

*А.С. Довбиш, д-р техн. наук, професор;  
М.С. Руденко, студент,  
Сумський державний університет, м. Суми*

*Розглядається синтез інформаційного і програмного забезпечення системи підтримки прийняття рішень для діагностування онкопатологій у рамках інформаційно-екстремальної інтелектуальної технології шляхом аналізу і синтезу морфологічних зображень тканин пацієнтів, отриманих за результатами біопсії.*

***Ключові слова:** система підтримки прийняття рішень, діагностування, оптимізація, навчання, критерій функціональної ефективності.*

*Рассматривается синтез информационного и программного обеспечения системы поддержки принятия решений для диагностирования онкопатологий в рамках информационно-экстремальной интеллектуальной технологии путем анализа морфологических изображений тканей пациента, полученных по результатам биопсии.*

***Ключевые слова:** система поддержки принятия решений, диагностирование, оптимизация, обучение, критерий функциональной эффективности.*

## ВСТУП

Проблема ранньої діагностики раку – одна з ключових в онкології, від вирішення якої залежать ефективність лікування, якість і тривалість життя онкологічних хворих. Один із шляхів підвищення достовірності діагностування пов'язаний із широким застосуванням в онкології комп'ютеризованих діагностичних систем та створенням на їх основі інтелектуальних систем підтримки прийняття рішень (СППР), здатних аналізувати великі обсяги діагностичної інформації в реальному темпі часу.

Сучасний розвиток теорії аналізу і синтезу інтелектуальних СППР ґрунтується на ідеях і методах розпізнавання образів і статистичної теорії прийняття рішень [1-3]. При цьому відомі алгоритми розпізнавання [4-6] в основному спрямовані на розв'язання модельних задач, які ігнорують апріорно нечітке розбиття класів розпізнавання. Для вирішення цієї проблеми перспективним є застосування ідей та методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технологія), що ґрунтується на максимізації інформаційної спроможності СППР шляхом введення додаткових обмежень у процесі оптимізації просторово-часових параметрів функціонування [7,8]. У працях [9,10] запропоновано інформаційно-екстремальні алгоритми розпізнавання морфологічних зображень тканин пацієнтів, одержаних за результатами біопсії. При цьому як параметри навчання розглядалися геометричні параметри контейнерів класів розпізнавання та контрольні допуски на ознаки розпізнавання, що не дозволило побудувати безпомилкові за навчальною матрицею вирішальні правила.

У статті досліджується можливість підвищення достовірності розпізнавання морфологічних зображень шляхом додаткової оптимізації в процесі навчання СППР кроку квантування за часом вхідних реалізацій образів.

## ПОСТАНОВКА ЗАДАЧІ СИНТЕЗУ СППР, ЩО НАВЧАЄТЬСЯ

Задано апріорно нечіткий алфавіт класів розпізнавання  $\{X_m^o \mid m = \overline{1, M}\}$ , який складається із морфологічних зображень тканин пацієнтів, одержаних за результатами методу біопсії. Відома навчальна матриця яскравості зображень  $\|y_{m,i}^{(j)}\|, i = \overline{1, N}, j = \overline{1, n}$ , де  $N, n$  – кількість ознак розпізнавання і випробувань відповідно. Рядок матриці  $\{y_{m,i}^{(j)} \mid i = \overline{1, N}\}$  утворює  $j$ -ту реалізацію образу, а стовпчик  $\{y_{m,i}^{(j)} \mid j = \overline{1, n}\}$  – навчальну вибірку з генеральної сукупності значень  $i$ -ї ознаки розпізнавання.

Потрібно на етапі навчання для структурованого вектора параметрів навчання СППР  $g = \langle x_m, d_m, \delta, \tau \rangle$ , де  $x_m$  – еталонний вектор-реалізація, вершина якого визначає геометричний центр контейнера класу  $X_m^o$ ;  $d_m$  – радіус контейнера класу  $X_m^o$ , що відновлюється в радіальному базисі простору ознак розпізнавання;  $\delta$  – параметр поля контрольних допусків на ознаки розпізнавання;  $\tau$  – крок квантування за часом вхідної реалізації класу, знайти оптимальні значення його координат, що забезпечують максимум усередненого за алфавітом класів розпізнавання інформаційного критерію функціональної ефективності (КФЕ) навчання СППР

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

де  $E_m$  – інформаційний КФЕ навчання системи розпізнавання реалізації класу  $X_m^o$ ;  $G_E$  – робоча (допустима) область визначення функції КФЕ, в якій здійснюється пошук її глобального максимуму.

