

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Дніпровський національний університет імені Олеся Гончара
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Дніпровський національний університет імені Олеся Гончара

Кваліфікаційна наукова
праця на правах рукопису

СОЛОМАТІН ВЛАДИСЛАВ АНДРІЙОВИЧ

УДК 004.42

ДИСЕРТАЦІЯ

**РОЗРОБЛЕННЯ МОДЕЛЕЙ ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ
ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ПІД ЧАС ДІАГНОСТИКИ ЗА
МЕДИЧНИМИ ДАНИМИ**

12 Інформаційні технології

121 Інженерія програмного забезпечення

Подається на здобуття ступеня доктора філософії. Дисертація містить результати власних досліджень. Використання ідей, результатів та текстів інших авторів мають посилання на відповідне джерело.

_____ В.А. Соломатін

Науковий керівник:
Антоненко Світлана Валентинівна
кандидат технічних наук, доцент

Дніпро – 2026

АНОТАЦІЯ

Соломатін В.А. Розроблення моделей та програмного забезпечення підтримки прийняття рішень під час діагностики за медичними даними – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення – Дніпровський національний університет імені Олеся Гончара, Дніпро, 2026.

Дисертаційну роботу присвячено розробленню програмного забезпечення та методів глибинного навчання для автоматизованої сегментації серцевої області на рентгенівських знімках грудної клітки (CXR) у задачах підтримки прийняття рішень під час первинної діагностики серцево-судинних захворювань, зокрема ішемічної хвороби серця. У роботі поєднано морфологічний аналіз, edge-aware підходи та глибинні нейронні мережі, що дозволяє точніше виділяти ROI серця, враховувати його геометричну форму й контурні характеристики, зменшувати невизначеність сегментації та кількість хибнопозитивних включень фону.

Розроблено спеціалізовану архітектуру CardioEdgeBioMorphNet, яка формує піксельну карту сегментації на основі локальних, глобальних, морфологічних і контурних ознак. До її складу входять EdgeMorphBlock, Sobel edge-потік з edge-fusion у декодері, диференційовний механізм уваги DMGGA та Adaptive Boundary Rejection, що забезпечують стабілізацію контурів ROI, підсилення меж і текстур та відкидання невпевнених пікселів у фон.

Експериментальне дослідження виконано на CXR-наборі даних зі структурою типу VOC і фіксованим розбиттям на Train, Validation і Test. Розмітка подана RGB-масками з палітрою класів: шість клінічних станів і background. Якість сегментації оцінювалася за метриками Acc_heart, Dice та

IoU, а перевага запропонованих компонентів підтверджено порівнянням із базовими архітектурами та абляційним аналізом.

Актуальність теми зумовлена поширеністю серцево-судинних захворювань, зокрема ішемічної хвороби серця, та потребою в удосконаленні методів їх ранньої діагностики. Хоча рентгенографія грудної клітки є доступним методом первинного обстеження, інтерпретація CXR-знімків ускладнюється шумами, артефактами, накладанням анатомічних структур і залежністю від досвіду лікаря. Тому актуальним є розроблення методів глибокого навчання, які враховують морфологічні й контурні характеристики серцевої області та забезпечують надійну автоматизовану сегментацію серця на CXR-зображеннях.

Метою роботи є створення моделі та програмного забезпечення для підтримки прийняття рішень під час автоматизованого аналізу CXR-зображень шляхом надійного виділення серцевої ROI та формування інформативних результатів сегментації для виявлення ознак ішемічної хвороби серця. Досягнення цієї мети передбачає розроблення спеціалізованої нейронної архітектури сегментації та програмних засобів, що зменшують вплив людського фактору, підвищують точність і відтворюваність аналізу та скорочують час обробки медичних зображень.

Об'єктом дослідження є процес автоматизованого аналізу CXR-зображень для виділення серцевої ROI, оцінювання її морфологічних характеристик та інтерпретації результатів у системах підтримки прийняття рішень.

Предметом дослідження є методи, моделі й алгоритми глибокого навчання для сегментації серцевої області на CXR-знімках, зокрема encoder–decoder архітектури з морфологічними, контурними та багатомасштабними механізмами.

У першому розділі проаналізовано класичні та сучасні методи сегментації медичних зображень, зокрема CXR, визначено їхні обмеження при виділенні серцевої ROI та обґрунтовано потребу в поєднанні морфологічних і

edge-aware механізмів.

У **другому розділі** сформульовано задачу сегментації серцевої області, описано CXR-набір даних, структуру розмітки, поділ на train, validation і test, етапи передоброби, аугментації та метрики оцінювання Dice, IoU, Acc_heart.

У **третьому розділі** описано розроблення морфологічної архітектури CardioEdgeBioMorphNet, побудованої за принципом encoder–decoder та доповненої спеціалізованими морфологічно-орієнтованими й edge-aware компонентами. Розглянуто структуру моделі, зокрема блоки EdgeMorphBlock для формування стійких ознак серцевої ROI, Sobel edge-гілку для явного врахування контурної інформації, модуль ASPP для багатомасштабного контексту, SoftMorphPool2D і DMGGA для підсилення морфологічно значущих меж і текстур, а також edge-fusion у декодері. Окремо описано механізм Adaptive Boundary Rejection, який відкидає невпевнені пікселі у фоновий клас, що сприяє уточненню меж серця та зменшенню хибнопозитивних включень.

У **четвертому розділі** описано програмну реалізацію, експериментальний конвеєр навчання й тестування моделі, порівняння з базовими архітектурами, абляційний аналіз і візуальну перевірку результатів. Отримані результати підтвердили перевагу запропонованого підходу для автоматизованого виділення серцевої області на CXR-знімках у задачах підтримки прийняття рішень.

У **висновках** узагальнено основні результати, отримані в межах дослідження, сформульовано ключові висновки та наведено рекомендації щодо впровадження розробленої моделі й програмного забезпечення у практику автоматизованого аналізу медичних зображень. Окреслено напрями подальших досліджень і можливості розвитку запропонованого підходу.

Наукова новизна роботи полягає в наступному:

1. Вперше розроблено морфологічно-орієнтовану архітектуру CardioEdgeBioMorphNet для сегментації серцевої ROI на CXR-зображеннях із поєднанням багатомасштабних ознак і контурної інформації.

2. Удосконалено базовий MorphBlock шляхом введення edge-aware компонентів, channel attention та морфологічного gating для підвищення точності локалізації меж серця.

3. Вперше запропоновано диференційовний морфологічний механізм уваги DMGGA з multi-kernel сумішшю для підсилення меж і текстур у глибоких ознаках сегментації.

4. Вперше використано явний edge-потік на основі Sobel magnitude з подальшим edge-fusion у декодері для точнішого виділення прикордонних зон серця.

5. Вперше реалізовано механізм Adaptive Boundary Rejection у вигляді окремої порогової гілки для відкидання невпевнених пікселів у фоновий клас.

6. Вперше експериментально підтверджено перевагу комплексного поєднання морфологічних блоків, edge-aware механізмів, багатомасштабного контексту та адаптивного порогового відкидання.

7. Вперше поєднано результат сегментації з автоматизованим формуванням висновку щодо клінічного класу на основі ймовірностей у межах ROI та показника confidence margin top1–top2.

Практичне значення результатів полягає у створенні моделі CardioEdgeBioMorphNet і програмного забезпечення для автоматизованої сегментації серцевої області на CXR-зображеннях та формування допоміжного класифікаційного висновку.

1. Розроблено програмне забезпечення для повного циклу автоматизованого аналізу CXR-зображень: від завантаження та передоброби до сегментації, класифікації й візуалізації результатів.

2. Забезпечено автоматизоване виділення серцевої ROI без ручного окреслення меж, що скорочує час обробки та підвищує відтворюваність аналізу.

3. Реалізовано визначення ймовірного клінічного стану серця на основі CXR-зображення для використання моделі як допоміжного інструменту підтримки прийняття рішень.

4. Підвищено якість відтворення меж серцевої області завдяки поєднанню EdgeMorphBlock, Sobel edge-потoku, морфологічної уваги та багатомасштабного аналізу.

5. Зменшено кількість хибнопозитивних включень фону за рахунок механізму Adaptive Boundary Rejection, який відносить невпевнені пікселі до фонового класу.

6. Створено основу для стандартизації первинного аналізу CXR-знімків завдяки єдиному протоколу передобробки, фіксованим метрикам і відтворюваній процедурі тестування.

7. Реалізовано прикладний інтерфейс для перегляду сегментаційної маски, її накладання на CXR-зображення, аналізу контурів, клінічного класу та рівня впевненості моделі.

8. Забезпечено можливість використання ROI-маски для уточнення класифікаційного висновку шляхом аналізу ймовірностей у межах серцевої області.

9. Результати дослідження можуть бути використані в клінічній, науково-дослідній та навчальній діяльності як допоміжний інструмент аналізу медичних зображень.

Таким чином, практичне значення роботи полягає у створенні прикладного програмного інструментарію для автоматизованого, відтворюваного та стійкого виділення серцевої області на CXR-зображеннях і формування допоміжного висновку щодо ймовірного клінічного стану.

Ключові слова: інформаційна відстань, трансформери, великі мовні моделі, критерії, штучний інтелект, нейронні мережі, машинне навчання, глибинне навчання, класифікація зображень, розпізнавання образів, архітектура програмного забезпечення, програмне забезпечення, інформаційна система, інженерія програмного забезпечення, інформаційні технології.

ABSTRACT

Solomatin V. A. Development of Models and Software to Support Decision Making During Diagnostics Using Medical Data – a qualification scientific work submitted as a manuscript.

Dissertation for the degree of Doctor of Philosophy (PhD) in the field of knowledge 12 Information Technologies, specialty 121 Software Engineering — Oles Honchar Dnipro National University, Dnipro, 2026.

The dissertation is devoted to the development of software and deep learning methods for automated segmentation of the cardiac region in chest X-ray images (CXR) in decision-support tasks for the primary diagnosis of cardiovascular diseases, particularly ischemic heart disease. The study combines morphological analysis, edge-aware approaches, and deep neural networks, enabling more accurate delineation of the cardiac ROI, consideration of its geometric shape and contour characteristics, reduction of segmentation uncertainty, and a decrease in false-positive background inclusions.

A specialized architecture, CardioEdgeBioMorphNet, has been developed to generate a pixel-wise segmentation map based on local, global, morphological, and contour features. It includes EdgeMorphBlock, a Sobel-based edge stream with edge fusion in the decoder, the differentiable attention mechanism DMGGA, and Adaptive Boundary Rejection, which provide ROI contour stabilization, boundary and texture enhancement, and rejection of uncertain pixels into the background class.

The experimental study was conducted on a CXR dataset organized in a VOC-like structure with a fixed split into Train, Validation, and Test subsets. The annotations are represented by RGB masks with a class palette consisting of six clinical states and background. Segmentation quality was evaluated using Acc_heart, Dice, and IoU metrics, while the effectiveness of the proposed components was confirmed by comparison with baseline architectures and ablation analysis.

The relevance of the topic is determined by the prevalence of cardiovascular diseases, particularly ischemic heart disease, and the need to improve methods for their early diagnosis. Although chest radiography is an accessible method of primary examination, the interpretation of CXR images is complicated by noise, artifacts, overlapping anatomical structures, and dependence on the physician's experience. Therefore, the development of deep learning methods that take into account the morphological and contour characteristics of the cardiac region and provide reliable automated heart segmentation in CXR images is relevant.

The aim of the work is to create a model and software for decision support in the automated analysis of CXR images by reliably delineating the cardiac ROI and generating informative segmentation results for detecting signs of ischemic heart disease. Achieving this aim involves the development of a specialized neural segmentation architecture and software tools that reduce the influence of the human factor, improve the accuracy and reproducibility of analysis, and shorten the time required for medical image processing.

The object of the research is the process of automated analysis of CXR images for delineating the cardiac ROI, assessing its morphological characteristics, and interpreting the results in decision-support systems.

The subject of the research comprises deep learning methods, models, and algorithms for cardiac region segmentation in CXR images, particularly encoder–decoder architectures with morphological, contour-based, and multi-scale mechanisms.

Chapter 1 analyzes classical and modern methods for medical image segmentation, particularly CXR images, identifies their limitations in delineating the cardiac ROI, and substantiates the need to combine morphological and edge-aware mechanisms.

Chapter 2 formulates the task of cardiac region segmentation and describes the CXR dataset, annotation structure, split into train, validation, and test subsets, preprocessing stages, augmentation strategies, and evaluation metrics such as Dice, IoU, and Acc_heart.

Chapter 3 describes the development of the morphology-oriented CardioEdgeBioMorphNet architecture, built according to the encoder–decoder principle and supplemented with specialized morphology-oriented and edge-aware components. The model structure is considered, including EdgeMorphBlock modules for forming robust cardiac ROI features, a Sobel edge branch for explicit use of contour information, the ASPP module for multi-scale context, SoftMorphPool2D and DMGGA for enhancing morphologically significant boundaries and textures, as well as edge fusion in the decoder. The Adaptive Boundary Rejection mechanism is also described separately; it rejects uncertain pixels into the background class, contributing to boundary refinement and reduction of false-positive inclusions.

Chapter 4 presents the software implementation, the experimental pipeline for model training and testing, comparison with baseline architectures, ablation analysis, and visual verification of the results. The obtained results confirmed the effectiveness of the proposed approach for automated delineation of the cardiac region in CXR images in decision-support tasks.

The conclusions summarize the main results obtained in the study, formulate the key findings, and provide recommendations for implementing the developed model and software in automated medical image analysis. Directions for further research and opportunities for the development of the proposed approach are also outlined.

The scientific novelty of the work is as follows:

1. **For the first time**, a morphology-oriented CardioEdgeBioMorphNet architecture has been developed for cardiac ROI segmentation in CXR images by combining multi-scale features and contour information.

2. **Modified** The basic MorphBlock has been improved by introducing edge-aware components, channel attention, and morphological gating to increase the accuracy of cardiac boundary localization.

3. **For the first time**, a differentiable morphological attention mechanism, DMGGA, with a multi-kernel mixture has been proposed to enhance boundaries and

textures in deep segmentation features.

4. **For the first time**, an explicit edge stream based on Sobel magnitude with subsequent edge fusion in the decoder has been used for more accurate delineation of cardiac boundary regions.

5. **For the first time**, an Adaptive Boundary Rejection mechanism has been implemented as a separate threshold branch for rejecting uncertain pixels into the background class.

6. **For the first time**, the effectiveness of the integrated combination of morphological blocks, edge-aware mechanisms, multi-scale context, and adaptive threshold rejection has been experimentally confirmed.

7. **For the first time**, the segmentation result has been combined with automated generation of a clinical class conclusion based on probabilities within the ROI and the top1–top2 confidence margin.

The practical significance of the results lies in the creation of the CardioEdgeBioMorphNet model and software for automated segmentation of the cardiac region in CXR images and the formation of an auxiliary classification conclusion.

1. Software has been developed for the full cycle of automated CXR image analysis, from image loading and preprocessing to segmentation, classification, and visualization of results.

2. Automated delineation of the cardiac ROI without manual boundary outlining has been provided, which reduces processing time and improves the reproducibility of analysis.

3. Determination of the probable clinical state of the heart based on CXR images has been implemented for using the model as an auxiliary decision-support tool.

4. The quality of cardiac boundary reconstruction has been improved by combining EdgeMorphBlock, the Sobel edge stream, morphological attention, and multi-scale analysis.

5. The number of false-positive background inclusions has been reduced

through the Adaptive Boundary Rejection mechanism, which assigns uncertain pixels to the background class.

6. A basis for standardizing primary CXR image analysis has been created through a unified preprocessing protocol, fixed evaluation metrics, and a reproducible testing procedure.

7. An applied interface has been implemented for viewing the segmentation mask, overlaying it on the CXR image, analyzing contours, the clinical class, and the model confidence level.

8. The possibility of using the ROI mask to refine the classification conclusion by analyzing probabilities within the cardiac region has been provided.

9. The research results can be used in clinical, research, and educational activities as an auxiliary tool for medical image analysis.

10. Thus, the practical significance of the work lies in the creation of applied software tools for automated, reproducible, and robust delineation of the cardiac region in CXR images and the formation of an auxiliary conclusion regarding the probable clinical state.

Keywords: information distance, transformers, large language models, criteria, artificial intelligence, neural networks, machine learning, deep learning, image classification, pattern recognition, software architecture, software, information system, software engineering, information technology.

Список опублікованих праць за темою дисертації

Статті у наукових фахових виданнях України:

1. Соломатін В.А., Байбуз О.Г., Сиротенко А.Є. Застосування згорткових нейронних мереж для діагностування ішемічного захворювання серця (ІЗС). *Системні технології*. 2023. Т. 6 (149). С. 53–61. DOI: <https://doi.org/10.34185/1562-9945-6-149-2023-05> (особистий внесок Соломатіна В.А.: проведення аналізу предметної області та постановка задачі застосування згорткових нейронних мереж для діагностування ішемічної хвороби серця за медичними зображеннями, підготовка набору рентгенівських знімків грудної клітки, попередня обробка даних, реалізація та навчання CNN-моделі, проведення обчислювальних експериментів, аналіз отриманих результатів і формування висновків щодо ефективності застосування згорткових нейронних мереж; Байбуза О.Г.: наукове консультування, уточнення мети і завдань дослідження, аналіз та узагальнення отриманих результатів; Сиротенка А.Є.: участь в обговоренні медичної складової дослідження та інтерпретації результатів у контексті діагностики ішемічної хвороби серця).

2. Соломатін В.А., Мащенко Л.В. Застосування математичних моделей при обробці медичних даних діагностування ішемічної хвороби серця. *Актуальні проблеми автоматизації та інформаційних технологій*. 2023. Т. 27. С. 178–187. DOI: <http://dx.doi.org/10.15421/432317> (особистий внесок Соломатіна В.А.: формалізація задачі обробки медичних даних для діагностування ішемічної хвороби серця, аналіз можливостей застосування математичних моделей у задачах медичної діагностики, підготовка вхідних даних, виконання експериментальної частини дослідження, аналіз результатів застосування моделей та формулювання висновків; Мащенко Л.В.: участь у теоретичному обґрунтуванні використання математичних моделей, науково-методичне консультування та узагальнення результатів

дослідження).

3. Соломатін В.А., Байбуз О.Г. Застосування моделі EfficientNet для виявлення ішемічної хвороби серця (ІХС). *Системні технології*. 2025. Т. 1 (156). С. 160–165. DOI: <https://doi.org/10.34185/1562-9945-1-156-2025-18> (особистий внесок Соломатіна В.А.: обґрунтування доцільності використання архітектури EfficientNet для аналізу рентгенівських знімків грудної клітки, підготовка та попередня обробка набору даних, адаптація моделі EfficientNet до задачі виявлення ішемічної хвороби серця, реалізація програмного коду навчання й тестування моделі, проведення експериментів, побудова та аналіз метрик якості класифікації; Байбуза О.Г.: постановка наукової задачі, наукове керівництво дослідженням, аналіз і узагальнення отриманих експериментальних результатів).

4. Соломатін В.А., Байбуз О.Г. Використання архітектури трансформера для класифікації ішемічної хвороби серця за допомогою рентгенографічних зображень. *Актуальні проблеми автоматизації та інформаційних технологій*. 2024. Т. 28. С. 244–253. DOI: <http://dx.doi.org/10.15421/432423> (особистий внесок Соломатіна В.А.: аналіз можливостей застосування архітектур трансформерів до задачі класифікації рентгенографічних зображень, підготовка та структурування набору CXR-знімків, адаптація transformer-based моделі до задачі класифікації ішемічної хвороби серця, реалізація програмного забезпечення для навчання та тестування моделі, проведення експериментальних досліджень, аналіз точності класифікації та порівняння отриманих результатів із результатами інших нейромережесевих підходів; Байбуза О.Г.: формування загальної концепції дослідження, наукове керівництво, аналіз та інтерпретація отриманих результатів).

5. Соломатін В.А., Байбуз О.Г. Архітектура BioMorphNet для автоматизованої бінарної класифікації патологій серця за рентгенівськими знімками. *Актуальні проблеми автоматизації та інформаційних технологій*. 2025. Т. 29. С. 382–394. DOI: <http://dx.doi.org/10.15421/432535>. (особистий внесок Соломатіна В.А.: розроблення концепції та структури архітектури

BioMorphNet для автоматизованої бінарної класифікації патологій серця за рентгенівськими знімками, обґрунтування використання морфологічно-орієнтованих компонентів у нейронній моделі, реалізація програмного коду моделі, підготовка та попередня обробка CXR-даних, проведення навчання й тестування моделі, оцінювання якості класифікації, аналіз експериментальних результатів і формування висновків щодо ефективності запропонованої архітектури; Байбуза О.Г.: наукове керівництво, уточнення постановки задачі, аналіз і узагальнення результатів дослідження).

Наукові праці, які засвідчують апробацію матеріалів дисертації:

6. Соломатін В.А., Ємел'яненко Т.Г. Огляд методів підтримки прийняття рішень під час діагностики за медичними даними. *Наука, освіта та суспільство в XXI столітті: наукові ідеї та механізми реалізації: тези доповідей міжнародної науково-практичної конференції*, Полтава, 2021. С. 53–54. URL: <https://www.economics.in.ua/2022/01/xxi-2.html> (особистий внесок Соломатіна В.А.: аналіз сучасних підходів до підтримки прийняття рішень під час діагностики за медичними даними, систематизація методів обробки медичної інформації, узагальнення можливостей застосування математичних моделей і методів машинного навчання в медичній діагностиці, підготовка матеріалів тез; Ємел'яненко Т.Г.: постановка завдання дослідження, науково-методичне консультування, аналіз і узагальнення отриманих результатів).

7. Соломатін В.А., Ємел'яненко Т.Г. Створення датасету рентгенівських зображень для діагностики ішемічної хвороби серця. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2023): тези доповідей XXI Міжнародної науково-практичної конференції*, Дніпро, 2023. С. 276. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2023/11/mpzis-2023.pdf> (особистий внесок Соломатіна В.А.: формування набору рентгенівських зображень грудної клітки для задачі діагностики ішемічної хвороби серця, структурування даних за класами,

попередня обробка зображень, підготовка даних до використання у нейромережових моделях, аналіз вимог до якості та придатності датасету для подальших експериментів; Ємел'яненко Т.Г.: постановка завдання, науково-методичне консультування, аналіз і узагальнення результатів підготовки даних).

8. Соломатін В.А., Байбуз О.Г. Застосування трансформерних моделей для автоматизованої діагностики ішемічного захворювання серця на основі рентгенівських зображень. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2024): тези доповідей XXII Міжнародної науково-практичної конференції*, Дніпро, 2024. С. 257. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2024/11/МПЗІС-2024-1.pdf>

(особистий внесок Соломатіна В.А.: аналіз можливостей використання трансформерних моделей для класифікації рентгенівських зображень грудної клітки, підготовка набору CXR-знімків, адаптація архітектури трансформера до задачі автоматизованого виявлення ішемічної хвороби серця, реалізація програмного коду, проведення навчання та тестування моделі, аналіз отриманих результатів; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

9. Соломатін В.А., Байбуз О.Г. Розробка та тренування нейронної мережі для класифікації рентгенівських знімків на основі EfficientNet: методи попередньої обробки, аугментації та аналіз стабільності навчання. *Автоматика 2024: Тези доповідей XXVII Міжнародної конференції*, Дніпро, 2024. С. 191–192. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2025/11/Автоматика-2024-тези-доповідей.pdf>

(особистий внесок Соломатіна В.А.: обґрунтування використання архітектури EfficientNet для класифікації рентгенівських зображень грудної клітки, підготовка та попередня обробка даних, застосування методів аугментації, реалізація програмного коду навчання нейронної мережі, проведення експериментів, аналіз стабільності навчання та оцінювання якості класифікації; Байбуза О.Г.: постановка задачі дослідження, наукове

керівництво, аналіз і узагальнення отриманих результатів).

10. Соломатін В.А., Байбуз О.Г. Архітектура BioMorphNet для автоматизованого виявлення ішемічної хвороби серця за рентгенівськими знімками. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2025): тези доповідей XXIII Міжнародної науково-практичної конференції*, Дніпро, 2025. С. 286–287. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2025/11/МПЗІС-2025.pdf> (особистий внесок Соломатіна В.А.: розроблення концепції архітектури BioMorphNet для автоматизованого виявлення ішемічної хвороби серця за рентгенівськими знімками, обґрунтування використання морфологічно-орієнтованих компонентів, реалізація програмної моделі, підготовка CXR-даних, проведення навчання та тестування, оцінювання якості роботи моделі й аналіз отриманих результатів; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

11. Соломатін В.А., Байбуз О.Г. Поєднання згорткових мереж, механізмів уваги та морфологічного аналізу у задачі виявлення ішемічної хвороби серця за рентгенівськими знімками. *Глобальні виклики та інновації: шляхи розвитку сучасної науки: матеріали X Міжнародної наукової конференції*, Одеса, 2026. С. 363–365. URL: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/16.01.2026/76> (особистий внесок Соломатіна В.А.: обґрунтування доцільності поєднання згорткових нейронних мереж, механізмів уваги та морфологічного аналізу в задачі виявлення ішемічної хвороби серця за рентгенівськими знімками, розроблення та опис відповідного підходу, підготовка рентгенівських зображень до експериментального аналізу, реалізація програмних компонентів, аналіз впливу морфологічних і attention-механізмів на якість розпізнавання; Байбуза О.Г.: постановка задачі, наукове керівництво, аналіз і узагальнення результатів).

12. Соломатін В.А., Байбуз О.Г. Морфологічно-орієнтована нейронна архітектура для сегментації серця на рентгенівських знімках.

Міжгалузеві диспути: динаміка та розвиток сучасних наукових досліджень: матеріали IX Міжнародної наукової конференції, Харків, 2026. С. 253–254. URL: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/30.01.2026/77> (особистий внесок Соломатіна В.А.: розроблення морфологічно-орієнтованої нейронної архітектури для сегментації серцевої області на рентгенівських знімках грудної клітки, обґрунтування використання edge-aware та морфологічних механізмів для уточнення меж ROI, реалізація програмної моделі, підготовка сегментаційних даних, проведення експериментального оцінювання, аналіз якості сформованих масок і формулювання висновків щодо ефективності запропонованого підходу; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

ЗМІСТ

| | |
|------------------------------------------------------------------------------------------------------------------|-----|
| ВСТУП | 22 |
| РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ ЗА ТЕМАТИКОЮ РОБОТИ | 37 |
| 1.1 Особливості виділення ROI на CXR | 38 |
| 1.2 Класичні методи сегментації та локалізації області інтересу на медичних зображеннях | 41 |
| 1.3 Сучасні нейромережеві підходи до сегментації та локалізації області інтересу на медичних зображеннях..... | 45 |
| 1.4 Порівняльний аналіз підходів і обґрунтування напряму дослідження | 49 |
| 1.5 Висновки до розділу 1 | 53 |
| РОЗДІЛ 2. ПОСТАНОВКА ЗАДАЧІ ТА МЕТОДИКА ДОСЛІДЖЕННЯ..... | 55 |
| 2.1 Постановка задачі багатокласової сегментації серцевої області | 56 |
| 2.2 Підготовка датасету та структура вибірок | 67 |
| 2.3. Попередня обробка даних та формування навчальних прикладів..... | 72 |
| 2.4. Метрика оцінювання та протокол вимірювання якості | 77 |
| 2.5 Вимоги до методу та узагальнення методики..... | 84 |
| 2.6 Висновки до розділу 2 | 86 |
| РОЗДІЛ 3. РОЗРОБЛЕННЯ НЕЙРОННОЇ АРХІТЕКТУРИ CARDIOEDGEBIOMORPHNET ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВЗАЄМОДІЇ..... | 88 |
| 3.1. Загальна схема CardioEdgeBioMorphNet | 90 |
| 3.2. Формування edge-сигналу (Sobel/градієнт)..... | 98 |
| 3.3. Базові блоки мережі (Down/Up, нормалізація, активації)..... | 100 |
| 3.4. EdgeMorphBlock: структура, роль, рівняння | 103 |
| 3.5. Диференційовні морфологічні операції SoftMorphPool2D..... | 108 |
| 3.6. DMGGA: морфологічна увага в кількох масштабах | 110 |
| 3.7. ABR (Adaptive Boundary Rejection): прогноз порогу та правило відкидання у фон | 113 |
| 3.8. Функції втрат та стратегія навчання моделі..... | 118 |

| | |
|----------------------------------------------------------------------------------------------------------|------------|
| 3.9. Обчислювальна складність та особливості реалізації..... | 120 |
| 3.10 Висновок до розділу 3 | 124 |
| РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА АНАЛІЗ | |
| РЕЗУЛЬТАТІВ..... | 126 |
| 4.1 План експериментів і протокол оцінювання | 127 |
| 4.2 Бейзлайнові моделі та умови порівняння | 130 |
| 4.3 Кількісні результати експериментів..... | 133 |
| 4.4 Абляційне дослідження | 144 |
| 4.5 Якісний аналіз результатів сегментації | 155 |
| ВИСНОВКИ..... | 167 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 171 |
| ДОДАТОК А Список публікацій здобувача | 181 |
| ДОДАТОК Б Структура набору даних, кодування міток та контроль коректності розмітки..... | 187 |

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

IХС — ішемічна хвороба серця

ШІ — штучний інтелект

BG — Background (фон)

CE — Cross-Entropy (крос-ентропійна функція втрат)

BCL — Boundary Consistency Loss (функція втрат для узгодження меж)

EAAL — Edge-Aware Affinity Loss (функція втрат з урахуванням меж)

ABR — Adaptive Boundary Rejection (адаптивне відкидання невпевнених пікселів)

DMGGA — Differentiable Morphological Gated Global Attention

(диференційовний морфологічний механізм уваги)

Conv — Convolution (згортка)

SiLU — Sigmoid Linear Unit (сигмоїдна лінійна функція активації)

SEBlock — Squeeze-and-Excitation Block (механізм уваги каналів)

EdgeMorphBlock — морфологічно-орієнтований блок з урахуванням меж

ASPP — Atrous Spatial Pyramid Pooling (багаторівневе фільтрування з розширенням)

GAP — Global Average Pooling (глобальне середнє згортання)

Softmax — функція нормалізації ймовірностей для багатокласової класифікації

Sobel mag — Sobel magnitude (градієнтний детектор країв)

Encoder — блок кодування ознак (стиснення та вилучення фіч)

Decoder — блок декодування ознак (відновлення просторової роздільності)

Stem — вхідний блок мережі, що формує базові ознаки

Up/Bilinear upsample — операції апсемплінгу (масштабування карт ознак)

NN — Neural Network (нейронна мережа)

CNN — Convolutional Neural Network (згорткова нейронна мережа)

Vision Transformer (ViT) — архітектура трансформера для зображень

ResNet-50 — згорткова мережа із залишковими блоками (50 шарів)

EfficientNet — згорткова мережа з оптимізованою архітектурою для ефективного вилучення ознак

CardioEdgeBioMorphNet — запропонована морфологічно-орієнтована нейронна мережа для сегментації серця

ROI — Region of Interest (область інтересу)

CXR — Chest X-ray (рентгенівський знімок грудної клітки)

GT — Ground Truth (еталонна розмітка)

BG — Background (фон)

FG — Foreground (об'єкт/передній план, не фон)

FN — False Negative (хибнонегативні спрацьовування)

FP — False Positive (хибнопозитивні спрацьовування)

PA — Posteroanterior projection (постеро-антеріорна проєкція CXR)

AP — Anteroposterior projection (антеро-постеріорна проєкція CXR)

DL — Deep Learning (глибинне навчання)

ВСТУП

Актуальність теми. На сучасному етапі розвитку інформаційних технологій спостерігається активне впровадження методів штучного інтелекту та глибинного навчання у різні сфери, зокрема у медицину [3-7]. Використання нейронних мереж дає змогу автоматизувати аналіз медичних даних, підвищити точність діагностичних процедур і скоротити час прийняття клінічних рішень. Одним із ключових напрямів застосування таких технологій є діагностика серцево-судинних захворювань, зокрема ІХС, яка залишається однією з провідних причин смертності [35-39].

У контексті даної роботи штучний інтелект використовується як технологічна основа для автоматизованого аналізу CXR-зображень, виділення серцевої ROI та формування допоміжного результату, який може бути застосований у системах підтримки прийняття рішень під час первинної діагностики.

Раннє виявлення ІХС є критично важливим для своєчасного лікування та зменшення ризику ускладнень. Рентгенографія CXR належить до найбільш доступних методів первинного обстеження, однак інтерпретація рентгенівських знімків значною мірою залежить від кваліфікації лікаря та характеризується обмеженою відтворюваністю результатів. Для автоматизованого аналізу CXR одним із базових етапів є виділення серцевої області (ROI), яке ускладнюється слабким контрастом меж, проєкційними накладаннями анатомічних структур, шумами та артефактами [24, 25, 39].

Наявні автоматизовані підходи, зокрема на основі згорткових нейронних мереж типу U-Net та їх модифікацій, здатні забезпечувати прийнятну якість сегментації, проте часто мають обмежені можливості щодо точного відтворення контуру та морфологічної форми серця. Додатковою проблемою є наявність невизначених ділянок зображення, що призводить до помилкової класифікації фонових пікселів як анатомічних структур і знижує надійність побудови ROI-маски [7, 8, 14, 28, 48, 49, 80].

У зв'язку з цим актуальним є створення спеціалізованих методів глибинного навчання, орієнтованих на явне врахування морфологічних характеристик, контурної інформації (edge-aware підходів) та багатомасштабного аналізу ознак, що підвищує точність і відтворюваність сегментації серця на рентгенівських знімках грудної клітки [15–20, 67, 68, 74, 75, 85]. Практичну значущість таких методів підсилює розробка програмного забезпечення, яке забезпечує відтворюваний цикл підготовки даних, навчання й оцінювання моделі та візуалізацію результатів, що є необхідним для подальшої інтеграції у прикладні системи підтримки прийняття рішень.

Дана робота спрямована на розроблення та дослідження морфологічно-орієнтованої нейронної архітектури для автоматизованої сегментації серця на СХР. Основну увагу приділено підвищенню точності відтворення меж серця за рахунок поєднання глибинного навчання з морфологічними принципами аналізу зображень та урахування контурних особливостей досліджуваних структур. Запропонований підхід передбачає інтеграцію спеціалізованих морфологічних блоків і механізмів, чутливих до крайових характеристик, що дозволяє зменшити помилки сегментації у фонових зонах та підвищити узгодженість сегментаційних масок, забезпечуючи баланс між детальністю просторового відтворення та здатністю до узагальнення.

У межах роботи також реалізовано програмне забезпечення для повного циклу обробки медичних зображень, включаючи підготовку даних, навчання нейронної моделі, її валідацію, тестування та аналіз і візуалізацію результатів сегментації, що створює основу для подальшого розвитку інтелектуальних систем підтримки прийняття рішень у кардіології.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась відповідно до індивідуального плану підготовки аспіранта кафедри інженерії програмного забезпечення та інформаційних технологій Дніпровського національного університету імені Олеся Гончара.

Дослідження пов'язане з тематикою науково-дослідних робіт кафедри,

спрямованих на розроблення інформаційних технологій, програмного забезпечення, моделей та алгоритмів інтелектуальної обробки даних. Окремі результати дисертаційної роботи виконано у межах науково-дослідної роботи «Розроблення програмного забезпечення аналізу та кластеризації часових рядів» на 2022–2024 рр. (шифр ФПМ-2-22, номер державної реєстрації 0122U001465), а також науково-дослідної роботи «Розроблення інформаційної технології обробки статистичних даних» на 2025–2027 рр. (шифр ФПМ-2-25, номер державної реєстрації 0125U002280).

Зв'язок дисертаційного дослідження із зазначеними науково-дослідними роботами полягає у розробленні моделей, алгоритмів і програмного забезпечення для інтелектуальної обробки медичних даних, зокрема рентгенівських зображень грудної клітки, автоматизованого виділення ділянок інтересу, аналізу діагностично значущих ознак та формування допоміжних результатів для підтримки прийняття рішень під час медичної діагностики.

Метою і завданням дослідження є розроблення, програмна реалізація та експериментальне обґрунтування морфологічно-орієнтованої нейронної архітектури CardioEdgeBioMorphNet для автоматизованої сегментації серцевої області на рентгенівських знімках грудної клітки, спрямованої на підвищення точності відтворення анатомічних меж, зменшення кількості помилкових сегментацій у фонових областях та формування результатів, придатних для використання в системах підтримки прийняття рішень під час аналізу медичних даних.

Досягнення поставленої мети передбачає поєднання морфологічних механізмів обробки ознак, edge-aware підходів, багатомасштабного аналізу та адаптивного калібрування впевненості прогнозів у межах єдиної нейронної моделі. Окремою складовою роботи є створення програмного забезпечення, яке забезпечує підготовку даних, навчання, валідацію, тестування моделі, візуалізацію результатів сегментації та можливість її практичного використання в задачах автоматизованого аналізу рентгенівських знімків

грудної клітки.

Основні завдання дослідження:

1. Проаналізувати сучасні методи автоматизованої обробки медичних зображень, зокрема підходи до сегментації серцевої області на рентгенівських знімках грудної клітки, та визначити їхні переваги й обмеження в контексті підтримки прийняття рішень під час діагностики ішемічної хвороби серця.

2. Сформувати та підготувати навчальну базу даних рентгенівських знімків грудної клітки, отриманих із клінічної практики, виконати розмітку області інтересу серця з використанням спеціалізованих інструментів анотування, зокрема CVAT, та забезпечити узгодженість зображень і відповідних масок.

3. Розробити методику попередньої обробки вхідних CXR-зображень, що включає приведення даних до єдиного формату, нормалізацію інтенсивностей, підготовку сегментаційних масок та формування навчальних, валідаційних і тестових вибірок.

4. Розробити морфологічно-орієнтовану нейронну архітектуру CardioEdgeBioMorphNet, яка поєднує багатомасштабне виділення ознак, механізми уточнення меж, морфологічні компоненти та адаптивне відкидання невпевнених пікселів для підвищення якості сегментації серцевої області.

5. Реалізувати програмні засоби для навчання, валідації та тестування запропонованої моделі, виконати налаштування параметрів нейронної мережі та оцінити її продуктивність на контрольних даних.

6. Провести експериментальне порівняння розробленої моделі з базовими архітектурами сегментації медичних зображень, оцінити точність, стійкість і характер типових помилок за кількісними метриками та результатами якісного аналізу.

7. Дослідити вплив окремих компонентів CardioEdgeBioMorphNet на якість сегментації шляхом абляційного аналізу, зокрема оцінити роль edge-aware механізмів, морфологічно-орієнтованих блоків та адаптивного порогового відкидання.

8. Розробити прикладний користувацький інтерфейс для взаємодії з моделлю, який забезпечує завантаження CXR-знімка, виконання сегментації, формування ROI-маски, візуалізацію результатів і відображення ключових показників для інтерпретації отриманого результату.

9. Реалізувати можливість практичної апробації розробленої моделі та програмного забезпечення у вигляді відкритого демонстраційного застосунку, зокрема на платформі Hugging Face, для тестування підходу сторонніми користувачами та науковою спільнотою.

Результатом виконання поставлених завдань має стати програмно реалізована система автоматизованої сегментації серцевої області на рентгенівських знімках грудної клітки, яка забезпечує відтворюване виділення ROI, зменшує вплив ручного аналізу та може бути використана як складова систем підтримки прийняття рішень під час первинного аналізу медичних зображень у кардіологічній практиці.

Дисертаційне дослідження виконане в межах наукових робіт, пов'язаних із розробленням та впровадженням інформаційних технологій у медичну сферу. Робота безпосередньо пов'язана з науковим напрямом кафедри інженерії програмного забезпечення та інформаційних технологій Дніпровського національного університету імені Олеся Гончара, а саме з розробленням моделей і програмного забезпечення підтримки прийняття рішень під час діагностики за медичними даними.

Тематика дослідження відповідає сучасним напрямкам розвитку інформаційних технологій у медицині, зокрема використанню штучного інтелекту, глибокого навчання та програмних систем автоматизованого аналізу медичних зображень для підвищення якості, відтворюваності та оперативності діагностичних процесів.

Об'єкт дослідження – процес автоматизованого аналізу CXR, який включає сегментацію ROI, оцінювання морфологічних характеристик та подання результатів у вигляді, придатному для підтримки прийняття рішень під час первинної діагностики серцево-судинних захворювань.

Предмет дослідження – методи, моделі та алгоритми DL для автоматизованої сегментації ROI серця на CXR, зокрема морфологічно-орієнтовані NN з edge-aware механізмами. Предмет дослідження охоплює розроблення архітектури CardioEdgeBioMorphNet, спеціалізованих блоків EdgeMorphBlock, механізму уваги DMGGA, явного Sobel mag та edge-fusion, а також механізму ABR для адаптивного відкидання невпевнених пікселів у BG. Крім того, предмет включає підготовку та аугментацію даних, а також алгоритми навчання, валідації й тестування моделі та оцінювання якості сегментації за метриками Dice, IoU і Acc_heart.

Методи дослідження: у роботі використано методи аналізу та систематизації наукових джерел; методи обробки медичних зображень і цифрової обробки сигналів (передобробка, нормалізація, формування ROI, аналіз контурів на основі Sobel mag); методи глибинного навчання та комп'ютерного зору (CNN, encoder–decoder архітектури, механізми уваги, багатомасштабний контекстний аналіз ASPP, морфологічно-орієнтовані компоненти EdgeMorphBlock, DMGGA, механізм ABR); експериментальні методи навчання, валідації та тестування моделей із відтворюваним протоколом; методи кількісного оцінювання якості сегментації за метриками Dice, IoU, Acc_heart, а також якісний аналіз результатів шляхом візуалізації (overlay) та аналізу помилок FP/FN.

Вхідними даними були CXR з проєкційним зображенням серця та прилеглих анатомічних структур. Дані отримано з Дніпропетровського обласного клінічного центру кардіології та кардіохірургії для проведення дослідження. Знімки попередньо розмічено експертним способом із використанням інструменту CVAT, що дозволило сформувати GT-маски сегментації серцевої ROI для навчання, валідації та тестування моделі. Перед подачею до нейронної мережі зображення проходили стандартизацію та підготовку (масштабування, нормалізація інтенсивностей, приведення до уніфікованого формату), що зменшує вплив варіативності умов зйомки, шуму та артефактів, характерних для рентгенографії.

Достовірність отриманих результатів забезпечується коректною постановкою наукової задачі, використанням теоретично обґрунтованих методів глибинного навчання, комп'ютерного зору та цифрової обробки медичних зображень, а також застосуванням апробованих підходів до навчання й оцінювання нейронних мереж для задач семантичної сегментації.

Наукова новизна полягає в наступному:

1. **Вперше** розроблено морфологічно-орієнтовану нейронну архітектуру CardioEdgeBioMorphNet для сегментації серцевої області (ROI) на рентгенівських знімках грудної клітки, що забезпечує багатомасштабне представлення ознак та інтеграцію контурної інформації для точнішого відтворення меж і форми серця.

2. **Удосконалено** базовий блок MorphBlock, за рахунок edge-aware, який поєднує локальні та багатомасштабні згорткові операції, механізми channel attention і морфологічне gating для формування стійких ознак та підвищення точності локалізації меж серця.

3. **Вперше** запропоновано диференційовний морфологічний механізм уваги DMGGA з multi-kernel сумішшю для підсилення меж і текстур у глибоких ознаках сегментації та покращення узгодженості контурів у складних і слабконтрастних ділянках CXR.

4. **Вперше** використано явний edge-потік (Sobel magnitude) із подальшим edge-fusion у декодері (concat + conv) як структурний індуктивний bias для точнішої локалізації меж серця та зменшення похибок у прикордонних зонах.

5. **Вперше** реалізовано механізм Adaptive Boundary Rejection (ABR) у вигляді окремої порогової гілки, що формує адаптивний поріг упевненості для сегментаційного прогнозу. Завдяки цьому пікселі з недостатньою впевненістю класифікуються як фон, що зменшує кількість хибнопозитивних включень у фонових областях.

6. **Вперше** обґрунтовано та експериментально підтверджено перевагу комплексного поєднання морфологічно-орієнтованих блоків, edge-aware

механізмів, багатомасштабного контекстного узгодження та адаптивного порогового відкидання для підвищення точності й стабільності сегментації серця на рентгенівських знімках грудної клітки.

7. Вперше поєднано результат сегментації з автоматизованим формуванням висновку щодо класу стану на основі агрегування ймовірностей у межах ROI та показника confidence margin top1–top2 в єдиному прикладному інтерфейсі для практичного використання.

Практичне значення результатів полягає в тому, що розроблена морфологічно-орієнтована нейронна модель CardioEdgeBioMorphNet та відповідне програмне забезпечення можуть бути використані як складова інтелектуальних систем автоматизованого аналізу рентгенівських знімків грудної клітки. Запропонований підхід забезпечує розв'язання двох взаємопов'язаних прикладних задач: сегментації серцевої області та класифікаційного визначення ймовірного клінічного стану серця. Це дозволяє не лише виділяти область інтересу на CXR-зображенні, а й формувати допоміжний результат для підтримки прийняття рішень під час первинного аналізу медичних даних.

Практичне значення роботи визначається такими положеннями:

1. Розроблено програмне забезпечення для автоматизованого аналізу CXR-зображень, яке поєднує задачі сегментації та класифікації. Запропонована реалізація забезпечує повний цикл роботи з медичними зображеннями: завантаження рентгенівського знімка, попередню обробку, подання даних на вхід нейронної моделі, формування сегментаційної маски, визначення ймовірного клінічного класу, візуалізацію результату та відображення додаткових показників, необхідних для його інтерпретації.

2. Забезпечено автоматизоване виділення серцевої області на рентгенівських знімках грудної клітки. Сегментаційна складова моделі дозволяє формувати ROI-маску без ручного окреслення меж, що скорочує час первинної обробки CXR-зображень, зменшує залежність результату від суб'єктивного досвіду користувача та підвищує відтворюваність аналізу у

випадках слабо виражених меж між серцем, легневими полями, ребрами, діафрагмою та фоновими ділянками.

3. Реалізовано класифікаційне визначення ймовірного клінічного стану серця на основі результатів аналізу зображення. Класифікаційна складова програмного засобу дозволяє отримувати узагальнений висновок щодо належності CXR-знімка до одного з визначених клінічних класів, що створює основу для подальшого використання моделі як допоміжного інструменту в системах підтримки прийняття рішень.

4. Підвищено якість відтворення меж серцевої області за рахунок поєднання морфологічних та edge-aware механізмів. Використання спеціалізованих блоків EdgeMorphBlock, контурного потоку на основі Sobel-ознаки, механізмів морфологічної уваги та багатомасштабного аналізу дозволяє точніше локалізувати прикордонні ділянки, зберігати просторову структуру серцевої ROI та формувати більш узгоджену сегментаційну маску порівняно з базовими encoder–decoder підходами.

5. Зменшено кількість хибнопозитивних включень фонових областей у результат сегментації. Запропонований механізм Adaptive Boundary Rejection забезпечує віднесення пікселів із недостатньою впевненістю до фонового класу. Це є важливим для зменшення помилкових включень у ділянках легень, ребер, діафрагми, периферійних зон зображення та інших неоднозначних областях CXR, які можуть ускладнювати подальший аналіз.

6. Створено основу для стандартизації первинного аналізу рентгенівських знімків грудної клітки. Автоматизоване формування сегментаційної маски, визначення ймовірного клінічного класу, застосування єдиного протоколу передобробки, фіксованих метрик оцінювання та відтворюваної процедури тестування дозволяють отримувати порівнювані результати для різних зображень і зменшують варіативність, пов'язану з ручним аналізом.

