

<https://doi.org/10.31891/2307-5732-2026-361-48>

УДК 004.932.2:616.31

РИБАК ВОЛОДИМИР

Національний лісотехнічний університет України

<https://orcid.org/0009-0006-8809-978X>

e-mail: volodymyr.rybak.personal@gmail.com

ШАБАТУРА ЮРІЙ

Національна академія сухопутних військ ім. гетьмана Петра Сагайдачного

<https://orcid.org/0000-0002-9961-1244>

e-mail: shabaturayuriy@gmail.com

**ДВОРІВНЕВИЙ МЕТОД РОЗПІЗНАВАННЯ СКЛАДНИХ ЦИФРОВИХ
ЗОБРАЖЕНЬ ІЗ ЗАСТОСУВАННЯМ ROI-МАСКУВАННЯ ТА ГІБРИДНОЇ
АРХІТЕКТУРИ U-NET 3+ / RESNET-U-NET, НА ПРИКЛАДІ ВИЯВЛЕННЯ
КАРІЄСУ ЗА РЕНТГЕНІВСЬКИМ ЗНІМКОМ**

Задача розпізнавання складних цифрових зображень є важливою і актуальною для багатьох галузей науки і техніки. Особливо вагоме її значення в проведенні медичної діагностики на основі аналізу рентгенологічних цифрових зображень. Застосування систем штучного інтелекту створює основу для успішної автоматизації вирішення цієї задачі. Однак, в ряді випадків, зокрема для раннього виявлення карієсу зубів на основі аналізу рентгенівських стоматологічних зображень стандартні методи розпізнавання цифрових зображень не дозволяють отримати хороші результати. Підвищення точності цієї діагностики потребує поєднання ефективних алгоритмів сегментації цифрових рентгенологічних стоматологічних зображень з адаптивними моделями глибокого навчання, здатними аналізувати лише релевантні ділянки зображення.

У роботі представлено дворівневий метод автоматичного розпізнавання складних цифрових зображень на прикладі вирішення задачі виявлення карієсу на стоматологічних рентгенівських знімках, який поєднує сегментаційний етап і цільову діагностику в межах області інтересу (ROI).

На першому етапі застосовано модифіковану архітектуру U-Net 3+, що інтегрує легковаговий енкодер MobileNet та Vision Mamba Layers у декодері, забезпечуючи високу точність і стабільність сегментації зображення при зниженні обчислювальної складності. Отримана маска зубного ряду використовується для формування ROI, яка обмежує аналіз клінічно значущою ділянкою зображення, виключаючи фон, м'які тканини та артефакти. На другому етапі виконується детекція каріозних уражень за допомогою оптимізованої моделі ResNet-U-Net, що поєднує залишкові блоки з U-подібною структурою для точного відтворення меж патологічних ділянок навіть за умов низького контрасту. Експериментальні дослідження, проведені на наборі даних DENTEX, підтвердили ефективність запропонованого методу: показники Dice = 0.74, Precision = 0.88 і Recall = 0.90 перевищують результати базових моделей U-Net, U-Net 3+ та MobileMamba-U-Net. Аналіз показав, що застосування ROI-маскування суттєво зменшує кількість хибнопозитивних результатів, підвищує стабільність моделі при варіаціях експозиції та скорочує час інференсу приблизно на третину.

Запропонований підхід забезпечує баланс між точністю та швидкодією, створюючи основу для розроблення інтелектуальних інформаційних систем здатних розпізнавати складні цифрові зображення, в тому числі, як показали проведені дослідження, він цілком придатний для створення інтелектуальних систем підтримки діагностичних рішень у стоматології. Метод має потенціал для впровадження у програмне забезпечення або мобільні додатки, спрямовані на автоматизований скринінг стоматологічних знімків, а подальші дослідження можуть бути орієнтовані на розширення навчальної вибірки, мультипатологічне навчання та адаптацію до інших типів рентгенівських даних.

Ключові слова: глибоке навчання; сегментація зображень; карієс; ROI-маскування; U-Net 3+; MobileNet; Vision Mamba Layers; ResNet-U-Net.

RYBAK VOLODYMYR

The National Forestry University of Ukraine

SHABATURA YURIY

Hetman Petro Sahaidachnyi National Ground Forces Academy

**A TWO-LEVEL METHOD FOR COMPLEX DIGITAL IMAGE RECOGNITION USING ROI
MASKING AND A HYBRID U-NET 3+ / RESNET-U-NET ARCHITECTURE ON CASE STUDY OF
DENTAL CARIES DETECTION ON X-RAY IMAGES**

The task of recognizing complex digital images is an important and relevant challenge in many scientific and technical domains. It has particular significance in medical diagnostics based on the analysis of radiographic digital images. The application of artificial intelligence systems provides a foundation for automating this process effectively. However, in several cases—particularly for the early detection of dental caries from radiographic dental images—standard digital image recognition methods fail to deliver accurate results. Increasing the accuracy of such diagnostics requires combining efficient segmentation algorithms for dental radiographs with adaptive deep learning models capable of analyzing only the relevant image regions.

This paper presents a two-level method for automatic recognition of complex digital images, demonstrated through the task of detecting carious lesions on dental X-ray images. The proposed system combines a segmentation stage with target analysis within a defined region of interest (ROI).

At the first stage, a modified U-Net 3+ architecture is applied, integrating a lightweight MobileNet encoder and Vision Mamba Layers in the decoder. This configuration ensures high segmentation accuracy and stability while reducing computational complexity. The resulting mask of the dental region is used to form the ROI, which constrains further analysis to clinically significant areas of the image while excluding background, soft tissues, and artifacts.

At the second stage, caries detection is performed using an optimized ResNet-U-Net model that combines residual blocks with a U-shaped structure, enabling precise reconstruction of object boundaries even under low-contrast conditions. Experimental studies conducted on the DENTEX dataset confirmed the effectiveness of the proposed approach: Dice = 0.74, Precision = 0.88, and Recall = 0.90, which outperform the baseline U-

Net, U-Net 3+, and MobileMamba-U-Net models. The analysis demonstrated that ROI masking significantly reduces the number of false positives, improves model stability under varying exposure conditions, and decreases inference time by approximately one-third.

The proposed approach provides a balance between accuracy and computational efficiency, forming the basis for developing intelligent information systems capable of recognizing complex digital images. As shown by the conducted research, it is particularly suitable for intelligent decision-support systems in dental diagnostics. The method has potential for integration into software and mobile applications for automated screening of dental radiographs. Future research may focus on expanding the training dataset, implementing multi-pathology learning, and adapting the approach to other types of radiographic data.

