

УДК 616.65; 616.62-008.22, 61:007, 004.931;  
004.932, 004.891

УКПП

№ держреєстрації № 0122U000773

Інв. №

**Міністерство освіти і науки України**  
Сумський державний університет (СумДУ)  
40007, м. Суми, вул. Римського-Корсакова, 2  
тел. (0542) 33-35-39 факс. (0542) 33-40-58  
e-mail: info@sci.sumdu.edu.ua

**ЗАТВЕРДЖУЮ**

Проректор з наукової роботи  
д-р фіз.-мат. наук, професор

\_\_\_\_\_ А.М. Черноус

**ЗВІТ**

**ПРО НАУКОВО-ДОСЛІДНИДНУ РОБОТУ**

Застосування штучного інтелекту для забезпечення автоматизації і стандартизації системи Глісона при діагностиці раку передміхурової залози

**РОЗРОБКА ДІАГНОСТИЧНОГО АЛГОРИТМУ НА ОСНОВІ  
ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНОГО МЕТОДУ КОМПАРАТОРНОГО  
РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ ТА СТВОРЕННЯ ДОДАТКОВИХ  
МОДУЛІВ УНІФІКАЦІЇ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ ОЦІНКИ  
БІОХІМІЧНИХ ПОКАЗНИКІВ ПАЦІЄНТІВ  
(проміжний)**

Науковий керівник  
проф., д.мед наук

А. М. Романюк

2022

Рукопис закінчено 25 грудня 2022 р.

Результати роботи розглянуто науковою радою СумДУ, протокол № \_ від \_ грудня 2022 р.

**СПИСОК АВТОРІВ**

Керівник НДР,  
гол.наук. співроб.,  
д.мед. наук

---

(підпис)

---

(дата)

Романюк Анатолій  
Миколайович  
(реферат, вступ, висновки,  
розділи 1,2)

Відповідальний  
виконавець  
наук.співроб.,

---

(підпис)

---

(дата)

Москаленко Роман  
Андрійович  
(реферат, вступ, висновки,  
розділи 1,2,3)

Ст. наук співроб.,  
к.мед.наук

---

(підпис)

---

(дата)

Коробов Артем Геннайдійович  
(розділ 2)

М.н.с.

---

(підпис)

---

(дата)

Денисенко Анастасія Петрівна  
(розділ 1)

Лаборант

---

(підпис)

---

(дата)

Чижма Руслана Анатоліївна  
(розділ 1)

Лаборант

---

(підпис)

---

(дата)

Казбан Надія Володимирівна  
(розділ 1)

Лаборант

---

(підпис)

---

(дата)

Голоп'орова Владлена  
Сергіївна  
(розділ 1)

## РЕФЕРАТ

Звіт про НДР: 48 с., 15 рис., 47 джерел.

Об'єкт дослідження – процес діагностування онкологічної патології.

Предмет досліджень – застосування засобів штучного інтелекту для діагностики раку передміхурової залози.

Мета роботи – розробка нового програмного забезпечення для покращення стандартизації градації пухлин передміхурової залози за шкалою Глісона з високою відтворюваністю результатів.

Для досягнення поставленої мети були визначені такі завдання:

- 1) створення бази даних референтних зображень гістологічних препаратів з різними діагностичними патернами шкали Глісона;
- 2) розробка математичного апарату автоматизованої діагностичної програми;
- 3) класифікація діагностичних паттернів різних ступенів раку передміхурової залози згідно оновленої шкали Глісона;
- 4) розробка додаткових модулів програми та її інтерфейсу.

Методи дослідження – бібліометричний аналіз наукових публікацій щодо застосування штучного інтелекту в морфологічній діагностиці раку передміхурової залози в електронних базах даних, таких як PubMed, Scopus, Web of Science та Google Scholar, за останні 35 років за ключовими словами «штучний інтелект» та «рак передміхурової залози». Для візуалізації результатів дослідження застосовувалися інструменти бібліометричного аналізу бази даних Scopus та система побудови бібліометричних мереж VOSviewer; інформаційно-екстремальний метод компараторного розпізнавання зображень.

Ключові слова: штучний інтелект, рак передміхурової залози, шкала Глісона, бібліометричний аналіз, система підтримки прийняття рішень, ієрархічне інформаційно-екстремальне машинне навчання.

## ЗМІСТ

ВСТУП .....	5
1. ОГЛЯД ЛІТЕРАТУРИ ТА БІБЛІОМЕТРИЧНИЙ АНАЛІЗ ПРОБЛЕМИ ДІАГНОСТИКИ РАКА ПЕРЕДМІХУРОВОЇ ЗАЛОЗИ .....	7
2. СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ДІАГНОСТУВАННЯ ОНКОПАТОЛОГІЙ ЗА ГІСТОЛОГІЧНИМИ ЗОБРАЖЕННЯМИ .....	17
ВИСНОВКИ .....	41
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ .....	43

## ВСТУП

Основною ідеєю проєкту є підвищення надійності, відтворюваності і точності діагностики раку передміхурової залози (РПЗ) за рахунок використання засобів штучного інтелекту (ШІ). Це дозволить позбутися суб'єктивізму в оцінці результатів гістологічного дослідження, створить умови для економії коштів та часу, який витрачається на діагностику.

Зараз головним підходом до розпізнавання і аналізу зображень у медицині є застосування методів машинного навчання та нейронних мереж.

Методика цифрової патології (ЦП) має певні етапи, які включають в себе первинну діагностику на екрані монітора комп'ютера, морфометрію за допомогою програмного забезпечення для аналізу зображень, а також певний рівень знань з галузі як патології так і програмного забезпечення. Тому спеціалісту з ЦП потрібно постійно удосконалювати свої знання та вміння на курсах тематичного удосконалення, конференціях, майстер-класах, для забезпечення більш якісного та ефективного процесу в галузі медицини.

Автоматизована програма діагностики, яка розробляється у нашому проєкті буде враховувати національні та локальні соціально-економічні особливості та буде адаптована до них. По-перше, для роботи програми не буде потрібно наявність повних сканів гістологічних препаратів – достатньо буде лише сфокусованих мікрофото патології. По-друге, автоматизована програма матиме модуль для стандартизації і нормалізації мікрофото. Зазначені інновації зроблять можливою роботу алгоритму для оцінки гістологічних зразків різної якості забарвлення та з усіх лабораторій і відділень, які мають цифрова камеру в світлооптичних мікроскопах.

Ми пропонуємо науково-технічний проєкт, який буде направлений на розробку нової автоматизованої діагностичної програми стандартизації системи Глісона при діагностиці раку передміхурової залози. У нашому рішенні завдання буде застосовано технологію інформаційно-екстремального методу компараторного розпізнавання зображень, що дозволить створити

автоматизовану систему, яка з накопиченням відповідних референтних зображень здатна самонавчатися і покращувати діагностичну ефективність. В автоматичній діагностичній програмі будуть інтегровані біохімічні результати досліджень простат-специфічного антигену пацієнтів, що підвищить точність та достовірність отримуваних результатів. Автоматизована диференційно-діагностична система підтримки прийняття рішень для стандартизації шкали Глісона у діагностиці РПЗ буде конкурентною на світовому ринку, в т.ч. для ринку Східної Європи та країн, що розвиваються.

Вхідний математичний опис системи діагностування РПЗ буде складається із: словника ознак розпізнавання, потужність якого визначається розміром кадру цифрового зображення; алфавіту класів розпізнавання, потужність якого визначається кількістю об'єктів розпізнавання; вхідної тривимірної навчальної матриці яскравості пікселів рецепторного поля кадру зображення; робочої бінарної навчальної матриці, яка в процесі машинного навчання адаптується до максимальної повної ймовірності прийняття правильних класифікаційних рішень. При розв'язанні задачі ідентифікації кадрів зображення тканин нами досліджувалася функціональна ефективність машинного навчання системи діагностування онкопатологій при формуванні вхідної навчальної матриці двома способами: шляхом формування із класів лінійного класифікатора, що на практиці використовується для не великої кількості алфавіту класів, і ієрархічний класифікатор, призначенням якого є боротьба з багатовимірністю алфавіту класів. Під час машинного навчання, згідно принципів ІЕІ-технології, формується система контрольних допусків. При цьому обирається базовий клас, на основі якого буде створюватися система контрольних допусків (СКД). Від цього етапу залежить подальша точність створених вирішальних правил. Оптимальна, в інформаційному розумінні, СКД повинна описувати як найбільше ознак обраних класів. Тому, було прийнято рішення, щоб обирати базовий клас за дисперсією.

## **1 ОГЛЯД ЛІТЕРАТУРИ ТА БІБЛІОМЕТРИЧНИЙ АНАЛІЗ ПРОБЛЕМИ ДІАГНОСТИКИ РАКА ПЕРЕДМІХУРОВОЇ ЗАЛОЗИ**

Рак передміхурової залози (РПЗ) є одним з найбільш поширених злоякісних новоутворень у населення розвинутих країн та є другою причиною смертності чоловіків від онкологічних захворювань [1].

Складність процесу діагностики і лікування РПЗ обумовлюється його генетичною нестабільністю. Іноді РПЗ має спадковий характер, однак більшість випадків виникає через соматичні мутації. Пізній початок РПЗ може свідчити, про накопичення генетичних змін задовго до клінічної маніфестації [2]. Дослідження показують, що хронічні інфекції та запалення можуть призвести до РПЗ через генетичні або епігенетичні зміни [3]. Покращення діагностики РПЗ відбулося за допомогою запровадження визначення простат-специфічного антигену (ПСА). До широкого визначення ПСА, яке почалося наприкінці 1980-х років, понад 50 % пацієнтів мали прогресуючий перебіг РПЗ [4]. Тепер же із поширеним використанням цього тесту постають інші проблеми – надмірної діагностики та лікування, а також рецидивів РПЗ у більше ніж 20 % пацієнтів після радикальної простатектомії, незважаючи на раннє втручання [5].

Використання штучного інтелекту (ШІ) може знизити рівень суб'єктивності та покращити ефективність діагностики РПЗ, використовуючи при цьому менше ресурсів, ніж стандартна схема діагностики зараз. Щодо соціально-економічного навантаження, то наприклад в США, десятирічні витрати на лікування пацієнтів із низьким та високим ризиком РПЗ оцінюються в 45 957 і 188 928 доларів США відповідно. Тому ширше залучення алгоритмів ШІ може допомогти пом'якшити цей тягар [6].

### **Матеріали та методи дослідження**

Автори провели пошук, аналіз та систематизацію наукових публікацій щодо застосування штучного інтелекту в морфологічній діагностиці раку передміхурової залози в електронних базах даних, таких як PubMed, Scopus,

Web of Science та Google Scholar, за останні 35 років за ключовими словами «штучний інтелект» та «рак передміхурової залози». Для візуалізації результатів дослідження застосовувалися інструменти бібліометричного аналізу бази даних Scopus та система побудови бібліометричних мереж VOSviewer.

### **Діагностика раку передміхурової залози**

Визначення рівня ПСА вважається одним із найпоширеніших лабораторних досліджень для діагностики РПЗ. Для раннього виявлення патологічним вважається рівень більше 4 нг/мл, але приблизно 20 % пацієнтів, у яких діагностовано РПЗ, не досягають цього порогу [6].

Іншим поширеним методом діагностики РПЗ є трансректальне ультразвукове дослідження із біопсією дванадцяти ділянок передміхурової залози, яке найчастіше рекомендується при підвищеному рівні ПСА або аномальному пальцевому ректальному обстеженні [7].

Протягом останніх років збільшилась роль магнітно-резонансної томографії, яка допомагає в діагностиці та виборі тактики лікування, а також нагляді за пацієнтами з РПЗ [8]. Причому застосування алгоритмів ШІ вже зараз зайняло свою нішу в радіологічній діагностиці та створює можливості для широкомасштабного скринінгу РПЗ без додаткового навантаження на лікарів-рентгенологів [9].