При цьому задано такі обмеження на параметр навчання:  $x_m$  – вектор,  $d_m \in [0; d(x_m \oplus x_c) - 1]$ , де  $d(x_m \oplus x_c)$  – кодова відстань центра класу  $X_m^o$  від центра найближчого (сусіднього) до нього класу  $X_c$ ,  $\delta \in [0; \delta_H / 2]$ , де  $\delta_H$  – нормоване (експлуатаційне) поле допусків на ознаку розпізнавання, і  $\tau \leq \Delta t = \frac{\pi}{\omega_{zp}}$ , де  $\omega_{zp}$  – гранична частота спектра

сигналу на вході СППР.

У режимі екзамену СППР повинна приймати рішення про належність реалізації образу, що розпізнається, до відповідного класу зі сформованого на етапі навчання алфавіту.

## МАТЕМАТИЧНА МОДЕЛЬ НАВЧАННЯ СППР

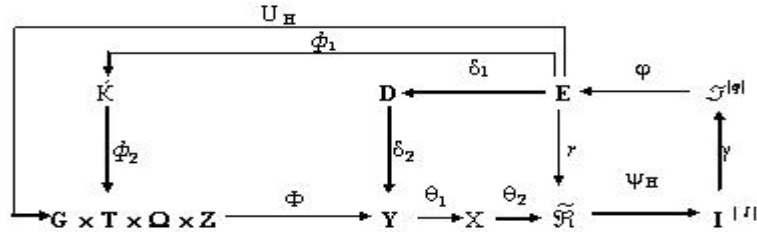
Математична модель інформаційно-екстремального навчання СППР як обов'язкову складову частину включає вхідний математичний опис, який подамо у вигляді теоретико-множинної структури:

$$\Delta = \langle G, T, \Omega, Z, \Theta, Y, X; \Phi_1, \Phi_2, \Phi_3 \rangle,$$

де  $G$  – простір факторів, що впливають на СППР;  $T$  – множина моментів часу зняття інформації;  $\Omega$  – простір ознак розпізнавання;  $Z$  – простір можливих функціональних станів СППР;  $\Theta$  – простір вхідних реалізацій образу;  $Y$  – вхідна вибіркова множина (вхідна навчальна матриця яскравості зображень);  $X$  – бінарна навчальна матриця яскравості;

$\Phi_1 : G \times T \times \Omega \times Z \rightarrow \Theta$  – оператор формування вхідних реалізацій образів;  
 $\Phi_2 : \Theta \rightarrow Y$  – оператор формування вибіркової множини  $Y$ ;  $\Phi_3 : Y \rightarrow X$  – оператор формування бінарної навчальної матриці  $X$ .

Категорійна модель у вигляді діаграми відображень множин, що застосовуються у процесі оптимізації кроку квантування за часом вхідних реалізацій, показана на рис. 1.



1 –

У діаграмі (рис. 1) оператор  $\theta : X \rightarrow \tilde{\mathfrak{R}}^{|M|}$  будує розбиття  $\tilde{\mathfrak{R}}^{|M|}$  простору ознак на класи розпізнавання, яке у загальному випадку є нечітким, а оператор класифікації  $\Psi : \tilde{\mathfrak{R}}^{|M|} \rightarrow I^{|I|}$  перевіряє основну статистичну гіпотезу про належність реалізацій  $\{x_m^{(j)} | j = \overline{1, n}\}$  класу  $X_m^o$  і формує множину гіпотез  $I^{|I|}$ , де  $I$  – кількість статистичних гіпотез. Оператор  $\gamma : I^{|I|} \rightarrow \mathfrak{Z}^{|q|}$  шляхом оцінки статистичних гіпотез формує множину точнісних характеристик  $\mathfrak{Z}^{|q|}$ , де  $q = I^2$  – кількість точнісних характеристик. Оператор  $\varphi : \mathfrak{Z}^{|q|} \rightarrow E$  обчислює множину значень інформаційного КФЕ, який є функціоналом точнісних характеристик. Контур діаграми, який складається з операторів  $\Psi, \gamma, \varphi$  і  $r$ , оптимізує геометричні параметри розбиття  $\tilde{\mathfrak{R}}^{|M|}$  шляхом пошуку глобального максимуму КФЕ в робочій (допустимій) області визначення його функції. Оптимізація системи контрольних допусків здійснюється за ітераційною процедурою, в якій задіяно оператори  $\theta, \psi, \gamma, \varphi, \delta_1$  і  $\delta_2$ . Контур оптимізації кроку квантування за часом вхідних реалізацій замикається операторами  $\tau_1 : E \rightarrow T$ , де  $T$  – терм-множина кроків квантування і  $\tau_2 : T \rightarrow G \times T \times \Omega \times Z$ , який змінює інтервали часу зняття інформації.

