

А.С. ДОВБИШ

Сумський державний університет, Суми, Україна, e-mail: a.dovbysh@cs.sumdu.edu.ua.

І.В. ШЕЛЕХОВ

Сумський державний університет, Суми, Україна, e-mail: i.shelehov@cs.sumdu.edu.ua.

А.М. РОМАНЮК

Сумський державний університет, Суми, Україна, e-mail: a.romanjuk@med.sumdu.edu.ua.

Р.А. МОСКАЛЕНКО

Сумський державний університет, Суми, Україна, e-mail: r.moskalenko@med.sumdu.edu.ua.

Т.Р. САВЧЕНКО

Сумський державний університет, Суми, Україна, e-mail: taras.savchenko01@gmail.com.

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ДІАГНОСТУВАННЯ ОНКОПАТОЛОГІЙ МОЛОЧНОЇ ЗАЛОЗИ ЗА ГІСТОЛОГІЧНИМИ ЗОБРАЖЕННЯМИ

Анотація. Запропоновано метод інформаційно-екстремального машинного навчання системи підтримки прийняття рішень для діагностування онкопатологій молочної залози за гістологічними зображеннями. Цей метод на відміну від відомих методів, включаючи нейроподібні структури, розроблено у межах функціонального підходу до моделювання когнітивних процесів формування і прийняття рішень природним інтелектом. Водночас побудовані у межах геометричного підходу вирішальні правила є практично інваріантними до багатовимірності простору діагностичних ознак. Розроблений метод дає змогу створити інформаційне, алгоритмічне та програмне забезпечення автоматизованого робочого місця лікаря-гістолога для діагностування онкопатологій різного генезу.

Ключові слова: інформаційно-екстремальне машинне навчання, інформаційний критерій оптимізації, система підтримки прийняття рішень, гістологічне зображення, рак молочної залози.

ВСТУП

Основним інформативним методом діагностування злоякісних пухлин досі залишається візуальний аналіз гістологічних зображень, отриманих за допомогою оптичного мікроскопа. При цьому достовірність діагнозу суттєво залежить від рівня професійної підготовки та досвіду лікаря-гістолога. За даними Всесвітньої організації охорони здоров'я 12 пацієнтам зі 100 помилковий діагноз ставлять або з помилкою першого роду, що зумовлює невинуватене лікування, або з помилкою другого роду, що призводить до пропуску злоякісного новоутворення на ранній стадії його розвитку. Проте сучасне обладнання дає змогу методом повнослایдової цифрової мікроскопії отримувати високоякісні гістологічні зображення морфології тканини, розміщеної на предметному склі. Отже, створення системи підтримки прийняття рішень (СППР), основною функцією якої є первинний аналіз вхідних даних, є актуальною задачею.

У цій статті розглянуто метод інформаційно-екстремального машинного навчання СППР для розпізнавання гістологічних зображень онкопатологій раку молочної залози.

АНАЛІЗ ПРОБЛЕМИ

Задача моделювання когнітивних процесів, притаманних людині під час формування та прийняття класифікаційних рішень, є однією з найскладніших задач сучасності. Розв'язати проблему підвищення функціональної ефективності комп'ютеризованих систем діагностування, зокрема, для розпізнавання гістологічних

зображень, можна шляхом застосування ідей і методів машинного навчання та розпізнавання образів. Приклад комп'ютеризованої системи діагностування раку молочної залози наведено у праці [1]. У цій системі вхідними даними для машинного навчання і тестування були результати цитологічного дослідження тканини молочної залози. При цьому структурований вектор складався всього з дев'яти характеристик клітин тканини молочної залози і однієї змінної, яка фіксувала доброякісність чи злоякісність пухлини. Оскільки доступ до бази даних був вільним, то дослідники мали можливість створювати класифікаційні алгоритми різними методами. Найбільшу достовірність для алфавіту з двох класів отримано під час реалізації методу опорних векторів [2]. Але варто зазначити, що у разі збільшення потужності алфавіту класів розпізнавання, які перетинаються у просторі діагностичних ознак, збіжність методу опорних векторів не гарантована. Значного прогресу в галузі розпізнавання медичних зображень досягнуто завдяки застосуванню швидких графічних процесорів, які дають можливість будувати та навчати штучні нейронні мережі (ШНМ) з великою кількістю шарів. У праці [3] вхідними даними для машинного навчання та тестування ШНМ був набір повнослайдових зображень, який складався з 400 мікроскопічних знімків тканин лімфовузлів. Найкращий варіант навченої ШНМ правильно ідентифікував 92.5 % пухлин на тестовому наборі слайдів, а лікар-патологоанатом — тільки 73.3 %. Але застосування традиційних методів інтелектуального аналізу даних Data Mining [4], включаючи ШНМ [5, 6], для інформаційного синтезу комп'ютеризованих систем діагностування онкопатологій за гістологічними зображеннями, не завжди забезпечує високу функціональну ефективність машинного навчання через низку науково-методологічних обмежень. Ці обмеження здебільшого пов'язані з довільними початковими умовами формування цифрових гістологічних зображень, значним перетином класів розпізнавання у просторі діагностичних ознак, а також багатовимірністю словника ознак і алфавіту класів розпізнавання. У праці [7] для зменшення впливу багатовимірності запропоновано використовувати побудовані на ШНМ екстрактори вхідних даних, але цей підхід неодмінно призводить до втрати інформації.