Щодо морфологічної діагностики, то шкала Глісона була розроблена ще у 1966 р. Дональдом Глісоном для оцінки стадії, клінічного ризику та прогнозу РПЗ і донині багато разів модифікувалася. З метою стандартизації діагностики РПЗ Міжнародним Товариством урологічної патології була створена база даних зображень, відома як Pathology Imagebase, яка у деяких випадках допомагає дійти консенсусу патологам зі всього світу [10].

Проблема зі стандартизацією морфологічної діагностики виникають коли патологу доводиться формувати своє заключення на основі суб'єктивного тлумачення набору правил, які самі по собі ще й нечітко сформульовані. Ситуація з РПЗ трохи краща, ніж з іншими пухлинами,



оскільки шкала Глісона містить численні архітектурні дескриптори та не включає інші непаралельні показники, такі як ядерна атипія, тобто є монопараметричною системою [11].

Класифікація за Глісоном дозволяє поділити РПЗ за агресивністю на три групи ризику: низький (група 1), середній (групи 2 і 3) та високий ризик (група 4 і 5). Група 1 позначається як GS6, група 2 — GS7 (3+4), група 3 — GS7 (4+3), група 4 — GS8 і група 5 — GS9-10 відповідно [6].

Різні гістологічні патерни допомагають віднести гістологічне зображення РПЗ до одного з п'яти шаблонів, де 1 — це добре диференційований, а 5 — погано. Шаблон 1 описує маленькі добре диференційовані клітини. У шаблоні 2 ці клітини знаходять на більшій відстані один від одного за рахунок більшої кількості стромы. Для шаблону 3 характерні добре сформовані та розділені залози різного розміру. Шаблон 4 включає зрощені залози, решітчасті та клубочкові структури та погано сформовані залози. Шаблон 5 описує слабодиференційовані окремі клітини, суцільні гнізда, канатики та лінійні масиви. Остаточний бал за Глісоном є сумою двох переважаючих шаблонів, присутніх у гістологічному зразку. У поточній клінічній практиці найнижчий бал за Глісоном, який призначається є GS6 (3 + 3) [11, 12].

Оцінка за Глісоном допомагає встановити остаточний діагноз РПЗ і є основою для вибору тактики лікування. Проте ручний і якісний характер аналізу гістологічних зображень РПЗ обмежує не тільки швидкість та пропускну здатність, а й перешкоджає точності через значні відмінності в оцінюванні між спостерігачами [13].

Зараз значний інтерес представляє автоматизація і стандартизація оцінки шкали Глісона. Ця необхідність зумовлена низькою відтворюваністю і значною суб'єктивністю результатів [14].

Крім того пандемія Covid-19 та війна в Україні викликала ще більший дефіцит патологів у віддалених місцях, який хоча б частково можна

компенсувати дистанційною роботою та максимальною автоматизацією діагностичного процесу.

Застосування штучного інтелекту для морфологічної діагностики раку передміхурової залози

Сьогодні ідея застосування ШІ для вирішення проблеми суб'єктивізму у патоморфологічній діагностиці захворювань набула значного поширення. Велика кількість програмних алгоритмів для аналізу зображень знаходиться у стадії розробки або тестування.

Перший підхід був застосований у роботі Källén et al та базувався на глибокому навчанні для прогнозування балів за шкалою Глісона. Проте дослідження було обмежене гістологічними зображеннями з однорідними патернами в межах одного шаблону, в той час як зазвичай зразки РПЗ характеризується одночасним поєднанням декількох шаблонів [15].

У новішій роботі Zhou et al. вдалося досягти загальної точності 75 % у диференціації GS7 (3+4) та GS7 (4+3), причому алгоритм був протестований на наборі зображень із бази The Cancer Genome Atlas [16]. А у роботі del Toro et al. на основі тої самої бази зображень алгоритм навчився розрізняти групи із низьким та середнім ризиком (7 та нижче за Глісоном) від групи високого ризику (8 та вище за Глісоном) [17].

Наведені вище дослідження показують, що автоматизація та стандартизація системи оцінювання за Глісоном за допомогою штучного інтелекту можлива, але всі вони були проведені в обмежених умовах — або на однорідних шаблонах, або як проблема бінарної класифікації на групи [15].

Для аналізу патогістологічних зображень зараз застосовується велика кількість алгоритмів машинного навчання: дерево рішень, бустінг, метод k-найближчих сусідів, метод опорних векторів, глибоке навчання, метод Баєсового структурного часового ряду та інші [13, 18-21].

Абсолютна більшість цих алгоритмів потребує повносканових слайдів гістологічних препаратів для діагностичного програмного забезпечення, а тому їх робота можлива лише у місцях, де вони наявні [22-27].

У роботі діагностичних алгоритмів, які широко розробляються і тестуються, відмічається значна залежність від сегментування тканини пухлини на епітеліальний, стромальний та люмінальний («просвітний», порожній) компонент. Наприклад, в одній із програм використовували підхід порівняння співвідношення компонентів тканини для диференціального діагнозу добро- та злоякісної патології простати [28].

У алгоритмі Gertych et al для діагностики РПЗ орієнтувались на кількісну оцінку епітеліального компоненту доброякісних та злоякісних пухлин [29]. А команда розробників Singh M et al намагалася застосовувати оцінку цілих епітеліальних структур і патернів для морфологічної класифікації РПЗ [30].

Але перед тим як почати автоматизований аналіз зображень зазвичай спочатку необхідне створення фрагментів з певними гістологічними патернами, які отримали оцінку досвідченим патологом. Під час аналізу зображення поділяється на «плитки», які потім класифікуються, наприклад, як строма, доброякісний елемент, злоякісний низько- чи високодиференційований елемент [29].

При деяких онкологічних захворюваннях автоматична якісна і кількісна оцінка компонентів тканини допомагає у прогнозуванні результатів лікування пацієнта. Наприклад, ядерні та стромальні особливості, що виявляються алгоритмами ШІ при аналізі гістологічних зображень, корелюють з прогнозом у пацієнтів із раком молочної залози [31]. Виявлення лімфоцитів, що інфільтрують пухлину, за допомогою моделей глибокого навчання дозволяє проводити просторовий аналіз імунного профілю, що дає можливість для прогнозування результатів лікування [32].

Однак при діагностиці РПЗ, де поки основна увага зосереджена на автоматичному оцінюванні зразка за шкалою Глісона та вирішенні проблеми суб'єктивності, про прогнозування за допомогою машинного навчання говорити поки що рано [33].

Одне із досліджень команди Nir et al із застосуванням кількох машинних методів і методів глибокого навчання демонструє 92 % точності при виявленні РПЗ та 79 % точності при його класифікації за Глісоном. Проте узгодженість результатів із заключеннями патологів була помірно низькою, ймовірно, через значну суб'єктивність у оцінюванні різними фахівцями [22].

У найбільшому на сьогодні дослідженні Nagpal et al. розробили свій алгоритм глибокого навчання та перевірили його ефективність за допомогою 29 сертифікованих патологів. Незважаючи на це, узгодженість із фахівцями не перевищує 70 % [34].

Отже, алгоритми ШІ зараз активно розвиваються та поступово займають свою нішу у діагностиці РПЗ. Не зважаючи на те, що на даний момент узгодженість між результатами автоматизованого аналізу та фахівцями-патологами не перевищує 70 %, безумовно, подальша безперервна взаємодія між людьми-спостерігачами та машинним навчанням має потенціал для автоматизації та стандартизації діагностики РПЗ.

### **Бібліометричний аналіз наукової літератури**

За період із 1987 до 2022 року до бази даних Scopus було включено 670 публікацій за ключовими словами «штучний інтелект» та «рак передміхурової залози». Відповідно до результатів бібліометричного аналізу, кількість публікацій на дану тему з 2016-го року постійно зростає, що свідчить про актуальність та перспективність застосування ШІ для діагностики РПЗ (рис. 1).

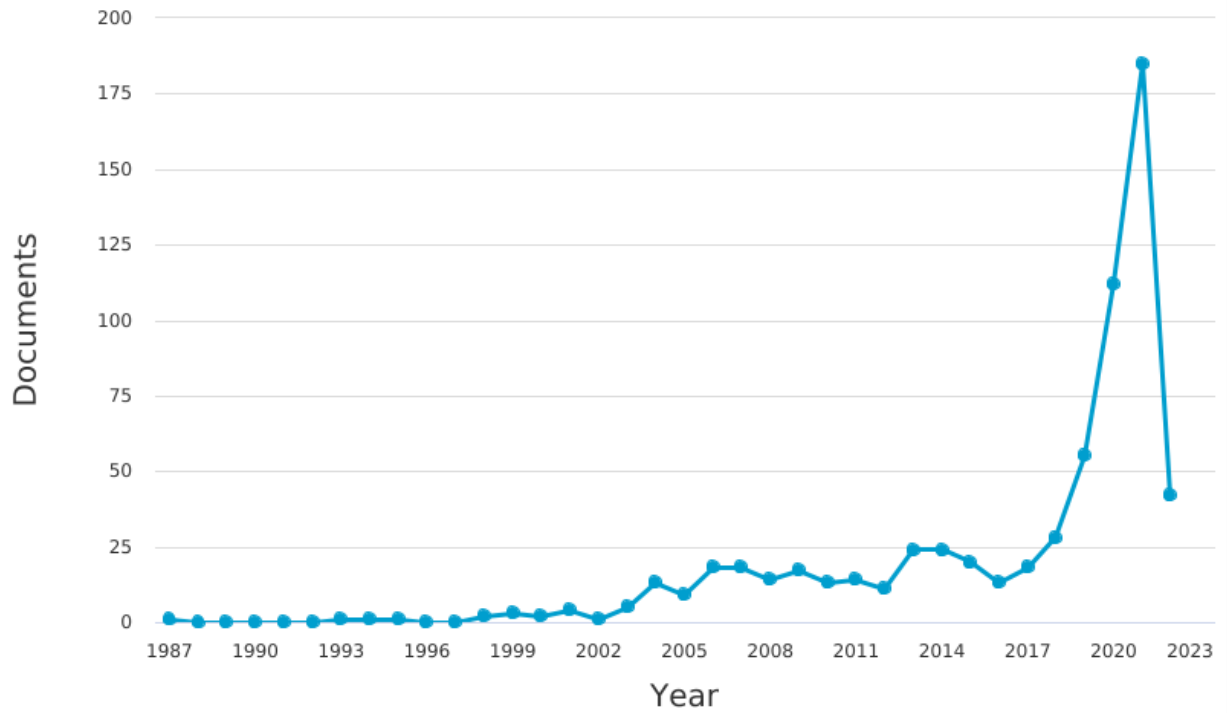


Рисунок 1 – Результат візуалізації хронології виходу публікацій на дану тему із використанням інструментів бібліометричного аналізу бази даних Scopus

Найпершими періодичними виданнями, у яких почали з'являтися публікації на дану тему є «Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics» та «BJU International», а в останні роки більшої популярності набрали «Cancers» та «Frontiers In Oncology».

Найбільш продуктивним та цитованим дослідником у галузі застосування ШІ для діагностики РПЗ можна назвати Anant Madabhushi (Case Western Reserve University, Cleveland, United States), який має 20 публікацій у базі даних Scopus на дану тему та індекс Гірша 58. Серед дослідників у цій галузі переважають вчені зі Сполучених Штатів Америки, Канади та Великої Британії, а серед найбільших спонсорських організацій можна виділити National Cancer Institute та National Institutes of Health (United States).

За результатами бібліометричного аналізу 670 публікацій у базі даних Scopus за ключовими словами «штучний інтелект» та «рак передміхурової залози», їх можна розподілити на 21 предметну область, переважна більшість

яких належить до медицини, комп'ютерних наук, біохімії, генетики та молекулярної біології (рис. 2).

