5. Дивак М. П. Ідентифікація параметрів моделі на основі аналізу інтервальних даних / М. П. Дивак, О. В. Крамар // Вісн. Тернопільського держ. технічн. унів. - Тернопіль, 1999. - Т. 4. – №1. - С.76–80.
6. Дивак М. П. Метод локалізації гарантованих оцінок в задачах параметричної ідентифікації / М. П. Дивак // Вимірювальна та обчислювальна техніка в технологічних процесах. - 2000. - №4. - С.12–17.
7. Дивак М. П. Метод ідентифікації вагових коефіцієнтів синаптичних зв'язків штучних нейронних мереж із радіально-базисними функціями на основі аналізу інтервальних даних / М. П. Дивак, Н.Я. Савка // Наукові праці Донецького національного технічного університету. Серія «Інформатика, кібернетика та обчислювальна техніка». - 2012. - Вип. 15 (203). - С.132-139.
8. Конюшенко В. В. Язык технических вычислений [Електронний ресурс]- Режим доступу - <http://www.mat.net.ua>