Одним із перспективних підходів до інформаційного синтезу комп'ютеризованих систем діагностування онкопатологій за гістологічними зображеннями є використання ідей та методів так званої інформаційно-екстремальної інтелектуальної технології (ІЕІТ), яка ґрунтується на максимізації інформаційної спроможності системи під час її машинного навчання [8–10]. Ідея методів ІЕІТ, як і в ШНМ, полягає в адаптації під час машинного навчання вхідного математичного опису до максимальної повної ймовірності прийняття правильних класифікаційних рішень. Але основною перевагою методів інформаційно-екстремального машинного навчання є те, що їх, на відміну від нейроподібних структур, розробляють у межах функціонального підходу до моделювання когнітивних процесів, притаманних людині під час формування та прийняття класифікаційних рішень. Цей підхід, на противагу структурним методам, дає змогу надати гнучкість методам інформаційно-екстремального машинного навчання під час перенавчання системи через розширення алфавіту класів розпізнавання. Крім того, побудовані у межах геометричного підходу вирішальні правила практично розв'язують проблему багатовимірності словника ознак розпізнавання. При цьому варто підкреслити, що методи інформаційно-екстремального машинного навчання реалізуються не в інтерактивному режимі, як це відбувається у ШНМ, а в автоматичному і потребують менше ніж на порядок зразків зображень для формування навчальної матриці.

ФОРМАЛІЗОВАНА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Розглянемо у межах ІЕІТ формалізовану постановку задачі інформаційного синтезу здатної навчатися СППР для діагностування онкопатологій за гістологічними зображеннями. Нехай сформовано алфавіт $\{X_m^o \mid m=1, M\}$ класів розпізнавання, які характеризують гістологічні зразки тканини з різними морфологічними структурами молочної залози. Для кожного класу розпізнавання сформовано тривимірну навчальну матрицю $\|y_{m,i}^{(j)}\|$ яскравості пікселів рецепторного поля зображення. Згідно з концепцією ІЕІТ вхідна навчальна матриця трансформується у задану в бінарному просторі Хеммінга робочу матрицю, яка під час машинного навчання адаптується до максимальної достовірності діагностичних рішень. Обмежимося дворівневою глибиною машинного навчання, за якого оптимізуються радіуси гіперсферичних контейнерів класів розпізнавання та контрольні допуски на діагностичні ознаки. Нехай для класу розпізнавання X_m^o задано вектор параметрів оптимізації:

$$g_m = \langle x_m, d_m, \delta \rangle, \quad (1)$$

де x_m — усереднений за реалізаціями двійковий вектор ознак, вершина якого визначає центр гіперсферичного контейнера класу розпізнавання X_m^o ; d_m — радіус гіперсферичного контейнера класу розпізнавання X_m^o , який відновлюється в радіальному базисі простору ознак розпізнавання; δ — параметр, величина якого дорівнює половині симетричного поля контрольних допусків на діагностичні ознаки.

На параметри машинного навчання накладають відповідні обмеження:

— область значень радіуса контейнера класу розпізнавання X_m^o задають нерівністю

$$d_m < d(x_m \oplus x_c),$$

де $d(x_m \oplus x_c)$ — кодова відстань між вектором x_m і усередненим вектором x_c найближчого сусіднього класу розпізнавання X_c^o , яка визначає міжцентрову відстань відповідних класів розпізнавання;

— область значень параметра δ задають нерівністю

$$0 \leq \delta < \delta_H / 2,$$

де δ_H — нормоване поле допусків, яке визначає область значень системи контрольних допусків на ознаки розпізнавання гістологічних зображень.

Потрібно під час машинного навчання системи діагностування оптимізувати параметри функціонування (1), які забезпечують максимальне значення інформаційного критерію оптимізації в робочій (допустимій) області визначення його функції:

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{G_E \cap \{k\}} E_m^{(k)}, \quad (2)$$

де $E_m^{(k)}$ — обчислене на k -му кроці машинного навчання значення інформаційного критерію; G_E — робоча (допустима) область визначення функції інформаційного критерію; $\{k\}$ — впорядкована множина кроків машинного навчання, на яких відновлюються контейнери класів розпізнавання в радіальному базисі простору діагностичних ознак.

У разі функціонування СППР у режимі екзамену потрібно перевірити функціональну ефективність машинного навчання.

Отже, задача інформаційного синтезу здатної навчатися діагностичної СППР полягає в оптимізації параметрів її функціонування шляхом наближення глобального максимуму інформаційного критерію (2) до його максимального граничного значення.