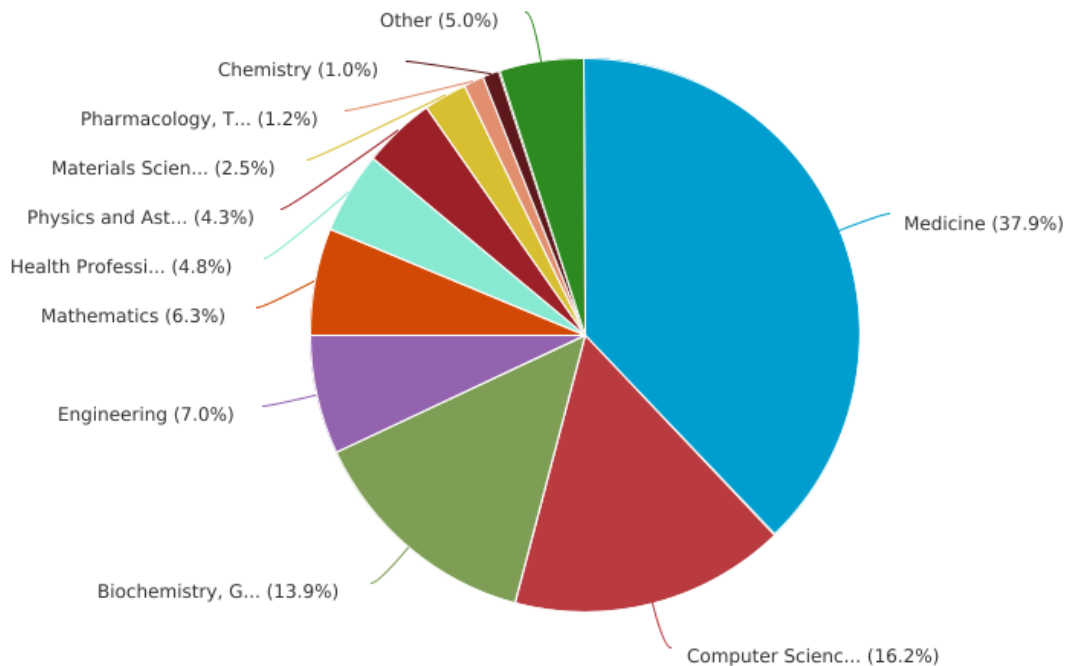


Рисунок 2 – Результат візуалізації розподілу публікацій на дану тему за предметними областями із використанням інструментів бібліометричного аналізу бази даних Scopus

За допомогою інструментів візуалізації бібліометричних мереж VOSviewer, 670 публікацій у базі даних Scopus за ключовими словами «штучний інтелект» та «рак передміхурової залози» можна розподілити на 3 тематичні кластери: ШІ, методи візуалізації, патогістологія РПЗ (рис. 3).

Також можна виділити 3 хронологічні етапи: дослідження за допомогою комп'ютерної та магнітно-резонансної томографії, застосування ШІ для автоматизації аналізу результатів методів візуалізації, початок залучення ШІ до аналізу патогістологічних зображень (рис. 4).

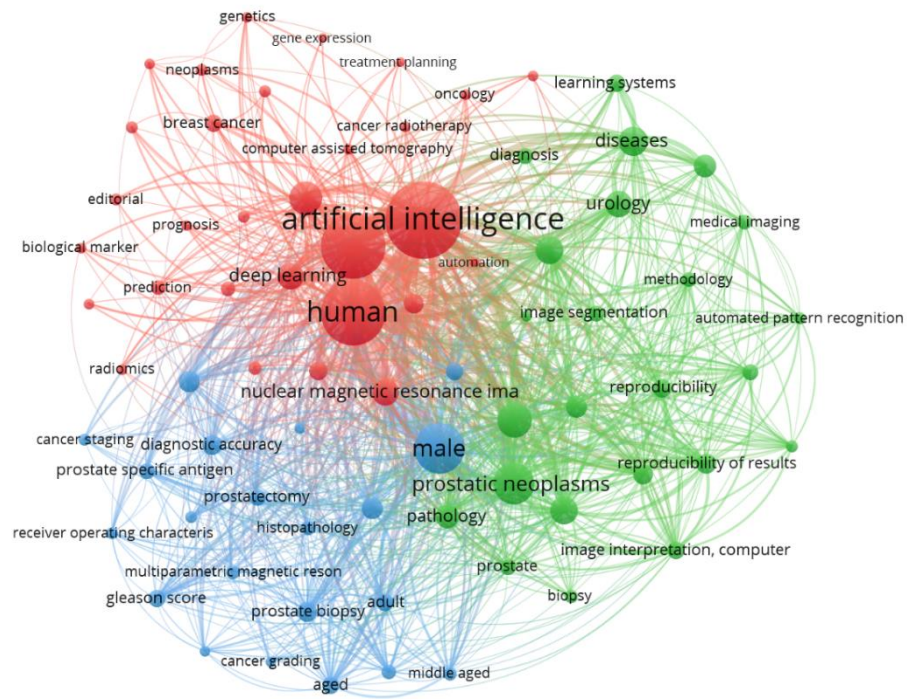


Рисунок 3 – Результат візуалізації закономірностей тематичного розподілу даної теми із використанням інструментів бібліометричного аналізу

VOSviewer

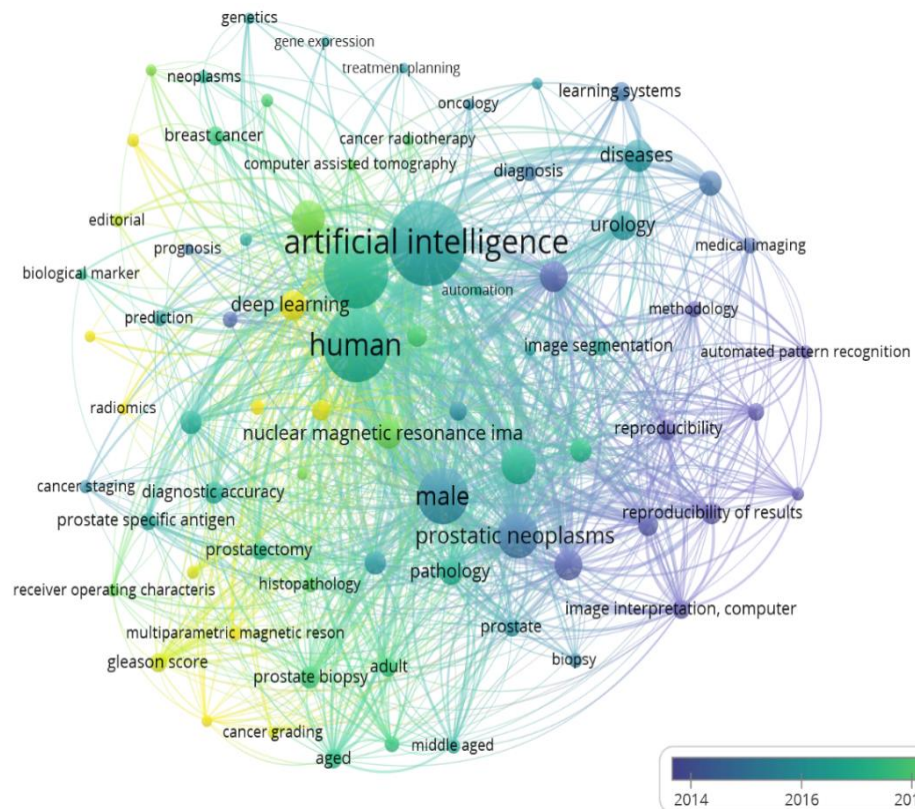


Рисунок 4 – Результат візуалізації закономірностей хронологічного розвитку даної теми із використанням інструментів бібліометричного аналізу

VOSviewer

## **Висновки до першого розділу**

Сьогодні ідея застосування ШІ для вирішення проблеми суб'єктивізму у патоморфологічній діагностиці захворювань набула значного поширення. Велика кількість програмних алгоритмів для аналізу зображень знаходиться у стадії розробки або тестування.

За результатами бібліометричного аналізу публікацій за ключовими словами «штучний інтелект» та «рак передміхурової залози» у базі даних Scopus опублікованих впродовж 1987-2022 років, було встановлено, що за останні 7 років значно збільшилась кількість статей у даній галузі, а найбільше досліджень проводиться в Сполучених Штатах Америки, Канаді та Великій Британії. Дані публікації можна розподілити на 21 предметну область, переважна більшість яких належить до медицини, комп'ютерних наук, біохімії, генетики та молекулярної біології.

За допомогою інструментів візуалізації бібліометричних мереж VOSviewer, 670 публікацій у базі даних Scopus за ключовими словами «штучний інтелект» та «рак передміхурової залози» можна розподілити на 3 тематичні кластери: штучний інтелект, методи візуалізації, патогістологія раку передміхурової залози. Також можна виділити 3 хронологічні етапи: дослідження за допомогою комп'ютерної та магнітно-резонансної томографії, застосування штучного інтелекту для автоматизації аналізу результатів методів візуалізації, початок залучення штучного інтелекту до аналізу патогістологічних зображень.

Отже, алгоритми ШІ зараз активно розвиваються та поступово займають свою нішу у діагностиці РПЗ. Не зважаючи на те, що на даний момент узгодженість між результатами автоматизованого аналізу та фахівцями-патологами не перевищує 70 %, безумовно, подальша безперервна взаємодія між людьми-спостерігачами та машинним навчанням має потенціал для автоматизації та стандартизації діагностики РПЗ.



## 2 СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ДІАГНОСТУВАННЯ ОНКОПАТОЛОГІЙ ЗА ГІСТОЛОГІЧНИМИ ЗОБРАЖЕННЯМИ

Застосування повно слайдової оптичної мікроскопії при гістологічних дослідженнях все ще залишається основним високо достовірним методом діагностування онкопатологій різного генезу. Основна перевага цього методу полягає в можливості отримати великий обсяг візуальної діагностичної інформації, але при цьому достовірність діагнозу суттєво залежить від рівня професійної підготовки та досвіду лікаря-гістолога.

Тому створення системи підтримки прийняття рішень (СППР) для діагностування онкопатологій за гістологічними зображеннями є актуальною задачею, оскільки дозволяє створити автоматизоване робоче місце лікаря-гістолога. Основним шляхом розв'язання цієї задачі є застосування для інформаційного синтезу діагностичної сппр ідей і методів машинного навчання. В дослідженні розглядається метод машинного навчання, який на відміну від існуючих методів інтелектуального аналізу даних дозволяє надати діагностичній СППР властивості адаптивності вирішальних правил до довільних умов формування гістологічних зображень і гнучкості при перенавчанні системи через розширення алфавіту класів розпізнавання, які характеризують різні структури тканини.

Моделювання когнітивних процесів, притаманних людині при формуванні та прийнятті класифікаційних рішень, є однією із найскладніших задач сучасності. Для вирішення цієї проблеми широко використовуються методи машинного навчання та розпізнавання образів [35, 36]. У праці [37] наведено приклад комп'ютеризованої системи діагностування раку молочної залози. У цій системі для двох класів розпізнавання (доброякісна та злоякісна пухлина) вхідними даними для машинного навчання і тестування були результати цитологічного дослідження тканини молочної залози. При цьому структурований вектор діагностичних ознак складався всього з дев'яти

характеристик клітин тканини молочної залози. Оскільки машинне навчання СППР здійснювалося за різними методами, то найбільшу достовірність для алфавіту із двох класів було отримано при реалізації методу опорних векторів [38]. Але варто зазначити, що у випадку збільшення потужності алфавіту класів розпізнавання при незмінному просторі діагностичних ознак збіжність методу опорних векторів не гарантується.

У праці [39] вхідними даними для машинного навчання та тестування багатошарової штучної мережи (ШНМ) був набір повно слайдових зображень, який складався із 400 мікроскопічних знімків тканин лімфовузлів. У результаті найкращий варіант машинного навчання ШНМ правильно розпізнавав 92,5 % пухлин на тестовому наборі слайдів, а лікар-патологоанатом тільки 73,3 %. Але основним недоліком ШНМ є чутливість до багатовимірності простору діагностичних ознак і алфавіту класів розпізнавання. У праці [40] для зменшення впливу багато вимірності пропонується використовувати побудовані на ШНМ екстрактори вхідних даних, але такий підхід неодмінно призводить до втрати інформації.

