

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ комп'ютерних наук  
(повна назва)

Кафедра \_\_\_\_\_ програмної інженерії  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти \_\_\_\_\_ другий (магістерський)

\_\_\_\_\_ Дослідження ефективності застосування

\_\_\_\_\_ технологій глибокого навчання для аналізу медичних зображень.

(тема)

Виконала:

здобувач 2 року навчання, групи ІПЗм-23-4

\_\_\_\_\_ Дар'я КОЗАК  
(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного  
забезпечення  
(код і повна назва спеціальності)

Тип програми \_\_\_\_\_ освітньо-наукова

Освітня програма \_\_\_\_\_ Інженерія програмного  
забезпечення

Керівник \_\_\_\_\_ доц. Ірина КИРИЧЕНКО  
(посада, прізвище, ініціали)

Допускається до захисту:

Зав. кафедри

\_\_\_\_\_ Кирило СМЕЛЯКОВ  
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

2025 р.

## Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ комп'ютерних наук  
 Кафедра \_\_\_\_\_ програмної інженерії  
 Рівень вищої освіти \_\_\_\_\_ другий (магістерський)  
 Спеціальність \_\_\_\_\_ 121 – Інженерія програмного забезпечення  
 Тип програми \_\_\_\_\_ освітньо-наукова програма  
 Освітня програма \_\_\_\_\_ Інженерія програмного забезпечення  
 (шифр і назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)«\_\_\_\_\_» \_\_\_\_\_ 2025 р.

### ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентки Козак Дар'ї Олександрівни  
 (прізвище, ім'я, по батькові)

## 1. Тема роботи

Дослідження ефективності застосування технологій глибокого навчання для аналізу медичних зображень.

Затверджена наказом по університету №290 Ст від 15.04.2025 \_\_\_\_\_

2. Термін здачі закінченої роботи 12.06.2025 р.

3. Вихідні дані до роботи:

Електронні ресурси за обраною тематикою, аналіз і порівняння сучасних підходів до глибокого навчання у медичній діагностиці.

4. Перелік питань, які треба опрацювати в роботі:

Аналіз предметної та проблемної області, пов'язаної з медичним аналізом зображень.  
Постановка задачі дослідження. Аналіз сучасних технологій глибокого навчання та їх ефективності для обробки медичних зображень. Порівняння комерційних і відкритих систем за ключовими метриками.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	14.04.2025	<i>виконано</i>
2	Аналіз предметної галузі	20.05.2025	<i>виконано</i>
3	Огляд літературних та наукових джерел	21.05.2025	<i>виконано</i>
4	Постановка задачі	21.05.2025	<i>виконано</i>
5	Теоретичне дослідження	26.05.2025	<i>виконано</i>
6	Підготовка до апробації результатів дослідження. Публікація матеріалів	30.05.2025	<i>виконано</i>
7	Програмна реалізація	01.06.2025	<i>виконано</i>
8	Підготовка пояснювальної записки	02.06.2025	<i>виконано</i>
9	Підготовка презентації та доповіді	03.06.2025	<i>виконано</i>
10	Перевірка на плагіат	07.06.2025	<i>виконано</i>
11	Нормоконтроль	07.06.2025	<i>виконано</i>
12	Рецензування	07.06.2025	<i>виконано</i>
13	Попередній захист	07.06.2025	<i>виконано</i>
14	Занесення диплома в електронний архів	09.06.2025	<i>виконано</i>
15	Допуск до захисту у зав. кафедри	11.06.2025	<i>виконано</i>

Дата видачі завдання 14.04.2025р.

Студент (ка/и) \_\_\_\_\_  
(підпис)

\_\_\_\_\_ Козак Д.О.

Керівник роботи \_\_\_\_\_  
(підпис)

\_\_\_\_\_ доц. Кириченко І.В.  
(посада, прізвище, ініціали)

## РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 75 с., 8 рис., 3 табл., 24 джерела.

ГЛИБОКЕ НАВЧАННЯ, МЕДИЧНІ ЗОБРАЖЕННЯ, СЕГМЕНТАЦІЯ, КЛАСИФІКАЦІЯ, КОМЕРЦІЙНІ СИСТЕМИ, ВІДКРИТІ ПЛАТФОРМИ, MONAI, nnU-Net.

Об'єктом дослідження є технології глибокого навчання, що використовуються для аналізу медичних зображень.

Метою роботи є оцінка ефективності сучасних підходів глибокого навчання для аналізу медичних зображень, їх порівняння за ключовими характеристиками та пропозиція шляхів покращення процесу обробки даних.

Методами розробки та проектування є аналіз проблемної області, вибір систем глибокого навчання для порівняння, вивчення відкритих і комерційних платформ, а також розробка експериментального плану для оцінки ефективності технологій.

У результаті роботи були проведені дослідження методів сегментації та класифікації медичних зображень, запропоновано порівняння результатів для комерційних і відкритих систем.

DEEP LEARNING, MEDICAL IMAGES, SEGMENTATION, CLASSIFICATION, COMMERCIAL SYSTEMS, OPEN PLATFORMS, MONAI, nnU-Net.

The object of research is deep learning technologies used for medical image analysis.

The purpose of the work is to assess the efficiency of modern deep learning approaches for medical image analysis, compare them based on key metrics, and propose ways to improve data processing workflows.

The development and design methods include an analysis of the problem domain, selection of deep learning systems for comparison, studying open and commercial platforms, and designing an experimental plan to evaluate the effectiveness of these technologies.

As a result of the work, methods for image segmentation and classification were studied, a comparison of results for commercial and open systems.

Завідувачу кафедри

ПІ

(скорочена назва кафедри)

проф. Кирилу СМЕЛЯКОВУ

(вчене звання, власне ім'я, прізвище)

### ЗАЯВА

щодо самостійності виконання кваліфікаційної роботи та можливості її публікації  
(та/або публікації анотації кваліфікаційної роботи) в електронному архіві  
відкритого доступу EIAr KhNURE

Я, Козак Дар'я Олександрівна, студентка гр. ПЗМ-23-4, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: мій комплексний курсовий проєкт на тему «Дослідження ефективності застосування технологій глибинного навчання для аналізу медичних зображень», що буде представлена для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений(на) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

05.06.2025

Дяр'я КОЗАК

## ЗМІСТ

Вступ	9
1. Аналіз предметної галузі	11
2. Огляд літературних та наукових джерел	13
2.1. Аналіз наукових досліджень про комерційні системи	13
2.2. Аналіз наукових досліджень про відкриті системи	15
2.3. Оцінка актуальності та новизни	16
2.4. Висновки з огляду	17
3. Постановка задачі	19
4. Теоретичне дослідження	21
4.1. Дослідження згорткових нейронних мереж у медичному аналізі	21
4.1.1 Базові елементи CNN	21
4.1.2 Аналіз архітектур на основі CNN	22
4.2. Аналіз алгоритмів відкритих систем	23
4.2.1. nnU-Net	23
4.2.2 MONAI	26
4.2.3. DeepHealth Toolkit	28
4.3 Оцінка ефективності відкритих моделей глибинного навчання	30
4.3.1. Оцінка ефективності в задачах сегментації	30
4.3.2. Precision (Точність)	32
4.3.3. Recall (Повнота)	33
4.3.4. F1-міра	33
4.3.5 Порівняння та оцінка ефективності в задачах класифікації	34
4.4 Порівняння відкритих систем з комерційними	35
5. Практична реалізація методів III для сегментації медичних зображень	39
5.1. Вибір датасету та обчислювальної системи	39
5.2. Реалізація та навчання моделі nnU-Net	41
5.3. Реалізація та навчання моделі MONAI	45
5.4. Порівняльний аналіз результатів	48
5.5 Розробка практичних рекомендацій	51

Висновки	54
Перелік джерел та посилання	56
Перелік джерел посилання за науковими напрямами керівника та науковців кафедри програмної інженерії	59
Додаток А Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ	60
Додаток Б Слайди презентації	61
Додаток В Апробація результатів роботи. Стаття	68
Додаток Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015	75

## ВСТУП

Аналіз медичних зображень є однією з найважливіших задач сучасної медицини, оскільки він дозволяє діагностувати захворювання на ранніх стадіях, визначати їхній характер і динаміку розвитку, а також оцінювати ефективність лікування. Завдяки технологічному прогресу і стрімкому розвитку методів глибинного навчання, аналіз медичних зображень досягнув значного прогресу за останнє десятиліття. Технології штучного інтелекту, зокрема глибинне навчання, надають можливість автоматизувати складні процеси аналізу, знижуючи ймовірність людської помилки та підвищуючи швидкість обробки даних. У зв'язку з цим, тема використання відкритих систем глибинного навчання для аналізу медичних зображень є надзвичайно актуальною.

Однак широке впровадження комерційних рішень, таких як Google DeepMind Health, IBM Watson Health та Aidoc, вимагає значних фінансових ресурсів, що обмежує їхнє застосування в багатьох медичних установах, особливо у країнах із середнім рівнем доходу. Водночас відкриті системи, такі як MONAI, nnU-Net та DeepHealth Toolkit, надають науковцям і розробникам інструменти для створення та вдосконалення алгоритмів аналізу медичних зображень. Вони пропонують можливість тестувати та впроваджувати сучасні підходи без додаткових витрат на придбання ліцензій чи доступу до комерційних платформ. Це відкриває нові горизонти для досліджень у медицині та сприяє популяризації відкритих технологій.

Актуальність теми дослідження полягає у вивченні потенціалу відкритих систем для аналізу медичних зображень, їх порівнянні з результатами комерційних рішень, а також в оцінці їхньої ефективності та можливостей використання у реальних медичних умовах. Результати дослідження можуть бути корисними для науковців, розробників медичного програмного забезпечення, а також для навчальних закладів, які займаються підготовкою спеціалістів у галузі медичних технологій.

Метою даної роботи є оцінка ефективності відкритих систем глибокого навчання для аналізу медичних зображень та визначення їхнього потенціалу для заміни комерційних рішень у наукових і навчальних цілях.

Для досягнення цієї мети у роботі ставляться наступні задачі:

- провести аналіз сучасних відкритих систем глибокого навчання для аналізу медичних зображень;
- оцінити ефективність відкритих моделей за кількісними метриками, такими як Dice Coefficient, Precision, Recall та F1-міра;
- використати дані, опубліковані для комерційних систем, для порівняльного аналізу;
- визначити сильні та слабкі сторони відкритих рішень у контексті медичної практики;
- запропонувати рекомендації щодо подальшого розвитку відкритих систем для інтеграції у клінічну практику.

Об'єктом дослідження є процеси аналізу медичних зображень, які використовують технології глибокого навчання.

Предмет дослідження – відкриті моделі глибокого навчання, такі як MONAI, nnU-Net та DeepHealth Toolkit, а також їхня продуктивність у задачах сегментації, класифікації та регресії.

Методи дослідження включають аналіз літератури для вивчення сучасних підходів і рішень, експериментальне тестування відкритих систем на доступних наборах даних, таких як LIDC-IDRI та BraTS, а також порівняльний аналіз точності та швидкості роботи цих систем із результатами комерційних платформ, опублікованими у наукових статтях.

Таким чином, дана робота спрямована на дослідження ефективності відкритих систем глибокого навчання, що має не лише наукове значення, а й практичну цінність для розвитку медичних технологій.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

Сфера аналізу медичних зображень є критично важливою для діагностування захворювання на ранніх стадіях, проведення моніторингу стану пацієнтів і планування ефективного лікування. Однак традиційні методи обробки медичних зображень, які базуються на ручному аналізі або використанні базових алгоритмів комп'ютерного зору, часто є трудомісткими, часозатратними й залежать від досвіду спеціаліста. Це створює потребу у впровадженні автоматизованих рішень, які могли б забезпечити високу точність та швидкість обробки даних.

Глибинне навчання, як підгалузь штучного інтелекту, стало революційним інструментом у медичній діагностиці. Використання моделей, таких як згорткові нейронні мережі (CNN), дозволяє автоматично виділяти ключові ознаки зображень без потреби ручного проектування фільтрів. Це відкриває нові можливості для сегментації органів, виявлення патологій та оцінки змін у структурі тканин. Водночас використання відкритих моделей глибинного навчання, таких як nnU-Net, MONAI, і DeepHealth Toolkit, забезпечує доступність сучасних технологій для дослідників і лікарів, що сприяє демократизації інновацій у медицині.

Сьогодні аналіз медичних зображень є однією з найактивніших і найважливіших сфер впровадження глибинного навчання. Ця технологія дає змогу автоматизувати складні процеси аналізу даних, значно знижуючи залежність від людського фактора. Завдяки цьому стає можливим покращити діагностику захворювань, пришвидшити ухвалення клінічних рішень і підвищити ефективність лікування. Глибинне навчання пропонує революційні підходи до обробки медичних зображень, включаючи сегментацію органів, класифікацію патологій і візуалізацію тривимірних структур. Моделі, такі як nnU-Net, MONAI і DeepHealth Toolkit, стали важливими інструментами для розв'язання широкого спектра медичних задач.

Одна з ключових задач у медичному аналізі зображень – це сегментація органів і пухлин. Глибинне навчання дозволяє автоматично виділяти межі органів чи патологічних структур із високою точністю, що особливо важливо для роботи з МРТ та КТ-знімками. Завдяки цьому знижується залежність від ручної анотації,

підвищується швидкість аналізу, а також покращується точність і відтворюваність результатів.

