

УДК 681.518:004.93.1



А.С. Довбиш, Г.А. Стадник, Г.А. Босенко

Сумський державний університет, м. Суми, Україна, kras@id.sumdu.edu.ua

## ОПТИМІЗАЦІЯ СИСТЕМИ КОНТРОЛЬНИХ ДОПУСКІВ НА ДІАГНОСТИЧНІ ОЗНАКИ РОЗПІЗНАВАННЯ

Запропоновано інформаційно-екстремальний алгоритм навчання комп'ютеризованої системи діагностування з паралельно-послідовною оптимізацією системи контрольних допусків на ознаки розпізнавання. Реалізацію алгоритму розглянуто на прикладі оцінки ефективності схеми лікування гострої кишкової інфекції.

ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНИЙ АЛГОРИТМ, КОМП'ЮТЕРИЗОВАНА СИСТЕМА ДІАГНОСТУВАННЯ, НАВЧАННЯ, ОПТИМІЗАЦІЯ, ОЗНАКА РОЗПІЗНАВАННЯ, ФУНКЦІОНАЛЬНА ЕФЕКТИВНІСТЬ, ГОСТРА КИШКОВА ІНФЕКЦІЯ

### Вступ

Складність раннього діагностування при лікуванні інфекційних патологій обумовлена довільними початковими умовами патологічного процесу та перетином класів розпізнавання, які характеризують функціональний стан керованого процесу. Одним із перспективних шляхів підвищення достовірності та оперативності діагностування за умови багатовимірності простору ознак є створення комп'ютеризованої системи діагностування на базі здатної навчатися системи підтримки прийняття рішень (СППР). Відомі методи розпізнавання образів [1, 2] в основному носять модельний характер, оскільки вони не враховують перетин класів розпізнавання і довільні умови формування навчальних матриць, що робить неефективним їх використання у практичних задачах діагностування та лікування.

Для підвищення функціональної ефективності комп'ютеризованих систем діагностування і прогнозування перебігу та наслідків інфекційних патологій перспективним є використання ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології), яка ґрунтується на максимізації інформаційної спроможності СППР у процесі її навчання [3-5]. У праці [6] у рамках ІЕІ-технології розглядався алгоритм навчання СППР з паралельною оптимізацією контрольних допусків на діагностичні ознаки розпізнавання для визначення схеми лікування гострої кишкової інфекції (ГКІ) за результатами лабораторних досліджень. При цьому одержані результати фізичного моделювання не дозволили побудувати безпомилкові за навчальною матрицею вирішальні правила.

У статті розглядається алгоритм інформаційно-екстремального навчання діагностичної СППР для визначення схеми лікування ГКІ за результатами лабораторних досліджень шляхом паралельно-послідовної оптимізації контрольних допусків на діагностичні ознаки розпізнавання.

### 1. Постановка задачі

Розглянемо формалізовану постановку задачі інформаційного синтезу СППР у рамках ІЕІ-

технології. Дано алфавіт класів розпізнавання, що характеризують  $M$  функціональних станів патологічного процесу. Сформовано апріорно класифіковану навчальну матрицю типу «об'єкт-властивість»  $\|y_{m,i}^{(j)}\|$ ,  $i = \overline{1, N}$ ,  $j = \overline{1, n}$ , де  $N, n$  — кількість ознак розпізнавання та реалізацій образу відповідно. Крім того, дано вектор параметрів навчання СППР  $g = \langle x_m, d_m, \delta \rangle$ , де  $x_m$  — еталонний вектор-реалізація класу  $X_m^o$ ,  $d_m$  — радіус контейнера класу  $X_m^o$ , що відновлюється в радіальному базисі простору ознак розпізнавання і  $\delta$  — параметр поля контрольних допусків. При цьому задано такі обмеження:  $X_m$  — вектор, вершина якого визначає геометричний центр контейнера класу  $X_m^o$ ,  $d_m \in [0; d(x_m \oplus x_c) - 1]$ , де  $d(x_m \oplus x_c)$  — кодова відстань центру класу  $X_m^o$  від центру найближчого (сусіднього) до нього класу  $X_c$  і параметр поля контрольних допусків  $\delta \in [0; \delta_H / 2]$ , де  $\delta_H$  — нормоване (експлуатаційне) поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра контрольного поля допусків  $\delta$ .

