

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ**

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Моделі сегментації для медичних зображень на
основі штучного інтелекту»

на здобуття освітнього ступеня бакалавра
зі спеціальності 122 Комп'ютерні науки
освітньо-професійної програми Штучний інтелект

*Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело*

(підпис) Ілля ФІЛІПАСЬ

Виконав: здобувач вищої освіти група ШІД-42
Ілля ФІЛІПАСЬ

Керівник: Максим ФЕСЕНКО
к.т.н., доцент

Рецензент:
науковий ступінь,
вчене звання

Ім'я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут інформаційних технологій

Кафедра Штучного інтелекту

Ступінь вищої освіти Бакалавр

Спеціальність 122 Комп'ютерні науки

Освітньо-професійна програма Штучний інтелект

ЗАТВЕРДЖУЮ

Завідувач кафедру Штучного інтелекту

_____ Ольга ЗІНЧЕНКО

«_____» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Філіпась Ілля Андрійович

1. Тема кваліфікаційної роботи: Моделі сегментації для медичних зображень на основі штучного інтелекту
керівник кваліфікаційної роботи Максим Фесенко к.т.н., доцент,
затверджені наказом Державного університету інформаційно-комунікаційних технологій від «24» лютого 2025 р. № 56
2. Строк подання кваліфікаційної роботи «23» травня 2025р.
3. Вихідні дані до кваліфікаційної роботи: науково-технічна література
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
 - Огляд методів сегментації медичних зображень і обґрунтування вибору
 - Підготовка LGG Dataset: розмітка, аугментація, нормалізація
 - Порівняння архітектур сегментації з одним енкодером
 - Налаштування та тренування моделей із єдиними гіперпараметрами
5. Перелік ілюстративного матеріалу: презентація
6. Дата видачі завдання «25» лютого 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз наукових статей, оглядів і репозиторіїв щодо сегментації медичних зображень і архітектур CNN/Transformer	25.02.2025 - 05.02.2025	Виконано
2	Підготовка та попередня обробка даних (аугментації, нормалізація інтенсивностей, розбиття на train/val/test)	06.03.2025 - 08.03.2025	Виконано
3	Реалізація єдиного пайплайна для навчання моделей із бекбоном EfficientNet-B3	09.03.2025 - 19.03.2025	Виконано
4	Налаштування та тренування моделей (U-Net, U-Net++, DeepLabV3, FPN, SegFormer) з однаковими гіперпараметрами	20.03.2025 - 29.03.2025	Виконано
5	Оцінка якості сегментації за метриками Dice, IoU, Precision, Recall та аналіз часу інференсу	30.03.2025 - 10.04.2025	Виконано
6	Порівняльний аналіз результатів, вибір оптимальної моделі й формулювання рекомендацій для клінічної практики	11.04.2025 - 25.04.2025	Виконано
7	Написання основних розділів роботи (методологія, результати, обговорення) та підготовка ілюстрацій і таблиць	26.04.2025 - 18.05.2025	Виконано
8	Остаточне редагування, перевірка оформлення згідно з вимогами університету та підготовка доповіді для захисту	19.05.2025 - 23.05.2025	Виконано

Здобувач вищої освіти

_____ (підпис)

Ілля ФІЛІПАСЬ

Керівник
кваліфікаційної роботи

_____ (підпис)

Максим ФЕСЕНКО

РЕФЕРАТ

Текстова частина бакалаврської кваліфікаційної роботи: 68 стор., 14 рис., 2 табл., 20 джерел.

Мета роботи - визначити оптимальну архітектуру для автоматизованої бінарної сегментації пухлин головного мозку на МРТ-знімках із використанням штучних нейронних мереж.

Об'єкт дослідження - процес автоматизованої сегментації патологічних зон на медичних зображеннях.

Предмет дослідження - різні підходи до побудови глибоких нейронних мереж для сегментації медичних зображень із уніфікованим механізмом витягання ознак.

Короткий зміст роботи: У вступі обґрунтовано актуальність автоматизованої сегментації головного мозку та огляд методів від класичних до глибинних мереж. Перший розділ аналізує теоретичні підходи: порогова сегментація, детектори контурів і CNN. У другому розділі описано LGG Segmentation Dataset, підготовку даних (аугментація, нормалізація, баланс) і єдиний пайплайн попередньої обробки. Третій розділ містить реалізацію кількох сучасних моделей із одним енкодером, їхнє навчання та налаштування гіперпараметрів. Проведено оцінку якості сегментації за Dice, IoU, Precision і Recall, а також порівняно швидкодію й обчислювальні витрати. У висновках визначено оптимальний баланс точності та продуктивності та рекомендовано використання моделі в клінічній практиці. Отримані результати можуть бути використані для покращення інструментів автоматизованої діагностики та моніторингу динаміки захворювання, а також слугувати основою для подальших наукових досліджень у галузі сегментації медичних зображень.

КЛЮЧОВІ СЛОВА: СЕГМЕНТАЦІЯ МЕДИЧНИХ ЗОБРАЖЕНЬ, ГЛИБОКЕ НАВЧАННЯ, БІНАРНА СЕГМЕНТАЦІЯ, АВТОМАТИЗОВАНА ДІАГНОСТИКА, МРТ, НЕЙРОННІ.

ABSTRACT

Textual part of the bachelor's qualification work: 68 pages, 14 fig., 2 tables, 20 sources.

Aim of the work: determine the optimal architecture for automated binary segmentation of brain tumors on MRI scans using artificial neural networks.

Object of research: the process of automated segmentation of pathological areas in medical images.

Subject of research: various approaches to constructing deep neural networks for medical image segmentation with a unified feature-extraction mechanism.

Summary of the work: The Introduction justifies the relevance of automated brain segmentation and reviews methods ranging from classical techniques to deep networks. Chapter 1 analyzes theoretical approaches: threshold-based segmentation, edge detectors, and convolutional neural networks (CNNs). In Chapter 2, the LGG Segmentation Dataset is described, along with data preparation (augmentation, normalization, balancing) and a single preprocessing pipeline. Chapter 3 presents the implementation of several state-of-the-art single-encoder models, their training, and hyperparameter tuning. Segmentation quality is evaluated using Dice, IoU, Precision, and Recall metrics, and throughput and computational costs are compared. The Conclusions identify the optimal trade-off between accuracy and performance and recommend the chosen model for clinical practice. The results obtained can be used to improve tools for automated diagnostics and disease-progress monitoring and can serve as a foundation for further scientific research in medical image segmentation.

KEYWORDS: MEDICAL IMAGE SEGMENTATION, DEEP LEARNING, BINARY SEGMENTATION, AUTOMATED DIAGNOSTICS, MRI, NEURAL NETWORKS.

ЗМІСТ

ВСТУП.....	11
1 ТЕОРЕТИЧНІ ОСНОВИ СЕГМЕНТАЦІЇ	13
1.1 Визначення та завдання сегментації.....	13
1.2 Класичні підходи (порогова обробка, контури)	15
1.3 Машинне навчання у сегментації	20
2 МЕТОДИ ТА РЕАЛІЗАЦІЯ	26
2.1 Опис LGG Segmentation Dataset	26
2.2 Підготовка даних і формування пайплайну обробки.....	29
2.3 Конволюційні нейронні мережі для сегментації медичних зображень	34
2.4 Вибір та порівняння енкодерів	42
2.5 Конфігурація навчання.....	43
3 ЕКСПЕРИМЕНТИ, РЕЗУЛЬТАТИ ТА АНАЛІЗ	53
3.1 Метрики оцінки якості сегментації.....	53
3.2 Результати сегментації для кожної архітектури.....	55
3.3 Статистичний аналіз та інтерпретація результатів	62
ВИСНОВКИ.....	73
ПЕРЕЛІК ПОСИЛАНЬ	75
ДОДАТОК А. ПРОГРАМНИЙ КОД АЛГОРИТМУ СЕГМЕНТАЦІЇ	77
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ	81

ВСТУП

На сьогоднішній день медична діагностика покладається на аналіз зображень, отриманих за допомогою різних методів візуалізації (МРТ, КТ, УЗД тощо). Особливо важливою в медичній діагностиці є сегментація патологічних зон на знімках головного мозку. Медична візуалізація – це основний інструмент діагностики неврологічних захворювань. Серед методів, що використовуються, особливо поширені МРТ із FLAIR-послідовностями, які дозволяють виявити вогнища ураження мозкової тканини. Однак самостійний аналіз десятків знімків є виснажливою і суб'єктивною роботою, що залежить від досвіду лікаря та займає багато часу. Цей процес не лише затримує ухвалення клінічних рішень, а й робить оцінку об'єму пухлини неточною.

Онкологічні захворювання мозку, зокрема гліоми низького та середнього ступеня агресивності (LGG), становлять серйозну проблему сучасної неврологічної практики. Швидке й точне визначення меж пухлини на МРТ-зображенні є ключовим етапом планування хірургічного втручання, радіотерапії та хіміотерапії.

Зростання обчислювальних потужностей та розвиток методів глибокого навчання відкрили нові можливості для побудови високоточних моделей сегментації, здатних працювати з неоднорідними та високо вимірними медичними даними. Саме тому розробка автоматизованих алгоритмів сегментації, що демонструють високу узгодженість із ручними розмітками експертів є вкрай важливою науковою та практичною задачею.

Сьогодні домінують методи глибокого навчання: від класичних U-Net і її модифікацій до сучасних трансформерних моделей, які автоматично виділяють ознаки й працюють із різномасштабними патернами. Проте висока обчислювальна складність цих підходів вимагає оцінки їх ефективності в кожному конкретному випадку. У цій роботі використовується відкритий «LGG Segmentation Dataset» із The Cancer Imaging Archive (TCIA), що включає МР-зображення головного мозку з FLAIR-послідовностями та відповідні ручні сегментаційні маски патологічних зон. Набір даних охоплює 110 пацієнтів із

колекції The Cancer Genome Atlas (TCGA) для пухлин нижчого ступеня, а також містить інформацію про геномні кластери та демографічні дані пацієнтів у файлі data.csv. Цей датасет вже використовувався у дослідженнях Budaet al. (2019) та Mazurowski et al. (2017).

Метою цієї бакалаврської роботи є порівняльний аналіз п'яти моделей сегментації (U-Net, U-Net++, DeepLabV3, FPN та SegFormer) з однаковим енкодером EfficientNet-B3 (без попереднього навчання) на базі LGG Segmentation Dataset. Для досягнення мети поставлено такі завдання:

1. Описати загальні принципи побудови моделей сегментації медичних зображень і механізми ключових архітектур.
2. Розробити єдиний процес попередньої обробки даних, таких як аугментації, нормалізація інтенсивностей, балансування класів.
3. Навчити та налаштувати всі чотири моделі з використанням однакових гіперпараметрів, наприклад оптимізатор, навчальна швидкість, функція втрат.
4. Оцінити якість сегментації за метриками F1, F2, IoU, Precision та Recall.
5. Визначити оптимальну з точки зору балансу точності та продуктивності модель та надати рекомендації щодо її застосування в клінічних задачах.

У цій роботі застосовуються методи аналізу наукової літератури, побудови та налаштування глибоких нейронних мереж, статистичного аналізу результатів експериментів, а також програмна реалізація на базі фреймворку PyTorch.

1 ТЕОРЕТИЧНІ ОСНОВИ СЕГМЕНТАЦІЇ

1.1 Визначення та завдання сегментації

Обробка зображень - це одна з ключових галузей штучного інтелекту, яка дозволяє автоматизувати аналіз великих обсягів візуальних даних. Сегментація зображень визначається як процес поділу зображення на області (сегменти), що мають однорідні ознаки: колір, текстуру, яскравість або форму. Кожен сегмент відображає семантично чи візуально цілісну частину сцени - від одного об'єкта до складних структур [1].

Сегментація зображень відіграє ключову роль у багатьох практичних сферах зокрема у медицині, геоінформаційних системах, автомобільних системах, автоматичному керуванні, розпізнаванні облич, аналізі відеопотоків тощо. Особливо значущою є сегментація у сфері медичної діагностики: вона лежить в основі автоматичного виявлення пухлин, обчислення їхнього об'єму, виявлення органів і судин. Висока точність сегментації дозволяє знизити похибки при плануванні операцій, оцінці ефективності лікування та довгостроковому моніторингу пацієнтів.

1.1.1 Фокус на двокласовій сегментації

У рамках цієї роботи обираємо бінарну (двокласову) сегментацію, де пікселі поділено на «фон» і «об'єкт». Такий підхід:

- Спрощує модель: зменшує кількість параметрів у вихідному шарі й дозволяє моделі фокусуватися саме на кордонах патології.
- Знижує потребу у даних: для якісного навчання бінарної моделі достатньо значно меншої кількості розмічених зображень порівняно з багатокласовою сегментацією.
- Прискорює інференс: менша кількість класів = швидша обробка кожного пікселя, що критично в реальному часі.

Проте двокласовий підхід не враховує деталізації внутрішньої будови об'єкта (наприклад, відділення некротичних ділянок від активних зон пухлини), тому в окремих застосунках можуть знадобитися багатокласові рішення.