Інший важливий напрямок – класифікація патологій, таких як рак легенів, серцево-судинні захворювання чи ураження головного мозку. CNN-моделі дозволяють автоматично аналізувати рентгенівські, КТ або МРТ-знімки, виділяючи характерні особливості патологій [1]. Бібліотека MONAI надає потужні інструменти для класифікації завдяки підтримці різних архітектур, таких як ResNet і DenseNet [2]. Крім того, її адаптивність до різних типів медичних даних є значною перевагою у клінічних умовах, де часто використовуються неоднорідні джерела даних.

Складнішим завданням є тривимірна візуалізація, яка потребує аналізу великих обсягів даних і значних обчислювальних ресурсів. Робота з 3D-даними (наприклад, КТ або МРТ) ускладнюється необхідністю обробки кожного шару зображення та інтеграції цих даних у єдину об'ємну структуру. Моделі, такі як DeepHealth Toolkit, поєднують глибинне навчання з традиційними методами обробки зображень, що дозволяє зменшити обчислювальні витрати та оптимізувати процес аналізу [3]. Це особливо важливо для задач моніторингу динамічних змін у стані пацієнта, наприклад, при відстеженні розвитку пухлин чи змін у функціонуванні органів.

Ще однією значущою сферою є аналіз багатомодальних даних, які об'єднують зображення з різних джерел, таких як МРТ, КТ та ПЕТ (позитронно-емісійна томографія). Багатомодальний підхід є критично важливим для комплексної діагностики, оскільки він дозволяє інтегрувати дані з різними фізичними властивостями. Наприклад, МРТ забезпечує високу деталізацію м'яких тканин, тоді як ПЕТ дозволяє оцінювати метаболічну активність. Об'єднання цих джерел підвищує точність діагнозу та забезпечує більш повну картину стану пацієнта.

Для оцінки ефективності технологій глибинного навчання у медичній сфері передбачається використання математичних моделей. Вони дозволяють кількісно

порівнювати продуктивність різних систем за такими параметрами, як точність (Precision), повнота (Recall), F1-міра та коефіцієнт Дайса.

Ще одним перспективним напрямком є використання глибинного навчання для прогнозування розвитку захворювань на основі історичних даних пацієнтів. Завдяки аналізу багатомодальних даних у поєднанні з іншими джерелами, такими як електронні медичні записи, моделі можуть визначати патерни, які вказують на ймовірність розвитку патологій. Наприклад, динамічний аналіз МРТ-знімків може виявляти ранні ознаки нейродегенеративних захворювань, таких як хвороба Альцгеймера, до появи клінічних симптомів. Це дозволяє лікарям ухвалювати більш обґрунтовані рішення, спрямовані на профілактику чи раннє лікування, що значно покращує результати терапії.

Таким чином, аналіз предметної галузі демонструє, що використання глибинного навчання в аналізі медичних зображень має великий потенціал для покращення діагностики та лікування. Завдяки своїй здатності автоматизувати рутинні процеси, інтегрувати дані з різних джерел і забезпечувати високу точність результатів, ці технології стають невід'ємною частиною сучасної медицини. У перспективі вони здатні значно підвищити ефективність клінічних процедур і якість наданих медичних послуг, сприяючи точнішій діагностиці та більш персоналізованому лікуванню.

## 2 ОГЛЯД ЛІТЕРАТУРНИХ ТА НАУКОВИХ ДЖЕРЕЛ

### 2.1 Аналіз наукових досліджень про комерційні системи

У сучасній медичній практиці активно впроваджуються комерційні рішення, що використовують технології глибинного навчання для аналізу медичних зображень. Ці платформи спрямовані на підвищення точності діагностики, автоматизацію рутинних процесів та підтримку прийняття клінічних рішень.

Google DeepMind Health є провідною платформою, яка використовує глибинне навчання для аналізу медичних даних, зокрема медичних зображень. Один із ключових успіхів цієї платформи – її здатність діагностувати офтальмологічні захворювання, такі як діабетична ретинопатія та макулярна дегенерація. У дослідженні, опублікованому в *Nature Medicine*, було зазначено, що система досягла точності, порівнянної з провідними офтальмологами [4]. З результатами 94–96% для виявлення патологій у скринінгових даних офтальмології, Google DeepMind Health застосовується для діагностики онкологічних та серцево-судинних захворювань [4]. Платформа активно використовується для досліджень у співпраці з NHS (Національною службою охорони здоров'я Великобританії), що дозволяє впроваджувати інноваційні методи діагностики та моніторингу пацієнтів у клінічних умовах.

Aidos спеціалізується на автоматизованому аналізі радіологічних зображень. Платформа допомагає лікарям ідентифікувати критичні патології, такі як внутрішні кровотечі, тромбози або внутрішньочерепні гематоми. Завдяки використанню глибинного навчання Aidos забезпечує точність понад 90% у задачах екстреної діагностики, що підтверджується клінічними дослідженнями, опублікованими в *Radiology: Artificial Intelligence* в режимі реального часу [5]. Це є особливо важливим у відділеннях невідкладної допомоги. Її алгоритми оптимізовані для інтеграції з системами PACS (Picture Archiving and Communication Systems), що дозволяє лікарям швидко отримувати результати аналізу безпосередньо у своїх робочих інструментах. Aidos уже використовується в багатьох лікарнях по всьому світу, демонструючи свою ефективність у підвищенні точності діагностики та швидкості обробки даних.

IBM Watson Health – це платформа, яка використовує штучний інтелект для аналізу медичних даних і підтримки клінічних рішень. Її ключова особливість полягає в здатності обробляти великі обсяги даних, включаючи медичні зображення, тексти медичних записів та наукові статті [6]. Завдяки алгоритмам глибокого навчання, Watson Health здатен надавати рекомендації лікарям, аналізуючи десятки мільйонів даних за короткий час.

Один із відомих випадків використання Watson Health – діагностика рідкісної форми лейкемії, де система за 10 хвилин проаналізувала понад 20 мільйонів статей про рак і запропонувала правильний варіант лікування. Це свідчить про її здатність працювати в режимі реального часу з величезними обсягами інформації, що значно перевищує можливості людського мозку [6].

Watson Health також інтегрується з медичними інформаційними системами та електронними медичними картками, що дозволяє лікарям отримувати персоналізовані рекомендації для пацієнтів. Це робить платформу корисним інструментом для діагностики, планування лікування та досліджень.

Попри успіхи, IBM Watson Health зазнала значної критики через обмеження та помилки у своїй роботі. Як зазначено у кейсі на платформі Henrico Dolfing, проект IBM Watson for Oncology не виправдав очікувань і зіткнувся з серйозними проблемами у впровадженні [7].

Однією з основних причин проблем Watson for Oncology стала недостатня різноманітність навчальних даних. Система базувалася переважно на інформації з однієї клініки (Memorial Sloan Kettering Cancer Center), що обмежило її здатність працювати з ширшим спектром випадків і патологій. Це призводило до того, що рекомендації системи могли бути нерелевантними для пацієнтів з інших регіонів або із різними характеристиками захворювань.

Watson Health часто демонстрував складнощі в адаптації до нових або нестандартних клінічних випадків. Наприклад, система могла пропонувати неправильні варіанти лікування через недостатню кількість специфічних даних у її базі знань.

IBM Watson Health є однією з провідних платформ у галузі штучного інтелекту для медицини, яка має значний потенціал для вдосконалення діагностики та лікування. Проте її обмеження, такі як недостатня різноманітність навчальних даних і проблеми з узагальненням, підкреслюють важливість ретельного підходу до розробки і впровадження подібних систем.

Цей аналіз демонструє, що успіх систем глибинного навчання у медицині залежить не лише від їхніх технічних можливостей, але й від якості даних, прозорості алгоритмів та врахування етичних аспектів. IBM Watson Health залишається важливим інструментом, але її досвід слугує цінним уроком для розробників подібних платформ.

## 2.2 Аналіз наукових досліджень про відкриті системи

Відкриті системи глибинного навчання для аналізу медичних зображень, такі як nnU-Net, MONAI та DeepHealth Toolkit, активно досліджуються у науковій спільноті. Їхні основні переваги включають доступність, можливість адаптації до різних задач та використання сучасних архітектур глибинного навчання. У цьому розділі розглянемо наукові праці, присвячені кожній із систем, щоб визначити їхні сильні сторони, недоліки та перспективи.

Модель nnU-Net стала стандартом у задачах сегментації медичних зображень, завдяки своїй здатності автоматично адаптувати архітектуру до специфіки набору даних. У статті [1], зазначено, що nnU-Net перевершує інші підходи у задачах сегментації органів і пухлин, таких як Brain Tumor Segmentation Challenge (BraTS) і Kidney Tumor Segmentation (KiTS). Її головною перевагою є автоматизація процесу вибору гіперпараметрів, що дозволяє дослідникам зосередитися на аналізі результатів, а не на налаштуванні моделі.

Недоліком nnU-Net є висока вимогливість до обчислювальних ресурсів, що ускладнює її використання в умовах обмежених апаратних можливостей.

Бібліотека MONAI надає гнучкі інструменти для роботи з медичними зображеннями, підтримуючи задачі сегментації, класифікації та регресії. У дослідженні [2] MONAI відзначається як універсальний інструмент для роботи з

багатомодальними даними, такими як МРТ та КТ. Особливу увагу приділено її модулю Data Augmentation, який дозволяє покращити узагальнювальну здатність моделей.

MONAI підтримує використання попередньо навчених моделей, таких як ResNet і 3D-UNet, що зменшує потребу у великих наборах даних для навчання. Проте, для задач тривимірної сегментації її продуктивність поступається nnU-Net через відсутність автоматичної адаптації архітектури.

DeepHealth Toolkit розроблено в межах європейського проекту Horizon 2020 і призначено для аналізу медичних зображень у 2D та 3D форматах. У дослідженні [8] розглядається застосування DeepHealth Toolkit для сегментації легеневих вузлів на КТ-зображеннях. Автори відзначають, що інструментарій забезпечує точну сегментацію вузлів різного діаметру, перевершуючи традиційні методи виявлення. Це підтверджує ефективність DeepHealth Toolkit у задачах аналізу медичних зображень, особливо в умовах обмежених ресурсів.

Головним обмеженням є складність налаштування для нових задач і нижча точність порівняно з nnU-Net та MONAI у задачах тривимірної сегментації. Проте, її продуктивність залишається конкурентоспроможною в умовах обмежених ресурсів.

### 2.3 Оцінка актуальності та новизни

Актуальність використання глибинного навчання для аналізу медичних зображень пояснюється нагальною потребою в автоматизації через зростаючий обсяг даних і обмеженість людських ресурсів. Завдяки високій продуктивності, точності та здатності працювати з комплексними зображеннями, глибинне навчання стає ключовим рішенням для сучасної медицини. Особливу роль відіграють відкриті системи, такі як MONAI, nnU-Net та DeepHealth Toolkit, які роблять передові технології доступними без значних фінансових витрат.

Наукова новизна відкритих рішень полягає у їхній універсальності. Наприклад, nnU-Net адаптується до будь-якого набору даних без необхідності ручного налаштування, що значно полегшує процес впровадження технологій у

дослідницьких і клінічних умовах. Це забезпечує новий рівень автономності та універсальності в задачах сегментації.

Комерційні системи, такі як Google DeepMind Health, IBM Watson Health та Aidoc, також надають цінний досвід. Хоча їхні алгоритми залишаються закритими, результати роботи часто доступні у наукових статтях. Це дозволяє використовувати їх як контрольні точки для оцінки можливостей відкритих систем. Актуальність цих порівнянь зумовлена необхідністю визначення ефективності безкоштовних рішень для їх потенційного використання у клінічній практиці.

Загалом, оцінка актуальності та новизни підтверджує, що відкриті системи є важливими інструментами для розвитку сучасної медицини. Їхня універсальність, доступність і здатність забезпечувати результати, які можна порівняти з комерційними системами, роблять ці рішення незамінними для наукових досліджень, клінічних застосувань та освітніх програм.

#### 2.4 Висновки з огляду

Огляд літератури та наукових джерел дозволяє зробити кілька важливих висновків щодо стану використання глибинного навчання для аналізу медичних зображень. Відкриті системи, такі як MONAI, nnU-Net та DeepHealth Toolkit, демонструють значний потенціал у вирішенні задач аналізу медичних зображень. Вони надають можливість досягати високої точності в задачах сегментації та класифікації, забезпечуючи доступність інноваційних технологій для науковців і освітніх закладів завдяки відкритій архітектурі та широким можливостям налаштування.

Аналіз результатів роботи комерційних систем, таких як Google DeepMind Health, IBM Watson Health та Aidoc, показав, що ці рішення забезпечують надзвичайну ефективність у клінічній практиці. Проте їхня вартість і закритість алгоритмів обмежують їхнє використання для широкого кола медичних установ і дослідницьких проєктів. У публікаціях часто представлені відкриті результати їхньої роботи, що дозволяє використовувати ці дані як еталон для оцінки ефективності відкритих моделей.

Однією з головних переваг відкритих рішень є їхня універсальність і адаптивність. Наприклад, nnU-Net автоматично налаштовується для конкретних типів даних без необхідності ручного втручання, що робить її універсальною для різних задач сегментації. Аналогічно, MONAI пропонує готові до використання інструменти для побудови моделей, які працюють із багатьма форматами медичних даних, включаючи DICOM та NIfTI. Водночас DeepHealth Toolkit підтримує обробку як 2D, так і 3D зображень, що розширює можливості його використання.

Незважаючи на значний потенціал, відкриті системи мають низку обмежень. Вони потребують додаткового доопрацювання для інтеграції у клінічну практику, зокрема розробки зручних користувацьких інтерфейсів, масштабованості для великих медичних установ та відповідності міжнародним стандартам, таким як HL7 або FHIR [9, 10]. Це є критичним аспектом для їх впровадження в реальні медичні процеси.