Таким чином, застосування традиційних методів інтелектуального аналізу даних Data Mining [41], включаючи багатозгорткові ШНМ [42,43], для інформаційного синтезу СППР для діагностування онкопатологій за гістологічними зображеннями, не завжди забезпечує високу функціональну ефективність машинного навчання через ряд науково-методологічних обмежень:

- 1) довільні умови формування гістологічних зображень;
- 2) суттєвий перетин класів розпізнавання в просторі діагностичних ознак;
- 3) багатовимірність простору діагностичних ознак повно слайдових гістологічних зображень;
- 4) вплив неконтрольованих збурюючих факторів.

Одним із перспективних підходів до інформаційного синтезу діагностичних СППР є використання ідей та методів так званої інформаційно-екстремальної інтелектуальної технології (ІЕІТ) аналізу даних, яка базується на максимізації інформаційної спроможності системи в процесі машинного навчання [44 – 46]. Ідея методів ІЕІТ як і в ШНМ полягає в адаптації в процесі машинного навчання вхідного математичного опису до максимальної повної ймовірності прийняття правильних діагностичних рішень. Але основна перевага методів інформаційно-екстремального машинного навчання полягає в тому, що вони на відміну від нейроподібних структур розробляються в рамках функціонального підходу до моделювання когнітивних процесів, притаманних людині при формуванні та прийнятті класифікаційних рішень.

Такий підхід на відміну від структурних методів дозволяє методам інформаційно-екстремального машинного навчання надати гнучкість при перенавчанні системи через розширення алфавіту класів розпізнавання. Крім того, побудовані в рамках геометричного підходу вирішальні правила практично вирішують проблему багато вимірності словника ознак розпізнавання. При цьому варто підкреслити, що методи інформаційно-екстремального машинного навчання здійснюються не в інтерактивному режимі, як це відбувається в ШНМ, а в автоматичному і потребують для формування навчальної матриці на порядок менше зразків зображень. В праці [46] наведено приклад інформаційного синтезу в рамках ІЕІТ діагностичної СППР для чотирьох класів розпізнавання. Але оскільки було реалізовано лінійний алгоритм інформаційно-екстремального машинного навчання, то не вдалося побудувати високо достовірні вирішальні правила.

Метою дослідження є підвищення функціональної ефективності СППР для діагностування онкопатологій за гістологічними зображеннями шляхом інформаційно-екстремального машинного навчання за ієрархічною структурою даних.

Розглянемо формалізовану постановку задачі ієрархічного інформаційно-екстремального машинного навчання СППР для діагностування

онкопатологій за гістологічними зображеннями. При цьому будемо розглядати ієрархічну структуру даних у вигляді так званого декурсивного бінарного дерева. Структуру даних у вигляді бінарного дерева будемо називати декурсивною, в якій атрибут з вершини верхнього ярусу передається у вершину своєї страти нижнього ярусу. У нашому випадку як атрибути вершин розглядаються навчальні матриці відповідних класів розпізнавання. Кінцеві страти, з яких атрибути не передаються, будемо називати фінальними. Таким чином, при збільшенні потужності алфавіту класів розпізнавання декурсивна ієрархічна структура розбивається на страти, кожна з яких складається із двох найближчих в бінарному просторі ознак Хеммінга класів розпізнавання. Це дозволяє для їх класифікації застосовувати лінійний алгоритм інформаційно-екстремального машинного навчання необхідної глибини. При цьому на відміну від нейроподібних структур глибина інформаційно-екстремального машинного навчання визначається не кількістю схованих шарів, а кількістю параметрів машинного навчання, що оптимізуються за інформаційним критерієм.

Нехай задано алфавіт класів розпізнавання у вигляді декурсивної ієрархічної структури  $\{X_{h,s,m}^o | h = \overline{1, N}; s = \overline{1, S}; m = \overline{1, 2}\}$ , де  $N$  – кількість ярусів ієрархічної структури;  $S$  – кількість страт на  $h$ -му ярусі. При цьому два класи розпізнавання кожної страти характеризують гістологічні зразки тканини з різними морфологічними структурами молочної залози. Для кожного класу розпізнавання шляхом оброблення гістологічних зображень сформовано вхідну навчальну матрицю яскравості  $\|y_{h,s,m,i}^{(j)} | i = \overline{1, N}, j = \overline{1, n}\|$ , де  $N$  – кількість діагностичних ознак в структурованому векторі класу розпізнавання  $X_{h,s,m}^o$ ;  $n$  – кількість структурованих векторів діагностичних ознак класу розпізнавання  $X_{h,s,m}^o$ , які утворюють навчальну матрицю.

Відомо, що концепція ІЕІТ полягає в перетворенні вхідної навчальної матриці типу «об'єкт – властивість» в задану в бінарному просторі Хеммінга робочу навчальну матрицю, яка в процесі машинного навчання шляхом

допустимих перетворень адаптується до максимально повної ймовірності прийняття правильних класифікаційних рішень. Нехай в просторі Хеммінга задано множину  $\{g_m\}$  структурованих векторів параметрів машинного навчання, які впливають на функціональну ефективність діагностичної СППР.

Обмежимося глибиною машинного навчання СППР другого рівня, яка дозволяє оптимізувати геометричні параметри гіперсферичних контейнерів і класів розпізнавання і систему контрольних допусків на діагностичні ознаки. У цьому випадку вектор параметрів машинного навчання для побудови вирішальних правил розпізнавання векторів діагностичних ознак, наприклад, класу  $X_{h,s,m}^0$  представимо у вигляді структури

$$g_{h,s} = \langle x_{h,s,m}, d_{h,s,m}, \delta_{h,s} \rangle \quad (1)$$

де  $x_{h,s,m}$  – двійковий усереднений вектор діагностичних ознак, вершина якого визначає геометричний центр контейнера класу розпізнавання  $X_{h,s,m}^0$ ,

$d_{h,s,m}$  – кодова відстань, яка визначає радіус гіперсферичного контейнера класу розпізнавання  $X_{h,s,m}^0$ ;

$\delta_{h,s}$  – параметр s-ї страти h-го ярусу декурсивного дерева, який дорівнює половині симетричного поля контрольних допусків на діагностичну ознаку.

При цьому задано обмеження на параметри машинного навчання:

1) областю значень радіуса контейнера класу розпізнавання  $X_{h,s,m}^0$  є:

$$d_{h,s,m} \in [0; d(x_{h,s,m} \oplus x_{h,s,c}) - 1],$$

де  $d(x_{h,s,m} \oplus x_{h,s,c})$  – кодова відстань Хеммінга між вектором діагностичних ознак  $x_{h,s,m}$  класу розпізнавання  $X_{h,s,m}^0$  і аналогічним вектором  $x_{h,s,m}$  найближчого сусіднього класу розпізнавання  $X_{h,s,m}^0$ ;

2) для двобічного симетричних поля допусків діагностичної ознаки має місце:

$$\delta_{h,s} \in [0; \delta_H / 2],$$

де  $\delta_H$  – поле нормованих допусків на діагностичні ознаки, яке визначає область значень відповідних контрольних допусків

На етапі машинного навчання необхідно:

1) оптимізувати параметри вектора (1) за усередненим за алфавітом класів розпізнавання інформаційним критерієм, який обчислюється на  $k$ -му кроці машинного навчання для  $s$ -ї страти  $h$ -го ярусу декурсивного дерева:

$$\bar{E}_{h,s}^{(k)} = \frac{1}{2} \sum_{m=1}^2 \max_{G_E \cap \{k\}} E_{h,s,m}^{(k)}, \quad (2)$$

де  $E_{h,s,m}^{(k)}$  – обчислений на  $k$ -му кроці машинного навчання критерій інформаційної спроможності СППР розпізнавати структуровані вектори діагностичних ознак класу  $X_{h,s,m}^0$ ;

$G_E$  – робоча (допустима) область визначення функції інформаційного критерію оптимізації параметрів машинного навчання;

$\{k\}$  – впорядкована множина кроків машинного навчання СППР.

2) за отриманими в процесі машинного навчання оптимальними геометричними параметрами контейнерів класів розпізнавання побудувати для кожної страти ієрархічної структури вирішальні правила, які гарантують високу повну ймовірність прийняття правильних діагностичних рішень.

3) на етапі екзамену необхідно прийняти класифікаційне рішення про належність вектора ознак до одного із класів розпізнавання відповідної фінальної страти.

У рамках ІЕІТ вхідний інформаційний опис діагностичної СППР складається із:

- 1) словника ознак розпізнавання, потужність якого визначається розміром кадра повно слайдового гістологічного зображення;
- 2) алфавіту класів розпізнавання, які характеризують різні морфологічні структури тканини молочної залози;
- 3) вхідної навчальної матриці яскравості пікселів кадру гістологічного зображення;
- 4) робочої бінарної навчальної матриці, яка в процесі машинного навчання адаптується до максимальної повної ймовірності прийняття правильних діагностичних рішень.

Гістологічне зображення морфології тканин із злоякісними новоутвореннями молочної залози було отримано за допомогою оптичної мікроскопії в Сумському обласному онкологічному центрі (Суми, Україна). Зображення мало розмір 1728x923 пікселів і розбивалося на 1400 кадрів розміром 54x54 пікселя, з яких шляхом експертної оцінки було сформовано алфавіт із чотирьох класів розпізнавання: клас  $X_1^0$  – гіперхромні ядра; клас  $X_2^0$  – інвазивний ріст клітин; клас  $X_3^0$  – скляні ядра і клас  $X_4^0$  – сполучна тканина. Представники цих класів розпізнавання показано на рисунку 5.

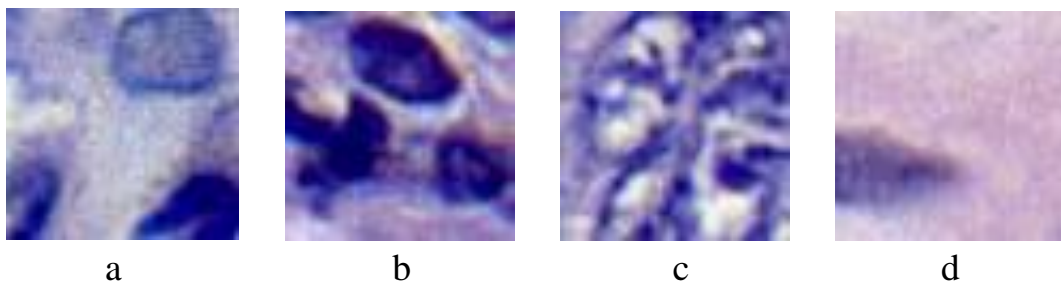


Рисунок 5 – Зображення кадрів зон інтересу: а – клас  $X_1^0$ ; б – клас  $X_2^0$ ; с – клас  $X_3^0$ ; d – клас  $X_4^0$

Вхідний інформаційний опис діагностичної СППР представимо у вигляді структури:

$$I_B = \langle G, T, \Omega, Z, H, Y^{|S|}, X^{|S|}; g, f_1, f_2 \rangle$$

де  $G$  – множина факторів, які впливають на діагностування;

$T$  – множина моментів часу зчитування інформації з гістологічного зображення;

$\Omega$  – простір діагностичних ознак;

$Z$  – алфавіт класів розпізнавання;

$H$  – ієрархічна структура даних у вигляді бінарного ієрархічного дерева;

$Y^{|S|}$  – множина вхідних (євклідових) навчальних матриць яскравості пікселів кадру гістологічного зображення для всіх страт декурсивного дерева;

$X^{|S|}$  – робоча бінарна навчальна матриця, трансформована в простір діагностичних ознак Хеммінга;

$g$  – оператор побудови декурсивного дерева;

$f_1$  – оператор формування множини  $Y^{|S|}$  навчальних матриць страт декурсивного дерева;

$f_2$  – оператор перетворення матриць  $Y^{|S|}$  в робочі бінарні матриці  $X^{|S|}$ .

Декартовий добуток множин  $G \times T \times \Omega \times Z$  задає джерело інформації.

Категорійну функціональну модель для другого рівня глибини інформаційно-екстремального машинного навчання діагностичної СППР за декурсивною ієрархічною структурою даних показано на рисунку 6.