КАТЕГОРІЙНА ФУНКЦІОНАЛЬНА МОДЕЛЬ МАШИННОГО НАВЧАННЯ

У межах функціонального підходу до моделювання когнітивних процесів категорійну функціональну модель інформаційно-екстремального машинного навчання будують у вигляді орієнтованого графу. Для цього вхідний математичний опис категорійної моделі представляють у вигляді структури

$$I_B = \langle G, T, \Omega, Z, K, Y, X; f_1, f_2 \rangle,$$

де G — множина факторів, які впливають на діагностування; T — множина моментів часу зчитування інформації; Ω — простір діагностичних ознак; Z — алфавіт класів розпізнавання; K — множина кадрів гістологічного зображення; Y — вхідна навчальна матриця яскравості пікселів кадру зображення; X — трансформована у простір Хеммінга робоча бінарна навчальна матриця; f_1 — оператор формування вхідної навчальної матриці Y ; f_2 — оператор перетворення матриці Y у робочу бінарну матрицю X .

Категорійну функціональну модель для другого рівня глибини інформаційно-екстремального машинного навчання діагностичної СППР наведено на рис. 1.

На рис. 1 декартовий добуток $G \times T \times \Omega \times Z \times K$ задає джерело інформації. Терм-множина E значень інформаційного критерію є спільною для всіх контурів оптимізації параметрів машинного навчання, заданих вектором (1). Оператор $r: E \rightarrow \mathbb{R}^{|M|}$ на кожному кроці навчання будує розбиття $\mathbb{R}^{|M|}$, яке відображається оператором θ на розподіл X двійкових векторів діагностичних ознак. Далі оператор $\psi: X \rightarrow I^{|S|}$, де $I^{|S|}$ — множина S гіпотез, перевіряє основну статистичну гіпотезу $\gamma_1: x_{m,i}^{(j)} \in X_m^o$. Оператор γ визначає множину $\mathfrak{T}^{|Q|}$ точнісних характеристик діагностичних рішень, де $Q = S^2$, а оператор φ обчислює множину значень E інформаційного критерію оптимізації, який є функціоналом від точнісних характеристик. Контур оптимізації контрольних допусків на діагностичні ознаки замикається через терм-множину D , елементами якої є значення системи контрольних допусків. Оператор u регламентує процес машинного навчання.

Перевірку функціональної ефективності інформаційно-екстремального

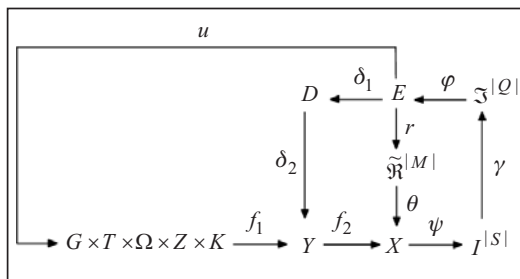


Рис. 1. Категорійна функціональна модель машинного навчання

машинного навчання здійснено для випадку функціонування СППР у режимі екзамену, алгоритм якого аналогічний алгоритму функціонування системи безпосередньо в режимі діагностування. Категорійна модель у вигляді орієнтованого графу відображень множин, які застосовуються у разі функціонування СППР у режимі екзамену, є типовою для методів ІЕІТ [9].