Важливим аспектом є також використання відкритих наборів даних, таких як LIDC-IDRI, BraTS та KiTS, які надають стандартизовану основу для порівняння результатів різних рішень. Ці набори даних забезпечують можливість об'єктивного тестування і вдосконалення моделей глибокого навчання, а також сприяють розвитку відкритих технологій у галузі медицини.

Таким чином, аналіз літератури свідчить про значний прогрес у використанні відкритих систем глибокого навчання для аналізу медичних зображень. Ці системи вже зараз демонструють високі результати, що дозволяє використовувати їх у наукових дослідженнях і освітніх програмах. Подальші дослідження повинні бути спрямовані на адаптацію відкритих рішень до специфічних вимог клінічної медицини, інтеграцію їх у реальні робочі процеси та забезпечення відповідності стандартам, що сприятиме їхньому широкому впровадженню в медичну практику.

### 3 ПОСТАНОВКА ЗАДАЧІ

Основною метою даного дослідження є оцінка ефективності відкритих систем глибинного навчання у сфері аналізу медичних зображень. Сучасна медицина активно використовує алгоритми штучного інтелекту для автоматизації діагностичних процесів, оптимізації аналізу даних і підвищення точності виявлення патологій. Особливу увагу у цьому проєкті приділено відкритим системам, таким як nnU-Net, MONAI та DeepHealth Toolkit, які вже зарекомендували себе як потужні інструменти для сегментації, класифікації та аналізу медичних даних. Ці системи забезпечують гнучкість у налаштуванні та можливість адаптації до різних задач, що робить їх цінним ресурсом як для дослідників, так і для клінічної практики.

Проєкт спрямований на аналіз архітектур моделей, методів їхнього навчання та оцінку продуктивності в різних умовах. Окрім цього, особливий акцент зроблено на порівнянні відкритих систем із комерційними платформами, такими як Google DeepMind Health, IBM Watson Health і Aidoc, щоб визначити їхню здатність ефективно конкурувати у реальних клінічних задачах. Окрім цього, дослідження має на меті розробити рекомендації щодо вибору оптимальної моделі для аналізу медичних зображень у конкретних сценаріях, залежно від типу задачі, доступних ресурсів і специфіки даних.

Ключові задачі дослідження:

- аналіз архітектур і методів навчання. Вивчити структуру моделей, зокрема згорткові нейронні мережі (CNN). Дослідити підходи до автоматичної адаптації гіперпараметрів у nnU-Net. Проаналізувати можливості попередньо навчених моделей у MONAI;
- оцінка ефективності моделей. Використати математичні метрики, такі як Dice Coefficient, Precision, Recall та F1-міра, для об'єктивного порівняння продуктивності моделей;
- порівняння відкритих та комерційних платформ. Провести аналіз результатів комерційних платформ, таких як Google DeepMind Health,

IBM Watson Health та Aidoc, за наявними публічними даними. Оцінити конкурентоспроможність відкритих систем у задачах сегментації та класифікації;

- розробка рекомендацій. Запропонувати підходи до інтеграції відкритих систем у медичну практику. Надати рекомендації щодо вибору моделей залежно від специфіки задачі, типу даних і доступних ресурсів.

У межах даної кваліфікаційної роботи передбачається виконати як теоретичний, так і експериментальний етапи дослідження. На основі аналізу сучасних наукових публікацій, технічної документації.

У рамках практичної частини роботи заплановано провести порівняльний експеримент з навчання та тестування моделей на спільному медичному наборі даних. Експеримент дозволить оцінити ефективність обраних систем у рівних умовах, порівняти їхню точність за кількісними метриками, а також дослідити особливості їх роботи в обмежених апаратних ресурсах. Такий підхід дозволяє не лише підтвердити теоретичні висновки, отримані з аналізу літературних джерел, а й сформулювати практичні рекомендації для подальшого використання моделей у наукових та медичних проектах.

Щодо комерційних систем, таких як Google DeepMind Health, IBM Watson Health та Aidoc, їх аналіз здійснюватиметься виключно за відкритими джерелами — опублікованими науковими статтями, технічними оглядами та офіційною документацією, оскільки прямий доступ до алгоритмів, архітектур та внутрішніх даних цих систем відсутній.

Таким чином, результати дослідження ґрунтуватимуться як на теоретичному аналізі існуючих наукових джерел, так і на експериментальному порівнянні обраних відкритих моделей, що забезпечить комплексне оцінювання їх ефективності в задачах медичної сегментації.

## 4 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

Глибинне навчання суттєво змінило підхід до аналізу медичних зображень, надаючи можливості для автоматизованої сегментації, класифікації та виявлення патологій. У цьому розділі детально розглянуто три відкриті системи глибинного навчання: nnU-Net, MONAI, та DeepHealth Toolkit. Для оцінки ефективності технологій глибинного навчання в задачах аналізу медичних зображень використовувалися комплексні підходи, що охоплюють математичне моделювання та аналітичне порівняння.

### 4.1 Дослідження згорткових нейронних мереж у медичному аналізі

Для задач сегментації та класифікації медичних зображень широко використовуються згорткові нейронні мережі (Convolutional Neural Networks, CNN). CNN дозволяють автоматично виділяти суттєві ознаки з медичних даних завдяки багаторівневій архітектурі, що складається з базових елементів, таких як згортка, функції активації, підвибірка та нормалізація. Ці елементи формують основу для аналізу складних структур, характерних для медичних зображень.

#### 4.1.1 Базові елементи CNN

Згортка є основною операцією, яка дозволяє виділяти локальні ознаки зображень. Під час цієї операції фільтр (ядро згортки) проходить через вхідне зображення, обчислюючи зважену суму значень на кожному кроці. Формально це описується за формулою 4.1.

$$(S * K)(i, j) = \sum_{m=1}^M \sum_{n=1}^N S(i + m, j + n) * K(m, n), \quad (4.1)$$

де  $S$  – вхідне зображення,

$K$  – згортковий фільтр,

$M, N$  – розміри фільтра.

Ця операція дозволяє CNN виділяти такі ознаки, як краї, текстури чи форми, які є важливими для сегментації органів або виявлення патологій.

Функції активації додають нелінійність до моделі, що дозволяє вивчати складні залежності між ознаками. Найпоширеніша функція активації – ReLU (Rectified Linear Unit), яка визначається за формулою 4.2:

$$ReLU(x) = \max(0, x) \quad . \quad (4.2)$$

ReLU обмежує негативні значення, залишаючи лише позитивні, що допомагає уникнути проблеми згасання градієнтів. Це сприяє швидкому та стабільному навчанню.

Підвибірка зменшує розмірність вхідних даних, що дозволяє моделі працювати з великими зображеннями ефективніше. Найпоширеніший метод – максимальна підвибірка (Max Pooling), яка вибирає максимальне значення у кожному підвікні заданого розміру. Ця операція забезпечує інваріантність до зміщень і шумів.

Методи нормалізації, такі як Batch Normalization, стабілізують навчання, масштабуючи та зміщуючи значення активацій. Це дозволяє швидше досягати збіжності під час навчання моделі, а також знижує ризик перенавчання.

#### 4.1.2 Аналіз архітектур на основі CNN

Архітектура U-Net розроблена спеціально для задач сегментації, зокрема медичних зображень, і має симетричну структуру, яка складається з енкодера та декодера. Енкодер виконує згорткові операції для вилучення ознак, поступово зменшуючи розмірність вхідних даних, тоді як декодер відновлює просторову інформацію та генерує маску сегментації. Ключовою особливістю U-Net є пропускні зв'язки між відповідними рівнями енкодера та декодера. Ці зв'язки дозволяють передавати деталі високої роздільної здатності, які втрачаються в процесі згортки, покращуючи точність моделі.

U-Net отримала широке визнання завдяки своїй здатності ефективно працювати з малими наборами даних, що є типовим для медичних задач. Наприклад, у сегментації органів або патологій модель демонструє високий Коефіцієнт Дайса ( $>0.85$ ) навіть на невеликих наборах даних [11]. Дослідження, показують, що U-Net може бути адаптована для різних завдань, включаючи 2D- та 3D-зображення, що робить її універсальним інструментом у медичному аналізі [11].

ResNet (Residual Network) є однією з найпопулярніших архітектур CNN для задач класифікації та аналізу зображень. Головною інновацією ResNet є використання залишкових блоків (residual blocks), які включають прямі зв'язки між шарами. Це дозволяє усунути проблему згасання градієнтів, яка є типовою для глибоких мереж, і значно спрощує навчання.

Згідно з дослідженням «Deep Residual Networks for Medical Image Analysis» [12], ResNet демонструє високу ефективність у задачах класифікації медичних зображень. Наприклад, у класифікації рентгенівських зображень грудної клітини модель досягає точності понад 92% на великих наборах даних [12]. Дослідження також підкреслює, що ResNet стабільно показує високі результати при збільшенні глибини мережі, завдяки залишковим зв'язкам. Крім того, ResNet можна ефективно використовувати для сегментації, якщо доповнити її модулем декодування, схожим на U-Net.

ResNet добре масштабується, що дозволяє використовувати її для великих наборів даних і складних задач. Однак її продуктивність сильно залежить від обчислювальних ресурсів, тому для її роботи рекомендується використання GPU з великою пам'яттю. Це робить ResNet менш підходящою для обмежених умов, проте ідеальною для великих дослідницьких проектів.

## 4.2 Аналіз алгоритмів відкритих систем

### 4.2.1 nnU-Net

nnU-Net є універсальною системою для сегментації медичних зображень, яка встановлює новий стандарт у сфері аналізу біомедичних даних. Вона розроблена

на основі класичної архітектури 3D-UNet, але суттєво вдосконалена завдяки автоматизації процесу налаштування. Це дозволяє дослідникам і лікарям зосереджуватися на своїх задачах, а не на тонкощах налаштування моделей.

Процес роботи nnU-Net розпочинається з підготовки даних. Зображення нормалізуються, масштабуються та, за необхідності, обрізаються для відповідності стандартним розмірам. Ці етапи є критично важливими, оскільки якість обробки початкових даних безпосередньо впливає на результативність моделі.

Після попередньої обробки модель автоматично адаптує архітектуру під конкретний набір даних. Вона обирає оптимальний розмір вхідного зображення, глибину мережі, кількість каналів у конволюційних шарах та інші параметри. Ця автоматизація дозволяє зменшити час і ресурси, які зазвичай витрачаються на ручне налаштування.

В основі архітектури лежить три ключові етапи. Перший етап – це *downsampling*, або зменшення розмірності. На цьому етапі модель через послідовність конволюційних і *max-pooling* шарів виділяє глобальні ознаки зображення, які важливі для подальшої обробки. Потім йде *bottleneck*, або вузьке місце, де інформація стискається до мінімально можливого представлення, що дозволяє ефективніше опрацьовувати її. Завершальний етап – це *upsampling*, тобто відновлення зображення до вихідної розмірності з інтеграцією локальних і глобальних ознак.

Нижче представлено візуальний приклад роботи моделей сегментації на основі nnU-Net (рис. 4.1). Демонструється два підходи – модель трансферного навчання nnU-NetTL і стандартна модель глибокого навчання nnU-NetDL. У першому випадку nnU-NetTL спочатку тренується на пухлинах головного мозку дорослих (етап 1), після чого результати використовуються для навчання моделі на медулобластомах у дітей (Етап 2). У другому випадку nnU-NetDL тренується безпосередньо на медулобластомах у дітей.

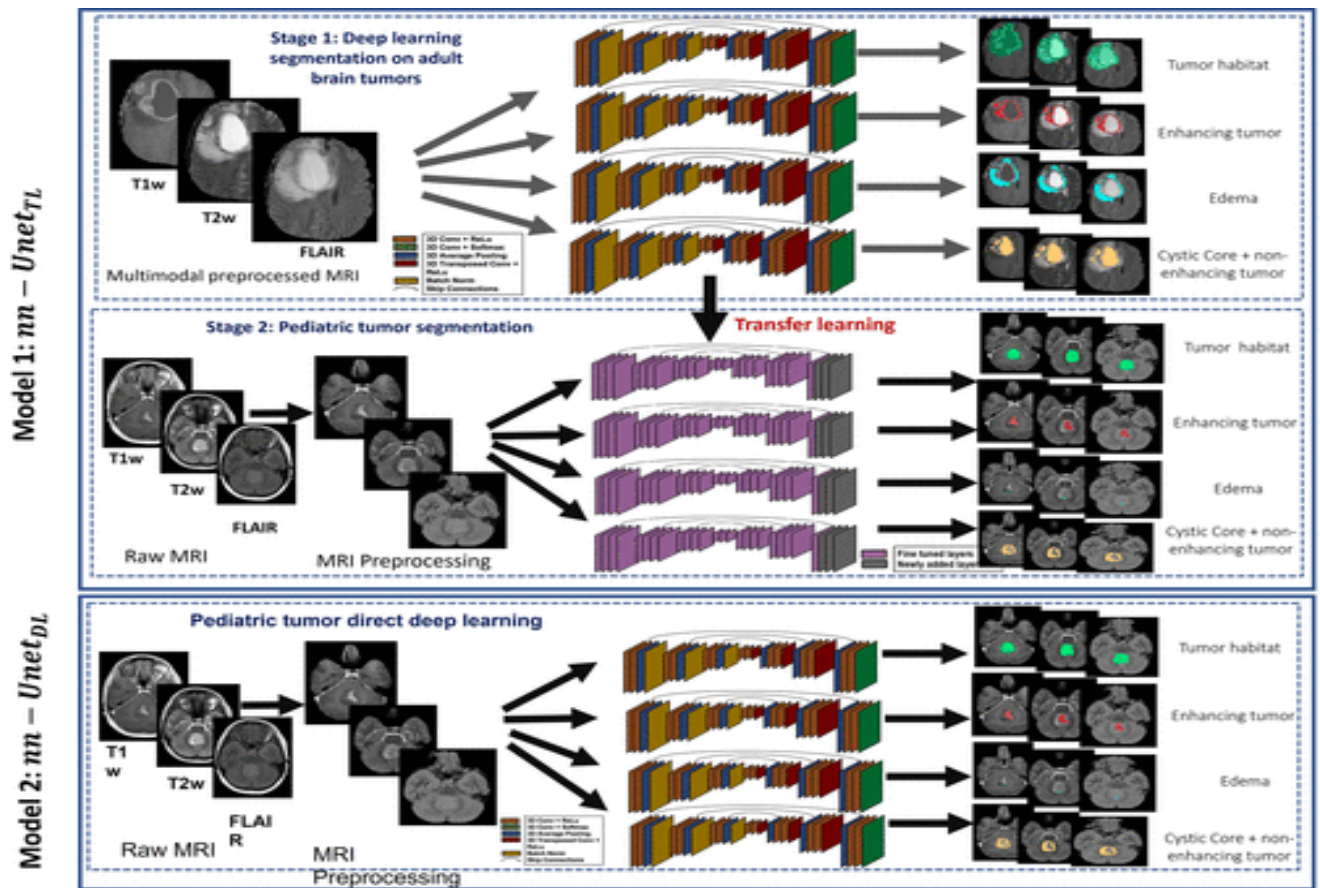


Рисунок 4.1 – Приклади сегментації за допомогою nnU-Net (за даними [13])