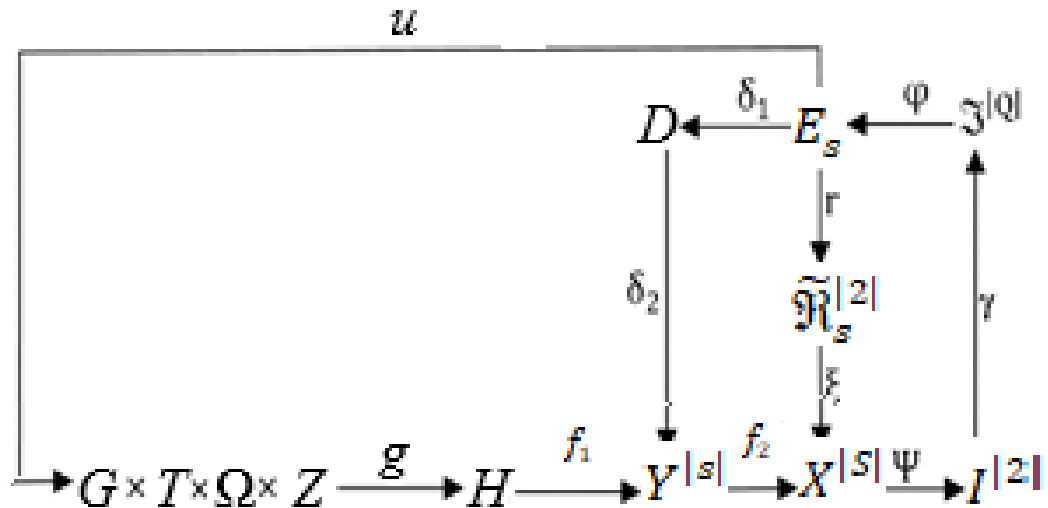


Рисунок 6 – Категорійна модель машинного навчання діагностичної СППР

Показаний на рисунку 6 оператор  $g$  з джерела інформації утворює декурсивне бінарне дерево  $H$ , а оператор  $f_1$  формує входні (євклідові) нечіткі навчальні матриці  $Y^{|S|}$  для всіх страт декурсивного дерева. Оператор  $f_2$  шляхом порівняння ознак розпізнавання із заданими контрольними допусками формує для всіх страт відповідно множину  $X^{|S|}$  бінарних робочих матриць, які в процесі машинного навчання адаптуються до максимальної повної ймовірності прийняття правильних класифікаційних рішень. Терм-множина  $E_S$ , елементами якої є обчислені на кожному кроці машинного навчання значення інформаційного критерію для кожної страти, є загальною для всіх контурів оптимізації параметрів навчання. Оператор  $\gamma: E \rightarrow \tilde{\mathfrak{R}}_S^{|2|}$  на кожному кроці машинного навчання відновлює в радіальному базисі Геммінгового простору діагностичних ознак контейнери класів розпізнавання, які утворюють для кожної страти розбиття  $\tilde{\mathfrak{R}}_S^{|2|}$ . Оператор  $\xi$  відображає розбиття  $\tilde{\mathfrak{R}}_S^{|2|}$  на розподіл двійкових векторів ознак робочої матриці  $X^{\|S\|}$ . Далі оператор  $\psi: X^{\|S\|} \rightarrow I^{|2|}$  формує множину  $I^{|2|}$  двох альтернативних статистичних гіпотез шляхом перевірки основної статистичної гіпотези  $\gamma_1: x_{h,s,n}^{(j)} \in X_{h,s,m}^o$ . Оператор  $\gamma$

визначає множину  $\mathfrak{Z}^{\|Q\|}$  точнісних характеристик діагностичних рішень, де  $Q = C^2$ , а оператор  $\varphi$  обчислює множину  $E$  значень інформаційного критерію оптимізації, якій є функціоналом від точнісних характеристик. Контур оптимізації контрольних допусків замикається через терм-множину  $D$ , елементами якої є значення контрольних допусків на діагностичні ознаки. Оператор  $u_H$  регламентує процес машинного навчання.

Таким чином, інформаційно-екстремальне машинне навчання діагностичної СППР полягає в цілеспрямованому пошуку глобального максимуму інформаційного критерію (2) оптимізації геометричних параметрів контейнерів класів розпізнавання, які відновлюються в радіальному базисі Хеммінгового простору діагностичних ознак.

При збільшенні потужності алфавіту класів розпізнавання декурсивна ієрархічна структура розбивається на фінальні страти, кожна з яких складається із двох найближчих в бінарному просторі ознак класів розпізнавання. Це дозволяє для їх класифікації застосовувати лінійний алгоритм інформаційно-екстремального машинного навчання необхідної глибини для побудови високо достовірних вирішальних правил. Згідно з категорійною моделлю (рис. 7) інформаційно-екстремальний алгоритм машинного навчання СППР з оптимізацією параметрів машинного навчання (1), представимо у вигляді ітераційної процедури пошуку глобального максимуму усередненого за алфавітом класів розпізнавання інформаційного критерію (2) для  $s$ -ї страти  $h$ -го ярусу декурсивного дерева:

$$\delta_{h,s}^* = \arg \max_{G_\delta} \{ \max_{G_E \cap (k)} \bar{E}_{h,s}^{(k)} \} \quad (3)$$

де  $\delta_{h,s}^*$  – оптимальний ( тут і далі в інформаційному розумінні) параметр поля контрольних допусків для класів розпізнавання  $s$ -ї страти  $h$ -го ярусу декурсивного дерева;

$G_\delta$  – допустима область значень параметра  $\delta_{h,s}$ .

Вхідною інформацією для алгоритму інформаційно-екстремального машинного навчання є масив  $\{y_{h,s,m,i}^{(j)}\}$  і система полів нормованих допусків  $\{\delta_{H,i}\}$  на діагностичні ознаки, яка задає область значень відповідних контрольних допусків. Розглянемо схему алгоритму інформаційно-екстремального машинного навчання з оптимізацією контрольних допусків на діагностичні ознаки за процедурою (3):

- 1) онулення лічильника класів розпізнавання:  $m:=0$
- 2)  $m:m+1$
- 3) онулення лічильника зміни параметра  $\delta_{h,s}$ :  $\delta_{h,s} = 0$ ;
- 4)  $\delta_{h,s} = \delta_{h,s} + 1$ ;
- 5) обчислюються нижні  $A_{H,i}$  і верхні  $A_{B,i}$  контрольні допуски на діагностичні ознак відповідно за правилами

$$A_{H,i} = y_{h,s,m,i} - \delta_{h,s}; \quad A_{B,i} = y_{h,s,m,i} + \delta_{h,s},$$

де  $y_{h,s,m,i}$  – усереднене значення  $i$ -ї діагностичної ознаки структурованого вектора ознак  $x_{h,s,m}$ ;

- 6) онулення лічильника кроків зміни радіуса гіперсферичного контейнера:  $k:=0$ ;
- 7)  $k:=k+1$ ;
- 8) формується бінарна навчальна матриця  $\{x_{h,s,m}\}$ , елементи якої обчислюються за правилом

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

9) формування масиву усереднених двійкових векторів-реалізацій  $\{x_{h,s,m}\}$ , елементи яких визначаються за правилом

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

де  $\rho_m$  – рівень селекції координат двійкового вектору  $\{x_{h,s,m}\}$ , який за замовчуванням дорівнює 0,5.

10) обчислюється інформаційний критерій оптимізації параметрів машинного навчання;

11) якщо  $k \leq N$ , то виконується пункт 7, інакше – пункт 12;

12) якщо  $\delta < \delta_H$ , то виконується пункт 4, інакше – пункт 13;

13) визначається максимальне значення інформаційного критерію в робочій області визначення його функції;

14) якщо  $m < M-1$ , то реалізується пункт 2, інакше – пункт 15;

15) визначається глобальний максимум усередненого інформаційного критерію  $\bar{E}_{h,s}^*$  в робочій області визначення його функції;

16) визначаються оптимальні значення параметра  $\delta_{h,s}^*$  і відповідно нижніх  $A_{H,i}^*$  і верхніх  $A_{B,i}^*$  контрольних допусків на всі діагностичні ознаки

Як критерій оптимізації параметрів машинного навчання СППР використовувалася модифікована інформаційна міра Кульбака у вигляді

$$E_{h,s,m}^{(k)} = \frac{1}{n_{\min}} \log_2 \left\{ \frac{2n_{\min} + 10^{-r} - [K_{1,h,s,m}^{(k)} + K_{2,h,s,m}^{(k)}]}{[K_{1,h,s,m}^{(k)} + K_{2,h,s,m}^{(k)}] + 10^{-r}} \right\} \\ [n_{\min} - (K_{1,h,s,m}^{(k)} + K_{2,h,s,m}^{(k)})], \quad (4)$$

де  $K_{1,h,s,m}^{(k)}$  – кількість подій, при яких вектори ознак класу розпізнавання  $X_{h,s,m}^0$  помилково до нього не відносяться;

$K_{2,h,s,m}^{(k)}$  – кількість подій, при яких помилково відносяться до класу розпізнавання  $X_{h,s,m}^0$  вектори ознак сусіднього в страті класу розпізнавання;

$n_{\min}$  – мінімальний обсяг репрезентативної навчальної вибірки;

$10^{-r}$  – достатньо мале число, що вводиться для уникнення поділу на нуль.

Для підвищення функціональної ефективності діагностування на другому рівні глибини інформаційно-екстремального машинного навчання СППР було реалізовано алгоритм послідовної оптимізацією контрольних допусків. При цьому отримані на етапі паралельної оптимізації контрольні допуски приймалися як стартові при послідовній оптимізації, яка здійснювалася за процедурою.

$$\delta_{h,s}^* = \arg \bigoplus_{l=1}^L \left\{ \max_{G_{\delta i}} \left[ \frac{1}{2} \sum_{m=1}^2 \max_{G_{Em} \cap \{k\}} E_{h,s,m}^{(k)}(l) \right] \right\} \quad i = \overline{1, N}, \quad (5)$$

де  $\bigoplus$  – символ операції повторення;

$L$  – кількість прогонів процедури послідовної оптимізації контрольних допусків, обумовлених неоптимальними стартовими величинами контрольних допусків для всіх ознак;

$E_{h,s,m}^{(k)}(l)$  – обчислений на  $k$ -му кроці  $l$ -го прогону машинного навчання інформаційний критерій.

За отриманими в процесі машинного навчання оптимальними геометричними параметрами контейнерів класів розпізнавання було побудовано вирішальні правила:

$$(\forall X_{h,s,m}^o \in \mathfrak{R}_{h,s}^{[2]}) (\forall x^{(j)} \in \mathfrak{R}_{h,s}^{[2]}) [if (\mu_{h,s,m} > 0) \& (\mu_{h,s,m} = \max_{\{m\}} \{\mu_{h,s,m}\})] \\ then x^{(j)} \in X_{h,s,m}^o \text{ else } x^{(j)} \notin X_{h,s,m}^o ],$$

де  $x^{(j)}$  – вектор, що розпізнається;

$\mu_{h,s,m}$  – функція належності вектора  $x^{(j)}$  контейнеру класу розпізнавання  $X_{h,s,m}^o$ .

У виразі (6) функція належності для гіперсферичного контейнера класу розпізнавання  $X_{h,s,m}^o$  визначається за формулою.

$$\mu_{h,s,m} = 1 - \frac{d(x_{h,s,m}^* \oplus x^{(j)})}{d_{h,s,m}^*},$$

де  $x_{h,s,m}^*$ ,  $d_{h,s,m}^*$  – оптимальні геометричні параметри машинного навчання: усереднена двійкова реалізація і радіус гіперсферичного контейнера відповідно.

Для автоматичної побудови ієрархічної структури у вигляді бінарного декурсивного дерева для вище наведених класів розпізнавання було сформовано варіаційний ряд за збільшенням середньої яскравості навчальних матриць кадрів зображення морфології тканини, як на рисунку 5. Як класи розпізнавання страти першого ярусу (за дендрологічною класифікацією таким є верхній ярус) було обрано найбільш віддалені класи  $X_1^o$  і  $X_4^o$ . Страти нижнього ярусу формувалися за принципом найближчого сусіда.

На рисунку 7 показана ієрархічна структура даних для чотирьох класів розпізнавання у вигляді бінарного декурсивного дерева.