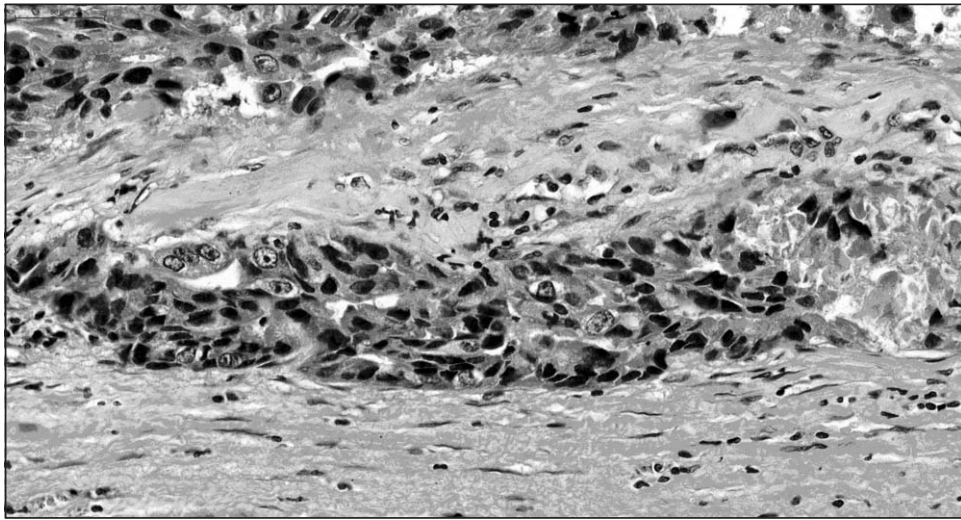


Рис. 2. Зразок тканини, ураженої злоякісними пухлинами

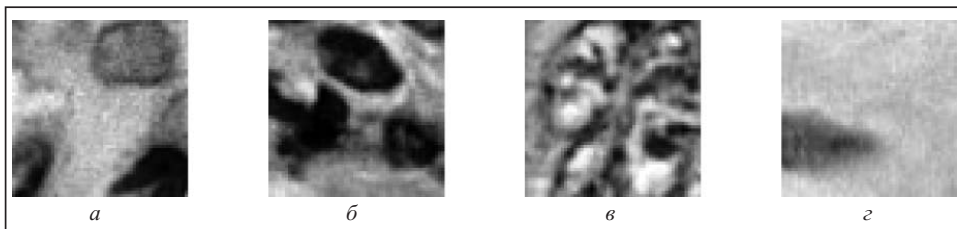


Рис. 3. Зображення кадрів зон інтересу: клас X_1^o (а), клас X_2^o (б), клас X_3^o (в), клас X_4^o (г)

МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕНЬ

У межах ІЕІТ вхідний математичний опис СППР для розпізнавання гістологічних зображень складається зі словника діагностичних ознак, потужність якого визначається розміром кадру цифрового зображення, алфавіту класів розпізнавання, вхідної тривимірної навчальної матриці яскравості пікселів рецепторного поля кадру зображення і робочої бінарної навчальної матриці. На рис. 2 наведено гістологічне зображення морфології тканини, отриманої методом біопсії.

Як класи розпізнавання вибрано кадри розміром 50×50 пікселів наведеного на рис. 2 зображення: клас X_1^o — інфільтративний ріст злоякісних клітин; клас X_2^o — гіперхромні ядра; клас X_3^o — скляні ядра; клас X_4^o — сполучна тканина. Вибрані кадри наведено на рис. 3.

Вхідну навчальну матрицю яскравості пікселів згідно з [8] сформовано обробленням у полярній системі координат кадрів кольорового гістологічного зображення, що дає змогу забезпечити інваріантність вирішальних правил відносно довільних початкових умов формування зображень.

У межах ІЕІТ контрольні допуски розглядають як рівні квантування під час перетворення на кожному кроці машинного навчання вхідної евклідової навчальної матриці в робочу бінарну навчальну матрицю, задану у просторі Хеммінга. При цьому інформаційно-екстремальне машинне навчання СППР з оптимізацією контрольних допусків реалізується за двоциклічною ітераційною процедурою пошуку глобального максимуму усередненого за алфавітом класів

розпізнавання інформаційного критерію в робочій (допустимій) області визначення його функції

$$\delta^* = \arg \max_{G_\delta} \left\{ \max_{G_E \cap \{k\}} \bar{E}^{(k)} \right\}, \quad (3)$$

де δ^* — оптимальний параметр поля контрольних допусків; G_δ — допустима область значень параметра δ поля контрольних допусків.

Як критерій оптимізації параметрів машинного навчання для двохальтернативної системи оцінок класифікаційних рішень розглянуто модифіковану міру Кульбака у вигляді

$$E_m^{(k)} = \frac{1}{n_{\min}} \log_2 \left\{ \frac{2n_{\min} + 10^{-r} - [K_1^{(k)} + K_2^{(k)}]}{[K_1^{(k)} + K_2^{(k)}] + 10^{-r}} \right\} [n - (K_1^{(k)} + K_2^{(k)})], \quad (4)$$

де $K_{1,h,s,m}^{(k)}$ — кількість подій, за яких реалізації, що належать класу $X_{h,s,m}^o$, помилково до нього не відносять; $K_{2,h,s,m}^{(k)}$ — кількість подій, за яких помилково відносять до класу розпізнавання $X_{h,s,m}^o$ реалізації сусіднього у страті класу розпізнавання; $n_{\min}$ — мінімальний обсяг репрезентативної навчальної вибірки; 10^{-r} — достатньо мале число, яке вводять, щоб уникнути ділення на нуль.

Внутрішній цикл процедури (3) реалізує так званий базовий алгоритм інформаційно-екстремального машинного навчання. Основними функціями цього алгоритму є обчислення на кожному кроці машинного навчання інформаційного критерію, пошук глобального максимуму його функції і визначення оптимальних геометричних параметрів контейнерів класів розпізнавання. При цьому постає задача вибору базового класу розпізнавання, відносно якого визначається система контрольних допусків на діагностичні ознаки. Було запропоновано як базовий клас розпізнавання вибирати зображення кадру з найбільшою дисперсією яскравості ознак розпізнавання. Обґрунтування цієї гіпотези полягає в тому, що клас розпізнавання, який має найбільший розкид яскравості ознак, є найближчим до всіх класів із заданого алфавіту.