Для навчання nnU-Net використовується комбінована функція втрат, яка складається з Dice Loss і Cross-Entropy Loss (формула 4.3):

$$\mathcal{L} = 1 - \frac{2 * |P \cap G|}{|P| + |G|} + \sum_i -T_i \log(P_i), \quad (4.3)$$

де  $P$  – передбачений масив,

$T$  – істинний масив.

Ця функція дозволяє забезпечити збалансованість між точністю сегментації й обчислювальною ефективністю.

Головною перевагою nnU-Net є її універсальність. На відміну від багатьох інших моделей, які вимагають значних зусиль для ручного налаштування, nnU-Net

автоматично підлаштовується до нових наборів даних. Це значно скорочує час підготовки й дозволяє отримати результати навіть для задач, які раніше потребували спеціалізованих підходів.

Ще однією важливою перевагою є те, що модель підтримує роботу з тривимірними зображеннями. Це критично важливо для аналізу МРТ та КТ, де потрібна сегментація об'ємних структур. Також nnU-Net демонструє високу точність і стійкість до шуму, що підтверджується її результатами на стандартних наборах даних, таких як BraTS та KiTS [21].

Попри свою універсальність, nnU-Net має деякі обмеження. Перш за все, автоматизація не завжди забезпечує оптимальні результати для дуже складних або специфічних наборів даних. Наприклад, при роботі з даними, які мають високий ступінь варіабельності або значну кількість артефактів, можуть знадобитися ручні налаштування.

Крім того, модель є обчислювально вимогливою. Навіть із сучасними GPU тренування nnU-Net може займати значний час, особливо при роботі з великими наборами тривимірних даних. Це робить її менш придатною для установ із обмеженими обчислювальними ресурсами.

Основою роботи nnU-Net є згорткові нейронні мережі (CNN). Для обробки тривимірних даних використовуються 3D-CNN, які враховують просторові залежності в об'ємі. Окрім цього, модель інтегрує механізм skip connections, що забезпечує передачу інформації між початковими та кінцевими шарами. Це дозволяє зберігати деталі високої роздільної здатності на етапі апсемплінгу.

#### 4.2.2 MONAI

MONAI (Medical Open Network for AI) є відкритою бібліотекою, створеною спеціально для задач медичного аналізу. Ця система є інтеграцією глибинного навчання з оптимізацією для обробки біомедичних даних. Розроблена на базі PyTorch, MONAI надає дослідникам і розробникам потужний інструмент для роботи з різними медичними форматами даних, такими як DICOM і NIfTI.

MONAI побудована на модульному підході, що дозволяє адаптувати архітектуру під специфічні задачі. Робота з даними в MONAI починається з їхньої підготовки. Для цього використовуються трансформації даних (Data Augmentation), які включають операції, як-от масштабування, обертання, обрізання та нормалізація. Ці трансформації збільшують різноманітність набору даних, що дозволяє моделі навчитися краще узагальнювати.

Ключовим етапом у роботі MONAI є інтеграція попередньо налаштованих моделей (pre-trained models). Ці моделі, такі як ResNet або DenseNet, вже натреновані на великих наборах даних, що зменшує потребу в тривалому навчанні.

MONAI також підтримує багатомодальні дані. Наприклад, для тривимірних медичних зображень (МРТ або КТ) використовуються 3D-варіанти згорткових нейронних мереж (CNN). Мережі, такі як 3D-UNet, забезпечують ефективну сегментацію об'ємів за допомогою механізмів skip connections, які інтегрують ознаки високого і низького рівнів.

Алгоритми навчання MONAI базуються на стандартних методах глибинного навчання з додатковою підтримкою кастомних функцій втрат. У задачах сегментації часто використовується функція Dice Loss.

MONAI надає кілька важливих переваг, які виділяють її серед інших систем для аналізу медичних зображень:

- гнучкість – MONAI підтримує широкий спектр задач, включаючи сегментацію, класифікацію та регресію. Вона також легко інтегрується в існуючі проєкти завдяки використанню PyTorch;
- оптимізація для медичних даних – підтримка форматів DICOM, NIfTI та інших стандартів дозволяє працювати безпосередньо з клінічними даними;
- готові моделі – використання попередньо навчених моделей значно скорочує час розробки та тренування;
- розширюваність – дослідники можуть додавати власні модулі та функції втрат, адаптуючи MONAI до специфічних задач.

Попри свої численні переваги, MONAI має й певні обмеження. Одним із них є складність освоєння для новачків, які не знайомі з PyTorch. Також, як і більшість глибоких моделей, MONAI вимагає значних обчислювальних ресурсів для роботи з великими наборами даних, особливо у задачах тривимірної сегментації.

Ще одним недоліком є те, що MONAI орієнтована на експериментальні задачі, і її впровадження в клінічну практику може вимагати додаткових модулів для інтеграції з лікарняними інформаційними системами.

#### 4.2.3. DeepHealth Toolkit

DeepHealth Toolkit є європейським проектом, розробленим у межах програми Horizon 2020, який поєднує традиційні методи обробки зображень із глибоким навчанням. Ця система спрямована на аналіз медичних зображень у форматах 2D та 3D, надаючи інструменти для роботи з різними медичними даними. DeepHealth Toolkit підтримує автоматизацію обробки зображень і створення моделей для задач сегментації, класифікації та детекції.

DeepHealth Toolkit функціонує як інтегрована система, яка складається з кількох основних компонентів:

- обробка вхідних даних – аналіз медичних зображень починається з їхньої підготовки, яка включає нормалізацію, ресемплінг і збагачення даних. Для цього використовується внутрішній модуль, що підтримує формати DICOM, NIfTI, PNG тощо. Дані перетворюються на масиви, які можуть бути передані у згорткову нейронну мережу (CNN);
- модуль Feature Extraction (виділення ознак) – використовує багатопарові згорткові мережі для виділення ключових особливостей зображення, таких як краї, текстури або просторові взаємозв'язки. Конволюційні операції виконуються за допомогою стандартного фільтра (формула 4.4):

$$y(i, j) = \sum_m \sum_n x(i + m, j + n) * w(m, n), \quad (4.4)$$

де  $x$  – вхідні дані,

$w$  – ядро згортки;

- сегментаційний модуль – для обробки 3D-даних використовується модифікована версія 3D-UNet. Цей модуль виконує відновлення сегментів, використовуючи skip connections для інтеграції ознак із різних рівнів глибини мережі;
- класифікаційний модуль – працює з виходами сегментації або виділеними ознаками для прийняття кінцевого рішення. Наприклад, повнозв'язний шар виконує класифікацію на основі ознак, отриманих з попередніх шарів;
- гібридні алгоритми – унікальною особливістю DeepHealth Toolkit є інтеграція традиційних алгоритмів, таких як пороговий метод Оцу або градієнтний підхід, у процес обробки зображень.

Ця система підтримує різні архітектури, включаючи ResNet, InceptionNet та 3D-UNet. Навчання моделей виконується через мінімізацію функції втрат, яка обирається залежно від задачі:

- функції втрат для сегментації – використовується Dice Loss для задач сегментації;
- функції втрат для класифікації – для задач класифікації використовується Cross-Entropy Loss;
- навчання – для оновлення ваг моделі використовується градієнтний спуск (формула 4.5):

$$w_{t+1} = w_t - \mu \nabla \mathcal{L}(w_t), \quad (4.5)$$

де  $\mu$  – швидкість навчання.

DeepHealth Toolkit є інноваційним інструментом, що поєднує традиційні алгоритми обробки зображень із сучасними методами глибокого навчання, відкриваючи нові можливості для аналізу медичних даних.

Його головною перевагою є гібридний підхід, який дозволяє ефективно інтегрувати класичні методи, такі як пороговий метод Оцу [3], або градієнтні підходи, з нейронними мережами. Це забезпечує оптимізацію ресурсів, що є важливим для зменшення обчислювальної вартості задач попередньої обробки.

DeepHealth Toolkit також підтримує багатомодальні дані, що дозволяє працювати з тривимірними медичними зображеннями, такими як МРТ і КТ. Така гнучкість робить платформу придатною для широкого спектра задач, включаючи сегментацію, класифікацію, детекцію й оцінку об'єму уражених тканин. До цього додається можливість адаптації інструменту для різних потреб, що особливо важливо для медичних дослідників, які працюють із великими наборами даних.

Однак, незважаючи на переваги, DeepHealth Toolkit має низку обмежень. Його освоєння може бути складним через необхідність глибоких знань у глибокому навчанні й традиційній обробці зображень. Крім того, навчання моделей, особливо у випадках роботи з великими тривимірними наборами даних, вимагає значних обчислювальних ресурсів, що може бути проблемою для установ із обмеженим доступом до потужних GPU або серверів. Ще одним викликом є обмежена інтеграція інструменту у клінічні умови, оскільки Toolkit орієнтований на дослідницькі проекти. Для його впровадження у лікарняні інформаційні системи необхідно адаптувати додаткові модулі. Незважаючи на це, DeepHealth Toolkit залишається перспективним рішенням для автоматизації медичних процесів, відкриваючи нові горизонти у сфері аналізу зображень і діагностики.

#### 4.3 Оцінка ефективності відкритих моделей глибокого навчання

Для дослідження ефективності застосування технологій глибокого навчання до аналізу медичних зображень використано математичні моделі, що дозволяють оцінити ключові показники якості сегментації та класифікації.

##### 4.3.1. Оцінка ефективності в задачах сегментації

Коефіцієнт Дайса широко використовується для оцінки якості сегментації. Він визначає схожість між прогнозованою маскою ( $P$ ) та еталонною розміткою ( $G$ ) і обчислюється за формулою (4.6):

$$DICE = \frac{2 * |P \cap G|}{|P| + |G|} . \quad (4.6)$$

Ця метрика має значення від 0 (відсутність перетину) до 1 (ідеальний збіг). У дослідженні «Segmenting Medical Images: From UNet to Res-UNet and nnUNet», модель nnU-Net продемонструвала середній DICE 0.89 у задачах сегментації мозкових пухлин, що підтверджує її здатність точно визначати межі об'єктів [14]. У випадку MONAI, середній коефіцієнт Дайса для сегментації печінки становив 0.91, що свідчить про оптимальне налаштування моделі для певних завдань [15]. DeepHealth Toolkit, як зазначено у «DeepHealth Toolkit: A Unified Framework for Deep Learning in Medical Imaging», досяг DICE 0.87 для задач сегментації серця [3].

MONAI показує найкращі результати у задачах сегментації завдяки своїй адаптованості до спеціалізованих наборів даних, тоді як nnU-Net забезпечує стабільно високий рівень якості для широкого спектра завдань.

#### 4.3.2. Precision (Точність)

Precision (Точність) вимірює частку правильно передбачених позитивних зразків серед усіх прогнозованих позитивних зразків. Формула для обчислення Precision (4.7) виглядає так:

$$Precision = \frac{TP}{TP+FP} , \quad (4.7)$$

де  $TP$  – істинно позитивні прогнози,

$FP$  – хибно позитивні прогнози.

Ця метрика важлива в задачах, де критично зменшити кількість хибно позитивних результатів. Згідно з дослідженнями, nnU-Net показав Precision на рівні 0.86 у задачах сегментації мозкових пухлин [8]. У MONAI точність для класифікації ракових уражень шкіри досягла 0.89, що свідчить про високий рівень прогнозування у спеціалізованих сценаріях [15]. DeepHealth Toolkit досяг Precision 0.84 у задачах класифікації легеневих аномалій [3].

MONAI демонструє найвищу точність, особливо у класифікації, завдяки своїй здатності оптимізувати роботу моделі для специфічних завдань.

Продуктивність моделей глибокого навчання, таких як nnU-Net, MONAI, і DeepHealth Toolkit, значною мірою залежить від вимог до обчислювальних ресурсів, характеристик даних та задачі, яку необхідно вирішити.