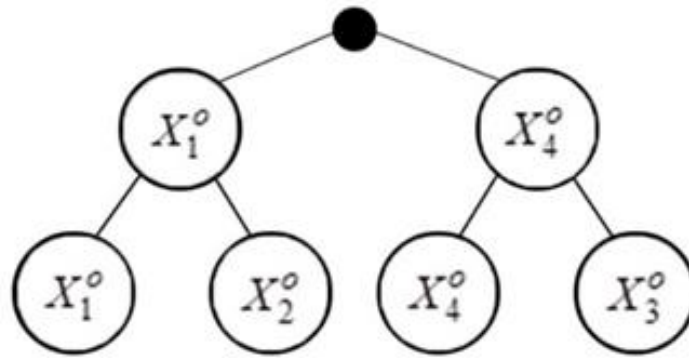


Рисунок 7 – Декурсивна структура даних

Для досягнення високої повної ймовірності прийняття правильних діагностичних рішень необхідно надати діагностичній СППР властивість персоналізації. Тобто СППР повинна діагностувати не за усередненими даними для групи різних пацієнтів, а враховувати індивідуальні показники кожного пацієнта. З цією метою було обчислено середні значення яскравості кожного кадру, які утворили варіаційний ряд за їх зростанням. Фрагмент варіаційного ряду із представників кожного із чотирьох класів розпізнавання представлено на рисунку 8.

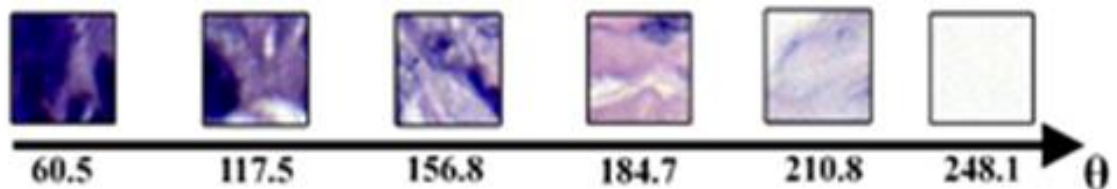


Рисунок 8 – Впорядкована послідовність кадрів

Спочатку визначалися атрибути класів розпізнавання  $X_1^o$  і  $X_4^o$ , які входять у першу страту ієрархічної структури (рис. 8). Для їх знаходження варіаційний ряд, в якому найменше усереднене значення яскравості дорівнювало 60.5 а найбільше 248, було розбито на 26 інтервалів, кожен із яких мав розмір 7 одиниць яскравості. Наступним кроком вибиралися кадри, яскравість яких була приблизно середньою для відповідного інтервалу. У результаті була сформована вибірка із усереднених значень яскравості кадрів. Далі вибірка

розбивалася на дві частини у місці найбільшої різниці між значеннями середньої яскравості кадрів (рис. 9).

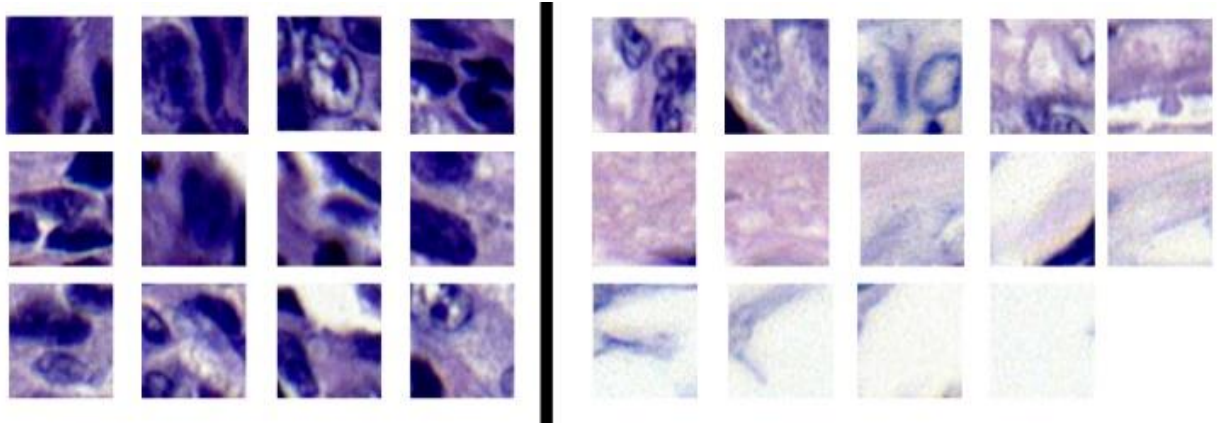


Рисунок 9 – Вибірка із кадрів морфології тканини і клітин

Останнім етапом відбувається перебір усіх можливих пар з різних частин вибірки (рис. 9). При цьому виконується процес інформаційно-екстремального машинного навчання спр з паралельною оптимізацією системи контрольних допусків на діагностичні ознаки за процедурою (3). Пара кадрів, яка має максимальне значення інформаційного критерію оптимізації (5), тобто ці кадри мають найбільшу міжкласову відстань, поміщається в декурсивну структуру даних (рис. 7) відповідно як атрибути класів розпізнавання  $X_1^o$  і  $X_4^o$  (рис. 10).

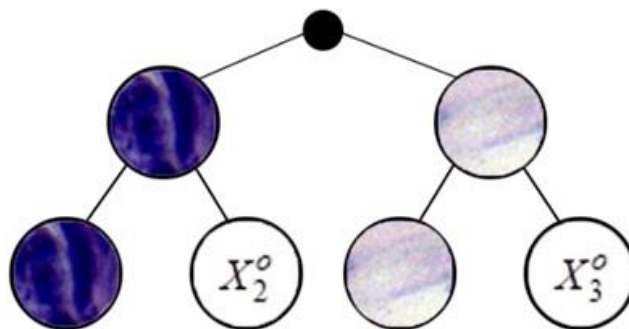


Рисунок 10 – Частково заповнена ієрархічна структура даних



Далі розглянемо визначення в ієрархічній структурі (рис. 7) атрибутів вершин класів розпізнавання  $X_2^0$  і  $X_3^0$ . Оскільки в декурсивному дереві атрибут вершини верхнього ярусу передається у відповідну страту нижнього ярусу, то необхідно вибирати репрезентативний атрибут вже не з 1400 кадрів, а лише з кадрів, які відносяться до цих класів. Оскільки атрибути страт нижніх ярусів обираються за принципом найближчого сусіда, то за атрибут класу розпізнавання  $X_2^0$  обирається навчальна матриця кадру за мінімальним максимальним значенням інформаційного критерію оптимізації, що свідчить про найбільшу схожість цього кадру з репрезентативним кадром класу розпізнавання  $X_1^0$ . Для другої фінальної страти відповідно обирається атрибут вершини класу розпізнавання  $X_3^0$ . У випадку розширення алфавіту класів розпізнавання за аналогічною процедурою визначаються атрибути вершин всіх страт нижніх ярусів. На рисунку 11 показано кінцевий вигляд ієрархічної декурсивної структури (рис. 7) з атрибутами всіх класів розпізнавання.

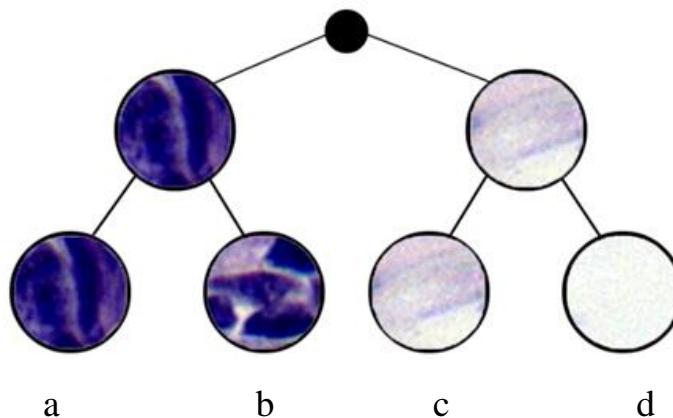


Рисунок 11 – Ієрархічна структура даних: а – клас  $X_1^0$ ; б – клас  $X_2^0$ ; с – клас  $X_4^0$ ;  
д – клас  $X_3^0$

Таким чином, формування ієрархічної структури даних у вигляді декурсивного бінарного дерева дозволяє розбити множину класів розпізнавання на пари найближчих сусідніх класів, для яких оптимізація здійснюється за вище наведеним лінійним алгоритмом інформаційно-

екстремального машинного навчання. При перенавчанні діагностичної спр через появу нового класу розпізнавання формується новий варіаційний ряд усереднених значень яскравості кадрів розширеного алфавіту і будується ієрархічна структура даних за розглянутою схемою.

Перевірка функціональної ефективності інформаційно-екстремального машинного навчання здійснювалася при функціонуванні спр в режимі екзамену за геометричними вирішальними правилами (6). При цьому алгоритм екзамену аналогічний алгоритму функціонування спр безпосередньо в режимі діагностування.

Інформаційний синтез діагностичної СППР здійснювався шляхом інформаційно-екстремального машинного навчання з оптимізацією контрольних допусків на діагностичні ознаки. Вхідна навчальна матриця формувалася шляхом скринінгу в декартовій системі координат кадрів гістологічного зображення (рис. 6). Як ознаки розпізнавання розглядалися значення яскравості RGB-складових, які зчитувалися в кожному пікселі рецепторного поля кадру. При цьому навчальна матриця для кожного класу розпізнавання складалася із 54 структурованих за часом зчитування векторів, кожний з яких мав 162 діагностичні ознаки. Інформаційно-екстремальне машинне навчання здійснювалося за ієрархічною структурою даних у вигляді декурсивного дерева (рис. 11). Для кожної фінальної страти визначалися оптимальні параметри машинного навчання (6).

На рисунку 12 показано графіки залежності усередненого нормованого критерію (4) від параметра поля контрольних допусків, отримані для кожної фінальної страти нижнього ярусу при реалізації алгоритму інформаційно-екстремального машинного навчання за процедурою (3) паралельної оптимізації контрольних допусків діагностичних ознак.

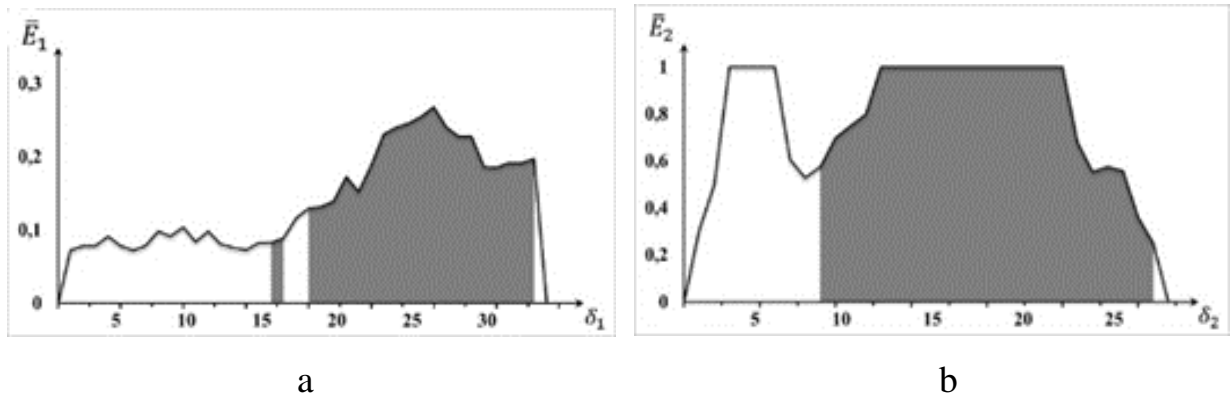


Рисунок 12 – Графіки залежності інформаційного критерію від параметра поля контрольних допусків: а – перша страта; б – друга страта

На рисунку 12 темними ділянками позначено робочі (допустимі) області визначення функції інформаційного критерію (4), в якій помилки першого та другого роду менше відповідно першої та другої достовірностей. Аналіз рисунку 12а показує, що в процесі інформаційно-екстремального машинного навчання за навчальними матрицями класів розпізнавання першої фінальної страти оптимальне значення параметра поля контрольних допусків: дорівнює  $\delta_1^* = 27$  (тут і далі в градаціях яскравості) при значенні усередненого за алфавітом класів розпізнавання цієї страти інформаційного критерію  $\bar{E}_1 = 0,26$ . На рисунку 12б максимальні значення інформаційного критерію належать ділянці типу плато. У цьому випадку за оптимальне береться значення параметра  $\delta_2$  при мінімальному значенню коефіцієнта  $\eta_\delta$ , який характеризує ступінь перетину двох класів розпізнавання:

$$\eta_s = \frac{\bar{d}_s(\delta)}{d[x_{h,s,1}(\delta) \oplus x_{h,s,2}(\delta)]} \rightarrow \min_{\{\delta\}}, \quad (7)$$

де  $\bar{d}_s(\delta)$  – усереднений радіус контейнерів класів розпізнавання  $X_{h,s,m}^0$ ;

$d[x_{h,s,1}(\delta) \oplus x_{h,s,2}(\delta)]$  – кодова міжцентрова відстань класів розпізнавання  $X_{h,s,1}^0$  і  $X_{h,s,2}^0$ .