1.1.2 Види сегментації

У межах цієї роботи ми зосереджуємося винятково на двокласовій (бінарній) сегментації, де кожний піксель належить або класу «фон» (здорові тканини), або класу «об'єкт» (пухлинна тканина). Проте, щоб зрозуміти цінність і обмеження цього підходу, доцільно розглянути й ширший контекст трьох класичних завдань сегментації:

Виділення об'єктів з фону (binarization): Найпростіший спосіб розділити зображення на дві області. Традиційні алгоритми (пороги, region growing) швидкі, але чутливі до шуму й нерівномірного освітлення.

Семантична сегментація: Кожен піксель відноситься до одного з кількох класів (пухлина, здорові тканини, судини тощо). Створює карту класів, але не розрізняє окремі екземпляри одного виду.

Інстанс-сегментація: Поєднує семантичну сегментацію з детекцією екземплярів: кожен піксель має мітку ідентифікатора конкретного об'єкта. Вимагає складної анотації й ресурсів, але дає найбільшу деталізацію.

Наш фокус перший рівень, оскільки для планування об'єму пухлини достатньо чіткого контуру «фон-об'єкт».

1.1.3 Особливості сегментації медичних зображень

Сегментація медичних зображень відрізняється від сегментації природних сцен через низку специфічних складнощів, що пов'язані з якістю та природою медичної візуальної інформації. По-перше, медичні зображення мають низьку роздільну здатність та низький контраст між різними тканинами, що значно ускладнює чітке виявлення меж між здоровими та патологічними ділянками. По-друге, зображення часто містять артефакти і нерівномірність інтенсивності (inhomogeneity), спричинені нерівномірним магнітним полем у МРТ або ефектом розсіювання у КТ, що призводить до важкопомітних спотворень у локалізації об'єктів. По-третє, цільові регіони (наприклад, пухлини) можуть бути невеликими й розрідженими, що створює проблему дисбаланс, оскільки фон домінує за площею та може «перекривати» сигнал малого об'єкта.

Різні модальності медичної візуалізації додають власні виклики.

- Комп'ютерна томографія (КТ): тут часто виникають артефакти та шум через низькі дози опромінення, що знижує SNR і ускладнює розрізнення тонких структур.

- Магнітно-резонансна томографія (МРТ): рух пацієнта та неоднорідність магнітного поля призводять до рухових артефактів і «затухання» сигналу у віддалених ділянках, що вимагає спеціальних методів корекції.

- Ультразвукова діагностика (УЗД): зображення сильно зашумлені спекл-шумом, який гранулярно «маскує» деталі тканин і знижує точність сегментації, особливо для невеликих чи нечітких утворень.

Щоб подолати ці труднощі, на стадії передобробки застосовують методи денойзингу, корекції неоднорідності освітлення та нормалізації інтенсивності. Наприклад, алгоритми усунення спекл-шуму на базі багаторівневих фільтрів вейвлет або CNN-підходів дозволяють підвищити контраст і зменшити артефакти без значної втрати дрібних деталей. Також поширені методи компенсації bias-поля в МРТ, які вирівнюють інтенсивність сигналу по всьому зрізу та покращують однорідність фону для наступної сегментації.

Цей набір складнощів і властивостей диктує вибір як традиційних, так і глибинних методів, що спеціалізуються на підвищенні стійкості сегментації в умовах низького контрасту, шумності та артефактів.

1.2 Класичні підходи (порогова обробка, контури)

Класичні методи сегментації зображень відіграли фундаментальну роль у розвитку комп'ютерного зору. Незважаючи на широке поширення глибоких нейронних мереж, вони й досі корисні для попередньої обробки, швидкого прототипування та як компоненти гібридних систем. У цьому розділі ми детально розглянемо найпоширеніші класичні методи, оцінюючи їхні принципи, різновиди, переваги й недоліки, а також досвід їх застосування в сегментації медичних зображень.

1.2.1 Порогова обробка (Thresholding)

Порогова обробка (thresholding) є одним із найпростіших і найстаріших методів сегментації зображень. Основна ідея полягає у виборі одного або декількох порогових значень інтенсивності пікселів, які розділяють зображення на класи. Найчастіше використовується бінарна порогова сегментація, де пікселі зображення класифікуються як «об'єкт» або «фон». У медичних застосунках, таких як МРТ, порогова обробка часто була першим кроком, що дозволяв швидко визначити грубі межі пухлин або інших патологічних структур.

Одним із перших відомих методів автоматичного підбору порогового значення є алгоритм Оцу[1]. Цей підхід передбачає максимізацію міжкласової дисперсії пікселів зображення, щоб досягнути оптимального порогу. Метод Оцу[1] є ефективним для зображень, які мають бімодальний гістограмний розподіл інтенсивностей, що досить часто трапляється у медичних знімках, отриманих у контрольованих умовах.

Проте порогова сегментація має низку суттєвих обмежень, особливо помітних у складних медичних задачах. По-перше, вона надзвичайно чутлива до шуму та нерівномірної освітленості. Якщо зображення має низький контраст або різні зони неоднорідної інтенсивності, використання фіксованого порогу стає неефективним. По-друге, цей підхід не враховує текстурні чи просторові особливості зображень, що дуже важливо при сегментації пухлин мозку. Наприклад, область набряку навколо пухлини може мати близьку до пухлини інтенсивність, але при цьому не є частиною самої пухлини, що призводить до помилок сегментації.

Для вирішення цієї проблеми було запропоновано адаптивні порогові методи, що змінюють поріг залежно від локальних характеристик зображення. Такі методи включають локальну порогову сегментацію за допомогою ковзного вікна, адаптивний алгоритм Sauvola (2000) [12], або використання гістограмних вирівнювань для поліпшення контрасту перед сегментацією. У контексті медичних зображень адаптивні порогові методи дозволяють частково

компенсувати неоднорідність інтенсивності, але часто залишаються занадто чутливими до розміру вікна або налаштувань алгоритму.

1.2.2 Регіональні методи (Region-based segmentation)

Регіональні методи сегментації ґрунтуються на ідеї об'єднання пікселів у ширші однорідні області на основі схожості їх характеристик - найчастіше інтенсивності або текстури. На відміну від простого порогоування, яке розглядає кожен піксель окремо, регіональні алгоритми враховують локальні просторові зв'язки, що робить їх стійкішими до шуму й нерівномірності освітлення. В Region Growing початковим кроком є вибір одного або декількох стартових пікселів (seed points), часто вручну або на основі порогових значень. Далі до регіону послідовно приєднуються сусідні пікселі, інтенсивність яких відрізняється від середньої інтенсивності регіону не більше ніж на заданий поріг δ . Цей ітеративний процес триває, поки не буде вичерпано всі відповідні пікселі.

Переваги: добре зберігає контури; мінімально вимагає обчислювальних ресурсів; дозволяє виділяти зони із плавними переходами.

Недоліки: надто чутливий до вибору початкових seed points і порогу δ ; схильний до «витікання» регіону в сусідні структури, якщо межі об'єкта нечіткі.

Щоб зменшити залежність від початкової точки, застосовують автоматичний вибір seed points на основі локальних мінімумів градієнта або кластерного аналізу гістограми [4]. Проте це додає складність і може призвести до надмірної фрагментації у складних медичних зображеннях.

Split-and-Merge алгоритм спочатку ділить зображення на великі блоки (наприклад, квадрати розміром 16×16 пікселів), у яких перевіряється однорідність (за дисперсією інтенсивності або іншою метрикою). Блоки, що містять різномірні пікселі, рекурсивно діляться на менші квадранти, доки однорідність не досягне заданого рівня. Потім суміжні блоки з подібними характеристиками об'єднуються (merge) [5].

Переваги: контролює рівень деталізації; дозволяє уникнути початкового seed selection.

Недоліки: залежить від встановлених порогів однорідності; складно узгодити розміри блоків і критерії зупинки розбиття, що може призвести як до надзвичайної дрібності сегментів, так і до надмірного об'єднання [6].

У медичних практиках *split-and-merge* використовують для попередньої розмітки великих зон тканини, наприклад при поділі мозку на зрілу та уражену зону, а потім уточнюють контури глибинними мережами.

Fuzzy region growing вводить нечіткі пороги схожості замість жорстких δ , що дозволяє плавніше приєднувати пікселі й знижує ризик «витікання» у зони з близькою інтенсивністю [7].

Marker-controlled region growing комбінує *region growing* із маркерним *watershed: seed points* визначаються як локальні мінімуми після гілчастого морфологічного фільтрування, що зменшує кількість хибних регіонів [8].

У задачі виділення пухлин мозку *region-based* методи допомагають швидко окреслити приблизні межі вогнищ, особливо коли пухлина має виразний контраст щодо навколишньої тканини (наприклад, після введення контрастного агента) [9]. Проте при поєднанні з аугментацією та *CNN-based* уточненням (гібридний підхід) ці методи покращують загальну стійкість системи до шуму й варіацій у інтенсивності знімків.

1.2.3 Детектування контурів (Edge-based methods)

Детектування контурів є одним із найінтуїтивніших підходів до сегментації зображень: воно базується на спостереженні, що реальні межі між об'єктами і фоном відповідають різким змінам яскравості. У традиційних медичних зображеннях з МРТ чи КТ ці переходи часто є єдиною надійною ознакою пухлини чи судини, коли текстурні й колірні відмінності мінімальні. Таким чином, *edge-based* методи продовжують відігравати важливу роль у попередній обробці перед застосуванням складніших алгоритмів [1].

Перші оператори градієнта, як-от Roberts (1965) і Sobel (1968), були розроблені для швидкого виявлення діагональних та орієнтовних переходів яскравості за допомогою малих згорток 2×2 і 3×3 відповідно [2][3]. Хоч обидва підходи прості в реалізації та маловитратні в обчисленнях, вони виявляються

недостатньо стійкими до шуму й нерівномірності освітлення, властивих медичним знімкам. Наприклад, результати Roberts-оператора можуть включати численні хибні контури через зміни яскравості в межах самої тканини.

Радикальною інновацією стало впровадження оператора Canny[2], який поєднав згладжування гаусовим фільтром, обчислення градієнта, подавлення не максимумів та подвійне порогоування з гістерезисом [4]. Ця послідовність операцій дозволяє точніше відокремити справжні межі від шуму: слабкі лінії приєднуються лише до надійних градієнтів, а переривчасті краї з'єднуються в суцільні контури. Завдяки цьому Canny[2] й досі вважається «золотим стандартом» для edge-detection у багатьох медичних застосуваннях.

Проте навіть Canny[2] не вирішує всіх проблем: у зонах неоднорідного контрасту або при слабкому розрізненні пухлини від навколишньої тканини межа може бути розмита й пропускатися. Щоби поліпшити роботу в таких умовах, використовують морфологічний градієнт різницю між дилатацією та ерозією зображення, що дає контур на основі просторово-морфологічних властивостей [5]. Цей підхід зменшує вплив дрібного шуму та часто використовується як доповнення до стандартних детекторів.

Попри свої недоліки, класичні підходи мають низку важливих переваг: високу швидкість роботи, прозорість і простоту реалізації, відсутність потреби у великій кількості розмічених даних. Водночас, враховуючи сучасні клінічні вимоги до точності, вони повинні поступатися місцем комплексним методам сегментації, що базуються на CNN та Transformer-архітектурах, здатних ефективно вирішувати задачу визначення меж навіть за умов низького контрасту та складної структури патологій.

Таким чином, класичні підходи до сегментації зображень залишаються важливими з теоретичного й історичного погляду, але в умовах сучасних медичних досліджень потребують поєднання з глибокими нейромережевими підходами для забезпечення клінічно значущої точності.

1.3 Машинне навчання у сегментації

Сегментація медичних зображень у контексті сучасних технологій значною мірою спирається на досягнення у сфері машинного навчання (Machine Learning, ML). У порівнянні з класичними підходами, які ґрунтуються на ручних або напівавтоматичних налаштуваннях параметрів, методи машинного навчання забезпечують автоматизацію процесу аналізу даних шляхом навчання алгоритмів на основі великої кількості попередньо підготовлених прикладів. ML-підходи спочатку з'явилися як рішення для задач класифікації та регресії, а згодом стали базою для сегментації завдяки здатності враховувати неявні ознаки зображень, що неможливо зробити вручну.

1.3.1 Класичні ML-методи

На початкових етапах автоматизованої сегментації медичних зображень домінували класичні методи машинного навчання, які дозволяли відносно швидко отримувати робочі рішення без потреби великих обчислювальних ресурсів. Найчастіше застосовували Support Vector Machines (SVM) та Random Forest (RF).

SVM створюють у багатовимірному ознаковому просторі гіперплощину, яка максимізує відстань між класами «фон» і «об'єкт» (пухлина) [1]. У медичній сегментації інтенсивність пікселя доповнюють текстурними й просторовими ознаками наприклад, гістограма градієнтів чи локальні бінарні патерни. Bauer et al. показали, що SVM із раціонально підібраним набором таких характеристик може виділити пухлину на МРТ із високою точністю. Проте відбір інформативних ознак (feature engineering) залишається ручним і трудомістким: необхідно знати, які саме характеристики найкраще розділяють тканини при різних послідовностях МРТ, що потребує досвіду як у медицині, так і в обробці зображень.

Random Forest ансамбль із сотень дерев рішень, в якому кожне дерево тренується на випадковому підмножині даних і ознак. Результати по деревах агрегаціюються голосуванням, що підвищує стійкість моделі до шуму й перетренування. Criminisi et al. використовували RF для сегментації КТ-знімків органів черевної порожнини, показавши, що автоматичний відбір релевантних ознак зменшує потребу в експертному втручанні. Однак через «побудову»

кожного дерева окремо RF не навчає моделювати довгі просторові залежності між пікселями, тому на переходах «здорова тканина - пухлина» можуть з'являтися артефакти, а контури виявляються нерівними.