Практична цінність категорійної моделі (рис. 1) в рамках евристичного підходу до аналізу і синтезу інтелектуальних СППР полягає в суттєвому спрощенні розроблення алгоритму їх навчання.

### АЛГОРИТМ НАВЧАННЯ

Оптимізація кроку квантування за часом вхідних реалізацій образу в рамках інформаційно-екстремального алгоритму навчання згідно з діаграмою відображень (рис. 1) здійснюється за трициклічним ітераційним алгоритмом пошуку глобального максимуму КФЕ в робочій області визначення його функції

$$\tau^* = \arg \max_{G_t} \{ \max_{G_s} \{ \max_{G_j} E \} \}, \quad (2)$$

де  $G_\tau$  – допустима область значень параметра функціонування  $\tau$ ;  
 $G_\delta$  – допустима область значень параметра поля допусків  $\delta$ ;  
 $G_d$  – допустима область значень параметра функціонування  $d_1$ .

З метою підвищення оперативності й точності обчислення інформаційного КФЕ навчання СППР розглянемо алгоритм паралельно-послідовної оптимізації системи контрольних допусків на ознаки розпізнавання. При цьому попередня реалізація алгоритму паралельної оптимізації дозволяє знайти стартові контрольні допуски на ознаки розпізнавання для запуску алгоритму послідовної оптимізації. Такий підхід дозволяє суттєво підвищити оперативність алгоритму послідовної оптимізації, оскільки визначені за алгоритмом паралельної оптимізації стартові квазіоптимальні допуски вже гарантують обчислення значень КФЕ в робочій області визначення його функції.

Вхідною інформацією для алгоритму оптимізації кроку (3) квантування функцій яскравості зображень є множина, яка утворює алфавіт класів розпізнавання  $\{X_m^o \mid m = \overline{1, M}\}$ . Наведемо етапи реалізації інформаційно-екстремального алгоритму навчання СППР з оптимізацією кроку квантування функцій яскравості морфологічних зображень тканин пацієнтів на базі алгоритму паралельної оптимізації системи контрольних допусків на ознаки розпізнавання.

1. Обробка вхідних зображень у полярній системі координат з метою забезпечення інваріантності алгоритму навчання до зсуву, повороту та масштабу.

2. Побудова вхідної реалізації зображення у вигляді усередненої функції яскравості від радіуса рецепторного поля.

3. Обнуління лічильника кроків квантування функції яскравості за радіусом рецепторного поля вхідної реалізації образу:  $\tau := 0$ .

4. Запуск лічильника:  $\tau := \tau + 1$ .

5. Формування вхідної цілої навчальної матриці  $\|y_{m,i}^{(j)}\|$ , яка складається з дискрет усереднених функцій яскравості від радіуса рецепторного поля для RGB-складових зображень.

6. Формування еталонного вектора  $y_1 \in X_1^0$  шляхом статистичного усереднення значень ознак у стовпчиках матриці  $\|y_{1,i}^{(j)}\|$ .

7. Реалізація алгоритму паралельної оптимізації системи контрольних допусків на ознаки розпізнавання [11].

8. Порівняння: якщо поточне значення критерію (1) дорівнює його глобальному максимуму  $E(\tau) = E^*$  за умови  $E(\tau) \in G_E$ , то виконується пункт 9, інакше – пункт 4.

9.  $\tau := \tau^*$ , де  $\tau^*$  – оптимальний крок квантування функції яскравості RGB-складової, і ЗУПИН.