Необхідно в процесі навчання оптимізувати координати вектора  $g$ , що забезпечують максимальне значення усередненого за алфавітом критерію функціональної ефективності (КФЕ) навчання системи розпізнавання:

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{\{k\}} E_m, \quad (1)$$

де  $E_m$  — інформаційний КФЕ навчання СППР розпізнавати реалізації класу  $X_m^o$ ;  $\{k\}$  — впорядкована множина кроків навчання (відновлення контейнерів класів розпізнавання).

На етапі екзамени необхідно з максимальною достовірністю прийняти рішення про належність реалізації образу, що розпізнається, одному із класів алфавіту.

### 2. Алгоритм навчання СППР

Розглянемо алгоритм навчання СППР в рамках ІЕІ-технології, в якому оптимізація контрольних допусків на ознаки розпізнавання спочатку здійснюється за паралельною, а потім — за послідовною

процедурами. При цьому паралельна оптимізація полягає в одночасній зміні контрольних допусків на всі ознаки розпізнавання, а послідовна – в по черговій зміні контрольних допусків для кожної ознаки при заданих допусках для наступних ознак. Алгоритм послідовної оптимізації контрольних допусків на ознаки розпізнавання у рамках ІЕІ-технології подамо у вигляді структурованої багаточиклічної ітераційної процедури:

$$\{\delta_{K,i}^*\} = \arg \left[ \bigotimes_{s=1}^S \max_{G_{\delta_i}} \{ \max_{G_E \cap G_d} \bar{E}^{(s)} \} \right], \quad i = \overline{1, N}, \quad (2)$$

де  $\bar{E}^{(s)}$  – усереднений КФЕ навчання СППР, обчислений на  $s$ -й ітерації послідовної процедури;  $G_{\delta_i}$ ,  $G_E$ ,  $G_d$  – області допустимих значень контрольних допусків для  $i$ -ї ознаки, критерію оптимізації і радіусів контейнерів відповідно;  $\otimes$  – символ операції повторення.

У процедурі (2) внутрішній цикл оптимізації на кожній  $s$ -й ітерації реалізує базовий алгоритм навчання СППР [3]. При цьому визначенні за результатами паралельної оптимізації квазіоптимальні контрольні допуски на ознаки розпізнавання приймаються для процедури послідовної оптимізації як стартові, що забезпечує знаходження значень інформаційного КФЕ в робочій області визначення його функції.

Вхідними даними для алгоритму паралельно-послідовної оптимізації контрольних допусків на ознаки розпізнавання є масив  $\{y_{m,i}^{(j)}\}$  – дійсна у загальному випадку багатовимірна навчальна матриця, а як робоча змінна – параметр поля контрольних допусків  $\delta$ , який дорівнює половині поля контрольних допусків і значення якого належать інтервалу  $[0; \delta_H / 2]$ .

Розглянемо схему алгоритму паралельно-послідовної оптимізації контрольних допусків на ознаки розпізнавання:

1. Обнуляється лічильник прогонів процедури оптимізації параметрів навчання  $s:=0$ .
2. Формується лічильник прогонів:  $s: s+1$ .
3. Обнуляється лічильник ознак розпізнавання:  $i:=0$ .
4. Формується лічильник ознак розпізнавання:  $i: i+1$ .
5. Обнуляється лічильник кроків зміни параметра  $\delta_i: l:=0$ .
6. Запускається лічильник:  $l:=l+1$  і обчислюються нижній  $A_{HK_i}[l]$  та верхній  $A_{BK_i}[l]$  контрольні допуски для  $i$ -тої ознаки розпізнавання

$$A_{HK_i}[l] = y_{1,i} - \delta_i \frac{\delta_{H_i}}{100}; \quad A_{BK_i}[l] = y_{1,i} + \delta_i \frac{\delta_{H_i}}{100}, \quad (3)$$

де  $y_{1,i}$  –  $i$ -та ознака еталонного вектора-реалізації  $y_1$  базового класу  $X_1^o$ , який характеризує найбільш бажаний для особи, що приймає рішення, функціональний стан СППР.