Ключова перевага цих методів прозорість і простота налаштування гіперпараметрів (коефіцієнт регуляризації у SVM, глибина і кількість дерев у RF). Водночас обидва підходи страждають від обмежених можливостей автоматичного захоплення складних просторових і текстурних патернів, притаманних медичним зображенням. Це і стало одним із головних стимулів розвитку глибокого навчання: згорткові нейронні мережі можуть напряду витягувати ознаки зі зображень, не вимагаючи ручного feature engineering, та краще справлятися з шумом, низьким контрастом і варіативністю патологій [2].

1.3.2 Перехід до глибокого навчання

На початку 2010-х років розробники медичних систем зіткнулися із межами класичних алгоритмів: ручний добір ознак виявлявся трудомістким, а обмежена здатність SVM і Random Forest захоплювати просторові залежності призводила до нерівних, фрагментованих масок. У відповідь на це з'явився новий парадигм глибоке навчання, зокрема згорткові нейронні мережі (CNN), здатні автоматично витягувати релевантні ознаки без явного програмування кожного кроку [1].

Перші FCN-архітектури[14] продемонстрували: якщо замінити повні зв'язки на згортки, мережа може видавати карту того ж розміру, що й вхідне зображення, передбачаючи клас кожного пікселя. Проте справжню революцію в сегментації зробив U-Net, який вперше поєднав симетричний енкодер-декодер із пропускними зв'язками, що забезпечують пряму передачу дрібних просторових ознак із рівня стискання на відповідний рівень відновлення. Завдяки такому рішення U-Net змогла досягти видатних результатів[15] навіть на невеликих обсягах медичних даних, де традиційні CNN швидко переобучалися.

Відтоді архітектура еволюціонувала: з'явилися DeepLabV3 з атрусними згортками для багатомасштабного контексту, SegNet для ефективного кодування й відновлення, Attention U-Net із механізмом уваги, що фокусується на найважливіших зонах зображення. Кожна нова модифікація намагається вирішити

конкретні виклики медичних знімків низький контраст, високий рівень шуму, неоднорідність патологій використовуючи глибші мережі, спеціальні блоки агрегації ознак або нові функції втрат.

Однак із впровадженням глибинних мереж з'явилися свої обмеження: велика кількість параметрів вимагає потужного обладнання, а якісна розмітка зображень залишається дефіцитною. У зв'язку з цим дослідники активно шукають компроміси: transfer learning із попереднім тренуванням на великих загальних датасетах, комбіновані втрати (наприклад, Dice + boundary loss) і складні аугментаційні схеми.

Моя авторська позиція полягає в тому, що перехід до глибокого навчання не замінює класичні методи, а компліментарно доповнює їх: CNN-архітектури беруть на себе складну задачу витягнення ознак і моделювання просторових залежностей, а класичні алгоритми зберігають роль у структурі попередньої обробки, ініціалізації сегментації та генерації маркерів для подальшого уточнення.

1.3.3 Трансформерні підходи в сегментації медичних зображень

У відповідь на обмеження згорткових мереж насамперед локальність сприйняття та потребу в об'ємних даних до сегментації почали застосовувати трансформери з механізмом самоуваги (self-attention). Ця ідея полягає в тому, що кожен фрагмент зображення «уважно» зважає взаємозв'язки з усіма іншими, дозволяючи моделі захоплювати довгі залежності й глобальний контекст.

Перші спроби, такі як Vision Transformer, розбивали зображення на патчі, але втрачаючи дрібні деталі. SegFormer подолав це, поєднавши адаптивну агрегацію ознак на різних масштабах із відмовою від жорсткого patch-embedding, тож модель одночасно враховує тонкі контури і широкі зони ураження.

Проте недоліки трансформерів не зникли: quadratically зростаюча вартість самоуваги ускладнює обробку великих тривимірних знімків, а без достатнього обсягу анотованих даних вони легко переобучаються. У практиці це стимулює створення гібридних схем, де CNN-енкодер спочатку формує компактні карти

ознак, а трансформерна частина уточнює сегментацію, поєднуючи локальні патерни й глобальні структури.

Попри ресурсні виклики, трансформерні моделі відкривають нові горизонти для автоматизованої діагностики, оскільки дозволяють враховувати складну та нерівномірну архітектуру патологій головного мозку без жорсткого ручного налаштування.

1.3.4 Трансферне навчання в сегментації медичних зображень

Трансферне навчання (англ. *Transfer Learning*) є потужним інструментом у сучасному машинному навчанні, що дозволяє використовувати знання, набуті в одній задачі, для покращення результатів в іншій, пов'язаній задачі. У контексті медичної сегментації цей підхід є особливо цінним через обмежений обсяг доступних розмічених даних, складність їх анотації та високу вартість медичних досліджень.

Суть підходу

Основна ідея трансферного навчання полягає у використанні попередньо натренованих моделей або їх частин, зазвичай на великих загальних датасетах (наприклад, ImageNet), як бази для подальшого донавчання на специфічному наборі даних. У випадку сегментації, зазвичай використовується лише енкодер частина моделі (наприклад, EfficientNet, ResNet тощо), яка навчається витягувати універсальні ознаки зображень (контури, текстур, форми). Декодер, який трансформує ці ознаки у сегментаційні маски, навчається з нуля або тонко адаптується. Причини ефективності трансферного навчання

- Недостатність медичних даних: Медичні зображення важко отримати у великих обсягах, і кожна анотація вимагає участі лікаря-експерта. Трансферне навчання дозволяє зменшити потребу в великих обсягах розмічених прикладів.

- Прискорення збіжності моделі: Ініціалізація ваг попередньо навченими значеннями забезпечує більш швидку і стабільну збіжність функції втрат, оскільки модель уже має сформовану ієрархію ознак.

- Універсальність базових ознак: Енкодери, натреновані на ImageNet, вчаться виявляти загальні патерни, які є корисними навіть у медичних зображеннях, наприклад: краї, кути, текстури тканин, локальні контрасти тощо.

- Менша схильність до перенавчання: Використання готових енкодерів із замороженими або частково замороженими шарами дозволяє уникнути перенавчання, особливо якщо обсяг цільового датасету обмежений.

Трансферне навчання стало ключовим інструментом у сегментації медичних зображень, особливо в умовах обмеженості даних. Системний огляд 2023 р. засвідчив, що попереднє тренування медичних трансформерів на природних зображеннях не знижує клінічну придатність моделей, якщо додати доменно-специфічний fine-tuning[17]. Воно дозволяє досягати високої якості сегментації з меншими витратами ресурсів на навчання та знижує бар'єр входу для застосування глибоких моделей у медичній практиці. У межах цієї роботи використання попередньо натренованих енкодерів забезпечило не лише швидшу збіжність, а й стабільніші результати в задачі виявлення пухлин головного мозку. Проте вибір моделі, стратегії адаптації та джерела попереднього навчання повинен бути обґрунтований з урахуванням специфіки даних, клінічних вимог і обчислювальних ресурсів.

2 МЕТОДИ ТА РЕАЛІЗАЦІЯ

2.1 Опис LGG Segmentation Dataset

2.1.1 Походження та склад колекції

LGG Segmentation Dataset є частиною більшого проекту TCGA (The Cancer Genome Atlas), який реалізується за підтримки Національного інституту раку США. Сам датасет належить до підкатегорії TCGA-LGG збірки даних пацієнтів із низькодиференційованими гліомами (low-grade gliomas, LGG), які є пухлинами II ступеня за класифікацією WHO. Зображення розміщено в архіві The Cancer Imaging Archive (TCIA), що спеціалізується на високоякісних клінічних візуалізаційних даних із відкритим доступом.

Усього в датасеті представлено 110 пацієнтів, кожен з яких має повний набір зображень у чотирьох модальностях МРТ:

- T1-зважене зображення до контрастування (T1),
- T1 з контрастуванням (T1-Gd),
- T2-зважене (T2),
- FLAIR (Fluid-Attenuated Inversion Recovery).

Ці модальності доповнюють одна одну і використовуються для точного виявлення підтипів пухлинної тканини. Приклад одного й того ж зрізу мозку в усіх чотирьох модальностях наведено на рис. 2.1.

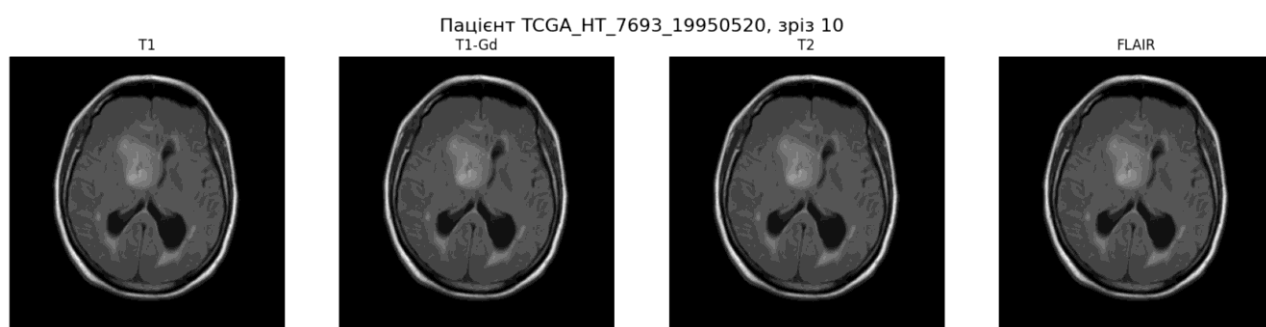


Рис. 2.1 Візуалізація одного і того ж МРТ-зрізу мозку пацієнта TCGA_HT_7693 у чотирьох різних модальностях: T1, T1-Gd, T2 та FLAIR. Кожна модальність забезпечує унікальну інформацію про структуру тканин і патологічні зміни.

Важливо, що всі зображення відносяться до преоперативного періоду, що є критичним для завдань діагностики та планування лікування.

2.1.2 Збір зображень та препроцесінг

Зображення МРТ були отримані з різних клінік, що використовували магнітно-резонансні томографи потужністю 1.5 або 3.0 Тесла. Вихідні знімки мали варіативні просторові характеристики: товщина зрізу коливалась між 1 мм і 5 мм, а розмір вокселів від 0.5×0.5 мм до 1×1 мм. Така неоднорідність потребувала стандартизації через серію препроцесінгових етапів:

Bias field correction (корекція інтенсивності): дозволяє прибрати градієнти яскравості, що виникають через фізичну неоднорідність магнітного поля в апараті. Як правило, використовується алгоритм N4ITK [13].

Skull stripping (видалення черепа): забезпечує ізоляцію мозкової тканини для виключення неінформативних структур з аналізу. Це особливо важливо при автоматичному аналізі, де череп може впливати на сегментацію.

Ресамплінг до ізотропних вокселів $1 \times 1 \times 1$ мм³: дозволяє уніфікувати розмірність даних, що є обов'язковим для глибоких нейронних мереж, чутливих до геометрії вхідного об'єму.

Жорстка реєстрація: всі модальності вирівнюються одна відносно одної, найчастіше за допомогою інструменту FLIRT (FMRIB's Linear Image Registration Tool), базуючись на T1-Gd як референсі.

Ці етапи формують єдину нормалізовану структуру даних, що уможливорює використання моделей глибокого навчання та зменшує розкид характеристик зображень між пацієнтами.

2.1.3 Анотація та мітки пухлинних підрегіонів

Однією з головних переваг LGG Segmentation Dataset є наявність ручної воксельної анотації, створеної досвідченими нейрорадіологами. Кожен об'єм позначено в трьох основних категоріях:

ЕТ (Enhancing Tumor): гіперінтенсивна область на T1-Gd, яка відображає активну, потенційно інвазивну частину пухлини, з проривом гематоенцефалічного бар'єру.

ТС (Tumor Core): включає ЕТ та некротизовані/неактивні області, як правило виявляються на T2-зображеннях як ізогіпоінтенсивні ділянки.

WT (Whole Tumor): охоплює TC разом із перифокальним набряком, добре помітним на FLAIR зображеннях.

Усі анотації представлені у форматі NIfTI (.nii.gz) з відповідними числовими мітками (1, 2, 4). Така деталізація дає змогу не лише сегментувати пухлину в цілому, а й аналізувати її структурну неоднорідність.

2.1.4 Застосування в дослідженнях

LGG Dataset посів ключове місце у дослідженнях автоматичної діагностики гліом. Основні напрями використання:

Сегментаційні моделі: різні варіації U-Net, такі як U-Net++, DeepLabV3 та SegFormer, використовуються для побудови точних піксельних масок. Показники Dice Score на Whole Tumor часто перевищують 0.85.

Трансформери: останнім часом активно застосовуються моделі типу TransUNet, які демонструють високу здатність до виявлення довгих залежностей і покращену локалізацію пухлин.

Клінічне прогнозування: результати сегментації використовуються як ознаки (features) у прогнозних моделях виживаності або класифікації за IDH/1p19q-статусом.

Фінетюнінг/пре-тренування моделей: LGG часто використовується як джерело для попереднього навчання або доанотації в мультимодальних системах діагностики.