Із урахуванням умови (7) для другої страти оптимальне значення параметра поля контрольних допусків дорівнює  $\delta^* = 18$  при максимальному граничному значенні нормованого інформаційного критерію  $\bar{E}_2 = 1$ .

Оскільки для першої фінальної страти (рис. 12.а) було досягнуто невисоке максимальне значення інформаційного критерію, то додатково було виконано послідовну оптимізацію системи контрольних допусків за процедурою (5). При цьому отримані на етапі паралельної оптимізації контрольні допуски приймалися як стартові при послідовній оптимізації. На рисунку 13 показано графік зміни усередненого нормованого інформаційного критерію (5) в процесі інформаційно-екстремального машинного навчання діагностичної СППР з послідовною оптимізацією системи контрольних допусків на діагностичні ознаки.

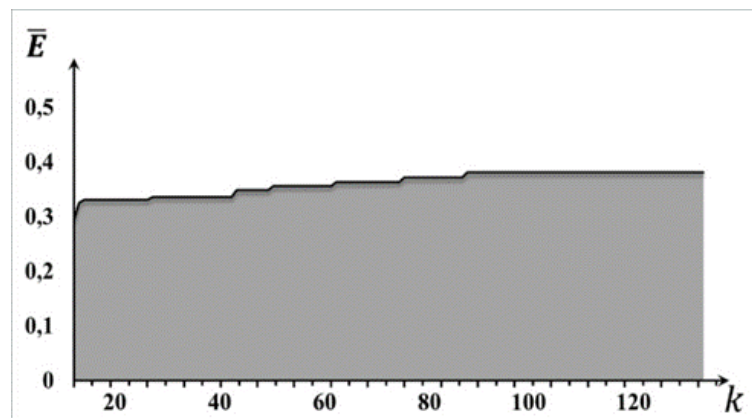


Рисунок 13 – Графік зміни інформаційного критерію в процесі послідовної оптимізації контрольних допусків для першої фінальної страти

Отриманий результат (рис. 13) показує, що інформаційний критерій оптимізації досягнув максимального значення  $\bar{E}^* = 0.38$  вже на другому прогоні процедури (5). При цьому порядковий номер прогону визначається відношенням кількості ітерацій  $k$  (кроків машинного навчання) до кількості ознак  $N$  в структурованому векторі ознак розпізнавання. Таким чином, в процесі машинного навчання з послідовною оптимізацією контрольних

допусків значення інформаційного критерію все ще залишається невисоким, але воно збільшилося у порівнянні з процедурою (4) паралельної оптимізації контрольних допусків.

Для побудови вирішальних правил (6) необхідно знання геометричних параметрів контейнерів класів розпізнавання. На рисунку 14 показано графіки залежності інформаційного критерію (5) від радіусів контейнерів класів розпізнавання фінальних страт декурсивного дерева.

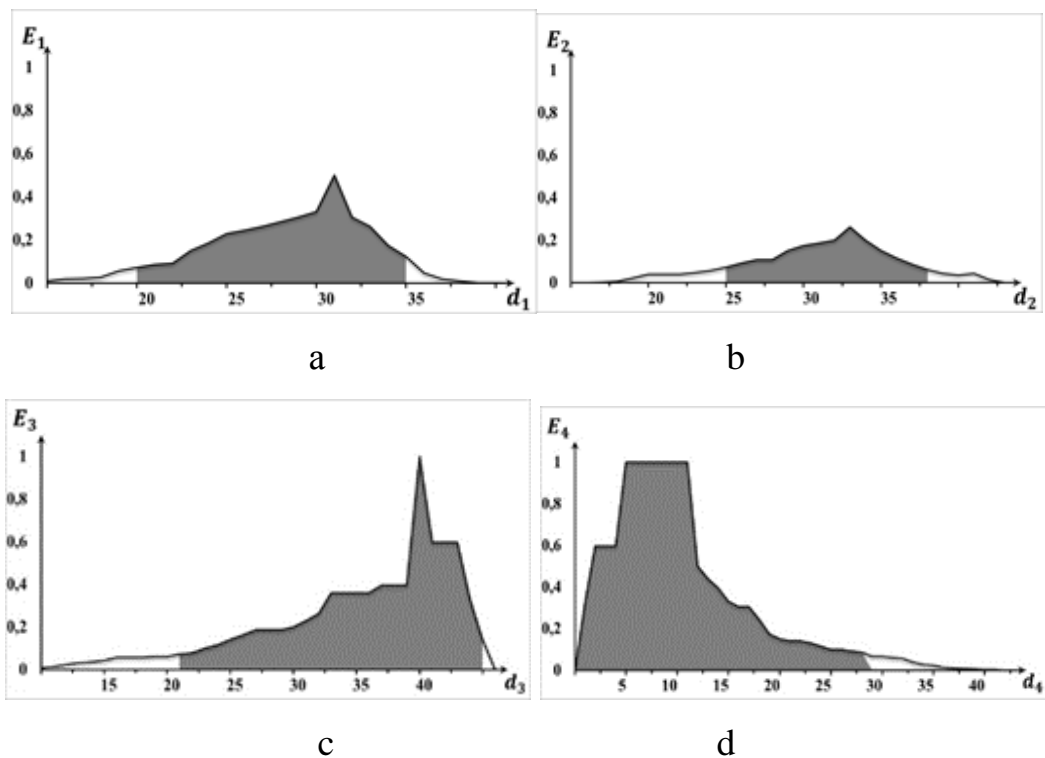


Рисунок 14 –. Графіки залежності інформаційного критерію оптимізації від радіусів контейнерів класів розпізнавання: а – клас  $X_1^0$ ; б– клас  $X_2^0$ ; с – клас  $X_3^0$ ; d – клас  $X_4^0$ ;

Аналіз рисунку 14 показує, що оптимальні радіуси контейнерів класів розпізнавання дорівнюють:  $d_1^* = 31$  (тут і далі в кодових одиницях Геммінга) для класу  $X_1^0$ ;  $d_2^* = 33$  для класу  $X_2^0$ ;  $d_3^* = 40$  для класу  $X_3^0$  і  $d_4^* = 8$  для класу  $X_4^0$ . При цьому середнє значення нормованого інформаційного критерію дорівнює  $\bar{E} = 0.7$ .

Перевірка функціональної ефективності машинного навчання СППР здійснювалася в режимі екзамену. На рисунку 15 показано оцифроване гістологічне зображення, отримане за вирішальними правилами (6). На рисунку цифри відповідають порядковим номерам класів розпізнавання.

Візуальний аналіз рисунків 6 і 15 підтверджує достатньо високу достовірність розпізнавання кадрів гістологічного зображення.

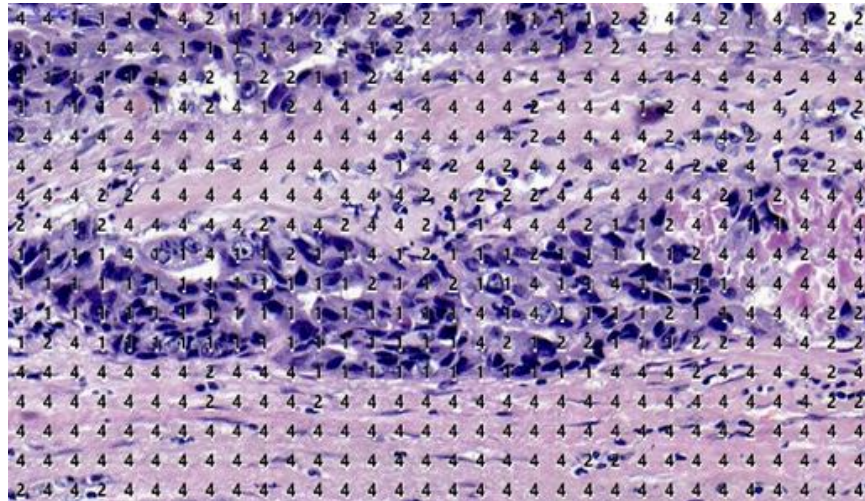


Рисунок 15 – Оцифроване гістологічне зображення результатів ідентифікації кадрів

Відсутність високостовірних СППР для діагностування онкопатологій за гістологічними зображеннями обумовлена недоліками сучасних методів машинного навчання, включаючи нейроподібні структури, що особливо проявляється при аналізі великих обсягів даних і алфавітів класів розпізнавання великої потужності. Перспективним напрямом підвищення функціональної ефективності СППР для діагностування онкопатологій за гістологічними зображеннями є застосування функціонального підходу до моделювання когнітивних процесів природнього інтелекту при формуванні та прийнятті класифікаційних рішень. У рамках такого підходу розроблено вище наведений метод інформаційно-екстремального машинного навчання СППР для діагностування онкопатологій за зображеннями морфології тканини, який є альтернативним нейроподібним структурам і усуває їх основні недоліки.

Побудовані за результатами ієрархічного інформаційно-екстремального машинного навчання вирішальні правила характеризуються достатньо високою достовірністю для чотирьох класів розпізнавання. Але аналіз рисунку 10 показує, що вирішальні правила не безпомилкові за навчальною матрицею, оскільки не вдалося досягнути максимально граничного значення інформаційного критерію оптимізації параметрів машинного навчання. Цей факт підтверджується візуальним аналізом сегментованого зображення (рис. 15) морфологічної структури тканини.

Тому предметом подальших досліджень є підвищення функціональної ефективності діагностичної СППР шляхом збільшення глибини інформаційно-екстремального машинного навчання. При цьому серед додаткових параметрів оптимізації важливу роль будуть відігравати параметри формування вхідного інформаційного опису діагностичної СППР. Такими параметрами машинного навчання можуть бути, наприклад, розмір кадру зображення, вагові коефіцієнти RGB-складових гістологічного зображення [46] тощо. Крім того, перспективним напрямом підвищення достовірності сегментації гістологічного зображення є оцінка інформативності діагностичних ознак за методом інформаційно-екстремального машинного навчання, запропонованим у праці авторів [47].

### **Висновки до другого розділу**

У рамках функціонального підходу до моделювання когнітивних процесів природнього інтелекту вперше розроблено метод інформаційно-екстремального машинного навчання діагностичної СППР за ієрархічною структурою даних у вигляді декурсивного бінарного дерева. Запропонований метод дозволяє автоматично формувати вхідну навчальну матрицю, враховувати індивідуальні особливості пацієнта і перенавчати систему при розширенні алфавіту класів розпізнавання. У результаті глибоке інформаційно-екстремальне машинне навчання діагностичної СППР здійснюється для кожної пари найближчих сусідніх класів лінійним алгоритмом.

Одержані результати дозволяють створити автоматизоване робоче місце лікаря-гістолога для діагностування онкопатологій різного патогенезу. Побудовані вирішальні правила є не безпомилковими за навчальною матрицею. Тому для підвищення функціональної ефективності інформаційно-екстремального машинного навчання діагностичної СППР необхідно збільшити глибину машинного навчання шляхом оптимізації додаткових параметрів, включаючи параметри формування вхідної навчальної матриці.