7. Формується бінарна навчальна матриця  $\|x_{m,i}^{(j)}\|$  за правилом

$$x_{m,i}^{(j)} = \begin{cases} 1, & \text{if } A_{HK_i}[l] < y_{m,i}^{(j)} < A_{BK_i}[l]. \\ 0, & \text{else.} \end{cases}$$

8. Формується множина  $\{x_m\}$  еталонних векторів-реалізацій класів розпізнавання за правилом

$$x_{m,i} = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_{m,i}^{(j)} > 0,5. \\ 0, & \text{else.} \end{cases}$$

9. Здійснюється розбиття множини  $\{x_m\}$  на пари найближчих еталонних векторів.

10. Для кожного класу розпізнавання відновлюється контейнер за схемою:

а) обнуляється лічильник класів розпізнавання:  $m:=0$ ;

б)  $m:=m+1$ ;

в) обнуляється лічильник кроків зміни радіуса контейнера:  $d:=0$ ;

г)  $d:=d+1$ ;

д) обчислюється значення інформаційного КФЕ (1);

е) якщо  $E_m \notin G_E$ , то виконується пункт 10г, інакше – пункт 10е;

є) виконується процедура пошуку глобального максимуму КФЕ (1) в робочій області визначення його функції;

ж)  $E_m^*[l] := \underset{\{d\}}{\text{extrem}} E_m[l, d]$ ;

з) визначається оптимальний радіус контейнера класу  $X_m^o: d_m^*[l] := \arg \underset{\{d\}}{\text{extrem}} E_m[l, d]$ .

11. Якщо  $m \leq M$ , то виконується пункт 10б, інакше – пункт 12.

12. Обчислюється за формулою (1) усереднений КФЕ  $\bar{E}[l]$ .

13. Якщо  $\delta_i[l] \leq \delta_{H_i} / 2$ , то виконується пункт 2, інакше – пункт 14.

14. Якщо  $\bar{E}[l] \notin G_E$ , то виконується пункт 15, інакше – пункт 6.

15. Виконується процедура пошуку глобального максимуму КФЕ  $\bar{E}[l]$  в робочій області визначення його функції.

17.  $\bar{E}^*[l] := \underset{\{s\}}{\text{extrem}} E_m[l]$ .

18. Визначається оптимальний параметр поля контрольних допусків на  $i$ -ту ознаку розпізнавання:  $\delta_i^* := \arg \bar{E}^*[l]$  і за формулою (3) обчислюються оптимальні контрольні допуски на  $i$ -ту ознаку розпізнавання

$$A_{HK_i}^* = y_{1,i} - \delta_i^* \frac{\delta_{H_i}}{100}; \quad A_{BK_i}^* = y_{1,i} + \delta_i^* \frac{\delta_{H_i}}{100}.$$

19. Якщо  $i \leq N$ , то виконується пункт 4, інакше – пункт 20.

20. Якщо на  $s$ -му прогоні ітераційної процедури послідовної оптимізації контрольних допусків на

ознаки розпізнавання різниці глобальних максимумів КФЕ відповідає умові  $|E_{\max}[s-1] - E_{\max}[s]| \leq \varepsilon$ , де  $\varepsilon$  – будь-яке мале позитивне число, то виконується пункт 21, інакше пункт 2.

21. ЗУПИН.