2.1.5 Обмеження

Незважаючи на високу якість і широку використаність, датасет має низку обмежень:

Дисбаланс класів: у багатьох випадках ET займає дуже малу частку від загального об'єму, що ускладнює навчання моделей і призводить до зсуву у бік WT.

Гетерогенність апаратного забезпечення: різні типи сканерів та МР-параметри можуть викликати доменні зсуви, які потребують додаткової нормалізації.

2.2 Підготовка даних і формування пайплайну обробки

Якісна підготовка даних є критичною складовою будь-якого проєкту з глибинного навчання, оскільки саме від неї залежить стабільність навчального процесу, швидкість збіжності та узагальнювальна здатність моделі. У задачах семантичної сегментації медичних зображень це особливо важливо через високу чутливість алгоритмів до артефактів, шумів та неточностей у розмітці, які можуть призвести до хибних діагнозів або втрати важливої клінічної інформації. У цьому розділі докладно розглянемо ключові кроки побудови надійного й масштабованого конвеєра обробки, починаючи від структурування й верифікації вхідного датасету, закінчуючи застосуванням комплексних препроцесингу та аугментації для підвищення різноманітності навчальних прикладів. Зокрема, ми продемонструємо, як організувати роботу з LGG Segmentation Dataset, щоб забезпечити коректне зчитування серій МРТ-знімків і відповідних масок пухлин, оптимізувати формат і масштаб даних, а також запровадити перевірки якості на кожному етапі конвеєра. Такі методологічні підходи гарантують, що подальші етапи розробки моделей базуватимуться на достовірній та високоякісній інформації, що сприяє отриманню надійних та клінічно значущих результатів.

2.2.1 Структура та організація датасету

У LGG Segmentation Dataset кожний фрейм медичної візуалізації напряду асоціюється зі своєю бінарною розміткою пухлини, що створює фундамент для коректного навчання моделей семантичної сегментації. Така подвійна відповідність «зображення ↔ маска» є критичною не лише для точності розмітки, але й для можливості інтерпретації результатів у клінічному контексті.

Ключові переваги ієрархії «пацієнт → знімки → маски»:

- Гарантія цілісності клінічних випадків. Розбиваючи дані так, щоб усі знімки одного пацієнта залишалися разом, ми уникаємо ситуацій, коли модель «побачила» частину даних про пухлину на етапі навчання й «впізнала» їх під час тестування. Це забезпечує реалістичну оцінку узагальнювальної здатності, адже справжні клінічні рішення будуть ґрунтуватися на незнайомих пацієнтах.

- Контроль за балансом класів. У класичних медичних наборах часто спостерігається сильна нерівномірність між кількістю знімків із пухлиною й без неї. Групуючи дані за пацієнтами, легко побачити, чи є в датасеті «важкі» випадки з одиничними позитивними масками або навпаки пацієнти без пухлин, і компенсувати це при генерації навчальних батчів.

- Масштабованість і відтворюваність. Наявність чіткої структури дозволяє автоматично генерувати таблицю метаданих із мінімальними зусиллями, зберігаючи інформацію про діагноз, спосіб сканування та часові точки. Це спрощує ведення версій набору даних, порівняння результатів різних ітерацій експериментів та обмін інформацією між дослідниками.

Аналітичні аспекти організації даних:

Використання пацієнт-орієнтованих груп у train/val/test сплітах дозволяє формувати збалансовані підвибірки з урахуванням клінічних характеристик (наприклад, віку або ступеню агресивності пухлини). Це особливо важливо, коли певні підгрупи пацієнтів (за профілем віку або комбінованих послідовностей МРТ) трапляються рідко, та можуть бути випадково «зашумлені» в одній із підвибірок.

Підрахунок кількості кадрів із позитивними та негативними масками для кожного пацієнта дозволяє виявити «важкі» навчальні кейси, де моделям може знадобитися спеціалізований підхід (наприклад, oversampling або фокусовані augmentation-стратегії).

Організовуючи дані за пацієнтами, легко поєднувати зображення з додатковими клінічними метаданими наприклад, з часом від первинного сканування до появи пухлини в II фазі. Це відкриває можливості для багатозадачного навчання (наприклад, сегментація + прогноз виживаності), що значно підвищує практичну цінність дослідження.

Практичні наслідки для конвеєра обробки:

- Передавання пацієнт-ідентифікатора в датасет при кожній ітерації гарантує коректне застосування трансформацій: аугментації застосовуються паралельно і до зображення, і до маски, а при необхідності лише в межах однієї групи кадрів.

- Таблична форма з полями `patient_id`, `image_path`, `mask_path` та `diagnosis` дозволяє вбудувати додаткові фільтри: наприклад, виключити з аналізу знімки низької контрастності або провести груповий аналіз якості сегментації по кожному пацієнту.

- Розподіл за пацієнтом підвищує ефективність логуювання та моніторингу експериментів: у кожному логу можна зберігати чужорідний ключ `patient_id` що спрощує відстеження помилок сегментації в конкретних клінічних випадках.

Таким чином, ієрархічна організація LGG Segmentation Dataset не просто спрощує технічну сторону побудови пайплайну, а створює міцну основу для достовірності, відтворюваності та практичної значущості результатів семантичної сегментації.

2.2.2 Оцінка та баланс класів

Першим кроком аналізу було визначення співвідношення зображень з наявною пухлиною та без неї. Після побудови табличної форми даних і позначення кожного випадку як «Positive» (наявність пухлини) або «Negative» (відсутність пухлини), було виконано підрахунок і візуалізацію кількості кадрів у кожній категорії. Результати демонструють приблизно рівномірний розподіл: близько 52 % зображень містять у собі сегментовану область пухлини, тоді як 48 % не містять. Така майже рівновага класів є надзвичайно важливою для навчання моделей семантичної сегментації, адже значне переважання одного з класів могло б призвести до упередженої роботи мережі й недостатньої чутливості до пухлинних областей.

Щоб зберегти цей баланс у кожному з піднаборів (навчальному, валідаційному та тестовому), при розділенні даних застосовано стратифіковану стратегію. Кожен пацієнт і кожне зображення віднесено до групи залежно від наявності пухлини, і пропорції «Positive/Negative» були відтворені у тестовому та валідаційному наборах з точністю до одного відсотка. Це гарантує, що жоден із піднаборів не виявиться з надто низькою або занадто високою часткою пухлинних знімків, а отже, навчання та оцінка моделі відбуваються на однорідних та порівнянних даних.

Додатково було перевірено розподіл по пацієнтах: кожен пацієнт виступає власною одиницею стратифікації, що виключає «протікання» інформації між піднабором (наприклад, коли кадри одного пацієнта потрапляють і в навчальний, і в валідаційний набори). Завдяки такому підходу мережа оцінюється на цілком нових прикладах, що належать іншим пацієнтам, що підвищує реалістичність і надійність отриманих результатів.

Отже, комбіноване використання детального аналізу розподілу класів та стратифікованого розбиття забезпечує рівномірну представленість обох категорій у всіх етапах роботи, мінімізуючи ризик упередженості моделі й підвищуючи точність сегментації найскладніших пухлинних областей.

2.2.3 Препроцесинг зображень і масок

Перший крок приведення всіх МР-зрізів до єдиного розміру 128×128 пікселів із наступною нормалізацією за середнім (0.485, 0.456, 0.406) і стандартним відхиленням (0.229, 0.224, 0.225) каналів, прийнятими в ImageNet. Це вирівнює розподіл інтенсивностей та прискорює збіжність оптимізатора, уникаючи “зависання” навчання через надто яскраві або темні кадри [16].

Щоб моделі не «запам'ятовували» конкретні положення пухлин та вміли розпізнавати їхню форму в різних варіаціях, до зображень і масок застосовано низку стохастичних геометричних і фотометричних перетворень:

- Еластичні деформації ($p=0.5$) і ShiftScaleRotate ($p=0.25$) імітація неточностей у положенні пацієнта під час сканування, зміни масштабу та зсувів, що виникають через рух та різні протоколи збору даних.
- Горизонтальні/вертикальні віддзеркалення ($p=0.5$) та округлі повороти на 90° ($p=0.5$) гарантують, що модель не залежатиме від сторони або орієнтації зрізу та розпізнаватиме пухлину незалежно від повороту голови пацієнта.
- Зміни контрасту (RandomBrightnessContrast, $p=0.5$) і гама (RandomGamma, $p=0.25$) захищають від відмінностей у налаштуваннях МРТ-сканера, забезпечуючи стійкість до варіацій інтенсивності сигналу.
- Легке розмивання (Blur, $p=0.01$) моделює шум та невеликі артефакти зображень.

Вибір інтерполяції NEAREST для масок забезпечує збереження чіткої бінарної кордони пухлин при трансформаціях. Такий конвеєр попередньої обробки підвищує здатність моделей глибокого навчання узагальнювати, покращує якість сегментації та зменшує ризик перенавчання. Випадкові приклади аугментованих МРТ-знімків, згенерованих у тренувальному пайплайні. Видно різноманіття позицій, орієнтацій і контрастності зображень. представлено на рис. 2.2.

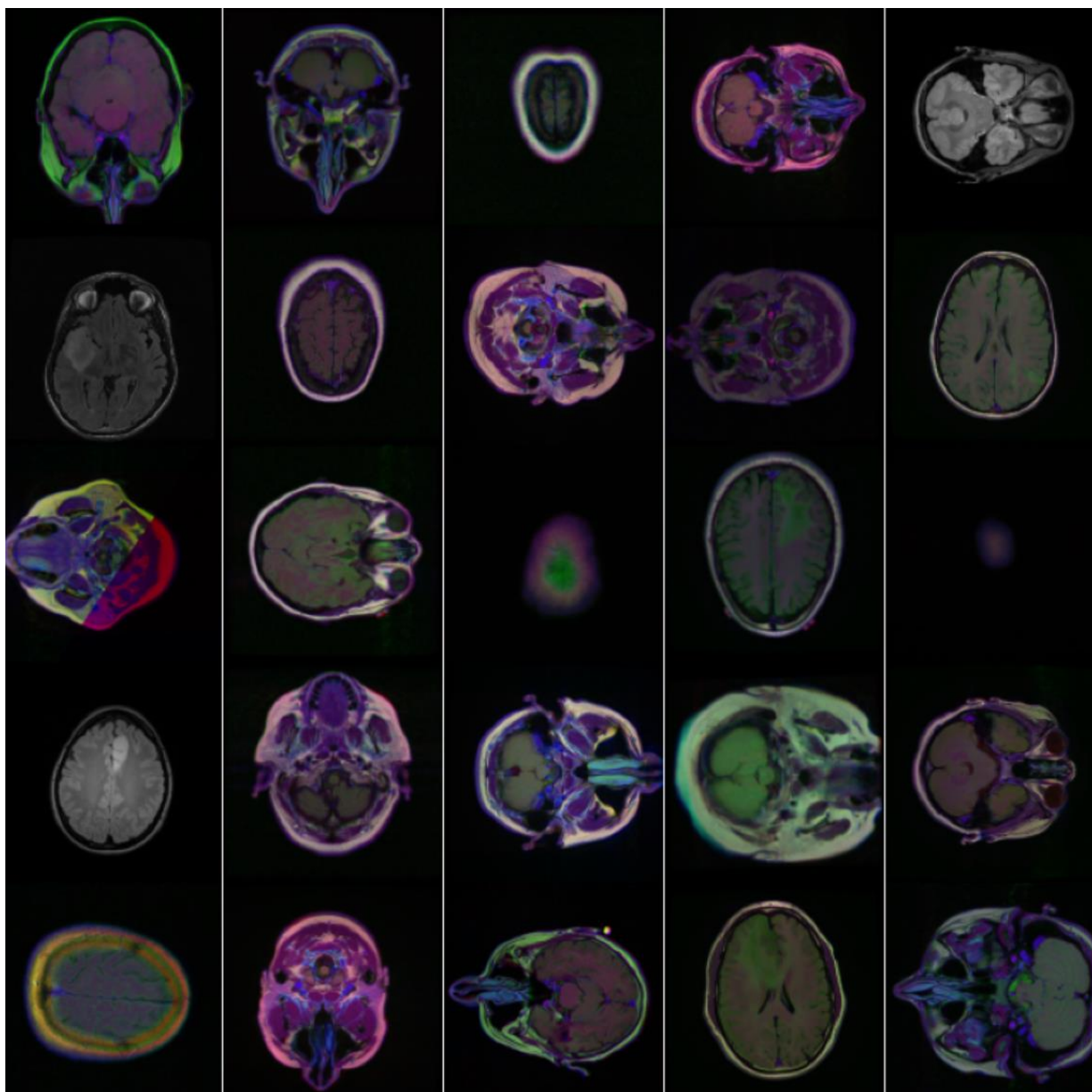


Рис. 2.2 Випадкові приклади аугментованих МРТ-знімків, згенерованих у тренувальному пайплайні. Видно різноманіття позицій, орієнтацій і контрастності зображень.

2.2.4 Розподіл на навчальну, валідаційну та тестову вибірки

З метою коректної оцінки узагальнювальної здатності моделі застосовується стратифіковане розбиття:

- Навчальна вибірка ($\approx 70 \%$)
- Валідаційна вибірка ($\approx 15 \%$)
- Тестова вибірка ($\approx 15 \%$)

Розподіл здійснюється на рівні пацієнтів, щоб жоден хворий не потрапив одночасно до тренування й оцінки. Такий підхід запобігає «витоку» інформації і дозволяє отримати більш реалістичну оцінку якості моделі на нових пацієнтах.