Розглянемо схему алгоритму машинного навчання з оптимізацією контрольних допусків на діагностичні ознаки за процедурою (3):

- 1) обнулення лічильника класів розпізнавання: $m := 0$;
- 2) $m := m + 1$;
- 3) обнулення лічильника зміни значень параметра поля контрольних допусків: $\delta := 0$;
- 4) $\delta := \delta + 1$;
- 5) обчислення нижніх $A_{H,i}$ і верхніх $A_{B,i}$ контрольних допусків на ознаки розпізнавання відповідно за правилами

$$A_{H,i} = y_i - \delta, \quad A_{B,i} = y_i + \delta; \quad (5)$$

- 6) обнулення лічильника кроків зміни радіуса гіперсферичного контейнера класу розпізнавання: $k := 0$;

- 7) $k := k + 1$;

- 8) формування тривимірного масиву бінарної навчальної матриці $\{x_{m,i}^{(j)}\}$, елементи якої обчислюються за правилом

$$x_{m,i}^{(j)}[k] = \begin{cases} 1, & \text{якщо } A_{HK,i}[k] < y_{m,i}^{(j)} < A_{BK,i}[k], \\ 0, & \text{якщо інакше;} \end{cases}$$

9) формування масиву усереднених двійкових векторів-реалізацій $\{x_m\}$, елементи яких визначаються за правилом

$$x_{m,i} = \begin{cases} 1, & \text{якщо } \frac{1}{n} \sum_{j=1}^n x_{m,i}^{(j)} > \rho_m, \\ 0, & \text{якщо інакше,} \end{cases}$$

де ρ_m — рівень квантування координат двійкового вектора x_m , який, як усталено, дорівнює 0.5;

- 10) розбиття множини векторів $\{x_m\}$ на пари найближчих «сусідів»;
- 11) обчислення інформаційного критерію оптимізації (4);
- 12) якщо $k \leq N$, то виконується п. 7, в іншому разі — п. 13;
- 13) якщо $\delta < \delta_H$, то виконується п. 4, в іншому разі — п. 14;
- 14) визначення максимального значення інформаційного критерію в робочій області визначення його функції;
- 15) якщо $m < M - 1$, то реалізується п. 2, в іншому разі — п. 16;
- 16) визначення глобального максимуму усередненого інформаційного критерію $\bar{E}^*$ в робочій області визначення його функції;
- 17) визначення за процедурою (3) оптимальних значень параметра δ^* і за правилами (8) відповідних нижніх $A_{H,i}^*$ і верхніх $A_{B,i}^*$ оптимальних контрольних допусків на всі ознаки розпізнавання

Для підвищення функціональної ефективності діагностичної СППР реалізовано алгоритм машинного навчання з послідовною оптимізацією контрольних допусків. При цьому отримані на етапі паралельної оптимізації контрольні допуски взято як стартові під час послідовної оптимізації, яку здійснено за процедурою

$$\delta_{K,i}^* = \arg \bigotimes_{l=1}^L \left\{ \max_{G_{\delta i}} \left[\frac{1}{M} \sum_{m=1}^M \max_{G_{Em} \cap \{k\}} E_{m,k}^{(l)} \right] \right\}, \quad i = \overline{1, N}, \quad (6)$$

де $E_{m,k}^{(l)}$ — значення інформаційного критерію, обчислене на k -му кроці машинного навчання під час l -го прогону процедури (6); L — кількість прогонів процедури (6), обумовлених неоптимальними стартовими величинами контрольних допусків для всіх ознак; $\otimes$ — символ операції повторення.

За отриманими у процесі машинного навчання оптимальними геометричними параметрами контейнерів класів розпізнавання побудовано вирішальні правила

$$(\forall X_m^o \in \mathfrak{R}^{|M|})(\forall x^{(j)} \in \mathfrak{R}^{|M|}) \{ \text{if } [(\mu_m > 0) \& \& (\mu_m = \max_{\{m\}} \{\mu_m\})] \text{ then } x^{(j)} \in X_m^o \text{ else } x^{(j)} \notin X_m^o \}, \quad (7)$$

де $x^{(j)}$ — вектор, що розпізнається; μ_m — функція належності вектора $x^{(j)}$ контейнеру класу розпізнавання X_m^o .

У виразі (7) функція належності для гіперсферичного контейнера класу розпізнавання X_m^o визначається за відомою формулою [9]

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

де x_m^* , d_m^* — оптимальні параметри машинного навчання: усереднена двійкова реалізація і радіус гіперсферичного контейнера відповідно.

РЕЗУЛЬТАТИ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ

На рис. 4 наведено графік залежності усередненого нормованого критерію (4) від параметра δ поля контрольних допусків, отриманий під час машинного навчання за процедурою (3). На рис. 4 і далі сірим кольором виділено робочі області визначення функції критерію, в яких перша і друга достовірності є відповідно більшими ніж помилки першого і другого роду.

Аналіз рис. 4 свідчить про те, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta^* = 34$ (у градаціях яскравості) для максимального значення інформаційного критерію $\bar{E}^* = 0.14$. Цей результат відтворено на рис. 5, на якому систему контрольних допусків позначено сірою зоною. При цьому лінією 1 позначено усереднений вектор діагностичних ознак класу розпізнавання X_1^o , лінією 2 — класу розпізнавання X_2^o , лінією 3 — класу розпізнавання X_3^o і лінією 4 — класу розпізнавання X_4^o .