Як критерій оптимізації параметрів функціонування системи розпізнавання може бути використано будь-який статистичний інформаційний критерій, який є функціоналом від точнісних характеристик. Наприклад, модифікований інформаційний критерій Кульбака має вигляд [8]

$$E_m^{(k)} = \log_2 \frac{2 - (\alpha^{(k)} + \beta^{(k)})}{\alpha^{(k)} + \beta^{(k)}} \cdot [1 - (\alpha^{(k)} + \beta^{(k)})], \quad (3)$$

де  $\alpha^{(k)}$ ,  $\beta^{(k)}$  – помилки першого та другого родів відповідно, обчислені на  $k$ -му кроці навчання.

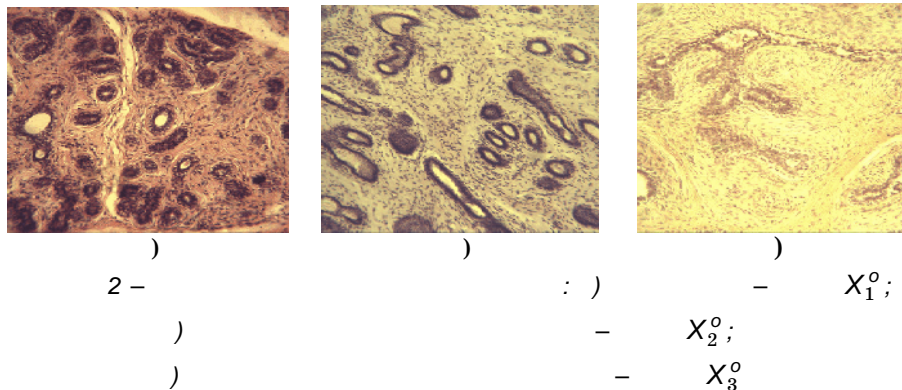
### РЕЗУЛЬТАТИ ФІЗИЧНОГО МОДЕЛЮВАННЯ

Вищенаведений алгоритм навчання СППР було реалізовано для діагностування раку молочної залози шляхом розпізнавання морфологічних зображень тканин пацієнтів, отриманих за методом біопсії. Як багатовимірна навчальна матриця  $\|y_{m,i}^{(j)}\|$  розглядалася матриця яскравості зображень, одержана шляхом зчитування значення яскравості в кожному пікселі рецепторного поля. Кожне зображення оброблялося в полярних координатах з метою забезпечення інваріантності алгоритму розпізнавання до зсуву, повороту та масштабу. Кожний рядок навчальної матриці складається з вектора-реалізації зображення, значення  $i$ -ї координати якого утворюється шляхом усереднення значень яскравості  $i$ -их пікселів відповідних складових RGB- складових за формулою

$$\Theta_{i,j} = \frac{1}{N_j} \sum_{l=1}^{N_j} \theta_{i,j,l}, \quad (4)$$

де  $\Theta_{i,j}$  – числове значення  $i$ -ї координати ( $i = \overline{0, R}$ )  $j$ -го вектора-реалізації;  $j = \overline{0, R}$ ;  $\theta_{i,j,l}$  – значення яскравості кольорової складової в  $l$ -му пікселі  $j$ -го кола зчитування яскравості;  $N_j$  – загальна кількість пікселів в  $j$ -му колі зчитування. Тут  $R$  – радіус кола зчитування яскравості RGB- складових зображень.