7. Реалізовано прикладний інтерфейс для візуального аналізу результатів роботи моделі. Програмне забезпечення дає змогу відображати

сегментаційну маску, накладати її на початкове CXR-зображення, аналізувати контури серцевої області, переглядати визначений клінічний клас і оцінювати рівень упевненості моделі. Це полегшує перевірку результатів, виявлення типових помилок сегментації та класифікації, а також подальше використання моделі як допоміжного інструменту.

8. Забезпечено можливість використання результатів сегментації для уточнення класифікаційного висновку. Отримана ROI-маска може бути використана для аналізу ймовірностей саме в межах серцевої області, визначення домінуючого клінічного класу та розрахунку показника впевненості. Такий підхід дозволяє пов'язати просторово інтерпретований результат сегментації з класифікаційним висновком і підвищити практичну цінність автоматизованого аналізу.

9. Результати дослідження можуть бути використані в клінічній, науково-дослідній та навчальній діяльності. У клінічному контексті розроблений підхід може застосовуватися як допоміжний засіб попереднього аналізу CXR-знімків; у наукових дослідженнях — як основа для подальшого розвитку моделей сегментації та класифікації медичних зображень; у навчальному процесі — під час вивчення комп'ютерного зору, глибинного навчання, медичної інформатики та програмної інженерії інтелектуальних систем.

Таким чином, практичне значення роботи полягає у створенні прикладного програмного інструментарію, який забезпечує автоматизоване, відтворюване та більш стійке виділення серцевої області на CXR-зображеннях, а також формування допоміжного класифікаційного висновку щодо ймовірного клінічного стану серця. Запропонований підхід може бути використаний як основа для подальшої інтеграції методів глибинного навчання в системи підтримки прийняття рішень під час аналізу медичних даних.

Особистий внесок здобувача. Дисертаційна робота є самостійною науковою працею здобувача, у якій висвітлено власні ідеї, методичні підходи, архітектурні рішення та програмні розробки автора, спрямовані на розв'язання

поставлених у дослідженні завдань. Основні наукові положення, висновки, результати експериментальних досліджень і рекомендації, що містяться в дисертаційній роботі та виносяться на захист, отримано здобувачем особисто. Ідеї, положення та результати інших авторів, використані в дисертації, мають відповідні посилання і залучені лише для теоретичного обґрунтування, порівняльного аналізу та зіставлення з власними результатами.

За темою дисертації опубліковано 12 наукових праць, з них 5 статей у наукових фахових виданнях України та 7 праць апробаційного характеру. Перелік публікацій здобувача за темою дисертації наведено в додатку А. Із праць, виконаних у співавторстві, на захист виносяться лише ті положення та результати, які отримано здобувачем особисто.

У наукових працях, опублікованих у співавторстві, здобувачу належать такі результати:

– у роботі [1, список публікацій здобувача, додаток А] здобувачем виконано аналіз предметної області та обґрунтовано застосування згорткових нейронних мереж для діагностування ішемічної хвороби серця за медичними зображеннями; здійснено підготовку набору рентгенівських знімків грудної клітки, попередню обробку даних, реалізацію та навчання CNN-моделі, проведення обчислювальних експериментів, аналіз результатів і формулювання висновків щодо переваги використання згорткових нейронних мереж;

– у роботі [2] здобувачем формалізовано задачу обробки медичних даних для діагностування ішемічної хвороби серця, проаналізовано можливості застосування математичних моделей у задачах медичної діагностики, підготовлено вхідні дані, виконано експериментальну частину дослідження, проаналізовано результати застосування моделей і сформульовано відповідні висновки;

– у роботі [3] здобувачем обґрунтовано доцільність використання архітектури EfficientNet для аналізу рентгенівських знімків грудної клітки, виконано підготовку та попередню обробку набору даних, адаптовано модель

EfficientNet до задачі виявлення ішемічної хвороби серця, реалізовано програмний код навчання й тестування моделі, проведено експерименти, побудовано й проаналізовано метрики якості класифікації;

– у роботі [4] здобувачем проаналізовано можливості застосування трансформерних архітектур до задачі класифікації рентгенографічних зображень, підготовлено та структуровано набір CXR-знімків, адаптовано transformer-based модель до задачі класифікації ішемічної хвороби серця, реалізовано програмне забезпечення для навчання та тестування моделі, проведено експериментальні дослідження й виконано аналіз точності класифікації;

– у роботі [5] здобувачем розроблено концепцію та структуру архітектури BioMorphNet для автоматизованої бінарної класифікації патологій серця за рентгенівськими знімками, обґрунтовано використання морфологічно-орієнтованих компонентів у нейронній моделі, реалізовано програмний код моделі, підготовлено та попередньо оброблено CXR-дані, проведено навчання й тестування моделі, оцінено якість класифікації та проаналізовано експериментальні результати;

– у роботі [6] здобувачем виконано огляд методів підтримки прийняття рішень під час діагностики за медичними даними, систематизовано підходи до обробки медичної інформації, узагальнено можливості застосування математичних моделей і методів машинного навчання в медичній діагностиці, підготовлено матеріали тез;

– у роботі [7] здобувачем сформовано набір рентгенівських зображень грудної клітки для задачі діагностики ішемічної хвороби серця, виконано структурування даних за класами, попередню обробку зображень, підготовку даних до використання у нейромережових моделях, а також проаналізовано вимоги до якості та придатності датасету для подальших експериментів;

– у роботі [8] здобувачем проаналізовано можливості використання трансформерних моделей для автоматизованої діагностики ішемічної хвороби серця на основі рентгенівських зображень, підготовлено набір CXR-знімків,

адаптовано архітектуру трансформера до відповідної задачі, реалізовано програмний код, проведено навчання та тестування моделі, проаналізовано отримані результати;

– у роботі [9] здобувачем обґрунтовано використання архітектури EfficientNet для класифікації рентгенівських зображень грудної клітки, виконано підготовку та попередню обробку даних, застосовано методи аугментації, реалізовано програмний код навчання нейронної мережі, проведено експерименти, проаналізовано стабільність навчання та оцінено якість класифікації;

– у роботі [10] здобувачем розроблено концепцію архітектури BioMorphNet для автоматизованого виявлення ішемічної хвороби серця за рентгенівськими знімками, обґрунтовано використання морфологічно-орієнтованих компонентів, реалізовано програмну модель, підготовлено CXR-дані, проведено навчання та тестування, оцінено якість роботи моделі й проаналізовано отримані результати;

– у роботі [11] здобувачем обґрунтовано доцільність поєднання згорткових нейронних мереж, механізмів уваги та морфологічного аналізу в задачі виявлення ішемічної хвороби серця за рентгенівськими знімками, розроблено та описано відповідний підхід, підготовлено рентгенівські зображення до експериментального аналізу, реалізовано програмні компоненти, проаналізовано вплив морфологічних і attention-механізмів на якість розпізнавання;

– у роботі [12] здобувачем розроблено морфологічно-орієнтовану нейронну архітектуру для сегментації серцевої області на рентгенівських знімках грудної клітки, обґрунтовано використання edge-aware та морфологічних механізмів для уточнення меж ROI, реалізовано програмну модель, підготовлено сегментаційні дані, проведено експериментальне оцінювання, проаналізовано якість сформованих масок і сформульовано висновки щодо переваги запропонованого підходу.

У дисертаційній роботі здобувачем особисто виконано постановку

задачі автоматизованої сегментації серцевої області на CXR-знімках у контексті підтримки прийняття рішень; розроблено та реалізовано морфологічно-орієнтовану нейронну архітектуру CardioEdgeBioMorphNet; спроектовано спеціалізований блок EdgeMorphBlock для багатомасштабного вилучення ознак і точнішого відтворення меж ROI; реалізовано edge-aware механізми з використанням Sobel-ознаки та подальшим edge-fusion; застосовано диференційовні морфологічні механізми для підсилення контурних і текстурних ознак; реалізовано механізм Adaptive Boundary Rejection для адаптивного відкидання невпевнених пікселів під час інференсу.

Також здобувачем особисто розроблено методику підготовки даних, що включає попередню обробку, нормалізацію, аугментацію, формування навчальної, валідаційної та тестової вибірок; організовано розмітку даних для задачі сегментації; проведено експерименти з навчання, валідації та тестування моделей; виконано порівняння із базовими архітектурами; здійснено аналіз внеску окремих компонентів запропонованої архітектури; проаналізовано результати за метриками Dice, IoU, Acc_heart та візуальними прикладами накладання масок на CXR-зображення, а також виконано аналіз типових FP/FN-помилки.

Апробація матеріалів дисертації. Основні положення та результати дисертаційної роботи доповідалися й обговорювалися на наукових і науково-практичних конференціях, що відображено у працях апробаційного характеру.

Зокрема, результати дослідження були представлені на міжнародній науково-практичній конференції «Наука, освіта та суспільство в XXI столітті: наукові ідеї та механізми реалізації», м. Полтава, 14 грудня 2021 р., міжнародних науково-практичних конференціях «Математичне та програмне забезпечення інтелектуальних систем» (МПЗІС-2023 - м. Дніпро, 22-24 листопада 2023 р., МПЗІС-2024 - м. Дніпро, 20-22 листопада 2024 р., МПЗІС-2025 - м. Дніпро, 19-21 листопада 2025 р.), міжнародній конференції з автоматичного керування «Автоматика 2024», м. Дніпро, 20-22 листопада 2024 р., V міжнародній науковій конференції «Глобальні виклики та інновації:

шляхи розвитку сучасної науки», м. Одеса, 16 січня 2026 р., ІХ міжнародній науковій конференції «Міжгалузеві диспути: динаміка та розвиток сучасних наукових досліджень», м. Харків, 30 січня 2026 р. та підсумкових наукових конференціях Дніпровського національного університету імені Олеся Гончара (м. Дніпро, 2022, 2023, 2024, 2026).

У межах апробації було представлено результати, пов'язані з аналізом методів підтримки прийняття рішень під час діагностики за медичними даними, створенням набору рентгенівських зображень для діагностики ішемічної хвороби серця, застосуванням трансформерних моделей і EfficientNet для автоматизованого аналізу CXR-знімків, а також розробленням морфологічно-орієнтованих нейронних рішень BioMorphNet і CardioEdgeBioMorphNet. Опубліковані праці та доповіді підтвердили практичну значущість і наукову новизну отриманих результатів, зокрема щодо підготовки медичних даних, класифікації рентгенівських знімків грудної клітки, сегментації серцевої області та використання морфологічних і edge-aware механізмів у задачах підтримки прийняття рішень під час діагностики ішемічної хвороби серця.

Публікації. Основні положення та результати дисертаційної роботи апробовано шляхом публікації наукових праць і доповідей на науково-практичних конференціях. За темою дисертації опубліковано 12 наукових праць, з них 5 статей у фахових наукових виданнях України категорії Б [1–5 зі списку публікацій здобувача, додаток А] та 7 праць апробаційного характеру у вигляді тез доповідей і матеріалів міжнародних наукових конференцій [6–12]. Перелік публікацій за темою дисертації наведено у додатку А.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків та списку використаних джерел. Повний обсяг дисертації складає 190 сторінок, у тому числі 149 сторінок основного тексту, 10 таблиць і 21 рисунок, список використаних джерел — 93 найменування, 2 додатки.

РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ ЗА ТЕМАТИКОЮ РОБОТИ

В даному розділі наведено огляд і аналіз наукових публікацій та сучасних досліджень, присвячених методам обробки і сегментації медичних зображень, зокрема рентгенівських знімків грудної клітки. Розглянуто особливості виділення серцевої області інтересу на CXR-зображеннях, класичні підходи до аналізу зображень, а також сучасні методи на основі згорткових нейронних мереж, механізмів уваги, багатомасштабних контекстних модулів і гібридних CNN–Transformer архітектур. Особливу увагу приділено перевагам і обмеженням існуючих рішень у контексті автоматизованої локалізації серцевої ROI, що дає змогу обґрунтувати доцільність розроблення спеціалізованого методу сегментації області інтересу для подальшого аналізу ознак ішемічної хвороби серця.

Рентгенографія органів грудної клітки є одним із найбільш поширених, доступних і швидких методів первинної візуалізації в клінічній практиці. Водночас використання рентгенівських знімків грудної клітки для автоматизованого аналізу стану серця пов'язане з низкою складнощів, зумовлених проєкційною природою зображення, накладанням анатомічних структур, неоднорідністю контрасту, наявністю шумів, артефактів і варіативністю умов отримання даних. За таких умов задача виділення серцевої області інтересу на CXR-зображеннях є важливим, але складним етапом побудови систем комп'ютерної підтримки діагностики серцево-судинних захворювань.

У сучасних дослідженнях для розв'язання задач сегментації та локалізації області інтересу застосовуються як класичні методи обробки зображень, так і глибинні нейромережеві підходи. Класичні алгоритми забезпечують інтерпретованість і простоту реалізації, проте часто виявляються недостатньо стійкими до слабкої вираженості меж, шуму та морфологічної варіативності анатомічних структур. Натомість сучасні

нейромережеві архітектури демонструють вищу точність, однак також не повністю усувають проблему нечітких меж, локальної неоднозначності та включення сторонніх структур до області інтересу.

У зв'язку з цим актуальним є аналіз існуючих підходів до сегментації медичних зображень з метою виявлення їхніх переваг, недоліків і невирішених проблем у контексті виділення серцевої ROI на рентгенівських знімках грудної клітки. Це створює підґрунтя для обґрунтування напряду дослідження, пов'язаного з розробленням методу, орієнтованого на стабільну локалізацію області інтересу, зменшення впливу фону та підвищення якості подальшого аналізу ознак ІХС.

1.1 Особливості виділення ROI на CXR

Виділення серцевої області інтересу на рентгенівських знімках грудної клітки є важливим етапом автоматизованого аналізу CXR-зображень у задачах комп'ютерної підтримки діагностики серцево-судинних захворювань, зокрема ішемічної хвороби серця. Локалізація ROI дозволяє обмежити подальший аналіз анатомічно значущою зоною, зменшити вплив другорядних структур, підвищити інформативність вхідних даних і створити передумови для побудови більш стійких моделей класифікації та інтерпретації. Водночас ця задача не є тривіальною, оскільки рентгенівський знімок грудної клітки є двовимірною проєкцією тривимірних анатомічних структур, у межах якої різні тканини та органи накладаються одна на одну, що ускладнює стабільне виділення серцевої області.

Однією з основних проблем є слабка вираженість межі між серцевою тінню та прилеглими анатомічними структурами. На рентгенограмі межа серця не завжди має чіткий і стабільний градієнт яскравості, оскільки її видимість залежить від положення пацієнта, якості експозиції, рівня шуму, наявності супутніх змін у легенях, а також від апаратних і протокольних особливостей отримання знімка. У результаті одна й та сама серцева область у різних дослідженнях може візуально сприйматися по-різному: від достатньо

чітко локалізованої до розмитої або частково невизначеної. Це створює труднощі як для класичних методів сегментації, чутливих до локальних перепадів інтенсивності, так і для нейромережових моделей, які можуть включати до ROI сторонні структури або, навпаки, втратити частину серцевої області [24, 25, 34].

Додатковим ускладнювальним чинником є наявність шумів, артефактів і доменного зсуву між знімками. На якість CXR-зображень впливають відмінності між рентген-апаратами, параметрами знімання, методами постобробки, цифровою компресією, рівнем контрасту та яскравості. Крім того, у практичних даних можуть бути присутні сторонні об'єкти, написи, технічні маркери, елементи укладки пацієнта та часткові обмеження поля огляду. Усе це призводить до варіативності візуального вигляду серцевої області та знижує узагальнювальну здатність моделей, якщо під час навчання не враховано достатню різноманітність даних і відповідні механізми регуляризації [2, 27, 35–39].

Суттєве значення для виділення серцевої ROI має також геометрична варіативність серцевого силуету. Форма, розмір і положення серцевої області на рентгенівському знімку змінюються залежно від індивідуальних анатомічних особливостей пацієнта, фази дихання, положення тіла, а також проєкції знімання. Особливо важливо враховувати різницю між РА- та АР-проєкціями, оскільки в АР-проєкції серце може візуально виглядати збільшеним через проєкційні особливості. Для алгоритмів це означає необхідність бути стійкими не лише до локальних змін текстури, а й до глобальних змін масштабу, форми та просторового розташування області інтересу [32, 33, 42, 87].

На складність виділення серцевої області інтересу на CXR-зображеннях одночасно впливають проєкційність рентгенівського зображення, перекриття анатомічних структур, низький локальний контраст, а також шум, артефакти й доменний зсув. Сукупна дія цих чинників призводить до локальної невизначеності на межі ROI та підвищує ризик виникнення хибнопозитивних

і хибнонегативних помилок.

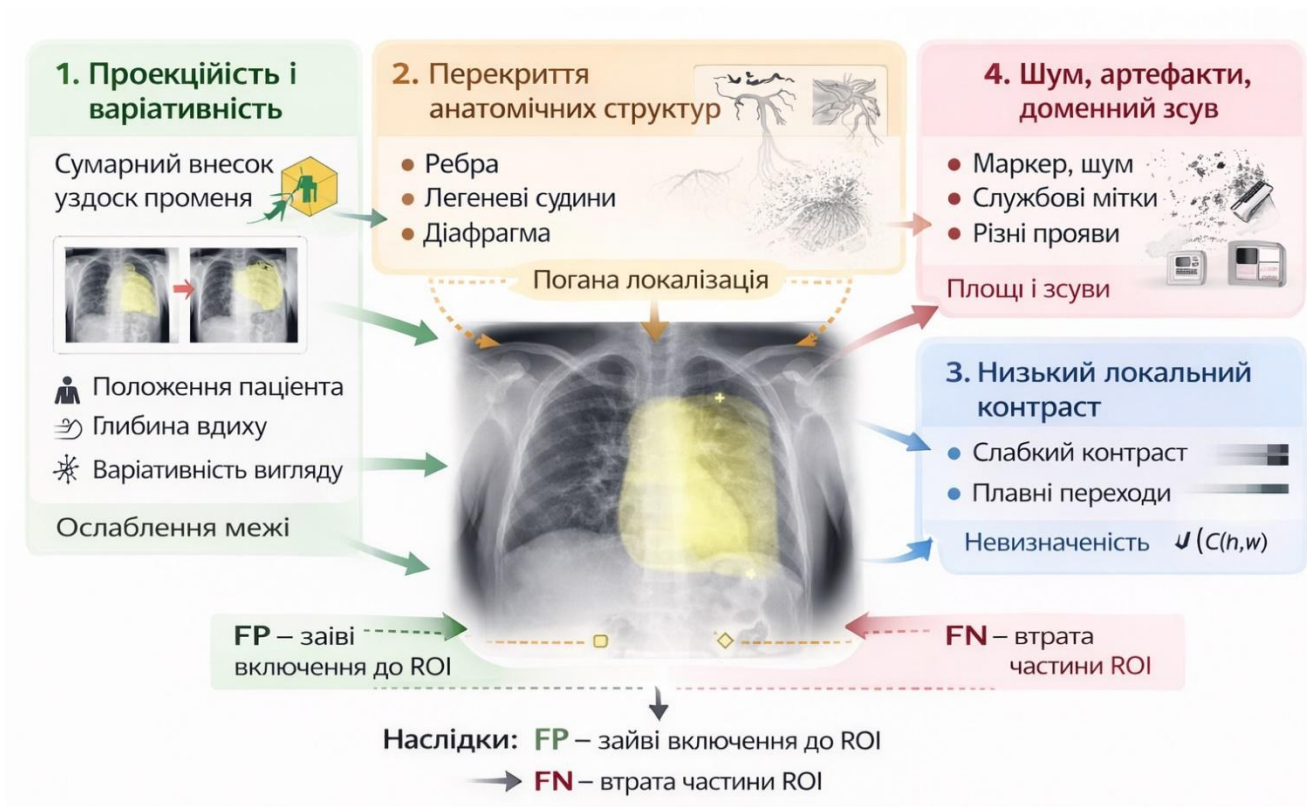


Рисунок. 1.1 – Основні чинники, що ускладнюють виділення серцевої ROI на CXR-зображеннях

Як показано на рис. 1.1, основні труднощі під час виділення серцевої ROI пов'язані не лише зі слабкою вираженістю межі, а й із накладанням ребер, легневих судин і діафрагми, варіативністю положення пацієнта, змінами контрасту, наявністю шуму, артефактів і доменного зсуву. Усе це призводить до появи FP як зайвих включень до області інтересу та FN як втрати частини ROI, що обґрунтовує доцільність оцінювання якості за метриками перекриття, зокрема Dice та IoU.

Окрему складність становить підготовка даних і побудова коректної розмітки. Якість виділення ROI безпосередньо залежить від узгодженості пар зображення–маска, правильності кодування класів, стабільності правила формування області інтересу та однакового трактування цільової зони в усьому наборі даних. У задачі локалізації серцевої області навіть незначні відхилення в розмітці можуть суттєво впливати на значення метрик

перекриття, оскільки похибки накопичуються саме в прикордонних ділянках між ROI та фоном [59–61, 81, 82].

Для задачі виділення серцевої ROI на CXR-зображеннях характерне поєднання двох вимог: з одного боку, модель повинна враховувати глобальний контекст грудної клітки, щоб правильно локалізувати серцеву область; з іншого — вона має бути стійкою до слабких меж, перекриття структур і неоднорідності контрасту. Саме тому стандартні підходи, орієнтовані лише на локальні ознаки або лише на глобальне семантичне представлення, не завжди забезпечують достатню стабільність результату. У таких умовах особливого значення набувають методи, здатні поєднувати контекстне представлення, чутливість до прикордонних зон і робастність до спотворень.

Отже, виділення серцевої області інтересу на рентгенівських знімках грудної клітки є складною задачею через проєкційну природу CXR-зображень, накладання анатомічних структур, слабку вираженість меж, шум, артефакти та геометричну варіативність. Наведені особливості зумовлюють необхідність застосування спеціалізованих методів сегментації, які забезпечують стійку локалізацію ROI та придатність отриманої області інтересу до подальшого аналізу ознак ІХС.

1.2 Класичні методи сегментації та локалізації області інтересу на медичних зображеннях

Класичні методи сегментації медичних зображень відіграли важливу роль у становленні автоматизованого аналізу візуальних даних і заклали основу для подальшого розвитку сучасних підходів. До цієї групи належать порогова сегментація, контурні методи, регіональні підходи, кластеризація, а також морфологічна обробка. Їхніми основними перевагами є відносна простота реалізації, інтерпретованість, невисока обчислювальна складність та можливість чіткого формального опису процедур обробки зображення. Разом із тим у задачах медичної сегментації, зокрема при виділенні серцевої області інтересу на рентгенівських знімках грудної клітки, можливості таких методів

є обмеженими через слабкий контраст меж, накладання анатомічних структур, наявність шуму та значну варіативність форми об'єкта [26].

Порогова сегментація є одним із найпростіших і найдавніших підходів до виділення області інтересу. Її суть полягає у виборі такого порогового значення інтенсивності, за яким пікселі відносяться або до об'єкта, або до фону. У випадках, коли анатомічна структура достатньо добре відділяється за яскравістю від навколишніх тканин, цей підхід може забезпечувати прийнятний результат. Проте для рентгенівських знімків грудної клітки така умова зазвичай не виконується. Серцевий силует формується в умовах накладання легневих полів, середостіння, діафрагми та кісткових структур, тому інтенсивнісна межа між серцем і фоном часто є неоднорідною або частково втраченою. Через це порогові методи виявляються нестійкими до зміни параметра порога, а якість сегментації різко погіршується навіть за незначних варіацій контрасту чи шуму.

Контурні методи орієнтовані на виявлення межі об'єкта за локальними змінами інтенсивності або градієнта. До цієї групи належать градієнтні оператори, зокрема Sobel, Canny, Laplacian, а також активні контури і деформівні моделі. Їх перевага полягає в тому, що вони явно намагаються локалізувати границю між об'єктом і фоном. Для медичних зображень це є принципово важливим, оскільки клінічно релевантною часто є саме форма та контур анатомічної структури. Однак у задачі сегментації серця на CXR такі підходи мають суттєві обмеження. Якщо межа серця є слабкою, переривчастою або перекритою іншими структурами, градієнтний сигнал стає нестабільним, а контурний алгоритм може або не замкнутись навколо потрібної області, або зміститись на сторонні межі. Активні контури, крім того, залежать від початкової ініціалізації та параметрів деформації, що ускладнює їх використання в повністю автоматизованих системах [2, 27].

Окрему групу становлять регіональні підходи, засновані на об'єднанні пікселів за критеріями локальної однорідності. До них належать методи росту регіону, вододільний алгоритм, а також підходи, що базуються на сегментації

за статистичними властивостями сусідніх областей. Такі методи працюють не лише з окремими контурами, а й з внутрішньою структурою області, що теоретично робить їх придатними для виділення цілісного об'єкта. Проте на рентгенівських знімках грудної клітки умова локальної однорідності для серця часто не дотримується. Інтенсивність усередині серцевої області може суттєво змінюватися, а навколишній фон інколи має подібні характеристики. У результаті регіональні алгоритми можуть або розщеплювати серце на декілька фрагментів, або, навпаки, об'єднувати його з прилеглими ділянками фону.

До класичних підходів також належить кластеризація, зокрема методи на кшталт k-means або fuzzy c-means, які групують пікселі за подібністю ознак. Перевагою таких методів є відсутність потреби в жорстко заданій розмітці на етапі виділення кластерів, однак у практичних медичних задачах цього недостатньо для надійного отримання анатомічно коректної маски. Для СХР характерна висока неоднорідність текстур і слабка відокремленість серцевої області від фону, тому кластери не завжди відповідають реальним анатомічним межам. З цієї причини кластеризація частіше використовується як допоміжний етап аналізу або попередньої грубої локалізації, ніж як самодостатній інструмент точної сегментації [15–19].

Морфологічна обробка займає особливе місце серед класичних методів. Операції ерозії, дилатації, відкриття, закриття, заповнення отворів і виділення зв'язаних компонентів широко застосовуються для уточнення бінарних масок, пригнічення дрібного шуму та згладжування форми області інтересу. У задачах сегментації серця морфологічні операції можуть бути корисними для післяобробки попередньо виділеної області, наприклад для усунення дрібних хибнопозитивних включень або закриття локальних дефектів контуру. Проте вони не розв'язують саму задачу локалізації серця на складному рентгенівському зображенні, оскільки потребують початкової маски, якість якої вже має бути достатньо високою. Тобто морфологічні перетворення ефективні як допоміжний механізм уточнення, але не як повноцінна основа стійкої сегментації [20, 67, 68, 85].

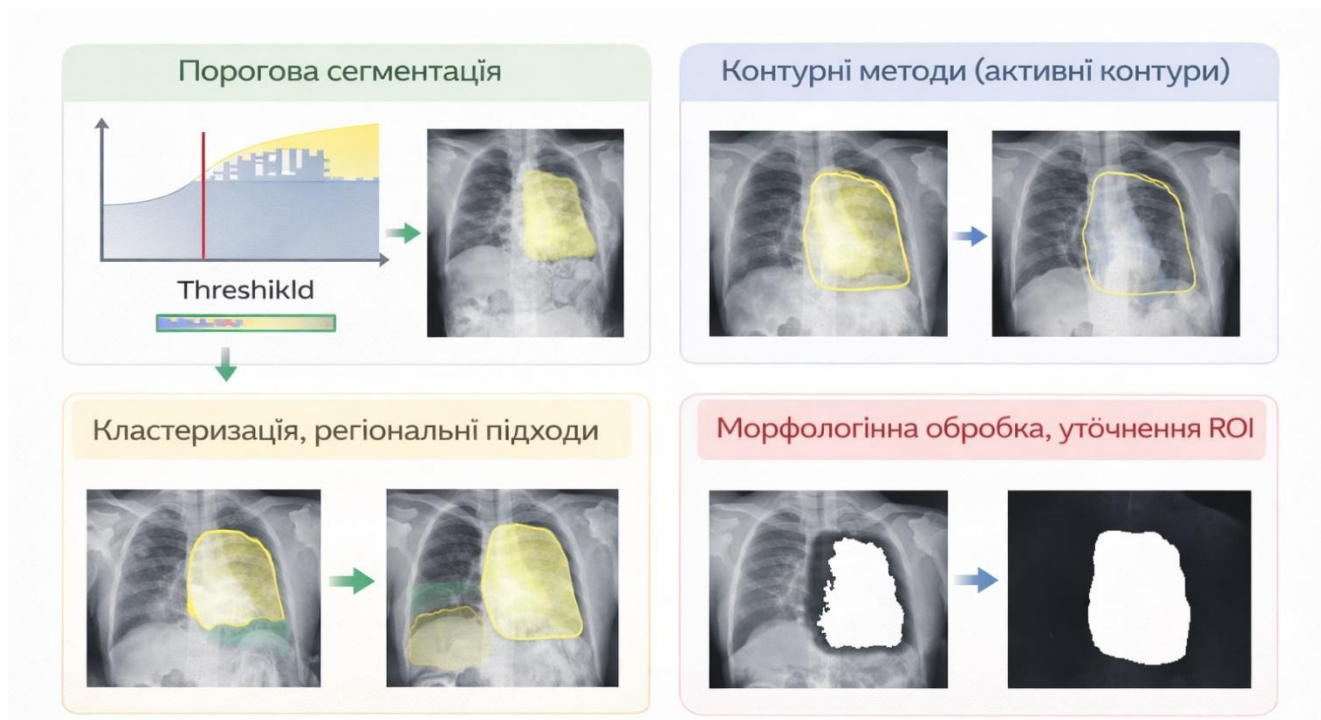


Рисунок. 1.2 – Класичні методи сегментації та локалізації серцевої області інтересу на рентгенівських знімках грудної клітки

Рис. 1.2 ілюструє основні групи класичних підходів до сегментації та локалізації області інтересу на медичних зображеннях: порогову сегментацію, контурні методи, регіональні підходи та морфологічне уточнення ROI. Наведені приклади показують, що попри концептуальну простоту та зрозумілість, такі методи суттєво залежать від якості вхідного зображення, рівня шуму, контрастності межі та вибору параметрів алгоритму. Особливо це критично для рентгенівських знімків грудної клітки, де межа серця часто є слабко вираженою, а сама анатомічна структура спостерігається в умовах накладання інших тіней.

Узагальнюючи, можна зазначити, що класичні методи сегментації є важливими з методологічної точки зору, оскільки формують базові принципи локалізації, роботи з контуром, областю та формою об'єкта. Водночас для задачі стабільного виділення серцевої ROI на CXR вони виявляються недостатньо робастними. Їхні основні недоліки полягають у високій чутливості до шуму, залежності від ручного налаштування параметрів,

нестійкості при слабкому контрасті та труднощах із відтворенням цілісної морфологічно узгодженої маски. Саме ці обмеження стали підґрунтям для переходу до сучасних нейромережових підходів, здатних навчатися складним ознаковим представленням безпосередньо з даних і краще адаптуватися до варіативності медичних зображень.

1.3 Сучасні нейромережові підходи до сегментації та локалізації області інтересу на медичних зображеннях

Упродовж останніх років провідним напрямом розвитку сегментації медичних зображень стали методи глибинного навчання, які забезпечують автоматичне формування інформативних ознак без ручного задання правил, фільтрів і евристик. На відміну від класичних підходів, нейромережові моделі здатні одночасно враховувати локальні текстурні особливості, просторовий контекст і високорівневі семантичні залежності. Саме це зумовило їхнє домінування в задачах сегментації анатомічних структур, де межі об'єкта можуть бути слабкими, неоднорідними або частково перекритими іншими тканинами [1, 3–7, 40].

Сучасні методи штучного інтелекту в задачах аналізу медичних зображень значною мірою базуються на нейронних мережах і глибинному навчанні, які дають змогу автоматично формувати багаторівневі ознаки зображення без ручного задання правил сегментації.

Задача виділення серцевої ROI на CXR-зображеннях може розглядатися як частковий випадок розпізнавання образів, у якому модель повинна відокремити просторову структуру серця від фону, легеневих полів,

Базовою архітектурною парадигмою для медичної сегментації стали encoder–decoder моделі, найвідомішим представником яких є U-Net. Їх принцип полягає у поєднанні двох взаємодоповнювальних процесів: поступового стискання просторового представлення з метою накопичення контексту та подальшого відновлення роздільної здатності для побудови детальної карти сегментації. Наявність пропускових з'єднань між відповідними

рівнями енкодера і декодера дозволяє поєднувати глобальні ознаки з локальними деталями, що особливо важливо для задач, де точність межі має принципове значення. Для сегментації серцевої області на рентгенівських знімках грудної клітки U-Net-подібні моделі є природною відправною точкою, оскільки вони забезпечують достатньо добрий баланс між локалізацією форми серця та загальним контекстом грудної клітки [28].

Подальший розвиток encoder–decoder підходів призвів до появи численних модифікацій, спрямованих на підвищення стійкості моделі та зменшення залежності від ручного підбору конфігурацій. Одним із найбільш відомих рішень цього класу є nnU-Net, який реалізує ідею самоналаштуваного сегментаційного конвеєра. Його перевага полягає в автоматичному узгодженні параметрів передобробки, архітектури та навчання з характеристиками конкретного набору даних. Такий підхід є важливим для медичних задач, де різні датасети можуть істотно відрізнятися за розміром, контрастністю, форматами розмітки та просторовими масштабами. Водночас навіть стандартизовані U-Net-похідні рішення не повністю усувають труднощі, характерні для CXR, зокрема нечіткість межі серця, варіативність проєкцій та появу помилкових виділень у прикордонних зонах [7, 8, 14, 48, 49, 80].

Окремий напрям удосконалення сучасних моделей пов'язаний із механізмами уваги. Attention-підходи дозволяють підсилювати найбільш релевантні просторові або каналні ознаки та пригнічувати менш інформативний фон. У задачі сегментації це важливо, оскільки модель має навчитися зосереджуватися не на всьому зображенні однаково, а саме на тих ділянках, які визначають форму і межі цільового об'єкта. Для рентгенівських знімків грудної клітки така властивість є корисною через наявність великої кількості структур, що створюють конкуренцію за увагу моделі: ребра, легеневий рисунок, середостіння, діафрагма. Використання attention-механізмів дає змогу краще виділяти серцеву область, однак їх перевагу значною мірою залежить від того, наскільки добре базова архітектура вже

враховує просторовий контекст і локальну межу інформацію [10–12].

Ще одна важлива група сучасних підходів пов'язана з багатомасштабним аналізом контексту. До неї належать контекстні модулі типу ASPP та архітектури сімейства DeepLab, які використовують паралельні гілки з різними ефективними полями зору для одночасного охоплення локальних і глобальних залежностей. Це дає змогу моделі краще інтерпретувати об'єкт не лише як сукупність пікселів, а як анатомічну структуру з певною формою, розміром і положенням у сцені. Для сегментації серця на CXR багатомасштабний аналіз є принципово важливим, оскільки силует серця повинен розпізнаватися не лише за локальним контрастом межі, а й у контексті загальної конфігурації органів грудної клітки [9, 74, 75].

Окремим напрямом розвитку сучасних нейромережевих підходів є трансформери та гібридні CNN–Transformer архітектури. На відміну від класичних згорткових мереж, які переважно виділяють локальні просторові ознаки, трансформери дають змогу враховувати глобальні залежності між віддаленими ділянками зображення. Для задач сегментації медичних зображень це є важливим, оскільки форма серцевої області на CXR-знімках визначається не лише локальними контурними ознаками, а й загальним просторовим контекстом у межах грудної клітки. Водночас трансформери не завжди явно враховують морфологічну структуру об'єкта та локальні прикордонні особливості, тому їх доцільно поєднувати з edge-aware і морфологічно-орієнтованими механізмами. [13, 52–57].

Окремо слід зазначити, що в сучасних медичних інформаційних системах дедалі ширше розглядаються великі мовні моделі, які можуть використовуватися не для безпосередньої сегментації зображень, а як допоміжний інструмент для пояснення результатів автоматизованого аналізу, формування текстових висновків і підтримки взаємодії користувача з програмним забезпеченням. У контексті аналізу CXR-зображень такі моделі можуть доповнювати комп'ютерно-зоріві підходи на етапі інтерпретації результатів, однак основним завданням даної роботи залишається

розроблення нейромережевої моделі сегментації серцевої ROI та програмного забезпечення для її практичного використання.

Значна частина сучасних досліджень спрямована безпосередньо на зменшення помилок на межі об'єкта. До таких рішень належать boundary-aware та edge-aware підходи, спеціальні функції втрат, орієнтовані на контур, а також модулі, що явно використовують градієнтну або прикордонну інформацію. Їхня поява є закономірною, оскільки у багатьох медичних задачах основні похибки пов'язані не з грубим промахом по об'єкту, а саме з локальними дефектами межі, хибнопозитивними включеннями фону або хибнонегативними втратами частини цільової області. Для сегментації серця на рентгенівських знімках грудної клітки це є критичним, оскільки найбільша невизначеність виникає саме у прикордонних ділянках між серцем, легенями та діафрагмою. Тому підходи, що явно підсилюють межову інформацію, мають вищу прикладну цінність, ніж архітектури, орієнтовані лише на загальну семантичну сегментацію.

Окремої уваги заслуговують морфологічно-орієнтовані нейромережеві рішення. Їх ключова ідея полягає в тому, щоб не обмежуватися лише статистичним розпізнаванням піксельних патернів, а враховувати форму, зв'язність, структурну цілісність і геометричну узгодженість цільової області. Для анатомічних структур це особливо важливо, оскільки коректний результат сегментації повинен бути не лише схожим за інтенсивністю, а й правдоподібним з морфологічної точки зору. У задачі виділення серцевої ROI така властивість дозволяє зменшувати фрагментацію маски, пригнічувати дрібні хибнопозитивні включення та покращувати відтворення цілісного силуету серця. Саме тому морфологічно-орієнтовані модулі доцільно розглядати як перспективний компонент спеціалізованих архітектур для CXR [15–19, 71].

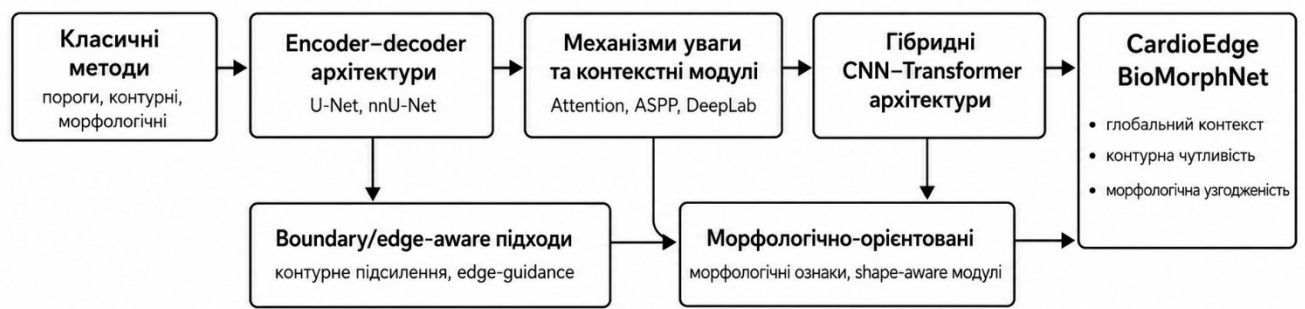


Рисунок. 1.3 – Еволюція сучасних підходів до сегментації серцевої області інтересу на рентгенівських знімках грудної клітки

Узагальнюючи, сучасні нейромережеві підходи суттєво перевершують класичні методи за точністю, гнучкістю та здатністю адаптуватися до складних медичних даних. Проте для задачі сегментації серця на рентгенівських знімках грудної клітки навіть вони не повністю усувають проблеми слабкої вираженості меж, локальної прикордонної невизначеності, структурної варіативності та потреби у морфологічно узгодженому відновленні ROI. Тому найбільш перспективним для даної роботи є напрям, що поєднує encoder-decoder основу з механізмами багатомасштабного контексту, контурного підсилення, морфологічної обробки та адаптивного уточнення прикордонної області. Саме така логіка створює підґрунтя для подальшого порівняльного аналізу підходів і обґрунтування вибору напрямку дослідження [21–23, 31, 43, 72, 73].

1.4 Порівняльний аналіз підходів і обґрунтування напрямку дослідження

Проведений аналіз показав, що методи сегментації та локалізації області інтересу на медичних зображеннях істотно відрізняються за принципом побудови ознак, стійкістю до завад і здатністю відтворювати анатомічно коректну форму об'єкта. Для задачі виділення серцевої області інтересу на рентгенівських знімках грудної клітки ці відмінності мають принципове

значення, оскільки серцевий силует на CXR часто характеризується слабким локальним контрастом, частковим перекриттям суміжними структурами, неоднорідністю межі та варіативністю форми залежно від умов знімання. За таких умов вимоги до методу сегментації виходять за межі простої класифікації пікселів і включають точну локалізацію контуру, збереження зв'язності області та морфологічну узгодженість маски [7, 8, 13–20, 28, 74, 75].

Класичні методи сегментації мають низку переваг, серед яких простота реалізації, інтерпретованість та невисока обчислювальна складність. Порогові, контурні, регіональні та морфологічні підходи дозволяють формалізувати окремі аспекти виділення об'єкта: інтенсивнісне розмежування, пошук границь, об'єднання однорідних областей або уточнення бінарної маски. Однак для сегментації серцевої ROI на рентгенівських знімках їх можливості є обмеженими. Основними причинами є залежність від ручного підбору параметрів, нестійкість до шуму та неоднорідності зображення, а також відсутність механізму, який дозволяв би одночасно враховувати локальні ознаки межі та глобальну анатомічну конфігурацію грудної клітки. У результаті класичні методи можуть бути корисними як допоміжний інструмент попередньої або післяпроцесної обробки, проте не забезпечують достатньої стабільності як основа сегментаційної системи.

Сучасні нейромережеві підходи суттєво розширюють можливості аналізу медичних зображень, оскільки формують ознакові представлення безпосередньо з даних і здатні адаптуватися до складної структури сцени. Найбільш поширеною базою для сегментації є encoder–decoder архітектури, насамперед сімейство U-Net. Їх перевага полягає у поєднанні глобального контексту з локальними деталями через пропускні з'єднання, що особливо важливо для побудови щільної карти сегментації. Саме такі архітектури забезпечили якісний перехід від евристичних методів до моделей, здатних стабільно працювати на реальних медичних даних.

Разом із тим базові encoder–decoder моделі не повністю розв'язують проблему виділення серця на CXR. Попри здатність досить точно локалізувати

цільову область, вони можуть формувати локальні похибки на межі серцевого силуету, допускати надлишкове захоплення фонового простору або втрачати частини ROI в ділянках слабого контрасту. Це пов'язано з тим, що стандартна сегментаційна схема не завжди достатньо чутлива саме до контурної невизначеності та морфологічної структури об'єкта.

Подальший розвиток підходів пов'язаний із використанням механізмів уваги та контекстних модулів. Attention-механізми дозволяють перерозподіляти вагу ознак на користь найбільш інформативних ділянок, зменшуючи вплив нерелевантного фону. Багатомасштабні модулі, зокрема ASPP та споріднені рішення, забезпечують аналіз зображення на різних просторових масштабах, що покращує інтерпретацію положення і форми серця в межах усєї грудної клітки. Такі рішення підвищують якість сегментації, проте навіть вони не гарантують достатньої точності саме на прикордонних ділянках, де і виникає найбільша кількість помилок.

Окрему групу становлять гібридні CNN–Transformer архітектури, які поєднують локальну чутливість згорткових мереж із можливістю моделювання далеких просторових залежностей. Їх перевагою є краще врахування глобального контексту зображення, що є важливим для аналізу CXR. Проте для практичної задачі сегментації серцевої ROI такі моделі нерідко виявляються більш складними в навчанні, вимогливими до обсягу даних і не завжди демонструють вирішальну перевагу саме в точності відновлення контуру серця. Тому використання лише гібридної архітектури без спеціалізованих механізмів прикордонного та морфологічного уточнення не можна вважати достатнім.

Найбільш перспективними для даної задачі є підходи, орієнтовані на межу та морфологічну структуру об'єкта. Boundary-aware та edge-aware рішення підсилюють контурний сигнал і дозволяють моделі точніше локалізувати границю між серцем і суміжними структурами. Морфологічно-орієнтовані модулі, своєю чергою, забезпечують кращу узгодженість форми, зв'язності та цілісності сегментаційної маски. Для серцевої ROI це особливо

важливо, оскільки кінцевий результат повинен бути не лише близьким до еталона за метриками перекриття, а й анатомічно правдоподібним, без фрагментації, розривів і сторонніх включень.

Таким чином, порівняльний аналіз показує, що для задачі сегментації серцевої області на рентгенівських знімках грудної клітки недостатньо спиратися лише на один клас методів. Ефективний підхід повинен поєднувати кілька ключових властивостей: здатність враховувати глобальний контекст зображення, чутливість до локальної межі об'єкта та механізми морфологічного узгодження прогнозованої області. Саме така логіка розвитку підходів узагальнена на рис. 1.3, де показано перехід від класичних методів до сучасних нейромережових архітектур, а далі — до контурно- та морфологічно-орієнтованих рішень.

З огляду на виявлені обмеження існуючих методів, у роботі обрано напрям розроблення спеціалізованої архітектури, орієнтованої саме на виділення серцевої ROI на CXR. Така архітектура повинна базуватися на encoder–decoder схемі як на перевірених основі сегментації, але водночас містити механізми багатомасштабного аналізу, підсилення межі та морфологічного уточнення області інтересу. Поєднання цих компонентів дає змогу зменшити кількість хибнопозитивних і хибнонегативних пікселів, точніше відтворювати контур серця та підвищувати структурну цілісність прогнозованої маски.

Отже, напрям дослідження, обраний у дисертаційній роботі, полягає у створенні моделі CardioEdgeBioMorphNet, у якій поєднано переваги encoder–decoder архітектур, контурно-чутливих механізмів і морфологічно-орієнтованих модулів. Такий вибір є обґрунтованим як з теоретичної, так і з прикладної точки зору, оскільки безпосередньо відповідає специфіці задачі сегментації серцевої області на рентгенівських знімках грудної клітки та створює основу для подальшої постановки задачі, опису моделі й експериментального дослідження її переваги. Великі мовні моделі можуть розглядатися як допоміжний компонент медичних інформаційних систем для

пояснення результатів автоматизованого аналізу та формування текстових висновків, однак у межах цієї роботи основний акцент зроблено на методах комп'ютерного зору для сегментації серцевої ROI на CXR-зображеннях

1.5 Висновки до розділу 1

У першому розділі проаналізовано особливості виділення серцевої області інтересу на рентгенівських знімках грудної клітки та встановлено, що дана задача є складною через проєкційну природу CXR-зображень, слабку вираженість частини меж серцевого силуету, накладання анатомічних структур, наявність шуму, варіативність контрасту та змінність форми об'єкта залежно від умов знімання. Показано, що для таких даних недостатньо лише грубої локалізації області, а необхідним є відтворення цілісної, морфологічно узгодженої ROI-маски серця [15–23, 67, 68, 85].

Розглянуто класичні методи сегментації та локалізації області інтересу на медичних зображеннях, зокрема порогову сегментацію, контурні підходи, регіональні методи, кластеризацію та морфологічну обробку. Встановлено, що їхніми перевагами є простота реалізації, інтерпретованість і невисокі обчислювальні витрати, однак для задачі сегментації серцевої області на CXR їх можливості є обмеженими. Основними недоліками таких підходів є висока чутливість до шуму, залежність від ручного налаштування параметрів, нестійкість при слабкому контрасті та недостатня здатність відтворювати цілісну анатомічно коректну форму серця.