Keywords: deep learning; image segmentation; dental caries; ROI masking; U-Net 3+; MobileNet; Vision Mamba Layers; ResNet-U-Net.

Стаття надійшла до редакції / Received 18.11.2025

Прийнята до друку / Accepted 11.01.2026

Опубліковано / Published 29.01.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Рибак Володимир, Шабатура Юрій

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Виявлення важливих елементів на складних цифрових зображеннях була, є і залишається досить нетривіальною задачею. Її вирішення є актуальним для багатьох галузей науки і техніки. Особливо важливою вона є для медичної сфери і найперше для проведення об'єктивної діагностики на основі аналізу рентгенографічних зображень.

Традиційні методи цифрової обробки зображень — порогове сегментування, методи кластеризації чи активних контурів — демонструють обмежену ефективність на рентгенограмах через неоднорідність тканин, варіації освітлення й наявність артефактів. Вони не можуть адаптивно враховувати складні просторові залежності, що є критично важливим для локалізації дрібних об'єктів, таких як каріозні порожнини.

Яскравим прикладом складності означеної проблеми є задача стоматологічної діагностики на основі рентгеновських знімків. Використання рентгеновських знімків стало невід'ємною частиною клінічного обстеження, а від точності їх інтерпретації безпосередньо залежить своєчасність виявлення патологій, зокрема карієсу. Попри суттєвий прогрес у цифрових технологіях, процес аналізу зображень у стоматології й досі переважно базується на ручному оцінюванні лікарем, що створює ризики суб'єктивності, пропуску дрібних уражень і різничитань між фахівцями.

Карієс є найпоширенішим стоматологічним захворюванням у світі, і його рання діагностика має вирішальне значення для запобігання ускладненням і збереження зубів. Водночас виявлення карієсу на рентгеновських знімках є складним завданням навіть для досвідчених спеціалістів через низький контраст уражених ділянок, накладання анатомічних структур і наявність шумів. Саме тому виникає потреба у створенні інтелектуальних систем, які здатні автоматично аналізувати рентгеновські стоматологічні зображення та виділяти патологічні області з високою точністю і стабільністю.

У свою чергу, глибокі нейронні мережі, зокрема архітектура типу U-Net і її модифікації, відкрили нові можливості в медичній сегментації завдяки здатності до багаторівневого вилучення ознак і точного відтворення анатомічних структур. Проте навіть ці моделі стикаються з низкою викликів у стоматологічних застосуваннях. По-перше, стандартні архітектури мають значну кількість параметрів і вимагають потужних обчислювальних ресурсів, що ускладнює їх використання на мобільних пристроях або у клінічних системах із обмеженою продуктивністю. По-друге, при аналізі панорамних рентгеновських знімків модель часто отримує надмірну кількість нерелевантної інформації, що призводить до помилкових рішень. Це знижує точність і практичну цінність автоматизованих систем.

Одним із перспективних шляхів розв'язання цих проблем є використання ROI-маскування (Region of Interest masking) — підходу, який дозволяє зосередити аналіз моделі лише на релевантних областях зображення. У контексті діагностики карієсу це означає обмеження пошуку патологій лише межами зубних структур, що суттєво зменшує кількість хибнопозитивних результатів і підвищує продуктивність аналізу. Для реалізації цього підходу важливо забезпечити точну сегментацію зображення зубів, що вимагає глибоких моделей зі збереженням контекстуальних залежностей та високої роздільності відтворення.

Модифікації архітектури U-Net 3+, інтегровані з MobileNet як легковаговим енкодером і Vision Mamba Layers, як блоками просторового моделювання, створюють основу для побудови ефективних систем автоматизованого аналізу стоматологічних зображень. Таке поєднання забезпечує високу точність сегментації зображення при суттєвому зниженні обчислювальних витрат. Наукова значущість цього підходу полягає у розробці технології розпізнавання складних зображень, здатної забезпечити баланс між продуктивністю, точністю та універсальністю, а практична — у можливості інтеграції в програмне забезпечення для стоматологічної діагностики.

Отже, розроблення методу розпізнавання складних зображень на основі ROI-маскування з використанням модифікованої архітектури U-Net 3+ із MobileNet та Vision Mamba Layers є актуальним завданням, що відповідає сучасним тенденціям розвитку інтелектуальних інформаційних систем. Його реалізація зокрема, для виявлення карієсу на основі аналізу рентгеновських знімків, сприятиме підвищенню об'єктивності діагностики, зменшенню навантаження на лікарів і створенню передумов для впровадження технологій штучного інтелекту в стоматологічну практику.

Аналіз досліджень та публікацій

Упродовж останніх років сегментація медичних зображень стала одним із найдинамічніших напрямів у галузі штучного інтелекту. Застосування глибоких нейронних мереж дозволило досягти істотного прогресу в

задачах виділення анатомічних структур, і зокрема у стоматології. Ключовим етапом у цьому процесі стала поява архітектури U-Net [1], яка стала стандартом де-факто для медичної сегментації завдяки симетричній структурі «енкодер–декодер» і використанню пропускних з'єднань, що забезпечують передачу просторових ознак між рівнями.

Подальший розвиток архітектури відбувався у напрямку підвищення точності та стабільності результатів: U-Net++ [2] удосконалила процес агрегації ознак між шарами через вкладені з'єднання, а U-Net 3+ [3] запровадила повношкальні пропускні з'єднання, які дозволяють моделі одночасно використовувати інформацію з усіх рівнів абстракції. Це значно покращило точність виділення дрібних структур і зробило U-Net 3+ особливо ефективною для медичних зображень з високим рівнем шуму.

Стоматологічна рентгенографія є специфічним типом медичних зображень, що поєднує складну анатомічну будову, тіньові перекриття та низький контраст між здоровими і ураженими тканинами. У цій галузі U-Net та її похідні застосовуються для сегментації зубів, щелепних кісток, кореневих каналів і каріозних уражень. Так, у роботі [4] було використано вкладену архітектуру Nested U-Net, де одна модель виконує сегментацію зубів, а друга — виявлення карієсу. Такий підхід продемонстрував високу точність локалізації, але залишався ресурсоємним і потребував значних обчислювальних потужностей.

В іншому дослідженні [5] запропонували ансамблеву систему CariSeg, яка поєднувала три архітектури — U-Net, DeepLabV3 і FPN — для покращення сегментації карієсу на панорамних знімках. Попри високі показники точності (Dice \approx 0,68–0,70), така система мала складну структуру та низьку швидкість, що обмежувало її застосування у клінічній практиці. Подібні результати отримали й Pornprasertsuk-Damrongsri et al. [6], які поєднали YOLOv5 і Attention U-Net для каскадного виявлення карієсу, досягнувши F1=0,85, однак із помітним зниженням precision через численні хибнопозитивні результати.