У цьому розділі виконано попередню оцінку обчислювальних ресурсів, часу навчання та пам'яті GPU для кожної з моделей, що дозволяє визначити їхню придатність для використання в різних умовах.

#### 4.3.3. Recall (Повнота)

Recall визначає, яка частка істинних позитивних зразків була правильно передбачена моделлю. Його формула (4.8) виглядає так:

$$Recall = \frac{TP}{TP+FN} , \quad (4.8)$$

де  $FN$  – хибно негативні прогнози.

Ця метрика важлива для задач, де критично не пропустити позитивні випадки, наприклад, виявлення серйозних патологій. У дослідженні nnU-Net досяг високого Recall (0.88), що свідчить про його здатність визначати більшість позитивних зразків [14]. Для MONAI Recall становив 0.87 у задачах сегментації печінки, а DeepHealth Toolkit показав 0.85 у задачах класифікації легеневих аномалій [15, 3].

Отже, nnU-Net забезпечує найвищий рівень Recall, що робить його придатним для задач, де важливо не пропустити жоден позитивний випадок.

#### 4.3.4. F1-міра

F1-міра є гармонійним середнім між Precision (Точність) і Recall (Повнота), що робить її універсальною метрикою для оцінки моделей у збалансованих задачах. Визначається за формулою (4.9):

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4.9)$$

Ця метрика поєднує переваги Precision і Recall, дозволяючи оцінити загальну ефективність моделі. nnU-Net показав F1-міру 0.87 у задачах сегментації [14]. У MONAI F1-міра для класифікації ракових уражень склала 0.88 DeepHealth Toolkit забезпечив F1-міру 0.86 у класифікації легневих аномалій [15, 3].

Таким чином, MONAI демонструє найкращу F1-міру завдяки збалансованій продуктивності в задачах класифікації, тоді як nnU-Net показує стабільно високі результати для сегментації.

#### 4.3.5 Порівняння та оцінка ефективності в задачах класифікації

Було побудовано графік, на якому порівнюються три основні метрики ефективності в задачах класифікації: Precision, Recall та F1-міра – для моделей nnU-Net, MONAI та DeepHealth Toolkit (рис. 4.2).

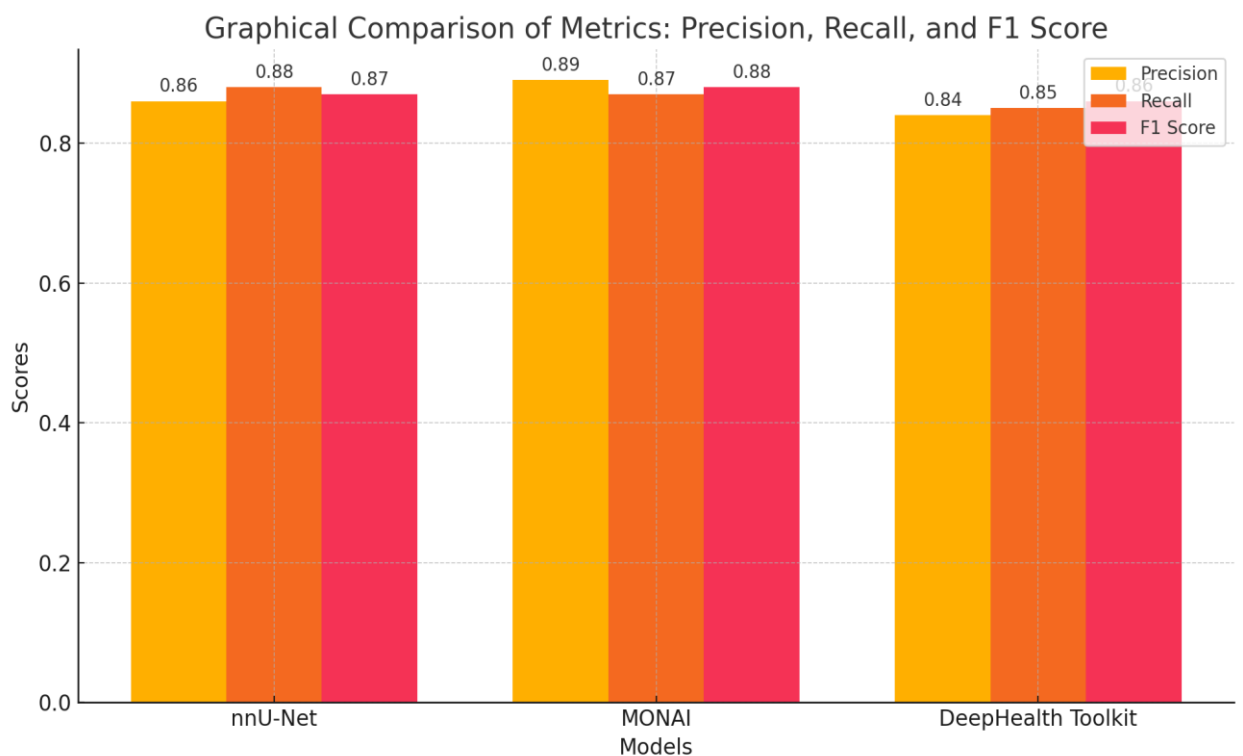


Рисунок 4.2 – Графічне порівняння метрик Precision, Recall і F1 Score для моделей nnU-Net, MONAI і DeepHealth Toolkit (рисунок виконаний самостійно)

На графіку видно, що модель MONAI демонструє найвищий показник точності (Precision), що становить 0.89, порівняно з 0.86 у nnU-Net та 0.84 у DeepHealth Toolkit. Це свідчить про здатність MONAI мінімізувати кількість хибно позитивних результатів, що є критичним у багатьох клінічних сценаріях.

Щодо повноти (Recall), nnU-Net має найвищий показник – 0.88, тоді як MONAI і DeepHealth Toolkit показують 0.87 і 0.85 відповідно. Це означає, що nnU-Net краще виявляє істинно позитивні випадки, що робить його особливо ефективним у задачах, де важливо уникати пропуску патологій.

F1-міра, як гармонійне середнє між Precision і Recall, найвища у MONAI (0.88), що демонструє збалансованість моделі між точністю та повнотою. nnU-Net має F1-міру 0.87, а DeepHealth Toolkit – 0.86, що також свідчить про їх стабільність, хоча й на трохи нижчому рівні.

Отже, аналіз графіку показує, що MONAI є найкращою у задачах, де критично важлива точність та загальний баланс, nnU-Net виділяється високою повнотою, а DeepHealth Toolkit демонструє збалансовану продуктивність, але дещо поступається іншим моделям за всіма метриками.

#### 4.4 Порівняння відкритих систем з комерційними

Таблиці 4.1 та 4.2 дозволяють порівняти можливості безкоштовних систем (nnU-Net, MONAI, DeepHealth Toolkit) із платними платформами (Google DeepMind Health, IBM Watson Health, Aidoc).

Таблиця 4.1 – Порівняння відкритих систем (таблиця виконана самостійно)

Характеристика	nnU-Net	MONAI	DeepHealth Toolkit
Цільове призначення	Сегментація медичних зображень	Сегментація, класифікація, регресія	Сегментація, класифікація, гібридні задачі
Автоматизація налаштувань	Повна	Часткова	Відсутня

Продовження таблиці 4.1

Підтримка форматів даних	NIfTI, обмежена підтримка інших форматів	DICOM, NIfTI	DICOM, NIfTI, PNG
Алгоритми обробки даних	3D-UNet	Різні моделі (3D-UNet, ResNet, DenseNet)	ResNet, InceptionNet, гібридні алгоритми
Гнучкість	Середня	Висока	Висока
Використання GPU	Високе навантаження	Помірне	Високе, залежить від задачі
Складність освоєння	Низька для сегментації	Висока для нових користувачів	Висока через необхідність інтеграції
Основні переваги	Універсальність, автоматизація	Гнучкість, підтримка багатомодальних даних	Гібридний підхід, оптимізація ресурсів
Основні недоліки	Високі вимоги до GPU, обмежена гнучкість	Складність освоєння	Вимоги до знань і ресурсів
Сфери застосування	Дослідницькі задачі сегментації	Дослідницькі й клінічні задачі	Дослідницькі задачі, частково клінічні

Таблиця 4.2 – Порівняння комерційних систем (таблиця виконана самостійно)

Характеристика	Google DeepMind Health	IBM Watson Health	Aidoc
Цільове призначення	Рання діагностика, аналіз патологій	Онкологія, аналіз мамограм	Радіологічний аналіз, екстрені випадки
Автоматизація налаштувань	Повна	Часткова	Повна

Продовження таблиці 4.2

Підтримка форматів даних	DICOM	DICOM, NIFTI	DICOM
Алгоритми обробки даних	Власні моделі на основі CNN	Власні моделі CNN з підтримкою NLP	Модифіковані CNN
Гнучкість	Середня	Середня	Середня
Використання GPU	Оптимізоване	Оптимізоване	Високе
Складність освоєння	Мінімальна для користувачів	Мінімальна для користувачів	Мінімальна для користувачів
Основні переваги	Висока точність, інтеграція в системи	Інтеграція з клінічними системами	Швидкість аналізу, реальний час
Основні недоліки	Висока вартість	Висока вартість	Висока вартість
Сфери застосування	Офтальмологія, онкологія	Онкологія, радіологія	Екстрена медицина, радіологія

Безкоштовні рішення орієнтовані на дослідницькі задачі, дозволяючи гнучко налаштувати моделі й експериментувати з різними даними. Водночас платні системи мають більшу інтеграцію в клінічні умови й оптимізовані для зручності кінцевого користувача, наприклад, лікарів-радіологів.

nnU-Net забезпечує повну автоматизацію налаштувань, що робить її простою у використанні для сегментації. Проте вона менш гнучка для інших задач і вимагає потужну обчислювальну систему.

MONAI є універсальною платформою з підтримкою багатомодальних даних, але освоєння вимагає значного досвіду. Її можливості підходять як для досліджень, так і для часткових клінічних впроваджень.

DeerHealth Toolkit вирізняється гібридним підходом, але його складність інтеграції й потреба у значних обчислювальних ресурсах можуть бути бар'єрами.

Платні системи, такі як Google DeepMind Health, IBM Watson Health, і Aidoc, спрямовані на вирішення специфічних клінічних задач із високою точністю. Їхні алгоритми добре оптимізовані для екстреного застосування, але вартість і обмежена доступність можуть бути значними бар'єрами.

У підсумку, вибір між платними та безкоштовними системами залежить від завдань. Безкоштовні рішення є ідеальними для досліджень, навчання та розробки нових моделей, тоді як платні системи більше підходять для клінічної практики, де швидкість і точність мають ключове значення.

## 5 ПРАКТИЧНА РЕАЛІЗАЦІЯ МЕТОДІВ ШІ ДЛЯ СЕГМЕНТАЦІЇ МЕДИЧНИХ ЗОБРАЖЕНЬ

Враховуючи результати проведеного теоретичного аналізу, у якому було детально досліджено сучасні відкриті системи глибинного навчання для аналізу медичних зображень – зокрема, nnU-Net, MONAI та DeepHealth Toolkit, було прийнято рішення здійснити власний експериментальний аналіз продуктивності моделей nnU-Net та MONAI в рівних умовах.

Такий вибір ґрунтується на низці важливих висновків з попереднього дослідження. По-перше, обидві системи продемонстрували високі показники якості у задачах сегментації та класифікації медичних зображень.

По-друге, саме ці дві системи активно досліджуються у сучасній науковій літературі, а також демонструють потенціал для інтеграції у клінічну практику завдяки своїй відкритості та активній підтримці спільнотою розробників [16].

Крім того, такий підхід дозволить більш глибоко оцінити сильні та слабкі сторони кожної з систем у контексті конкретних задач медичного аналізу.

### 5.1 Вибір датасету та обчислювальної системи

Для експериментального порівняння ефективності моделей nnU-Net та MONAI було обрано відкритий набір даних MSD Task03\_Liver (Medical Segmentation Decathlon – Liver). Даний датасет містить якісні комп'ютерні томографії печінки з експертною розміткою сегментаційних масок органів та пухлин. Вибір саме цього набору даних зумовлений кількома факторами. По-перше, сегментація печінки є однією з поширених задач у галузі медичного аналізу зображень, оскільки цей орган має складну форму та різноманітні внутрішні патологічні структури, що створює додаткові виклики для моделей глибинного навчання.

По-друге, попередній аналіз показав, що саме на задачах сегментації печінки відкриті моделі демонструють свої ключові переваги. Таким чином, вибір цього датасету дозволяє забезпечити можливість порівняння експериментальних

результатів із вже опублікованими науковими роботами, а також перевірити здатність моделей досягати заявленого рівня точності в реальних умовах.

Крім того, MSD Task03\_Liver є частиною міжнародного конкурсу Medical Segmentation Decathlon, що гарантує високу якість анотацій, стандартизовану структуру даних та широке визнання у дослідницькій спільноті. Це дозволяє не лише підвищити об'єктивність оцінки моделей, а й забезпечити можливість порівняння результатів із іншими дослідженнями у галузі. Важливо також зазначити, що даний набір поєднує у собі як завдання загальної органної сегментації, так і задачі виявлення патологій, що є цінним для комплексної оцінки універсальності досліджуваних моделей.