Аналіз рис. 5 показує, що поле контрольних допусків охоплює ознаки всіх класів розпізнавання із заданого алфавіту, що дає змогу використовувати для розв'язання задач класифікації критерії близькості, до яких відносять й інформаційні критерії.

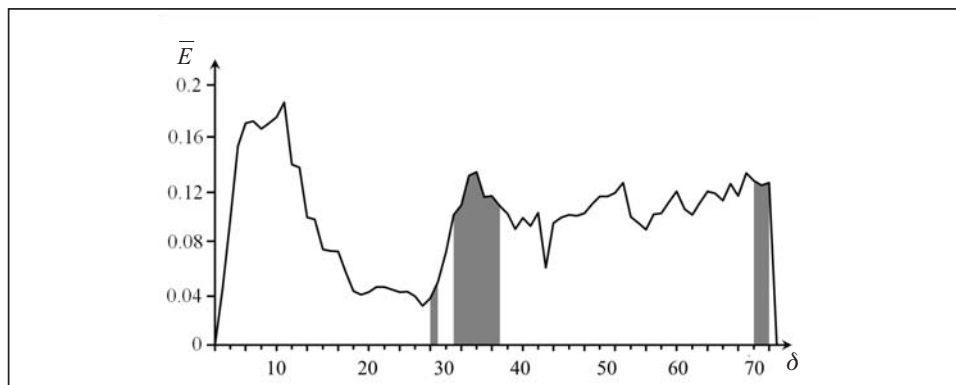


Рис. 4. Графік залежності усередненого інформаційного критерію оптимізації від параметра поля контрольних допусків

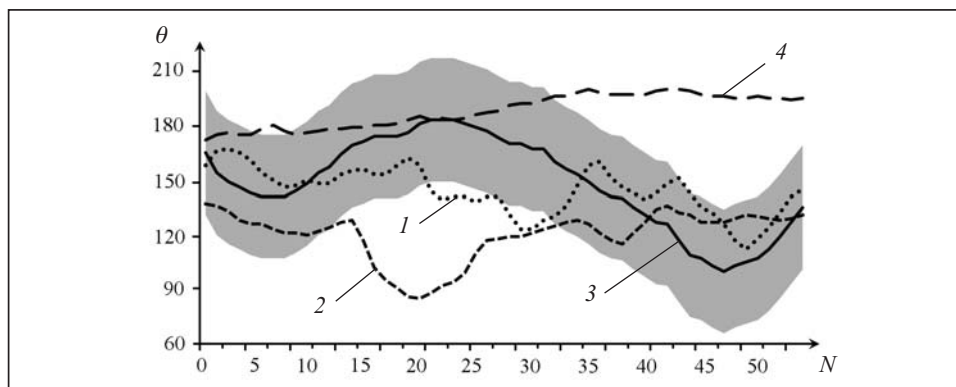


Рис. 5. Система контрольних допусків для заданого алфавіту класів розпізнавання

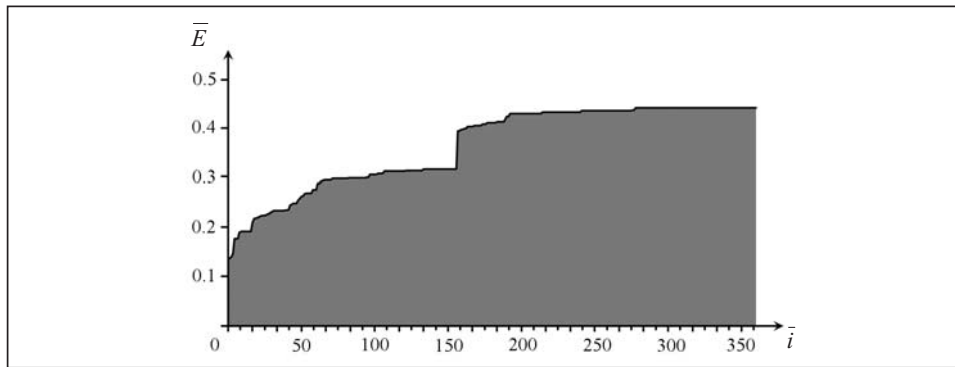


Рис. 6. Графік зміни інформаційного критерію під час послідовної оптимізації контрольних допусків