Найбільш помітною тенденцією останніх років стало прагнення оптимізувати архітектури сегментаційних моделей з метою забезпечення їхньої працездатності на пристроях із обмеженими ресурсами. Зокрема, MobileNet [7] і подібні легковагові моделі стали основою для створення ефективних енкодерів, які забезпечують високу швидкість завдяки використанню глибоких сепарабельних згорток (Depthwise Separable Convolutions). На основі MobileNet створено модифікації, що поєднують просторову ефективність з високою точністю, наприклад Mobile-U-Net і Squeeze U-Net [8]. Вони дозволяють зменшити кількість параметрів у кілька разів без значної втрати продуктивності, що є критичним для клінічних систем реального часу.

Додатковий імпульс розвитку легковагових архітектур дало впровадження Vision Mamba Layers — нового класу модулів на основі лінійних моделей стану (State Space Models), здатних ефективно моделювати глобальні просторові залежності. Роботи [9] та [10] показали, що додавання Mamba-блоків до архітектур типу U-Net або Swin-UNet підвищує точність сегментації на 2–3% при збереженні компактності мережі. Цей напрям отримав подальший розвиток у гібридних моделях MobileMamba-U-Net, де MobileNet використовується як енкодер, а Vision Mamba Layers — у ролі декодера для покращення контекстної інтеграції ознак.

У попередніх дослідженнях авторів [11–12] було запропоновано саме таку архітектуру для сегментації стоматологічних зображень. Вона поєднала легковагову структуру з високими показниками точності (Dice \approx 0,91, IoU \approx 0,84), що довело ефективність підходу для реальних клінічних умов. Водночас у цих роботах сегментація обмежувалася анатомічними структурами, без переходу до виявлення патологій. Наступним логічним кроком стало використання отриманих сегментаційних моделей для побудови каскадних систем діагностики, зокрема для ROI-маскування та подальшого виявлення карієсу.

Таким чином, сучасні дослідження підтверджують доцільність поєднання точних сегментаційних архітектур (U-Net 3+, Attention U-Net) із оптимізованими енкодерами (MobileNet) і контекстними шарами (Vision Mamba Layers). Водночас у науковій літературі відсутні системні рішення, які б об'єднували ці компоненти в єдину технологію для виявлення карієсу на стоматологічних рентгеновських знімках. Це формує наукову нішу для запропонованої в даній роботі методики — дворівневого підходу ROI-маскування з використанням модифікації U-Net 3+ із MobileNet та Vision Mamba Layers.

Формулювання цілей статті

Метою роботи є: розробка дворівневого методу розпізнавання складних цифрових зображень, який базується на поєднанні ROI-маскування та модифікованої архітектури U-Net 3+, інтегрованої з MobileNet і Vision Mamba Layers, та його експериментальна перевірка на прикладі автоматичного виявлення карієсу на основі аналізу стоматологічних рентгеновських знімків.

Для досягнення поставленої мети в роботі вирішуються такі завдання:

1. Проаналізувати сучасні підходи до сегментації цифрових стоматологічних зображень та існуючі методи виявлення карієсу за такими зображеннями на основі використання глибоких нейронних мереж.
2. Розробити архітектуру дворівневої моделі, де перший рівень виконує сегментацію зображення зубів (ROI-маскування) на основі модифікованої U-Net 3+ з MobileNet-енкодером, а другий — ідентифікацію на аналізованих зображеннях власне карієсних уражень у межах ROI.
3. Провести навчання та тестування моделі на відкритому датасеті стоматологічних рентгеновських знімків із розміткою для оцінки її точності за стандартними метриками (Dice, IoU, Precision, Recall).
4. Порівняти результати запропонованого підходу з базовими моделями (та показати його переваги в аспектах обчислювальної ефективності й точності).

Досягнення поставлених завдань дозволить обґрунтувати доцільність використання дворівневих систем ROI-маскування для розпізнавання складних цифрових зображень в тому числі для вирішення клінічних задач

автоматизованої діагностики карієсу зубів за аналізом цифрових рентгенологічних зображень, та створити основу для подальшої розробки інтелектуальних інформаційних систем, зокрема для підтримки діагностичних рішень в стоматології.

Виклад основного матеріалу

Загальна концепція методу. Розроблений метод ґрунтується на дворівневій структурі обробки цифрових зображень, де перший рівень відповідає за сегментацію об'єктів і формування області інтересу (ROI), а другий — за подальший локальний аналіз цієї області. Такий підхід дозволяє зосередити обчислювальні ресурси лише на релевантних ділянках зображення, усуваючи вплив фонових артефактів і шумів, що підвищує загальну ефективність системи розпізнавання.

Розроблений метод протестовано на прикладі аналізу стоматологічних рентгеновських зображень, які характеризуються високим рівнем шуму, низьким контрастом та неоднорідністю освітлення. Такі умови створюють типовий сценарій для оцінки ефективності алгоритмів сегментації та локального аналізу у складних медичних даних. Використання цього прикладу дало змогу продемонструвати здатність системи працювати зі зображеннями, де критично важливим є точне відтворення меж і стійкість до артефактів.

Сегментаційна частина методу реалізується на основі модифікованої архітектури MobileMamba-U-Net 3+ [11, 12], яка поєднує переваги повношкальної структури U-Net 3+ із легковаговими обчисленнями. Архітектура складається з двох основних компонентів: енкодера на базі MobileNetV2 та декодера з Vision Mamba Layers. Енкодер використовує глибинні сепарабельні згортки (Depthwise Separable Convolutions), що дозволяє суттєво скоротити кількість параметрів і обчислювальних операцій порівняно з класичним U-Net. Кожен рівень енкодера вилучає ознаки різного рівня абстракції та передає їх до відповідних рівнів декодера через пропусні з'єднання (skip connections), зберігаючи детальну просторову інформацію.

Декодер побудований на основі Vision Mamba Layers (VSS Blocks), які реалізують лінійні моделі стану (State Space Models) для моделювання просторово-контекстних залежностей між пікселями. На відміну від self-attention механізмів, Mamba-блоки мають лінійну обчислювальну складність $O(n)$, що забезпечує стабільну роботу з великими зображеннями та зменшує витрати пам'яті. Завдяки цьому архітектура зберігає високу продуктивність навіть за умов значного шуму або низького контрасту вхідних даних.