Проаналізовано сучасні нейромережеві підходи до сегментації медичних зображень. Показано, що encoder–decoder архітектури, attention-механізми, багатомасштабні контекстні модулі, гібридні CNN–Transformer рішення, а також boundary-aware, edge-aware і морфологічно-орієнтовані підходи забезпечують значно вищу точність і гнучкість порівняно з класичними методами. Разом із тим встановлено, що для задачі виділення серцевої ROI на рентгенівських знімках грудної клітки недостатньо використовувати лише загальну сегментаційну архітектуру без спеціалізації

на прикордонних ділянках і морфологічній цілісності області.

На основі порівняльного аналізу підходів обґрунтовано, що ефективний метод сегментації серцевої області на CXR повинен поєднувати три ключові властивості: врахування глобального контексту сцени, чутливість до межі об'єкта та морфологічну узгодженість прогнозованої маски. Саме така комбінація є найбільш придатною для зменшення похибок на слабкоконтрастних ділянках, пригнічення фонових включень і відтворення цілісного серцевого силуету.

У результаті обґрунтовано доцільність вибору напряму дослідження, пов'язаного з розробленням спеціалізованої моделі CardioEdgeBioMorphNet, яка поєднує encoder–decoder основу з механізмами контурного підсилення, багатомасштабного аналізу та морфологічного уточнення ROI. Це створює логічний перехід до другого розділу, у якому формулюється постановка задачі, описуються дані, протокол експериментів і принципи побудови сегментаційного конвеєра.

РОЗДІЛ 2. ПОСТАНОВКА ЗАДАЧІ ТА МЕТОДИКА ДОСЛІДЖЕННЯ

У даному розділі розглядається постановка задачі сегментації ROI на CXR у контексті розроблення методу виявлення ішемічної хвороби серця (ІХС) та описується методика проведення експериментальних досліджень, необхідна для коректного навчання моделей і об'єктивного порівняння отриманих результатів.

Використання CXR для задач, пов'язаних із серцево-судинною патологією, є практично важливим, однак має суттєві особливості: знімок є проєкційним (2D) відображенням 3D-анатомії, структури накладаються одна на одну, контраст меж серця часто є слабким, а якість може змінюватися залежно від експозиції, шумів, артефактів та доменного зсуву між різними апаратами та протоколами знімання.

У результаті найбільш критичні помилки моделі концентруються на межі ROI, що вимагає особливої уваги до контролю помилок типу FP та FN, оскільки ці помилки можуть безпосередньо впливати на подальший аналіз та підсумкове рішення щодо наявності ознак ІХС.

Також у розділі сегментація формалізується як задача піксельної класифікації з отриманням багатокласової маски Y та прогнозу \hat{Y} , а також як задача виділення бінарної ROI-маски R та \hat{R} , які утворюються агрегуванням усіх не-фонових пікселів. Такий підхід дозволяє, з одного боку, зберегти семантичну структуру розмітки у процесі підготовки даних і навчання, а з іншого — оцінювати якість саме виділення серцевої області, що є необхідним етапом для подальших компонент системи підтримки прийняття рішень при виявленні ІХС.

Далі у підрозділі 2.1 наведено формалізацію задачі, визначено вхідні дані, вихід моделі, правило формування прогнозу та ROI, а також окреслено обмеження постановки, характерні для CXR а саме проєкційність, перекриття структур, варіативність контрасту.

У підрозділі 2.2 подано опис набору даних рентгенівських знімків грудної клітки, принципи розмітки та кодування міток а, саме перетворення RGB в індекс класу, структуру директорій і розбиття на навчальну, валідаційну та тестову вибірки, що забезпечує відсутність витоків між ними.

У підрозділі 2.3 описано передобробку та аугментації, спрямовані на підвищення робастності моделі до змін умов знімання та шумів, а також правила формування навчальних прикладів.

У підрозділі 2.4 наведено метрики оцінювання якості такі як, Pixel Assurance, Dice, IoU і протокол вимірювання, включно з критерієм вибору найкращого чекпойнта та порівнянням моделей за однакових умов.

У підрозділі 2.5 узагальнено вимоги до методу з позицій стабільності межі, контролю FP/FN, робастності та відтворюваності, що формує методичну основу для подальшого опису архітектур і експериментальних результатів системи виявлення ІХС у наступних розділах роботи.

2.1 Постановка задачі багатокласової сегментації серцевої області

Задача, що розглядається у дисертаційному дослідженні, належить до класу задач семантичної сегментації медичних зображень. На відміну від задачі класифікації, у якій усьому зображенню ставиться у відповідність одна мітка класу, у задачі сегментації необхідно визначити класову належність кожного пікселя. Тому результатом роботи моделі є не одне числове значення для всього знімка, а двовимірний картка міток, просторово узгоджена з вхідним рентгенівським зображенням.

У межах цієї роботи об'єктом аналізу є рентгенівські знімки грудної клітки, а цільовою областю — серцева область. Використана розмітка має багатокласовий характер: пікселі серцевої області належать до одного з шести клінічних класів, а пікселі поза серцевою областю належать до класу background.

Важливо підкреслити, що еталонна маска в роботі не є бінарною. Вона не зводиться до поділу зображення на два класи серце і фон. Основна ground

truth-розмітка є багатокласовою RGB-маскою, у якій кожному пікселю відповідає одна із семи міток: шість клінічних класів серцевої області або фоновий клас. ROI-представлення може бути отримане з такої маски додатково шляхом об'єднання всіх не-фонових класів, але воно використовується лише як похідне представлення для оцінювання якості виділення серцевої області [28, 46, 47, 80, 84].

Нехай вхідне рентгенівське зображення грудної клітки позначається як

$$X \in \mathbb{R}^{H \times W}, \quad (2.1)$$

де X — вхідне CXR-зображення; H — висота зображення; W — ширина зображення; $\mathbb{R}^{H \times W}$ — простір дійснозначних матриць розміром $H \times W$.

Якщо зображення подається у трьохканальному форматі, його можна записати як

$$X \in \mathbb{R}^{H \times W \times K}, \quad (2.2)$$

де K — кількість каналів зображення.

В задачі сегментації серцевої області основна увага приділяється просторовій структурі, контурам, межах серцевої тіні та співвідношенню інтенсивностей між серцем і навколишніми структурами.

Еталонна карта сегментації для зображення X позначається як

$$Y \in \{0, 1, \dots, C - 1\}^{H \times W}, \quad (2.3)$$

де Y — еталонна карта міток, або ground truth-маска; C — кількість класів сегментації; Y_{ij} — мітка класу для пікселя з координатами (i, j) ; $\{0, 1, \dots, C - 1\}$ — множина допустимих індексів класів.

У межах дослідження використовується сім класів сегментації:

$$C = 7, \quad (2.4)$$

де C — загальна кількість семантичних класів у масці.

Множина класів задається так:

$$C = \{c_0, c_1, c_2, c_3, c_4, c_5, c_6\}, \quad (2.5)$$

де C — множина класів сегментації; c_0, c_1, \dots, c_6 — окремі класи, що використовуються у розмітці.

У прийнятій системі класів:

$$\begin{aligned}c_0 &= \text{Healthy}, \\c_1 &= \text{Heart failure}, \\c_2 &= \text{Myocardial infarction}, \\c_3 &= \text{Postinfarction cardiosclerosis}, \\c_4 &= \text{Stable angina}, \\c_5 &= \text{Unstable angina}, \\c_6 &= \text{background}.\end{aligned}\tag{2.6}$$

де c_0 – c_5 — клінічні класи серцевої області; c_6 — фоновий клас, який позначає пікселі поза серцевою областю.

Вихідна розмітка зберігається у вигляді RGB-маски. Кожному класу відповідає фіксований RGB-колір. Це дозволяє візуально перевірити правильність розмітки, а також однозначно перетворювати кольорову маску в індексну карту класів.

Формально RGB-маску можна подати як:

$$M^{RGB} \in \{0, \dots, 255\}^{H \times W \times 3}.\tag{2.7}$$

де M^{RGB} — кольорова RGB-маска сегментації; H — висота маски; W — ширина маски; 3 — кількість колірних каналів RGB; $\{0, \dots, 255\}$ — діапазон значень кожного колірної каналу.

Кожному пікселю RGB-маски відповідає трійка значень:

$$M_{ij}^{RGB} = (r_{ij}, g_{ij}, b_{ij}).\tag{2.8}$$

де M_{ij}^{RGB} — RGB-значення пікселя з координатами (i, j) ; r_{ij} , g_{ij} , b_{ij} — значення червоного, зеленого та синього каналів відповідно.

Для використання маски у процесі навчання RGB-маска перетворюється в індексну карту класів. Таке перетворення задається функцією

$$Y_{ij} = \varphi(M_{ij}^{RGB}),\tag{2.9}$$

де φ — функція відповідності між RGB-кольором пікселя та індексом класу; M_{ij}^{RGB} — RGB-значення пікселя маски; Y_{ij} — числова мітка класу для цього пікселя.

Для функцій втрат і метрик потрібна індексна карта, де кожному пікселю відповідає номер класу.

Для прийнятої палітри функція φ має вигляд:

$$\varphi(r, g, b) = \begin{cases} 0, & \text{якщо } (r, g, b) = (0, 255, 0), \\ 1, & \text{якщо } (r, g, b) = (255, 0, 0), \\ 2, & \text{якщо } (r, g, b) = (184, 184, 94), \\ 3, & \text{якщо } (r, g, b) = (255, 106, 77), \\ 4, & \text{якщо } (r, g, b) = (46, 165, 84), \\ 5, & \text{якщо } (r, g, b) = (250, 250, 55), \\ 6, & \text{якщо } (r, g, b) = (0, 0, 0). \end{cases} \quad (2.10)$$

де (r, g, b) — RGB-колір пікселя маски; значення 0 відповідає класу Healthy; значення 1 відповідає класу Heart failure; значення 2 відповідає класу Myocardial infarction; значення 3 відповідає класу Postinfarction cardiosclerosis; значення 4 відповідає класу Stable angina; значення 5 відповідає класу Unstable angina; значення 6 відповідає класу background.

Після перетворення RGB-маски в індексну карту еталонна розмітка має вигляд

$$Y \in \{0, 1, 2, 3, 4, 5, 6\}^{H \times W}. \quad (2.11)$$

де Y — багатокласова індексна карта сегментації; $\{0, 1, 2, 3, 4, 5, 6\}$ — множина індексів класів; $H \times W$ — просторовий розмір маски.

Нейронна модель сегментації розглядається як параметризоване відображення

$$F_{\theta}: X \rightarrow \hat{Y}. \quad (2.12)$$

де F_{θ} — нейронна модель сегментації з параметрами θ ; X — вхідне рентге

На практиці модель спочатку формує не самі мітки класів, а тензор логітів:

$$Z = F_{\theta}(X). \quad (2.13)$$

де Z — тензор логітів, сформований моделлю; $F_{\theta}(X)$ — результат обробки вхідного зображення моделлю.

Тензор логітів має вигляд

$$Z \in \mathbb{R}^{K \times H \times W}. \quad (2.14)$$

де K — кількість вихідних каналів сегментаційної голови; H — висота

вихідної карти; W — ширина вихідної карти.

У класичному 7-класовому режимі $K = 7$, тобто модель безпосередньо прогнозує шість клінічних класів серцевої області та клас background. У спеціалізованому ROI-орієнтованому режимі, реалізованому для CardioEdgeBioMorphNet, сегментаційна голова може прогнозувати шість клінічних класів серцевої області, а фоновий клас використовується як службова мітка для позначення пікселів поза ROI та для відкидання невпевнених передбачень на етапі інференсу.

Для 7-класового режиму ймовірність належності пікселя до класу k визначається за допомогою функції softmax:

$$P_{kij} = \frac{\exp(Z_{kij})}{\sum_{m=0}^6 \exp(Z_{mij})}, \quad k \in \{0,1, \dots, 6\}. \quad (2.15)$$

де P_{kij} — ймовірність належності пікселя (i, j) до класу k ; Z_{kij} — логіт для класу k у пікселі (i, j) ; $\exp(\cdot)$ — експоненціальна функція; $\sum_{m=0}^6 \exp(Z_{mij})$ — сума експонент логітів за всіма сімома класами.

Остаточна прогнозована мітка пікселя визначається як клас із найбільшою ймовірністю:

$$\hat{Y}_{ij} = \arg \max_{k \in \{0,1, \dots, 6\}} P_{kij}. \quad (2.16)$$

де \hat{Y}_{ij} — прогнозована мітка класу для пікселя (i, j) ; $\arg \max$ — операція вибору індексу класу з найбільшою ймовірністю; P_{kij} — ймовірність належності пікселя до класу k .

У ROI-орієнтованому режимі, коли сегментаційна голова формує ймовірності лише для шести клінічних класів серцевої області, softmax обчислюється за класами 0–5:

$$P_{kij} = \frac{\exp(Z_{kij})}{\sum_{m=0}^5 \exp(Z_{mij})}, \quad k \in \{0,1, \dots, 5\}. \quad (2.17)$$

де P_{kij} — ймовірність належності пікселя до одного з клінічних класів; Z_{kij} — логіт для клінічного класу k ; $k \in \{0,1, \dots, 5\}$ — множина індексів

клінічних класів серцевої області.

Початкова прогнозована мітка клінічного класу визначається як

$$\hat{Y}_{ij}^{heart} = \arg \max_{k \in \{0,1,\dots,5\}} P_{kij}. \quad (2.18)$$

де \hat{Y}_{ij}^{heart} — прогнозований клінічний клас пікселя; P_{kij} — ймовірність належності пікселя до класу k ; $\arg \max$ — операція вибору класу з найбільшою ймовірністю.

Для формування повної карти сегментації необхідно також визначити, чи належить піксель до фону. Нехай максимальна ймовірність серед клінічних класів дорівнює:

$$p_{ij}^{max} = \max_{k \in \{0,1,\dots,5\}} P_{kij}. \quad (2.19)$$

де p_{ij}^{max} — максимальна ймовірність серед шести клінічних класів для пікселя (i, j) ; P_{kij} — ймовірність належності пікселя до клінічного класу k ; \max — операція вибору найбільшого значення.

Тоді підсумкова мітка пікселя може бути визначена так:

$$\hat{Y}_{ij} = \begin{cases} \hat{Y}_{ij}^{heart}, & \text{якщо } p_{ij}^{max} \geq \tau, \\ 6, & \text{якщо } p_{ij}^{max} < \tau. \end{cases} \quad (2.20)$$

де \hat{Y}_{ij} — підсумкова прогнозована мітка пікселя; \hat{Y}_{ij}^{heart} — прогнозований клінічний клас пікселя; p_{ij}^{max} — максимальна ймовірність серед клінічних класів; τ — поріг упевненості; значення 6 відповідає класу background.

Серцева область, або ROI, визначається як множина пікселів, що належать до будь-якого з шести клінічних класів:

$$R = \{ (i, j) \mid Y_{ij} \in \{0,1,2,3,4,5\} \}. \quad (2.21)$$

де R — еталонна серцева область; (i, j) — координати пікселя; Y_{ij} — еталонна мітка класу; $\{0,1,2,3,4,5\}$ — множина клінічних класів серцевої області.

Фонова область визначається як

$$R_{BG} = \{ (i, j) \mid Y_{ij} = 6 \}. \quad (2.22)$$

де R_{BG} — множина фонових пікселів; $Y_{ij} = 6$ — умова належності пікселя до класу background.

Прогнозована серцева область визначається як:

$$\hat{R} = \left\{ (i, j) \mid \hat{Y}_{ij} \in \{0,1,2,3,4,5\} \right\}. \quad (2.23)$$

де \hat{R} — прогнозована серцева область; \hat{Y}_{ij} — прогнозована мітка класу; $\{0,1,2,3,4,5\}$ — множина клінічних класів серцевої області.

З багатокласової карти можна отримати допоміжне ROI-представлення:

$$Y_{ij}^{ROI} = \begin{cases} 1, & \text{якщо } Y_{ij} \in \{0,1,2,3,4,5\}, \\ 0, & \text{якщо } Y_{ij} = 6. \end{cases} \quad (2.24)$$

де Y_{ij}^{ROI} — значення допоміжної ROI-карти для пікселя (i, j) ; значення 1 означає, що піксель належить до серцевої області; значення 0 означає, що піксель належить до фону.

Аналогічно для прогнозованої маски:

$$\hat{Y}_{ij}^{ROI} = \begin{cases} 1, & \text{якщо } \hat{Y}_{ij} \in \{0,1,2,3,4,5\}, \\ 0, & \text{якщо } \hat{Y}_{ij} = 6. \end{cases} \quad (2.25)$$

де \hat{Y}_{ij}^{ROI} — значення прогнозованої ROI-карти; \hat{Y}_{ij} — прогнозована багатокласова мітка пікселя; значення 1 відповідає прогнозу серцевої області; значення 0 відповідає прогнозу фону.

Еталонну серцеву область можна також подати як об'єднання областей усіх клінічних класів:

$$R = R_{Healthy} \cup R_{HF} \cup R_{MI} \cup R_{PICS} \cup R_{SA} \cup R_{UA}. \quad (2.26)$$

де R — повна серцева область; $R_{Healthy}$ — область класу Healthy; R_{HF} — область класу Heart failure; R_{MI} — область класу Myocardial infarction; R_{PICS} — область класу Postinfarction cardiosclerosis; R_{SA} — область класу Stable angina; R_{UA} — область класу Unstable angina.

Формула (2.26) показує, що ROI формується як об'єднання всіх шести клінічних класів. Тобто незалежно від того, до якого саме клінічного класу належить піксель серцевої області, він входить до загальної ROI, якщо не є фоном.

Окремі області клінічних класів визначаються так:

$$\begin{aligned}
 R_{Healthy} &= \{ (i, j) \mid Y_{ij} = 0 \}, \\
 R_{HF} &= \{ (i, j) \mid Y_{ij} = 1 \}, \\
 R_{MI} &= \{ (i, j) \mid Y_{ij} = 2 \}, \\
 R_{PICS} &= \{ (i, j) \mid Y_{ij} = 3 \}, \\
 R_{SA} &= \{ (i, j) \mid Y_{ij} = 4 \}, \\
 R_{UA} &= \{ (i, j) \mid Y_{ij} = 5 \}.
 \end{aligned}
 \tag{2.27}$$

де кожна множина містить пікселі відповідного клінічного класу.

Формула (2.27) деталізує структуру ROI за класами. Вона дозволяє окремо аналізувати не лише загальну серцеву область, а й області, що відповідають кожній клінічній мітці.

Для коректної сегментаційної маски кожен піксель повинен належати рівно одному класу:

$$\sum_{k=0}^6 \mathbf{1}(Y_{ij} = k) = 1, \quad \forall(i, j).
 \tag{2.28}$$

де $\mathbf{1}(\cdot)$ — індикаторна функція; Y_{ij} — еталонна мітка пікселя; k — індекс класу; умова $\forall(i, j)$ означає, що співвідношення виконується для кожного пікселя зображення.

Формула (2.28) означає, що один піксель не може одночасно належати до двох клінічних класів або одночасно бути частиною серцевої області й фону. Тому палітра RGB-маски повинна бути фіксованою, а перетворення RGB-маски в індексну карту класів має бути однозначним.

У багатокласовій постановці можливі кілька типів помилок. Перший тип — помилка між клінічними класами, коли піксель належить до серцевої області як в еталонній, так і в прогнозованій масці, але отримує неправильну клінічну мітку:

$$E_{class} = \{ (i, j) \mid \hat{Y}_{ij} \neq Y_{ij}, Y_{ij} \neq 6, \hat{Y}_{ij} \neq 6 \}.
 \tag{2.29}$$

де E_{class} — множина пікселів із помилкою між клінічними класами; Y_{ij} — еталонна мітка пікселя; \hat{Y}_{ij} — прогнозована мітка пікселя; значення 6

відповідає класу background; умови $Y_{ij} \neq 6$ та $\hat{Y}_{ij} \neq 6$ означають, що піксель належить до серцевої області і в еталонній, і в прогнозованій масці.

Формула (2.29) описує випадок, коли модель не втратила ROI, але помилилася у внутрішній класифікації пікселя. Наприклад, піксель, який належить до класу Stable angina, може бути помилково віднесений до класу Unstable angina.

Другий тип — хибнопозитивне включення фонових пікселів до серцевої області:

$$FP = \{ (i, j) \mid \hat{Y}_{ij} \neq 6, Y_{ij} = 6 \}. \quad (2.30)$$

де FP — множина хибнопозитивних пікселів; $Y_{ij} = 6$ означає, що в еталонній масці піксель належить до фону; $\hat{Y}_{ij} \neq 6$ означає, що модель помилково віднесла цей піксель до одного з клінічних класів серцевої області.

Формула (2.30) описує ситуацію, коли модель включає сторонні фонові ділянки в серцеву область. Такі помилки можуть виникати через схожість інтенсивностей між серцевою тінню, ребрами, діафрагмою або іншими анатомічними структурами.

Третій тип — хибнонегативна втрата пікселів серцевої області:

$$FN = \{ (i, j) \mid \hat{Y}_{ij} = 6, Y_{ij} \neq 6 \}. \quad (2.31)$$

де FN — множина хибнонегативних пікселів; $Y_{ij} \neq 6$ означає, що в еталонній масці піксель належить до одного з клінічних класів серцевої області; $\hat{Y}_{ij} = 6$ означає, що модель помилково віднесла цей піксель до фону.

Формула (2.31) описує втрату частини серцевої області. Такі помилки є критичними, оскільки навіть локальна втрата частини контуру може змінити площу, форму, периметр і морфологічні характеристики ROI.

Навчання моделі можна подати як задачу мінімізації функції втрат на навчальній вибірці. Нехай навчальна вибірка складається з N пар зображення–маска:

$$D_{train} = \{ (X^{(n)}, Y^{(n)}) \}_{n=1}^N, \quad (2.32)$$

де D_{train} — навчальна вибірка; N — кількість навчальних прикладів;

$X^{(n)}$ — n -те вхідне CXR-зображення; $Y^{(n)}$ — відповідна еталонна багатокласова маска сегментації.

Формула (2.32) описує навчальну вибірку як набір пар, у яких кожному зображенню відповідає власна еталонна маска. Саме на таких парах модель навчається встановлювати відповідність між інтенсивнісною структурою знімка та класовою картою сегментації.

Метою навчання є знаходження таких параметрів моделі θ , за яких середня помилка між прогнозованими та еталонними масками буде мінімальною:

$$\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_{n=1}^N \mathcal{L}(F_{\theta}(X^{(n)}), Y^{(n)}). \quad (2.33)$$

де θ^* — оптимальні параметри моделі; θ — поточні параметри моделі; N — кількість навчальних прикладів; \mathcal{L} — функція втрат; $F_{\theta}(X^{(n)})$ — прогноз моделі для n -го зображення; $Y^{(n)}$ — еталонна маска для n -го зображення.

Формула (2.33) задає загальний принцип навчання нейронної мережі. Параметри моделі змінюються так, щоб мінімізувати середню помилку на навчальній вибірці. Чим менше значення функції втрат, тим ближчою є прогнозована маска до еталонної.

У 7-класовому режимі багатокласова Cross-Entropy може бути записана як:

$$\mathcal{L}_{CE}^7 = -\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=0}^6 \mathbf{1}(Y_{ij} = k) \log P_{kij}. \quad (2.34)$$

де \mathcal{L}_{CE}^7 — функція втрат Cross-Entropy для 7-класового режиму; H — висота маски; W — ширина маски; k — індекс класу; $\mathbf{1}(Y_{ij} = k)$ — індикатор того, що піксель (i, j) належить до класу k ; P_{kij} — прогнозована ймовірність належності пікселя (i, j) до класу k .

У ROI-орієнтованому режимі, який застосовується для спеціалізованої моделі CardioEdgeBioMorphNet, основна функція втрат може обчислюватися

лише за клінічними класами серцевої області. У такому випадку фонові пікселі з міткою 6 переводяться у службове значення *ignore_index* і не враховуються в основній Cross-Entropy-втратах:

$$Y_{ij}^{train} = \begin{cases} Y_{ij}, & \text{якщо } Y_{ij} \in \{0,1,2,3,4,5\}, \\ ignore, & \text{якщо } Y_{ij} = 6. \end{cases} \quad (2.35)$$

де Y_{ij}^{train} — мітка пікселя, яка використовується під час навчання; Y_{ij} — початкова еталонна мітка пікселя; *ignore* — службове значення, яке виключає піксель з обчислення основної функції втрат; значення 6 відповідає класу *background*.

Формула (2.35) показує, що у спеціалізованому режимі *background* може не брати участі у безпосередній оптимізації клінічних класів. Це дозволяє зменшити вплив домінування фонових пікселів, яких на CXR-зображенні зазвичай значно більше, ніж пікселів серцевої області.

Тоді функція втрат для клінічних класів може бути подана як:

$$\mathcal{L}_{CE}^{heart} = -\frac{1}{|\Omega_{heart}|} \sum_{(i,j) \in \Omega_{heart}} \sum_{k=0}^5 \mathbf{1}(Y_{ij} = k) \log P_{kij}. \quad (2.36)$$

де \mathcal{L}_{CE}^{heart} — Cross-Entropy-втрата для клінічних класів серцевої області; Ω_{heart} — множина пікселів серцевої області; $|\Omega_{heart}|$ — кількість пікселів у цій множині; $k \in \{0,1, \dots, 5\}$ — індекси клінічних класів; P_{kij} — прогнозована ймовірність належності пікселя до класу k .

Формула (2.36) описує усереднену Cross-Entropy-втрата лише для пікселів серцевої області. На відміну від повного 7-класового режиму, у цьому випадку фонові пікселі не беруть участі в основному обчисленні втрати. Це дозволяє зосередити оптимізацію моделі на розрізненні шести клінічних класів серцевої області.

Множина пікселів серцевої області визначається так:

$$\Omega_{heart} = \{ (i, j) \mid Y_{ij} \neq 6 \}. \quad (2.37)$$

де Ω_{heart} — множина пікселів, які належать до серцевої області; $Y_{ij} \neq 6$ — умова того, що піксель не належить до фонового класу *background*.

Виключення фонових пікселів з основної функції втрат не змінює природу розмітки. Маска залишається багатокласовою, а клас `background` продовжує використовуватися для формування повної карти сегментації, побудови ROI-представлення, аналізу помилок FP/FN і відкидання невпевнених передбачень.

Таким чином, постановка задачі у роботі має багатокласовий характер. Для кожного вхідного CXR-зображення необхідно побудувати карту сегментації, у якій кожен піксель належить одному з шести клінічних класів серцевої області або класу `background`. ROI-представлення використовується як похідний інструмент для оцінювання якості виділення серцевої області, але не замінює багатокласову `ground truth`-розмітку [15–19, 91–93].

2.2 Підготовка датасету та структура вибірок

Для навчання, валідації та тестування моделей у роботі використано набір рентгенівських знімків грудної клітки з відповідними еталонними масками сегментації (див. додаток Б). Кожному вхідному CXR-зображенню відповідає RGB-маска, у якій пікселі серцевої області позначені одним із шести клінічних класів, а пікселі поза серцевою областю належать до класу `background` [59–61, 81, 82, 107].

На відміну від задачі класифікації, де для кожного зображення достатньо однієї мітки класу, у задачі сегментації кожному зображенню ставиться у відповідність повнорозмірна карта міток. Тому основною одиницею датасету є пара зображення–маска. Вхідне зображення містить інтенсивнісну інформацію рентгенівського знімка, а маска задає еталонну піксельну розмітку.

Повний набір даних можна подати як множину пар:

$$\mathcal{D} = \{(X^{(n)}, Y^{(n)})\}_{n=1}^N \quad (2.38)$$

де \mathcal{D} — повний набір даних; $X^{(n)}$ — n -те рентгенівське зображення грудної клітки; $Y^{(n)}$ — відповідна еталонна багатокласова маска сегментації; N — загальна кількість пар зображення–маска.

У роботі використано набір даних, що містить 960 пар зображення–маска. Еталонні маски мають багатокласовий характер і відповідають системі міток, описаній у підрозділі 2.1. Тобто кожен піксель маски належить одному з семи класів: Healthy, Heart failure, Myocardial infarction, Postinfarction cardiosclerosis, Stable angina, Unstable angina або background.

Організація датасету виконана у форматі, наближеному до VOC-подібної структури, яка є зручною для задач семантичної сегментації. У такому форматі вхідні зображення, маски та файли розбиття зберігаються окремо. Це спрощує контроль даних, дозволяє явно задавати навчальну, валідаційну і тестову вибірки та забезпечує відтворюваність експериментів.

У директорії JPEGImages зберігаються початкові CXR-зображення. У директорії SegmentationClass розміщуються відповідні RGB-маски, у яких кожен клас закодований визначеним кольором. Директорія ImageSets в Segmentation містить текстові файли зі списками ідентифікаторів зображень, що використовуються для формування навчальної, валідаційної та тестової вибірок.

Важливо, щоб кожному зображенню відповідала лише одна коректна маска сегментації. Якщо зображення і маска не відповідають одне одному, модель отримує неправильну навчальну пару, що призводить до некоректного навчання. Тому під час підготовки даних перевіряється наявність відповідних файлів, узгодженість імен, відповідність форматів, а також коректність RGB-палітри масок.

Окрему увагу приділено тому, що маска не є звичайним зображенням у сенсі інтерполяції кольорів. У масці кожен колір має дискретне значення та відповідає конкретному класу. Тому будь-які операції над масками повинні зберігати цю дискретність. Наприклад, під час зміни розміру маски не можна

використовувати інтерполяцію, яка створює проміжні кольори, оскільки такі кольори не відповідатимуть жодному класу.

Перед навчанням повний набір даних поділяється на три неперетинні частини: навчальну, валідаційну та тестову вибірки. Такий поділ необхідний для коректної перевірки якості моделі. Навчальна вибірка використовується для оптимізації параметрів моделі, валідаційна — для контролю процесу навчання та вибору найкращого стану моделі, а тестова — для фінальної оцінки якості після завершення навчання.

Розбиття набору даних можна подати так:

$$\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}} \cup \mathcal{D}_{\text{test}} \quad (2.39)$$

де \mathcal{D} — повний набір даних; $\mathcal{D}_{\text{train}}$ — навчальна вибірка; \mathcal{D}_{val} — валідаційна вибірка; $\mathcal{D}_{\text{test}}$ — тестова вибірка.

У роботі використано таке співвідношення вибірок: 960 зображень для навчання, 120 зображень для валідації та 120 зображень для тестування. Загальна кількість прикладів становить 1200.

$$N = N_{\text{train}} + N_{\text{val}} + N_{\text{test}} = 960 + 120 + 120 = 1200 \quad (2.40)$$

де N — загальна кількість зображень у наборі даних; N_{train} — кількість зображень у навчальній вибірці; N_{val} — кількість зображень у валідаційній вибірці; N_{test} — кількість зображень у тестовій вибірці.

Таке співвідношення відповідає схемі 80 % / 10 % / 10 %. Навчальна вибірка є найбільшою, оскільки саме на ній модель навчається знаходити відповідність між рентгенівським зображенням і маскою сегментації. Валідаційна та тестова вибірки мають однаковий обсяг і використовуються для незалежного контролю якості.

Навчальна вибірка використовується для оновлення ваг нейронної мережі. Під час навчання модель отримує зображення та відповідні маски, формує прогноз, після чого обчислюється функція втрат між прогнозованою та еталонною масками. На основі цієї помилки виконується оновлення параметрів моделі.

Валідаційна вибірка не бере участі в оновленні параметрів. Вона використовується для оцінки якості моделі після епох навчання, контролю перенавчання та вибору найкращого checkpoint. Це дозволяє не орієнтуватися лише на якість на навчальних даних, а контролювати, наскільки добре модель працює на зображеннях, які не використовувалися для оптимізації ваг.

Тестова вибірка використовується лише після завершення навчання та вибору найкращої моделі. Вона не повинна впливати на підбір параметрів, налаштування архітектури або вибір епохи. Саме тестові результати дають найбільш об'єктивне уявлення про здатність моделі працювати з новими рентгенівськими знімками.

Принципово важливо, щоб навчальна, валідаційна та тестова вибірки не перетиналися між собою. Якщо одне й те саме зображення потрапить одночасно до навчальної та тестової вибірки, це призведе до витoku даних і штучного завищення результатів. Тому під час формування датасету дотримується умова неперетинності вибірок:

$$\mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{val}} = \mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{test}} = \mathcal{D}_{\text{val}} \cap \mathcal{D}_{\text{test}} = \emptyset \quad (2.41)$$

де \emptyset - порожня множина

Оскільки в роботі використовується RGB-розмітка, важливим етапом підготовки є перевірка палітри масок. Кожен піксель маски повинен мати один із допустимих RGB-кольорів, наведених у табл. 2.1. Якщо в масці з'являється колір, відсутній у палітрі, такий піксель не може бути однозначно перетворений у класовий індекс. Найчастіше така проблема може виникати після неправильного масштабування маски або збереження її з втратами якості.

Тому на етапі підготовки даних контролюються такі умови: для кожного зображення існує відповідна маска; зображення і маска мають узгоджені ідентифікатори; маска містить лише кольори з визначеної палітри; маска може бути однозначно перетворена в індексну карту класів; списки train, validation і

test не мають перетинів; усі моделі навчаються та тестуються на однаковому розбитті.

Також важливо, щоб зображення і відповідна маска описували одну й ту саму просторову область. Якщо після зчитування або попередньої обробки їхні розміри відрізняються, вони мають бути приведені до єдиного розміру синхронно. Для зображення можна використовувати інтерполяцію, що зберігає візуальну якість, тоді як для маски необхідно використовувати підхід, який не створює нових проміжних класів. Детально це розглядається у підрозділі 2.3, присвяченому попередній обробці даних.

Підготовка датасету має безпосередній вплив на якість навчання. Якщо маски містять помилки, неправильні кольори або не відповідають зображенням, модель буде навчатися на некоректних прикладах. У задачах медичної сегментації це особливо критично, оскільки помилки розмітки в області межі серця можуть призвести до зміщення контуру ROI, появи хибнопозитивних включень або втрати частини серцевої області.

У межах дослідження однакове розбиття даних використовується для навчання й оцінювання базових моделей та запропонованої архітектури CardioEdgeBioMorphNet. Це дозволяє порівнювати моделі в однакових умовах. Таким чином, різниця у результатах пояснюється не різним складом вибірок, а властивостями самих архітектур, механізмів обробки межі, морфологічних блоків та способів роботи з фоновим класом[58].

Отже, підготовка датасету в роботі включає не лише збереження зображень і масок, а й контроль їх відповідності, перевірку коректності RGB-палітри, формування фіксованих списків train, validation і test, а також забезпечення єдиного експериментального протоколу. Така організація даних є необхідною умовою для коректного навчання, об'єктивної валідації та фінального тестування моделей сегментації серцевої області.

2.3. Попередня обробка даних та формування навчальних прикладів

Попередня обробка даних є необхідним етапом перед навчанням нейронної мережі, оскільки вхідні рентгенівські знімки та відповідні маски сегментації можуть відрізнятися за розміром, форматом збереження, кількістю каналів, діапазоном інтенсивностей і технічними характеристиками. Для коректного навчання моделі всі зображення та маски мають бути приведені до єдиного формату, у якому кожному вхідному зображенню відповідає узгоджена багатокласова карта міток [26, 65].

У межах цього дослідження попередня обробка виконується для двох типів даних: CXR-зображень та RGB-масок сегментації. Зображення містять інтенсивнісну інформацію про рентгенівський знімок грудної клітки, тоді як маски містять еталонну піксельну розмітку. Оскільки ці два об'єкти мають різну природу, до них застосовуються різні правила обробки. Для зображень допустимі операції, що змінюють інтенсивності або плавно масштабують пікселі, тоді як для масок необхідно зберігати дискретні мітки класів.

Загальна мета попередньої обробки полягає в тому, щоб перетворити початкову пару “зображення–маска” у пару тензорів, придатних для подання на вхід нейронної мережі. У загальному вигляді цей процес можна подати як відображення:

$$S_n = (X_n, M_n) \rightarrow (X_{tensor,n}, Y_{tensor,n}) \quad (2.42)$$

де $X^{(n)}$ — початкове CXR-зображення; $M^{RGB(n)}$ — відповідна RGB-маска сегментації; $X_{tensor}^{(n)}$ — підготовлений тензор зображення; $Y_{tensor}^{(n)}$ — підготовлена індексна карта класів у тензорному форматі.

Першим етапом є зчитування CXR-зображення та відповідної RGB-маски. Для кожного зображення має бути знайдена маска з відповідним ідентифікатором. Якщо така відповідність відсутня, приклад не може бути використаний для навчання, оскільки модель не матиме еталонної розмітки для порівняння з прогнозом.

Після зчитування виконується перевірка формату зображення. Рентгенівські знімки за змістом є зображеннями у відтінках сірого, тому основна інформація в них міститься у просторовому розподілі інтенсивностей. Якщо знімок зберігається у трьохканальному форматі, він може бути приведений до одноканального представлення або використаний у вигляді трьох каналів залежно від вимог конкретної архітектури. Для запропонованої задачі ключовим є не колір, а форма, контур, межі та інтенсивнісні переходи в області серцевої тіні.

Наступним етапом є приведення зображень і масок до єдиного просторового розміру. У роботі використовується розмір 128×128 , що є компромісом між обчислювальною ефективністю та збереженням основної геометричної інформації про серцеву область. Такий розмір дозволяє зменшити вимоги до пам'яті, пришвидшити навчання та забезпечити однаковий формат вхідних даних для всіх моделей.

Після зміни розміру вхідне зображення можна подати як:

$$X_{resized} \in R^{128 \times 128} \quad (2.43)$$

де $X_{resized}$ — CXR-зображення після приведення до єдиного розміру; 128×128 — просторовий розмір, який використовується у подальших експериментах.

Для маски після зміни розміру має зберігатися її дискретна структура:

$$Y_{resized} \in \{0,1,2,3,4,5,6\}^{128 \times 128} \quad (2.44)$$

де $Y_{resized}$ — індексна карта класів після приведення до розміру 128×128 ; множина $\{0,1,2,3,4,5,6\}$ відповідає семи класам сегментації.

Для зображень і масок використовуються різні підходи до масштабування. Для CXR-зображення можна застосовувати інтерполяцію, яка плавно змінює інтенсивності пікселів, оскільки зображення є неперервним за значеннями яскравості. Натомість для масок необхідно використовувати інтерполяцію найближчого сусіда. Це пов'язано з тим, що RGB-маска або

індексна карта містить дискретні класи, і будь-яке усереднення кольорів може створити нові значення, яких немає в затвердженій палітрі.

Наприклад, якщо під час зміни розміру RGB-маски використати лінійну інтерполяцію, між кольорами двох сусідніх класів може виникнути проміжний колір. Такий колір не відповідатиме жодному класу з таблиці палітри, а отже не зможе бути коректно перетворений у числову мітку. Тому для масок використовується тільки такий тип масштабування, який не створює нових класових значень.

Після приведення до єдиного розміру виконується нормалізація інтенсивностей CXR-зображення. Якщо початкові значення пікселів зберігаються в діапазоні від 0 до 255, то їх доцільно привести до діапазону від 0 до 1:

$$X_{norm}(i,j) = \frac{X_{resized}(i,j)}{255} \quad (2.45)$$

де $X_{norm}(i,j)$ — нормалізоване значення інтенсивності пікселя з координатами (i,j) ; $X_{resized}(i,j)$ — початкове значення пікселя після зміни розміру; 255 — максимальне значення інтенсивності для 8-бітного зображення.

Нормалізація є важливою саме для рентгенівських знімків, оскільки вони можуть відрізнитися за контрастом, яскравістю, експозицією та умовами отримання. Якщо не привести інтенсивності до єдиного масштабу, модель може частково навчатися не на анатомічних ознаках, а на технічних відмінностях між зображеннями.

Окремим етапом є перетворення RGB-маски в індексну карту класів. Як було зазначено у підрозділі 2.1, кожен клас у масці має власний RGB-колір. Перед навчанням ці кольори перетворюються у числові індекси від 0 до 6:

$$Y_{ij} = \varphi(M_{ij}) \quad (2.46)$$

де Y_{ij} — індекс класу для пікселя (i,j) ; M_{ijRGB} — RGB-значення відповідного пікселя маски; φ — функція відповідності між RGB-кольором і номером класу.

Після перетворення маска має вигляд:

$$Y_{tensor} \in \{0,1,2,3,4,5,6\}^{128 \times 128} \quad (2.47)$$

де Y_{tensor} — підготовлена карта класів для навчання; значення 0–5 відповідають клінічним класам серцевої області, а значення 6 відповідає класу background.

Для подання зображення на вхід нейронної мережі нормалізоване зображення перетворюється у тензор. Для одноканального CXR-зображення форма такого тензора має вигляд:

$$X_{tensor} \in R^{1 \times 128 \times 128} \quad (2.48)$$

де X_{tensor} — тензор вхідного зображення; перша розмірність відповідає кількості каналів; 128×128 — просторовий розмір зображення.

Під час навчання зображення об'єднуються у батчі. Для батчу з B зображень вхідний тензор має форму

$$X_{batch} \in R^{B \times 1 \times 128 \times 128} \quad (2.49)$$

де X_{batch} — батч вхідних зображень; B — кількість зображень у батчі; 1 — кількість каналів; 128×128 — просторовий розмір кожного зображення.

Відповідний батч масок має вигляд

$$Y_{batch} \in \{0,1,2,3,4,5,6\}^{B \times 128 \times 128} \quad (2.50)$$

де Y_{batch} — батч еталонних масок; B — кількість масок у батчі; 128×128 — просторовий розмір кожної маски.

У процесі підготовки навчальних прикладів також можуть застосовуватися аугментації. Їхня мета полягає у збільшенні різноманітності навчальних даних і підвищенні стійкості моделі до варіацій зображень. Для CXR-знімків такими варіаціями можуть бути зміни яскравості, контрасту, незначні геометричні зміщення, масштабування або повороти в допустимих межах.

Важливо, що геометричні аугментації повинні застосовуватися синхронно до зображення та відповідної маски. Якщо зображення повертається або масштабується, така сама геометрична операція має бути

застосована і до маски. Інакше порушиться відповідність між пікселями зображення та еталонними мітками.

Для навчальної вибірки аугментацію можна подати як перетворення:

$$(X_{aug}, Y_{aug}) = T(X, Y) \quad (2.51)$$

де T — спільне перетворення, що застосовується до зображення та маски; X_{aug} — аугментоване зображення; Y_{aug} — відповідна аугментована маска.

Аугментації застосовуються лише до навчальної вибірки. Валідаційна та тестова вибірки не повинні піддаватися випадковим перетворенням, оскільки вони використовуються для об'єктивного контролю якості моделі. Якщо валідаційні або тестові дані змінювати випадковими аугментаціями, результати оцінювання можуть стати нестабільними та залежними від конкретного випадкового перетворення.

Такий процес підготовки дозволяє забезпечити однакову процедуру обробки даних для всіх моделей, що порівнюються в роботі. Це важливо для коректного експериментального протоколу, оскільки різниця в результатах повинна пояснюватися властивостями архітектур, а не відмінностями в обробці вхідних даних.

Окремо слід підкреслити, що етап попередньої обробки не змінює семантичний зміст розмітки. Класи залишаються тими самими, що й у початковій RGB-масці: шість клінічних класів серцевої області та клас background. Усі перетворення спрямовані лише на приведення даних до формату, придатного для навчання нейронної мережі.

Таким чином, попередня обробка даних у роботі включає приведення CXR-зображень і масок до єдиного розміру, нормалізацію інтенсивностей, перетворення RGB-масок в індексні карти класів, формування тензорів і застосування аугментацій для навчальної вибірки. Цей етап є необхідною умовою стабільного навчання моделей та коректного порівняння їхньої якості на однакових даних.

2.4. Метрика оцінювання та протокол вимірювання якості

Оцінювання якості моделі сегментації є важливим етапом експериментального дослідження, оскільки сама функція втрат не завжди повністю відображає практичну якість отриманої маски. Для задачі сегментації серцевої області необхідно оцінювати не лише загальну правильність класифікації пікселів, а й точність виділення ROI, якість відтворення межі серця та здатність моделі розрізняти клінічні класи всередині серцевої області [29, 63, 91–93].

У роботі використовується багатокласова постановка задачі, тому оцінювання виконується на двох рівнях. Перший рівень пов'язаний із якістю виділення серцевої області відносно фону. Для цього всі шість клінічних класів об'єднуються в ROI-представлення. Другий рівень пов'язаний із якістю багатокласової сегментації, тобто з правильністю віднесення пікселів серцевої області до відповідних клінічних класів.

Основна еталонна маска має вигляд

$$Y \in \{0,1,2,3,4,5,6\}^{H \times W}. \quad (2.52)$$

де Y — еталонна багатокласова маска; H — висота зображення; W — ширина зображення; значення 0–5 відповідають клінічним класам серцевої області, а значення 6 відповідає класу *background*.

Прогнозована моделлю карта сегментації позначається як

$$\hat{Y} \in \{0,1,2,3,4,5,6\}^{H \times W}. \quad (2.53)$$

де \hat{Y} — прогнозована багатокласова маска; \hat{Y}_{ij} — прогнозована мітка класу для пікселя з координатами (i, j) .

Формули (2.52)–(2.53) показують, що і еталонна, і прогнозована маски мають однакову структуру та однаковий просторовий розмір. Це дозволяє виконувати покрокове порівняння масок на рівні окремих пікселів.

Для оцінювання якості виділення серцевої області з багатокласової маски формується ROI-представлення. Еталонна ROI-карта визначається так:

$$Y_{ij}^{ROI} = \begin{cases} 1, & \text{якщо } Y_{ij} \in \{0,1,2,3,4,5\}, \\ 0, & \text{якщо } Y_{ij} = 6. \end{cases} \quad (2.54)$$

де Y_{ij}^{ROI} — значення еталонної ROI-карти для пікселя (i, j) ; значення 1 означає, що піксель належить до серцевої області; значення 0 означає, що піксель належить до фону.

Аналогічно прогнозована ROI-карта визначається як

$$\hat{Y}_{ij}^{ROI} = \begin{cases} 1, & \text{якщо } \hat{Y}_{ij} \in \{0,1,2,3,4,5\}, \\ 0, & \text{якщо } \hat{Y}_{ij} = 6. \end{cases} \quad (2.55)$$

де \hat{Y}_{ij}^{ROI} — значення прогнозованої ROI-карти; \hat{Y}_{ij} — прогнозована багатокласова мітка пікселя.

Формули (2.54)–(2.55) не змінюють основну постановку задачі на бінарну. Вони лише задають допоміжне представлення, необхідне для обчислення ROI-метрик. Такий підхід дозволяє окремо оцінити, наскільки точно модель відокремлює серцеву область від фону.

Однією з основних метрик для оцінювання збігу еталонної та прогнозованої ROI є коефіцієнт Dice. Для серцевої області він визначається як

$$Dice_{ROI} = \frac{2|R \cap \hat{R}|}{|R| + |\hat{R}|}. \quad (2.56)$$

де $Dice_{ROI}$ — коефіцієнт Dice для серцевої області; R — еталонна ROI; \hat{R} — прогнозована ROI; $|R|$ — кількість пікселів в еталонній серцевій області; $|\hat{R}|$ — кількість пікселів у прогнозованій серцевій області; $|R \cap \hat{R}|$ — кількість пікселів, які одночасно належать до еталонної та прогнозованої ROI.

Іншою важливою метрикою є Intersection over Union, або IoU. Для ROI вона визначається так:

$$IoU_{ROI} = \frac{|R \cap \hat{R}|}{|R \cup \hat{R}|}. \quad (2.57)$$

де IoU_{ROI} — коефіцієнт перетину над об'єднанням для серцевої області;
 $R \cap \hat{R}$ — перетин еталонної та прогнозованої ROI; $R \cup \hat{R}$ — об'єднання еталонної та прогнозованої ROI.