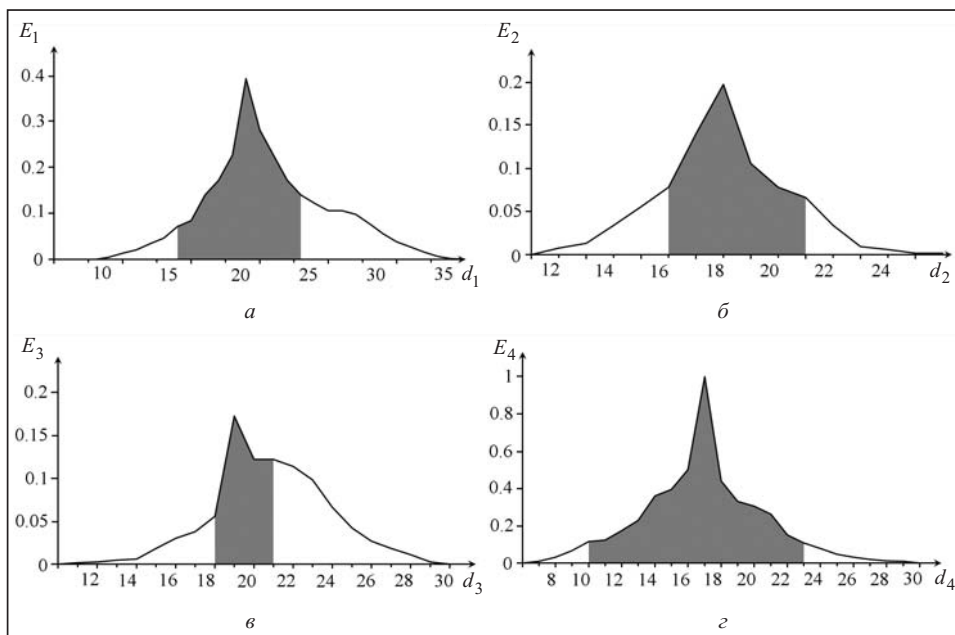


Рис. 7. Графіки залежності критерію (4) від радіусів контейнерів класів розпізнавання: клас X_1^o (а), клас X_2^o (б), клас X_3^o (в), клас X_4^o (г)

Для підвищення максимального значення інформаційного критерію оптимізації реалізовано машинне навчання за алгоритмом послідовної оптимізації (6). На рис. 6 наведено графік зміни нормованого критерію (4) під час послідовної оптимізації контрольних допусків на ознаки розпізнавання.

Аналіз рис. 6 свідчить про те, що інформаційний критерій оптимізації на п'ятому прогоні, який визначається відношенням кількості ітерацій до кількості ознак N у реалізації, збільшився і досягнув значення $\bar{E}^* = 0.45$.

На рис. 7 наведено графіки залежності інформаційного критерію (4) від радіусів контейнерів класів розпізнавання.

Аналіз рис. 7 показує, що оптимальні радіуси контейнерів класів розпізнавання дорівнюють $d_1^* = 21$ (тут і далі в кодових одиницях) для класу X_1^o ; $d_2^* = 18$ для класу X_2^o ; $d_3^* = 19$ для класу X_3^o і $d_4^* = 17$ для класу X_4^o .

На рис. 8 наведено отримане в режимі екзамену за вирішальними правилами (7) оцифроване зображення результатів розпізнавання кадрів гістологічного зображення, де цифрами позначено відповідні класи розпізнавання.

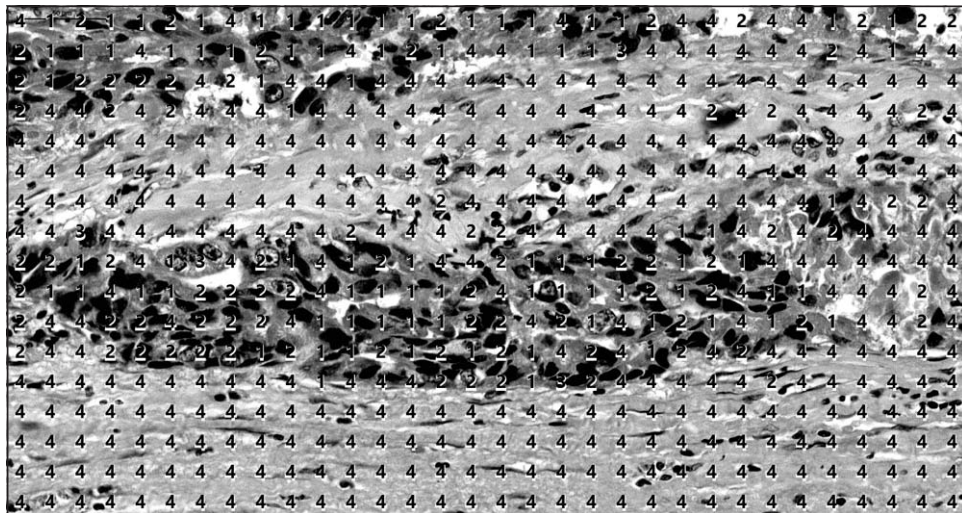


Рис. 8. Оцифроване зображення результатів розпізнавання кадрів гістологічного зображення

Візуальний аналіз рис. 8 підтверджує достатньо високу достовірність розпізнавання кадрів гістологічного зображення.

ОБГОВОРЕННЯ РЕЗУЛЬТАТІВ

Відповідно до запропонованої категорійної функціональної моделі алгоритм інформаційно-екстремального машинного навчання другого рівня глибини полягає в реалізації двоциклічної процедури ітераційного пошуку глобального максимуму інформаційного критерію оптимізації параметрів функціонування діагностичної СППР. На відміну від ШНМ запропонований у статті метод інформаційно-екстремального машинного навчання не передбачає інтерактивного режиму, тобто здійснюється автоматично без участі оператора, який вводить тільки вхідні дані.