Для підвищення узагальнювальної здатності використовується механізм повношкальних пропусних з'єднань, притаманний U-Net 3+, який об'єднує ознаки з усіх рівнів енкодера та декодера. Така структура забезпечує баланс між локальною деталізацією та глобальним контекстом сцени. У результаті MobileMamba-U-Net 3+ демонструє високу точність сегментації при зменшенні кількості параметрів більш ніж утричі порівняно з класичною U-Net, зберігаючи при цьому обчислювальну ефективність і стійкість до артефактів.

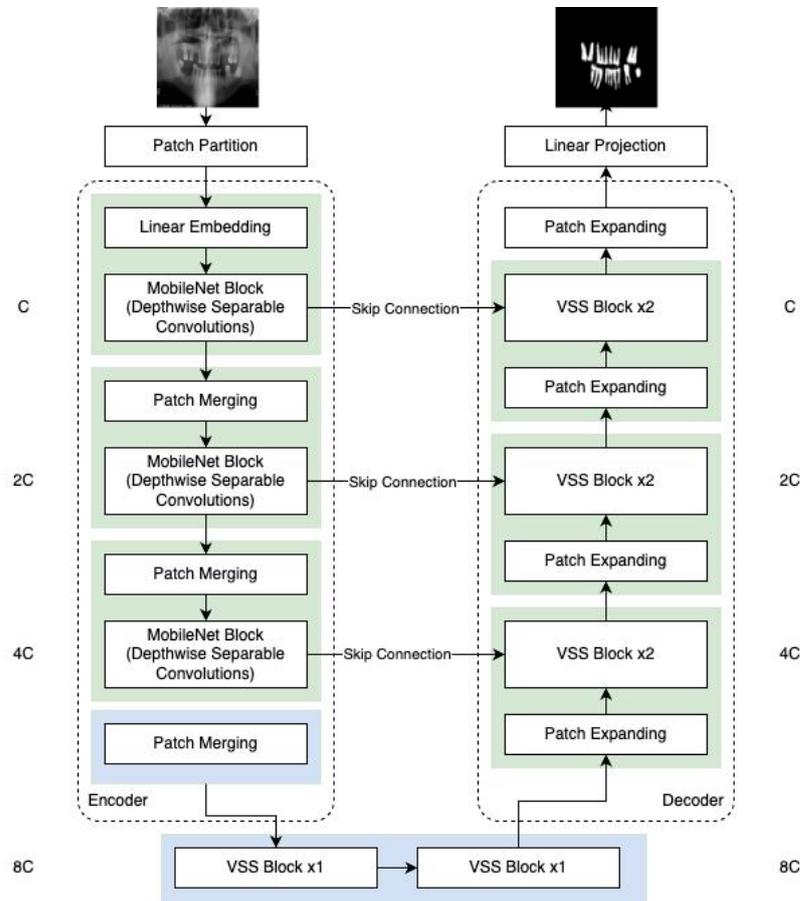


Рис. 1. Архітектура сегментаційної моделі MobileMamba-U-Net 3+.

Метод формування області інтересу (ROI) базується на результатах сегментації, отриманих від першої моделі. Його мета — автоматично визначити релевантну ділянку зображення для подальшого локального аналізу. Процес складається з кількох етапів обробки сегментаційної маски:

1. Морфологічна фільтрація. До бінарної маски застосовується операція розширення (dilation), що дозволяє об'єднати сусідні об'єкти та заповнити дрібні проміжки між ними. Потім виконується операція closing для згладження контурів і усунення локальних розривів.

2. Видалення шумів. Невеликі окремі компоненти, які не належать до основної структури, видаляються за допомогою аналізу зв'язаних областей (connected components analysis). Це забезпечує стабільність подальшої локалізації.

3. Побудова ROI. На основі очищеної маски визначається мінімальний обмежувальний прямокутник (bounding box), у межах якого вирізається підмасив вхідного зображення. Для складних протяжних сцен допускається часткове розширення меж на 5–10% для збереження контексту при подальшому аналізі.

4. Нормалізація. Отримане ROI-зображення масштабується до сталої роздільної здатності (512×512 пікселів) із збереженням пропорцій. Водночас проводиться нормалізація інтенсивностей для усунення варіацій контрасту між різними джерелами зображень.

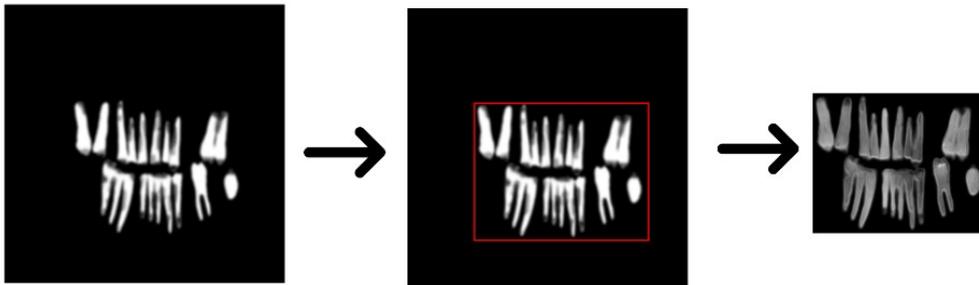


Рис. 2. Послідовність формування області ROI після сегментації рентгенологічного зображення зубів

Таким чином, ROI формується не шляхом простого множення маски, а за допомогою структурованої морфологічної обробки, що забезпечує точну локалізацію області аналізу та підвищує надійність подальших етапів розпізнавання.

Архітектура моделі локального аналізу. Для аналізу в межах ROI використовується легковагова модифікація U-Net із ResNet-енкодером [4, 5]. Архітектура запропонованої моделі ResNet-U-Net складається з п'яти рівнів енодера та симетричного декодера, з'єднаних між собою пропусковими зв'язками (skip connections), що забезпечують відновлення дрібних структурних деталей під час реконструкції зображення. Основною особливістю архітектури є застосування ResNet-блоків як базових компонентів енодера, що стабілізують процес навчання та зменшують ризик затухання градієнтів у глибоких мережах.

Енодер (Encoder Path) побудований на базі ResNet18, який складається з п'яти послідовних блоків. Кожен блок містить дві згортки 3×3, операції нормалізації (Batch Normalization або Group Normalization) та функцію активації ReLU. Після кожного блоку виконується операція підвибірки (stride = 2), що зменшує розмір карти ознак удвічі та збільшує кількість каналів у два рази. Завдяки залишковим з'єднанням між входом і виходом блоку модель зберігає інформацію про низькорівневі ознаки навіть на глибоких шарах, що підвищує точність локалізації меж об'єктів.