## ВИСНОВКИ

Сьогодні ідея застосування ШІ для вирішення проблеми суб'єктивізму у патоморфологічній діагностиці захворювань набула значного поширення. Велика кількість програмних алгоритмів для аналізу зображень знаходиться у стадії розробки або тестування.

За результатами бібліометричного аналізу публікацій за ключовими словами «штучний інтелект» та «рак передміхурової залози» у базі даних Scopus опублікованих впродовж 1987-2022 років, було встановлено, що за останні 7 років значно збільшилась кількість статей у даній галузі, а найбільше досліджень проводиться в Сполучених Штатах Америки, Канаді та Великій Британії. Дані публікації можна розподілити на 21 предметну область, переважна більшість яких належить до медицини, комп'ютерних наук, біохімії, генетики та молекулярної біології.

За допомогою інструментів візуалізації бібліометричних мереж VOSviewer, 670 публікацій у базі даних Scopus за ключовими словами «штучний інтелект» та «рак передміхурової залози» можна розподілити на 3 тематичні кластери: штучний інтелект, методи візуалізації, патогістологія раку передміхурової залози. Також можна виділити 3 хронологічні етапи: дослідження за допомогою комп'ютерної та магнітно-резонансної томографії, застосування штучного інтелекту для автоматизації аналізу результатів методів візуалізації, початок залучення штучного інтелекту до аналізу патогістологічних зображень.

Отже, алгоритми ШІ зараз активно розвиваються та поступово займають свою нішу у діагностиці РПЗ. Не зважаючи на те, що на даний момент узгодженість між результатами автоматизованого аналізу та фахівцями-патологами не перевищує 70 %, безумовно, подальша безперервна взаємодія між людьми-спостерігачами та машинним навчанням має потенціал для автоматизації та стандартизації діагностики РПЗ.

У рамках функціонального підходу до моделювання когнітивних процесів природнього інтелекту вперше розроблено метод інформаційно-екстремального машинного навчання діагностичної СППР за ієрархічною структурою даних у вигляді декурсивного бінарного дерева. Запропонований метод дозволяє автоматично формувати вхідну навчальну матрицю, враховувати індивідуальні особливості пацієнта і перенавчати систему при розширенні алфавіту класів розпізнавання. У результаті глибоке інформаційно-екстремальне машинне навчання діагностичної СППР здійснюється для кожної пари найближчих сусідніх класів лінійним алгоритмом.

Одержані результати дозволяють створити автоматизоване робоче місце лікаря-гістолога для діагностування онкопатологій різного патогенезу. Побудовані вирішальні правила є не безпомилковими за навчальною матрицею. Тому для підвищення функціональної ефективності інформаційно-екстремального машинного навчання діагностичної СППР необхідно збільшити глибину машинного навчання шляхом оптимізації додаткових параметрів, включаючи параметри формування вхідної навчальної матриці.

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ**

1. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. *Eur Urol.* 2016;70(1):106-119. doi:10.1016/J.EURURO.2016.02.028
2. Knudsen BS, Vasioukhin V. Mechanisms of prostate cancer initiation and progression. *Adv Cancer Res.* 2010;109(C):1-50. doi:10.1016/B978-0-12-380890-5.00001-6
3. Bianco-Miotto T, Chiam K, Buchanan G, et al. Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2611-2622. doi:10.1158/1055-9965.EPI-10-0555
4. Veltri RW, Christudass CS. Nuclear morphometry, epigenetic changes, and clinical relevance in prostate cancer. *Adv Exp Med Biol.* 2014;773:77-99. doi:10.1007/978-1-4899-8032-8\_4
5. Carleton NM, Lee G, Madabhushi A, Veltri RW. Advances in the computational and molecular understanding of the prostate cancer cell nucleus. *J Cell Biochem.* 2018;119(9):7127-7142. doi:10.1002/JCB.27156
6. Van Booven DJ, Kuchakulla M, Pai R, et al. A Systematic Review of Artificial Intelligence in Prostate Cancer. *Res reports Urol.* 2021;13:31-39. doi:10.2147/RRU.S268596
7. Costa DN, Pedrosa I, Donato F, Roehrborn CG, Rofsky NM. MR Imaging-Transrectal US Fusion for Targeted Prostate Biopsies: Implications for Diagnosis and Clinical Management. *Radiographics.* 2015;35(3):696-708. doi:10.1148/RG.2015140058
8. Samtani S, Burotto M, Roman JC, Cortes-Herrera D, Walton-Diaz A. MRI and Targeted Biopsy Essential Tools for an Accurate Diagnosis and Treatment Decision Making in Prostate Cancer. *Diagnostics (Basel, Switzerland).* 2021;11(9). doi:10.3390/diagnostics11091551

9. Greer MD, Lay N, Shih JH, et al. Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study. *Eur Radiol.* 2018;28(10):4407-4417. doi:10.1007/S00330-018-5374-6
10. Egevad L, Cheville J, Evans AJ, et al. Pathology Imagebase-a reference image database for standardization of pathology. *Histopathology.* 2017;71(5):677-685. doi:10.1111/HIS.13313
11. Epstein JI, Amin MB, Fine SW, et al. The 2019 genitourinary pathology society (GUPS) white paper on contemporary grading of prostate cancer. *Arch Pathol Lab Med.* 2021;145(4):461-493. doi:10.5858/arpa.2020-0015-ra
12. Epstein JI. Prostate cancer grading: a decade after the 2005 modified system. *Mod Pathol.* 2018;31(S1):47-63. doi:10.1038/MODPATHOL.2017.133
13. Kwak JT, Hewitt SM. Multiview boosting digital pathology analysis of prostate cancer. *Comput Methods Programs Biomed.* 2017;142:91-99. doi:10.1016/J.CMPB.2017.02.023
14. Harmon SA, Tuncer S, Sanford T, Choyke PL, Türkbey B. Artificial intelligence at the intersection of pathology and radiology in prostate cancer. 2019;25(3):183-188. Accessed March 24, 2022. <https://pubmed.ncbi.nlm.nih.gov/31063138/>
15. Arvaniti E, Fricker KS, Moret M, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep.* 2018;8(1). doi:10.1038/S41598-018-30535-1
16. Zhou N, Fedorov A, Fennessy F, Kikinis R, Gao Y. Large scale digital prostate pathology image analysis combining feature extraction and deep neural network. Published online May 7, 2017. doi:10.48550/arxiv.1705.02678
17. Jiménez del Toro O, Atzori M, Otálora S, et al. Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score. *Med Imaging 2017 Digit Pathol.* 2017;10140:101400O. doi:10.1117/12.2255710
18. Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology:

Gleason grading of prostate cancer. *BMC Bioinformatics*. 2012;13(1). doi:10.1186/1471-2105-13-282

19. Doyle S, Feldman M, Tomaszewski J, Madabhushi A. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Trans Biomed Eng*. 2012;59(5):1205-1218. doi:10.1109/TBME.2010.2053540

20. Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*. 2016;191:214-223. doi:10.1016/J.NEUCOM.2016.01.034

21. Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep*. 2016;6. doi:10.1038/SREP26286

22. Nir G, Hor S, Karimi D, et al. Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Med Image Anal*. 2018;50:167-180. doi:10.1016/J.MEDIA.2018.09.005

23. Nagpal K, Foote D, Tan F, et al. Development and Validation of a Deep Learning Algorithm for Gleason Grading of Prostate Cancer From Biopsy Specimens. *JAMA Oncol*. 2020;6(9):1372-1380. doi:10.1001/JAMAONCOL.2020.2485

24. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301-1309. doi:10.1038/S41591-019-0508-1

25. Bulten W, Balkenhol M, Belinga JJA, et al. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Mod Pathol*. 2021;34(3):660-671. doi:10.1038/S41379-020-0640-Y

26. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol*. 2020;21(2):222-232. doi:10.1016/S1470-2045(19)30738-7

27. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Heal.* 2020;2(8):e407-e416. doi:10.1016/S2589-7500(20)30159-X
28. Rashid S, Fazli L, Boag A, Siemens R, Abolmaesumi P, Salcudean SE. Separation of benign and malignant glands in prostatic adenocarcinoma. *Med Image Comput Comput Assist Interv.* 2013;16(Pt 3):461-468. doi:10.1007/978-3-642-40760-4\_58
29. Gertych A, Ing N, Ma Z, et al. Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Comput Med Imaging Graph.* 2015;46 Pt 2(Pt 2):197-208. doi:10.1016/J.COMPMEIMAG.2015.08.002
30. Singh M, Kalaw EM, Giron DM, Chong K-T, Tan CL, Lee HK. Gland segmentation in prostate histopathological images. *J Med imaging (Bellingham, Wash).* 2017;4(2):027501. doi:10.1117/1.JMI.4.2.027501
31. Lu C, Romo-Bucheli D, Wang X, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Lab Invest.* 2018;98(11):1438-1448. doi:10.1038/S41374-018-0095-7
32. Saltz J, Gupta R, Hou L, et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* 2018;23(1):181-193.e7. doi:10.1016/J.CELREP.2018.03.086
33. Rezaeilouyeh H, Mollahosseini A, Mahoor MH. Microscopic medical image classification framework via deep learning and shearlet transform. *J Med imaging (Bellingham, Wash).* 2016;3(4):044501. doi:10.1117/1.JMI.3.4.044501
34. Nagpal K, Foote D, Liu Y, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digit Med* 2019 21. 2019;2(1):1-10. doi:10.1038/s41746-019-0112-2
35. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 12, pp. 2481-2495.

36. Avramenko V.V., Demianenko V.M. Operative recognition of standard signal types. *Радіоелектроніка, інформатика, управління*, 2020. 2. С. 75 - 83. DOI: 10.15588/1607-3274-2020-2-8.
37. Luo, W. Hassell, L. A. Use cases for digital pathology. *Digital Pathology: Historical Perspectives, Current Concepts and Future Applications*. Springer Intern. Publ., 2016, pp. 5–15.
38. Gerrit van den Burg, P. J. Groenen. GenSVM A Generalized Multiclass Support Vector Machine. *Journal of Machine Learning Research*, 2016, vol. 17, no 224, pp. 1 – 42.
39. Beckwith, B. A. Standards for digital pathology and whole slide imaging. *Digital Pathology: Historical Perspectives, Current Concepts and Future Applications*. Springer Intern. Publ., 2016. pp. 87–97.
40. Moskalenko, V.V., Korobov, A. G. Extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extraction. *Radio Electronics, Computer Science, Control*, 2017, no 2, pp. 38 – 45.
41. Xu, G., Zong, Y., Yang, Y.Z. *Applied Data Mining*. CRC Press, 2013, 284 p
42. Rozhkova N. I., Roitberg P. G., Varfolomeeva A. A. Mazo M. M., Dobrenkii A. N., Blinov D. S., Sushkov E. V., Deryabina O. N., Sokolov A. I. Neural network-based segmentation model for breast cancer X-ray screening. *Sechenov Medical Journal*, 2020, no 11(3), pp. 4 – 14. Doi.org/10.47093/2218-7332.2020.11.3.4-14
43. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015, pp. 234-241
44. Dovbysh, A.S., Rudenko, M. S. Information-extreme learning algorithm for a system of recognition of morphological images in diagnosing oncological pathologies / *Cybernetiks and Systems Analysis*, 2014, v. 50, No. 1, pp. 157-163.: DOI: 10.15587/1729-4061.2016.71930

45. 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(4), pp 534-543. DOI:10.1007/s10559-020-00269-y.

46. Igor Naumenko, Mykyta Myronenko, Taras Savchenko. Information-extreme machine training of on-board recognition system with optimization of RGB-component digital images. *Radioelectronic and Computer Systems*, 2021, no 4, pp. 59-70. – DOI: 10.32620/reks.2021.4.05.

47. Dovbysh, I. Shelehov, S. Pylypenko, O. Berest. Estimation of informativeness of recognition signs at extreme information machine learning of knowledge control system. *CEUR Workshop Proceedings*, 2019, vol. 2362, pp. 143-152.