Як критерій оптимізації параметрів навчання у рамках ІЕІ-технології може розглядатися будь-яка статистична інформаційна міра, яка є функціоналом від точнісних характеристик. Наприклад, для двохальтернативних рішень і рівноймовірних гіпотез розглянемо модифіковану інформаційну міру Кульбака, в якій логарифмічне відношення правдоподібності має вигляд [4]:

$$E = \log_2 \frac{P_t^k}{P_f^k} = \log_2 \frac{p(\mu_m)p(\gamma_{1,k} / \mu_m) + p(\mu_{m+1})p(\gamma_{2,k} / \mu_{m+1})}{p(\mu_m)p(\gamma_{2,k} / \mu_m) + p(\mu_{m+1})p(\gamma_{1,k} / \mu_{m+1})}$$

де  $P_t^k$  – повна ймовірність правильного прийняття рішень про належність реалізацій класів  $X_m^o$  і  $X_{m+1}^o$  контейнеру  $K_{m,k}^o \in X_m^o$ , побудованому на  $k$ -му кроці навчання СППР розпізнавати реалізації класу  $X_m^o$ ;  $P_f^k$  – повна ймовірність помилкового прийняття рішень, гіпотези  $\gamma_{1,k}$ ,  $\gamma_{2,k}$  – апіорні основна і альтернативна гіпотези про належність контейнеру  $K_{m,k}^o$  реалізацій класів  $X_m^o$  і  $X_{m+1}^o$  відповідно;  $\mu_m, \mu_{m+1}$  – апостеріорні гіпотези про належність контейнеру  $K_{m,k}^o$  реалізацій класів  $X_m^o$  і  $X_{m+1}^o$  відповідно;  $p(\mu_m), p(\mu_{m+1})$  – безумовні ймовірності прийняття гіпотез  $\mu_m, \mu_{m+1}$  відповідно;  $p(\gamma_{l,k} / \mu)$  – апіорна умовна ймовірність прийняття гіпотези  $\gamma_{l,k}$  ( $l=1,2$ ) за умови, що має місце апостеріорна гіпотеза  $\mu$  ( $\mu_m$  або  $\mu_{m+1}$ ).

За умов апіорної невизначеності згідно з принципом Бернуллі-Лапласа приймемо допущення, що ймовірності  $p(\mu_m) = p(\mu_{m+1}) = 0,5$ . Оскільки  $p(\gamma_{1,k} / \mu_m)$  – перша достовірність  $D_{1,m}^{(k)}$ ;  $p(\gamma_{2,k} / \mu_{m+1})$  – друга достовірність  $D_{2,m}^{(k)}$ ;  $p(\gamma_{2,k} / \mu_m)$  – помилка першого роду  $\alpha_m^{(k)}$  і  $p(\gamma_{1,k} / \mu_{m+1})$  –  $\beta_m^{(k)}$  помилка другого роду, остаточно міра Кульбака набуває вигляду

$$E_m^{(k)} = 0,5 \log_2 \left( \frac{D_{1,m}^{(k)} + D_{2,m}^{(k)}}{\alpha_m^{(k)} + \beta_m^{(k)}} \right) * [(D_{1,m}^{(k)} + D_{2,m}^{(k)}) - (\alpha_m^{(k)} + \beta_m^{(k)})] = \log_2 \left( \frac{2 - (\alpha_m^{(k)} + \beta_m^{(k)})}{\alpha_m^{(k)} + \beta_m^{(k)}} \right) [1 - (\alpha_m^{(k)} + \beta_m^{(k)})]. \tag{4}$$

Таким чином, у рамках ІЕІ-технології алгоритм навчання діагностичної СППР з паралельно-послідовною оптимізацією контрольних допусків на ознаки розпізнавання полягає в ітераційній процедурі наближення інформаційного КФЕ до глобального максимуму, що обчислюється в робочій (допустимій) області визначення його функції.