2.2.5 Формування батчів та оптимізація продуктивності

Для прискорення тренування використовуються:

- Batch size у межах 16-32 зображень, що є компромісом між конвергенцією та обсягом пам'яті GPU.
- Parallel DataLoader (кілька робітників) для одночасного читання та препроцесингу зображень.
- Shuffle перемішування даних у тренувальній вибірці на кожній епосі для зменшення кореляції між першими та останніми прикладами в батчі.

Такі налаштування забезпечують безперервну подачу даних у GPU та мінімізують простої під час тренування.

2.3 Конволюційні нейронні мережі для сегментації медичних зображень

У прагненні знайти баланс між точністю, швидкістю та ресурсною ефективністю в задачі бінарної сегментації пухлин головного мозку, ми обрали п'ять архітектур, які репрезентують основні педагогічні парадигми сучасного комп'ютерного зору.

2.3.1 U-Net

U-Net є однією з найпопулярніших моделей для задач семантичної сегментації зображень, зокрема в медичній сфері. Базова архітектура сегментації U-Net із симетричним енкодер-декодером та скіп-зв'язками забезпечує високу якість масок навіть на невеликих датасетах [3]. Вона має характерну "U"-подібну

форму (рис. 2.3), яка забезпечує симетричну структуру мережі: шлях енкодера (звуження) та шлях декодера (розширення). Енкодер виконує послідовне зменшення просторових розмірів зображення та витяг ознак за допомогою згорткових шарів і max pooling. У свою чергу, декодер поступово відновлює просторову роздільність, застосовуючи апсемплінг і згортки для уточнення результату.

Важливою особливістю архітектури є наявність скіп-з'єднань (skip connections), які напряду передають проміжні ознаки з відповідних рівнів енкодера до декодера. Це дозволяє поєднувати глибокий контекст із просторовою точністю, що значно покращує якість сегментації. На виході мережа формує карту ймовірностей, яка порівнюється з розміткою зображення (ground truth) для обчислення функції втрат. Завдяки такій побудові U-Net демонструє високу ефективність навіть за умов обмеженого обсягу навчальних даних, що робить її особливо придатною для практичного застосування в умовах, де анотовані дані є рідкісними.

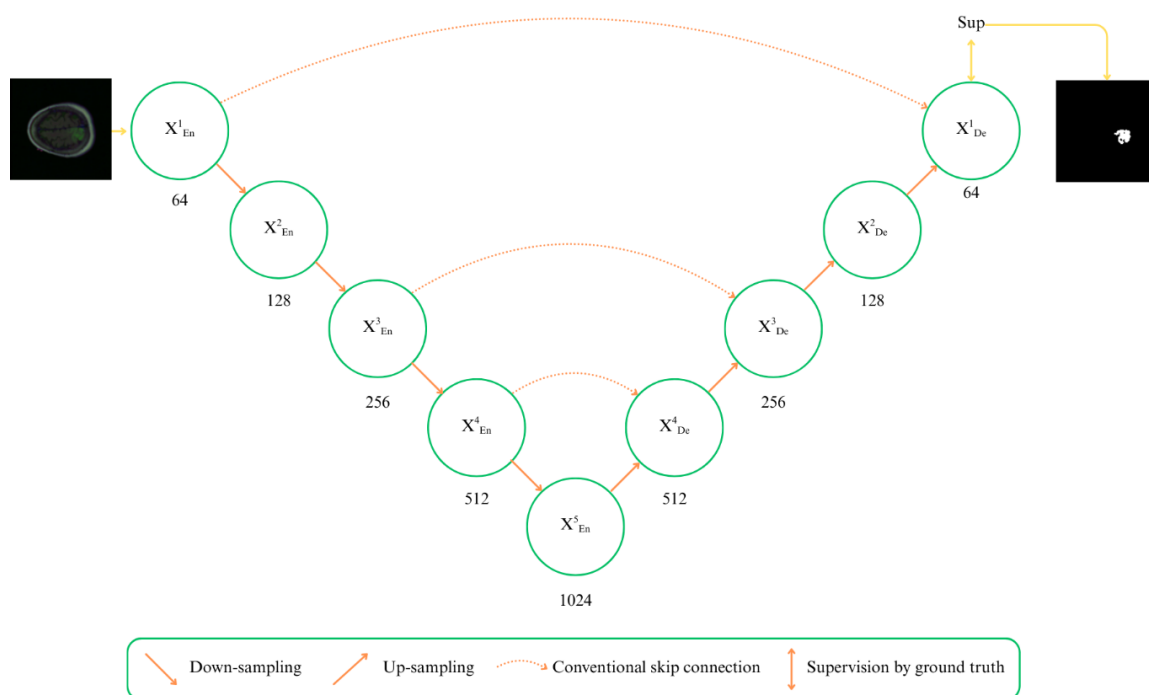


Рис. 2.3 Архітектура U-Net

2.3.2 U-Net++

Архітектура U-Net++ є покращеною версією базової моделі U-Net, розробленою для підвищення якості сегментації за рахунок кращої агрегації ознак

(features). U-Net++ з додатковими внутрішніми скіп-зв'язками дозволяє краще узгоджувати ознаки на різних рівнях просторової деталізації [4]. Основною відмінністю є наявність додаткових проміжних вузлів і розгалужених скіп-з'єднань (див. рис. 2.4), які з'єднують не лише відповідні рівні енкодера та декодера, як у звичайному U-Net, а також проміжні вузли між різними рівнями. Це дозволяє гнучко передавати інформацію між шарами, поліпшуючи навчання глибших ознак та зменшуючи втрату просторової інформації.

Модель генерує карти сегментації з кількох рівнів і об'єднує їх перед обчисленням функції втрат, що сприяє кращому поєднанню локальних і глобальних ознак. Попри вищу точність, U-Net++ із приблизно 53 млн параметрів потребує більше пам'яті та часу на тренування, ніж класичний U-Net.

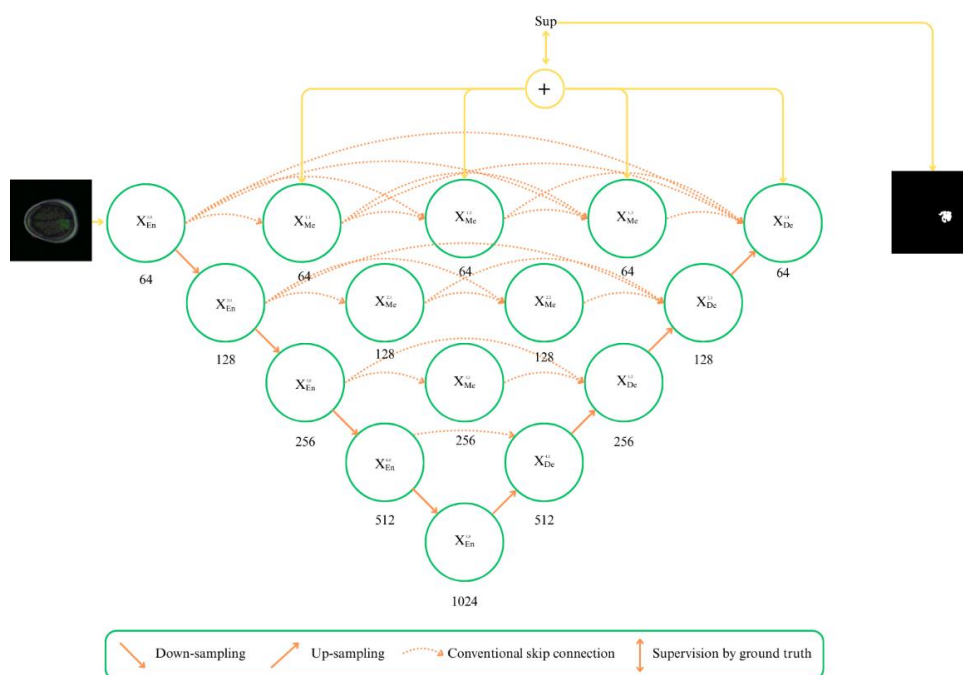


Рис. 2.4 Архітектура U-Net++

2.3.3 FPN

Архітектура Feature Pyramid Network (FPN) була запропонована як компромісне рішення між точністю сегментації та швидкістю виконання, особливо у задачах, де важлива низька затримка. Вона використовує багаторівневі ознаки для одночасного захоплення глобального та локального контексту [5]. Як видно на рис. 2.5, FPN використовує згортковий енкодер (backbone), який генерує

карти ознак на різних рівнях масштабування. Це дозволяє моделі враховувати як локальні, так і глобальні особливості зображення.

На етапі декодування застосовується ієрархічне злиття ознак через прості операції апсемплінг та згортки 1×1 , що представлені в нижньому блоці рисунка. Таке злиття формує піраміду ознак, яку можна ефективно використовувати для багатомасштабних передбачень (Scale 1-3). Завдяки малій обчислювальній складності, FPN добре підходить для використання в мобільних додатках та в реальному часі, де критичними є обмеження за ресурсами та швидкістю обробки.

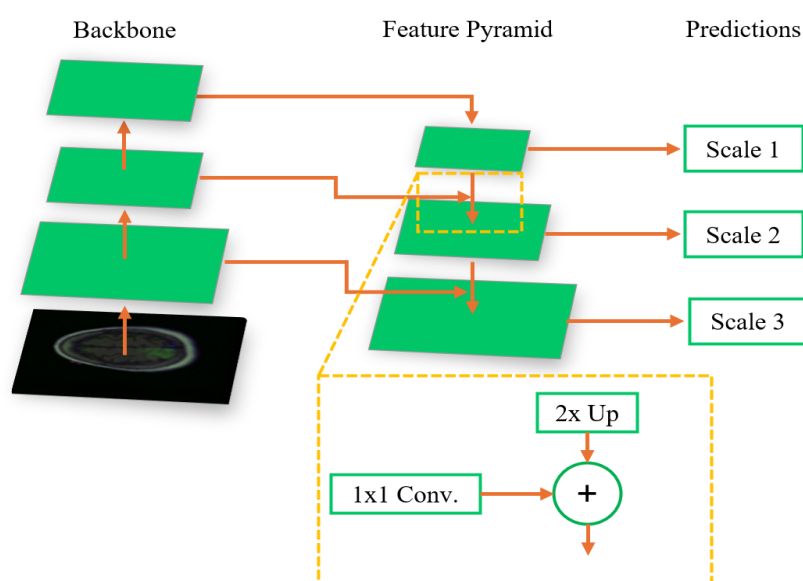


Рис. 2.5 Архітектура FPN

2.3.4 DeepLabV3

DeepLabV3 вирізняється серед сучасних архітектур семантичної сегментації завдяки використанню модуля Atrous Spatial Pyramid Pooling (ASPP), дозволяє захоплювати багатомасштабний контекст[6] для багатомасштабного аналізу пухлинних зон [6]. Як показано на рис. 2.6, ASPP складається з кількох паралельних згорткових шарів: 1×1 згортка, кілька 3×3 згорток з різними степенями дилатації (rate 6, 12, 18) та глобальний пулінг. Ці шари обробляють ознаки з різною щільністю, що дає змогу мережі «бачити» як дрібні, так і великі об'єкти без втрати роздільної здатності.

Після паралельної обробки, ознаки об'єднуються й пропускаються через декілька 1×1 згорток для об'єднання інформації та формування карти сегментації. Такий підхід дозволяє DeepLabV3 ефективно розпізнавати об'єкти зі складними, розмитими або неоднорідними межами, зокрема пухлинні структури. Водночас архітектура потребує більше обчислювальних ресурсів, що варто враховувати при її застосуванні у системах з обмеженими можливостями.

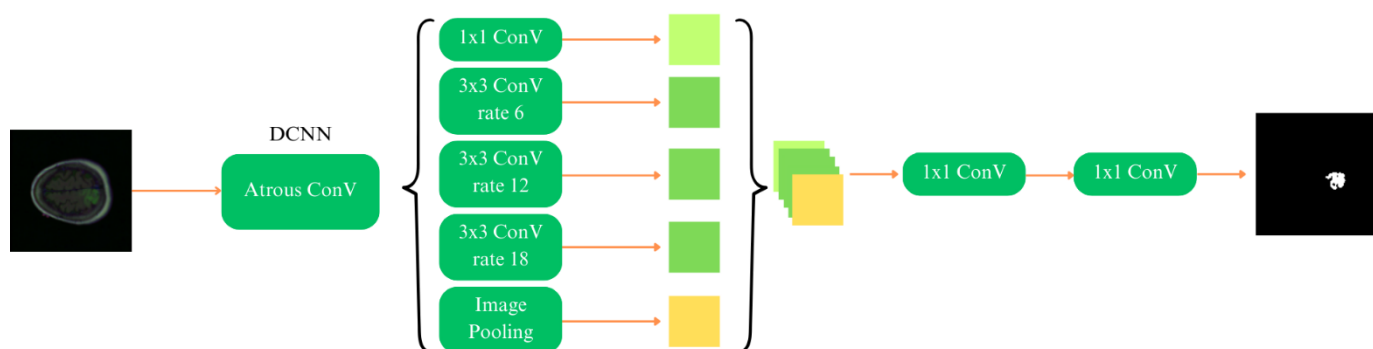


Рис. 2.6 Архітектура DeepLabV3

2.3.5 SegFormer

SegFormer представляє сучасний підхід до семантичної сегментації на основі трансформерів, глобальної самоуваги у поєднанні з легким багаторівневим декодером для деталізованої сегментації [7]. Як показано на рис. 2.7, енкодер складається з послідовних трансформерних блоків, які опрацьовують вхідне зображення через багаторівневу агрегацію ознак. На відміну від класичних ViT, де зображення розбивається на фіксовані патчі, SegFormer використовує перекривні патч-ембедінги, що дозволяє зберегти локальну структуру зображення.