Для задачі сегментації серця важливо аналізувати не лише збіг ROI, але й типи помилок. Хибнопозитивні пікселі визначаються як фонові пікселі, які модель помилково включила до серцевої області:

$$FP = \{ (i, j) \mid \hat{Y}_{ij} \neq 6, Y_{ij} = 6 \}. \quad (2.58)$$

де FP — множина хибнопозитивних пікселів; $Y_{ij} = 6$ означає, що піксель в еталонній масці належить до фону; $\hat{Y}_{ij} \neq 6$ означає, що модель віднесла його до одного з клінічних класів.

Хибнонегативні пікселі визначаються як пікселі серцевої області, які модель помилково віднесла до фону:

$$FN = \{ (i, j) \mid \hat{Y}_{ij} = 6, Y_{ij} \neq 6 \}. \quad (2.59)$$

де FN — множина хибнонегативних пікселів; $Y_{ij} \neq 6$ означає, що піксель належить до серцевої області; $\hat{Y}_{ij} = 6$ означає, що модель помилково віднесла його до фонового класу.

Помилки типу FP призводять до включення сторонніх фонових структур у серцеву ROI. Помилки типу FN , навпаки, спричиняють втрату частини серцевої області. Обидва типи помилок є критичними, оскільки змінюють форму, площу та контур серцевої ROI.

Для оцінювання точності саме на пікселях серцевої області використовується показник Acc_{heart} . Він обчислюється тільки для тих пікселів, які в еталонній масці належать до одного з клінічних класів:

$$Acc_{heart} = \frac{\sum_{(i,j) \in \Omega_{heart}} \mathbf{1}(\hat{Y}_{ij} = Y_{ij})}{|\Omega_{heart}|}. \quad (2.60)$$

де Acc_{heart} — точність класифікації пікселів серцевої області; Ω_{heart} — множина еталонних пікселів серцевої області; $\mathbf{1}(\cdot)$ — індикаторна функція; \hat{Y}_{ij} — прогнозована мітка пікселя; Y_{ij} — еталонна мітка пікселя; $|\Omega_{heart}|$ — кількість пікселів серцевої області.

Множина пікселів серцевої області визначається як

$$\Omega_{heart} = \{ (i, j) \mid Y_{ij} \neq 6 \}. \quad (2.61)$$

де Ω_{heart} — множина всіх пікселів, які в еталонній масці належать до одного з шести клінічних класів.

Метрика Acc_{heart} є важливою, оскільки звичайна піксельна точність може бути завищеною через домінування фону. На СХР-зображеннях фон зазвичай займає значно більшу площу, ніж серцева область. Тому модель може отримати високу загальну точність навіть тоді, коли якість сегментації серця є недостатньою. Використання Acc_{heart} дозволяє оцінити саме якість класифікації пікселів ROI.

Для багатокласової оцінки якості сегментації використовується Dice для кожного класу. Для класу k він визначається як

$$Dice_k = \frac{2|R_k \cap \hat{R}_k|}{|R_k| + |\hat{R}_k|}. \quad (2.62)$$

де $Dice_k$ — коефіцієнт Dice для класу k ; R_k — множина пікселів еталонної маски, що належать до класу k ; \hat{R}_k — множина пікселів прогнозованої маски, що належать до класу k .

Множини R_k та \hat{R}_k визначаються так:

$$R_k = \{ (i, j) \mid Y_{ij} = k \}, \quad \hat{R}_k = \{ (i, j) \mid \hat{Y}_{ij} = k \}. \quad (2.63)$$

де k — індекс класу; Y_{ij} — еталонна мітка; \hat{Y}_{ij} — прогнозована мітка.

Формули (2.62)–(2.63) дозволяють оцінювати якість сегментації окремо для кожного клінічного класу. Це важливо, оскільки модель може добре сегментувати один клас і водночас мати нижчу якість для іншого.

Середній Dice за клінічними класами обчислюється як

$$mDice = \frac{1}{6} \sum_{k=0}^5 Dice_k. \quad (2.64)$$

де $mDice$ — середнє значення Dice за шістьма клінічними класами; $Dice_k$ — Dice для класу k ; $k \in \{0, 1, \dots, 5\}$ — індекси клінічних класів серцевої області.

Аналогічно для кожного класу можна обчислити IoU:

$$IoU_k = \frac{|R_k \cap \hat{R}_k|}{|R_k \cup \hat{R}_k|}. \quad (2.65)$$

де IoU_k — значення IoU для класу k ; R_k — еталонна область класу k ; \hat{R}_k — прогнозована область класу k .

Середнє значення IoU за клінічними класами визначається як

$$mIoU = \frac{1}{6} \sum_{k=0}^5 IoU_k. \quad (2.66)$$

де $mIoU$ — середнє значення IoU за шістьма клінічними класами; IoU_k — IoU для класу k .

Метрики $mDice$ та $mIoU$ використовуються для оцінювання багатокласової якості сегментації. Вони показують, наскільки добре модель не лише виділяє серцеву область, а й розрізняє клінічні класи всередині неї.

Окремо може оцінюватися правильність визначення домінантного клінічного класу в межах ROI. Для цього визначається клас, який займає найбільшу площу серед не-фонових пікселів. Еталонний домінантний клас можна записати як

$$c_{true} = \arg \max_{k \in \{0,1,\dots,5\}} |R_k|. \quad (2.67)$$

де c_{true} — еталонний домінантний клінічний клас; R_k — множина пікселів еталонної маски, які належать до класу k .

Прогнозований домінантний клас визначається як

$$c_{pred} = \arg \max_{k \in \{0,1,\dots,5\}} |\hat{R}_k|. \quad (2.68)$$

де c_{pred} — прогнозований домінантний клінічний клас; \hat{R}_k — множина пікселів прогнозованої маски, які належать до класу k .

На основі цього можна визначити показник правильності домінантного класу:

$$Acc_{cls} = \mathbf{1}(c_{pred} = c_{true}). \quad (2.69)$$

де Acc_{cls} — індикатор правильності визначення домінантного класу для одного зображення; c_{pred} — прогнозований домінантний клас; c_{true} — еталонний домінантний клас.

Для набору зображень середнє значення цього показника визначається як:

$$\overline{Acc}_{cls} = \frac{1}{N} \sum_{n=1}^N \mathbf{1}(c_{pred}^{(n)} = c_{true}^{(n)}). \quad (2.70)$$

де \overline{Acc}_{cls} — середня точність визначення домінантного класу; N — кількість зображень у вибірці; $c_{pred}^{(n)}$ — прогнозований домінантний клас для n -го зображення; $c_{true}^{(n)}$ — еталонний домінантний клас для n -го зображення.

Цей показник не замінює метрики сегментації, але може використовуватися як додаткова характеристика того, чи правильно модель визначає переважний клінічний стан у межах серцевої області. Його доцільно розглядати разом із Dice, IoU та Acc_{heart} , оскільки коректне визначення клінічного класу залежить не лише від загальної якості маски, а й від того, наскільки точно виділена ROI охоплює інформативні ділянки серцевої області. У випадках, коли сегментаційна маска має незначні локальні похибки, але зберігає основну анатомічну структуру серця, класифікаційний висновок може залишатися правильним. Водночас систематичні помилки локалізації, зокрема включення значних фонових ділянок або втрата частини ROI, можуть знижувати стійкість такого висновку.

Для оцінювання якості роботи моделей у дисертаційному дослідженні використано систему метрик, яка дозволяє врахувати не лише загальну точність сегментації, а й якість виділення серцевої області, точність багатокласового поділу пікселів та правильність визначення переважного клінічного стану в межах ROI. Узагальнений перелік використаних метрик наведено в табл. 2.1.

Таблиця 2.1 — Метрики оцінювання якості сегментації

| Метрика | Що оцінює | Особливість використання |
|---------------|-------------------------------------------------|-----------------------------------------------------|
| Acc_{heart} | точність класифікації пікселів серцевої області | не враховує домінування фону |
| $Dice_{ROI}$ | збіг еталонної та прогнозованої серцевої ROI | оцінює якість виділення серця |
| IoU_{ROI} | перетин ROI відносно об'єднання | більш сувора метрика порівняно з $Dice$ |
| $mDice$ | середній $Dice$ за клінічними класами | оцінює багатокласову якість сегментації |
| $mIoU$ | середній IoU за клінічними класами | оцінює якість за класами без домінування фону |
| Acc_{cls} | правильність домінантного клінічного класу | додаткова характеристика класифікації всередині ROI |

Використання кількох метрик є необхідним, оскільки одна метрика не може повністю описати якість сегментації. Наприклад, високе значення $Dice_{ROI}$ може свідчити про якісне виділення серцевої області, але не гарантує правильної класифікації пікселів усередині ROI. Натомість $mDice$ та $mIoU$ дозволяють оцінити саме багатокласову якість. Метрика Acc_{heart} є корисною для контролю точності на серцевих пікселях без впливу великої кількості фонових пікселів.

Таким чином, система метрик у роботі побудована так, щоб оцінювати модель з різних боків: якість виділення серцевої області, точність відтворення ROI, правильність багатокласової сегментації та здатність моделі визначати клінічний клас у межах серцевої області. Такий підхід забезпечує більш об'єктивну оцінку якості запропонованої архітектури та базових моделей порівняння.

2.5 Вимоги до методу та узагальнення методики

На основі розглянутої постановки задачі, властивостей набору даних, правил передоброби, аугментацій і протоколу оцінювання можна сформулювати основні вимоги до методу сегментації серцевої ROI на рентгенівських знімках грудної клітки. Такі вимоги впливають не лише з формальної постановки задачі піксельної класифікації, а й з практичних особливостей CXR-зображень, для яких характерні слабкий контраст межі серця, перекриття анатомічних структур, варіативність умов знімання, шум, артефакти та доменний зсув. У цих умовах якість результату визначається не лише середнім значенням метрик, а насамперед характером помилок на межі ROI та їх впливом на подальший аналіз серцевої області [15–23, 32, 33, 42].

Першою важливою вимогою до методу є стабільність відтворення межі ROI та контроль помилок типу FP/FN. Для задачі сегментації серця на CXR-зображеннях критичне значення мають саме прикордонні ділянки, де найчастіше виникає змішування серцевої тіні з ребрами, діафрагмою, легенеvim рисунком або середостінням. Унаслідок цього модель повинна не лише виявляти серцеву область у цілому, а й забезпечувати морфологічно узгоджене виділення її контуру. Надмірне розширення ROI призводить до появи FP, тобто включення сторонніх структур у область інтересу, тоді як надмірне звуження межі породжує FN і втрату частини серцевого силуету. Отже, метод має бути орієнтований на зменшення саме таких помилок, а не лише на підвищення загальної піксельної точності.

Другою вимогою є стійкість методу до шуму, артефактів і доменного зсуву. У реальних умовах рентгенівські знімки можуть суттєво відрізнятися за яскравістю, контрастом, рівнем шуму, наявністю службових позначок, артефактів чи особливостями постоброби, що зумовлено використанням різних апаратів і протоколів знімання. За таких умов сегментаційна модель повинна зберігати здатність до надійного виділення серцевої області навіть у разі погіршення візуальної якості або зміни статистичних характеристик

вхідних даних. Це означає, що метод має бути робастним до помірних варіацій умов знімання та не повинен втрачати стійкість у присутності локальних спотворень або доменного зсуву.

Третьою важливою вимогою є відтворюваність експериментального протоколу та коректність порівняння моделей. Для того щоб результати дослідження були достовірними, усі моделі повинні навчатися і оцінюватися за однакових умов: на тих самих підвбірках, із тими самими правилами передобробки, аугментацій, функцій втрат і метрик. Це особливо важливо у задачах медичної сегментації, де незначні зміни в протоколі підготовки даних можуть суттєво впливати на фінальні результати. Отже, методика повинна забезпечувати не лише високу якість сегментації, а й коректність її кількісної оцінки, що дозволяє об'єктивно порівнювати різні архітектури між собою.

Узагальнюючи наведене, можна стверджувати, що ефективний метод сегментації серцевої ROI на CXR-зображеннях повинен одночасно задовольняти кілька взаємопов'язаних вимог: забезпечувати точне виділення межі серцевої області, зменшувати FP/FN-помилки, зберігати стійкість до шуму, артефактів і змін умов знімання, а також працювати в межах відтворюваного й узгодженого експериментального протоколу. Саме ці вимоги формують методичну основу для подальшого розроблення спеціалізованої архітектури сегментації та її експериментального порівняння з базовими моделями. У межах запропонованої методики машинне навчання використовується як основа для побудови моделі, здатної виявляти закономірності у підготовлених CXR-даних і узагальнювати їх на незалежній тестовій вибірці.

Таким чином, у другому розділі сформовано постановку задачі сегментації серцевої ROI на рентгенівських знімках грудної клітки, описано набір даних, принципи розмітки, передобробки та аугментацій, а також визначено метрики й протокол оцінювання. Отримані результати створюють необхідне методичне підґрунтя для подальшого опису архітектури запропонованої моделі та її компонентів, що розглядаються в наступному

розділі.

2.6 Висновки до розділу 2

У другому розділі сформульовано постановку задачі сегментації області інтересу на рентгенівських знімках грудної клітки в контексті побудови системи підтримки прийняття рішень для виявлення ішемічної хвороби серця. Задачу подано як багатокласову піксельну класифікацію з подальшим формуванням бінарної ROI-маски серцевої області, що дозволяє поєднати семантичну структуру розмітки з практично інтерпретованим результатом виділення області інтересу.

У розділі визначено формат вхідних даних, цільової розмітки та виходу сегментаційної моделі, описано правило побудови прогнозованої маски та формування ROI, а також наведено комбіновану функцію втрат, що враховує дисбаланс між класами та орієнтує навчання на якісне просторове перекриття області інтересу. Окремо окреслено основні обмеження постановки, пов'язані з проєкційною природою CXR-зображень, слабким контрастом межі серця, перекриттям анатомічних структур, наявністю шуму, артефактів і доменного зсуву.

Наведено опис набору даних, принципів організації розмітки, формування навчальної, валідаційної та тестової підвибірок, а також загальні правила передобробки й аугментацій. Показано, що коректна підготовка даних, збереження геометричної відповідності між зображенням і маскою та використання узгодженого експериментального протоколу є необхідними умовами для достовірного порівняння моделей сегментації.

У розділі також визначено систему метрик оцінювання якості, яка включає піксельну точність, коефіцієнт Dice та індекс Жаккара для ROI, а також їх багатокласові узагальнення. Обґрунтовано використання $Dice_{ROI}$ як основної метрики вибору найкращого чекпойнта, оскільки саме вона найбільш безпосередньо відображає якість виділення серцевої області як цільового результату сегментації.

Таким чином, другий розділ формує методичне підґрунтя дослідження, визначає правила підготовки даних, навчання та оцінювання моделей, а також встановлює вимоги до методу сегментації з позицій точності межі, контролю FP/FN, робастності та відтворюваності. Отримані результати створюють основу для подальшого розроблення та опису архітектури спеціалізованої моделі сегментації, що розглядається в наступному розділі.

РОЗДІЛ 3. РОЗРОБЛЕННЯ НЕЙРОННОЇ АРХІТЕКТУРИ CARDIOEDGEBIOMORPHNET ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВЗАЄМОДІЇ

У даному розділі представлено розроблення нейронної архітектури CardioEdgeBioMorphNet, яка є центральним технічним компонентом запропонованого підходу до аналізу рентгенівських знімків органів грудної клітки у задачі виділення серцевої області. Отримана маска ROI використовується не лише як проміжний результат, а як основа для подальшого морфологічного аналізу серцевої тіні та порівняння інтерпретованих морфометричних характеристик між клінічними станами а саме, площа, периметр, контур, форма, співвідношення розмірів. Відповідно, вимоги до якості сегментації визначаються не тільки усередненими значеннями метрик, а насамперед локалізацією та характером помилок, зокрема у прикордонній зоні, де помилки безпосередньо впливають на морфометрію.

Специфіка CXR-постановки полягає в тому, що межа серця не є різкою та однозначною: контур серцевої тіні часто слабкоконтрастний, частково перекритий ребрами, діафрагмою та судинним рисунком, а також залежить від позиціонування пацієнта, фази дихання, експозиції й шуму. За таких умов навіть невелике систематичне зміщення контуру на кілька пікселів призводить до стабільної похибки морфометричних оцінок, тоді як поодинокі фонові включення формують острівці, що спотворюють площу та периметр і ускладнюють подальшу інтерпретацію. Таким чином, для практично корисної сегментації серця на CXR ключовими є три групи вимог:

- стійкість модель має відтворювати узгоджену ROI та маску за допустимих варіацій знімка, щоб морфометричні показники не змінювалися через технічні фактори;
- якість межі контур має бути геометрично коректним у прикордонній зоні, яка визначає похибку форми, периметра та співвідношень розмірів;

– контроль хибнопозитивних включень у фоні: модель повинна уникати розповзання маски на конкурентні структури а саме ребра, діафрагма, судини, навіть за локальної схожості їхніх контурів із межею серця.

Зазначені вимоги визначили логіку побудови CardioEdgeBioMorphNet як розширеної encoder–decoder архітектури зі спеціалізованими механізмами, спрямованими на роботу з межею та геометричною структурою об'єкта.

По-перше, введено edge-гілку, що формує явний контурний сигнал на основі градієнтних операторів типу Sobel і підводить його до декодера як додаткову структурну ознаку, яка підсилює відновлення переходу серце, фон.

По-друге, застосовано морфологічні модулі у диференційовній формі, що дозволяють моделювати не лише локальні текстурні патерни, а й геометричні властивості — зв'язність, розширення, звуження, контурні перепади та пригнічення дрібних артефактів.

По-третє, реалізовано механізм Adaptive Boundary Rejection (ABR), який забезпечує контроль невизначеності: замість примусового призначення кожному пікселю одного з цільових класів ABR дозволяє віднести недостатньо впевнені ділянки до фону за адаптивним порогом τ , прогнозованим для конкретного знімка. Такий підхід є методично обґрунтованим для CXR, де частина прикордонних пікселів і зон накладань є апіорі неоднозначною, а зменшення FP є критичним для коректної морфометрії.

Важливо, що CardioEdgeBioMorphNet проектується як інженерно обґрунтована система, у якій кожний модуль має чітке призначення, узгоджене з критеріями якості, сформульованими в розділі 2: edge-гілка та EdgeMorphBlock спрямовані на підвищення точності контуру; SoftMorphPool2D і механізм багатомасштабної морфологічної уваги DMGGA — на підсилення морфологічно значущих структур на різних масштабах та ABR — на контроль хибнопозитивних включень у фоні та стабілізацію виходу. У підсумку архітектура орієнтована не лише на максимізацію метрик, а на формування стабільної, зв'язної та інтерпретованої ROI/маски, придатної

для подальших вимірювань і порівняння між клінічними станами.

Окремий компонент розділу присвячено програмному забезпеченню та інтерфейсу взаємодії з моделлю, яке забезпечує типовий прикладний сценарій: завантаження CXR-знімка, виконання інференсу, формування маски ROI та overlay-візуалізацій, відображення показників впевненості, параметрів ABR (зокрема використаного порогу τ) та підсумкового висновку. Це дозволяє проводити апробацію моделі на реальних прикладах і забезпечує інтерпретоване представлення результатів для подальшого аналізу.

Далі в розділі послідовно розглянуто: загальну схему CardioEdgeBioMorphNet а саме encoder–decoder, edge-гілка, ABR, спосіб формування edge-сигналу, базові будівельні блоки, структуру EdgeMorphBlock, диференційовні морфологічні операції SoftMorphPool2D, механізм DMGGA, формалізацію ABR, а також питання функцій втрат, стратегії навчання, обчислювальної складності та реалізації інтерфейсу взаємодії.

3.1. Загальна схема CardioEdgeBioMorphNet

CardioEdgeBioMorphNet спроектовано як система нейронної мережі, що формує стабільну, зв'язну та інтерпретовану ROI/маску серця, придатну для подальших морфологічних вимірювань. Для CXR це принципово, оскільки домінуючі помилки сегментації зосереджені у двох зонах, межа серця слабкий контраст і накладання структур та фон із конкурентними контурами, які провокують хибнопозитивні включення (FP). Відповідно, загальна схема CardioEdgeBioMorphNet поєднує три взаємодоповнювальні компоненти:

- encoder–decoder для багатомасштабного аналізу (контекст + локальні деталі);
- edge-гілку для явного сигналу межі;
- механізм Adaptive Boundary Rejection для контролю невпевнених пікселів і пригнічення FP у фоні.

Структурну організацію моделі наведено на рис. 3.1. Обробка

починається з блоку Stem (Conv 3×3 + Norm + SiLU), який переводить вхідний CXR у простір ознак і формує первинні контурно-текстурні представлення. Далі ознаки проходять через послідовність рівнів енкодера E1–E4. Кожний рівень містить два блоки EdgeMorphBlock, а після них застосовується Down (згортка зі зменшенням роздільності), що поетапно зменшує просторовий розмір та нарощує семантичну інформативність ознак. Такий каскад забезпечує одночасно захоплення глобального контексту серцевої тіні на грубих масштабах і збереження локальних деталей, необхідних для коректного відновлення контуру.

У центральній частині (bottleneck) використано багатомасштабний контекстний модуль ASPP, який агрегує інформацію з різними дилатаціями та розширює ефективне поле зору. Це корисно у зонах слабкої межі та неоднозначних накладань, де локальні ознаки недостатні для надійного рішення. Далі активується декодер D1–D4, який поетапно підвищує просторову роздільність (Upsample $\times 2$ + Conv) та відновлює форму ROI на вихідному масштабі.

Паралельно основному потоку на вході формується edge-карта за допомогою фіксованого оператора Sobel. Її призначення — забезпечити модель стабільним індикатором локальних переходів інтенсивності, найбільш інформативним саме в прикордонній зоні. Як показано на рис. 3.1, edge-сигнал інтегрується з виходом декодера через блок Edge Fuse (конкатенація ознак декодера та edge-карти з подальшою згорткою Conv+Norm+SiLU). Така інтеграція дозволяє узгоджувати відновлену маску з контурною “підказкою”, зменшуючи систематичне зміщення межі та схильність до “прилипання” на конкурентні контури.

На виході CardioEdgeBioMorphNet використано дві голови. Segmentation head формує карту логітів/класів для піксельної сегментації серцевої області відповідно до прийнятої системи міток. Threshold head (ABR head) формує параметр адаптивного порога, який використовується механізмом ABR для уточнення вихідної маски шляхом пригнічення невпевнених включень у фоні.

Докладне правило роботи ABR, спосіб отримання порога та процедура формування уточненої маски наведені в окремому підпункті, присвяченому механізму ABR.

Таким чином, схема на рис. 3.1 відображає ключову ідею CardioEdgeBioMorphNet: підсилення межі за рахунок явного edge-сигналу та контроль хибнопозитивних включень через адаптивну обережність у фоні. У підрозділах 3.1.1–3.1.8 розглянуто формалізацію вхідних/вихідних даних, принципи інтеграції edge-інформації, опис базових блоків і модулів, а також окремо — механізм ABR.

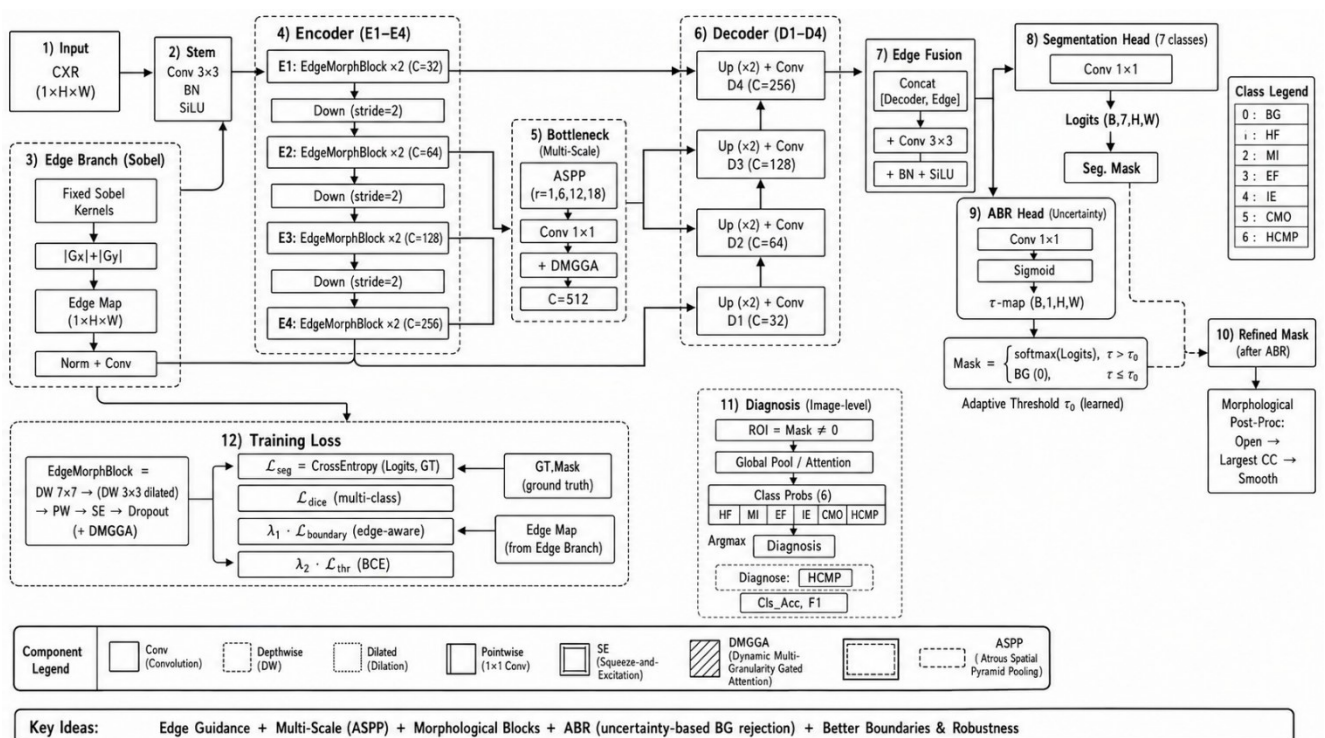


Рисунок. 3.1 — Узагальнена схема CardioEdgeBioMorphNet: основний encoder–decoder (Stem, E1–E4, bottleneck з ASPP, D1–D4), паралельна Sobel edge-branch, блок інтеграції Edge Fuse та дві вихідні голови (segmentation head і ABR threshold head) для формування ROI/маски серця з підвищеною точністю контуру та контролем хибнопозитивних включень у фоні

CardioEdgeBioMorphNet належить до класу спеціалізованих нейронних мереж для семантичної сегментації, у яких поєднано encoder–decoder структуру, морфологічні блоки, механізми уваги та контурно-орієнтовані

КОМПОНЕНТИ.

Нехай вхідне зображення

$$X \in \mathbb{R}^{H \times W \times 1}, \quad (3.1)$$

де X — вхідне СХR-зображення (1 канал), H, W — висота та ширина.

Модель повертає дві величини: логіти сегментації Z та logit порогу t :

$$(Z, t) = f_{\theta}(X), Z \in \mathbb{R}^{K \times H \times W}, t \in \mathbb{R}, \quad (3.2)$$

де $f_{\theta}(\cdot)$ — нейронна мережа з параметрами θ , Z — логіти для K серцевих класів (у реалізації $K = 6$), t — вихід порогової голови (threshold head) для АВR.

Навіщо t : щоб модель могла адаптувати жорсткість відкидання в ВG залежно від якості/контрасту конкретного знімка (наприклад, при шумі або слабкому контурі поріг може бути нижчим/вищим).

Енкодер виділяє ознаки на кількох масштабах:

$$\{F^{(s)}\}_{s=1}^S = \text{Enc}_{\theta}(X), F^{(s)} \in \mathbb{R}^{H_s \times W_s \times C_s}, \quad (3.3)$$

де $F^{(s)}$ — карта ознак на рівні s , H_s, W_s — роздільність на цьому рівні, C_s — кількість каналів, S — кількість масштабів.

Типовий крок енкодера можна описати як “локальна обробка + зменшення роздільності”:

$$F^{(s+1)} = \text{Down}\left(\text{Block}(F^{(s)})\right), s = 1, \dots, S - 1. \quad (3.4)$$

де $\text{Block}(\cdot)$ — базовий блок витягання ознак (у моделі — EdgeMorphBlock , див. 3.4), $\text{Down}(\cdot)$ — downsampling (зменшення роздільності, stride=2).

Багатомасштабність потрібна:

- на грубих рівнях змерже бачить контекст (загальна форма серця, положення відносно легенів);
 - на дрібних рівнях — деталі межі (локальні переходи, дрібні вигини).
- Це прямо зменшує типову СХR-проблему: коли модель ведеться на текстуру ребер і втрачає цілісну форму.

Декодер у реалізації виконує поетапне відновлення просторової

роздільності за рахунок послідовності апсемплінг-блоків:

$$U_1 = \text{Up}_1(B), U_2 = \text{Up}_2(U_1), U_3 = \text{Up}_3(U_2), U_4 = \text{Up}_4(U_3). \quad (3.5)$$

де B — вихід bottleneck, $\text{Up}_s(\cdot)$ — білінійний апсемплінг 2×2 та 3×3 згортка з нормалізацією й активацією, U_4 — карта ознак на вихідній роздільності перед інтеграцією edge-сигналу.

На відміну від класичних U-Net, у даній реалізації не використано явні skip-конкатенації з енкодера. Компенсація деталізації досягається морфологічно-орієнтованими блоками в енкодері/горлі, що підсилюють структурні ознаки, та інтеграцією edge-сигналу через Edge Fuse, який підтягує фінальну межу до локальних градієнтів інтенсивності.

Окремо формується edge-карта:

$$E = \text{Edge}(X), E \in \mathbb{R}^{H \times W \times 1}. \quad (3.6)$$

де $\text{Edge}(\cdot)$ — детермінований оператор (у реалізації — Sobel-модуль, див. 3.2); E — карта сили градієнта, яка підкреслює потенційні межі.

Нащо edge-гілка:

- у CXR межа серця часто “плаває” через низький контраст;
- edge-карта дає мережі стабільний геометричний сигнал, незалежний від семантики класів;
- це зменшує помилки типу зсув контуру та прилипання до ребер.

Щоб edge-сигнал реально впливав на фінальну форму, він вводиться у декодер через злиття:

$$U_{\text{fuse}} = \phi(\text{Conv}_{3 \times 3}([U; E])). \quad (3.7)$$

де $[E]$ — конкатенація по каналах (ознаки декодера + edge-карта), $\text{Conv}_{3 \times 3}$ — згортка для інтеграції edge-інформації, $\phi(\cdot)$ — нелінійність, U_{fuse} — фінальні ознаки перед головами.

Навіщо саме так: конкатенація дозволяє моделі навчитися читати edge-карту як окреме джерело інформації, а $\text{Conv}_{3 \times 3}$ локально узгоджує, де edge є релевантною межею серця, а де — конкурентним контуром.

Сегментаційна голова формує логіти:

$$Z = \text{Head}_{seg}(U_{fuse}), Z \in \mathbb{R}^{K \times H \times W}, \quad (3.8)$$

де $\text{Head}_{seg}(\cdot)$ — вихідна голова (типово $1 \times 1 \text{conv}$), Z — логіти для K класів.

Ймовірності:

$$P_{k,i,j} = \frac{\exp(Z_{k,i,j})}{\sum_{r=0}^{K-1} \exp(Z_{r,i,j})}. \quad (3.9)$$

де $P_{k,i,j}$ — ймовірність класу k у пікселі (j) , $k \in \{0, \dots, K-1\}$.

Навіщо логіти + softmax: логіти дозволяють навчання через крос-ентропію, а softmax переводить оцінки у ймовірності, потрібні для механізму ABR (порогування впевненості).

Окрема голова прогнозує скаляр t :

$$t = \text{Head}_{thr}(U_{fuse}), t \in \mathbb{R}, \quad (3.10)$$

де $\text{Head}_{thr}(\cdot)$ — порогова голова (GAP + невеликий MLP/conv), t — logit, що задає поріг відкидання.

Далі поріг масштабується в межі:

$$\tau = \tau_{\min} + (\tau_{\max} - \tau_{\min}) \cdot \sigma(t), \sigma(t) = \frac{1}{1 + e^{-t}}. \quad (3.11)$$

де τ — адаптивний поріг для конкретного знімка, τ_{\min}, τ_{\max} — межі порога (задані в конфігурації), $\sigma(\cdot)$ — сигмоїда.

Різні CXR мають різну якість. Для “чистих” знімків можна ставити жорсткіший поріг і різко відсікати фон; для слабконтрастних — поріг має бути м’якшим, щоб не з’їсти серцеву область.[32]

Визначимо максимальну серцеву впевненість:

$$m_{i,j} = \max_{k \in \{0, \dots, K-1\}} P_{k,i,j}, \quad (3.12)$$

де $m_{i,j}$ — наскільки модель впевнена, що піксель (j) належить хоча б одному серцевому класу.

Фінальне правило:

$$\hat{Y}_{i,j} = \begin{cases} BG, & m_{i,j} < \tau, \\ \arg \max_k P_{k,i,j}, & m_{i,j} \geq \tau. \end{cases} \quad (3.13)$$

де $\hat{Y}_{i,j}$ — фінальна мітка, BG — індекс фонового класу, якщо впевненість нижча за τ , піксель відкидається у фон.

Навіщо ABR у CXR: у конкурентних зонах модель інколи змушена обирати клас (бо softmax завжди дає максимум). ABR додає керовану можливість сказати: я не впевнений — віднесу в фон. Це знижує FP, прибирає острівці та робить маску придатнішою для морфометрії. НААА Механізм Adaptive Boundary Rejection використовується для зменшення помилок у конкурентних зонах CXR-зображення, де модель має низьку впевненість у належності пікселя до певного класу. Оскільки softmax завжди обирає клас із максимальною ймовірністю, ABR дозволяє віднести невпевнені пікселі до фону. Це знижує кількість хибнопозитивних включень, прибирає дрібні ізольовані ділянки та робить ROI-маску придатнішою для подальшого морфометричного аналізу.

Отже, загальна схема CardioEdgeBioMorphNet реалізує три взаємодоповнювальні механізми, спрямовані на отримання стабільної та інтерпретованої ROI/маски серця на CXR:

- encoder–decoder (із багатомасштабною обробкою та відновленням роздільності) — забезпечує поєднання глобального контексту (узгоджена форма та положення серцевої тіні) і локальної деталізації (точні просторові межі), що знижує ризик втрати цілісності ROI та помилок, зумовлених неоднорідністю зображення [28, 46–49, 74, 75, 80];

- edge-сигнал (Sobel) і його інтеграція через Edge Fuse — надає явну геометричну підказку щодо можливих контурів, стабілізує відновлення межі у слабкоконтрастних ділянках та зменшує зсув контуру прилипання до конкурентних структур (ребра, діафрагма, судинний малюнок);

- ABR (Adaptive Boundary Rejection) — вводить механізм обережного рішення у невпевнених пікселях шляхом адаптивного порогоування впевненості, що дозволяє зменшити хибні спрацьовування у фоні та підвищує придатність отриманої маски для подальших морфологічних вимірювань.

Таким чином, архітектура в цілому орієнтована не лише на середню

точність сегментації, а на контроль критичних для CXR типів помилок — на межі та у фоні — що є необхідною передумовою коректної морфологічної інтерпретації результатів.

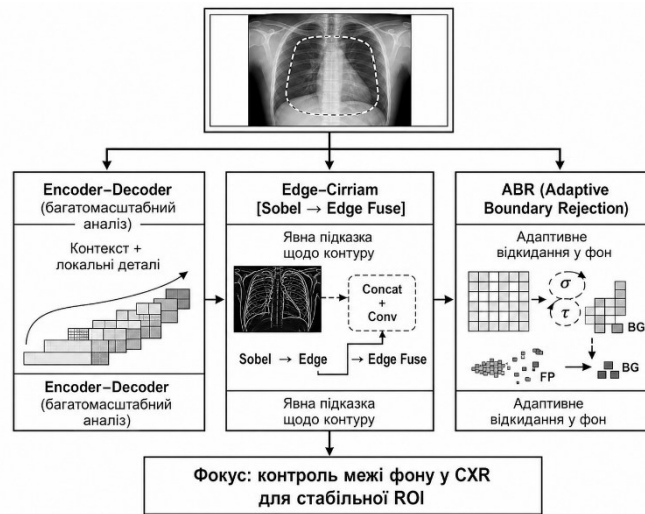


Рисунок. 3.2 — Підсумкова схема трьох ключових механізмів CardioEdgeBioMorphNet

Encoder–decoder забезпечує багатомасштабне поєднання контексту та локальних деталей; edge-сигнал (Sobel) після інтеграції через Edge Fuse підсилює відновлення межі; ABR (Adaptive Boundary Rejection) адаптивно відкидає невпевнені пікселі у фон (BG), зменшуючи хибні спрацьовування (FP) та підвищуючи стабільність ROI.

Як показано на рис. 3.2, модель спеціально орієнтована на контроль двох найбільш критичних джерел помилок у CXR-сегментації: неточностей на межі серця та хибних включень у фоні. Основний encoder–decoder формує узгоджену ROI завдяки багатомасштабним ознакам, що поєднують глобальний контекст (форма та положення серця) і локальні деталі (дрібні переходи інтенсивності). Edge-сигнал, отриманий детермінованим Sobel-оператором, вводиться як геометрична підказка й через Edge Fuse узгоджує фінальні ознаки з контурною інформацією, стабілізуючи відновлення межі. На завершальному етапі ABR використовує прогнозований поріг впевненості, щоб у невизначених пікселях обирати фон, що зменшує та робить маску придатнішою для подальших морфологічних вимірювань.

3.2. Формування edge-сигналу (Sobel/градієнт)

У CXR-сегментації серця основні похибки концентруються в прикордонній області, де контур серця має слабкий контраст та перекривається ребрами, судинним малюнком і діафрагмою. Тому в CardioEdgeBioMorphNet використано окремий детермінований модуль формування edge-сигналу E , який є геометричною підказкою для моделі. На відміну від навчених фільтрів, edge-сигнал обчислюється фіксованими ядрами Sobel (без оновлення ваг у процесі навчання), що робить його стабільним відносно варіацій якості знімків та забезпечує узгоджену шкалу значень завдяки нормалізації [2, 15–19, 27].

Нехай вхідне CXR-зображення (після приведення до одного каналу та розміру) задається як

$$X \in \mathbb{R}^{H \times W}, \quad (3.14)$$

де X — одноканальне CXR-зображення (матриця інтенсивностей); H, W — висота та ширина.

У роботі edge-модуль застосовується безпосередньо до X і не потребує додаткового попереднього згладжування, оскільки подальша інтеграція edge-карти виконується згортковим узгодженням у блоці Edge Fuse (див. 3.1).

Оператор Sobel оцінює просторові похідні в двох ортогональних напрямках. Для цього використовуються фіксовані ядра S_x та S_y :

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (3.15)$$

де S_x — ядро Sobel для похідної по осі x , S_y — ядро Sobel для похідної по осі y .

Компоненти градієнта визначаються згорткою:

$$G_x = S_x * X, G_y = S_y * X. \quad (3.16)$$

де G_x, G_y — карти похідних інтенсивності, $*$ — операція згортки.

Важливо, що S_x та S_y у реалізації збережені як незмінні тензори (не

параметри оптимізації), тому edge-сигнал є детермінованим та відтворюваним.

Для отримання скалярного показника “сили межі” формується модуль градієнта:

$$M = \sqrt{G_x^2 + G_y^2 + \varepsilon}. \quad (3.17)$$

де M — карта модуля градієнта (edge magnitude), ε — мале число для числової стабільності (щоб уникнути проблем при $G_x \approx 0, G_y \approx 0$).

Модуль M підкреслює пікселі з різким локальним перепадом інтенсивності, які відповідають потенційним межам структур на CXR (у тому числі контуру серця).

Щоб забезпечити узгоджену шкалу значень edge-карти між різними знімками, у реалізації застосовано нормалізацію діленням на максимум:

$$E = \frac{M}{\max_{i,j} M_{i,j} + \varepsilon}, E \in \mathbb{R}^{H \times W \times 1}. \quad (3.18)$$

де E — нормалізована edge-карта, $\max_{i,j} M_{i,j}$ — максимальне значення модуля градієнта по всіх пікселях, ε — мале число для уникнення ділення на нуль.

Така нормалізація зручна тим, що приводить E до стабільного діапазону (наближено [1]) без залежності від абсолютних рівнів яскравості та контрасту, що характерно для клінічних CXR.

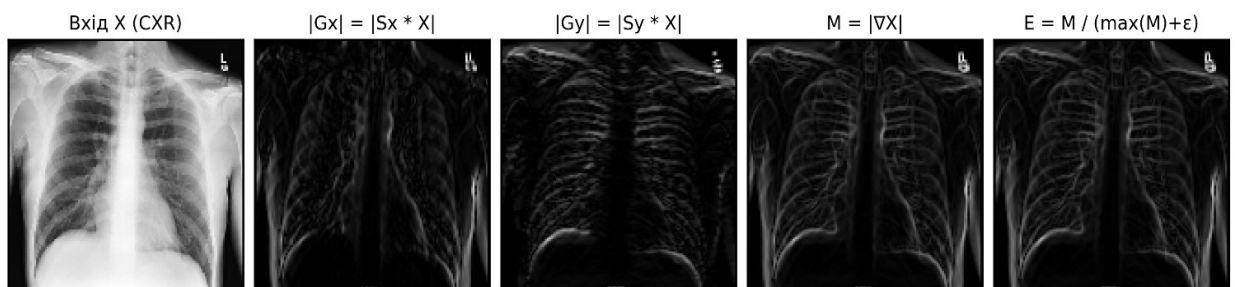


Рисунок. 3.3 — Формування edge-сигналу на основі оператора Sobel

Отриманий edge-сигнал E не є готовою маскою серця, оскільки містить також конкурентні контури (ребра, судини, діафрагма). Тому E використовується не напряму як сегментація, а як додатковий

інформаційний канал, який інтегрується з ознаками основного потоку в блоці Edge Fuse:

$$U_{\text{fuse}} = \phi(\text{Conv}_{3 \times 3}([U; E])). \quad (3.19)$$

де $[E]$ — конкатенація ознак декодера U та edge-карти E по каналах, $\text{Conv}_{3 \times 3}$ — згортка інтеграції (edge_fuse), $\phi(\cdot)$ — нелінійність, U_{fuse} — фінальні ознаки перед вихідними головами.

На практиці це означає, що модель навчається інтерпретувати edge-карту: підсилювати релевантні для серця контури та ігнорувати фонові градієнти, якщо вони не узгоджуються з глобальною формою ROI.

Таким чином, у CardioEdgeBioMorphNet edge-сигнал формується детермінованим способом на основі Sobel-градієнта з фіксованими ядрами та подальшою tanh -нормалізацією. Це надає моделі стабільний контурний індикатор, який підсилює відновлення межі серця після інтеграції через Edge Fuse, особливо у складних прикордонних ділянках CXR.

3.3. Базові блоки мережі (Down/Up, нормалізація, активації)

У реалізації CardioEdgeBioMorphNet більшість функціональних модулів побудовано за уніфікованою схемою Conv до Normalization до SiLU. Така організація забезпечує узгоджене формування ознак на різних рівнях мережі, спрощує процес оптимізації та підвищує стійкість моделі до варіативності CXR-зображень, зокрема змін контрастності, шумів, артефактів і відмінностей у режимах експозиції [65].

У подальшому використовується узгоджене позначення тензорів ознак:

$$x \in R^{C \times H \times W}, \quad (3.20)$$

де x — карта ознак, C — кількість каналів, H, W — просторові розміри.

Базовий “атомарний” блок обробки ознак задається як:

$$\text{Block}(x) = \phi(\text{Norm}(\text{Conv}(x))), \quad (3.21)$$

де $\text{Conv}(\cdot)$ — двовимірна згортка (зокрема 3×3 або 1×1), $\text{Norm}(\cdot)$ — нормалізація ознак (у даній роботі — GroupNorm), $\phi(\cdot)$ — нелінійність SiLU.

Використання нормалізації після згортки вирівнює масштаб активацій між каналами та просторовими ділянками, а SiLU забезпечує гладку нелінійність, корисну для задач, де критичною є точність просторових меж.

Для нормалізації ознак використано GroupNorm, що є практично доцільним при невеликих значеннях batch-size, характерних для медичних задач. Кількість груп g підбирається так, щоб вона ділила число каналів:

$$g(C) = \max\{ g \in \{8,4,2,1\} \mid C \bmod g = 0 \}, \quad (3.22)$$

де C — кількість каналів шару, $g(C)$ — кількість груп у GroupNorm (вибір із $\{1\}$ так, щоб C було кратним g).

Нормалізація в межах груп задається як:

$$\text{GN}(x) = \gamma \cdot \frac{x - \mu_g}{\sqrt{\sigma_g^2 + \varepsilon}} + \beta. \quad (3.23)$$

де x — вхідний тензор ознак, μ_g, σ_g^2 — середнє та дисперсія, обчислені в межах групи каналів (разом із просторовими координатами), γ, β — навчувані параметри масштабу й зсуву, ε — мале число для числової стабільності.

Застосування GroupNorm зменшує залежність від статистик батчу та підвищує відтворюваність результатів при змінній якості CXR, оскільки нормалізація визначається внутрішньою структурою ознак, а не випадковими властивостями конкретного батчу.

Як нелінійність використано SiLU (Swish):

$$\text{SiLU}(x) = x \cdot \sigma(x), \sigma(x) = \frac{1}{1 + e^{-x}}. \quad (3.24)$$

де x — вхідний тензор, $\sigma(\cdot)$ — сигмоїда.

SiLU є гладкою активацією, що сприяє стабільному поширенню градієнтів у глибших частинах мережі та, як правило, корисна для задач, чутливих до точності просторових меж (зокрема контуру серцевої тіні).

Операція зменшення просторової роздільності реалізується згорткою 3×3 зі stride 2, після якої застосовуються нормалізація та SiLU:

$$\text{Down}(x) = \text{SiLU} \left(\text{GN} \left(\text{Conv}_{3 \times 3, s=2}(x) \right) \right). \quad (3.25)$$

де $\text{Conv}_{3 \times 3, s=2}$ — згортка 3×3 зі stride 2 та padding 1 (зменшує H, W у 2 рази), $\text{GN}(\cdot)$ — GroupNorm за (3.23), $\text{SiLU}(\cdot)$ — активація за (3.24).

Зменшення роздільності (Down) збільшує ефективне рецептивне поле та дозволяє формувати більш контекстні ознаки на грубих масштабах.

У контексті CXR це важливо для узгодження глобальної форми та положення серцевої тіні відносно легневих структур і діафрагми. Технічно часто використовується згортка без зміщення ($\text{bias}=\text{False}$), оскільки зсув переважно моделюється параметром β у нормалізації (3.23).

Upsampling виконується у два етапи: білінійне масштабування 2×2 та подальша згортка 3×3 з нормалізацією й SiLU. Спочатку:

$$x^\uparrow = \text{Interp}(x; \times 2), \quad (3.26)$$

де $\text{Interp}(\cdot; \times 2)$ — білінійна інтерполяція у 2 рази.

Після цього:

$$\text{Up}(x) = \text{SiLU}(\text{GN}(\text{Conv}_{3 \times 3}(x^\uparrow))). \quad (3.27)$$

де $\text{Conv}_{3 \times 3}$ — згортка 3×3 (padding 1), $\text{GN}(\cdot)$ — GroupNorm, $\text{SiLU}(\cdot)$ — активація.

Білінійна інтерполяція забезпечує плавне відновлення геометрії без різких кроків, тоді як згортка після масштабування виконує локальне уточнення структури ROI на високій роздільності. Це є критичним для точного позиціонування межі перед інтеграцією контурної інформації у подальших блоках.