3. Результати фізичного моделювання

Розглянемо застосування алгоритму навчання СППР з паралельно-послідовною оптимізацією контрольних допусків на ознаки розпізнавання для визначення схеми лікування ГКІ, викликаної умовно-патогенними мікроорганізмами за результатами лабораторних досліджень. Вхідний математичний опис СППР сформовано наступним чином: клас  $X_1^o$  характеризує контрольну групу осіб (здорові особи); клас  $X_2^o$  – групу пацієнтів, для яких необхідне комбіноване лікування з включенням до схеми колоїдного срібла (10 мг/л) і клас  $X_3^o$  – групу пацієнтів, для яких необхідне одночасне призначення пробіотика та колоїдного срібла на тлі базисної терапії. Навчальні матриці класів мали по 40 реалізацій, кожна з яких складалася з 19 ознак розпізнавання. При цьому структурована реалізація функціонального стану патологічного процесу складалася із значень наступних ознак розпізнавання:

- 1) лейкоцитарний індекс інтоксикації (ЛІІ);
- 2) швидкість осідання еритроцитів (ШОЕ) (мм/год);
- 3) кількість лейкоцитів ( $10^9$ /л);
- 4) гематологічний показник інтоксикації (ГПІ);
- 5) індекс зсуву лейкоцитів (ІЗЛК);
- 6) лімфоцитарний індекс (ЛІмф);
- 7) біфітобактерії (lg КУО/г);
- 8) лактобацили (lg КУО/г);
- 9) кишкова паличка зі слабо вираженими ферментними властивостями (lg КУО/г);
- 10) загальна кількість кишкової палички (lg КУО/г);
- 11) гемолізуюча кишкова паличка (lg КУО/г);
- 12) умовно патогенні ентеробактерії (УПЕ) (lg КУО/г);
- 13) золотистий стафілокок (lg КУО/г);
- 14) гемолізуючий стрептокок (lg КУО/г);
- 15) гриби роду Candida (lg КУО/г);
- 16) секреторний імуноглобулін А (пг/л);
- 17) інтерлейкін 1 бета (пг/л);
- 18) інтерлейкін 4 (мг/л);
- 19) патологічні мікроби сімейства «Кишкові».

На рис. 1 наведено залежність усередненого за алфавітом класів розпізнавання КФЕ від кількості ітерацій (прогонів)  $s$  алгоритму навчання СППР.

Аналіз рис. 1 показує, що вже на третьому прогоні алгоритму послідовної оптимізації отримано максимальне значення усередненого значення КФЕ, яке дорівнює  $\bar{E}^* = 6,54$ .

На рис. 2 наведено результати паралельно-послідовної оптимізації контрольних допусків на ознаки розпізнавання.

Аналіз рис. 2 показує, що перша ознака (ЛІІ), четверта (ГПІ), дев'ята (кишкова паличка зі слабо

вираженими ферментними властивостями), одинадцята (гемолізувальна кишкова паличка), дванадцята (УПЕ), тринадцята (золотистий стафілокок), чотирнадцята (гемолізувальний стрептокок) та п'ятнадцята (гриби роду *Candida*) є неінформативними. Це пояснюється тим, що при заданому алфавіті класів розпізнавання їх вилучення із структурованого вектора ознак не змінює максимальне усереднене значення КФЕ.

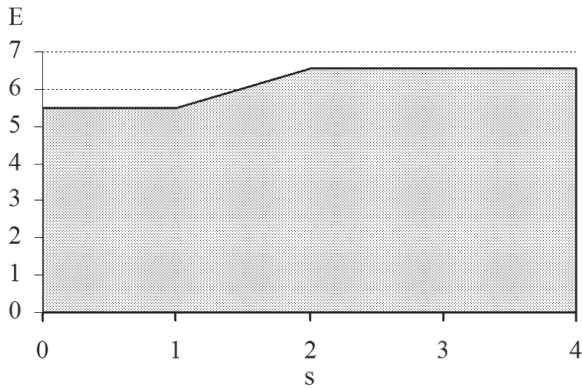


Рис. 1. Графік залежності критерію Кульбака від кількості ітерацій алгоритму навчання СППР