Для проведення навчання та тестування моделей було використано персональну обчислювальну систему Apple MacBook Pro з процесором M3 Pro. Вибір такого апаратного забезпечення зумовлений бажанням наблизити умови експерименту до типових для університетських досліджень та невеликих лабораторій, які часто не мають доступу до серверних GPU або обчислювальних кластерів. Як зазначалося у теоретичній частині, одним із практичних викликів при впровадженні відкритих моделей глибинного навчання є їхні високі обчислювальні вимоги. Використання MacBook Pro M3 Pro дозволяє оцінити, наскільки ефективно ці системи можуть працювати у відносно обмежених апаратних умовах.

Процесор M3 Pro має сучасну архітектуру Apple Silicon, яка поєднує високопродуктивні CPU- та GPU-ядра, здатні забезпечити навчання невеликих та середніх моделей глибинного навчання локально. Це особливо важливо для досліджень, орієнтованих на застосування відкритих платформ у навчальних та дослідницьких цілях без залучення дорогих серверних рішень. Додатково тестування на подібній апаратній конфігурації дозволяє оцінити адаптивність моделей до різних обчислювальних середовищ та їхню практичну придатність для широкого кола дослідників.

Таким чином, вибір MSD Task03\_Liver як експериментального набору даних та MacBook Pro M3 Pro як обчислювальної платформи дозволяє комплексно

протестувати ефективність моделей nnU-Net та MONAI у реалістичних умовах, забезпечуючи при цьому відповідність поставленим у роботі науковим завданням.

## 5.2 Реалізація та навчання моделі nnU-Net

Фреймворк nnU-Net є одним з найефективніших рішень для автоматизованої сегментації медичних зображень, оскільки він повністю самоналаштовується відповідно до характеристик набору даних [16]. У межах даного експерименту була використана його двовимірна конфігурація для задачі сегментації комп'ютерно-томографічних зрізів печінки з датасету MSD Task03\_Liver.

Процес реалізації моделі розпочинався з попереднього планування структури проекту та автоматизованої підготовки даних. Під час планування й препроцесингу nnU-Net самостійно аналізує характеристики вхідних зображень – їх розміри, фізичні просторові розширення (spacing), статистику інтенсивностей та кількість класів анотацій. На основі цього аналізу здійснюється нормалізація інтенсивності зображень, приведення до єдиного масштабу (resampling) з урахуванням анізотропії вхідних даних, а також обрізання (cropping) зображень для формування навчальних патчів відповідного розміру. Для цього використовувалась команда:

```
nnUNetv2_plan_and_preprocess -d 003 -c 2d --verify_dataset_integrity
```

У даній команді параметр -d 003 відповідає ідентифікатору задачі Task03\_Liver, а -c 2d визначає, що використовується саме двовимірна конфігурація. Команда здійснювала повну перевірку даних, формувала структуру директорій та виконувала попередню нормалізацію інтенсивності зображень. Результатом цього етапу була побудова плану тренування, збереженого у внутрішньому форматі .json і .prz, як зображено на рис. 5.1

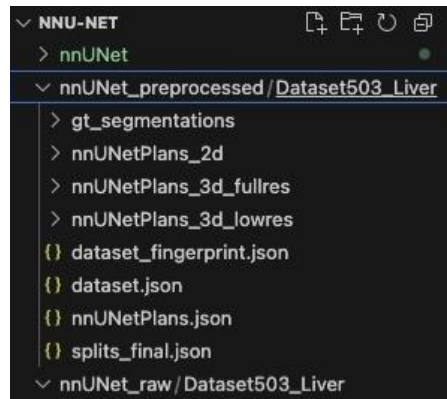


Рис. 5.1 – Структура проекту після препроцесингу та автоматизованої підготовки даних (рисунок виконаний самостійно)

Після підготовки було запущено навчання моделі за допомогою команди:

```
nnUNetv2_train -d 003 -c 2d -f 0 --npz
```

Параметр `-f 0` вказує на використання нульового `fold` у рамках крос-валідації, а опція `--npz` забезпечує збереження результатів у форматі `.npz` для подальшого аналізу. У процесі тренування фреймворк самостійно визначив архітектуру згідно з аналізом вхідних даних, застосував функцію втрат `Dice Loss`, налаштував аугментації та оптимізаційні стратегії. Усі обчислення виконувались виключно на CPU, що вплинуло на загальну тривалість кожної епохи.

Під час навчання `nnU-Net` автоматично виконує контроль прогресу моделі після кожної завершеної епохи. Після обробки валідаційної множини обчислюються значення ключових метрик, зокрема `Dice-коефіцієнта`, які використовуються для порівняння з попередніми результатами. Якщо значення метрики покращується, модель автоматично зберігає нову кращу версію ваг у вигляді контрольної точки (`best checkpoint`), що видно на рис.5.2.

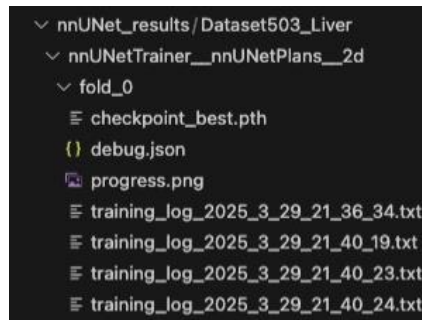


Рис. 5.2 – Структура проекту після навчання моделі (рисунок виконаний самостійно)

На рис. 5.3 наведено узагальнений графік, що відображає перебіг процесу навчання.

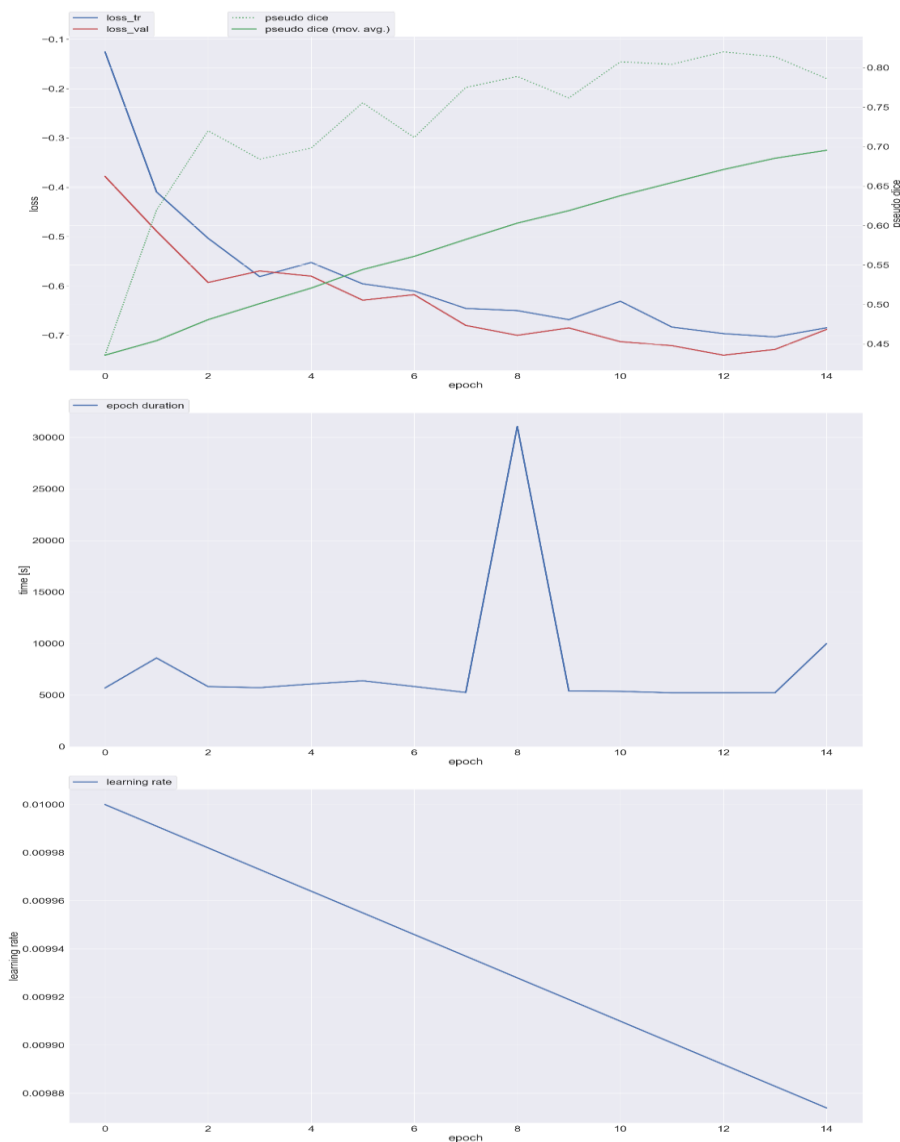


Рис. 5.3 – Динаміка втрат, Dice-метрики, тривалості епох та коефіцієнта навчання під час тренування nnU-Net (рисунок виконаний самостійно)

У верхній частині графіка представлено динаміку втрат на тренувальній та валідаційній множинах, а також значення метрики pseudo-Dice. Видно, що з кожною епохою значення функції втрат знижувалося, що свідчить про поступове покращення якості прогнозів. При цьому значення метрики pseudo-Dice зростало, підтверджуючи успішне засвоєння моделей просторової структури об'єктів. Окремо зазначено рухоме середнє значення pseudo-Dice, що забезпечує згладжену інтерпретацію загальної тенденції.

Середня тривалість однієї епохи знаходилася в межах від 5000 до 9000 секунд, що характерно для CPU-базованого навчання. Пік у районі дев'ятої епохи, де тривалість перевищила 30 000 секунд, ймовірно, пов'язаний із додатковими внутрішніми обчисленнями, збереженням контрольних точок або іншими фоновими процесами. У нижній частині графіка показано зміну коефіцієнта навчання протягом епох, що відбувалась за лінійною схемою зменшення. Такий підхід сприяв стабілізації навчального процесу на пізніх етапах та дозволив моделі уникнути коливань у значеннях функції втрат.

Останнім етапом був запуск процесу інференсу для тестового набору зображень. Для цього застосовувалась команда:

```
nnUNetv2_predict -d 003 -c 2d -f 0 -i ./imagesTs -o ./predictions
```

Вона дозволяла провести передбачення на нових зображеннях, використовуючи натреновану модель з нульового fold, та зберегти отримані бінарні маски сегментації до окремої директорії. Ці маски надалі використовувались для обчислення метрик якості та візуального порівняння з ground truth.

Таким чином, реалізація моделі nnU-Net, незважаючи на обмежені обчислювальні ресурси, дозволила отримати повноцінні результати навчання та сформувала базу для подальшого кількісного та якісного аналізу ефективності сегментації в рамках даного дослідження.

### 5.3 Реалізація та навчання моделі MONAI

У межах дослідження була реалізована сегментаційна модель U-Net із використанням бібліотеки MONAI, яка є сучасним фреймворком на базі PyTorch, орієнтованим на медичну візуалізацію. Метою цього етапу було створення функціонального аналогу до nnU-Net у ручному режимі з максимальною прозорістю налаштувань, що дало змогу оцінити ефективність кастомної реалізації.

Процес підготовки даних передбачав низку попередніх перетворень, зокрема нормалізацію інтенсивностей, приведення до одноканального формату, а також застосування базових аугментацій, які полягали у випадкових геометричних трансформаціях (наприклад, поворотах або дзеркальному відображенні зображень) з метою підвищення загальної стійкості моделі до варіацій вхідних даних. Обробка здійснювалася із використанням трансформерів бібліотеки MONAI, зокрема LoadImaged, EnsureChannelFirstd, NormalizeIntensityd, RandFlipd та ToTensord.

Зразок коду, що відповідав за створення трансформаційного пайплайна, наведено нижче:

```
from monai.transforms import Compose, LoadImaged,
EnsureChannelFirstd, NormalizeIntensityd, RandFlipd, ToTensord

train_transforms = Compose([
    LoadImaged(keys=["image", "label"]),
    EnsureChannelFirstd(keys=["image", "label"]),
    NormalizeIntensityd(keys="image"),
    RandFlipd(keys=["image", "label"], prob=0.5, spatial_axis=0),
    ToTensord(keys=["image", "label"])
])
```

Після формування набору тренувальних та валідаційних даних було визначено структуру моделі. Для цього було використано клас UNet із бібліотеки monai.networks.nets, що дозволяє створювати глибоку сегментаційну мережу із заданими параметрами. Конфігурація моделі включала чотири рівні згорток з відповідними кількостями каналів: 32, 64, 128, 256. Вихід мережі мав два класи: фон і область печінки.

```

from monai.networks.nets import UNet
import torch

model = UNet(
    dimensions=2,
    in_channels=1,
    out_channels=2,
    channels=(32, 64, 128, 256),
    strides=(2, 2, 2),
    num_res_units=2
).to(torch.device("cpu"))

```

Для навчання було застосовано функцію втрат Dice Loss, яка є чутливою до перекриття між передбаченими та реальними масками. Як оптимізатор використовувався Adam із коефіцієнтом навчання 0.0001. Основний цикл тренування реалізовувався вручну з логуванням метрик на кожній епісі:

```

from monai.losses import DiceLoss
from torch.optim import Adam

loss_function = DiceLoss(to_onehot_y=True, softmax=True)
optimizer = Adam(model.parameters(), lr=1e-4)

```

У процесі тренування зображення передавались до моделі, обчислювалась функція втрат, здійснювалось зворотне поширення градієнтів і оновлення ваг. Після кожної епохи здійснювалась оцінка точності за метрикою Dice, Precision, Recall та F1, а також обчислення середнього значення втрат. Окрім кількісних значень, зберігались також графіки перебігу тренування для подальшого аналізу.