Кожен рівень енкодера на основі MLP об'єднує інформацію з усіх рівнів без складної ієрархії, що забезпечує швидкість і компактність архітектури.

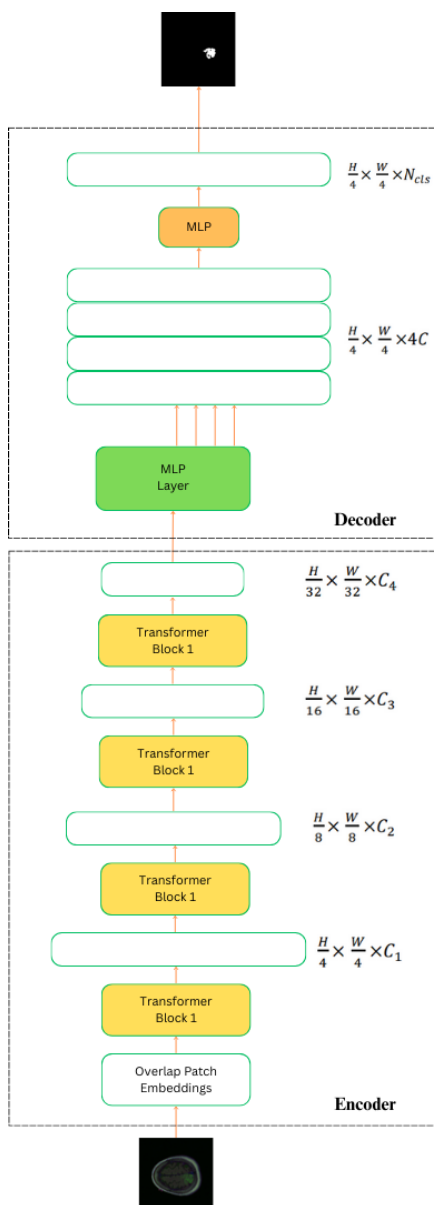


Рис. 2.7 Архітектура SegFormer

2.3.6 Порівняльна характеристика архітектур

Таким чином, кожна із п'яти архітектур пропонує свій унікальний набір сильних сторін і компромісів. U-Net і її модифікації залишаються надійною базою для швидкого прототипування, FPN забезпечує легку масштабованість, DeepLabV3 додає багатомасштабний контекст, а SegFormer доводить ефективність глобальних механізмів уваги. Порівняльний аналіз цих моделей на єдиному енкодері дозволить виявити, який підхід найкраще задовольняє клінічні вимоги до точності, швидкості та стабільності сегментації пухлин головного мозку.

2.3.7 Порівняння параметрів та розміру моделей

У цьому підрозділі проведено детальний аналіз п'яти архітектур сегментації, реалізованих у проєкті: U-Net, U-Net++, Feature Pyramid Network (FPN), DeepLabV3 та SegFormer. Порівняння охоплює кількість параметрів, розмір моделей у форматі float32, а також орієнтовне споживання відеопам'яті (VRAM) під час інференсу на зображеннях розміром 512×512 пікселів (табл. 2.1).

Таблиця 2.1

Розміри моделей

Модель	Параметри (млн)	Розмір (MB)	VRAM ($\approx$ GB)
U-Net	41.2	164.8	~ 1.0
U-Net++	52.7	210.8	~ 1.2
FPN	38.9	155.6	~ 0.9
DeepLabV3	40.7	162.8	~ 1.0
SegFormer	46.5	186.0	~ 1.1

U-Net є базовою архітектурою з помірною кількістю параметрів (~ 41 млн), що забезпечує ефективну сегментацію при відносно низькому споживанні ресурсів. Її симетрична структура з "U"-подібною формою дозволяє ефективно обробляти зображення, зберігаючи просторову інформацію через скіп-з'єднання між енкодером та декодером.

U-Net++ розширює базову архітектуру U-Net, додаючи додаткові скіп-з'єднання та декодерні блоки, що підвищує кількість параметрів до ~ 53 млн. Це дозволяє моделі краще агрегувати ознаки та покращує точність сегментації, особливо при обробці складних структур. Однак, це також призводить до збільшення споживання пам'яті та часу на тренування.

FPN має найменшу кількість параметрів серед розглянутих моделей (~ 39 млн), зберігаючи при цьому здатність обробляти багаторівневі ознаки. Її ефективна структура дозволяє зменшити обчислювальні витрати, що робить його привабливим вибором для систем з обмеженими ресурсами.

DeepLabV3 поєднує ASPP-блок для захоплення багаторівневого контексту з легким декодером, маючи близько 41 млн параметрів. Це дозволяє моделі

ефективно обробляти об'єкти різного масштабу та складності, забезпечуючи високу точність сегментації.

SegFormer використовує трансформерний декодер, що підвищує кількість параметрів до ~46 млн. Це забезпечує моделі здатність захоплювати глобальний контекст, покращуючи якість сегментації. Однак, це також призводить до збільшення споживання відеопам'яті порівняно з традиційними CNN-моделями.

У процесі вибору оптимальної архітектури для задачі сегментації медичних зображень важливо враховувати баланс між точністю моделі та її обчислювальними вимогами. Моделі U-Net та FPN демонструють ефективність у системах з обмеженими ресурсами, забезпечуючи задовільну точність при помірному споживанні пам'яті та обчислювальних потужностей. Це робить їх придатними для застосування в умовах, де ресурси обмежені, наприклад, на мобільних пристроях або в реальному часі.

З іншого боку, архітектури U-Net++ та SegFormer, завдяки своїй складнішій структурі, досягають вищої точності сегментації. Проте це супроводжується збільшенням кількості параметрів та, відповідно, підвищеними вимогами до обчислювальних ресурсів. Такі моделі доцільно використовувати в середовищах з достатніми ресурсами, де пріоритетом є максимальна точність, наприклад, у високоточних медичних діагностичних системах.

DeerLabV3 виступає компромісним варіантом, поєднуючи відносно помірну складність з високою якістю сегментації. Його здатність ефективно обробляти об'єкти різного масштабу робить його універсальним вибором для широкого спектру задач.

Отже, вибір конкретної моделі повинен базуватися на специфічних вимогах завдання, доступних обчислювальних ресурсах та необхідному рівні точності. Ретельний аналіз цих факторів дозволить обрати оптимальну архітектуру, яка забезпечить ефективну та точну сегментацію в конкретному контексті застосування.

2.4 Вибір та порівняння енкодерів

У задачах семантичної сегментації медичних зображень, особливо при побудові моделей на основі архітектур типу U-Net, FPN чи DeepLabV3, критичну роль відіграє енкодер. Саме ця частина мережі відповідає за витягування ознак як базових (контури, текстури), так і високорівневих (структурні патерни, глобальний контекст), на яких далі ґрунтується рішення про клас кожного пікселя. Від якості енкодера залежить, чи зможе модель «побачити» малопомітні межі пухлини або відрізнити її від фону з низьким контрастом. Тому вибір енкодера повинен бути не випадковим, а стратегічно обґрунтованим.

У межах даного дослідження було прийнято рішення застосувати єдиний енкодер для всіх архітектур, щоб забезпечити чесне порівняння моделей. Такий підхід дозволяє ізолювати вплив структури декодера та механізмів агрегації ознак від здатності витягувати інформацію на початкових етапах обробки. Після розгляду кількох популярних варіантів (ResNet-34, ResNet-50, DenseNet-121, MobileNet, EfficientNet-B0/B4). ResNet-архітектури слугували основою багатьох енкодерів завдяки полегшеній оптимізації у глибоких мережах але було обрано EfficientNet-B3 через оптимальне співвідношення точності й кількості параметрів [8].

Цей вибір продиктований як практичними, так і методологічними міркуваннями. По-перше, EfficientNet-B3 забезпечує відмінне співвідношення між точністю і кількістю параметрів завдяки принципу compound scaling, за яким одночасно масштабуються глибина, ширина та роздільна здатність моделі. Це дає змогу досягати високої точності навіть на складних задачах при порівняно невеликому розмірі моделі. По-друге, архітектура побудована на мобільних згорткових блоках (MBConv), які демонструють хорошу здатність до узагальнення, навіть якщо кількість навчальних прикладів обмежена, як це часто буває з медичними зображеннями.

Крім того, попередньо навчені на ImageNet ваги для EfficientNet-B3 є широко доступними й стабільними у використанні. Це дозволяє не лише пришвидшити процес збіжності, але й підвищити якість сегментації за рахунок

перенесення вже сформованих у моделі базових ознак. Практика показує, що енкодери, попередньо навчені на великій кількості зображень, навіть не медичних, ефективно працюють у домені медичної візуалізації завдяки здатності розпізнавати текстури, контури та форми, що є універсальними ознаками.

З огляду на вищезазначене, вибір EfficientNet-B3 як спільного енкодера для всіх розглянутих архітектур є обґрунтованим компромісом між потужністю, ефективністю та сумісністю з наявним пайплайном. Це рішення створює рівні умови для порівняльного аналізу архітектур і дозволяє зосередитися на оцінці реального впливу різних стратегій декодування та агрегації ознак на якість сегментації пухлин головного мозку.

2.5 Конфігурація навчання

Успішне навчання моделей глибокого навчання для задач семантичної сегментації медичних зображень значною мірою залежить від ретельної конфігурації процесу тренування. Цей процес включає вибір відповідної функції втрат, оптимізатора, стратегії зміни швидкості навчання, налаштування гіперпараметрів, а також впровадження методів запобігання перенавчанню, таких як рання зупинка та збереження найкращих версій моделі.

У даному розділі детально розглядаються ключові компоненти конфігурації навчання, які були застосовані при розробці моделей сегментації на основі датасету LGG Segmentation Dataset. Особлива увага приділяється вибору функції втрат, що впливає на здатність моделі ефективно розрізняти класи, оптимізатора, який визначає швидкість та напрямок оновлення ваг, а також стратегіям адаптації швидкості навчання, що сприяють досягненню оптимального мінімуму функції втрат.

Крім того, обговорюються гіперпараметри, такі як розмір пакету даних, кількість епох, початкова швидкість навчання та інші, які були налаштовані з урахуванням специфіки задачі та характеристик даних. Також розглядаються методи, спрямовані на підвищення стабільності та ефективності навчання, включаючи механізми ранньої зупинки, що запобігають перенавчанню, та

збереження моделей, що дозволяє зберегти найкращі результати навчання для подальшого використання.

2.5.1 Вибір loss-функції

У задачах семантичної сегментації правильний вибір функції втрат (loss-функції) відіграє ключову роль для коректного навчання моделей. Вона визначає, як обчислюється різниця між передбаченими сегментаціями та істинними мітками, і, відповідно, впливає на градієнти під час зворотного поширення помилки. У цій роботі були розглянуті декілька популярних функцій втрат, але остаточний вибір було зупинено на Cross-Entropy Loss з огляду на її стабільність, простоту інтерпретації та здатність добре працювати на завданнях двокласової класифікації пікселів.

Порівняння альтернативних функцій втрат:

- Dice Loss: ця функція орієнтована на перекриття між передбаченими і справжніми масками, що особливо корисно при сильному дисбалансі класів. Проте вона має нестійку поведінку на ранніх етапах навчання, особливо коли перетину ще практично немає. Крім того, Dice Loss не є функцією ймовірності і може мати проблеми з градієнтами при дуже малих значеннях.

- Focal Loss: модифікація Cross-Entropy, яка додає ваги до важких прикладів, щоб приділяти більше уваги неправильно класифікованим пікселям. Цей підхід часто корисний у задачах з рідкісними класами, але в нашому випадку двокласового поділу пухлина/фон така складність не обґрунтована.

- BCEWithLogitsLoss (бінарна крос-ентропія з логітами): функція, що також часто використовується в задачах бінарної сегментації. Вона подібна до Cross-Entropy, але безпосередньо працює з логітами та поєднує Sigmoid-функцію активації всередині. Її переваги доречні в задачах без softmax-нормалізації, що не є нашим випадком.

Обґрунтування вибору Cross-Entropy Loss:

Cross-Entropy Loss є стандартом де-факто у багатьох задачах класифікації, включно з піксельною сегментацією. Вона вимірює відстань між справжнім розподілом класів і передбаченим, що математично формулюється як (2.1):

$$\mathcal{L}_{CE} = \sum_{i=1}^N \sum_{c=1}^C y_{i,c} * \log(\hat{y}_{i,c}) \quad (2.1)$$

де:

N- кількість пікселів,

C - кількість класів (у нашому випадку 2),

$y_{i,c}$ - істинна мітка (one-hot encoding),

$\hat{y}_{i,c}$ - передбачене значення softmax для класу ccc.

Функція має добре визначені градієнти, стабільну поведінку і добре підходить для випадків, коли мітки є точними, а сегментація виконується на рівні окремих пікселів. Оскільки наша задача - це двокласова сегментація пухлини проти фону, Cross-Entropy є логічним і обґрунтованим вибором.