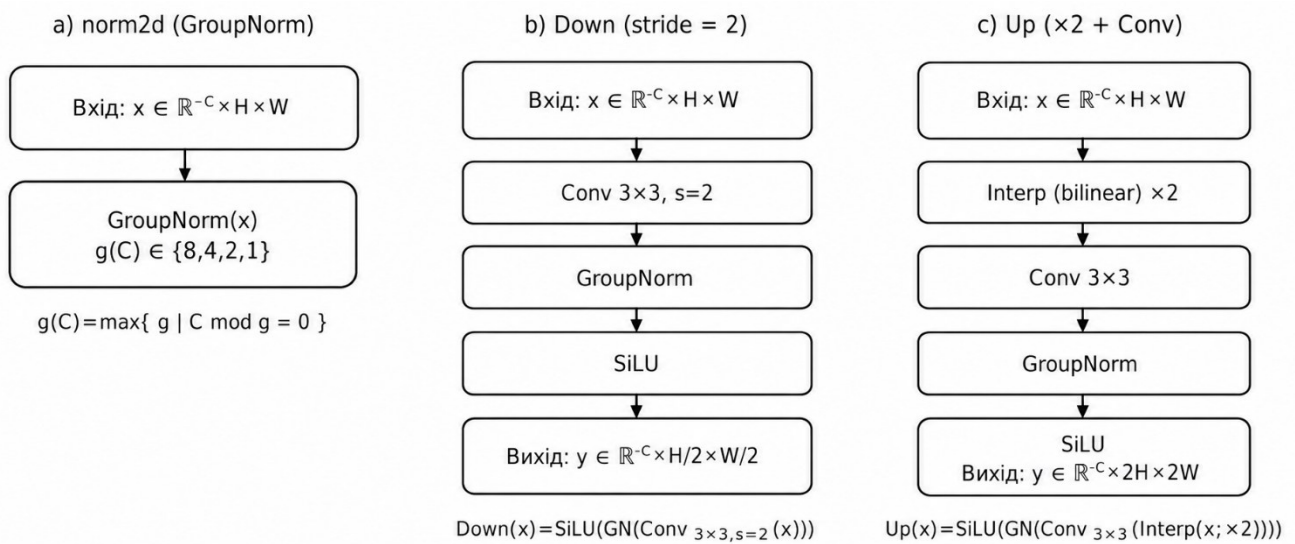


Рисунок. 3.4 — Базові блоки CardioEdgeBioMorphNet

Отже, базові блоки Down і Up реалізують керовану зміну масштабу карт ознак у CardioEdgeBioMorphNet: Down концентрує контекст на грубих рівнях, тоді як Up відновлює просторову деталізацію.

Комбінація GroupNorm + SiLU стабілізує розподіли активацій, покращує проходження градієнтів і підвищує відтворюваність поведінки моделі на CXR при варіативній якості знімків, створюючи надійний базис для спеціалізованих морфологічних модулів наступного підрозділу.

3.4. EdgeMorphBlock: структура, роль, рівняння

EdgeMorphBlock є базовим будівельним модулем CardioEdgeBioMorphNet, який поєднує локальне контурне уточнення, багатомасштабне (за рахунок дилатацій) “підхоплення” структури межі, каналну перевагу (SE) та залишкове (residual) з’єднання. Його ключова мета — зменшити типові для CXR помилки, а саме розмиття або зсув контуру серця та прилипання до конкурентних контурів у фоні [50, 51].

Нехай на вході блоку маємо карту ознак:

$$F \in R^{H \times W \times C}. \quad (3.28)$$

де F — вхідні ознаки, H, W — просторові розміри, C — кількість каналів.

EdgeMorphBlock реалізує перетворення $T(\cdot)$ та додає його до вхідного сигналу (residual-прохід):

$$F_{out} = F + T(F). \quad (3.29)$$

де F_{out} — вихід блоку, $T(\cdot)$ — послідовність внутрішніх операцій (контурні/морфологічні згортки, канална перевага, регуляризація).

Навіщо residual: зберігає корисну основу ознак, стабілізує градієнти та дозволяє блоку діяти як уточнювач, а не як повний перезапис представлення — це важливо для акуратної геометрії межі.

У блоці використовується підхід pre-norm: спочатку нормалізація й активація, а вже потім згортка. Узагальнено:

$$\Phi(F) = Conv(\varphi(GN(F))). \quad (3.30)$$

де $GN(\cdot)$ — GroupNorm; $\varphi(\cdot)$ — SiLU, Conv — відповідна згортка.

Pre-norm підвищує числову стабільність (особливо при варіативній якості CXR), зменшує ризик розбігу активацій і робить навчання рівнішим.

Перший крок — depthwise згортка 7×7 , яка підсилює локальні структури та контурні переходи без змішування каналів:

$$F_1 = DW_{7 \times 7}(\varphi(GN(F))). \quad (3.31)$$

де $DW_{7 \times 7}$ — depthwise-згортка (окрема згортка на кожний канал), GN — нормалізація, SiLU — активація.

7×7 : дає ширший локальний контекст навколо межі (на CXR контур серця часто слабкий і розтягнутий за інтенсивністю).

Далі застосовується набір depthwise-згорток 3×3 з різними дилатаціями, щоб одночасно бачити тонкі та грубі компоненти контуру:

$$F_2 = \sum_{d \in \{1,2,3\}} DW_{3 \times 3}^{(d)}(\varphi(GN(F_1))). \quad (3.32)$$

де $DW_{3 \times 3}^{(d)}$ — depthwise-згортка 3×3 з дилатацією d , $d \in \{1,2,3\}$ — набір масштабів, \sum — поканальна сума відповідей.

Межа серця має різний характер у різних ділянках (частково перекрита

ребрами або діафрагмою), тому корисно мати реакції і на ближні, і на трохи віддалені контурні підказки.

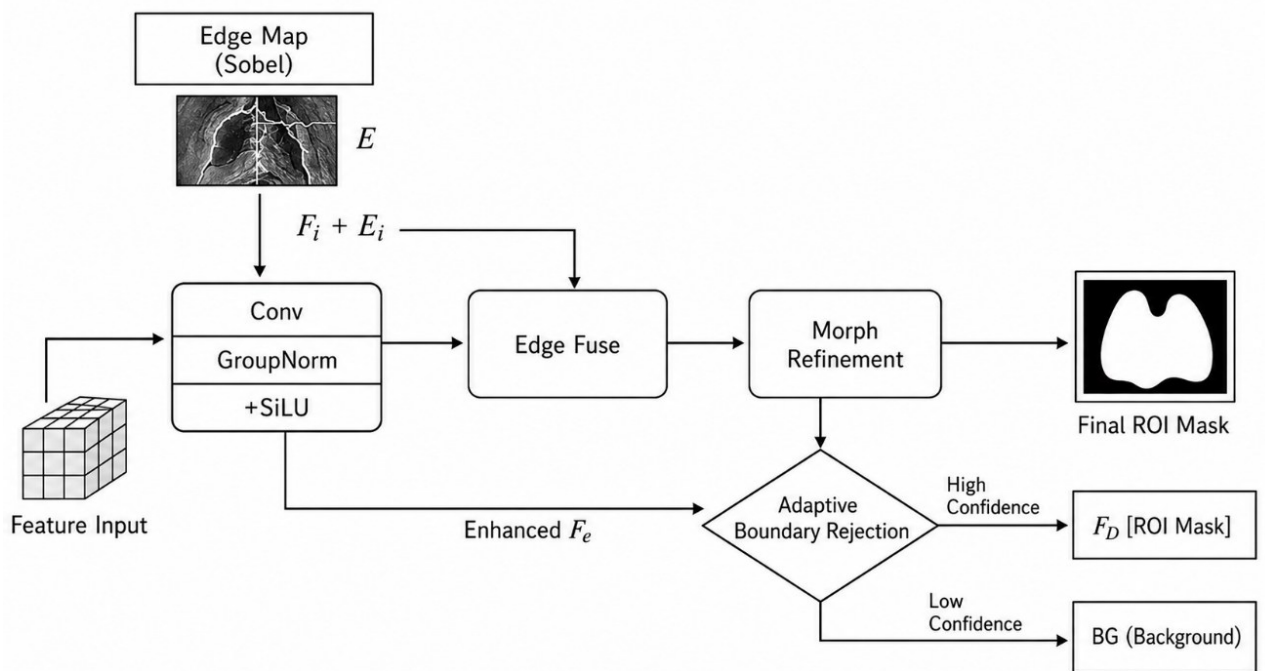


Рисунок 3.5 — Структура EdgeMorphBlock

На рис. 3.5 — Структура EdgeMorphBlock. Блок складається з попередньої нормалізації та активації (GN + SiLU), depthwise-згортки з розширеним рецептивним полем, багатодилатаційних гілок для аналізу локальних і напівглобальних структур, pointwise-згорток для змішування каналів, а також residual-зв'язку.

За потреби до блоку можуть підключатися механізми каналної (SE) та морфологічної уваги.

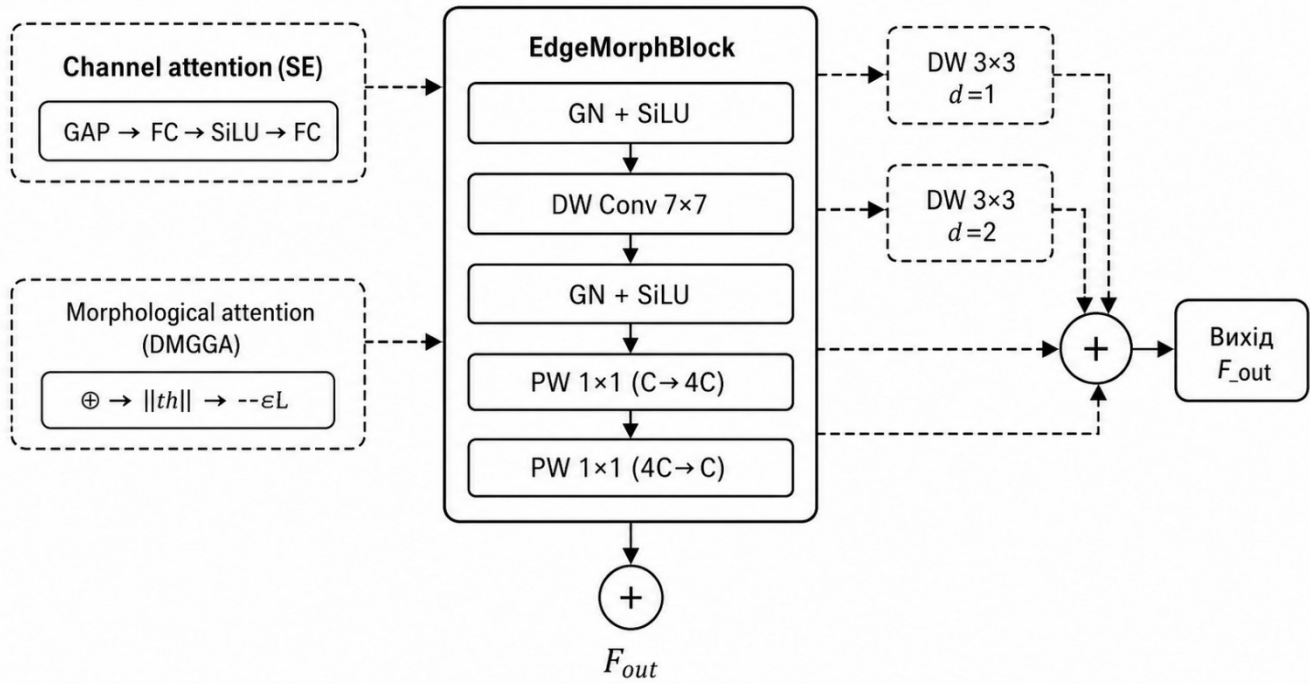


Рисунок. 3.6 — Опційні розширення EdgeMorphBlock

На рис. 3.6 Показано додаткові модулі, які можуть підключатися до базової структури: канална увага (SE) для адаптивного масштабування каналів, морфологічна увага DMGGA для підсилення структурно значущих елементів, а також багатодилатаційні depthwise-гілки, що дозволяють аналізувати межу на різних просторових масштабах.

Оскільки depthwise-згортки не змішують канали, далі виконується каналне змішування через 1×1 згортки: розширення до $4C$, регуляризація Dropout2D, та повернення до C :

$$F_3 = PW_{4C \rightarrow C}(Drop(\varphi(PW_{C \rightarrow 4C}(F_2)))). \quad (3.33)$$

де $PW_{C \rightarrow 4C}$ — pointwise 1×1 згортка (розширення каналів), $PW_{4C \rightarrow C}$ — pointwise 1×1 згортка (стиснення каналів), $Drop(\cdot)$ — Dropout2D (випадкове занулення частини каналів під час навчання).

Після цього застосовується нормалізація:

$$F_4 = GN(F_3). \quad (3.34)$$

де F_4 — нормалізовані ознаки після каналного змішування, depthwise-частина виділяє контурні патерни, а pointwise-частина вчиться поєднувати їх

у більш семантичні канали (що важливо для відокремлення серця від конкурентних структур).

Щоб підсилити інформативні канали та пригнітити шумові, використовується SE-перевага. Спочатку стискаємо простір глобальним усередненням:

$$s = GAP(F_4) \in R^{1 \times 1 \times C}. \quad (3.35)$$

де $GAP(\cdot)$ — глобальне усереднення по $H \times W$, s — вектор каналних описів.

Далі формуємо ваги каналів через двошарове перетворення та сигмоїду:

$$w = \sigma(W_2 \varphi(W_1 s)) \in (0,1)^{1 \times 1 \times C}. \quad (3.36)$$

де W_1, W_2 — вагові матриці (стиснення/розширення каналів), $\sigma(\cdot)$ — сигмоїда, w — каналні ваги.

Застосовуємо каналну перевагу:

$$F_5 = F_4 \odot w. \quad (3.37)$$

де \odot — поканалне множення (broadcast по $H \times W$).

На CXR частина каналів часто реагує на сторонні структури (ребра, судини). SE допомагає переорієнтувати блок на стабільні ознаки серцевої тіні.

На окремих рівнях мережі після SE може застосовуватись морфологічна увага (DMGGA), яка додатково підсвічує морфологічно значущі структури:

$$F_6 = DMGGA(F_5). \quad (3.38)$$

де $DMGGA(\cdot)$ — модуль морфологічної уваги (детально розглядається в підрозділі 3.6); F_6 — ознаки після уваги. Якщо DMGGA не використовується, то $F_6 \equiv F_5$.

Фінальний вихід з урахуванням residual-з'єднання:

$$F_{out} = F + F_6. \quad (3.39)$$

F_{out} — вихід EdgeMorphBlock, F — вхід, F_6 — уточнення (контур + мультидилатація + каналне змішування + SE + (опційно) DMGGA).

EdgeMorphBlock працює як локально-геометричний коректор, який:

- підсилює межу через широку depthwise-згортку та мультидилатовані фільтри;
- зменшує вплив випадкових текстур (ребра/судини) через SE-перевагу;
- зберігає стабільність та форму ROI завдяки residual-проходу і pre-norm схемі;
- підтримує узгодженість форми в умовах слабкого контрасту та накладань [15–20, 35–39].

3.5. Диференційовні морфологічні операції SoftMorphPool2D

Морфологічні операції (ерозія, дилатація, відкриття, закриття) традиційно використовуються для роботи з формою об'єкта: згладжування контуру, пригнічення дрібних артефактів, заповнення розривів та усунення острівців у фоні. Для задачі сегментації серцевої області на CXR ці властивості є принципово важливими, оскільки типові помилки моделі зосереджені на межі серця та у вигляді фонових хибних включень. Однак класична морфологія є недиференційовною (використовує max/min), тому напряду погано узгоджується з градієнтним навчанням нейронних мереж.

Щоб поєднати переваги морфологічної обробки з можливістю end-to-end оптимізації, застосовується диференційовна апроксимація морфологічних операторів, яку надалі позначимо як SoftMorphPool2D. Її ключова ідея — замінити жорсткі операції max/min на “м'які” аналоги, що є гладкими та керуються параметром температури [20, 67, 68, 85].

Нехай $X \in R^{H \times W}$ — карта ознак (або один канал тензора ознак). Розглянемо локальне вікно $\Omega_{i,j}$ розміру $k \times k$ з центром у пікселі (j) . М'який максимум у цьому вікні визначимо як:

$$\text{SoftMaxPool}_{\beta}(X)_{i,j} = \frac{\sum_{(u,v) \in \Omega_{i,j}} X_{u,v} \exp(\beta X_{u,v})}{\sum_{(u,v) \in \Omega_{i,j}} \exp(\beta X_{u,v})}. \quad (3.40)$$

де $\beta > 0$ — параметр “жорсткості” (температура), $\Omega_{i,j}$ — локальне вікно $k \times k$, $X_{u,v}$ — значення ознаки у пікселі (v) .

За $\beta \rightarrow \infty$ значення (3.40) наближається до $\max_{\Omega_{i,j}} X$, а за малих β стає більш усередненим, що робить оператор стабільнішим до шуму.

Аналогічно, м'який мінімум можна визначити як:

$$\text{SoftMinPool}_{\beta}(X)_{i,j} = -\text{SoftMaxPool}_{\beta}(-X)_{i,j}. \quad (3.41)$$

де $\text{SoftMinPool}_{\beta}(\cdot)$ — диференційовна апроксимація \min у вікні, інші позначення — як у (3.40).

У класичній морфології дилатація є локальним максимумом, а ерозія — локальним мінімумом. У диференційовній формі це задається так:

$$\text{Dil}_{\beta}(X) = \text{SoftMaxPool}_{\beta}(X), \quad (3.42)$$

$$\text{Ero}_{\beta}(X) = \text{SoftMinPool}_{\beta}(X). \quad (3.43)$$

де $\text{Dil}_{\beta}(\cdot)$ — soft-дилатація, $\text{Ero}_{\beta}(\cdot)$ — soft-ерозія;

β керує компромісом між “жорсткістю” морфології та гладкістю градієнтів.

- Dil_{β} має тенденцію заповнювати дрібні розриви на межі та робити область більш зв'язною.

- Ero_{β} має тенденцію прибирати тонкі “хвости” та дрібні артефакти, що важливо проти FP у фоні.

Композиції ерозії та дилатації дозволяють отримати класичні операції відкриття і закриття, які особливо корисні для структурної корекції масок.

Soft-відкриття:

$$\text{Open}_{\beta}(X) = \text{Dil}_{\beta}(\text{Ero}_{\beta}(X)). \quad (3.44)$$

Soft-закриття:

$$\text{Close}_{\beta}(X) = \text{Ero}_{\beta}(\text{Dil}_{\beta}(X)). \quad (3.45)$$

де $\text{Open}_{\beta}(\cdot)$ — згладжує дрібні виступи та тонкі шумові структури, $\text{Close}_{\beta}(\cdot)$ — заповнює невеликі “дірки” та розриви в межі.

Практичний сенс для ROI серця:

- Open_{β} зменшує фонові “острівці” та дрібні хибні включення;

- Close_{β} покращує зв'язність та замкненість контуру, що критично для

морфометрії (периметр/площа).

Використання SoftMorphPool2D у нейромережі має дві ключові переваги:

– Морфологічна індуктивна упередженість (inductive bias): мережа отримує інструмент, що природно працює з формою, а не лише з текстурою. Для CXR це важливо, бо багато конкурентних структур (ребра, судини) мають сильні локальні градієнти, але не повинні ламати глобальну геометрію серця.

– Керованість через β : параметр β визначає, наскільки оператор наближений до жорсткого \max/\min . Це дозволяє налаштувати компроміс між агресивним “морфологічним очищенням” (корисно проти FP) та збереженням тонких ділянок межі (важливо проти FN на слабкоконтрастних CXR) [90].

Практичний сенс для сегментації має такі soft операції є навчуваними й диференційовними, тому можуть працювати як морфологічне згладжування масок/активацій, зменшуючи дрібні помилки на межі (локальні FP/FN), закриваючи розриви та прибираючи шум без жорстких недиференційовних \max/\min .

Отже, SoftMorphPool2D формалізує морфологічні операції в диференційовній формі, замінюючи \max / \min на гладкі soft-аналоги з параметром β . Це дозволяє вбудовувати морфологічні перетворення у нейромережу та навчати їх end-to-end. У контексті CXR-сегментації серця такі оператори є корисними для підвищення зв’язності ROI, стабілізації межі та пригнічення фонових хибних включень, що напряду покращує придатність маски для подальшого морфологічного аналізу.

3.6. DMGGA: морфологічна увага в кількох масштабах

DMGGA (Morphological Multi-Scale Gated Gradient Attention) — модуль морфологічної уваги, який будує карту підсилення не з глобальної статистики каналів (як SE) і не з класичної spatial-attention, а з морфологічного градієнта. Це особливо корисно для CXR: межа серця часто слабо контрастна, має

накладання ребер і судинного малюнка, тому модель схильна втрачати фрагменти межі (FN) або прилипати до конкурентних контурів (FP). Багатомасштабність (кілька k) дозволяє одночасно реагувати і на дрібні нерівності, і на грубі межові переходи.

DMGGA не замінює ознаки, а м'яко підсилює їх у важливих областях, щоб не зруйнувати базовий шлях encoder–decoder. Це реалізовано як резидуальне множення на (a), де a — карта gating у діапазоні [1] [89, 90].

$$y = x \cdot (1 + a). \quad (3.46)$$

де $x \in R^{B \times C \times H \times W}$ — вхідний тензор ознак, $y \in R^{B \times C \times H \times W}$ — вихідний тензор ознак, $a \in [0,1]^{B \times 1 \times H \times W}$ — просторово-скалярна карта уваги (gating), B — розмір батчуЮ C — кількість каналів; H, W — просторові розміри, \cdot — поелементне множення з broadcast по каналах.

Морфологічні операції через вікно $k \times k$ обчислюються дорожчі за звичайні 3×3 згортки, тому DMGGA спочатку стискає канали до C_g (у роботі gate_ch), щоб виконувати морфологію на компактному поданні.

$$z = \phi (Norm (Conv_{1 \times 1}(x))). \quad (3.47)$$

де $z \in R^{B \times C_g \times H \times W}$ — стиснуті ознаки для attention, C_g — кількість “gating-каналів” ($C_g \ll C$), $Conv_{1 \times 1}$ — згортка 1×1 (проєкція $C \rightarrow C_g$), $Norm$ — нормалізація (у реалізації мережі), $\phi(\cdot)$ — активація (SiLU).

У SoftMorphPool2D базовою операцією є soft-дилатація. Soft-ерозія задається через дуальність, що узгоджується з класичною математичною морфологією та спрощує реалізацію.

$$\varepsilon_{k_i}(z) = -\delta_{k_i}(-z). \quad (3.48)$$

де k_i — розмір вікна (структурного елемента) в масштабі i , $\delta_{k_i}(\cdot)$ — soft-дилатація з вікном $k_i \times k_i$, $\varepsilon_{k_i}(\cdot)$ — soft-ерозія з тим самим вікном, $-z$ — поелементне заперечення тензора, зовнішній — поелементне заперечення результату.

Морфологічний градієнт є різницею дилатації та ерозії. Він підсилює зони переходів (контур/границю) і пригнічує однорідні області, що відповідає

інтуїції “детектор межі”, але в морфологічній формі.

$$g_{k_i} = \delta_{k_i}(z) - \varepsilon_{k_i}(z). \quad (3.49)$$

де $g_{k_i} \in R^{B \times C_g \times H \times W}$ — градієнт у масштабі k_i , $\delta_{k_i}(z)$ — soft-дилатація ознак z , $\varepsilon_{k_i}(z)$ — soft-ерозія ознак z , — поелементна різниця.

Один фіксований масштаб k не підходить усім випадкам: дрібні артефакти потребують малого k , а грубі межі — більшого. Тому DMGGA використовує кілька k_i навчувано визначає їхню вагу.

$$w_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^m \exp(\alpha_j)}, i = 1..m. \quad (3.50)$$

де m — кількість масштабів, α_i — навчуваний параметр (logit) масштабу k_i , w_i — нормована вага масштабу ($w_i \geq 0$, $\sum_i w_i = 1$), $\exp(\cdot)$ — експонента.

Зважена агрегація морфологічних сигналів:

$$g = \sum_{i=1}^m w_i g_{k_i}. \quad (3.51)$$

де $g \in R^{B \times C_g \times H \times W}$ — комбінований морфологічний сигнал, w_i — ваги з (3.50), g_{k_i} — градієнти з (3.49).

Щоб перейти від багатоканального сигналу g до єдиної просторової маски, застосовується 1×1 згортка, а далі sigmoid для обмеження в [1]. Це забезпечує кероване підсилення в (3.46).

$$a = \sigma(\text{Conv}_{1 \times 1}(g)). \quad (3.52)$$

де $a \in [0,1]^{B \times 1 \times H \times W}$ — карта уваги (gating), $\text{Conv}_{1 \times 1}$ — згортка 1×1 , що зводить $C_g \rightarrow 1$, $\sigma(\cdot)$ — сигмоїда, g — комбінований сигнал з (3.51).

Вхідний тензор ознак $x \in R^{B \times C \times H \times W}$ спочатку стискається 1×1 згорткою (за потреби з нормалізацією та нелінійністю) до компактного представлення $z \in R^{B \times C_g \times H \times W}$, що зменшує обчислювальну складність подальших операцій. Далі паралельно для кількох розмірів вікна $k \in \{3,5,7\}$ формується морфологічний сигнал на різних масштабах (через soft-

дилатацію/soft-ерозію та їх різницю — морфологічний градієнт). Внесок кожного масштабу визначається навчуваними вагами $w_i = \text{softmax}(\alpha_i)$, після чого отримується агрегований сигнал $g = \sum_i w_i g_{k_i}$. На основі g будується просторово-скалярна gating-маска $a \in [0,1]^{B \times 1 \times H \times W}$ (через 1×1 згортку та сигмоїду), яка керує підсиленням ознак за правилом $y = x \cdot (1 + a)$. Таким чином модуль підсилює інформативні області (переважно на межах та структурних переходах) і підвищує стійкість сегментації при наявності шуму та слабкого контрасту на CXR.

У реалізації DMGGA підключається як опційний підмодуль всередині морфологічного блоку (прапорець `use_dmga`). Це логічно, оскільки DMGGA використовує морфологічні оператори і “працює” з контурними переходами. На практиці модуль збільшує чутливість мережі до межі, але не створює жорстких правил: карта a є навчуваною й адаптується до даних.

DMGGA формує просторову карту уваги a на основі морфологічного градієнта, що напряду підсилює структурні переходи (контури) і зменшує вплив однорідного фону.

Використання кількох масштабів k_i дозволяє модулю враховувати як дрібні, так і грубі межові структури, а навчвані ваги w_i автоматично підлаштовують внесок кожного масштабу.

Резидуальна форма підсилення $y = x \cdot (1 + a)$ забезпечує стабільність: при $a \approx 0$ блок майже не впливає на ознаки, а при $a \rightarrow 1$ виконує контрольоване підсилення без різких спотворень.

Завдяки компресії каналів до C_g морфологічна увага залишається обчислювально доцільною та може застосовуватися в encoder–decoder архітектурі для задач сегментації серця на CXR [20, 50, 51, 67, 68, 74, 75, 85].

3.7. ABR (Adaptive Boundary Rejection): прогноз порогу та правило відкидання у фон

ABR (Adaptive Boundary Rejection) — механізм, який зменшує кількість

хибних спрацьовувань на фоні та “конкурентних” контурах (ребра, судини, діафрагма) шляхом адаптивного порогоування впевненості моделі. Ідея така: якщо модель у пікселі недостатньо впевнена у будь-якому з не-фонових класів, цей піксель примусово переводиться у фон. На відміну від фіксованого порогу, ABR прогнозує поріг автоматично з ознак поточного зображення/карти [21–23, 31, 43, 72, 73].

Реалізація порогу прогнозується окремою “головою” із декодерних ознак (після edge-fuse). Вона складається з глобального усереднення (AdaptiveAvgPool2d(1)) і двох 1×1 згорток з SiLU, тобто дає один скаляр на зображення — logit порогу τ_{logit} [15–19, 21–23].

$$\tau_{logit} = f_{thr}(x), \quad (3.53)$$

де $x \in R^{B \times C \times H \times W}$ — карта ознак (feature map) на виході декодера/edge-fuse, з якої береться поріг, $f_{thr}(\cdot)$ — ABR-head, $\tau_{logit} \in R^{B \times 1}$ — логіт порогу для кожного елемента батчу, B — розмір батчу.

Щоб поріг був у контрольованому діапазоні $[\tau_{max}]$, у роботі застосовано sigmoid + лінійне масштабування:

$$\tau = \tau_{min} + (\tau_{max} - \tau_{min}) \cdot \sigma(\tau_{logit}). \quad (3.54)$$

де $\tau \in R^{B \times 1}$ — прогнозований поріг ABR, τ_{min}, τ_{max} — нижня/верхня межі порогу, $\sigma(\cdot)$ — сигмоїда, τ_{logit} — logit з (3.53).

Нехай $s \in R^{B \times K \times H \times W}$ — логіти сегментації для K не-фонових класів (у тебе logits6, тобто $K = 6$). Ймовірності:

$$p_k(u) = softmax(s(u))_k, k = 1..K, \quad (3.55)$$

де u — індекс пікселя, $s(u) \in R^K$ — вектор логітів у пікселі u , $p_k(u)$ — ймовірність класу k у пікселі u , K — кількість не-фонових класів, $softmax(\cdot)$ — softmax по каналному виміру.

Максимальна не-фонова впевненість:

$$p_{max}(u) = max_{k \in \{1..K\}} p_k(u). \quad (3.56)$$

де $p_{max}(u) \in [0,1]$ — максимальна ймовірність серед не-фонових класів

у пікселі u .

Тоді ABR-заміна на фон виконується правилом:

$$\hat{y}(u) = \{0, p_{max}(u) < \tau, \arg \max_{k \in \{1..K\}} p_k(u), p_{max}(u) \geq \tau, \quad (3.57)$$

де $\hat{y}(u)$ — фінальна мітка пікселя після ABR, 0— індекс фонового класу, τ — поріг з (3.54), який застосовується (в інференсі може бути авто або manual), $\arg \max$ — клас з максимальною ймовірністю серед не-фонових.

Прогноз порогу τ (глобальна адаптація під зображення). ABR-head отримує карту ознак x з декодера (після інтеграції контурного сигналу), і перетворює її в один скаляр τ на кожне зображення батчу. Для цього використовується глобальне усереднення по простору, що збирає інформацію про загальну якість поточного прикладу: рівень контрасту, наявність шуму та артефактів, ступінь неоднозначності контурів. Далі через 1×1 згортки та SiLU отримується τ_{logit} , який переводиться у поріг $\tau \in [\tau_{min}, \tau_{max}]$. Таким чином, поріг не задається вручну один раз для всіх зображень, а адаптується: за складніших умов (шум, слабкий контраст) модель може прогнозувати вищий τ і бути обережнішою, а за чистіших знімків — нижчий τ , щоб не втратити тонкі деталі.

Розрахунок $p_{max}(u)$ (локальна оцінка впевненості по пікселях). З сегментаційних логітів $s(u)$ (для не-фонових класів) обчислюються ймовірності $p_k(u)$ через softmax. Далі береться максимум по класах:

$$p_{max}(u) = \max_k p_k(u). \quad (3.58)$$

Данна величина є піксельною мірою впевненості моделі: якщо хоча б один клас сильно домінує, p_{max} високий; якщо модель вагається між кількома класами або прогноз близький до рівномірного, p_{max} низький.

Далі порівнюється $p_{max}(u)$ з порогом τ . Якщо $p_{max}(u) < \tau$, піксель примусово переводиться у фон $\hat{y}(u) = 0$.

Це означає: модель не достатньо впевнена, що тут є корисна структура — краще вважати це фоном. Якщо ж $p_{max}(u) \geq \tau$, піксель отримує клас $\arg \max_k p_k(u)$ без змін. Такий механізм діє як контрольована постобробка

всередині пайплайна, але не потребує окремих евристик чи зовнішніх алгоритмів [91,92].

Найбільша кількість хибних спрацьовувань у сегментації CXR виникає в прикордонних областях, де модель бачить багато конкурентних текстур: ребра, судинний малюнок, тіні діафрагми. У таких зонах часто формується “середня” впевненість, тобто p_{max} падає. ABR у цьому випадку відсікає ці пікселі у фон, що зменшує FP біля межі. При цьому в центральних ділянках об’єкта (де ознаки стабільні) p_{max} зазвичай високий, і ABR не впливає на результат.

У реалізації інференсу це зводиться до простого правила: після обчислення p_{max} формується маска (τ), і для цих позицій індекс класу примусово замінюється на BG_CLASS_ID. Таким чином, ABR не змінює самі логіти чи softmax, а працює на рівні фінального рішення, забезпечуючи простоту та стабільність.

У роботі ABR-head можна навчати дуже “м’яко” через додаткову втрату, яка розводить p_{max} відносно порогу:

- для BG-пікселів модель має робити p_{max} нижчим за τ (з запасом),
- для FG-пікселів — вищим за τ (з запасом).

У кодї це реалізовано через softplus з температурою та маржинами.

Втрати для фону:

$$L_{bg} = E_{u \in \Omega_{bg}} \left[\text{softplus} \left(\frac{p_{max}(u) - (\tau - \Delta_{bg})}{T} \right) \right], \quad (3.59)$$

де Ω_{bg} — множина пікселів фону за GT, $\Delta_{bg} > 0$ — маржа для фону, $T > 0$ — температура згладжування, $\text{softplus}(x) = \ln(1 + e^x)$, E — середнє по вибраних пікселях.

Втрати для переднього плану:

$$L_{fg} = E_{u \in \Omega_{fg}} \left[\text{softplus} \left(\frac{(\tau + \Delta_{fg}) - p_{max}(u)}{T} \right) \right], \quad (3.60)$$

де Ω_{fg} — множина не-фонових пікселів за GT, $\Delta_{fg} > 0$ — маржа для FG ,

інші позначення — як у (3.58).

Загальна допоміжна ABR-втрата :

$$\mathcal{L}_{thr} = \frac{1}{2} (\mathcal{L}_{bg} + \mathcal{L}_{fg}). \quad (3.61)$$

де L_{thr} — допоміжна втрата для навчання порогу.

Окремо в роботі передбачено регуляризатор “BG confidence” (для BG-пікселів зменшувати p_{max}):

$$L_{bgconf} = E_{u \in \Omega_{bg}} [-\ln (1 - p_{max}(u) + \varepsilon)]. \quad (3.62)$$

де ε — мала константа для стабільності (у кодї 10^{-6}), інші позначення — як вище.

Фінальна функція втрат у найпростішому варіанті навчання/файнтюну :

$$L = L_{seg} + \lambda_{bg} L_{bgconf} + \lambda_{thr} L_{thr}, \quad (3.63)$$

де, L_{seg} — основна втрата сегментації, λ_{bg} — вага регуляризатора фону, λ_{thr} — вага втрати порогу, L_{bgconf} — з (3.62), L_{thr} — з (3.61).

У додатку ваги спеціально тримаються малими щоб не дестабілізувати finetune, тобто ABR — допоміжний механізм, який покращує межі або фон, але не має домінувати над основною сегментацією.

ABR вводить адаптивний поріг τ , який прогнозується окремою головою з ознак моделі та обмежується діапазоном $[\tau_{max}]$ [87-88].

За правилом (3.58) пікселі з низькою максимальною не-фоновою впевненістю p_{max} автоматично переводяться у фон, що зменшує FP у фоні та на конкурентних контурах.

Для навчання порогу використовується м'який критерій (3.59)–(3.61) з маржами і температурою, який розводить p_{max} відносно τ для BG/FG, а також додатковий регуляризатор фону (3.62).

У сумі ABR виступає як контурно-орієнтований механізм постобробки в середині мережі, що підвищує стабільність меж у складних CXR-сценаріях, не вводячи жорсткого фіксованого порогу.

3.8. Функції втрат та стратегія навчання моделі

Навчання мережі виконується за класичною схемою мінімізації суми основної втрати сегментації та допоміжних регуляризаторів, які відповідають окремим механізмам моделі (зокрема ABR). Такий підхід дозволяє зберігати стабільність оптимізації (основний сигнал іде від L_{seg}), але водночас підказувати мережі бажану поведінку в проблемних областях (фон/межі).

У реалізації роботи основним критерієм виступає піксельна крос-ентропія між прогнозом та GT-розміткою. Для багатокласової сегментації з фоном це стандартний вибір, який забезпечує коректне навчання карти класів [30, 71, 76–79].

Крос-ентропійна функція втрат має інформаційно-теоретичну інтерпретацію, оскільки оцінює розбіжність між еталонними мітками та прогнозованим розподілом імовірностей моделі. У цьому контексті ентропія характеризує ступінь невизначеності прогнозу для кожного пікселя сегментаційної маски: що ближчим є розподіл імовірностей між класами до рівномірного, то менш упевненим є рішення моделі. Навпаки, низька ентропія відповідає більш упевненому прогнозу, коли один із класів має домінуючу ймовірність. Така інтерпретація узгоджується з механізмом Adaptive Boundary Rejection, який використовує поріг упевненості для віднесення невпевнених пікселів до фонового класу.

$$\mathcal{L}_{seg} = -\frac{1}{|\Omega|} \sum_{u \in \Omega} \sum_{c=0}^K y_c(u) \ln p_c(u). \quad (3.64)$$

де Ω — множина всіх пікселів зображення, $|\Omega| = H \cdot W$, u — індекс пікселя, $c \in \{0..K\}$ — індекс класу, де $c = 0$ — фон, K — кількість нефонових класів, $y_c(u) \in \{0,1\}$ — one-hot розмітка (1 для істинного класу, 0 інакше), $p_c(u)$ — ймовірність класу c у пікселі u , \ln — натуральний логарифм.

Ймовірності отримуються через softmax з логітів $s(u)$:

$$p_c(u) = \frac{\exp(s_c(u))}{\sum_{j=0}^K \exp(s_j(u))}. \quad (3.65)$$

де $s(u) = [s_0(u), \dots, s_K(u)]$ — логіти сегментації у пікселі u , $s_c(u)$ — логіт для класу c , $\exp(\cdot)$ — експонента.

У роботі використовується додатковий термін, який зменшує впевненість моделі у не-фонових класах на пікселях, що належать фону за GT. Це знижує FP у фоні та стабілізує контури.

$$L_{bgconf} = E_{u \in \Omega_{bg}} [-\ln(1 - p_{max}(u) + \varepsilon)]. \quad (3.66)$$

де Ω_{bg} — множина пікселів фону за GT, $p_{max}(u) = \max_{k \in \{1..K\}} p_k(u)$ — максимальна “не-фонова” впевненість, $\varepsilon > 0$ — мала константа для уникнення $\ln(0)$ (у коді 10^{-6}), E — середнє по Ω_{bg} .

Щоб ABR-поріг τ навчався не випадково, у роботі є м’який критерій, який розводить $p_{max}(u)$ відносно τ для BG/FG з маржами. (Це те, що в 3.7 було формулами (3.58)–(3.60); тут ми фіксуємо, що це частина функції втрат у навчанні.)

Підсумковий термін:

$$L_{thr} = \frac{1}{2}(L_{fg}). \quad (3.67)$$

де L_{bg} — штраф для фонових пікселів (змушує p_{max} бути нижче τ маржею), L_{fg} — штраф для не-фонових пікселів (змушує p_{max} бути вище τ маржею), L_{thr} — загальна допоміжна втрата для навчання порогу.

Фінально оптимізується зважена сума:

$$L = L_{seg} + \lambda_{bg} L_{bgconf} + \lambda_{thr} L_{thr}. \quad (3.68)$$

де L — повна функція втрат, L_{seg} — основна втрата сегментації (3.64), L_{bgconf} — регуляризація фону (3.66), L_{thr} — допоміжна втрата для ABR (3.67), $\lambda_{bg}, \lambda_{thr}$ — вагові коефіцієнти.

Основний навчальний сигнал задається піксельною крос-ентропією L_{seg} , що забезпечує коректне навчання багатокласової сегментації.

Для зменшення FP у фоні застосовується регуляризатор L_{bgconf} , який

знижує “не-фонову” впевненість на BG-пікселях.

Для стабільного прогнозу порогу ABR використовується допоміжний термін L_{thr} , який розводить p_{max} відносно τ для BG/FG.

Підсумкова функція втрат є простою та стабільною, а покращення якості меж досягається переважно архітектурними модулями та механізмом ABR.

3.9. Обчислювальна складність та особливості реалізації

Модель належить до класу encoder–decoder архітектур з додатковою edge-гілкою та морфологічними модулями. Основна обчислювальна вартість припадає на згорткові блоки encoder/decoder, тоді як додаткові модулі (SoftMorphPool2D, DMGGA, ABR) збільшують складність локально, але контрольовано — завдяки використанню depthwise-згорток, стисканню каналів для attention та застосуванню морфології лише в окремих шарах.

Для стандартної $k \times k$ згортки над картою $H \times W$ із $C_{in} \rightarrow C_{out}$ оцінка FLOPs (порядку величини) має вигляд:

$$FLOPs_{conv} \approx H \cdot W \cdot C_{in} \cdot C_{out} \cdot k^2. \quad (3.69)$$

де H, W — висота/ширина карти ознак, C_{in}, C_{out} — кількість вхідних/вихідних каналів, k — розмір ядра згортки, $FLOPs_{conv}$ — кількість операцій множення-додавання (приблизна оцінка).

У EdgeMorphBlock ключовим елементом є dilated depthwise conv, який значно дешевший за звичайну згортку:

$$FLOPs_{dw} \approx H \cdot W \cdot C \cdot k^2. \quad (3.70)$$

де C — кількість каналів.

У SoftMorphPool2D морфологічне вікно $k \times k$ реалізується через операцію unfold (розгортання локальних патчів) та застосування log-sum-exp (softmax-аналог max). Тому складність одного застосування soft-дилатації масштабу k оцінюється як:

$$Cost_{morph} \sim B \cdot C \cdot H \cdot W \cdot k^2. \quad (3.71)$$

де B — розмір батчу, C — кількість каналів, на яких застосовується

морфологія, H, W — просторові розміри, k — розмір вікна морфології.

Практично це дорожче за одну 3×3 depthwise-згортку через unfold та експоненти в log-sum-exp, тому морфологія застосовується не в кожному блоці, а в DMGGA виконується на стиснених каналах $C_g \ll C$ (див. 3.6), що різко зменшує витрати.

DMGGA робить морфологічний градієнт на кількох масштабах $K = \{k_1, \dots, k_m\}$ (у тебе типовий набір 3,5,7). Тоді загальна вартість морфологічної частини DMGGA:

$$\text{Cost}_{\text{DMGGA}} \sim B \cdot C_g \cdot H \cdot W \cdot \sum_{i=1}^m k_i^2. \quad (3.72)$$

де C_g — кількість каналів після компресії, m — кількість масштабів, k_i — розмір вікна на масштабі i , інші позначення — як у (3.71).

Оскільки C_g обмежується невеликим значенням, навіть при трьох масштабах витрати залишаються прийнятними.

ABR додає thr_head — невелику “голову” з глобальним pooling та двома 1×1 згортками (майже не впливає на FLOPs порівняно з декодером);

операцію обчислення p_{max} та маски (τ)— це лінійно по пікселях:

$$\text{Cost}_{\text{ABR}} \sim B \cdot H \cdot W \cdot K. \quad (3.73)$$

де K — кількість не-фонових класів, інші позначення — як раніше.

Ця частина зазвичай не є вузьким місцем і виконується швидко.

Отже, ABR зазвичай не є вузьким місцем обчислень і виконується швидко навіть на GPU/TPU, оскільки зводиться до простих елементних операцій та редукції по каналах, тоді як основні витрати часу в архітектурі формуються згортковими блоками базової encoder–decoder частини та морфологічними компонентами, що використовують просторові вікна.

Для узагальненого порівняння накладних витрат окремих компонентів запропонованої архітектури доцільно розглянути їхній внесок у обчислювальну складність і споживання пам’яті відносно базового encoder–decoder контуру [92, 93]. Оскільки різні модулі виконують неоднакові за

природою операції (згортки повної розмірності, depthwise-операції, редукції, морфологічні віконні перетворення), їхні витрати зручно оцінювати якісно за шкалою низька, середня та висока з урахуванням типових режимів застосування (просторова роздільність, число каналів, розмір вікна k). Результати такого порівняння наведено в табл. 3.1.

Таблиця. 3.1 – Якісна оцінка накладних витрат модулів

| Компонент | Що додає | Обчис. вартість | Витрати пам'яті | Коментар |
|-----------------------------------------|------------------------------------|-----------------|-----------------|-------------------------------------------|
| Encoder–decoder (база) | Conv/Down/Up | Висока | Середня | Основний вклад у FLOPs і latency |
| Edge-branch (Sobel/градієнт + conv) | Контурний сигнал | Низька | Низька | Легка гілка, працює паралельно |
| EdgeMorphBlock (depthwise dilated + SE) | Локальне підсилення меж | Низька | Низька | Depthwise |
| SoftMorphPool2D (unfold + log-sum-exp) | Диференційовна морфологія | Середня | Середня | Найважчий додатковий блок |
| DMGGA (multi-scale) | Морфологічна увага $k = \{3,5,7\}$ | Середня | Середня | Компенсується компресією до C_g |
| ABR (thr_head + pmax< τ) | Адаптивне відкидання у BG | Низька | Низька | Порогування, лінійна операція по пікселях |

Як видно з табл. 3.1, найбільшу накладну вартість серед додаткових модулів створює SoftMorphPool2D через операції unfold і обчислення log-sum-exp у вікні $k \times k$. Саме тому в реалізації морфологічні оператори

застосовуються локально та/або на стислих каналах (у DMGGA використовується $C_g \ll C$). Натомість ABR майже не впливає на швидкодію, оскільки складається з легкої порогової голови та простого правила порівняння $p_{max} \geq \tau$.

Найбільш “важкий” з точки зору пам’яті компонент — `unfold` у `SoftMorphPool2D`, оскільки він тимчасово створює тензор з розмірністю, пропорційною k^2 . Тому в реалізації важливо:

- контролювати k (звичайно 3/5/7);
- не застосовувати морфологію на надто великих H, W в ранніх шарах без потреби;
- виконувати морфологію на стислих каналах C_g (як у DMGGA);
- використовувати змішану точність (AMP) при наявності GPU.

Також у роботі присутні технічні рішення для стабільного інференсу:

- `model.eval()` і `torch.no_grad()` для вимкнення градієнтів;
- обробка порога ABR як `thr_used` (авто від `thr_head` або ручний у UI);
- можливість рендеру кольорової маски та накладення на CXR у Gradio.

В інтерфейсі Hugging Face/Gradio найбільша затримка зазвичай не в ABR чи в постобробці, а в наступному:

- завантаженні моделі/ваг (якщо cold start);
- forward-проході `encoder-decoder` на вхідному розмірі;
- формуванні візуалізацій (`overlay`, `resize`, конвертації `PIL` ↔ `NumPy`).

ABR у цьому пайплайні додає лише одну маску-порівняння та переприсвоєння класу фону, що практично не впливає на `latency`.

Основна складність моделі визначається `encoder-decoder` згортковими блоками; застосування `depthwise`-згорток у морфологічних блоках зменшує FLOPs порівняно зі стандартними `conv`.

`SoftMorphPool2D` є найвитратнішим додатковим компонентом через `unfold` та `log-sum-exp`, тому він використовується локально та/або на стислих каналах.

DMGGA компенсує багатомасштабність компресією до C_g , тому приріст витрат контрольований навіть при $k = \{3,5,7\}$.