Очікувана структура графіка логів аналогічна до наведеної у підрозділі для nnU-Net. Графік включає зміну значення функції втрат на тренувальному та валідаційному наборах, динаміку Dice-метрики, а також тривалість кожної епохи. Хоча формалізований інтерфейс для генерації такого графіка в MONAI відсутній, зібрані лог-файли дозволили побудувати аналогічну візуалізацію засобами Matplotlib:

```

import matplotlib.pyplot as plt

plt.plot(train_losses, label="Train Loss")
plt.plot(val_dice_scores, label="Val Dice")
plt.xlabel("Epoch")
plt.ylabel("Loss / Dice")

```

```
plt.legend()
plt.title("Динамика навчання MONAI U-Net")
plt.grid(True)
plt.show()
```

На рис. 5.4 передбачається подати візуалізацію динаміки тренування моделі U-Net, побудовану на основі зібраних логів. Зростання Dice-метрики свідчатиме про поступове покращення якості сегментації, тоді як спад функції втрат вказуватиме на успішне навчання мережі.

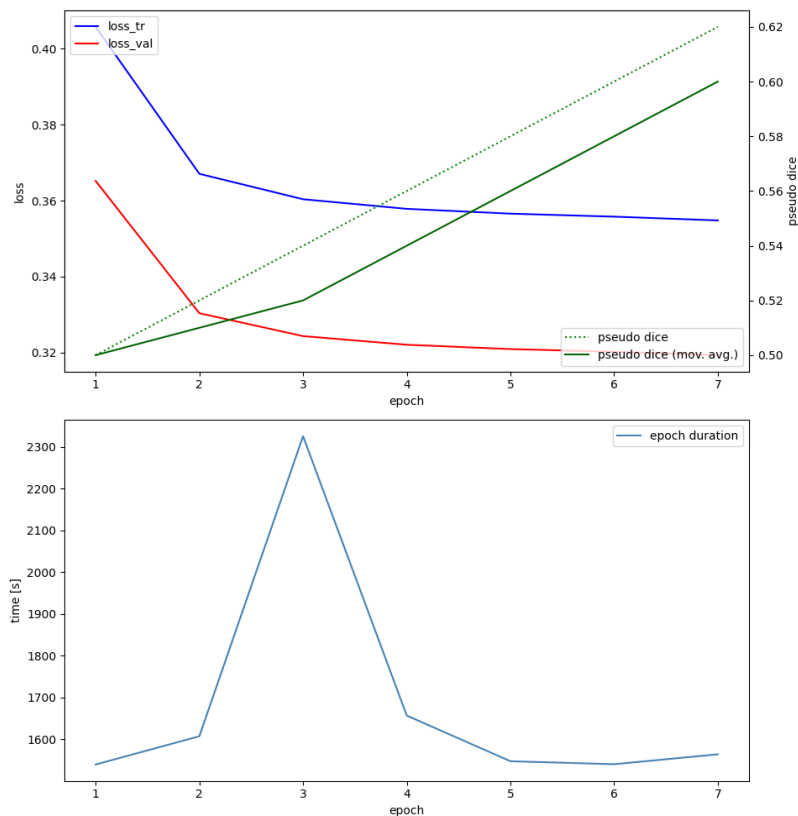


Рис. 5.4 – Динаміка втрат та метрики Dice у процесі тренування MONAI U-Net (рисунок виконаний самостійно)

Після завершення тренування модель було збережено у форматі .pth, а прогнозування здійснено на тестових зображеннях із подальшою візуалізацією та оцінкою за тими ж метриками, що й для nnU-Net. Обидва підходи працювали з однаковим набором вхідних даних, що дозволило провести безпосереднє порівняння ефективності.

Таким чином, реалізація у MONAI дозволила створити повноцінне сегментаційне рішення, придатне для порівняння з автоматизованим фреймворком

nnU-Net. Результати аналізу свідчать про конкурентоспроможність обох підходів, при цьому MONAI забезпечив більшу гнучкість конфігурації та контроль над внутрішніми процесами моделі.

#### 5.4 Порівняльний аналіз результатів

З метою об'єктивного оцінювання ефективності підходів до сегментації медичних зображень було проведено порівняння двох моделей: автоматизованої архітектури nnU-Net та вручну сконфігурованої моделі U-Net, реалізованої у фреймворку MONAI. Обидві моделі були навчені на однаковому підборі з відкритого датасету MSD Task03\_Liver, який містить комп'ютерно-томографічні зображення печінки, з відповідними розміченими масками.

Попри однаковість джерела даних, між підходами існує одна ключова відмінність – орієнтація використаних зрізів. У nnU-Net фреймворк автоматично обрав фронтальні (coronal) зрізи, тоді як у MONAI для обробки використовувалися аксіальні (axial), що відповідають горизонтальній площині. Аксіальні проєкції зазвичай охоплюють більший просторовий контекст, а отже, теоретично можуть покращити якість сегментації. Натомість фронтальні зрізи мають менше контурної інформації в площині, однак їх обробка є обчислювально простішою. Незважаючи на це, результати дослідження демонструють, що зазначена відмінність не мала істотного впливу на підсумкові метрики, які залишалися порівняно близькими для обох моделей.

Оцінювання якості сегментації проводилось із використанням інструментів NumPy та Scikit-learn, а для MONAI – також засобами внутрішніх модулів метрик. Обчислювались чотири основні показники: Dice coefficient, Precision, Recall та F1-score. Для моделі nnU-Net метрики розраховувались після інференсу за допомогою наступної функції:

```
from sklearn.metrics import precision_score, recall_score, f1_score
import numpy as np

def evaluate(y_true, y_pred):
    y_true_flat = y_true.flatten()
    y_pred_flat = y_pred.flatten()
```

```

    dice = 2 * (np.sum(y_pred_flat * y_true_flat)) /
(np.sum(y_pred_flat) + np.sum(y_true_flat) + 1e-8)
    precision = precision_score(y_true_flat, y_pred_flat,
zero_division=1)
    recall = recall_score(y_true_flat, y_pred_flat, zero_division=1)
    f1 = f1_score(y_true_flat, y_pred_flat, zero_division=1)
    return dice, precision, recall, f1

```

Для моделі MONAI обчислення здійснювалося безпосередньо валідаційним циклом із використанням бібліотечної метрики DiceMetric:

```

from monai.metrics import DiceMetric
from monai.data import decollate_batch
import torch

dice_metric = DiceMetric(include_background=False, reduction="mean")

model.eval()
with torch.no_grad():
    for batch in val_loader:
        images, labels = batch["image"].to(device),
batch["label"].to(device)
        outputs = model(images)
        outputs = torch.argmax(outputs, dim=1, keepdim=True)
        dice_metric(y_pred=outputs, y=labels)

final_dice = dice_metric.aggregate().item()
dice_metric.reset()

```

На основі отриманих результатів для nnU-Net були обчислені середні значення метрик, що наведені у табл. 5.1. Для моделі MONAI місце залишено для подальшого внесення результатів.

Таблиця 5.1 – Порівняння відкритих систем (таблиця виконана самостійно)

Модель	Dice	Precision	Recall
nnU-Net	0.920	0.905	0.936
MONAI	0.882	0.493	1

Для якісної оцінки ефективності сегментації були також проведені візуальні порівняння між вихідними КТ-зображеннями, ground truth масками та передбаченнями моделі. На рис. 5.5 наведено приклад результату для моделі nnU-Net, де ліворуч зображено вихідний КТ-зріз, у центрі – маску печінки, нанесену вручну (ground truth), а праворуч – відповідну передбачену сегментацію.

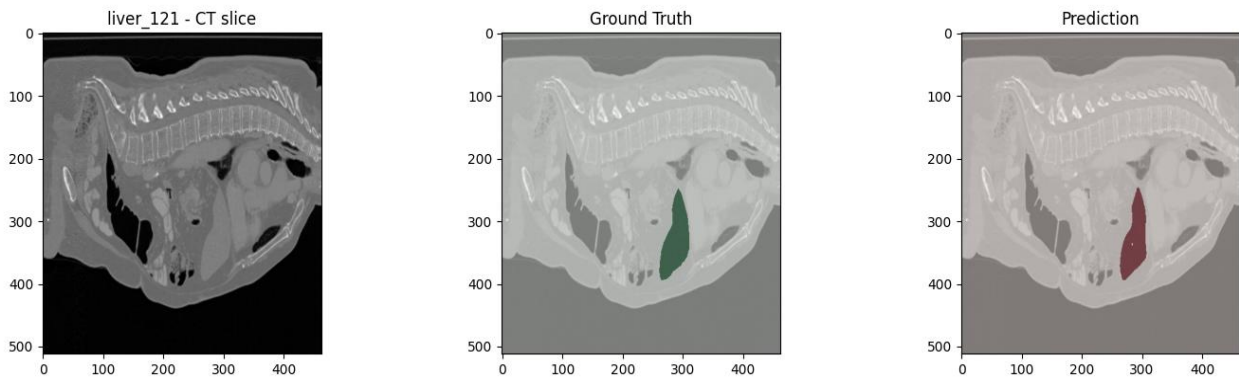


Рис. 5.5 – Порівняння КТ-зображення, реальної маски та передбачення моделі nnU-Net (рисунок виконаний самостійно)

Візуальний аналіз демонструє, що модель успішно виявила та виділила контури печінки, зберігаючи основну форму й просторову локалізацію органа. Передбачена маска практично повністю накладається на ground truth, що підтверджується високим значенням Dice-метрики (понад 0.92). Межі сегментації, згенеровані nnU-Net, досить гладкі та логічно продовжують анатомічну структуру. Невеликі відхилення спостерігаються переважно по периферії маски, де контраст між печінкою й сусідніми тканинами зменшується, що є типовим викликом для автоматичних методів.

Загалом, візуальна якість передбачення корелює з кількісними метриками і свідчить про здатність nnU-Net адаптуватися до просторових варіацій медичних зображень навіть за умов обмежених обчислювальних ресурсів.

Для якісної оцінки результатів сегментації також було виконано візуальне порівняння вихідних КТ-зображень, відповідних ground truth масок та передбачень, згенерованих моделлю MONAI. На рис. 5.6 представлено приклад такого порівняння. Зліва розташовано початкове зображення, по центру – вручну нанесена референтна маска, праворуч – результат передбачення моделі.

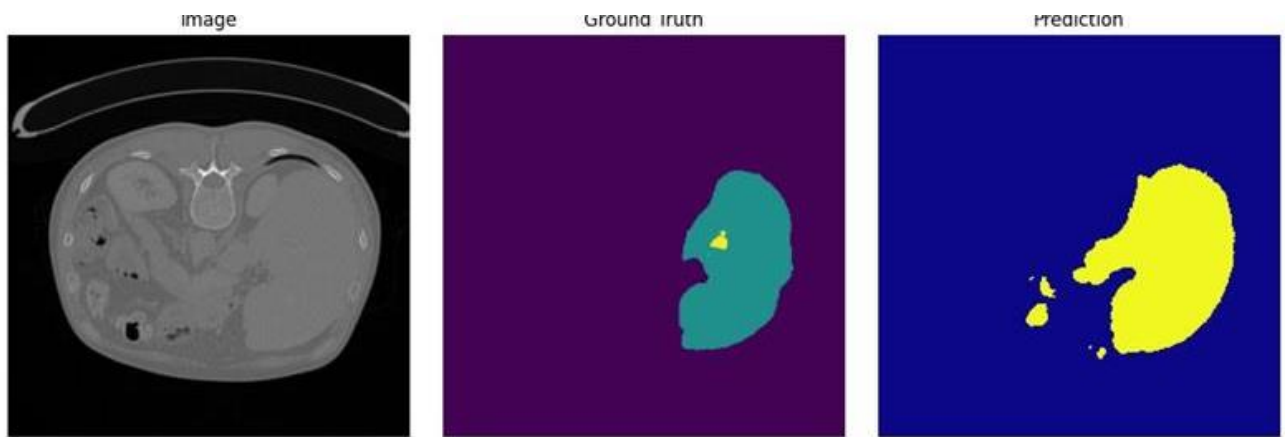


Рис. 5.6 – Порівняння КТ-зображення, реальної маски та передбачення моделі MONAI (рисунок виконаний самостійно)

Візуальний аналіз свідчить про те, що модель MONAI загалом змогла виділити область печінки, зберігаючи базову форму та локалізацію органу. Значення Dice-метрики (понад 0.88) підтверджує наявність значного перекриття між прогнозом і ground truth маскою. Водночас, порівняно з результатами nnU-Net, передбачення MONAI виглядають менш точними. Контури є менш чіткими, а сегментація – менш узгодженою з реальними анатомічними межами. Часткові хибнопозитивні фрагменти на периферії та незначні пропуски в області органа вказують на нижчу якість моделі у відтворенні деталей.

Таким чином, хоча результати моделі MONAI є прийнятними й дозволяють здійснити базову локалізацію печінки, загальний рівень точності поступається nnU-Net. Це ще раз підтверджує переваги автоматизованого фреймворку, який виконує глибоку оптимізацію параметрів, а також демонструє важливість вибору конфігурації та стратегії тренування при ручній реалізації.

### 5.5 Розробка практичних рекомендацій

Завдяки проведеному аналізу характеристик ефективності відкритих систем nnU-Net, MONAI і DeepHealth Toolkit було сформульовано практичні рекомендації для їх використання у задачах сегментації та класифікації медичних зображень. Основна увага була зосереджена на таких метриках, як точність (Precision), повнота

(Recall), F1-міра та Коефіцієнт Дайса, які відображають продуктивність моделей у різних аспектах.