2.5.2 Вибір оптимізатора

Вибір оптимізатора є важливим етапом у побудові ефективного процесу навчання глибоких нейронних мереж. У задачах сегментації медичних зображень, де якість передбачення може напряму впливати на клінічну цінність моделі, стабільність і швидка збіжність оптимізатора набувають критичного значення. Для навчання обрано AdamW, який на сьогодні є одним із найефективніших оптимізаторів для сегментаційних задач у комп'ютерному зорі, із розділеною регуляризацією weight decay що покращує узагальнення моделей [9]..

Порівняння популярних оптимізаторів:

- SGD (Stochastic Gradient Descent) - класичний варіант стохастичного градієнтного спуску, який демонструє хорошу узагальнювальну здатність, особливо при великих наборах даних. Однак у випадку складних моделей з високою кількістю параметрів та складним ландшафтом функції втрат, він часто збігається повільніше, особливо без правильної настройки learning rate, momentum і decay.

- Adam - адаптивний оптимізатор, що поєднує переваги RMSprop та Momentum, автоматично коригуючи величини градієнтів для кожного параметра. Adam здатен забезпечити швидке зменшення втрат на початкових етапах навчання, однак його застосування може призводити до переобучення або нестійкої поведінки, особливо в умовах надмірної регуляризації.

- AdamW - модифікація Adam з *decoupled weight decay*, яка чітко розділяє регуляризацію та оновлення градієнтів. Це усуває проблему некоректного впливу weight decay на моменти градієнта, що було недоліком класичного Adam. Таким чином, AdamW забезпечує кращу узагальнюючу здатність, швидшу збіжність і стабільнішу оптимізацію у складних задачах.

Формалізація оновлення параметрів в AdamW:

AdamW оновлює ваги за наступною формулою (2.2):

$$\Theta_t = \Theta_{t-1} - \eta * \left(\frac{m_t}{\sqrt{v_t + \epsilon}} + \lambda * \Theta_{t-1} \right) \quad (2.2)$$

де:

Θ_t - параметри моделі на ітерації t ,

η - швидкість навчання,

m_t, v_t - перші та другі моменти градієнтів (адаптивна частина),

λ - коефіцієнт weight decay (розділена регуляризація),

ϵ - стабілізаційна константа для уникнення ділення на нуль.

На відміну від класичного Adam, регуляризація через $\lambda \cdot \Theta$ не впливає на адаптивні моменти, що дозволяє ефективніше контролювати переобучення.

2.5.3 Scheduler та стратегії зміни learning rate

Під час навчання глибоких нейронних мереж вибір правильної стратегії зміни швидкості навчання (*learning rate*) відіграє ключову роль у досягненні ефективної збіжності та уникненні локальних мінімумів. Занадто високий *learning rate* може призвести до нестабільного навчання, тоді як занадто низький до повільної збіжності або застрягання в неглибоких мінімальних точках. Для плавного зниження швидкості навчання застосовано Cosine Annealing Learning Rate Scheduler, який показав високу ефективність у задачах сегментації [10].

Загальні стратегії регулювання learning rate

Існує кілька популярних підходів до динамічного регулювання швидкості навчання під час тренування:

- StepLR: зменшує *learning rate* на фіксовану кількість через певну кількість епох. Має жорсткі пороги, що можуть спричинити нестабільність при зміні режиму.
- ReduceLROnPlateau: реагує на плато в метриці валідації, знижуючи *learning rate* після кількох епох без покращення. Гнучкий, але вимагає ретельного налаштування *patience* та *threshold*.
- ExponentialLR: поступово зменшує *learning rate* за експоненціальним законом. Добре підходить для задач зі стабільною динамікою втрат.
- CosineAnnealingLR: зменшує *learning rate* за косинусоподібною функцією від початкового значення до мінімального, плавно знижуючи агресивність оновлень до завершення навчання.

CosineAnnealingLR: теоретичне обґрунтування

У рамках даного проєкту було обрано CosineAnnealingLR у зв'язці з оптимізатором AdamW. Такий підхід забезпечує плавне зменшення швидкості навчання згідно з функцією(2.3):

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right)) \quad (2.3)$$

де:

η_t - поточний *learning rate*,

η_{max} - початковий *learning rate*,

η_{min} - мінімальний *learning rate*,

T_{cur} - кількість минулих епох,

T_{max} - загальна кількість епох.

Цей метод дозволяє на початку інтенсивно оновлювати ваги, а ближче до завершення тренування стабілізувати навчання. Це особливо корисно у задачах медичної сегментації, де перенавчання на незначні деталі може негативно вплинути на генералізовану здатність моделі.

2.5.4 Гіперпараметри та протокол тренування

У задачах глибокого навчання правильний вибір гіперпараметрів є ключовим для досягнення стабільного й ефективного навчання моделей. У

даному проєкті конфігурація гіперпараметрів була ретельно підібрана, виходячи зі специфіки медичних зображень, характеристик LGG Segmentation Dataset і потреб рівного порівняння архітектур.

Основні гіперпараметри

Кількість епох (epochs): 30

Попередні експерименти показали, що при використанні попередньо натренованих енкодерів (на ImageNet) більшість моделей досягають збіжності вже після 20-25 епох. Аналіз кривих функції втрат і метрик показав насичення якості ближче до 30-ї епохи. Тому було обрано саме це значення як компроміс між достатньою тривалістю навчання та уникненням перенавчання. Збільшення кількості епох не давало суттєвого приросту метрик, а отже було нераціональним з точки зору ресурсів.

Початкове значення learning rate: $1e-3$

Це популярне стартове значення для багатьох моделей, зокрема при використанні оптимізатора AdamW. Воно забезпечує швидке зменшення функції втрат на ранніх етапах навчання. Вищі значення могли призвести до нестабільної динаміки градієнтів, особливо у сегментаційних моделях із високою чутливістю до структури зображення.

Мінімальний learning rate: $1e-5$

Встановлений як нижня межа в циклі CosineAnnealing, щоб уникнути «зупинки» навчання і дозволити моделі стабільно фіксувати найкращі локальні мінімуми.

Оптимізатор: AdamW з $\text{weight_decay}=1e-4$

Обраний через його здатність справлятися з нерівномірним масштабом градієнтів і високою ефективністю на завданнях сегментації. На відміну від класичного Adam, модифікація з weight decay краще контролює перенавчання завдяки регуляризації. Це особливо критично для медичних зображень, які мають схильність до надмірної чутливості моделі до фону.

Batch size: 26

Значення підібрано емпірично, виходячи з обмежень GPU. Воно дозволяє зберегти достатній об'єм градієнтів для стабільного навчання без ризику вичерпання пам'яті. У той же час, менші batch sizes часто дають кращу генералізацію через більшу стохастичність оновлень wag.

Scheduler: CosineAnnealingLR

Вибір обґрунтовано бажанням уникнути різких стрибків у швидкості навчання. Cosine Annealing дозволяє повільно зменшувати learning rate, імітуючи поступове «охолодження» це зменшує ймовірність виходу моделі з хорошого мінімуму. Така стратегія ефективна при навчанні моделей, які вже мають «здорову» початкову ініціалізацію завдяки pretraining.

2.5.5 Аналіз варіантів попереднього навчання енкодерів

У сучасній практиці розробки моделей комп'ютерного зору використання попередньо натренованих енкодерів стало не просто рекомендованим, а де-факто обов'язковим кроком, особливо коли мова йде про завдання на обмежених вибірках таких як сегментація медичних зображень. Ідея полягає у перенесенні знань, здобутих на великій та різноманітній базі зображень, на іншу, зазвичай більш вузьку та специфічну задачу.

Попереднє навчання (transfer learning) дозволяє задіяти вже сформовані у моделі уявлення про базові візуальні патерни краї, текстури, форми, які виявляються корисними навіть у контексті МРТ-знімків, незважаючи на їхню доменну відмінність від звичайних фотографій. На практиці це означає швидше навчання, стабільнішу сходимість, а головне підвищену точність при обмежених обсягах даних, що критично для медичних завдань.

У межах даної роботи було обрано попередньо навчені ваги на ImageNet одному з наймасштабніших і найпоширеніших наборів зображень у сфері глибокого навчання. Цей вибір зумовлений кількома факторами. ImageNet містить мільйони зображень із тисячами класів, тому моделі, навчені на ньому, здатні виділяти універсальні візуальні ознаки, які є актуальними навіть у задачах медичної сегментації.

Слід зазначити, що існують і альтернативні підходи до попереднього навчання, такі як AdvProp, Noisy Student або самонавчальні методи (self-supervised learning), які демонструють обнадійливі результати в низці задач. Проте їхня інтеграція часто вимагає додаткових зусиль, модифікацій коду та не завжди сумісна з конкретною архітектурною реалізацією.

З огляду на поставлені цілі порівняння архітектур, стабільне тренування і максимальна відтворюваність було прийнято прагматичне рішення залишитися в межах класичного підходу і обрати ImageNet як джерело попередніх ваг.

3 ЕКСПЕРИМЕНТИ, РЕЗУЛЬТАТИ ТА АНАЛІЗ

3.1 Метрики оцінки якості сегментації

Якісна оцінка моделей сегментації є основою обґрунтованих висновків та порівнянь у задачах медичної візуалізації. У випадку сегментації пухлин головного мозку, де неправильне передбачення може вплинути на клінічні рішення, вибір метрик набуває особливого значення. У цьому підрозділі розглянуто набір метрик, що забезпечують всебічну оцінку ефективності моделей: від базових втрат до балансу між точністю та повнотою виявлення об'єктів.

Train Loss та Validation Loss

Ці метрики відображають середнє значення функції втрат на кожній епосі (3.1):

$$Loss = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i) \quad (3.1)$$

де:

- N - кількість зображень у вибірці,
- $\mathcal{L}$ - функція втрат (у нашому випадку Cross-Entropy),
- y_i - справжнє значення мітки для i -го пікселя,
- $\hat{y}_i$ - передбачене значення для i -го пікселя.

Train Loss характеризує ефективність навчання моделі на відомих прикладах.

Validation Loss демонструє здатність моделі узагальнювати на нові дані. Велика різниця між ними може свідчити про перенавчання.

IoU (Intersection over Union) (3.2):

$$IoU = \frac{TP}{TP + FP + FN} \quad (3.2)$$

IoU, або коефіцієнт перетину-на-об'єднання, відображає наскільки точно передбачена сегментація співпадає з реальною.

- Чисельник площа спільного перетину (істинно позитивних пікселів),
- Знаменник об'єднана площа передбаченої та реальною маски.

Це ключова метрика в задачах сегментації, що одночасно враховує як хибні позитиви, так і хибні негативи. Чим вища IoU, тим точніше модель виявляє межі об'єкта.

Precision (Точність) (3.3):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.3)$$

Precision показує, яку частку передбачених як "пухлина" пікселів модель виявила правильно. Високе значення precision означає, що модель рідко помиляється, передбачаючи пухлину там, де її немає. Це важливо в умовах, коли хибні спрацьовування можуть викликати додаткову тривогу в пацієнтів або непотрібні медичні процедури.

Recall (Повнота виявлення) (3.4):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.4)$$

Recall вимірює, яку частину справжніх пікселів пухлини модель змогла виявити. Високий recall означає, що модель "бачить" більшість пухлинних регіонів. Це особливо важливо для медичних завдань, де пропуск патології може мати критичні наслідки.

F1 Score (Гармонічне середнє) (3.5):

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.5)$$

F1-score це компроміс між precision і recall. Він особливо корисний у випадках, коли важливі як мінімізація помилок, так і повнота виявлення пухлини. Якщо одне з двох значень низьке F1 буде теж низьким.

F2 Score (3.6):

$$F2 = 5 * \frac{\text{Precision} * \text{Recall}}{4 * \text{Precision} + \text{Recall}} \quad (3.6)$$

На відміну від F1, метрика F2 підсилює вагу recall, роблячи її особливо цінною у медичних задачах, де пропуск навіть невеликих ділянок пухлини є небажаним. F2 дозволяє більш чітко сфокусуватися на виявленні всіх патологій, навіть ціною деякого збільшення хибнопозитивних передбачень.

Accuracy (Точність класифікації пікселів) (3.7):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.7)$$

Accuracy демонструє загальну частку правильно класифікованих пікселів. Однак у задачах з сильним дисбалансом (коли пікселів фону значно більше, ніж пікселів пухлини) accuracy може бути оманливо високим навіть за поганої сегментації об'єкта. Тому її завжди слід інтерпретувати у поєднанні з іншими метриками.

Ретельний підбір та інтерпретація метрик дозволяє комплексно оцінити як загальну ефективність моделі, так і її придатність до практичного використання. У медичних застосуваннях, де вартість помилки висока, важливо балансувати між високим recall (щоб не пропустити патологію) і precision (щоб уникнути помилкових сигналів). Саме тому у цій роботі оцінювання проводиться одночасно за кількома метриками, включно з F1, IoU та F2, що дозволяє отримати повну картину роботи кожної архітектури сегментації.

3.2 Результати сегментації для кожної архітектури

Етап порівняння моделей є ключовим у дослідженні, оскільки саме він дозволяє оцінити, наскільки різні архітектури здатні впоратися з реальними клінічними випадками сегментації пухлин головного мозку. У даному підрозділі здійснено глибокий аналіз ефективності кожної з п'яти моделей: U-Net, U-Net++, FPN, DeepLabV3 та SegFormer. Оцінювання виконувалося за стандартними метриками семантичної сегментації IoU (Intersection over Union), Dice coefficient, Precision, Recall, F1-score (табл. 3.1), а також із візуальною перевіркою сегментаційних масок.