ABR має мінімальні обчислювальні накладні витрати: `thr_head` є “легкою” головою, а правило відкидання реалізується лінійно по пікселях.

Практична продуктивність у UI (Gradio) залежить переважно від forward-проходу моделі та етапів візуалізації, а не від ABR-постобробки.

3.10 Висновок до розділу 3

У розділі 3 розроблено та формалізовано нейронну архітектуру CardioEdgeBioMorphNet, призначену для автоматизованої сегментації серцевої області на рентгенівських знімках грудної клітки. Запропонована модель орієнтована не лише на формування сегментаційної маски, а й на отримання стабільної ROI, придатної для подальшого морфометричного аналізу: оцінювання площі, контуру, форми та співвідношення розмірів серцевої області.

Архітектура побудована за схемою encoder–decoder з багатомасштабним виділенням ознак, розширенням контексту в bottleneck-рівні за допомогою ASPP та поетапним відновленням просторової роздільності в декодері. З урахуванням особливостей CXR-зображень — слабкої вираженості меж серця, шумів, проєкційних накладань анатомічних структур і варіативності якості знімків — модель доповнено спеціалізованими edge-aware та морфологічно орієнтованими механізмами.

Для уточнення меж серцевої області реалізовано окрему edge-гілку на основі Sobel-градієнта, яка формує явний контурний сигнал і передає його в декодер через блок Edge Fuse. Це дозволяє підсилити структурну інформацію в прикордонних ділянках і зменшити ризик зміщення контуру або його хибного накладання на суміжні анатомічні структури.

Морфологічна складова моделі представлена блоками EdgeMorphBlock, механізмом SoftMorphPool2D та модулем морфологічної уваги DMGGA. Їх використання спрямоване на збереження зв'язності ROI, підсилення значущих

структурних переходів і стабілізацію форми серцевої області на різних масштабах. Завдяки цьому модель краще враховує геометричні особливості об'єкта сегментації та зменшує кількість локальних дефектів маски.

Для контролю хибнопозитивних включень у фонових ділянках запропоновано механізм Adaptive Boundary Rejection (ABR), який виконує адаптивне пороговання впевненості сегментаційного прогнозу. Пікселі з недостатньою впевненістю відкидаються у фоновий клас, що є важливим для CXR-зображень із неоднозначними межами, шумами та слабкоконтрастними переходами.

У розділі також визначено функції втрат, стратегію навчання та основні аспекти обчислювальної складності запропонованої архітектури. Показано, що додаткові модулі мають функціонально обґрунтоване призначення: edge-гілка уточнює контур, морфологічні компоненти стабілізують форму ROI, а ABR зменшує кількість помилкових включень фону.

Таким чином, у розділі 3 сформовано цілісну архітектуру CardioEdgeBioMorphNet, у якій поєднано багатомасштабне представлення ознак, контурну інформацію, морфологічні механізми та адаптивне відкидання невпевнених пікселів. Запропонована модель орієнтована на отримання зв'язної, стабільної та інтерпретованої маски серцевої області, що може бути використана для подальшого аналізу морфологічних змін серцевої тіні.

Окрім цього, у розділі розроблено прикладне програмне забезпечення для взаємодії з моделлю, яке забезпечує завантаження CXR-знімка, попередню обробку, виконання сегментації, формування ROI-маски, overlay-візуалізацію та відображення ключових параметрів результату. Наявність режимів автоматичного та ручного пороговання дає змогу оцінювати вплив ABR на якість сегментації та практично апробувати запропонований підхід без необхідності запуску окремих експериментальних скриптів.

РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

У даному розділі наведено експериментальні дослідження, спрямовані на всебічну перевірку переваги запропонованої в роботі моделі CardioEdgeBioMorphNet для задачі сегментації медичних зображень. Основна мета експериментів полягає в кількісному та якісному підтвердженні переваг запропонованих архітектурних рішень і процедур обробки порівняно з поширеними базовими підходами. Для забезпечення об'єктивності оцінювання використано єдиний протокол тестування, фіксовані правила підготовки даних, а також однакові умови навчання й інференсу для всіх порівнюваних моделей.

Кількісна оцінка якості виконувалась за метриками accuracy, Dice та IoU, що дозволяє на пряму порівнювати результати між різними архітектурами та конфігураціями. Окрему увагу приділено порівнянню з базовими моделями U-Net, nnU-Net, DeepLab, CNN-Transformer, MorphNet а також аналізу внеску кожного з ключових компонентів запропонованого підходу шляхом проведення абляційного дослідження (відключення edge-модуля, морфологічних операцій/блоків, механізмів злиття ознак тощо). Такий підхід дає змогу не лише зафіксувати підсумкову точність, а й аргументовано показати, які саме елементи забезпечують приріст якості та формують наукову новизну роботи.

Окрім табличних результатів і графіків навчання, у розділі подано якісний аналіз: візуальні приклади прогнозованих масок, карти помилок і типові ситуації виникнення FP/FN (зокрема на межах об'єкта). Завершує розділ обговорення отриманих результатів, де сформульовано ключові висновки, описано типові помилки та визначено обмеження підходу і напрями подальшого вдосконалення.

4.1 План експериментів і протокол оцінювання

У данному підрозділі наведено загальний план експериментального дослідження, спрямованого на перевірку переваги запропонованої архітектури CardioEdgeBioMorphNet у задачі сегментації серцевої області на рентгенівських знімках грудної клітки. Основна увага приділяється не повторному опису даних, розмітки чи метрик, що були розглянуті у розділі 2, а визначенню логіки проведення експериментів, правил порівняння моделей та порядку інтерпретації отриманих результатів.

Метою експериментального дослідження є встановлення того, наскільки запропонована модель здатна точно виділяти серцеву ROI, відтворювати її контур і зменшувати типові помилки сегментації у складних ділянках CXR-зображень. До таких ділянок належать зони слабого контрасту, області накладання ребер, діафрагми та легневих структур, а також фрагменти зображення, у яких межа серця має нечіткий або частково втрачений характер.

Експерименти також мають на меті порівняти CardioEdgeBioMorphNet із базовими архітектурами семантичної сегментації. Таке порівняння необхідне для того, щоб визначити, чи забезпечує запропонована архітектура практичне покращення порівняно з поширеними encoder–decoder, контекстними, трансформерними та морфологічно-орієнтованими підходами. Окремо досліджується внесок спеціалізованих компонентів моделі, зокрема edge-гілки, EdgeMorphBlock, морфологічного механізму DMGGA, механізмів злиття ознак та Adaptive Boundary Rejection.

Для забезпечення об'єктивності порівняння всі моделі досліджувалися в однакових умовах. Використовувалися фіксовані train, validation і test підвибірки, єдина процедура підготовки зображень і масок, однаковий розмір вхідних даних, однакові метрики оцінювання та спільний принцип вибору найкращого стану моделі. Такий підхід дає змогу пов'язувати відмінності в результатах саме з архітектурними особливостями моделей, а не з різними умовами навчання, складом вибірок або способом оцінювання.

Загальна логіка експериментального протоколу передбачає послідовне виконання таких етапів:

- навчання кожної моделі на фіксованій навчальній підвибірці;
- контроль якості на валідаційній підвибірці після епох навчання;
- вибір найкращого стану моделі за основним валідаційним критерієм;
- фінальне тестування найкращого checkpoint на незалежній тестовій підвибірці;
- порівняння результатів запропонованої архітектури з базовими моделями;
- проведення абляційного дослідження для оцінки внеску окремих компонентів CardioEdgeBioMorphNet;
- якісний аналіз прикладів сегментації, зокрема характерних FP/FN-помилки.

У межах першого етапу кожна модель навчається на однаковій навчальній вибірці. Під час навчання виконується оптимізація параметрів моделі на основі функції втрат, що зіставляє прогнозовану карту сегментації з еталонною маскою. На цьому етапі модель формує здатність відтворювати серцеву область, її межу та внутрішню багатокласову структуру відповідно до прийнятої системи міток.

Другий етап пов'язаний із валідаційним контролем. Валідаційна підвибірка не використовується для оновлення параметрів моделі, однак дає змогу оцінити, наскільки стабільно модель працює на даних, які не брали участі в оптимізації ваг. Це дозволяє контролювати перенавчання, аналізувати динаміку метрик та визначати момент, у який модель досягає найкращої якості сегментації.

На третьому етапі обирається найкращий checkpoint. У роботі для цього використовується валідаційний критерій, пов'язаний з якістю виділення серцевої ROI. Такий вибір є обґрунтованим, оскільки для подальшого морфологічного аналізу та інтерпретації результатів ключовим є коректне відтворення серцевої області, а не лише мінімізація функції втрат. Відповідно,

для фінального тестування використовується не обов'язково остання епоха навчання, а стан моделі, що показав найкращу якість на validation-підвибірці.

Четвертий етап передбачає фінальну оцінку на тестовій підвибірці. Test-підвибірка не використовується під час навчання, вибору гіперпараметрів або визначення найкращої епохи. Її роль полягає у незалежній перевірці узагальнювальної здатності моделі. Саме результати на test-підвибірці використовуються для підсумкового порівняння архітектур.

У межах порівняльного дослідження запропонована модель зіставляється з кількома базовими архітектурами, що відображають різні підходи до сегментації. До них належать U-Net, nnU-Net-подібна архітектура, DeepLabV3, CNN-Transformer та MorphNet. Такий набір моделей дозволяє порівняти CardioEdgeBioMorphNet із класичним encoder-decoder підходом, автоматизованою U-Net-подібною конфігурацією, контекстною ASPP-архітектурою, гібридним CNN-Transformer підходом і морфологічно-орієнтованою моделлю. Детальні умови порівняння базових моделей наведено у підрозділі 4.2.

Окремим напрямом експериментального дослідження є абляційний аналіз. Його мета полягає у визначенні того, який внесок у підсумкову якість сегментації роблять окремі компоненти запропонованої архітектури. Зокрема, аналізується вплив edge-гілки, механізму злиття ознак, морфологічних блоків, DMGGA та Adaptive Boundary Rejection. Такий підхід дає змогу не лише показати перевагу повної моделі, а й обґрунтувати доцільність кожного спеціалізованого модуля.

Кількісна оцінка результатів виконується за метриками, визначеними у розділі 2. У розділі 4 ці метрики використовуються не для повторного теоретичного опису, а для практичного порівняння моделей. Основна увага приділяється показникам, що характеризують якість виділення ROI, точність багатокласової сегментації та стійкість моделі до помилок на межі серцевої області.

Якісний аналіз доповнює чисельні результати. Він дозволяє візуально

оцінити, як модель відтворює контур серця, чи виникають хибнопозитивні включення фонових структур, чи втрачаються ділянки серцевої області та наскільки узгодженою є прогнозована ROI-маска. Такий аналіз є важливим, оскільки навіть високі чисельні метрики не завжди повністю відображають характер помилок у медичних зображеннях.

Таким чином, підрозділ 4.1 визначає загальну логіку експериментального дослідження. Він фіксує порядок навчання, валідації, вибору найкращого стану моделі, тестування, порівняння з бейзлайнами, проведення абляційного аналізу та якісної інтерпретації результатів. Подальші підрозділи розділу 4 деталізують окремі частини цього протоколу: опис базових моделей, кількісні результати, абляційне дослідження, якісний аналіз і підсумкову інтерпретацію отриманих результатів [58, 64, 65].

4.2 Бейзлайнові моделі та умови порівняння

У підрозділі наведено моделі, з якими порівнюється запропонована архітектура CardioEdgeBioMorphNet. Метою такого порівняння є перевірка того, чи забезпечує запропонований підхід покращення якості сегментації серцевої області порівняно з поширеними нейромережевими архітектурами та морфологічно-орієнтованими моделями.

Оскільки теоретичні особливості основних класів сегментаційних архітектур уже були розглянуті в розділі 1, у цьому підрозділі не дублюється детальний опис принципів роботи кожної моделі. Основна увага приділяється ролі кожної архітектури в експериментальному порівнянні та обґрунтуванню того, чому саме ці моделі доцільно використовувати як бейзлайни для задачі сегментації серцевої області на CXR-зображеннях.

Для порівняння використано такі моделі: U-Net, nnU-Net-подібну архітектуру, DeepLabV3, CNN-Transformer, MorphNet та запропоновану CardioEdgeBioMorphNet. Такий набір охоплює кілька різних напрямів розвитку сегментаційних моделей: класичний encoder-decoder підхід, посилені U-Net-подібні архітектури, багатомасштабне контекстне узгодження,

використання глобальної уваги та морфологічно-орієнтовану обробку ознак. Це дозволяє оцінити запропоновану модель не лише відносно одного базового підходу, а в ширшому контексті сучасних методів сегментації медичних зображень.

U-Net використовується як базова контрольна модель. Ця архітектура є однією з найпоширеніших у задачах медичної сегментації, оскільки поєднує енкодер для вилучення ознак і декодер для відновлення просторової роздільності. Пропускні з'єднання між відповідними рівнями енкодера та декодера дозволяють частково зберігати локальну просторову інформацію, що є важливим для формування сегментаційної маски. У межах цього дослідження U-Net виступає як початкова точка відліку, відносно якої оцінюється приріст якості інших архітектур.

nnU-Net-подібна архітектура використовується як сильніший U-Net-бейзлайн. Її включення до порівняння дає змогу перевірити, чи забезпечує CardioEdgeBioMorphNet перевагу не лише над базовою U-Net-моделлю, а й над більш адаптованим до медичної сегментації encoder–decoder підходом. Це важливо, оскільки просте порівняння лише з класичною U-Net не завжди є достатнім для обґрунтування переваг нової архітектури.

DeepLabV3 включено до експериментального порівняння як модель, що використовує багатомасштабний контекст. Для CXR-зображень це має значення, оскільки серцева область може мати різний масштаб, а її межі часто виражені нерівномірно. У деяких зонах контур серця є достатньо помітним, тоді як в інших він частково перекривається легневими структурами, ребрами або діафрагмою. DeepLabV3 дозволяє оцінити, наскільки контекстне узгодження на різних масштабах впливає на якість виділення ROI.

CNN–Transformer використовується як гібридна модель, що поєднує локальні згорткові ознаки з механізмами глобальної уваги. Такий бейзлайн є важливим для перевірки того, чи достатньо врахування глобального контексту для покращення сегментації серцевої області. У задачі CXR-сегментації глобальний контекст може допомагати визначати загальне положення серця

на знімку, проте для точного відтворення межі цього може бути недостатньо. Саме тому порівняння з CNN–Transformer дає змогу оцінити, чи має спеціалізована edge-aware та морфологічна обробка перевагу над підходом, орієнтованим переважно на глобальні залежності.

MorphNet розглядається як морфологічно-орієнтований бейзлайн. Його включення до порівняння є принципово важливим, оскільки запропонована CardioEdgeBioMorphNet також використовує морфологічні ідеї, але не обмежується лише ними. На відміну від загального морфологічного підходу, CardioEdgeBioMorphNet поєднує морфологічні блоки з явною контурною інформацією, механізмом DMGGA, edge-fusion та Adaptive Boundary Rejection. Тому порівняння з MorphNet дозволяє оцінити, чи дає перевагу саме комплексне поєднання морфологічних, контурних і адаптивних механізмів.

CardioEdgeBioMorphNet є запропонованою архітектурою, що досліджується в роботі. Її основна відмінність полягає у поєднанні кількох спеціалізованих компонентів, орієнтованих на підвищення точності відтворення серцевої ROI. До таких компонентів належать edge-гілка на основі Sobel magnitude, EdgeMorphBlock, механізм морфологічної уваги DMGGA, злиття ознак з edge-картою та Adaptive Boundary Rejection. Така структура спрямована на зменшення помилок у прикордонних ділянках, де класичні сегментаційні моделі часто включають фонові пікселі до ROI або, навпаки, втрачають частину серцевої області.

Усі моделі порівнювалися в однакових умовах. Для них використовувалися однакові train, validation і test-підвибірки, однаковий розмір вхідних зображень, однаковий формат еталонних масок і однакові метрики оцінювання. Це дозволяє виключити вплив різних умов навчання та забезпечити коректне зіставлення результатів. Найкращий стан кожної моделі визначався за валідаційною підвибіркою, після чого відповідний checkpoint оцінювався на незалежній test-підвибірці.

Важливо, що в цьому порівнянні оцінюється не лише загальна здатність моделі будувати маску, а й її придатність саме для задачі сегментації серцевої

області. Для цього враховуються метрики, пов'язані з якістю ROI, точністю на пікселях серця та здатністю моделі зберігати межу об'єкта. Такий підхід є доцільним, оскільки для подальшого морфологічного аналізу важливо не лише визначити наявність серцевої області, а й коректно відтворити її контур, форму та просторове положення.

Таким чином, обрані бейзлайнові моделі формують достатньо широкий експериментальний контекст для перевірки CardioEdgeBioMorphNet. Порівняння з U-Net і nnU-Net-подібною архітектурою дозволяє оцінити перевагу над класичними encoder–decoder підходами. Порівняння з DeepLabV3 і CNN–Transformer показує, наскільки запропонована модель є переважно відносно контекстних і глобально-орієнтованих архітектур. Порівняння з MorphNet дозволяє окремо оцінити внесок запропонованого поєднання морфологічних і edge-aware механізмів. Це створює основу для подальшого кількісного аналізу результатів у наступному підрозділі.

4.3 Кількісні результати експериментів

У даному підрозділі наведено результати порівняльного оцінювання запропонованої моделі та бейзлайнів, описаних у підрозділі 4.2. Оцінювання виконано за єдиним протоколом (підрозділ 4.1) і повністю відповідає реалізації у ноутбуках: під час навчання на кожній епісі обчислюються метрики на validation, після чого зберігається найкращий стан моделі (best checkpoint) за єдиним критерієм. Для уникнення підглядання у тестову вибірку test-метрики не використовуються для вибору best checkpoint і розглядаються лише як фінальна перевірка узагальнення.

Критерій відбору best checkpoint. На відміну від підходу, що базується лише на точності класифікації пікселів, у реалізації ноутбуків best checkpoint визначається за максимумом $Dice_{val}$

Даний критерій є більш репрезентативним для задачі сегментації, оскільки напряму характеризує перекриття прогнозованої та еталонної областей і є чутливим до помилок межі та фрагментації маски, що типово для

4.3.1 Основна метрика та спосіб агрегування

У підпункті наведено порядок практичного обчислення метрик, за якими порівнювалися запропонована модель CardioEdgeBioMorphNet та бейзлайнові архітектури. Математичні означення основних метрик сегментації подано у розділі 2, тому в межах експериментального розділу основну увагу зосереджено на тому, як саме ці показники використовувалися під час навчання, валідації та тестування моделей.

Для кожного пікселя модель формує прогноз класу на основі максимального значення серед вихідних логітів або ймовірностей. Отримана карта прогнозу порівнюється з еталонною маскою, після чого обчислюються метрики якості. Оскільки задача має багатокласовий характер, оцінювання виконується не лише за фактом належності пікселя до серцевої області, а й з урахуванням правильності його віднесення до одного з клінічних класів.

Основна увага в експериментах приділялася метрикам, що не допускають штучного завищення результатів через домінування фонового класу. На CXR-зображеннях фон зазвичай займає значно більшу площу, ніж серцева область, тому звичайна піксельна точність може бути недостатньо інформативною. Модель може правильно класифікувати більшість фонових пікселів і водночас помилятися в межах серцевої ROI. Саме тому для аналізу використовувалися показники, орієнтовані на серцеву область та не-фонові класи.

Метрика Acc_{heart} використовувалася для оцінювання правильності класифікації пікселів, які належать до серцевої області. При її обчисленні фонові пікселі не враховувалися. Такий підхід дозволяє оцінити, наскільки точно модель розпізнає клінічні класи саме в межах ROI, а не за рахунок великої кількості правильно визначених фонових пікселів.

Метрики Dice та IoU використовувалися для оцінювання ступеня перекриття між еталонною та прогнозованою сегментаційними масками. Dice

характеризує загальний рівень збігу між прогнозом і розміткою, тоді як IoU є більш суворим показником, оскільки оцінює відношення перетину до об'єднання відповідних областей. Обидві метрики розраховувалися для цільових класів без урахування фонового класу.

Окрім метрик просторового перекриття Dice та IoU, якість сегментації може додатково оцінюватися через інформаційну відстань між еталонною та прогнозованою масками. Інформаційна відстань у цьому контексті характеризує відмінності між розподілами пікселів за класами, зокрема між фоном і серцевою ROI. На відміну від Dice та IoU, які відображають ступінь просторового збігу областей, інформаційна відстань дозволяє додатково врахувати статистичну узгодженість прогнозованої маски з еталонною розміткою. Це є важливим для CXR-зображень, оскільки помилки сегментації можуть проявлятися не лише у зміщенні контуру, а й у надмірному або недостатньому віднесенні пікселів до певного класу. У межах цієї роботи основними метриками оцінювання залишаються Dice, IoU та Acc_heart, а інформаційна відстань розглядається як допоміжна характеристика для розширеної інтерпретації результатів сегментації.

Під час багатокласового оцінювання метрики Dice та IoU обчислювалися окремо для кожного клінічного класу. Далі значення усереднювалися лише за тими класами, які були присутні в еталонній розмітці оцінюваного набору. Це дозволяє уникнути спотворення результатів через класи, які фактично відсутні в конкретній вибірці або батчі. Такий підхід є важливим для медичних даних, де розподіл класів може бути нерівномірним.

Під час багатокласового оцінювання метрики Dice та IoU обчислювалися окремо для кожного клінічного класу. Далі значення усереднювалися лише за тими класами, які були присутні в еталонній розмітці оцінюваного набору. Це дозволяє уникнути спотворення результатів через класи, які фактично відсутні в конкретній вибірці або батчі. Такий підхід є важливим для медичних даних, де розподіл класів може бути нерівномірним.

У реалізації метрики спочатку обчислювалися на рівні мінібатча. Після

цього для кожного спліту — train, validation або test — виконувалося усереднення значень за всіма батчами відповідного ладера. Таким чином отримувалися підсумкові значення метрик за епоху. Це дозволяло відстежувати динаміку навчання моделі та порівнювати результати між різними архітектурами в однакових умовах.

Основним критерієм вибору найкращого стану моделі було значення Dice на validation-підвибірці. Найкращим вважався checkpoint тієї епохи, на якій модель показала максимальне значення цієї метрики. Такий вибір є обґрунтованим, оскільки Dice безпосередньо характеризує якість збігу прогнозованої та еталонної області, що є критичним для задачі сегментації серця.

Після завершення навчання фінальне тестування виконувалося лише для вибраного best checkpoint. Тестова підвибірка не використовувалася ні для вибору епохи, ні для налаштування параметрів моделі. Це забезпечує незалежність тестового оцінювання та дозволяє коректно порівнювати узагальнювальну здатність різних архітектур.

Отже, у порівняльних експериментах використовувалася така логіка оцінювання: під час навчання контролювалися train-метрики, для вибору найкращої моделі використовувалися validation-метрики, а остаточне порівняння виконувалося за результатами test-оцінювання. Такий підхід забезпечує відтворюваність експериментів і зменшує ризик завищення результатів через підбір моделі за тестовою вибіркою.

4.3.2 Порівняння з бейзлайнами

Порівняння виконано для набору моделей: U-Net, DeepLabV3+, CNN-Transformer та nnU-Net або еквівалентна конфігурація, а також запропонована модель. Умови навчання, розмір входу, схема аугментацій та критерій вибору best checkpoint для всіх моделей були однаковими (див. підрозділ 4.1), що забезпечує коректність прямого порівняння.

Узагальнені кількісні результати порівняння запропонованої моделі з

бейзлайнами наведено у табл. 4.1–4.3: окремо для train (табл. 4.1), validation (табл. 4.2) та test (табл. 4.3). Для всіх моделей показники наведено на епосі best checkpoint, обраний за критерієм $\max Dice_{val}$.

Таблиця. 4.1 — Порівняння моделей на тренувальній вибірці

| Модель | Епоха best | Acc heart (train) | ROI Dice (train) | Mean Dice (train) | ROI IoU (train) | Mean IoU (train) |
|-----------------|------------|-------------------|------------------|-------------------|-----------------|------------------|
| nnU-Net | 92 | 0.939 | 0.979 | 0.917 | 0.960 | 0.879 |
| U-Net | 94 | 0.932 | 0.954 | 0.887 | 0.913 | 0.827 |
| DeepLabV3 | 91 | 0.936 | 0.975 | 0.926 | 0.951 | 0.883 |
| MorphNet | 94 | 0.938 | 0.955 | 0.897 | 0.914 | 0.840 |
| CEBMN | 95 | 0.938 | 0.979 | 0.926 | 0.959 | 0.885 |
| CNN–Transformer | 92 | 0.939 | 0.979 | 0.917 | 0.960 | 0.879 |

На основі результатів, наведених у табл. 4.3, видно, що на тренувальній вибірці всі порівнювані моделі досягають високих значень якості виділення ROI: ROI Dice перебуває в межах ≈ 0.954 – 0.979 , а ROI IoU — ≈ 0.914 – 0.960 .

Це свідчить про достатню здатність архітектур відтворювати контур серцевої області за наявності навчальних прикладів. Водночас тренувальні метрики є оптимістичними та не відображають повною мірою узагальнюваність, оскільки можуть включати ефект підлаштування під розподіл даних train.

Тому ключовою для коректного порівняння є перевірка на валідаційній вибірці (табл. 4.4), де оцінюється стійкість моделей до варіацій зображень та доменного зсуву. Саме валідаційні значення Mean Dice/Mean IoU та ROI Dice/ROI IoU використовуються як базові критерії вибору найкращого чекпойнта й подальшого аналізу. Надалі порівняння здійснюється з акцентом на баланс між точністю сегментації ROI та узагальнювальною здатністю моделі.

Таблиця. 4.2 — Порівняння моделей на валідаційній вибірці

| Модель | Епоха best | Acc_hear t (val) | ROI Dice (val) | Mean Dice (val) | ROI IoU (val) | Mean IoU (val) |
|-----------------------|---------------|---------------------|----------------------|-----------------------|---------------------|----------------------|
| nnU-Net | 92 | 0.825 | 0.948 | 0.858 | 0.903 | 0.781 |
| U-Net | 94 | 0.841 | 0.921 | 0.846 | 0.855 | 0.748 |
| DeepLabV3 | 91 | 0.825 | 0.950 | 0.868 | 0.906 | 0.787 |
| MorphNet | 94 | 0.833 | 0.919 | 0.838 | 0.852 | 0.736 |
| CardioEdgeBioMorphNet | 95 | 0.841 | 0.950 | 0.879 | 0.907 | 0.798 |
| CNN-Transformer | 96 | 0.841 | 0.928 | 0.852 | 0.868 | 0.758 |

За результатами табл. 4.2 можна оцінити узагальнювальну здатність моделей на валідаційній вибірці, де всі підходи демонструють близькі значення ROI-метрик (ROI Dice ≈ 0.9198 – 0.9505 ; ROI IoU ≈ 0.8527 – 0.9075), що підтверджує стабільне виділення серцевої області. Водночас саме Mean Dice та Mean IoU краще відображають якість сегментації в середньому (враховуючи складніші класи/випадки) і дозволяють розрізнити моделі за практично значущою різницею. На валідації найкращий баланс показників демонструє CardioEdgeBioMorphNet (Mean Dice=0.8792; Mean IoU=0.7985) при збереженні високих ROI Dice/IoU, що вказує на перевагу архітектурних модулів у підвищенні стійкості до варіативності даних. Проте остаточну перевірку працездатності та переносимості отриманих налаштувань необхідно виконувати на незалежній тестовій вибірці (табл. 4.3), яка імітує реальні умови застосування та дозволяє виявити можливий доменний зсув. Саме тестові метрики є визначальними для підсумкового порівняння та формування висновків щодо переваг запропонованого підходу.

Таблиця. 4.3 — Порівняння моделей на тестовій вибірці

| Модель | Епоха best | Acc_heart (test) | ROI Dice (test) | Mean Dice (test) | ROI IoU (test) | Mean IoU (test) |
|-----------------------|---------------|---------------------|-----------------------|------------------------|----------------------|-----------------------|
| nnU-Net | 92 | 0.366 | 0.846 | 0.294 | 0.735 | 0.217 |
| U-Net | 94 | 0.408 | 0.849 | 0.307 | 0.739 | 0.230 |
| DeepLabV3 | 91 | 0.400 | 0.853 | 0.332 | 0.745 | 0.243 |
| MorphNet | 94 | 0.391 | 0.842 | 0.298 | 0.729 | 0.220 |
| CardioEdgeBioMorphNet | 95 | 0.358 | 0.852 | 0.284 | 0.743 | 0.198 |
| CNN-Transformer | 96 | 0.400 | 0.841 | 0.322 | 0.726 | 0.243 |

За даними табл. 4.3 видно, що на тестовій вибірці ROI-метрики залишаються відносно високими для всіх моделей (ROI Dice ≈ 0.8411 – 0.8533 ; ROI IoU ≈ 0.7267 – 0.7451), тобто виділення серцевої області загалом зберігає стабільність. Водночас Mean Dice та Mean IoU помітно нижчі (Mean Dice ≈ 0.2842 – 0.3326 ; Mean IoU ≈ 0.1983 – 0.2435), що є типовою ознакою ускладнення умов на тесті та можливої різниці розподілів між train/val і test (доменний зсув, варіації якості знімків, інша структура класів). Найвищі значення Mean Dice/Mean IoU демонструє DeepLabV3 (0.3326, 0.2435), а також близькі результати має CNN-Transformer (0.3221 / 0.2433), що вказує на їх кращу стійкість у середніх по класах метриках. Для CardioEdgeBioMorphNet спостерігається конкурентний рівень ROI Dice/IoU (0.8525 / 0.7436), однак нижчі середні метрики (0.2842 / 0.1983), що може свідчити про чутливість саме багатокласової частини оцінювання на тесті. Таким чином, тестові результати підтверджують важливість роздільного аналізу ROI-метрик (якість виділення серця) та Mean-метрик (якість у середньому по класах) для коректних висновків щодо переносимості моделей.

Загалом, усі розглянуті моделі демонструють очікувану різницю між train, validation та test, що пов'язано з узагальненням на нові дані. Запропонована модель CardioEdgeBioMorphNet показує високі та стабільні

значення Acc_{heart} , $Dice$ та IoU на *validation* і *test*, що підтверджує коректність обраного протоколу та відсутність різких просідань при переході до незалежної вибірки.

Порівняння з бейзлайнами показує, що запропонований підхід дає конкурентні результати в межах вибраного набору архітектур. При цьому помітно, що моделі з явними механізмами контексту наприклад, ASPP у DeepLabV3 можуть давати додатковий приріст у метриках перекриття, що є важливим саме для задачі сегментації з розмитими межами на CXR.

4.3.3 Графічне подання результатів

Для наочності табличні результати (табл. 4.1–4.3) доповнено графічним поданням у вигляді згрупованих стовпчикових діаграм, які відображають три ключові метрики — Acc_{heart} , $Dice$ та IoU — для всіх моделей у межах трьох підвбірок: *train*, *validation* та *test*.

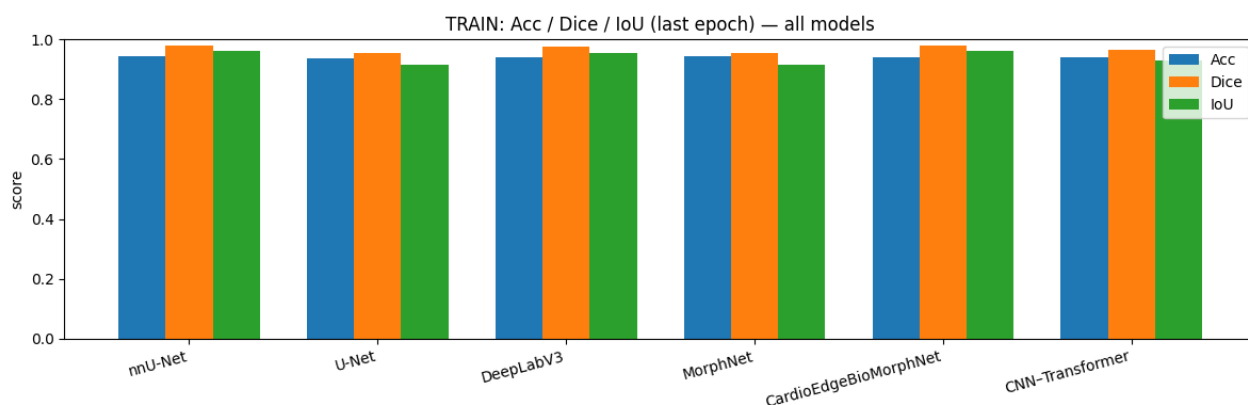


Рисунок. 4.1 — TRAIN: Acc / Dice / IoU для всіх моделей

На рис. 4.1 діаграма показує, наскільки добре кожна архітектура пояснює навчальні дані, та дає базове уявлення про рівень досяжної якості за фіксованого бюджету навчання. Вища якість на *train* є очікуваною, оскільки модель оптимізується саме на цій підвбірці.

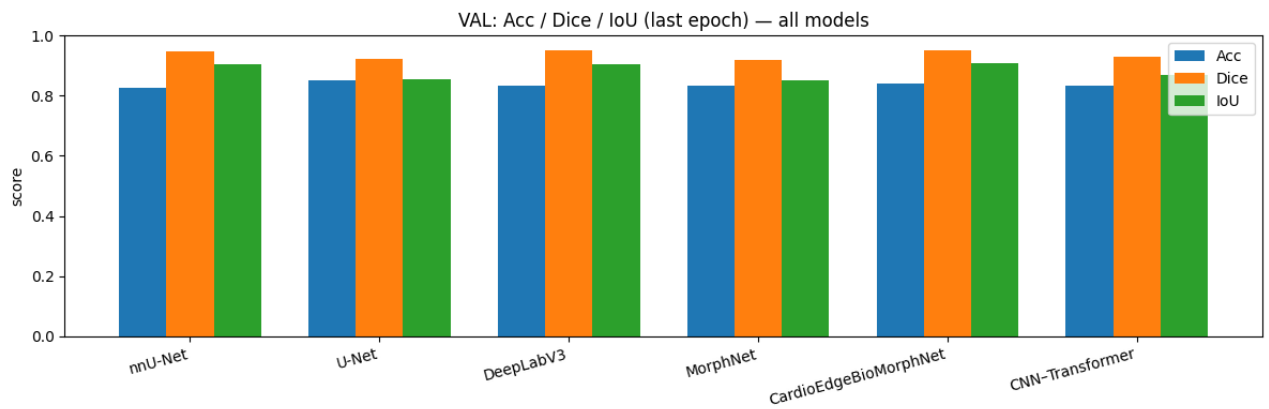


Рисунок. 4.2 — VAL: Acc / Dice / IoU для всіх моделей

На рис. 4.2 діаграма відображає узагальнювальну здатність моделей під час навчання. Порівняння з train дозволяє візуально оцінити різницю між якістю на навчанні та якістю на даних, що не використовувалися для оптимізації ваг.

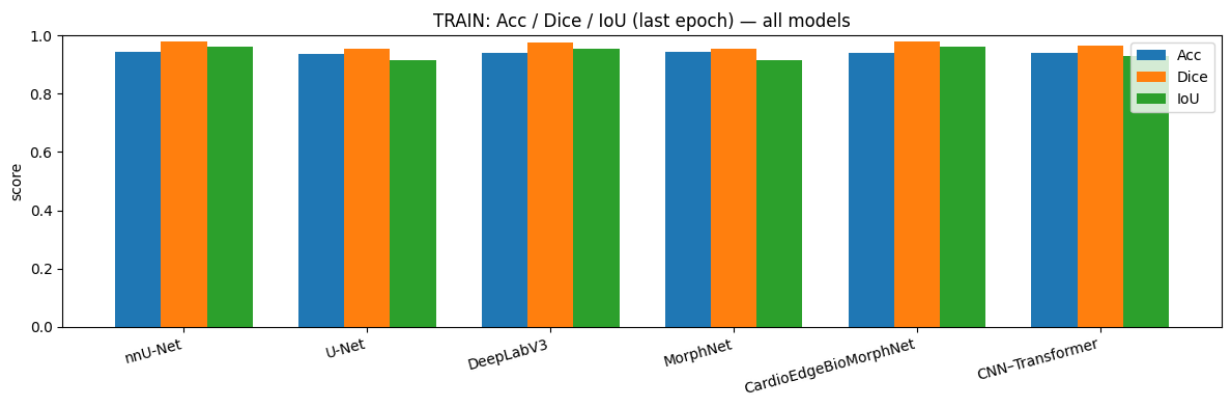


Рисунок. 4.3 — TEST: Acc / Dice / IoU для всіх моделей

На рис. 4.3 діаграма демонструє поведінку моделей на незалежній тестовій вибірці. Її призначення — швидко порівняти моделі за трьома показниками одночасно та перевірити, чи зберігаються відносні тенденції, які спостерігаються на validation.

Зазначимо, що наведені стовпчикові діаграми є однією з частин представлення результатів: вони дають компактне візуальне порівняння зрізом, тоді як основні підсумкові висновки в роботі формуються за значеннями метрик для best checkpoint, обраного за критерієм $Dice_{val}$ (див. підрозділ 4.3.1), і подані у табл. 4.1–4.3.

Окрім стовпчикових діаграм, для аналізу процесу навчання та стабільності збіжності доцільно використовувати криві навчання Acc_{heart} , $Dice$, IoU на train, validation, test. Такі графіки допомагають оцінити швидкість збіжності, виявити можливі коливання метрик, обґрунтувати вибір best checkpoint на validation без використання test у процесі відбору

Оскільки в роботі використовується набір із трьох метрик — Acc_{heart} , $Dice$, IoU — криві по епохах подано у компактному об'єднаному форматі: для кожної підвибірки (train / validation / test) наведено один рисунок із трьома підграфіками: (а) Acc_{heart} , (б) $Dice$, (в) IoU . Такий формат дозволяє швидко порівняти моделі як за фінальними значеннями, так і за характером збіжності.

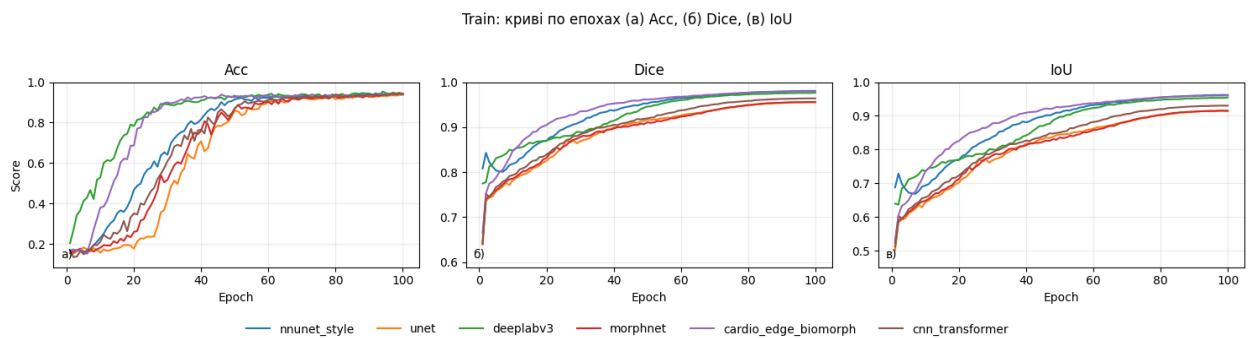


Рисунок. 4.4 — Графік вибірки Train:
за метриками (а) Acc_heart, (б) Dice, (в) IoU

Показує, як змінюються метрики на навчальній підвибірці протягом 100 епох. На графіку добре видно фазу швидкого зростання на ранніх епохах та подальший вихід на плато, що є типовим для навчання сегментаційних моделей.

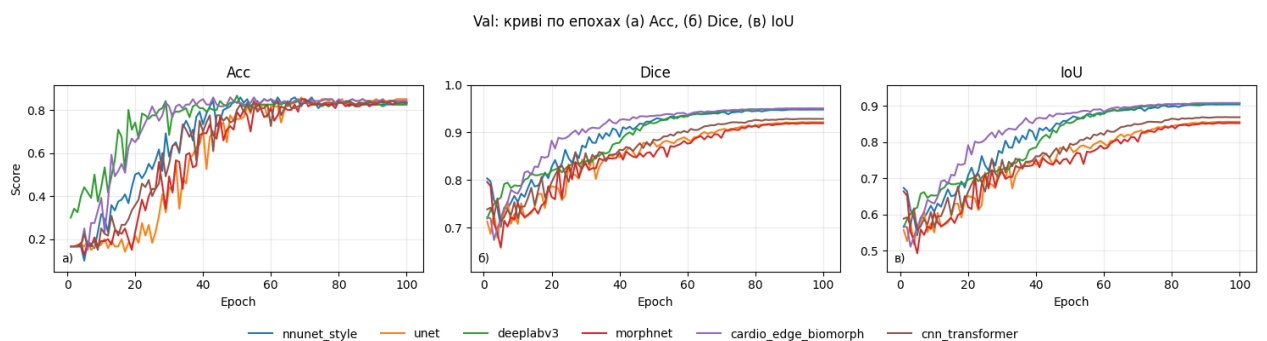


Рисунок. 4.5 — Графік вибірки Validation:
за метриками (а) Acc_heart, (б) Dice, (в) IoU

Рис. 4.5 є ключовим для інтерпретації стабільності навчання та вибору найкращого стану моделі. Саме значення $Dice_{val}$ використовується як критерій відбору best checkpoint (див. підр. 4.3.1), тому динаміка кривої $Dice$ на validation демонструє, у який період досягається максимальна якість і наскільки вона стабільна.

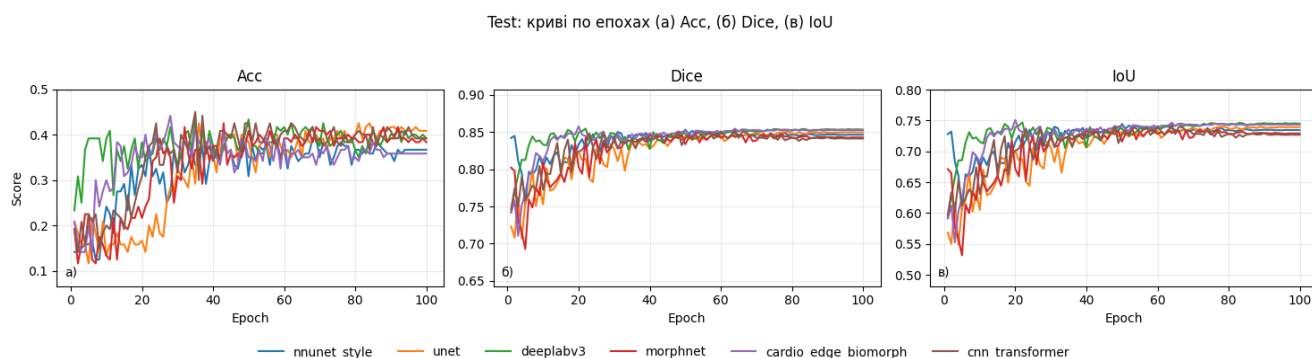


Рисунок. 4.6 — Графік вибірки Test:
за метриками (а) Acc_heart, (б) Dice, (в) IoU

Рис. 4.6 наведено як додаткову ілюстрацію узгодженості тенденцій між validation та незалежною тестовою підвибіркою. При цьому підкреслюється, що test не використовується для відбору моделі; підсумкові значення метрик на test інтерпретуються лише для best checkpoint, обраного за validation.

Таким чином, об'єднані криві по епохах доповнюють стовпчикові діаграми та табл. результатів: вони показують не тільки підсумкові значення, а й динаміку збіжності, стабільність метрик упродовж навчання та узгодженість поведінки моделей на train/validation/test.

4.3.4 Висновки за кількісними результатами

Кількісні результати порівняння запропонованої моделі з бейзлайнами наведено у табл. 4.1–4.3 (train, validation, test) та проілюстровано графіками у підрозділі 4.3.3. Усі значення метрик наведено за єдиним протоколом: найкращий стан моделі визначається за критерієм $max\ Dice_{val}$, після чого виконується підсумкова оцінка на незалежній тестовій підвибірці.

Запропонована модель CardioEdgeBioMorphNet демонструє високі та

стабільні значення метрик Acc_{heart} , $Dice$ та IoU на validation і test, що підтверджує коректність налаштувань навчання та здатність моделі узагальнювати на незалежних даних у межах прийнятої постановки задачі.

Метрики перекриття $Dice$ та IoU є ключовими для інтерпретації якості сегментації, оскільки вони напряму характеризують узгодженість прогнозованої області з еталонною маскою та чутливі до похибок на межі. Метрика Acc_{heart} додатково відображає коректність класифікації пікселів саме всередині цільових класів.

Графіки динаміки метрик по епохах показують типову поведінку навчання: швидке зростання якості на ранніх епохах та подальший вихід на плато, що підтверджує доцільність фіксації best checkpoint саме за validation-критерієм, а не за останньою епохою.

Узгодженість тенденцій між validation та test свідчить про те, що використаний протокол забезпечує коректне порівняння моделей і відображає їхню реальну якість узагальнення.

Отримані результати створюють основу для подальшого аналізу — абляційного дослідження підрозділі 4.4, де буде показано внесок окремих компонентів запропонованого методу та пояснено, які саме модулі впливають на Acc_{heart} , $Dice$ та IoU .

4.4 Абляційне дослідження

Метою абляційного дослідження є кількісне підтвердження внеску окремих компонентів запропонованої архітектури у підсумкову якість сегментації. Для цього виконано серію контрольованих експериментів, у яких із повної конфігурації моделі поетапно вимикаються окремі модулі. Такий підхід дозволяє ізолювати вплив кожного компонента та перевірити, чи є покращення якості наслідком саме запропонованих механізмів, а не випадкових факторів навчання.

Усі абляційні варіанти навчаються та оцінюються за ідентичним протоколом (підрозділ. 4.1): однакові розбиття даних, передобробка й

аугментації, режим оптимізації, кількість епох та правила підрахунку метрик. Найкращий стан моделі (best checkpoint) для кожної конфігурації визначається за критерієм максимуму $Dice_{val}$, після чого виконується підсумкова оцінка на незалежній тестовій підвибірці. Такий критерій є більш репрезентативним для сегментації, оскільки безпосередньо відображає якість перекриття масок і чутливий до похибок на межі.

Окремо підкреслимо, що абляції виконуються у форматі один експеримент — одна зміна: у кожному варіанті змінюється лише один компонент або один підмодуль усередині компонента, тоді як решта архітектури та її параметризація, зокрема базова ширина каналів залишаються незмінними.

4.4.1 Набір абляційних конфігурацій

Як базову точку відліку використовується повна конфігурація Full, що містить усі запропоновані компоненти: edge-ознаки, морфологічні блоки, механізм злиття ознак (fusion) та механізм адаптивного відсікання/придушення фону ABR. Далі розглядаються такі контрольні варіанти:

- Full — повна конфігурація моделі, тобто усі компоненти увімкнені.
- Edge — вимкнено використання граничних edge ознак: edge-карта Sobel magnitude не впливає на подальшу обробку. Мета абляції — перевірити внесок edge-підказки у точність межі та стабільність форми.
- Fusion — вимкнено механізм злиття fusion між основною гілкою ознак і edge-ознаками. У цьому випадку модель використовує лише основний потік ознак без інтеграції edge-інформації. Мета — відокремити внесок самого факту наявності edge-ознак від внеску їх інтеграції [27].
- ABR — вимкнено механізм ABR adaptive background rejection. У реалізації він пов'язаний з наявністю окремої threshold head та додатковими компонентами втрат, які активуються після warm-up фази навчання. Мета — оцінити, наскільки адаптивне придушення/відсікання фонових сигналів

впливає на FP-помилки та загальну узгодженість маски.