У середині архітектури розташований **bottleneck-блок**, який виконує найглибше вилучення ознак. Він містить дві послідовні згортки 3×3, нормалізацію, ReLU і Dropout (0,3), що зменшує ризик перенавчання. Цей рівень акумулює просторовий контекст ROI і формує узагальнене представлення, яке потім передається у декодер.

Декодер (Decoder Path) відновлює просторову роздільність зображення за допомогою транспонованих згорток (TransConv). Після кожного етапу upsampling-а відбувається конкатенація відповідних карт ознак із енодера, що забезпечує відновлення локальних контурів і структур. Кожен рівень декодера містить дві згортки 3×3 з активацією ReLU. Для зменшення кількості параметрів частина згорток замінена на 1×1-фільтри, що знижує обчислювальне навантаження без втрати точності.

Вихідний шар (Output Layer) декодера складається із згортки 1×1, яка зменшує кількість каналів до одного. Використовується функція активації $\sigma(x)$ (сигмоїда), що перетворює вихідну ймовірнісну маску приналежності пікселів до цільового класу. Отримана маска може бути візуалізована у вигляді теплової карти (heatmap), накладеної на оригінальне зображення.

Ініціалізація ваг здійснюється методом He-normal, що забезпечує швидку збіжність при глибоких архітектурах. Для стабільності на малих batch застосовується Group Normalization. Використання Dropout у bottleneck-рівні сприяє узагальненню моделі. Всі шари оптимізуються за допомогою алгоритму Adam із початковою швидкістю навчання $1e-4$.

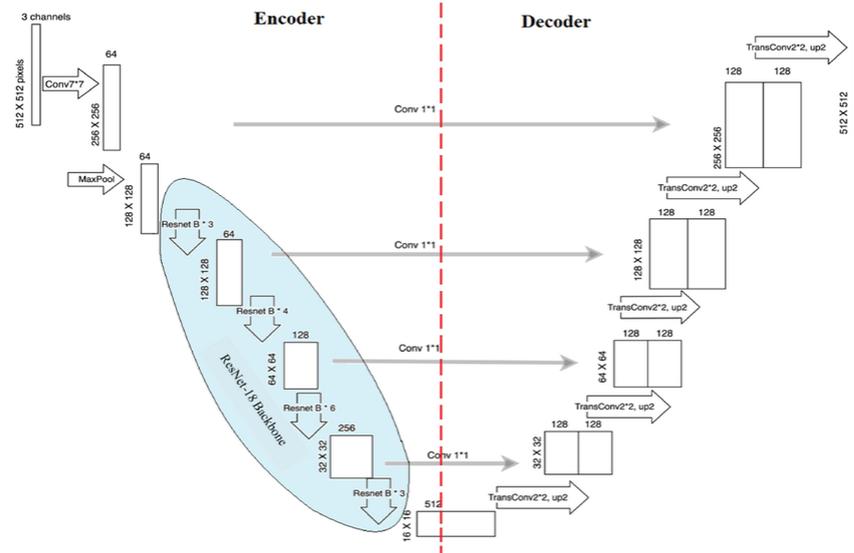


Рис. 3. Архітектура моделі локального аналізу (ResNet-U-Net) у межах ROI

Вибір архітектурних і навчальних гіперпараметрів здійснювався експериментально з урахуванням компромісу між точністю та швидкодією. Як енкодер обрано ResNet18, оскільки він забезпечує достатню глибину для ефективного вилучення ознак при помірній кількості параметрів. Більш глибокі варіанти, такі як ResNet34 або ResNet50, не давали суттєвого приросту точності, але значно збільшували обчислювальні витрати. Значення Dropout = 0.3 визначене емпірично: менші призводили до перенавчання, тоді як більші сповільнювали збіжність. Межі ROI розширювалися на 5–10% відносно обмежувального прямокутника (bounding box), що дозволяло зберігати контекст навколишніх структур для коректнішого розпізнавання об'єктів у межах сцени. Такий набір параметрів забезпечив оптимальний баланс між стабільністю, узагальненням і обчислювальною ефективністю моделі.

Принцип роботи дворівневої системи. Після навчання обидві моделі — сегментаційна та модель локального аналізу — працюють послідовно (рис. 4). Спочатку перша модель виконує сегментацію зображення та формує ROI-маску, яка визначає релевантну область для подальшої обробки. Потім вихідне зображення обрізається за межами маски, і отримане ROI подається на вхід другої моделі, що виконує детальний аналіз і локалізацію цільових об'єктів. Така дворівнева структура дозволяє мінімізувати вплив фонових елементів і підвищити точність розпізнавання в умовах складної структури зображень.

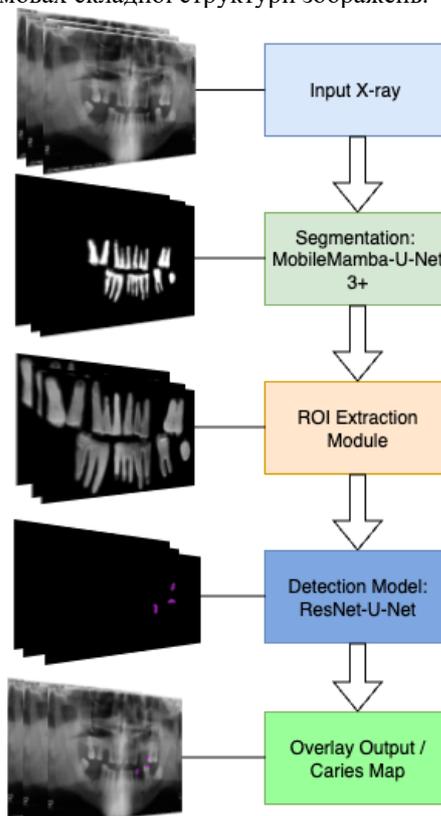


Рис. 4. Каскадна схема роботи системи ROI-маскування та локалізації цільових об'єктів

Отримані карти ймовірності можуть бути візуалізовані у вигляді накладення на оригінальне зображення, що підвищує інтерпретованість результатів та зручність їх подальшої експертної оцінки. Такий підхід дозволяє користувачу швидко зіставити локалізацію виявлених об'єктів із початковими даними. У перспективі дворівнева система може бути інтегрована в програмні комплекси аналізу цифрових зображень або інформаційні системи технічного та медичного призначення для автоматизованої попередньої обробки та візуальної підтримки прийняття рішень.