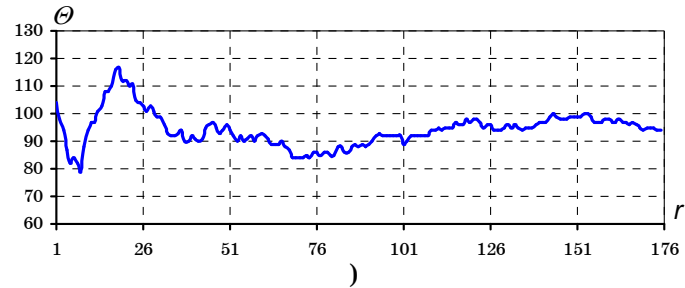
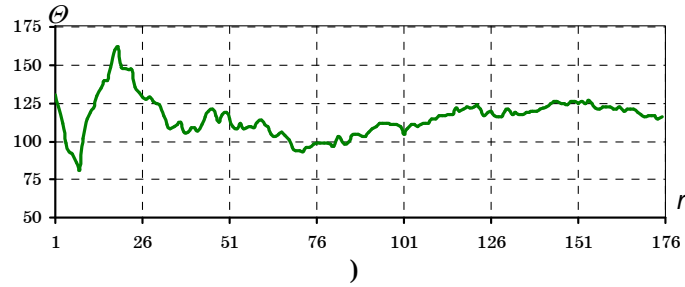
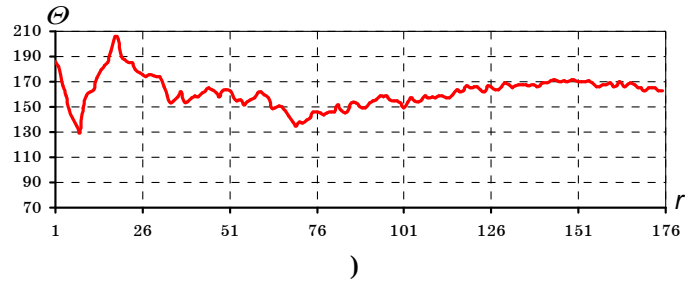
При формуванні навчальної матриці для трьох класів використовувалося 90 зображень розміром  $351 \times 351$  піксель. Як вхідні дані було використано морфологічні зображення трьох класів захворювань, які було отримано за методом біопсії при діагностуванні пацієнтів на рак молочної залози (рис. 2).



У результаті оброблення зображень у полярних координатах за формулою (4) було сформовано функції яскравості для трьох RGB-складових, які розглядалися як реалізації відповідних образів. На рис. 3 наведено такі функції яскравості залежно від радіуса поляризації  $r$ , одержані після оброблення зображення базового класу  $X_1^0$  (мазоплазія).

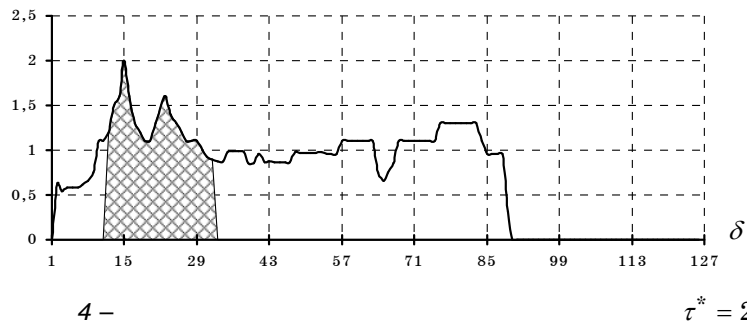
З метою оптимізації кроку квантування функцій яскравості RGB-складових, який визначає оптимальне в інформаційному розумінні

рецепторне поле, було реалізовано алгоритм навчання (2) з паралельною оптимізацією контрольних допусків на ознаки розпізнавання. При цьому крок квантування змінювався від 0 до 5 пікселів.



3 –  $X_1^0$  ; )

У результаті найбільше значення КФЕ, який обчислювався на кожному кроці навчання за формулою (3), було одержано для кроку квантування  $\tau^* = 2$  пікселі. На рис. 4 показано графік залежності критерію Кульбака від параметра поля допусків на ознаки при реалізації алгоритму навчання з паралельною оптимізацією контрольних допусків при  $\tau^* = 2$  пікселі.



4 –  $\tau^* = 2$

На рис. 4 заштрихована ділянка позначає робочу (допустиму) область визначення функції критерію (3). Аналіз рис. 4 показує, що при значенні кроку квантування  $\tau^* = 2$  оптимальне значення поля контрольних допусків дорівнює  $\delta^* = \pm 15$  градацій яскравості.

З метою збільшення значення інформаційного КФЕ навчання було реалізовано ітераційний алгоритм (2) з паралельно-послідовною оптимізацією контрольних допусків на ознаки розпізнавання при  $\tau^* = 2$  пікселі. При цьому одержані за алгоритмом паралельної оптимізації контрольні допуски розглядалися як стартові для алгоритму послідовної оптимізації. Результати реалізації алгоритму паралельно-послідовної оптимізації контрольних допусків на ознаки розпізнавання наведено в табл. 1.