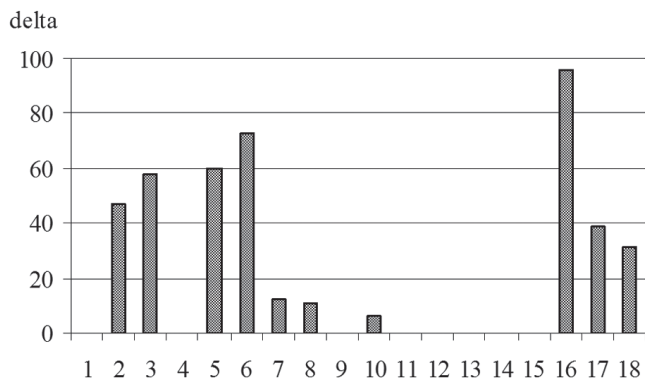


Рис. 2. Система оптимальних контрольних допусків на ознаки розпізнавання

На рис. 3 показано графіки залежності КФЕ (4) від радіусів гіперсферичних контейнерів класів розпізнавання, одержаних при застосуванні системи оптимальних контрольних допусків на ознаки розпізнавання, одержаної за алгоритмом (2). Тут темні ділянки графіків позначають робочі (допустимі) області, в яких значення достовірностей перевищують відповідні помилки першого та другого роду.

Аналіз рис. 3 показує, оптимальні радіуси контейнерів класів розпізнавання дорівнюють: для класу  $X_1^o$  –  $d_1^* = 2$  (тут і далі у кодових одиницях), для класу  $X_2^o$  –  $d_2^* = 3$  і для класу  $X_3^o$  –  $d_3^* = 3$ , а міжцентрові відстані між парами найближчих класів –  $d(x_1 \oplus x_2) = 4$ ,  $d(x_2 \oplus x_1) = 4$  та  $d(x_3 \oplus x_2) = 5$ .

У режимі екзамену рішення приймалося шляхом визначення максимального значення геометричної функції належності, яка для гіперсферичного класифікатора і реалізацій класу  $X_m^o$  має вигляд:

$$\mu_m = 1 - \frac{d(x_m^* \oplus x_e)}{d_m^*},$$

де  $x_e$  – вектор-реалізація образу, що розпізнається.

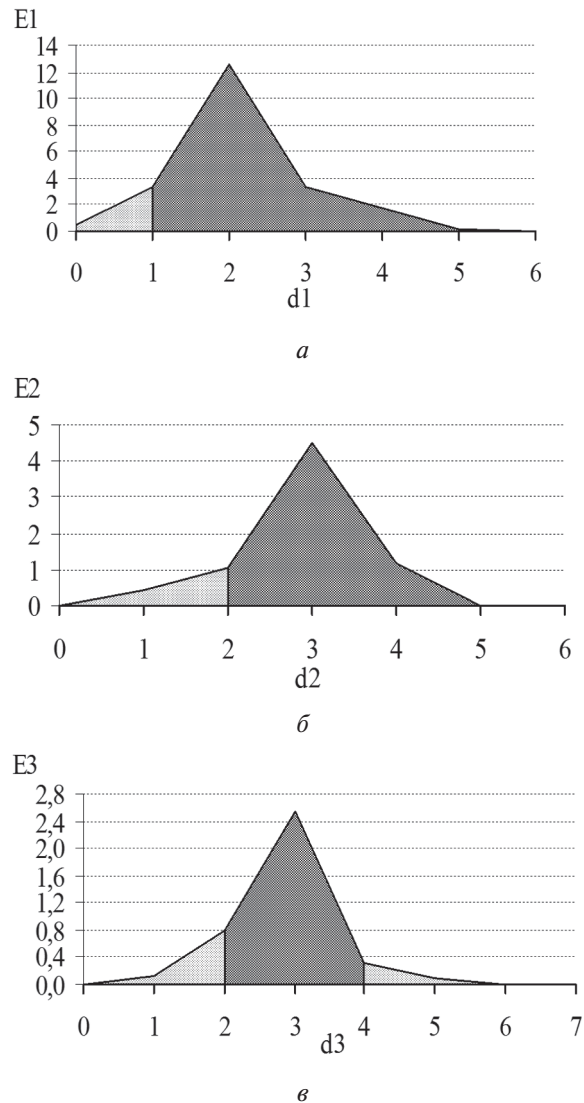


Рис. 3. Графіки залежності критерію Кульбака від радіусів контейнерів класів розпізнавання: а – клас  $X_1^o$ ; б – клас  $X_2^o$ ; в – клас  $X_3^o$

За результатами фізичного моделювання в режимі екзамену СППР підтверджено працездатність та надійність розробленого інформаційного та програмного забезпечення.