Опис експериментальних умов

Для навчання та тестування запропонованої технології використовувався відкритий набір даних DENTEX [10], який містить 700 високороздільних панорамних рентгенівських зображень. Для кожного зразка надано дві типи розмітки: (1) маски основних структур об'єктів, підготовлені для сегментаційного етапу, (2) додаткові маски локальних областей, що використовувались для другого рівня аналізу. Розмітка виконувалась вручну експертами, із подальшою верифікацією для забезпечення узгодженості даних.

Перед початком навчання всі зображення були масштабовані до розміру 512×512 пікселів і нормалізовані за інтенсивністю в діапазоні [0, 1]. Для покращення локального контрасту застосовано алгоритм CLAHE (Contrast Limited Adaptive Histogram Equalization), який підвищує видимість дрібних структур при різному освітленні. Кожне зображення оброблялося попередньо натренованою сегментаційною моделлю MobileMamba-U-Net 3+, яка формувала ROI-маску основної області аналізу. Отримані ROI-зони зберігалися у вигляді окремих підмасивів і використовувались як вхідні дані для другої моделі — ResNet-U-Net, що виконувала локальний аналіз.

Для забезпечення узагальнювальної здатності моделей дані було розділено за схемою train/validation/test = 70/15/15%. Формування підвбірок відбувалося випадковим чином, при цьому однакові об'єкти не потрапляли до різних множин, щоб уникнути витоків даних. У межах навчальної вибірки застосовувались аугментації: горизонтальне віддзеркалення, повороти $\pm 15^\circ$, масштабування в межах 0.9–1.1, зміна контрасту та яскравості, а також випадкове Gaussian-зашумлення.

Обидва етапи — сегментаційний і локальний — навчались окремо. Сегментаційна модель MobileMamba-U-Net 3+ була попередньо натренована та заморожена під час навчання другої моделі. Для обох етапів використовувався оптимізатор Adam із початковою швидкістю навчання $1e-4$, scheduler із поетапним зменшенням LR на 0.1 кожні 15 епох, Dropout (0.3) та L2-регуляризація ($1e-5$). Навчання другої моделі тривало 100 епох при batch size = 8. Для компенсації дисбалансу між класами «карієс» та «норма» використовувалася комбінована функція втрат, що об'єднує Binary Cross-Entropy (BCE) та Dice Loss. Її вираз має вигляд:

$$L = \alpha L_{BCE} + (1 - \alpha) L_{Dice}, \quad (1)$$

де $\alpha = 0.5$ забезпечує рівновагу між глобальною точністю класифікації та якістю локалізації меж об'єктів. Таке поєднання сприяє стабільному навчанню за наявності дисбалансних даних і покращує збіжність на ранніх етапах тренування.

Для оцінювання ефективності запропонованого методу використовувався відкритий набір зображень із відповідними масками сегментації, створеними вручну експертами. Набір включає приклади з різними варіаціями контрасту, шуму та геометрії об'єктів, що дозволяє оцінити стійкість алгоритму до зовнішніх факторів. Перед початком навчання всі дані проходили попередню обробку: нормалізацію інтенсивності, масштабування до розміру 512×512 пікселів і аугментацію (повороти, віддзеркалення, зміни контрасту та яскравості).

Навчання та інференс виконувались у середовищі Python 3.10 з використанням фреймворку PyTorch 2.1.0, бібліотек CUDA 12.1 і cuDNN 8.7. Обчислення проводилися на робочій станції з GPU NVIDIA RTX 3080 Ti (12 GB VRAM), процесором AMD Ryzen 9 5900X (12 ядер, 24 потоки) та 64 GB оперативної пам'яті. Середній час навчання сегментаційної моделі становив приблизно 4,5 години (100 епох, batch size = 16), а моделі локального аналізу — 2,7 години (100 епох, batch size = 8). Час інференсу одного зображення розміром 512×512 пікселів становив близько 112 мс, що дозволяє використовувати систему в режимі, близькому до реального часу.

Оцінювальні метрики

Для кількісної оцінки якості роботи моделей використовувались такі метрики:

Dice coefficient (DSC) – основна метрика, що відображає схожість між передбаченою моделлю маскою та еталонною, ідеально підходить для оцінки сегментації нерівномірно розподілених об'єктів. Визначається як:

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \quad (2)$$

де TP — кількість правильно знайдених пікселів, FP — хибно знайдених, FN — пропущених. Значення Dice = 1 відповідає ідеальному збігу масок.

Intersection over Union (IoU) — коефіцієнт Жаккара, альтернативна метрика перекриття масок, яка враховує співвідношення площі перетину та об'єднання двох множин:

$$IoU = \frac{TP}{TP + FP + FN}. \quad (3)$$

Використовується як додатковий індикатор точності сегментації.

Precision — частка правильно передбачених позитивних пікселів серед усіх, які модель класифікувала як позитивні:

$$Precision = \frac{TP}{TP + FP}. \quad (4)$$

Ця метрика характеризує схильність моделі до генерації хибних позитивних результатів (низьке Precision означає велику кількість FP).

Чутливість (Recall) – частка знайдених позитивних пікселів відносно всіх істинно позитивних:

$$Recall = \frac{TP}{TP+FN}. \quad (5)$$

Показує здатність моделі не пропускати об'єкти цільового класу.

F1-міра – гармонійне середнє precision та recall, застосовується для узагальненої оцінки балансу між ними.

Для оцінювання першого рівня (сегментації) застосовувались метрики Dice та IoU, оскільки клас об'єкта займав значну частку зображення, і показники Precision/Recall були менш інформативними. Для оцінювання другого рівня (локального аналізу в межах ROI) використовувались усі зазначені метрики, з акцентом на Dice (загальна якість локалізації) та Precision (зменшення хибних спрацьовувань поза ROI).

Для перевірки ефективності дворівневого підходу проведено серію експериментів, у яких запропонована система порівнювалася з трьома базовими моделями: U-Net, U-Net 3+ та MobileMamba-U-Net (без ROI-маскування). Кожна модель навчалась і тестувалась на однакових даних із тими самими функціями втрат і параметрами оптимізації.

Для об'єктивності оцінки використовувався підхід 5-fold cross-validation, де результати усереднювались за п'ятьма незалежними експериментами. Для кожної моделі визначалися середні значення метрик (Dice, IoU, Precision, Recall, F1) та стандартні відхилення. Додатково вимірювався час інференсу (inference time) для одного зображення, щоб оцінити обчислювальну ефективність. Це дозволило продемонструвати, що запропонований дворівневий підхід не лише підвищує точність сегментації, а й залишається придатним для застосування в реальних системах автоматизованого аналізу зображень.