1–

№ ітер.	Клас №1				Клас №2				Клас №3			
	E	D1	D2	d*	E	D1	D2	d*	E	D1	D2	d*
1.	2,9	0,83	1	45	3,3	0,87	1	44	4,5	0,93	1	2
2.	5,7	0,97	1	34	3,3	0,87	1	33	4,5	0,93	1	1
3.	<b>5,7</b>	<b>0,97</b>	<b>1</b>	<b>32</b>	<b>5,7</b>	<b>0,97</b>	<b>1</b>	<b>28</b>	<b>4,5</b>	<b>0,93</b>	<b>1</b>	<b>1</b>
4.	5,7	0,97	1	32	5,7	0,97	1	28	4,5	0,93	1	1

Аналіз табл. 1 показує, що одержане вже на третьому прогоні процедури послідовної оптимізації контрольних допусків максимальне усереднене значення КФЕ ( $E^* = 5,3$ ) суттєво перевищує його значення, одержане при реалізації алгоритму паралельної оптимізації, яке, як видно з рис. 4, дорівнює  $E^* = 2,0$ .

## ВИСНОВКИ

1. За результатами фізичного моделювання доведено, що крок квантування функцій яскравості RGB-складових морфологічних зображень тканин пухлин суттєво впливає на функціональну ефективність навчання СППР для діагностування онкопатологій.

2. Запропонований алгоритм оптимізації кроку квантування функцій яскравості RGB-складових морфологічних зображень досягає максимального значення інформаційного КФЕ навчання СППР за умови реалізації паралельно-послідовного алгоритму визначення контрольних допусків на ознаки розпізнавання.

## SUMMARY

### OPTIMIZATION PARAMETERS OF LEARNING INTELLIGENCE DECISION SUPPORT SYSTEM FOR ONCOPATHOLOGY DIAGNOSIS

*A.S. Dovbysh, M.S. Rudenko,  
Sumy State University, Sumy*

*The article presents synthesis of dataware and software of intelligence decision support system for oncopathology diagnosis, in bounds of information-extreme intelligence technology, on basis of analysis and synthesis of morphological tissue images of patient, gained by biopsy method.*

*Keywords: intelligence decision support system, diagnosis, optimization, learning, functional efficiency criteria.*

1. Фукунага К. Введение в статистическую теорию распознавания образов.: пер. с англ. / К. Фукунагаю – М.: Наука. Главная редакция физико-математической литературы, 1979. – 368 с.
2. Vapnik V. Statistical Learning Theory/ V. Vapnik. - New York: John Wiley&Sons, 1998. – 732 p.
3. Фомин Я.А. Статистическая теория распознавания образов / Я.А. Фомин, Г.Р. Тарловский.- М.: Радио и связь, 1986.– 264 с.
4. Duda R.O. Pattern Classification, second ed./ R.O. Duda, P.E. Hart, D.G. Stork. - New York: John Wiley & Sons, 2001. – 738 p.
5. Shalkoff R.J. Digital image processing and computer vision/ R.J. Shalkoff. – New York-Chichester-Brisbane-Toronto-Singapore: John Wiley & Sons, Inc., 1989. – 489 p.
6. Васильев В.И. Распознающие системы: справочник. 2-е изд., перераб. и доп. / В.И. Васильев. – К.: Наукова думка, 1983. – 422 с.
7. Довбиш А.С. Основы проектування інтелектуальних систем: навчальний посібник / А.С. Довбиш. – Суми: Видавництво СумДУ, 2009.– 171 с.
8. Краснопопосовський А.С. Інформаційний синтез інтелектуальних систем керування: Підхід, що ґрунтується на методі функціонально-статистичних випробувань / А.С. Краснопопосовський. – Суми: Видавництво СумДУ, 2004. – 261 с.
9. Довбиш А.С. Інтелектуальна система діагностування онкопатологій / А.С. Довбиш, О.П. Чекалов, С.С. Мартиненко // Радіоелектронні і комп'ютерні системи. – 2009.– №3(37).–С.92-96.
10. Довбиш А.С. Інтелектуальна система підтримки прийняття рішень для діагностування онкопатологій / А.С. Довбиш, В.О. Жаловага // Вісник СумДУ. Серія Технічні науки.– 2009.–№3.– С.5-12.
11. Довбиш А.С. Оптимізація контрольних допусків на ознаки розпізнавання в інформаційно-екстремальних методах автоматичної класифікації / А.С. Довбиш, М.В. Козинець, С.М. Котенко // Вісник Сумського державного університету. Серія Технічні науки. - 2007. –№1.– С. 169-178.