Отримані результати інформаційно-екстремального машинного навчання діагностичної СППР дають змогу ідентифікувати чотири кадри різної морфологічної структури гістологічного зображення раку молочної залози. Візуальний аналіз рис. 8 показує, що з 512 кадрів рецепторного поля гістологічного зображення правильно було ідентифіковано 482 кадри. Отже, повна ймовірність прийняття правильних діагностичних рішень дорівнює $P_t = 0.94$, що перевищує відомий результат $P_t = 0.92$, отриманий із застосуванням ШНМ для двох класів розпізнавання.

ВИСНОВКИ

1. На основі запропонованої категорійної функціональної моделі розроблено та програмно реалізовано алгоритм інформаційно-екстремального машинного навчання СППР для діагностування раку молочної залози, який дає змогу отримати для алфавіту з чотирьох класів розпізнавання усереднену повну ймовірність правильного прийняття діагностичних рішень 0.94.

2. Запропоновано як базовий клас розпізнавання, відносно якого визначається система контрольних допусків на діагностичні ознаки, вибирати клас за найбільшою дисперсією його вхідної навчальної матриці яскравості.

3. Подальше підвищення функціональної ефективності інформаційно-екстремального машинного навчання системи розпізнавання гістологічних зображень потребує збільшення глибини навчання шляхом оптимізації додаткових параметрів функціонування системи, включаючи параметри оброблення зображень. У разі збільшення потужності алфавіту класів розпізнавання доцільним є

перехід від лінійної структури вхідних даних до ієрархічної, що буде предметом подальших досліджень.

СПИСОК ЛІТЕРАТУРИ

1. UCI machine learning repository: Datasets. URL: <http://archive.ics.uci.edu/ml/machine-learningdatabases/breast-cancer-wisconsin>.
2. van den Burg G.J.J., Groenen P.J.F. GenSVM: A generalized multiclass support vector machine. *Journal of Machine Learning Research*. 2016. Vol. 17, N 224. P. 1–42.
3. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *Proc. 18th International Conference on Medical image computing and computer-assisted intervention MICCAI 2015* (5–9 October 2015, Munich, Germany). Munich, 2015. P. 234–241.
4. Xu G., Zong Y., Yang Y.Z. Applied data mining. Boca Raton: CRC Press, 2013. 284 p.
5. Moskalenko A.S., Moskalenko V.V., Pimonenko S.V., Korobov A.G. Development of the method of features learning and training decision rules for the prediction of violation of service level agreement in a cloud-based environment. *Eastern-European Journal of Enterprise Technologies*. 2017. Vol. 5, N 2 (89). P. 26–33.
6. Ammour N., Alhichri H., Bazi Y., Benjdira B., Alajlan N., Zuair M. Deep learning approach for car detection in UAV imagery. *Remote Sensing*. 2017. Vol. 9, N 4. P. 1–15.
7. Moskalenko V.V., Korobov A.G. Information-extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extractor. *Radio Electronics, Computer Science, Control*. 2017. N 2. P. 61–69. <https://doi.org/10.15588/1607-3274-2017-2-7>.
8. Dovbysh A.S., Rudenko M.S. Information-extreme learning algorithm for a system of recognition of morphological images in diagnosing oncological pathologies. *Cybernetics and Systems Analysis*. 2014. Vol. 50, N 1. P. 157–163. <https://doi.org/10.1007/s10559-014-9603-y>.
9. Dovbysh A.S., Budnyk M.M., Piatachenko V.Yu., Myronenko M.I. Information-extreme machine learning of on-board vehicle recognition system. *Cybernetics and Systems Analysis*. 2020. Vol. 56, N 4. P. 534–543. <https://doi.org/10.1007/s10559-020-00269-y>.
10. Naumenko I.V., Myronenko M.I., Savchenko T.R. Information-extreme machine training of onboard recognition system with optimization of RGB-component digital images. *Radioelectronic and Computer Systems*. 2021. N 4. P. 59–70. <https://doi.org/10.32620/reks.2021.4.05>.

**A.S. Dovbysh, I.V. Shelehov, A.M. Romaniuk,
R.A. Moskalenko, T.R. Savchenko**

A DECISION SUPPORT SYSTEM FOR DIAGNOSIS OF BREAST ONCOPATHOLOGIES BY HISTOLOGICAL IMAGES

Abstract. The authors propose the method of information-extreme machine learning of the decision support system for diagnosing breast oncopathologies by histological images. In contrast to the available methods, including neuro-like structures, this method was developed as part of a functional approach to modeling cognitive processes of generating and making decisions by natural intelligence. At the same time, the decision rules constructed within the framework of the geometric approach are practically invariant to the multidimensionality of the diagnostic feature space. The developed method allows the creation of information and algorithmic support and software of the automated workplace of a histologist for diagnosing oncopathologies of different genesis.

Keywords: information-extreme machine learning, information criterion of optimization, decision support system, histological image, breast cancer.

Надійшла до редакції 22.08.2022