Аналіз отриманих результатів

Для оцінювання ефективності запропонованого методу було проведено серію експериментів, у яких результати дворівневої системи порівнювалися з базовими моделями: U-Net, U-Net 3+ та MobileMamba-U-Net (без ROI-маскування). Усі моделі навчалися на однакових даних і перевірялися за однаковими метриками. Результати узагальнено у таблиці 1.

Таблиця 1.

Порівняння показників ефективності різних моделей сегментації та локального аналізу зображень.

Модель	Dice	IoU	Precision	Recall	F1-score
U-Net	0.65	0.48	0.78	0.85	0.81
U-Net 3+	0.68	0.51	0.82	0.87	0.84
MobileMamba-U-Net	0.70	0.54	0.86	0.88	0.87
Запропонований метод (ROI + ResNet-U-Net)	0.74	0.59	0.88	0.90	0.89

Отримані результати демонструють, що впровадження етапу ROI-маскування забезпечує стабільне підвищення всіх основних метрик. Зокрема, показник Dice зріс на 4–6% порівняно з MobileMamba-U-Net, що свідчить про точніше відтворення контурів цільових областей. Підвищення Precision до 0.88 вказує на зменшення кількості хибнопозитивних виявлень, тоді як Recall = 0.90 підтверджує здатність системи коректно розпізнавати більшість об'єктів у межах ROI.

Результати показують, що поєднання сегментаційного етапу та ROI-маскування є ключовим для підвищення точності. ROI-маска обмежує область аналізу лише релевантними ділянками, що суттєво зменшує кількість помилкових передбачень у фонових зонах або регіонах, не пов'язаних із цільовими структурами. Завдяки цьому друга модель (ResNet-U-Net) фокусується на інформативних областях, що забезпечує більш чисте вилучення ознак та покращення загальної ефективності.

Аналіз показників Precision і Recall свідчить, що основне покращення відбулося саме в показнику точності (Precision), тобто система стала менш схильною до хибних класифікацій фонових ділянок. Це підтверджує ефективність ROI-маскування як механізму зниження впливу шуму. Підвищення Recall при цьому демонструє здатність моделі виявляти навіть дрібні локальні об'єкти, що підтверджує її стійкість до варіацій контрасту.

Отримані результати узгоджуються з висновками попередніх досліджень [11, 12], у яких показано ефективність MobileMamba-U-Net у задачах сегментації складних структур цифрових зображень. Запропонований у цій роботі підхід є розвитком попередніх рішень, адже переносить акцент із суто сегментації на багаторівневий локальний аналіз із каскадною архітектурою. Інтеграція ResNet-блоків у модель другого рівня дозволила зберегти швидкодію при високій точності локалізації.

Порівняно з одноетапними системами, що виконують розпізнавання без попереднього ROI, запропонований метод демонструє стабільнішу роботу в умовах варіацій експозиції, якості та рівня шуму

зображень. Скорочення кількості параметрів і використання легковагових архітектур робить можливим застосування моделі в системах реального часу або мобільних додатках для автоматизованого аналізу цифрових зображень.

Основними перевагами запропонованого методу є поєднання точності та обчислювальної ефективності. Завдяки попередньому виділенню ROI модель уникає зайвих обчислень, що скорочує час інференсу приблизно на 35% порівняно з одноетапними моделями. Крім того, метод показує високу стабільність при зміні контрасту, освітлення та якості вхідних даних.

Серед обмежень можна відзначити, що якість локалізації залежить від точності первинної сегментації. У випадках некоректного формування маски (наприклад, при сильних артефактах або часткових втратах об'єктів) ROI може бути визначена неточно, що знижує ефективність другого рівня. Крім того, поточна конфігурація орієнтована на повнорозмірні зображення, тому для зменшених фрагментів або специфічних типів даних може знадобитися адаптація параметрів.

Запропонований дворівневий метод довів свою ефективність у задачах автоматизованого розпізнавання складних цифрових зображень. Використання ROI-маскування дозволяє оптимізувати процес обробки, підвищуючи точність і швидкість без збільшення кількості параметрів. Отримані результати перевищують показники базових моделей, що підтверджує доцільність каскадного підходу «сегментація → локальний аналіз» для побудови інтелектуальних інформаційних систем комп'ютерного зору.

Висновки з даного дослідження

і перспективи подальших розвідок у даному напрямі

У результаті проведеного дослідження розроблено та експериментально перевірено дворівневий метод автоматизованого аналізу цифрових зображень, який поєднує ROI-маскування з модифікованою архітектурою U-Net 3+, що інтегрує MobileNet та Vision Mamba Layers. Запропонована система реалізує концепцію поетапного розпізнавання, де перший рівень відповідає за сегментацію основних структур і формування області інтересу, а другий — за локальний аналіз ROI за допомогою легковагової моделі ResNet-U-Net. Такий підхід дозволив ефективно обмежити область обчислень лише релевантними ділянками, зменшивши вплив фону та шумів.

Отримані результати показують, що запропонована система перевершує традиційні архітектури U-Net та U-Net 3+ за більшістю метрик точності. Підвищення Dice і Precision досягнуто завдяки поєднанню повношкільних зв'язків U-Net 3+, оптимізованих блоків MobileNet, а також контекстних механізмів Vision Mamba Layers, які ефективно моделюють просторові залежності між пікселями. Інтеграція ROI-маскування зменшила кількість хибнопозитивних визначень, водночас зберігаючи високу чутливість (Recall). Таким чином, система продемонструвала стабільну роботу навіть за наявності шумів, варіацій контрасту та неідеальних умов отримання даних.

Розроблений метод поєднує високу точність розпізнавання з помірною обчислювальною складністю. Завдяки використанню легковагових архітектур кількість параметрів моделі вдалося зменшити більш ніж утричі порівняно з класичними рішеннями U-Net, що відкриває перспективи інтеграції технології в системи комп'ютерного зору, мобільні застосунки та програмні комплекси для автоматизованої обробки зображень.

Результати дослідження також підтвердили важливість коректного формування ROI, оскільки точність первинної сегментації безпосередньо впливає на ефективність другого рівня. Подальше вдосконалення методу передбачає використання attention-механізмів або просторових трансформерів (Spatial Transformers) для динамічного виділення областей інтересу. Це дозволить враховувати індивідуальні особливості структури зображень і підвищити стійкість до артефактів.