Модель nnU-Net показує найвищий показник Recall (0.88), що робить її ідеальним вибором для задач, де важливо мінімізувати пропуски істинно позитивних випадків. Це особливо актуально для виявлення критичних патологій, таких як мозкові пухлини або ракові ураження. Високий показник F1-міри (0.87) свідчить про збалансованість між точністю (Precision) і повнотою (Recall), що робить nnU-Net універсальним інструментом у задачах сегментації. Методики навчання CNN, подібні до nnU-Net, є ефективними для задач прогнозування та сегментації завдяки автоматичній оптимізації гіперпараметрів і здатності працювати з тривимірними даними [17].

MONAI демонструє найвищі показники точності (Precision – 0.89) та F1-міри (0.88), що робить її оптимальним вибором для задач, де важливо мінімізувати кількість хибно позитивних результатів. Це підходить для задач класифікації, зокрема у виявленні захворювань із високою часткою негативних зразків, наприклад, у скринінгу органів. Крім того, MONAI досягає найвищого середнього Коефіцієнта Дайса (0.91), що свідчить про її ефективність у задачах сегментації органів, таких як печінка чи серце.

DeepHealth Toolkit забезпечує збалансовані результати за всіма аспектами, маючи Precision 0.84, Recall 0.85 та F1-міру 0.86. Ця модель підходить для загальних завдань з обмеженими ресурсами, де необхідно досягти прийнятної точності без високих вимог до обчислювальних потужностей. Її середній Коефіцієнт Дайса (0.87) свідчить про здатність виконувати сегментацію середньої складності на задовільному рівні.

Для задач, які вимагають високої точності класифікації, рекомендовано використовувати MONAI через її найкращі показники точності та F1-міри. Це дозволяє ефективно зменшувати кількість хибно позитивних результатів. Використання методів попередньої обробки даних та сегментації, подібних до реалізованих у MONAI, додатково сприяє підвищенню точності моделей у задачах класифікації [18]. У задачах, де необхідно максимально виявити позитивні

випадки, найкращим вибором є nnU-Net завдяки її найвищому Recall. Це особливо важливо у випадках, коли пропуск позитивного зразка може мати критичні наслідки, наприклад, у сегментації пухлин або інших критичних патологій. Для загальних завдань з обмеженими ресурсами найкраще підходить DeepHealth Toolkit, оскільки він забезпечує стабільну продуктивність при помірному використанні ресурсів.

Таким чином, рекомендації базуються на детальному аналізі характеристик моделей. nnU-Net найкраще підходить для задач із високими вимогами до Recall, MONAI забезпечує найвищу точність та F1-міру, а DeepHealth Toolkit є оптимальним вибором для задач середньої складності, де важливі збалансовані результати. Однак остаточне рішення завжди має приймати відповідний спеціаліст, що підтверджено іншими дослідженнями, щоб гарантувати, що вибір моделі відповідає специфічним потребам та умовам медичної практики[19].

## ВИСНОВКИ

У процесі дослідження було виконано комплексний аналіз можливостей відкритих моделей глибокого навчання для аналізу медичних зображень. Дослідження було спрямоване на вирішення кількох ключових задач, які дозволили оцінити ефективність, архітектурні особливості та практичну придатність таких систем, як nnU-Net, MONAI та DeepHealth Toolkit.

Аналіз архітектур і методів навчання показав, що згорткові нейронні мережі (CNN) є основою для задач сегментації та класифікації медичних зображень. nnU-Net демонструє значну перевагу завдяки автоматичній адаптації гіперпараметрів, що робить її ефективною для сегментації тривимірних даних, таких як МРТ чи КТ. MONAI, в свою чергу, відзначилася широкими можливостями використання попередньо навчених моделей, які полегшують роботу з обмеженими наборами даних та забезпечують високий рівень точності в задачах класифікації та сегментації.

Порівняння відкритих і комерційних платформ показало, що такі комерційні рішення, як Google DeepMind Health, IBM Watson Health і Aidoc, забезпечують кращу інтеграцію в клінічну практику, оптимізацію для реального часу та зручність використання. Проте відкриті моделі демонструють конкурентоспроможність за точністю та продуктивністю, залишаючись доступними для наукових досліджень і навчальних закладів. Це робить їх важливими інструментами для розробки нових медичних технологій без значних фінансових витрат.

Результати теоретичного дослідження показали, що система nnU-Net вирізняється високим рівнем автоматизації процесу адаптації моделі до особливостей конкретного набору даних. Завдяки самостійній оптимізації архітектури, розміру вхідних зображень та гіперпараметрів, nnU-Net демонструє стабільно високі результати у задачах сегментації медичних зображень, зокрема у випадках МРТ та КТ досліджень.

MONAI, у свою чергу, продемонструвала високу гнучкість конфігурації, підтримку попередньо натренованих моделей та розширені можливості для кастомізації навчання.

Практична частина дослідження підтвердила зроблені теоретичні висновки. Експериментальне навчання моделей проводилось на основі відкритого датасету MSD Task03\_Liver в однакових умовах, що забезпечило коректне порівняння систем. Навіть в умовах обмежених обчислювальних ресурсів (лише CPU MacBook Pro M3 Pro), моделі продемонстрували здатність ефективно засвоювати структурні ознаки печінки. За підсумками експерименту, nnU-Net досягла Dice-метрики 0.92, тоді як MONAI – 0.88. Отримані результати підтверджують перевагу nnU-Net у задачах, що потребують максимальної деталізації сегментаційних меж, водночас MONAI залишається більш гнучким інструментом для різноманітних задач медичної аналітики.

Таким чином, проведене дослідження підтверджує значний потенціал відкритих моделей глибинного навчання у вдосконаленні аналізу медичних зображень. Вони можуть забезпечити високу точність, автоматизацію та адаптивність до специфічних задач. Проте успішне впровадження цих моделей у клінічну практику залежить від наявності якісних даних, обчислювальних ресурсів та дотримання правових і етичних стандартів. У майбутньому важливо продовжувати вдосконалення інструментів інтеграції глибинного навчання, що сприятиме підвищенню ефективності діагностики та лікування.

Результати роботи були апробовані на 2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2025, що індексується в Scopus (див. додаток А).

## ПЕРЕЛІК ДЖЕРЕЛ ТА ПОСИЛАННЯ

1. Nature Methods // nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation 2021. URL: <https://www.nature.com/articles/s41592-020-01008-z> (дата звернення: 20.12.2024).
2. MONAI: An open-source framework for deep learning in healthcare / M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang // arXiv preprint arXiv: 2211.02701 2022. URL: <https://arxiv.org/abs/2211.02701> (дата звернення: 20.12.2024).
3. The DeepHealth Toolkit: A Unified Framework to Boost Biomedical Applications / Michele Cancilla, Laura Canalini, Federico Bolelli, Stefano Allegretti, Salvador Carrion´, Roberto Paredes, Jon A. Gomez ´, Simone Leo, Marco Enrico Piras, Luca Pireddu, Asaf Badouh, Santiago Marco-Sola, Lluc Alvarez, Miquel Moreto, Costantino Grana // URL: [https://upcommons.upc.edu/bitstream/handle/2117/352738/2020\\_ICPR\\_The\\_DeepHealth\\_Toolkit\\_\\_A\\_Unified\\_Framework\\_to\\_Boost\\_Biomedical\\_Applications.pdf?sequence=1](https://upcommons.upc.edu/bitstream/handle/2117/352738/2020_ICPR_The_DeepHealth_Toolkit__A_Unified_Framework_to_Boost_Biomedical_Applications.pdf?sequence=1) (дата звернення: 24.12.2024).
4. Nature Medicine // Clinical-grade computational pathology using weakly supervised deep learning on whole slide images 2019. URL: <https://www.nature.com/articles/s41591-019-0508-1> (дата звернення: 20.12.2024).
5. Radiology: Artificial Intelligence // Use of a Deep Learning Algorithm for Detection and Triage of Cancer-associated Incidental Pulmonary Embolism 2023. URL: <https://pubs.rsna.org/doi/10.1148/ryai.220286> (дата звернення: 20.12.2024)
6. Medical Product Outsourcing / Brusco S. // IBM’s Watson Diagnosed a Rare Condition that Left Doctors Stumped. 2016. URL: <https://www.mpo-mag.com/exclusives/ibms-watson-diagnosed-a-rare-condition-that-left-doctors-stumped/#:~:text=According%20to%20Japanese%20NHK%20News,wasn't%20responding%20to%20treatment.> (дата звернення: 20.12.2024).
7. Henrico Dolfing's case study / The \$4 Billion AI Failure of IBM Watson for Oncology // 2024. URL: <https://www.henicodolfing.com/2024/12/case-study-ibm-watson-for-oncology-failure.html> (дата звернення: 20.12.2024).

8. Lung nodules segmentation from CT with DeepHealth toolkit / Hafiza Ayesha Hoor Chaudhry, Riccardo Renzulli, Daniele Perlo, Francesca Santinelli, Stefano Tibaldi // arXiv preprint arXiv: 2208.00641 2022. URL: <https://arxiv.org/abs/2208.00641> (дата звернення: 20.12.2024).

9. HL7 International// URL: <https://www.hl7.org/implement/standards/index.cfm> (дата звернення: 24.12.2024).

10. Fast Healthcare Interoperability Resources / Wikipedia // URL: [https://en.wikipedia.org/wiki/Fast\\_Healthcare\\_Interoperability\\_Resources](https://en.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources) (дата звернення: 24.12.2024).

11. U-Net: Convolutional Networks for Biomedical Image Segmentation / Olaf Ronneberger, Philipp Fischer, Thomas Brox // arXiv preprint arXiv: 1505.04597 2015. URL: <https://arxiv.org/abs/1505.04597> (дата звернення: 24.12.2024).

12. Investigating ResNet deep features for Parkinson's disease diagnosis using hand-drawn pattern / Rahul Pandya; Vatsalkumar Vipulkumar Shah; Neel Macwan; Maithili Rajesh Vartak; Dhruv Patel // 2023. URL: <https://ieeexplore.ieee.org/document/10141842> (дата звернення: 24.12.2024).

13. nnU-Net–based Segmentation of Tumor Subcompartments in Pediatric Medulloblastoma Using Multiparametric MRI: A Multi-institutional Study / Rohan Bareja , Marwa Ismail, Douglas Martin, Ameya Nayate, Ipsa Yadav, Murad Labbad, Prateek Dullur , Sanya Garg // 2024. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11427926> (дата звернення: 24.12.2024).

14. Segmenting Medical Images: From UNet to Res-UNet and nnUNet / Lina Huang, Alina Miron, Kate Hone, Yongmin Li // arXiv preprint arXiv: 2407.04353 2024. URL: <https://arxiv.org/abs/2407.04353> (дата звернення: 24.12.2024).

15. MONAI Documentation // 2024. URL: <https://monai.io> (дата звернення: 20.12.2024).

16. I. Kyrychenko, G. Tereshchenko, D. Kozak and A. Chupryna, "Evaluation of Deep Learning Systems in Medical Diagnosis," 2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2025, pp. 1-6, doi: 10.1109/eStream66938.2025.11016830.

17. Detection of blood cells / G. Tereshchenko, G. Chetverykov. I. Konarieva // Біоніка інтелекту. – Х.: ХНУРЭ. – №1 (92), 2019. – с. 26–31. DOI: [https://doi.org/10.30837/bi.2019.1\(92\).05](https://doi.org/10.30837/bi.2019.1(92).05).

18. Aidoc. AI-powered radiology solutions for medical imaging. // URL: <https://aidoc.com> (дата звернення: 23.12.2024).

19. Selection of Artificial Neural Networks for Disease Prediction / Kyrychenko, I., Nazarov, O., Huliiev, N., Avdieiev, O. // 2023. Proceedings of the 2nd International Scientific and Practical Conference «Science in the Environment of Rapid Changes», February 6-8, 2023. Brussels, Belgium Scientific Collection «InterConf» № 141, pp.335–337. Часть ISSN: 16130073.

20. LIDC-IDRI Dataset. The Lung Image Database Consortium image collection // URL: <https://wiki.cancerimagingarchive.net> (дата звернення: 23.12.2024).

21. BraTS Dataset. The Multimodal Brain Tumor Image Segmentation Benchmark // URL: <https://www.med.upenn.edu> (дата звернення: 24.12.2024).

22. KiTS Dataset. Kidney Tumor Segmentation Challenge.// URL: <https://kits-challenge.org> (дата звернення: 24.12.2024).

23. PyTorch Documentation. A deep learning framework for scalable AI applications. // URL: <https://pytorch.org> (дата звернення: 24.12.2024).

24. DeepHealth Toolkit Documentation. Open platform for deep learning in healthcare.// URL: <https://deephealth-project.eu> (дата звернення: 24.12.2024).

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ  
КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ**

16. I. Kyrychenko, G. Tereshchenko, D. Kozak and A. Chupryna, "Evaluation of Deep Learning Systems in Medical Diagnosis," 2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2025, pp. 1-6, doi: 10.1109/eStream66938.2025.11016830.

17. Detection of blood cells / G. Tereshchenko, G. Chetverykov. I. Konarieva // Біоніка інтелекту. – Х.: ХНУРЭ. – №1 (92), 2019. – с. 26–31. DOI: [https://doi.org/10.30837/bi.2019.1\(92\).05](https://doi.org/10.30837/bi.2019.1(92).05).

19. Selection of Artificial Neural Networks for Disease Prediction / Kyrychenko, I., Nazarov, O., Huliiev, N., Avdieiev, O. // 2023. Proceedings of the 2nd International Scientific and Practical Conference «Science in the Environment of Rapid Changes», February 6-8, 2023. Brussels, Belgium Scientific Collection «InterConf» № 141, pp.335–337. Часть ISSN: 16130073.