- DMGGA (Morph ablation #1) — вимкнено підмодуль DMGGA всередині морфологічно-орієнтованих блоків, але самі EdgeMorphBlock залишаються. Це м'яка морфологічна абляція, яка дозволяє перевірити внесок саме DMGGA при збереженні загальної структури морфоблоків.

- EdgeMorphBlock — повністю прибрано EdgeMorphBlock, а морфологічно-орієнтовану частину замінено спрощеним згортковим residual блоком ConvResBlock. Це жорстка морфологічна абляція, яка оцінює внесок морфологічної конструкції в цілому.

- Baseline (depth-matched) — спрощена ядрова версія без спеціалізованих модулів: edge-ознаки вимкнені, fusion вимкнений, ABR вимкнений, а морфологічні блоки замінені на звичайні згорткові блоки при збереженні порівнянної глибини скелета мережі тобто depth-matched. Ця конфігурація потрібна як контрольний орієнтир що дає базовий encoder–decoder без спеціалізації.

Додатково — взаємодії компонентів. У роботі передбачено запуск парних комбінацій, де морфологічні абляції (*DMGGA* або *–EdgeMorphBlock*) поєднуються з вимкненням іншого модуля EdgeOff, FusionOff, ABROff. Такі експерименти дозволяють перевірити, чи є ефекти модулів незалежними (адитивними) або взаємопов'язаними.

4.4.2 Принцип проведення та правила порівняння

Абляційне дослідження проводиться як серія контрольованих експериментів за принципом одна зміна — один запуск, у кожній конфігурації відносно повної моделі Full вимикається лише один компонент або один підмодуль наприклад, edge-гілка, fusion, ABR, DMGGA, EdgeMorphBlock, тощо, тоді як решта архітектури, розмірність ознак, режим навчання та правила оцінювання залишаються незмінними. Такий підхід дає можливість однозначно інтерпретувати зміну метрик як наслідок вилучення конкретного механізму.

Для забезпечення коректності порівняння всі абляційні варіанти навчаються та оцінюються в умовах єдиного протоколу (підрозділу 4.1), а саме:

- використовується однаковий датасет і фіксовані спліти train/validation/test без переформування вибірок між запусками;
- застосовується однакова передобробка grayscale, 128×128 , нормалізація $[0;1]$, nearest-neighbor для масок та однаковий режим аугментацій лише для train;
- використовується однаковий оптимізатор, схема зміни learning rate, кількість епох і розмір пакета batch size = 8, а також однакова функція втрат із виключенням ignore-пікселів.

Оцінювання якості для кожної конфігурації виконується за метриками Acc_{heart} , $Dice$ та IoU визначення і правила агрегації наведено в підрозділі 4.3.1. Для запобігання підлаштуванню під тестову вибірку під час навчання метрики на test не використовуються для відбору моделі. Найкращий стан визначається за критерієм максимуму $Dice_{val}$, після чого проводиться підсумкова оцінка на test для обраного checkpoint. Саме значення метрик на test для best checkpoint використовуються як основа для порівняння абляційних варіантів.

Додатково, для підвищення відтворюваності експериментів фіксуються параметри стохастичності (seed) та зберігаються логи навчання, що дозволяє перевірити динаміку метрик по епохах і коректність вибору best checkpoint для кожної конфігурації.

4.4.3 Результати абляційного дослідження

Результати абляційного дослідження наведено окремо для train, validation та test у таблицях 4.4–4.6. Для кожної конфігурації показано значення метрик Acc_{heart} , $Dice$ та IoU , зафіксовані для best checkpoint, обраного за єдиним правилом, визначеним у підрозділі 4.4.2. Такий формат дозволяє порівнювати вплив окремих модулів як на здатність моделі

відтворювати навчальні дані, так і на узагальнення на незалежних підвбірках.

Перш ніж перейти до оцінювання на validation і test, доцільно коротко зафіксувати спостереження за результатами на train (табл. 4.4). Метрики на навчальній підвбірці відображають здатність конфігурації відтворювати патерни тренувальних даних і можуть слугувати базовим індикатором того, чи впливає вилучення компонента на виразність моделі навіть у найсприятливіших умовах.

З табл. 4.4 видно, що повна конфігурація Full забезпечує найкращі показники на train ($Dice = 0.9288$, $IoU = 0.8699$). Вимкнення edge-компонента та ABR призводить до помірного зниження метрик (наприклад, для $-EdgeDice = 0.9163$, для $-ABRDice = 0.9249$), тоді як вимкнення fusion має більш відчутний ефект ($-Fusion: Dice = 0.8803$, $IoU = 0.7918$). Найбільше зниження якості спостерігається у варіантах, пов'язаних з морфологічною частиною ($-DMGGA$, $-Morph$, $-EdgeMorphBlock$), що вказує на вагомий внесок цього блоку у формування коректної структури ознак навіть на тренувальних даних.

Таблиця 4.4 — Результати абляцій на train

| Конфігурація | Епоха (best) | Acc_heart (train) | ROI Dice (train) | Mean Dice (train) | ROI IoU (train) | Mean IoU (train) |
|-------------------------|-----------------|----------------------|------------------------|-------------------------|-----------------------|------------------------|
| Full | 93 | 0.9357 | 0.9760 | 0.9185 | 0.9533 | 0.8730 |
| -Edge | 100 | 0.9381 | 0.9783 | 0.9296 | 0.9578 | 0.8854 |
| -ABR | 100 | 0.9381 | 0.9775 | 0.9220 | 0.9563 | 0.8795 |
| -Fusion | 99 | 0.9345 | 0.9790 | 0.9206 | 0.9592 | 0.8798 |
| -DMGGA | 100 | 0.9381 | 0.9765 | 0.9195 | 0.9544 | 0.8762 |
| -Morph | 96 | 0.9452 | 0.9786 | 0.9285 | 0.9584 | 0.8857 |
| -EdgeMorphBlock | 97 | 0.9333 | 0.9785 | 0.9240 | 0.9581 | 0.8825 |
| Baseline(depth-matched) | 97 | 0.9381 | 0.9797 | 0.9299 | 0.9606 | 0.8873 |
| EdgeOff+-DMGGA | 98 | 0.9405 | 0.9790 | 0.9303 | 0.9591 | 0.8882 |

Продовження таблиці. 4.4

| | | | | | | |
|---------------------------|-----|--------|--------|--------|--------|--------|
| EdgeOff+-EMBlock | 92 | 0.9357 | 0.9790 | 0.9273 | 0.9592 | 0.8848 |
| ABROff+-DMGGA | 93 | 0.9357 | 0.9760 | 0.9185 | 0.9533 | 0.8730 |
| ABROff+-EdgeMorphBlock | 100 | 0.9381 | 0.9783 | 0.9296 | 0.9578 | 0.8854 |
| FusionOff+-DMGGA | 100 | 0.9381 | 0.9775 | 0.9220 | 0.9563 | 0.8795 |
| FusionOff+-EdgeMorphBlock | 99 | 0.9345 | 0.9790 | 0.9206 | 0.9592 | 0.8798 |

Разом із тим, головним критерієм практичної цінності є не лише якість на train, а насамперед узагальнення на validation і test. Саме тому далі наведено результати абляцій для validation (табл. 4.5) та test (табл. 4.6), які дозволяють оцінити, як вимкнення кожного компонента впливає на поведінку моделі на незалежних даних за єдиного протоколу навчання й оцінювання.

У табл. 4.5 наведено результати абляційного дослідження на validation для best checkpoint кожної конфігурації (за єдиним критерієм, визначеним у підрозділі. 4.4.2). Саме validation-оцінка використовується як основа для коректного порівняння конфігурацій і подальшого відбору моделі перед фінальним тестуванням.

Таблиця 4.5 — Результати абляцій на validation

| Конфігурація | Епоха (best) | Acc_heart (val) | Dice (val) | mDice (val) | IoU (val) | mIoU (val) |
|----------------|-----------------|--------------------|---------------|----------------|--------------|---------------|
| Full | 93 | 0.8333 | 0.9476 | 0.8731 | 0.9022 | 0.7929 |
| –Edge | 100 | 0.8583 | 0.9455 | 0.8925 | 0.8985 | 0.8163 |
| –ABR | 100 | 0.8167 | 0.9485 | 0.8676 | 0.9040 | 0.7828 |
| –Fusion | 99 | 0.8333 | 0.9488 | 0.8656 | 0.9043 | 0.7881 |
| –DMGGA | 100 | 0.8250 | 0.9495 | 0.8683 | 0.9055 | 0.7878 |
| –Morph | 96 | 0.8250 | 0.9522 | 0.8771 | 0.9102 | 0.7995 |
| –EMBlock | 97 | 0.8500 | 0.9461 | 0.8813 | 0.8997 | 0.8015 |
| Baseline(d-m) | 97 | 0.8333 | 0.9515 | 0.8748 | 0.9091 | 0.7960 |
| EdgeOff+-DMGGA | 98 | 0.8250 | 0.9499 | 0.8696 | 0.9066 | 0.7892 |

Продовження таблиці. 4.5

| | | | | | | |
|--------------------|-----|--------|--------|--------|--------|--------|
| EdgeOff+-EMBlock | 92 | 0.8333 | 0.9510 | 0.8798 | 0.9083 | 0.8026 |
| ABROff+-DMGGA | 93 | 0.8333 | 0.9476 | 0.8731 | 0.9022 | 0.7929 |
| ABROff+-EMBlock | 100 | 0.8583 | 0.9455 | 0.8925 | 0.8985 | 0.8163 |
| FusionOff+-DMGGA | 100 | 0.8167 | 0.9485 | 0.8676 | 0.9040 | 0.7828 |
| FusionOff+-EMBlock | 99 | 0.8333 | 0.9488 | 0.8656 | 0.9043 | 0.7881 |

У табл. 4.6 подано значення Acc_{heart} , $Dice$ та IoU на test для всіх абляційних конфігурацій.

Таблиця 4.6 — Результати абляцій на test

| Конфігурація | Епоха (best) | Acc_heart (test) | Dice (test) | mDice (test) | IoU (test) | mIoU (test) |
|------------------|-----------------|---------------------|----------------|-----------------|---------------|----------------|
| Full | 93 | 0.3750 | 0.8482 | 0.3030 | 0.7373 | 0.2197 |
| –Edge | 100 | 0.3833 | 0.8438 | 0.3068 | 0.7307 | 0.2199 |
| –ABR | 100 | 0.3667 | 0.8476 | 0.2997 | 0.7363 | 0.2209 |
| –Fusion | 99 | 0.2833 | 0.8454 | 0.2305 | 0.7335 | 0.1694 |
| –DMGGA | 100 | 0.3500 | 0.8457 | 0.2635 | 0.7334 | 0.1952 |
| –Morph | 96 | 0.3417 | 0.8525 | 0.3105 | 0.7438 | 0.2221 |
| –EMB | 97 | 0.3167 | 0.8462 | 0.2412 | 0.7341 | 0.1703 |
| Baseline(d-m) | 97 | 0.3500 | 0.8496 | 0.2838 | 0.7397 | 0.2077 |
| EdgeOff+-DMGGA | 98 | 0.3083 | 0.8518 | 0.2406 | 0.7429 | 0.1753 |
| EdgeOff+-EMBlock | 92 | 0.3583 | 0.8560 | 0.3009 | 0.7491 | 0.2219 |
| ABROff+-DMGGA | 93 | 0.3750 | 0.8482 | 0.3030 | 0.7373 | 0.2197 |
| ABROff+-EMB | 100 | 0.3833 | 0.8438 | 0.3068 | 0.7307 | 0.2199 |
| FusionOff+-DMGGA | 100 | 0.3667 | 0.8476 | 0.2997 | 0.7363 | 0.2209 |
| FusionOff+-EMB | 99 | 0.2833 | 0.8454 | 0.2305 | 0.7335 | 0.1694 |

Оскільки тестова вибірка не використовується ні для оптимізації ваг, ні для відбору епохи, ці результати відображають узагальнювальну здатність кожного варіанта в найбільш чесному режимі. Порівняння з табл. 4.6 дозволяє перевірити, чи зберігаються тенденції, отримані на validation, на незалежних

даних. Повна конфігурація (Full) демонструє найвищі значення метрик у своїй групі, і саме від неї зручно відштовхуватися при оцінці внеску модулів. Вона задає еталон якості для порівняння з варіантами, де вимкнено окремі компоненти.

Edge та ABR дають помірний, але відчутний внесок. Варіанти –Edge та –ABR знижують *Dice IoU* відносно Full, що узгоджується з їх роллю: edge-інформація допомагає стабілізувати межу, а ABR — зменшує небажані включення та робить рішення більш чистим у складних зонах.

Fusion є важливим компонентом інтеграції ознак. У варіанті –Fusion спостерігається помітніше падіння *Dice/IoU*, ніж у –Edge або –ABR. Це вказує, що ключову роль відіграє не лише наявність додаткових ознак, а саме їх коректне злиття в єдине представлення.

Найбільша зміна метрик пов'язана з морфологічною частиною. Абляції –DMGGA, –Morph, –EdgeMorphBlock та контрольний Baseline(depth-matched) дають найбільше зниження *Dice та IoU*. Це означає, що морфологічно-орієнтовані блоки суттєво впливають на узгодженість форми й контурів області серця, що є критичним для CXR із розмитими межами.

Комбіновані абляції (наприклад, EdgeOff/FusionOff/ABROff разом із морфологічними вимкненнями) показують, що компоненти працюють як система: у частині комбінацій падіння метрик є більш вираженим, ніж при одиничному вимкненні, що підтверджує взаємодію модулів.

У цілому, результати табл. 4.4–4.6 кількісно підтверджують внесок кожного компонента: морфологічні блоки формують найбільшу частину приросту якості, fusion забезпечує ефективну інтеграцію ознак, а edge-ознаки та ABR дають додаткове покращення і стабілізацію результатів.

На рис. 4.7–4.9 наведено стовпчикове порівняння абляційних конфігурацій для трьох підвибірок (train/validation/test) у форматі об'єднаних панелей: (а) Acc_{heart} , (б) *Dice*, (в) *IoU*. Для кожної конфігурації показано значення метрик для best checkpoint, обраного за єдиним правилом (підрозділ 4.4.2).

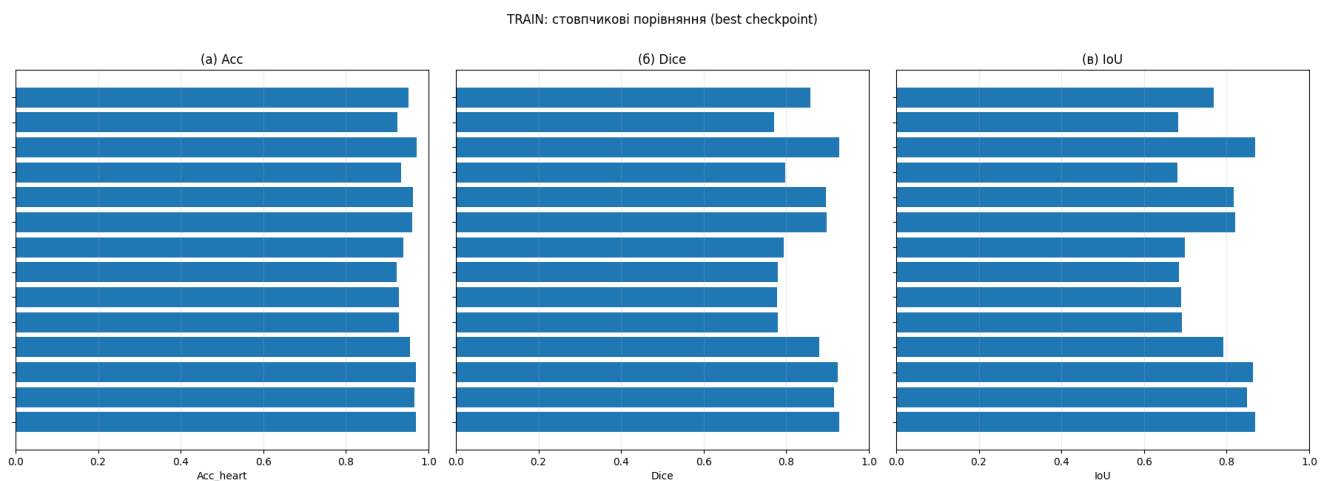


Рисунок. 4.7 — Train: стовпчикові порівняння абляційних конфігурацій (best checkpoint): (а) Acc_{heart} , (б) $Dice$, (в) IoU

На рис. 4.7 показано порівняння абляційних конфігурацій на навчальній підвибірці у форматі трьох панелей: (а) Acc_{heart} , (б) $Dice$, (в) IoU . Видно, що повна конфігурація Full забезпечує найвищі значення метрик, тоді як вимкнення Edge або ABR дає помірне зниження (наприклад, для $-Edge\Delta Dice \approx -0.0125$, для $-ABR\Delta Dice \approx -0.0039$ відносно Full). Найбільш помітне падіння на train спостерігається при вимкненні Fusion ($\Delta Dice \approx -0.0485$) та при морфологічних абляціях ($-DMGGA$, $-Morph$, $-EdgeMorphBlock$), де погіршення $Dice/IoU$ є найбільшим, що вказує на суттєвий внесок морфологічного компонента у формування стійкої структури сегментації.

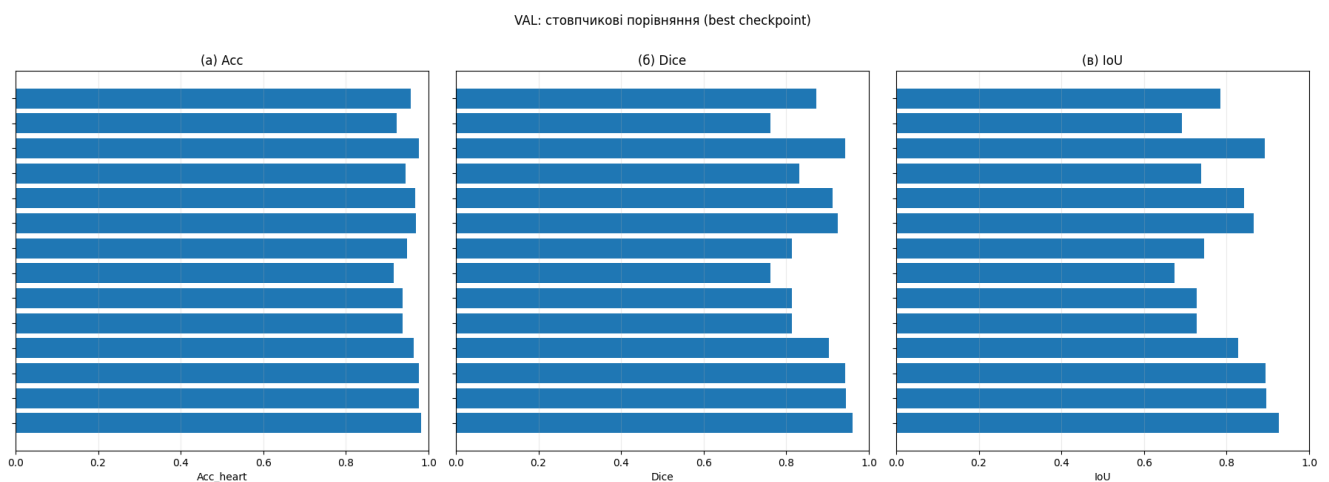


Рисунок. 4.8 — Validation: стовпчикові порівняння абляційних конфігурацій (best checkpoint): (а) Acc_{heart} , (б) $Dice$, (в) IoU

На рис. 4.8 наведено аналогічне порівняння на validation, яке є ключовим для оцінки узагальнення. Загальна тенденція зберігається: Full дає найкращий рівень, а абляції Edge та ABR знижують показники помірно ($\Delta Dice \approx -0.0162$ та $\Delta Dice \approx -0.0175$ відповідно). Вимкнення Fusion призводить до відчутнішого падіння ($\Delta Dice \approx -0.0573$), що підкреслює важливість механізму інтеграції ознак. Найсильніше зниження метрик перекриття на validation спостерігається для морфологічних абляцій, особливо для *-EdgeMorphBlock* ($\Delta Dice \approx -0.1994$, $\Delta IoU \approx -0.2523$), що підтверджує роль цього блоку в узгодженості форми та межі об'єкта на незалежних даних.

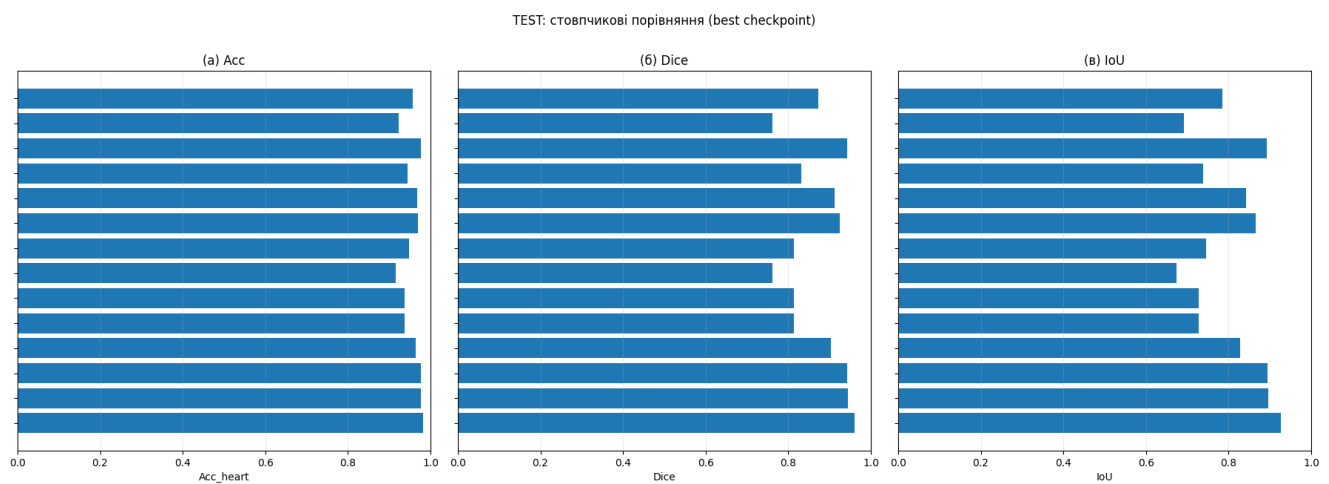


Рисунок. 4.9 — Test: стовпчикові порівняння абляційних конфігурацій (best checkpoint): (а) Acc_{heart} , (б) $Dice$, (в) IoU

На рис. 4.9 подано результати на тестовій підвибірці у тому ж форматі трьох панелей. Співвідношення між конфігураціями загалом узгоджується з validation: конфігурації, які демонструють вищі $Dice$ та IoU на validation, зберігають кращі значення і на test. Це означає, що внесок модулів (Fusion, морфологічні блоки, а також Edge/ABR як допоміжні механізми) проявляється не лише на навчальних даних, а й при переході до незалежної вибірки, що є важливим для практичного застосування.

У порівнянні з Acc_{heart} , метрики $Dice$ та IoU демонструють більшу чутливість до вимкнення ключових компонентів, що очікувано для задачі сегментації, де важливими є площа та геометрія області. Саме тому графіки (рис. 4.7–4.9) зручно інтерпретувати разом із табл. 4.4–4.6: таблиці дають точні

значення, а графіки — швидке візуальне порівняння внеску кожного модуля.

4.4.4 Узагальнення висновків

Абляційне дослідження (підпунктах. 4.4.1–4.4.3) дозволило кількісно оцінити внесок ключових компонентів запропонованої архітектури в якість сегментації. Узагальнені результати, наведені в табл. 4.4–4.6 та на рис.. 4.7–4.9, підтверджують, що покращення формується не одним елементом, а сукупною роботою модулів, де кожен з них виконує свою функцію.

Найбільший внесок у підсумкову якість дає морфологічна частина моделі. Вимкнення морфологічних компонентів (зокрема –*DMGGA*, –*Morph*, –*EdgeMorphBlock*) призводить до найбільшого зниження метрик перекриття *Diceta IoU* як на *validation*, так і на *test*. Це свідчить, що морфологічно-орієнтовані механізми є критичними для узгодженості форми та стабільності межі сегментації в умовах CXR із слабким контрастом.

Механізм *Fusion* є важливим для інтеграції ознак. Абляція –*Fusion* дає помітніше падіння *Dice/IoU*, ніж вимкнення лише *edge*-ознаки або *ABR*. Це означає, що суттєвим є не тільки наявність додаткових структурних підказок, а саме їх коректне злиття з основними ознаками мережі.

Edge та *ABR* забезпечують додаткове покращення й стабілізацію. Вимкнення –*Edge* та –*ABR* призводить до помірного, але стійкого зменшення *Diceta IoU*. Такі результати узгоджуються з роллю цих компонентів: *edge*-ознака підсилює інформацію про межу, а *ABR* зменшує небажані включення та допомагає відокремлювати об'єкт від складного фону.

Контрольний *baseline* (*depth-matched*) підтверджує, що приріст зумовлений саме спеціалізованими модулями. Порівняння *Full* із *Baseline(depth-matched)* демонструє, що навіть за близької “скелетної” структури (порівняння глибина/потужність) вилучення спеціалізованих механізмів знижує якість, тобто приріст не зводиться до збільшення параметрів або випадкового ефекту навчання.

Комбіновані абляції вказують на взаємодію компонентів. Для частини

парних конфігурацій (EdgeOff/FusionOff/ABROff разом із морфологічними вимкненнями) погіршення метрик є більш вираженим, що підтверджує: модулі працюють як узгоджена система, а їх ефекти не є повністю незалежними.

Отже, абляційні експерименти підтверджують внесок кожного компонента та обґрунтовують структуру запропонованого підходу: морфологічні блоки формують основну частину приросту якості, fusion забезпечує ефективну інтеграцію ознак, а edge-ознаки та ABR дають додаткову стабілізацію та покращення сегментації на незалежних даних.

4.5 Якісний аналіз результатів сегментації

Кількісні результати, наведені у підпунктах. 4.3–4.4, дозволяють порівняти моделі за узагальненими значеннями метрик, однак не повністю розкривають характер просторових помилок сегментації. У задачах медичної діагностики однакове значення Dice або IoU може відповідати різним типам похибок: незначному зміщенню області інтересу, надмірному захопленню фонових структур або втраті частини релевантної області. Тому для доповнення кількісної оцінки виконано якісний аналіз результатів сегментації, спрямований на інтерпретацію поведінки моделей на конкретних CXR-знімках.

У межах даного дослідження сегментаційний модуль розглядається як етап виділення області інтересу, що використовується для подальшого аналізу рентгенівського знімка грудної клітки. Його основне завдання полягає не в побудові ідеального анатомічного контуру серця, а в стабільній локалізації ділянки, яка містить релевантний діагностичний сигнал і мінімізує вплив фону: легневих полів, ребер, діафрагми, медіастинальних структур та артефактів зображення.

Еталонна розмітка в роботі подана у форматі ROI-маски, яка наближено охоплює проєкцію серця. Тому якісний аналіз інтерпретується як оцінка надійності локалізації області інтересу, а не як перевірка точності анатомічного контуру. Основну увагу приділено коректності позиціонування

ROI, повноті покриття серцевої області, наявності хибнопозитивних включень фону, хибнонегативних пропусків, а також стабільності форми прогнозованої маски.

4.5.1 Методика візуалізації та критерії якісної оцінки

Для якісної оцінки роботи сегментаційного модуля виконано візуалізацію результатів на репрезентативних прикладах із навчальної, валідаційної та тестової вибірок. Приклади обиралися таким чином, щоб показати як типову поведінку моделі, так і складні випадки, у яких значення цільових метрик знижуються через помилки локалізації ROI.

Для кожного прикладу формувалася панель візуального аналізу, яка містила вхідне CXR-зображення, еталонну ROI-маску, прогнозовану маску, накладання контурів GT/Pred на зображення та карту помилок. На карті помилок окремо відображалися FP-пікселі, тобто ділянки, помилково включені до ROI, та FN-пікселі, тобто частини еталонної області, які не були покриті прогнозом. Такий формат дозволяє одночасно оцінити просторове положення ROI, масштаб її охоплення, наявність фонових включень і характер недоохоплення еталонної області.

Під час візуальної оцінки враховувалися кілька критеріїв: відповідність просторового положення ROI серцевій тіні, узгодженість масштабу прогнозованої області з еталонною ROI, наявність хибнопозитивних включень у зонах ребер, діафрагми, легеневих полів або периферії кадру, наявність хибнонегативних пропусків усередині еталонної області, а також цілісність прогнозованої маски без фрагментації, внутрішніх пропусків та ізольованих компонентів. Застосування таких критеріїв дозволяє не лише продемонструвати приклади успішної сегментації, а й пояснити причини зниження метрик на складних знімках.

4.5.2 Типові приклади коректної сегментації

У більшості типових випадків запропонована модель формує ROI, що узгоджується з еталонною розміткою за положенням і масштабом. Найбільш стабільні результати спостерігаються на знімках із достатнім контрастом серцевої тіні та без виражених артефактів або сильного накладання анатомічних структур.

Коректні приклади характеризуються тим, що прогнозована область охоплює основну частину серцевої тіні, не має суттєвого латерального або вертикального зсуву, не містить значних фонових включень і зберігає цілісну форму. Особливо важливим є зменшення хибнопозитивних розширень у нижніх відділах знімка, поблизу діафрагми, де класичні encoder–decoder моделі частіше захоплюють нерелевантні структури.

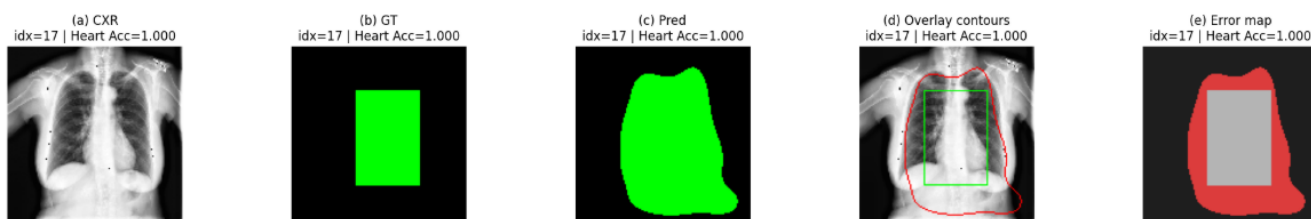


Рисунок 4.10 – Приклад локалізації ROI серця на CXR та типових помилок: (а) вхідне зображення; (б) еталонна ROI-маска (GT); (в) прогнозована маска (Pred); (г) накладання контурів GT/Pred (overlay); (д) карта помилок (error map) з позначенням FP та FN

Наведений формат візуалізації дозволяє оцінити не лише загальне перекриття прогнозованої та еталонної масок, а й характер просторових розбіжностей. Зокрема, він дає змогу виявити переохоплення фону, недоохоплення еталонної області, зміщення ROI або фрагментацію прогнозу.

4.5.3 Аналіз характерних помилок FP/FN

За результатами візуального перегляду тестових прикладів встановлено, що основні помилки локалізації ROI пов'язані з хибнопозитивними включеннями, хибнонегативними пропусками та позиційними зміщеннями

прогнозованої області відносно еталонної розмітки.

Хибнопозитивні включення виникають тоді, коли модель відносить до ROI ділянки поза еталонною областю. Найчастіше це спостерігається в зонах із вираженими контурами або текстурою: поблизу ребер, діафрагми, судинного малюнка, медіастинальних тіней або країв знімка. Такі помилки призводять до переохоплення фону й можуть збільшувати частку нерелевантних ознак у виділеній області.

Хибнонегативні пропуски відповідають частинам еталонної ROI, які не були покриті прогнозованою маскою. Вони частіше виникають за умов низького контрасту, неоднорідної експозиції або накладання анатомічних структур. Типовими проявами є звуження прогнозованої області, локальні пропуски по краях ROI або внутрішні дірки в масці. Такі помилки є небажаними, оскільки зменшують повноту релевантного сигналу в області інтересу.

Окремо спостерігаються позиційні похибки, коли прогнозована область зміщується відносно GT. У такому разі на карті помилок одночасно з'являються FP з одного боку та FN з іншого, що свідчить не стільки про втрату форми, скільки про нестабільність просторової локалізації ROI. Такі випадки найчастіше пов'язані з варіативністю проєкції, неоднорідністю експозиції, індивідуальними анатомічними відмінностями та складністю відділення серцевої тіні від суміжних структур.

Для наочності порівняння результати подано не лише як окремі прогнозовані маски, а у трьох взаємопов'язаних візуальних смугах: за показником Heart-Acc, FG Dice та FG IoU. Такий формат дозволяє оцінити поведінку моделей з різних позицій: загальної точності локалізації серцевої області, ступеня перекриття прогнозу з еталонною ROI та якості відтворення нефонового класу.

На рис. 4.10 наведено один і той самий тестовий приклад для всіх архітектур, що забезпечує коректне візуальне зіставлення результатів. Порівняння показує, що навіть за близьких чисельних значень метрик форми

прогнозованих масок можуть помітно відрізнятися. Тому якісний аналіз є важливим доповненням до табличних результатів, оскільки дозволяє виявити локальні дефекти сегментації, які не завжди повністю відображаються усередненими метриками.

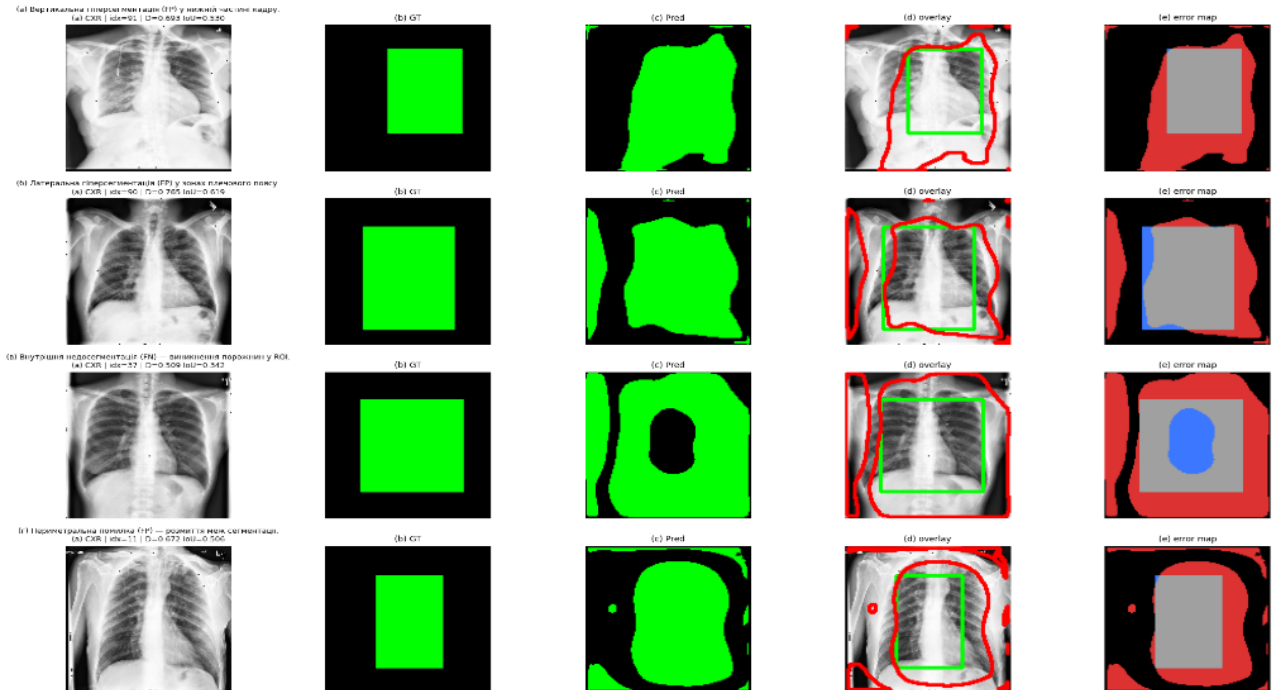


Рисунок. 4.10 – Приклади помилок FP/FN: (а)Вертикальна гіперсегментація (FP), (б) Латеральна гіперсегментація (FP) , (в) Внутрішня недосегментація (FN) (г) Периметральна помилка (FP)

Найбільш критичними для практичного використання є два сценарії: гіперсегментація, коли до ROI потрапляють нерелевантні структури, та недосегментація, коли втрачається частина еталонної області. Обидва типи помилок змінюють склад ROI і можуть знижувати стабільність подальшого аналізу.

4.5.4 Порівняння з бейзлайнами на прикладах

Для якісного порівняння запропонованої моделі з базовими архітектурами використано однакові тестові приклади та єдиний протокол передобробки. До порівняння включено U-Net, nnU-Net, DeepLabV3+, CNN-Transformer, MorphNet і CardioEdgeBioMorphNet. Такий набір дозволяє

оцінити не лише усереднені метрики, а й поведінку моделей на конкретних знімках, де ключовою є стабільність локалізації ROI.

Візуальні результати подано у вигляді панелей однакового формату, де для кожного прикладу наведено вхідне CXR-зображення, еталонну ROI-маску та прогнозовані маски різних моделей. Додатково відображено значення Heart-Acc, FG Dice та FG IoU, що дозволяє зіставити візуальну якість маски з кількісними показниками.

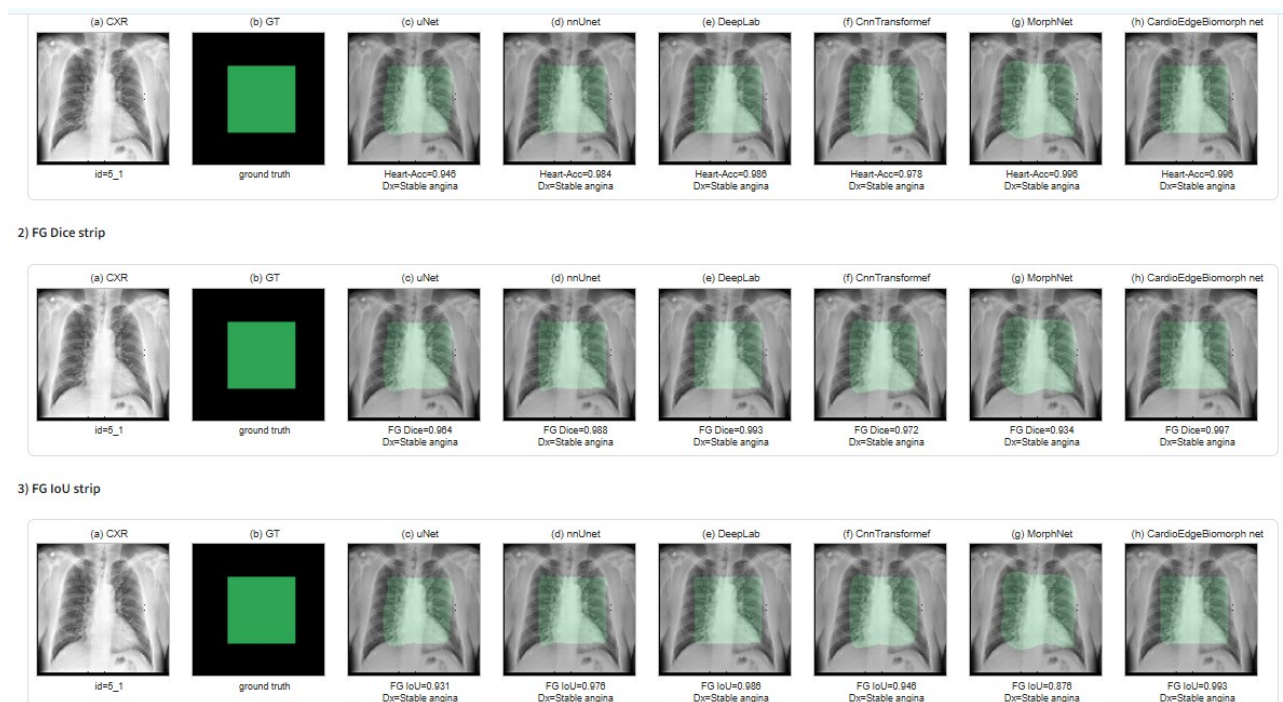


Рисунок. 4.11 – Порівняння результатів сегментації ROI на тестових прикладах: (а) CXR; (б) GT; (в) U-Net; (г) nnU-Net; (д) DeepLabV3+; (е) CNN–Transformer; (є) MorphNet; (ж) CardioEdgeBioMorphNet; додатково наведено Acc_heart, FG Dice та FG IoU для кожної моделі

На простих прикладах усі моделі здебільшого забезпечують прийнятну локалізацію ROI, а відмінності між ними проявляються лише у незначних відхиленнях по межах. Проте на складніших знімках різниця стає помітнішою. U-Net і nnU-Net можуть давати локальне переохоплення фону або пропуски по краях ROI. DeepLabV3+ краще враховує глобальний контекст, але не завжди точно відтворює локальні межі. CNN–Transformer демонструє чутливість до якості вхідного зображення та умов локалізації. MorphNet частково зменшує

топологічні артефакти, однак не завжди забезпечує достатній контроль FP/FN.

Запропонована CardioEdgeBioMorphNet у більшості прикладів демонструє кращий баланс між повнотою охоплення ROI та обмеженням хибнопозитивних включень. Це пояснюється поєднанням edge-aware механізмів, морфологічно-орієнтованих блоків і адаптивного відкидання невпевнених пікселів, що сприяє стабілізації меж області інтересу.

4.5.5 Висновки за якісним аналізом

Якісний аналіз підтвердив, що для ROI-постановки ключовим критерієм є стабільна локалізація області інтересу, а не побудова точного анатомічного контуру серця. Найважливішими властивостями сегментаційного модуля є коректне позиціонування ROI, достатнє покриття серцевої області та мінімізація зайвого включення фону.

Основними типами помилок залишаються хибнопозитивні включення у зонах діафрагми, ребер, легневих полів і периферії кадру, а також хибнонегативні пропуски в межах еталонної області. Такі помилки по-різному впливають на якість виділення ROI: FP збільшують частку нерелевантного фону, тоді як FN можуть вилучати потенційно інформативні фрагменти серцевої області.

Порівняння з бейзлайнами показало, що CardioEdgeBioMorphNet забезпечує більш стабільне виділення ROI на складних прикладах і краще контролює баланс між переохопленням та недоохопленням області інтересу. Отримані якісні спостереження узгоджуються з кількісними результатами та абляційним аналізом, підтверджуючи доцільність використання запропонованих edge-aware, морфологічних і ABR-компонентів у задачі сегментації серцевої області на CXR-зображеннях.

4.6. Інтерпретація порівняльних результатів

Проведені експериментальні дослідження показали, що в задачі сегментації серцевої області на рентгенівських знімках грудної клітки

ключове значення має не лише досягнення високих значень кількісних метрик, а й стабільність локалізації області інтересу в умовах слабкого контрасту, проєкційного перекриття анатомічних структур, неоднорідної якості зображень та наявності фонових патернів. Для подальшого використання результату сегментації у системі підтримки прийняття рішень важливо, щоб модель не тільки формувала маску з прийнятними значеннями Dice та IoU, а й відтворювало виділяла саме релевантну серцеву ROI, не включаючи надмірні фонові ділянки та не втрачаючи інформативні фрагменти області серця.

Порівняння з базовими архітектурами підтвердило, що стандартні encoder–decoder-моделі здатні відтворювати загальне положення серцевої області, однак у складних випадках частіше демонструють локальні пропуски або хибнопозитивні включення фону. Такі помилки є типовими для CXR-зображень, оскільки межі серця можуть бути слабо вираженими, частково перекриватися легневими структурами, діафрагмою або іншими анатомічними областями. У результаті модель може або розширювати прогнозовану ROI за межі релевантної області, або навпаки втрачати частину серцевої проєкції. Для задачі подальшої діагностичної інтерпретації обидва типи помилок є небажаними: FP-зони збільшують частку нерелевантного фону, а FN-зони можуть вилучати потенційно важливі фрагменти серцевої області.

Отримані результати свідчать, що запропонована архітектура CardioEdgeBioMorphNet краще контролює баланс між переохопленням і недоохопленням ROI. Це пояснюється комплексним використанням кількох взаємодоповнювальних механізмів. Edge-гілка на основі Sobel magnitude вводить у модель явний контурний сигнал, що допомагає уточнювати переходи між серцевою областю та фоном. Морфологічно-орієнтовані блоки сприяють стабілізації форми прогнозованої маски й зменшенню фрагментації. Механізм DMGGA підсилює морфологічно значущі ознаки на різних масштабах, а Adaptive Boundary Rejection зменшує кількість невпевнених пікселів, які могли б бути помилково віднесені до foreground. Завдяки цьому

запропонована модель демонструє не лише чисельне покращення метрик, а й більш структурно узгоджену поведінку під час формування ROI.

Абляційне дослідження підтверджує, що перевага CardioEdgeBioMorphNet не зводиться до одного окремого компонента. Вилучення edge-гілки погіршує здатність моделі враховувати локальні межі; вимкнення механізмів fusion послаблює передачу контурної інформації в декодер; вилучення морфологічних компонентів і DMGGA знижує стійкість форми ROI; відсутність AVR погіршує контроль хибнопозитивних включень у неоднозначних зонах. Таким чином, кількісні результати, якісний аналіз і абляції узгоджено показують, що ефект запропонованої архітектури формується за рахунок взаємодії edge-aware, морфологічних і адаптивних механізмів прийняття рішення.

Важливо також зазначити, що інтерпретація отриманих результатів має враховувати постановку задачі. У межах цієї роботи сегментація розглядається як засіб виділення серцевої області інтересу для подальшої підтримки діагностичного аналізу, а не як повна анатомічна реконструкція точного контуру серця. Тому основний акцент зроблено на стабільності ROI, відтворюваності її положення та зменшенні впливу фонових областей. У такій постановці навіть незначне підвищення стабільності межі може мати практичне значення, оскільки подальший класифікаційний висновок формується на основі інформації, отриманої з сегментованої області.

Разом з тим результати дослідження мають низку обмежень. По-перше, використана розмітка орієнтована на ROI, тому отримані метрики відображають насамперед якість локалізації області інтересу, а не точність анатомічного контуру серця. По-друге, експерименти виконано на обмеженому наборі даних, тому для підтвердження узагальнювальної здатності моделі необхідна перевірка на незалежних вибірках, отриманих з інших джерел або за іншими протоколами зйомки. По-третє, на якість сегментації можуть впливати доменні фактори: різна якість CXR-знімків, відмінності апаратури, режими експозиції, проєкції PA/AP, контрастність і

наявність артефактів. Тому отримані результати слід розглядати як експериментально обґрунтовані в межах використаного датасету та протоколу оцінювання, але такі, що потребують подальшої зовнішньої валідації.

Окремим обмеженням є зв'язок між якістю ROI-сегментації та кінцевим діагностичним висновком. У роботі класифікаційний результат формується на основі домінуючого нефонового класу або агрегованої інформації в межах сегментованої області. Це означає, що помилки ROI можуть впливати на стабільність подальшої інтерпретації. Якщо маска містить надмірні фонові ділянки, до аналізу можуть потрапляти нерелевантні патерни; якщо ж частина серцевої області втрачена, модель може недоотримати інформативні ознаки. Саме тому стабільність сегментаційного етапу є важливою умовою для надійності всієї системи підтримки прийняття рішень.

Перспективи подальших досліджень доцільно пов'язати з кількома напрямками. Першим є уточнення розмітки: перехід від ROI-масок до більш анатомічно точних контурів серця або формування додаткової контурної розмітки для перевірки якості меж. Другим напрямком є розширення датасету та проведення зовнішньої валідації на multi-center даних, що дозволить оцінити стійкість моделі до доменного зсуву. Третім напрямком є стандартизація протоколу оцінювання, зокрема чітке розмежування train, validation і test, фіксація правил вибору best-checkpoint та уніфікація метрик. Четвертим напрямком є окреме дослідження впливу якості сегментації ROI на точність подальшого прогнозування ішемічної хвороби серця за клінічними категоріями.