### Висновки

1. Запропоновано інформаційне та програмне забезпечення здатної навчатися СППР для визначення схеми лікування ГКІ за апріорно класифікованою навчальною матрицею у рамках ІЕІ-технології.

2. Досліджено вплив системи контрольних допусків на ознаки розпізнавання на функціональну ефективність навчання СППР і визначено оптимальні контрольні допуски, що дозволило підвищити достовірність діагностування.

3. З метою підвищення достовірності функціонування СППР перспективними є оптимізація інших параметрів функціонування СППР та перехід на ієрархічну структуру алгоритмів навчання та екзамону, що дозволить збільшити потужність алфавіту класів розпізнавання.

**Список літератури:** 1. *Васильев В.И.* Распознающие системы. Справочник. 2-е изд., перераб. и доп. [Текст] / В.И. Васильев – Киев: Наукова думка, 1983. – 422 с. 2. *Айвазян С. А.* Прикладная статистика: Классификация и снижение размерности. Справ. изд. [Текст] / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин / Под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с. 3. *Довбиш А. С.* Основи проектування інтелектуальних систем: Навчальний посібник [Текст] / А. С. Довбиш – Суми: Видавництво СумДУ, 2009. – 171 с. 4. *Довбиш А. С.* Оптимізація контрольних допусків на ознаки розпізнавання в інформаційно-екстремальних методах автоматичної класифікації [Текст] / А.С. Довбиш, М.В. Козинець, С.М. Котенко // Вісник Сумського державного університету. Серія «Техніка». – 2007. – № 1. – С. 169-178. 5. *Довбиш А. С., Руденко М. С.* Оптимізація параметрів навчання СППР для діагностування онкопатологій [Текст] / А. С. Довбиш, М. С. Руденко // Вісник СумДУ. Серія «Технічні науки». – 2010. – №2. – С.48-55. 6. *Довбиш А. С.* Система підтримки прийняття рішень для визначення схеми лікування гострої кишкової інфекції [Текст] / А.С. Довбиш, Г.А. Стадник, К.С. Полов'ян // Вісник Сумського державного університету. Серія «Технічні науки». – 2012. – №1. – С. 20-26.

*Надійшла до редколегії 16.05.2012*

УДК 681.518:004.93.1'

**Параллельно-последовательная оптимизация системы контрольных допусков на признаки распознавания / А.А. Стадник // Бионика интеллекта: науч.-техн. журнал. – 2012. – № 2 (79). – С. 38–42.**

В статье рассматривается метод определения схемы лечения острой кишечной инфекции в рамках информационно-экстремальной интеллектуальной технологии, основанной на максимизации информационной способности СППР в процессе ее обучения. Предложено информационное и программное обеспечение способной обучаться СППР с параллельно-последовательной оптимизацией системы контрольных допусков на признаки распознавания.

Ил. 3. Библиогр.: 6 назв.

UDK 681.518:004.93.1'

**Parallel-sequential optimization of the control tolerances for recognition attributes / H.A. Stadnyk // Bionics of Intelligence: Sci. Mag. – 2012. – № 2 (79). – P. 38–42.**

The article presents determination of acute enteric infection treatment regimen within the bounds of information-extreme intellectual technology, which is based on maximization of informational capability of recognition system in the process of its learning. Dataware and software of recognition system which is capable of learning with the parallel-sequential optimization of the control tolerances for recognition attributes are presented.

Fig.: 3. Ref.: 6 items.