У перспективі планується адаптація методу для різних типів зображень, зокрема промислових і технічних, а також розширення навчальної вибірки для покращення узагальнення. Окрему увагу буде приділено розробці multi-task архітектур, які зможуть одночасно виконувати сегментацію, класифікацію та локалізацію кількох типів об'єктів. Подальша оптимізація коду та апаратна реалізація дозволять забезпечити роботу системи в реальному часі, що робить її придатною для інтеграції в інтелектуальні системи аналізу цифрових даних.

Проведені експерименти на рентгеновських зображеннях стоматологічного типу підтвердили ефективність запропонованої технології в контексті обробки медичних даних. Отримані результати демонструють, що метод здатен зберігати високу точність локалізації навіть у випадках низького контрасту та наявності шумів, що робить його перспективним для інтеграції у системи автоматизованого аналізу медичних зображень.

Таким чином, запропонований метод довів свою ефективність і перспективність у контексті сучасних підходів до побудови високоефективних систем комп'ютерного зору. Він формує основу для створення нових поколінь алгоритмічних рішень, орієнтованих на підвищення точності, швидкості та адаптивності процесів аналізу зображень у різних галузях.

Література

1. U-Net: Convolutional Networks for Biomedical Image Segmentation / Olaf Ronneberger, Philipp Fischer, Thomas Brox // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). – 2015. – С. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
2. UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Zongwei Zhou [та ін.] // Deep

- Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. – 2018. – С. 3–11. DOI: 10.1007/978-3-030-00889-5_1.
3. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation / Heng Huang [та ін.] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – С. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405.
 4. Deep Learning Based Detection of Dental Caries in Panoramic Radiographs Using Nested U-Net / Shuaa Saud Alharbi [та ін.] // Computers in Biology and Medicine. – 2023. – Т. 160. – Article № 106952. DOI: 10.1016/j.combiomed.2023.106952.
 5. Teeth Segmentation and Carious Lesions Segmentation in Panoramic X-ray Images Using CariSeg / Andrei C. Mărginean [та ін.] // Heliyon. – 2024. – Т. 10, № 2. – e25133. DOI: 10.1016/j.heliyon.2024.e25133.
 6. Clinical Application of Deep Learning for Enhanced Multistage Caries Detection in Panoramic Radiographs / S. Pornprasertsuk-Damrongsri [та ін.] // Scientific Reports. – 2025. – Т. 15, № 7. – С. 1243–1252. DOI: 10.1038/s41598-025-12432-y.
 7. MobileNetV2: Inverted Residuals and Linear Bottlenecks / Mark Sandler [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – С. 4510–4520. DOI: 10.1109/CVPR.2018.00474.
 8. Squeeze U-Net: A Memory and Energy Efficient Image Segmentation Network / Negin Beheshti [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2020. – С. 194–201. DOI: 10.1109/CVPRW50498.2020.00190.
 9. Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation / Zhiqiang Wang [та ін.] // arXiv preprint arXiv:2402.05079. – 2024. DOI: 10.48550/arXiv.2402.05079.
 10. LightM-UNet: Mamba Assists in Lightweight UNet for Medical Image Segmentation / Wei Liao [та ін.] // arXiv preprint arXiv:2403.05246. – 2024. DOI: 10.48550/arXiv.2403.05246.
 11. Шабатура Ю. В., Рыбак В. Р. Оцінювання ефективності застосування моделі U-Net 3+ у процесі сегментації анатомічних структур рентгенівських знімків зубів // Науковий вісник ХНТУ України. – 2024. – №34(7). – С. 145–152. DOI: 10.36930/40340718.
 12. Шабатура Ю. В., Рыбак В. Р. Інформаційна технологія сегментації стоматологічних зображень на основі інтеграції легковагової модифікації U-Net із MobileNet та Vision Mamba // Вісник Хмельницького національного університету. Технічні науки. – 2025. – №2. – С. 371–378. DOI: 10.31891/2307-5732-2026-349-54.

References

1. U-Net: Convolutional Networks for Biomedical Image Segmentation / Olaf Ronneberger, Philipp Fischer, Thomas Brox // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). – 2015. – С. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
2. UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Zongwei Zhou [та ін.] // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. – 2018. – С. 3–11. DOI: 10.1007/978-3-030-00889-5_1.
3. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation / Heng Huang [та ін.] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – С. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405.
4. Deep Learning Based Detection of Dental Caries in Panoramic Radiographs Using Nested U-Net / Shuaa Saud Alharbi [та ін.] // Computers in Biology and Medicine. – 2023. – Vol. 160. – Article № 106952. DOI: 10.1016/j.combiomed.2023.106952.
5. Teeth Segmentation and Carious Lesions Segmentation in Panoramic X-ray Images Using CariSeg / Andrei C. Mărginean [та ін.] // Heliyon. – 2024. – Vol. 10, № 2. – e25133. DOI: 10.1016/j.heliyon.2024.e25133.
6. Clinical Application of Deep Learning for Enhanced Multistage Caries Detection in Panoramic Radiographs / S. Pornprasertsuk-Damrongsri [та ін.] // Scientific Reports. – 2025. – Vol. 15, № 7. – P. 1243–1252. DOI: 10.1038/s41598-025-12432-y.
7. MobileNetV2: Inverted Residuals and Linear Bottlenecks / Mark Sandler [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – С. 4510–4520. DOI: 10.1109/CVPR.2018.00474.
8. Squeeze U-Net: A Memory and Energy Efficient Image Segmentation Network / Negin Beheshti [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2020. – С. 194–201. DOI: 10.1109/CVPRW50498.2020.00190.
9. Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation / Zhiqiang Wang [та ін.] // arXiv preprint arXiv:2402.05079. – 2024. DOI: 10.48550/arXiv.2402.05079.
10. LightM-UNet: Mamba Assists in Lightweight UNet for Medical Image Segmentation / Wei Liao [та ін.] // arXiv preprint arXiv:2403.05246. – 2024. DOI: 10.48550/arXiv.2403.05246.
11. Shabatura Yu. V., Rybak V. R. Evaluation of the Effectiveness of the U-Net 3+ Model in the Process of Segmenting Anatomical Structures in Dental X-ray Images // Scientific Bulletin of UNFU. – 2024. – № 34(7). – P. 145–152. DOI: <https://doi.org/10.36930/40340718>.
12. Shabatura Yu. V., Rybak V. R. Information Technology for Segmentation of Dental Images Based on the Integration of a Lightweight U-Net Modification with MobileNet and Vision Mamba Layers // Herald of Khmelnytskyi National University. Technical Sciences. – 2025. – №2. – P. 371–378. DOI: <https://doi.org/10.31891/2307-5732-2026-349-54>.