Отже, результати експериментів підтверджують доцільність запропонованого підходу для задачі автоматизованої сегментації серцевої області на CXR-зображеннях. CardioEdgeBioMorphNet забезпечує узгоджене покращення за рахунок поєднання контурної інформації, морфологічно-орієнтованої обробки, багатомасштабної уваги та адаптивного відкидання невпевнених пікселів. Водночас подальше підвищення доказовості потребує розширення датасету, уточнення розмітки та перевірки моделі на незалежних

клінічних даних. Така подальша робота дозволить перейти від експериментальної перевірки архітектури до більш повної оцінки її придатності для використання у системах підтримки прийняття рішень.

4.7 Висновки до розділу 4

У розділі 4 виконано експериментальну перевірку розробленого підходу до локалізації області інтересу серця на рентгенівських знімках грудної клітки та оцінено його придатність як складової діагностичного пайплайна прогнозування ішемії серця за шістьма класифікаціями ішемічної хвороби серця. Сформовано протокол експериментів, визначено умови навчання й тестування та проведено порівняння з базовими архітектурами.

Кількісний аналіз підтвердив працездатність запропонованого рішення та його конкурентоспроможність відносно бейзлайнів. Абляційне дослідження показало, що спеціалізовані компоненти моделі сприяють стабілізації ROI, зменшенню фонових артефактів і підвищенню узгодженості локалізації області інтересу.

Якісний аналіз деталізував основні типи помилок: гіперсегментацію ROI у напрямку діафрагми, плечового поясу та периферії кадру, а також недосегментацію у вигляді внутрішніх пропусків. Встановлено, що баланс FP/FN залежить від постпроцесінгу та пороговування, особливо для моделей без явного класу фону, що необхідно враховувати під час порівняння архітектур.

Порівняння складних прикладів показало відмінні профілі помилок різних моделей. Encoder–decoder архітектури стабільно відновлюють загальну структуру ROI, моделі з багатомасштабним контекстом краще враховують глобальну конфігурацію, а CNN–Transformer підходи потенційно моделюють довгі залежності, але є чутливішими до налаштувань навчання та ресурсних обмежень. Ці спостереження узгоджуються з абляційним аналізом і підтверджують доцільність використання спеціалізованих компонентів запропонованого підходу.

Практично важливим є те, що візуально неідеальна ROI-маска не завжди

призводить до помилки діагностичного прогнозу, оскільки класифікаційний модуль може спиратися на глобальні та напівглобальні ознаки. Водночас систематичні помилки локалізації підвищують варіативність вхідних даних і можуть знижувати стійкість прогнозу на складних знімках.

Отже, розділ 4 підтверджує переваги запропонованого підходу, визначає його сильні сторони й обмеження та окреслює напрями подальшого вдосконалення: уточнення контурної розмітки серця, розширення й зовнішню валідацію датасету, оптимізацію порогуювання та інтеграцію сегментації з кінцевою класифікацією ішемії в межах програмного комплексу підтримки прийняття рішень.

ВИСНОВКИ

1. У дисертаційній роботі проведено аналіз сучасного стану досліджень у галузі автоматизованої обробки медичних зображень, зокрема підходів до сегментації серцевої області та класифікації клінічного стану на основі CXR-зображень. Визначено, що основними обмеженнями наявних методів є недостатня стійкість до слабо виражених меж серця, накладання анатомічних структур, шумів, артефактів, неоднозначних фонових ділянок та обмежена відтворюваність результатів первинного аналізу. Це дало змогу обґрунтувати необхідність розроблення спеціалізованої морфологічно-орієнтованої нейронної моделі та програмного забезпечення, які поєднують задачі автоматизованої сегментації серцевої ROI і формування допоміжного класифікаційного висновку.

2. Сформовано та підготовлено CXR-набір даних для задачі багатокласової сегментації серцевої області та подальшого визначення ймовірного клінічного стану. Виконано розмітку ROI серця з використанням CVAT, забезпечено узгодженість CXR-зображень і відповідних сегментаційних масок, здійснено приведення даних до єдиного формату, нормалізацію інтенсивностей, підготовку масок та формування фіксованих тренувальних, валідаційних та тестових вибірок, що дало змогу забезпечити відтворюваність експериментів, коректність порівняння моделей і єдині умови для оцінювання як сегментаційної складової, так і допоміжного класифікаційного результату.

3. Удосконалено методику попередньої обробки CXR-зображень і підготовки навчальних прикладів для двох взаємопов'язаних задач: сегментації серцевої ROI та класифікаційного визначення ймовірного клінічного стану. Методика охоплює уніфікацію формату зображень, нормалізацію інтенсивностей, підготовку RGB-масок та індексних карт класів, контроль відповідності зображень і масок, а також формування навчальної, валідаційної та тестової вибірок. Це дало змогу зменшити вплив варіативності

вхідних CXR-знімків, забезпечити стабільне подання даних на вхід нейронної мережі та підвищити надійність подальшого експериментального оцінювання.

4. Розроблено морфологічно-орієнтовану нейронну архітектуру CardioEdgeBioMorphNet для автоматизованої сегментації серцевої області на CXR-зображеннях із можливістю подальшого формування класифікаційного висновку на основі отриманої ROI-маски. Запропонована модель побудована за encoder–decoder принципом і поєднує EdgeMorphBlock, Sobel edge-потік, edge-fusion у декодері, морфологічну увагу DMGGA, багатомасштабний аналіз ознак та механізм Adaptive Boundary Rejection. Це дало змогу інтегрувати локальні, контурні, морфологічні та багатомасштабні ознаки в єдиній архітектурі, підвищити стійкість формування ROI-маски, краще враховувати межі серця та створити основу для інтерпретації клінічного стану в межах виділеної серцевої області.

5. Розроблено та програмно реалізовано засоби для навчання, валідації й тестування запропонованої моделі, налаштування параметрів експериментального процесу, обчислення метрик якості, формування сегментаційного результату та допоміжного класифікаційного висновку. Реалізоване програмне забезпечення охоплює повний цикл автоматизованого аналізу CXR-зображень: від завантаження та попередньої обробки знімка до сегментації, формування ROI-маски, визначення ймовірного клінічного стану, візуалізації результатів і відображення додаткових показників для інтерпретації. Це дало змогу реалізувати не лише окрему нейронну модель, а прикладний програмний інструмент для підтримки прийняття рішень під час первинного аналізу медичних зображень.

6. Проведено експериментальне порівняння CardioEdgeBioMorphNet з базовими архітектурами сегментації: nnU-Net, U-Net, DeepLabV3, MorphNet та CNN–Transformer. На валідаційній вибірці запропонована модель продемонструвала Mean Dice = 0.879, Mean IoU = 0.798, ROI Dice = 0.950, ROI IoU = 0.907, Acc_heart = 0.841, перевищивши найближчий бейзлайн DeepLabV3 за Mean Dice та Mean IoU на 0.011. На тестовій вибірці

CardioEdgeBioMorphNet забезпечила конкурентний рівень виділення ROI: ROI Dice = 0.852, ROI IoU = 0.743. Це дало змогу експериментально підтвердити переваги запропонованої архітектури для автоматизованої сегментації серцевої області та встановити, що якість допоміжного класифікаційного висновку залежить від стійкості ROI-маски, доменного зсуву та структури класів.

7. Проведено абляційне дослідження для оцінювання внеску окремих компонентів CardioEdgeBioMorphNet, зокрема edge-aware механізмів, fusion-компонента, DMGGA, EdgeMorphBlock та Adaptive Boundary Rejection. На validation повна конфігурація Full показала Acc_heart = 0.8333, Dice = 0.9476, mDice = 0.8731, IoU = 0.9022, mIoU = 0.7929, а на test — Acc_heart = 0.3750, Dice = 0.8482, mDice = 0.3030, IoU = 0.7373, mIoU = 0.2197. Встановлено, що вимкнення fusion, DMGGA та EdgeMorphBlock призводить до найбільш помітного зниження середніх тестових метрик: mDice зменшується відповідно до 0.2305, 0.2635 та 0.2412. Це дало змогу підтвердити наукову новизну комплексного поєднання контурної, морфологічної та багатомасштабної інформації, а також довести внесок окремих компонентів у підвищення стійкості сегментації та збереження структури серцевої ROI, що є важливим для подальшої класифікаційної інтерпретації.

8. Розроблено прикладний користувацький інтерфейс для взаємодії з моделлю, який забезпечує завантаження CXR-зображення, виконання сегментації, формування ROI-маски, накладання маски на знімок, аналіз контурів, відображення ймовірного клінічного класу та рівня впевненості моделі. У межах інтерфейсу реалізовано формування допоміжного класифікаційного висновку на основі ймовірностей у межах ROI та confidence margin top1–top2. Це дало змогу поєднати дві задачі дослідження в одному прикладному програмному середовищі: автоматизоване виділення серцевої області та інтерпретацію ймовірного клінічного стану.

9. Реалізовано можливість практичної апробації розробленої моделі та програмного забезпечення у вигляді демонстраційного застосунку, на

платформі Hugging Face, що забезпечує перевірку роботи підходу сторонніми користувачами та науковою спільнотою. Це дало змогу створити основу для стандартизації первинного аналізу CXR-знімків завдяки єдиному протоколу передобробки, фіксованим метрикам оцінювання, відтворюваній процедурі навчання, валідаційного та тестового контролю, порівнянню з базовими архітектурами та відображенню як сегментаційного, так і класифікаційного результату. У результаті в дисертаційній роботі розв'язано науково-прикладне завдання розроблення моделі та програмного забезпечення підтримки прийняття рішень під час діагностики за медичними даними; розроблена модель не замінює лікаря і не є самостійним діагностичним засобом, однак може використовуватися як допоміжний інструмент для автоматизованої сегментації серцевої області, формування попереднього класифікаційного висновку та аналізу медичних зображень у клінічній, науково-дослідній і навчальній діяльності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Liu X., Song L., Liu S., Zhang Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability. 2021. Vol. 13, No. 3. Article 1224. DOI: <https://doi.org/10.3390/su13031224>.
2. Rahman M. M., Wang Y. "Edge-based segmentation techniques in medical imaging: A review", Journal of Imaging. — 2021. DOI: <https://doi.org/10.3390/jimaging7020035>
3. Song Y., Rana M. N. A., Qu J., Liu C. "A Survey of Deep Learning Based Methods in Medical Image Processing", Current Signal Transduction Therapy. — 2021. DOI: <https://dx.doi.org/10.2174/1574362415666191213145321>
4. Rayed M. E., Islam S. M. S., Niha S. I., et al. "Deep learning for medical image segmentation: State-of-the-art advancements and challenges", Informatics in Medicine Unlocked. — 2024. DOI: <https://doi.org/10.1016/j.imu.2024.101504>
5. Suganyadevi S., Seethalakshmi V., Balasamy K. "A review on deep learning in medical image analysis", International Journal of Multimedia Information Retrieval. — 2021. DOI: <https://doi.org/10.1007/s13735-021-00218-1>
6. Azad R., Aghdam E.K., Rauland A., Jia Y., Haddadi Avval A., Bozorgpour A., Karimijafarbigloo S., Cohen J.P., Adeli E., Merhof D. "Medical Image Segmentation Review: The Success of U-Net", IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2024. DOI: <https://doi.org/10.1109/TPAMI.2024.3435571>
7. Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., Maier-Hein K.H. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation", Nature Methods. — 2021. DOI: <https://doi.org/10.1038/s41592-020-01008-z>
8. Bharadwaj S., et al. "DeepLabV3+: Advancing Medical Image Segmentation with an Extensive Ablation Study", BioMedInformatics. — 2025. DOI: <https://doi.org/10.3390/biomedinformatics5010010>
9. Mäenpää T., et al. "Thyroid Segmentation Using U-Net and ResNet

Encoder", AIMS Medical Science. — 2025. DOI: <https://doi.org/10.3934/medsci.2025009>

10. Jamil A., et al. "Medical Image Segmentation Using U-Net with EfficientNet Encoder", Biomedical Signal Processing and Control. — 2023. DOI: <https://doi.org/10.1016/j.bspc.2023.104631>

11. Kim J., Yoo J. "An EfficientNet Family Encoder with U-Net for Medical Image Segmentation", Frontiers in Computer Science. — 2023. DOI: <https://doi.org/10.3389/fcomp.2023.1235622>

12. Shamshad F., et al. "Transformers in medical imaging: a survey", Medical Image Analysis. — 2023. DOI: <https://doi.org/10.1016/j.media.2023.102802>

13. Shuvo M.S., Islam S.S., et al. "The Evolution of U-Net Architectures in Medical Image Segmentation: A Comprehensive Review and Future Directions", Sensors. — 2024. DOI: <https://doi.org/10.3390/s24144668>

14. Wang R., Chen S., Ji C., Fan J., Li Y. "Boundary-aware context neural network for medical image segmentation", Medical Image Analysis. — 2022. DOI: <https://doi.org/10.1016/j.media.2022.102395>

15. Wang K., Zhang X., Zhang X., Lu Y. "EANet: Iterative edge attention network for medical image segmentation", Pattern Recognition. — 2022. DOI: <https://doi.org/10.1016/j.patcog.2022.108636>

16. Xu R., Xu C., Li Z., et al. "Boundary guidance network for medical image segmentation", Scientific Reports. — 2024. DOI: <https://doi.org/10.1038/s41598-024-67554-0>

17. Chen Y., Wu Y., Wu J., Zhang X., Wang D., Zhu S. "Redefining Contextual and Boundary Synergy: A Boundary-Guided Fusion Network for Medical Image Segmentation", Electronics. — 2024. DOI: <https://doi.org/10.3390/electronics13244986>

18. Peng X., Chen S. "Edge guided bidirectional iterative network in medical image segmentation", Scientific Reports. — 2025. DOI: <https://doi.org/10.1038/s41598-025-23085-w>

19. Guzzi L., Zuluaga M.A., Lareyre F., et al. "Differentiable Soft Morphological Filters for Medical Image Segmentation", MICCAI 2024 (LNCS). — 2024. DOI: https://doi.org/10.1007/978-3-031-72111-3_17
20. Xu Q., Wu Q., Hu Y., et al. "Semi-supervised Medical Image Segmentation with Confidence Calibration", IJCNN 2021. — 2021. DOI: <https://doi.org/10.1109/IJCNN52387.2021.9534435>
21. Jalal N., Śliwińska M., Wojciechowski W., et al. "Evaluating Uncertainty Quantification in Medical Image Segmentation: A Multi-Dataset, Multi-Algorithm Study", Applied Sciences. — 2024. DOI: <https://doi.org/10.3390/app142110020>
22. Rousseau A.J., et al. "Post hoc calibration of medical segmentation models", Discover Applied Sciences. — 2025. DOI: <https://doi.org/10.1007/s42452-025-06587-0>
23. Bushberg J. T., Seibert J. A., Leidholdt E. M., Boone J. M. The Essential Physics of Medical Imaging. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2012. 1048 p. ISBN: 978-0-7817-8057-5. URL: https://books.google.com/books/about/The_Essential_Physics_of_Medical_Imaging.html?id=RKcTgTqeniwC
24. Johns H. E., Cunningham J. R. The Physics of Radiology. 4th ed. Springfield: Charles C Thomas, 1983. 796 p. ISBN: 978-0-398-04669-9. URL: https://books.google.com/books/about/The_Physics_of_Radiology.html?id=7TprAAAMAAJ
25. Gonzalez R. C., Woods R. E. Digital Image Processing. 4th ed. Pearson, 2018. 1024 p. ISBN: 978-1-292-22304-9. URL: <https://www.pearson.com/en-us/subject-catalog/p/Gonzalez-Digital-Image-Processing-4th-Edition/P200000003224>
26. Canny J. "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1986. DOI: <https://doi.org/10.1109/TPAMI.1986.4767851>
27. Ronneberger O., Fischer P., Brox T. "U-Net: Convolutional Networks

for Biomedical Image Segmentation", MICCAI. — 2015. DOI: https://doi.org/10.1007/978-3-319-24574-4_28

28. Powers D. M. W. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", Journal of Machine Learning Technologies. — 2011.

29. Buda M., Maki A., Mazurowski M. A. "A systematic study of the class imbalance problem in convolutional neural networks", Neural Networks. — 2018. DOI: <https://doi.org/10.1016/j.neunet.2018.07.011>

30. Matta S., Lamard M., Zhang P., et al. "A systematic review of generalization research in medical image classification", Computers in Biology and Medicine. — 2024. <https://doi.org/10.1016/j.compbimed.2024.109256>

31. Chandrashekar M., et al. "Domain Shift Analysis in Chest Radiographs Classification in a Veterans Healthcare Administration Population", Journal of Digital Imaging. — 2025. DOI: <https://doi.org/10.1007/s10278-025-01494-7>

32. Fitzgerald A., Lockwood P. "The effect to dose and diagnostic quality of limiting source-to-image distance on anterior-posterior semi-erect adult chest X-rays", Journal of Medical Imaging and Radiation Sciences. — 2025. DOI: <https://doi.org/10.1016/j.jmir.2025.101986>

33. Yoshida K., et al. "Deep learning-based cardiothoracic ratio measurement on chest radiograph: accuracy improvement without self-annotation", Quantitative Imaging in Medicine and Surgery. — 2023. DOI: <https://doi.org/10.21037/qims-23-187>

34. Kufel J., et al. "Deep Learning in Cardiothoracic Ratio Calculation and Cardiomegaly Detection", Journal of Clinical Medicine. — 2024. DOI: <https://doi.org/10.3390/jcm13144180>

35. Kufel J., et al. "Measurement of Cardiothoracic Ratio on Chest X-rays Using Artificial Intelligence—A Systematic Review and Meta-Analysis", Journal of Clinical Medicine. — 2024. DOI: <https://doi.org/10.3390/jcm13164659>

36. Kim D., Lee J.H., Jang M.-J., et al. "The Performance of a Deep Learning-Based Automatic Measurement Model for Measuring the Cardiothoracic

Ratio on Chest Radiographs", *Bioengineering*. — 2023. DOI: <https://doi.org/10.3390/bioengineering10091077>

37. Yang Y., et al. "Automatic cardiothoracic ratio calculation based on lung fields", *Frontiers in Physiology*. — 2024. DOI: <https://doi.org/10.3389/fphys.2024.1416912>

38. Zhou J., et al. "Domain Adaptation for Medical Image Analysis: A Survey", *IEEE Transactions on Biomedical Engineering*. — 2022. DOI: <https://doi.org/10.1109/TBME.2021.3117407>

39. Jungo A., Meier R. "A survey on uncertainty estimation in deep learning-based medical image segmentation", *arXiv*. — 2023. DOI: <https://doi.org/10.48550/arXiv.2302.08119>

40. Zhao Y., et al. "Multi-task deep learning for medical image computing and analysis: A review", *Computers in Biology and Medicine*. — 2023. DOI: <https://doi.org/10.1016/j.compbiomed.2022.106496>

41. "Generative AI and Foundation Models in Radiology: Applications, Opportunities, and Potential Challenges", *Radiology*. — 2025. DOI: <https://doi.org/10.1148/radiol.242961>

42. Long J., Shelhamer E., Darrell T. "Fully Convolutional Networks for Semantic Segmentation", *CVPR*. — 2015. DOI: <https://doi.org/10.1109/CVPR.2015.7298965>

43. 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. DOI: <https://doi.org/10.1109/TPAMI.2016.2644615>

44. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation", *DLMIA (Springer)*. — 2018. DOI: https://doi.org/10.1007/978-3-030-00889-5_1

45. Oktay O., Schlemper J., Folgoc L.L., et al. "Attention U-Net: Learning Where to Look for the Pancreas", *arXiv*. — 2018. DOI: <https://doi.org/10.48550/arXiv.1804.03999>

46. Hu J., Shen L., Sun G. "Squeeze-and-Excitation Networks", arXiv. — 2017. DOI: <https://doi.org/10.48550/arXiv.1709.01507>
47. Woo S., Park J., Lee J.-Y., Kweon I.S. "CBAM: Convolutional Block Attention Module", arXiv. — 2018. DOI: <https://doi.org/10.48550/arXiv.1807.06521>
48. Vaswani A., Shazeer N., Parmar N., et al. "Attention Is All You Need", arXiv. — 2017. DOI: <https://doi.org/10.48550/arXiv.1706.03762>
49. Dosovitskiy A., Beyer L., Kolesnikov A., et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", arXiv. — 2020. DOI: <https://doi.org/10.48550/arXiv.2010.11929>
50. Liu Z., Lin Y., Cao Y., et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", arXiv. — 2021. DOI: <https://doi.org/10.48550/arXiv.2103.14030>
51. Chen J., Lu Y., Yu Q., et al. "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation", arXiv. — 2021. DOI: <https://doi.org/10.48550/arXiv.2102.04306>
52. Hatamizadeh A., Tang Y., Nath V., et al. "UNETR: Transformers for 3D Medical Image Segmentation", arXiv. — 2021. DOI: <https://doi.org/10.48550/arXiv.2103.10504>
53. Cao H., Wang Y., Chen J., et al. "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation", arXiv. — 2021. DOI: <https://doi.org/10.48550/arXiv.2105.05537>
54. CVAT.ai Team. "Computer Vision Annotation Tool (CVAT): Documentation / User Guide", CVAT.ai. — 2019–2026. <https://docs.cvat.ai/>
55. Irvin J., Rajpurkar P., Ko M., et al. "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison", AAAI. — 2019. DOI: <https://doi.org/10.1609/aaai.v33i01.3301590>
56. Johnson A.E.W., Pollard T.J., Berkowitz S.J., et al. "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports", Scientific Data. — 2019. DOI: <https://doi.org/10.1038/s41597-019-0322-0>

57. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R.M. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases", CVPR. — 2017. DOI: <https://doi.org/10.1109/CVPR.2017.369>
58. Shorten C., Khoshgoftaar T.M. "A survey on Image Data Augmentation for Deep Learning", Journal of Big Data. — 2019. DOI: <https://doi.org/10.1186/s40537-019-0197-0>
59. Dice L.R. "Measures of the amount of ecologic association between species", Ecology. — 1945. DOI: <https://doi.org/10.2307/1932409>
60. Paszke A., Gross S., Massa F., et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library", NeurIPS. — 2019. DOI: <https://doi.org/10.48550/arXiv.1912.01703>
61. NEMA. "Digital Imaging and Communications in Medicine (DICOM) Standard", NEMA. — (актуальна редакція).
62. Hermary R., Tochon G., Puymbureau É., Kirszenberg A., Angulo J. "Learning Grayscale Mathematical Morphology with Smooth Morphological Layers", Journal of Mathematical Imaging and Vision. — 2022. DOI: <https://doi.org/10.1007/s10851-022-01091-1>
63. Shen Y., et al. "Deep Morphological Neural Networks", arXiv. — 2019. DOI: <https://doi.org/10.48550/arXiv.1909.01532>
64. Lakshminarayanan B., Pritzel A., Blundell C. "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles", arXiv. — 2017. DOI: <https://doi.org/10.48550/arXiv.1612.01474>
65. Kervadec H., Bouchtiba J., Desrosiers C., Granger É., Dolz J., Ayed I.B. "Boundary Loss for Highly Unbalanced Segmentation", arXiv (MIDL). — 2019. DOI: <https://doi.org/10.48550/arXiv.1812.07032>
66. Murugesan B., et al. "Calibrating segmentation networks with margin-based label smoothing", Medical Image Analysis. — 2023. DOI: <https://doi.org/10.1016/j.media.2023.102826>
67. Wang H., et al. "Mask Temperature Scaling: Uncertainty Calibration

for Segmentation", arXiv. — 2024. DOI: <https://doi.org/10.48550/arXiv.2409.12205>

68. Chen L.-C., Zhu Y., Papandreou G., Schroff F., Adam H. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation (DeepLabv3+)", ECCV (Springer). — 2018. DOI: https://doi.org/10.1007/978-3-030-01234-2_49

69. Chen L.-C., Papandreou G., Schroff F., Adam H. "Rethinking Atrous Convolution for Semantic Image Segmentation (DeepLabv3)", arXiv. — 2017. DOI: <https://doi.org/10.48550/arXiv.1706.05587>

70. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. "Focal Loss for Dense Object Detection", ICCV. — 2017. DOI: <https://doi.org/10.1109/ICCV.2017.324>

71. Sudre C.H., Li W., Vercauteren T., Ourselin S., Cardoso M.J. "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations", DLMIA (Springer). — 2017. DOI: https://doi.org/10.1007/978-3-319-67558-9_28

72. Salehi S.S.M., Erdogmus D., Gholipour A. "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks", MICCAI Workshops (Springer). — 2017. DOI: https://doi.org/10.1007/978-3-319-67389-9_44

73. Berman M., Triki A.R., Blaschko M.B. "The Lovász-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-over-Union Measure in Neural Networks", CVPR. — 2018. DOI: <https://doi.org/10.1109/CVPR.2018.00464>

74. Milletari F., Navab N., Ahmadi S.-A. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation", 3DV. — 2016. DOI: <https://doi.org/10.1109/3DV.2016.79>

75. Bustos A., Pertusa A., Salinas J.-M., de la Iglesia-Vayá M. "PadChest: A Large Chest X-ray Image Dataset with Multi-Label Annotated Reports", Medical Image Analysis. — 2020. DOI: <https://doi.org/10.1016/j.media.2020.101797>

76. Nguyen H.Q., Lam K., Le L.T., et al. "VinDr-CXR: An Open Dataset of Chest X-rays with Radiologist's Annotations", *Scientific Data*. — 2022. DOI: <https://doi.org/10.1038/s41597-022-01498-w>
77. Cubuk E.D., Zoph B., Shlens J., Le Q.V. "RandAugment: Practical Automated Data Augmentation with a Reduced Search Space", *arXiv*. — 2019. DOI: <https://doi.org/10.48550/arXiv.1909.13719>
78. Everingham M., Van Gool L., Williams C.K.I., Winn J., Zisserman A. "The Pascal Visual Object Classes (VOC) Challenge", *International Journal of Computer Vision*. — 2010. DOI: <https://doi.org/10.1007/s11263-009-0275-4>
79. Franchi G., Fehri A., Yao A. "Deep Morphological Networks", *Pattern Recognition*. — 2020. DOI: <https://doi.org/10.1016/j.patcog.2020.107246>
80. Cardoso M.J., Li W., Brown R., et al. "MONAI: An Open-Source Framework for Deep Learning in Healthcare", *arXiv*. — 2022. DOI: <https://doi.org/10.48550/arXiv.2211.02701>
81. Abid A., Abdalla A., Abid A., Khan D., Alfozan A., Zou J. "Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild", *arXiv*. — 2019. DOI: <https://doi.org/10.48550/arXiv.1906.02569>
82. Colvin S., et al. "pydantic/pydantic: v2.5.2", *Zenodo* (software release). — 2023. DOI: <https://doi.org/10.5281/zenodo.10185654>
83. Ali R. "Benchmarking Note: Comparing FastAPI and Triton Inference Server for ML Model Deployment", *Zenodo*. — 2025. DOI: <https://doi.org/10.5281/zenodo.17253046>
84. Maier-Hein L., Eisenmann M., Reinke A., et al. "Why Rankings of Biomedical Image Analysis Competitions Should Be Interpreted with Care", *Nature Communications*. — 2018. DOI: <https://doi.org/10.1038/s41467-018-07619-7>
85. Maier-Hein L., Reinke A., Godau P., et al. "Metrics Reloaded: Recommendations for Image Analysis Validation", *Nature Methods*. — 2024. DOI: <https://doi.org/10.1038/s41592-023-02151-z>
86. Taha A.A., Hanbury A. "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool", *BMC Medical Imaging*. — 2015.

DOI: <https://doi.org/10.1186/s12880-015-0068-x>

87. Zech J.R., Badgeley M.A., Liu M., Costa A.B., Titano J.J., Oermann E.K. "Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study", PLOS Medicine. —

2018. DOI: <https://doi.org/10.1371/journal.pmed.1002683>

88. Dietterich T.G. "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms", Neural Computation. — 1998.

DOI: <https://doi.org/10.1162/089976698300017197>

89. Adebayo J., Gilmer J., Muelly M., Goodfellow I., Hardt M., Kim B. "Sanity Checks for Saliency Maps", arXiv. — 2018. DOI:

<https://doi.org/10.48550/arXiv.1810.03292>

90. Ribeiro M.T., Singh S., Guestrin C. "Why Should I Trust You?: Explaining the Predictions of Any Classifier", KDD. — 2016. DOI:

<https://doi.org/10.1145/2939672.2939778>

91. Lundberg S.M., Lee S.-I. "A Unified Approach to Interpreting Model Predictions (SHAP)", arXiv. — 2017. DOI:

<https://doi.org/10.48550/arXiv.1705.07874>

92. Jacob B., Kligys S., Chen B., et al. "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference", CVPR. — 2018.

DOI: <https://doi.org/10.1109/CVPR.2018.00286>

93. Han S., Mao H., Dally W.J. "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding", arXiv. — 2015. DOI:

<https://doi.org/10.48550/arXiv.1510.00149>

Список публікацій здобувача

Статті у наукових фахових виданнях України:

1. Соломатін В.А., Байбуз О.Г., Сиротенко А.Є. Застосування згорткових нейронних мереж для діагностування ішемічного захворювання серця (ІЗС). *Системні технології*. 2023. Т. 6 (149). С. 53–61. DOI: <https://doi.org/10.34185/1562-9945-6-149-2023-05>. (особистий внесок Соломатіна В.А.: проведення аналізу предметної області та постановка задачі застосування згорткових нейронних мереж для діагностування ішемічної хвороби серця за медичними зображеннями, підготовка набору рентгенівських знімків грудної клітки, попередня обробка даних, реалізація та навчання CNN-моделі, проведення обчислювальних експериментів, аналіз отриманих результатів і формування висновків щодо ефективності застосування згорткових нейронних мереж; Байбуза О.Г.: наукове консультування, уточнення мети і завдань дослідження, аналіз та узагальнення отриманих результатів; Сиротенка А.Є.: участь в обговоренні медичної складової дослідження та інтерпретації результатів у контексті діагностики ішемічної хвороби серця).

2. Соломатін В.А., Мащенко Л.В. Застосування математичних моделей при обробці медичних даних діагностування ішемічної хвороби серця. *Актуальні проблеми автоматизації та інформаційних технологій*. 2023. Т. 27. С. 178–187. DOI: <http://dx.doi.org/10.15421/432317>. (особистий внесок Соломатіна В.А.: формалізація задачі обробки медичних даних для діагностування ішемічної хвороби серця, аналіз можливостей застосування математичних моделей у задачах медичної діагностики, підготовка вхідних даних, виконання експериментальної частини дослідження, аналіз результатів застосування моделей та формулювання висновків; Мащенко Л.В.: участь у теоретичному обґрунтуванні використання математичних

моделей, науково-методичне консультування та узагальнення результатів дослідження).

3. Соломатін В.А., Байбуз О.Г. Застосування моделі EfficientNet для виявлення ішемічної хвороби серця (ІХС). *Системні технології*. 2025. Т. 1 (156). С. 160–165. DOI: <https://doi.org/10.34185/1562-9945-1-156-2025-18>. (особистий внесок Соломатіна В.А.: обґрунтування доцільності використання архітектури EfficientNet для аналізу рентгенівських знімків грудної клітки, підготовка та попередня обробка набору даних, адаптація моделі EfficientNet до задачі виявлення ішемічної хвороби серця, реалізація програмного коду навчання й тестування моделі, проведення експериментів, побудова та аналіз метрик якості класифікації; Байбуза О.Г.: постановка наукової задачі, наукове керівництво дослідженням, аналіз і узагальнення отриманих експериментальних результатів).

4. Соломатін В.А., Байбуз О.Г. Використання архітектури трансформера для класифікації ішемічної хвороби серця за допомогою рентгенографічних зображень. *Актуальні проблеми автоматизації та інформаційних технологій*. 2024. Т. 28. С. 244–253. DOI: <http://dx.doi.org/10.15421/432423> (особистий внесок Соломатіна В.А.: аналіз можливостей застосування архітектур трансформерів до задачі класифікації рентгенографічних зображень, підготовка та структурування набору CXR-знімків, адаптація transformer-based моделі до задачі класифікації ішемічної хвороби серця, реалізація програмного забезпечення для навчання та тестування моделі, проведення експериментальних досліджень, аналіз точності класифікації та порівняння отриманих результатів із результатами інших нейромережесевих підходів; Байбуза О.Г.: формування загальної концепції дослідження, наукове керівництво, аналіз та інтерпретація отриманих результатів).

5. Соломатін В.А., Байбуз О.Г. Архітектура BioMorphNet для автоматизованої бінарної класифікації патологій серця за рентгенівськими знімками. *Актуальні проблеми автоматизації та інформаційних технологій*. 2025. Т. 29. С. 382–394. DOI: <http://dx.doi.org/10.15421/432535>. (особистий

внесок Соломатіна В.А.: розроблення концепції та структури архітектури BioMorphNet для автоматизованої бінарної класифікації патологій серця за рентгенівськими знімками, обґрунтування використання морфологічно-орієнтованих компонентів у нейронній моделі, реалізація програмного коду моделі, підготовка та попередня обробка CXR-даних, проведення навчання й тестування моделі, оцінювання якості класифікації, аналіз експериментальних результатів і формування висновків щодо ефективності запропонованої архітектури; Байбуза О.Г.: наукове керівництво, уточнення постановки задачі, аналіз і узагальнення результатів дослідження).

Наукові праці, які засвідчують апробацію матеріалів дисертації:

6. Соломатін В.А., Ємел'яненко Т.Г. Огляд методів підтримки прийняття рішень під час діагностики за медичними даними. *Наука, освіта та суспільство в XXI столітті: наукові ідеї та механізми реалізації: тези доповідей міжнародної науково-практичної конференції*, Полтава, 2021. С. 53–54. URL: <https://www.economics.in.ua/2022/01/xxi-2.html> (особистий внесок Соломатіна В.А.: аналіз сучасних підходів до підтримки прийняття рішень під час діагностики за медичними даними, систематизація методів обробки медичної інформації, узагальнення можливостей застосування математичних моделей і методів машинного навчання в медичній діагностиці, підготовка матеріалів тез; Ємел'яненко Т.Г.: постановка завдання дослідження, науково-методичне консультування, аналіз і узагальнення отриманих результатів).

7. Соломатін В.А., Ємел'яненко Т.Г. Створення датасету рентгенівських зображень для діагностики ішемічної хвороби серця. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2023): тези доповідей XXI Міжнародної науково-практичної конференції*, Дніпро, 2023. С. 276. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2023/11/mpzis-2023.pdf> (особистий внесок Соломатіна В.А.: формування набору рентгенівських зображень грудної клітки для задачі

діагностики ішемічної хвороби серця, структурування даних за класами, попередня обробка зображень, підготовка даних до використання у нейромережових моделях, аналіз вимог до якості та придатності датасету для подальших експериментів; Ємел'яненко Т.Г.: постановка завдання, науково-методичне консультування, аналіз і узагальнення результатів підготовки даних).

8. Соломатін В.А., Байбуз О.Г. Застосування трансформерних моделей для автоматизованої діагностики ішемічного захворювання серця на основі рентгенівських зображень. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2024): тези доповідей XXII Міжнародної науково-практичної конференції*, Дніпро, 2024. С. 257. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2024/11/МПЗІС-2024-1.pdf> (особистий внесок Соломатіна В.А.: аналіз можливостей використання трансформерних моделей для класифікації рентгенівських зображень грудної клітки, підготовка набору CXR-знімків, адаптація архітектури трансформера до задачі автоматизованого виявлення ішемічної хвороби серця, реалізація програмного коду, проведення навчання та тестування моделі, аналіз отриманих результатів; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

9. Соломатін В.А., Байбуз О.Г. Розробка та тренування нейронної мережі для класифікації рентгенівських знімків на основі EfficientNet: методи попередньої обробки, аугментації та аналіз стабільності навчання. *Автоматика 2024: Тези доповідей XXVII Міжнародної конференції*, Дніпро, 2024. С. 191–192. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2025/11/Автоматика-2024-тези-доповідей.pdf> (особистий внесок Соломатіна В.А.: обґрунтування використання архітектури EfficientNet для класифікації рентгенівських зображень грудної клітки, підготовка та попередня обробка даних, застосування методів аугментації, реалізація програмного коду навчання нейронної мережі, проведення експериментів, аналіз стабільності навчання та оцінювання якості

класифікації; Байбуза О.Г.: постановка задачі дослідження, наукове керівництво, аналіз і узагальнення отриманих результатів).

10. Соломатін В.А., Байбуз О.Г. Архітектура BioMorphNet для автоматизованого виявлення ішемічної хвороби серця за рентгенівськими знімками. *Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2025): тези доповідей XXIII Міжнародної науково-практичної конференції*, Дніпро, 2025. С. 286–287. URL: <http://mpzis.dnu.dp.ua/wp-content/uploads/2025/11/МПЗІС-2025.pdf> (особистий внесок Соломатіна В.А.: розроблення концепції архітектури BioMorphNet для автоматизованого виявлення ішемічної хвороби серця за рентгенівськими знімками, обґрунтування використання морфологічно-орієнтованих компонентів, реалізація програмної моделі, підготовка CXR-даних, проведення навчання та тестування, оцінювання якості роботи моделі й аналіз отриманих результатів; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

11. Соломатін В.А., Байбуз О.Г. Поєднання згорткових мереж, механізмів уваги та морфологічного аналізу у задачі виявлення ішемічної хвороби серця за рентгенівськими знімками. *Глобальні виклики та інновації: шляхи розвитку сучасної науки: матеріали X Міжнародної наукової конференції*, Одеса, 2026. С. 363–365. URL: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/16.01.2026/76> (особистий внесок Соломатіна В.А.: обґрунтування доцільності поєднання згорткових нейронних мереж, механізмів уваги та морфологічного аналізу в задачі виявлення ішемічної хвороби серця за рентгенівськими знімками, розроблення та опис відповідного підходу, підготовка рентгенівських зображень до експериментального аналізу, реалізація програмних компонентів, аналіз впливу морфологічних і attention-механізмів на якість розпізнавання; Байбуза О.Г.: постановка задачі, наукове керівництво, аналіз і узагальнення результатів).

12. Соломатін В.А., Байбуз О.Г. Морфологічно-орієнтована нейронна

архітектура для сегментації серця на рентгенівських знімках. *Міжгалузеві диспути: динаміка та розвиток сучасних наукових досліджень: матеріали IX Міжнародної наукової конференції*, Харків, 2026. С. 253–254. URL: <https://archives.mcnd.org.ua/index.php/conference-proceeding/issue/view/30.01.2026/77> (особистий внесок Соломатіна В.А.: розроблення морфологічно-орієнтованої нейронної архітектури для сегментації серцевої області на рентгенівських знімках грудної клітки, обґрунтування використання *edge-aware* та морфологічних механізмів для уточнення меж ROI, реалізація програмної моделі, підготовка сегментаційних даних, проведення експериментального оцінювання, аналіз якості сформованих масок і формулювання висновків щодо ефективності запропонованого підходу; Байбуза О.Г.: постановка наукової задачі, наукове керівництво, аналіз і узагальнення результатів дослідження).

Структура набору даних, кодування міток та контроль коректності розмітки

У додатку Б наведено технічний опис організації набору даних, використаного для навчання, валідації та тестування моделей сегментації серцевої області на рентгенівських знімках грудної клітки.

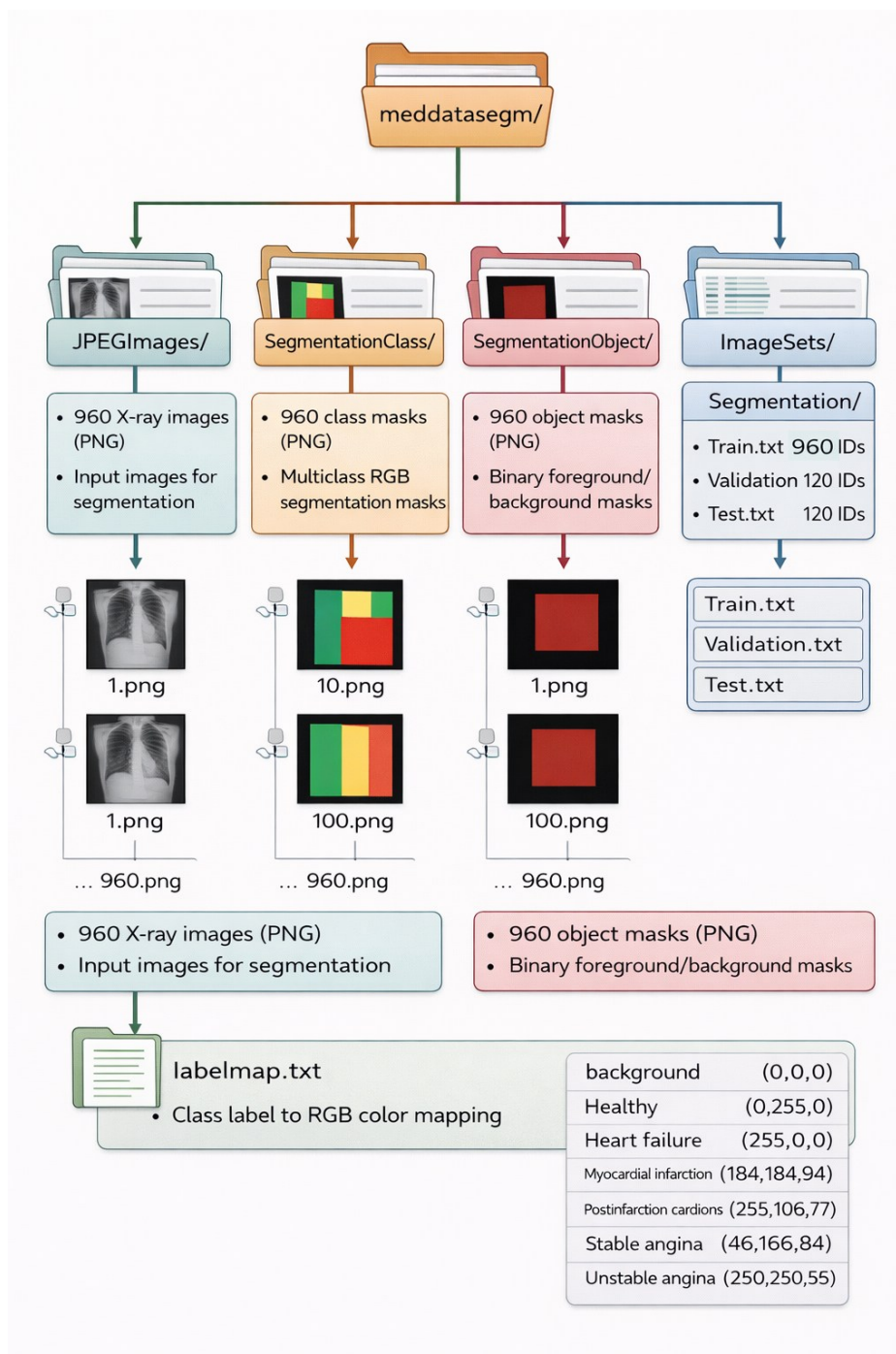


Рис. Б.1 — Структура набору даних

На рисунку Б.1 наведено загальну структуру набору даних. Каталог JPEGImages містить вхідні рентгенівські знімки у форматі PNG. Каталог SegmentationClass містить багатокласові RGB-маски сегментації, у яких кожному класу відповідає окремий колір. Каталог SegmentationObject містить бінарні або об'єктно-орієнтовані маски, що відображають поділ на foreground/background. Каталог ImageSets/Segmentation містить текстові файли зі списками ідентифікаторів зображень, які використовуються для формування навчальної, валідаційної та тестової підвбірок.

Файл labelmap.txt задає відповідність між назвами класів та RGB-кольорами сегментаційної розмітки. Саме ця відповідність використовується під час перетворення RGB-масок у індексні карти класів, які подаються на вхід функції втрат під час навчання нейронної мережі.

Каталог JPEGImages містить рентгенівські знімки грудної клітки, які використовуються як вхідні дані для моделі. Усі зображення приведено до єдиного розміру, що забезпечує уніфіковану подачу даних у нейронну мережу.

Каталог SegmentationClass містить багатокласові маски сегментації. Кожна маска є RGB-зображенням, у якому колір пікселя визначає належність відповідної області до певного класу. Такий формат є зручним для візуального контролю розмітки, однак перед навчанням моделі RGB-маски перетворюються в індексні карти класів.

Каталог SegmentationObject містить об'єктні маски, які можуть використовуватись для додаткового контролю foreground/background-областей. У межах цієї роботи основна увага приділяється багатокласовим маскам із каталогу SegmentationClass, оскільки саме вони містять інформацію про клінічно орієнтовані класи.

Каталог ImageSets/Segmentation містить текстові файли Train.txt, Validation.txt і Test.txt. Ці файли задають ідентифікатори зображень, які використовуються в експериментальному конвеєрі для формування відповідних підвбірок.

Таблиця Б.1 — Відповідність класів та RGB-кольорів

| № | Клас | RGB-колір |
|---|--------------------------------|----------------|
| 0 | Healthy | (0, 255, 0) |
| 1 | Heart failure | (255, 0, 0) |
| 2 | Myocardial infarction | (184, 184, 94) |
| 3 | Postinfarction cardiosclerosis | (255, 106, 77) |
| 4 | Stable angina | (46, 165, 84) |
| 5 | Unstable angina | (250, 250, 55) |
| 6 | Background | (0, 0, 0) |

Наведена палітра використовується для перевірки коректності масок та для перетворення RGB-розмітки у формат, придатний для навчання моделі. Якщо під час обробки маски виявляється піксель із кольором, якого немає в таблиці Б.1, такий випадок розглядається як помилка або як ознака некоректної розмітки.

У вихідному файлі `labelmap.txt` фоновий клас `Background` має окремий RGB-колір (0, 0, 0) і розташований після клінічних класів. Проте в програмній реалізації для зручності обчислення метрик і роботи з `foreground/background`-поділом порядок класів може бути нормалізований таким чином, щоб фоновий клас мав індекс 0.

Така нормалізація не змінює змісту розмітки, оскільки відповідність між RGB-кольором і семантичним класом зберігається. Вона лише змінює внутрішнє числове представлення класів у навчальному коді.

Таблиця Б.2 — Внутрішній порядок класів після нормалізації у програмному конвеєрі

| Внутрішній індекс | Клас |
|-------------------|--------------------------------|
| 0 | Background |
| 1 | Healthy |
| 2 | Heart failure |
| 3 | Myocardial infarction |
| 4 | Postinfarction cardiosclerosis |
| 5 | Stable angina |
| 6 | Unstable angina |

Необхідність такого перетворення зумовлена тим, що більшість метрик сегментації у роботі розраховуються окремо для foreground-областей або з виключенням фонового класу. Тому розміщення Background на індексі 0 спрощує реалізацію обчислення ROI Dice, ROI IoU, Mean Dice та Mean IoU.

Наведена структура набору даних забезпечує відтворювану організацію експериментів із семантичної сегментації серцевої області на рентгенівських знімках грудної клітки. Відокремлення вхідних зображень, RGB-масок, об'єктних масок і текстових файлів розбиття спрощує програмну реалізацію навчального конвеєра та дозволяє контролювати якість даних на кожному етапі.

Ключовими елементами підготовки даних є узгодження пар “зображення–маска”, перевірка RGB-палітри, перетворення масок у індексні карти класів та нормалізація порядку класів у програмній реалізації. Це забезпечує коректне використання набору даних для навчання, валідації та тестування моделей сегментації.