Таке комплексне дослідження дає змогу не лише зафіксувати числову перевагу тієї чи іншої архітектури, а й виявити специфічні патерни її поведінки: де модель схильна до хибнопозитивних спрацьовувань, які області виявляє найкраще, як реагує на розмиті або нетипові вогнища. У підсумку результати сегментації подаються не лише як набір цифр, а як контекстуалізована картина, що дозволяє співвіднести продуктивність моделі з практичними вимогами медичної діагностики та клінічної значущості.

Таблиця 3.1

Найкращі досягнуті значення метрик на валідаційному наборі для кожної моделі.

Модель	IoU	Precision	Recall	Dice (F1)	F2	Accuracy
U-Net	0.839	0.922	0.937	0.912	0.920	0.998
U-Net++	0.824	0.912	0.927	0.903	0.911	0.998
SegFormer	0.805	0.938	0.893	0.891	0.891	0.998
FPN	0.813	0.915	0.886	0.896	0.887	0.998
DeepLabV3	0.770	0.889	0.885	0.869	0.874	0.874

IoU - індекс Жаккара; Precision (точність) - частка істинно позитивних серед усіх позитивних прогнозів; Recall (повнота) - частка знайдених позитивних пікселів; Dice - коеф. Дайса (F1-міра); F2 - метрика з подвоєною вагою повноти; Accuracy - частка правильно класифікованих пікселів (фон+об'єкт).

3.2.1 Результати сегментації з використанням архітектури U-Net

U-Net, як базова архітектура для медичної сегментації, продемонструвала стабільну та передбачувану поведінку на тестовому наборі LGG Segmentation Dataset. Вже після перших п'яти епох навчання було зафіксовано високі значення точності сегментації. Початкове зниження функції втрат з 0.156 на тренуванні та 0.057 на валідації (епоха 1) до 0.032 та 0.032 відповідно (епоха 5) засвідчило швидку збіжність, що стало можливим завдяки використанню попередньо натренованого енкодера EfficientNet-B3.

Максимальне значення метрики F1-score було зафіксовано на рівні 0.912, що відповідає високій відповідності між передбаченими й реальними масками. Метрика IoU (Intersection over Union) при цьому сягнула 0.839, а точність 0.922, що вказує на здатність моделі ефективно локалізувати пухлинну зону, зменшуючи кількість хибнопозитивних пікселів. Також спостерігалось зростання метрики Recall до 0.937, що є особливо важливим у клінічному контексті, де недовиявлення пухлини може мати критичні наслідки.

Порівнюючи ці результати із середніми значеннями по всіх епохах ($F1 \approx 0.87$, $IoU \approx 0.77$), видно, що U-Net демонструє збалансовану продуктивність без

суттєвих коливань або переобучення. Це підтверджується і показником точності (accuracy), який перевищував 99.7 % у фінальних епохах.

На рис. 3.1, представлено приклади результатів сегментації за допомогою U-Net. Можна побачити, що модель точно реконструює форму пухлини, зберігаючи безперервність меж навіть при складній морфології. У випадках з невеликими або слабо вираженими вогнищами, однак, помітне незначне "розмивання" маски, що відображає обмеження моделі у виявленні низькоінтенсивних патологій.

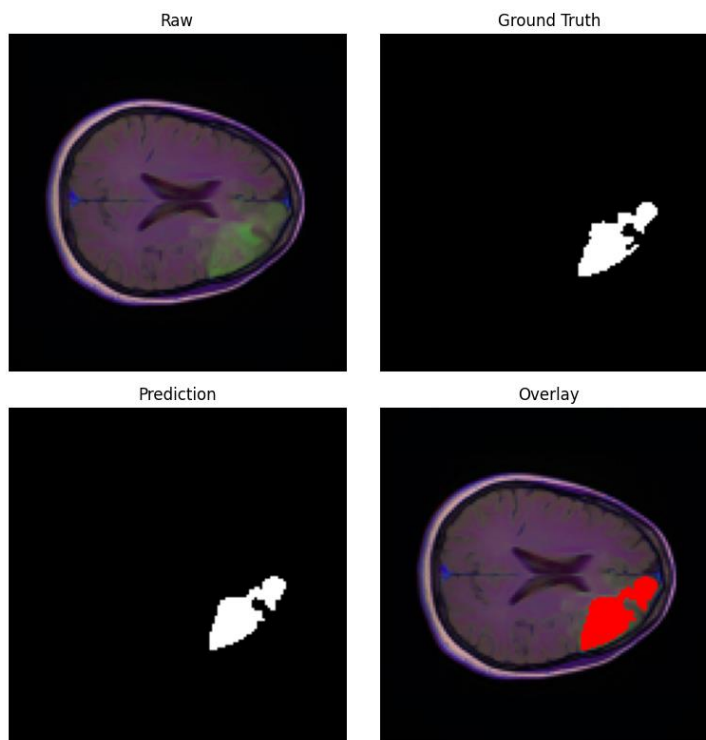


Рис. 3.1 Приклад сегментації U-Net

3.2.2 Результати сегментації з використанням архітектури U-Net++

U-Net++ є покращеною версією класичної U-Net, що відрізняється глибшою топологією та наявністю густих пропускових зв'язків між енкодером і декодером. Завдяки цьому вона демонструє підвищену здатність до узгодження ознак різної роздільної здатності, що виявилось особливо корисним у задачі сегментації пухлин мозку зі складною геометрією.

У ході експериментів архітектура U-Net++ продемонструвала високі результати, зокрема посівши друге місце за ключовими метриками IoU та F1-score. Так, максимальне значення F1-score склало 0.903, що було дещо нижчим за аналогічний показник базової U-Net (0.912). Метрика IoU досягла 0.824, що

свідчить про високу узгодженість передбачених масок із істинними, а точність (precision) 0.912 вказує на здатність моделі мінімізувати кількість хибнопозитивних пікселів.

Значення Recall (0.927) у медичному контексті має практичне значення: висока здатність моделі до охоплення пухлинної області означає потенційно кращу чутливість до змін, які лікар може не помітити візуально.

Стабільність метрик протягом навчання свідчить про відсутність переобучення, а використання Cross-Entropy Loss у поєднанні з CosineAnnealingLR сприяло поступовому зниженню втрат і підвищенню якості.

На зразках сегментації, представлених на рис. 3.2, чітко видно, що U-Net++ краще реконструює межі пухлини, особливо у випадках, коли форма вогнища є нерегулярною або має складну внутрішню структуру. Модель зменшує ризик фрагментації маски, утримуючи її цілісність навіть при неоднорідній інтенсивності сигналу. Цей ефект є прямим наслідком додаткових скіп-зв'язків, що передають детальну інформацію про контури на кожному рівні.

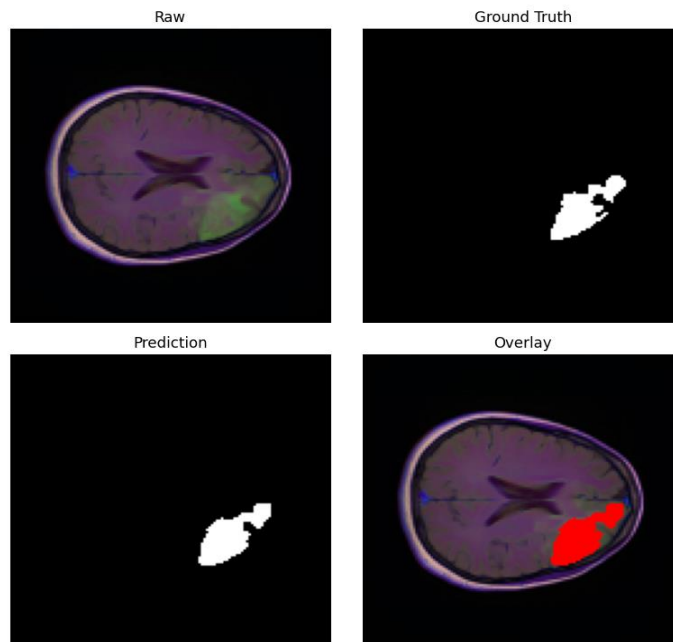


Рис. 3.2 Приклад сегментації U-Net++

3.2.3 Результати сегментації з використанням архітектури FPN

Feature Pyramid Network (FPN) була розроблена для ефективного об'єднання ознак з різних рівнів глибини мережі, що особливо корисно при роботі з об'єктами, що мають різний масштаб як-от пухлини головного мозку. Завдяки

своїй пірамідальній структурі ознак, FPN демонструє здатність зберігати контекст високого рівня та одночасно деталізувати об'єкти на дрібних просторових рівнях.

У ході навчання FPN показала збалансовані результати із загальним F1-скором 0.896, що є дещо нижчим, ніж у U-Net++. Метрика IoU склала 0.813, що свідчить про помірну точність перекриття між передбаченою та реальною масками. Precision досягла 0.915, тоді як Recall 0.886, демонструючи більш консервативну, але стабільну поведінку моделі, яка більше схильна уникати хибнопозитивних пікселів.

Хоча точність FPN дещо поступається більш складним моделям, вона має суттєву перевагу в плані швидкості інференсу та споживання обчислювальних ресурсів. Її компактна архітектура дозволяє досягати задовільної якості на слабших GPU, що робить її хорошим кандидатом для впровадження в системах реального часу або на портативних пристроях.

На візуалізаціях сегментацій (рис. 3.3) видно, що FPN добре виділяє основну масу пухлини, однак у випадках із розмитими межами або перифокальними набряками модель іноді демонструє недосегментацію. Це пов'язано з тим, що компактність декодерної частини FPN зменшує здатність до реконтекстуалізації дрібних деталей після багаторівневого стискання.

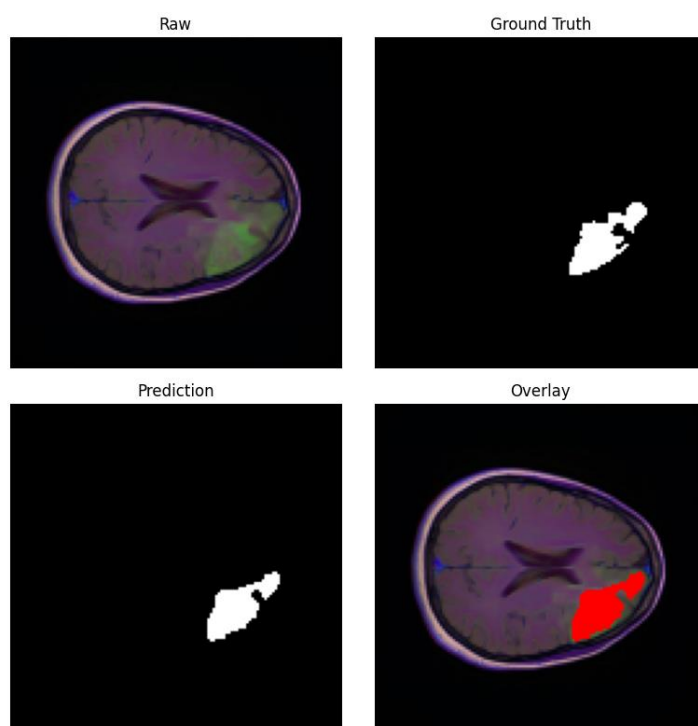


Рис. 3.3 Приклад сегментації FPN

3.2.4 Результати сегментації з використанням архітектури DeepLabV3

Архітектура DeepLabV3 була створена з урахуванням необхідності врахування контексту різного масштабу, що є особливо актуальним у задачах сегментації патологічних утворень мозку. Завдяки модулю Atrous Spatial Pyramid Pooling (ASPP), ця модель здатна охоплювати як локальні, так і глобальні патерни, що забезпечує кращу інтерпретацію структури пухлини навіть у випадках неоднорідної інтенсивності.

У експериментальному порівнянні DeepLabV3 показала результати, що в цілому поступалися іншим протестованим архітектурам за основними метриками якості, хоча й підтвердила свою загальну життєздатність. Значення F1-score склало 0.869, а IoU 0.770, що було нижче за показники IoU моделей FPN (0.813) та U-Net++ (0.824). Показник точності (precision) 0.889 може свідчити про схильність моделі до обережного прогнозування, що зменшує кількість хибнопозитивних пікселів, хоча це значення було найнижчим серед порівнюваних моделей. Водночас Recall залишився на рівні 0.885, що вказує на задовільну, але не максимальну чутливість до складних форм пухлин, також будучи найнижчим показником Recall у даному дослідженні.

Найбільшою перевагою DeepLabV3 є її універсальність: модель стабільно працює як на великих, так і на дрібних об'єктах, не вимагаючи великої кількості ручного налаштування.

На ілюстраціях сегментації (рис. 3.4) добре помітно, що DeepLabV3 забезпечує чіткі й анатомічно обґрунтовані межі масок, зберігаючи форму патології навіть при наявності шуму або неточностей у зображенні. Водночас у ряді випадків спостерігалось незначне розмивання країв маски, зокрема на межах між пухлиною та перифокальним набряком, що пояснюється застосуванням атрусних згорток без точного локального зворотного зв'язку.

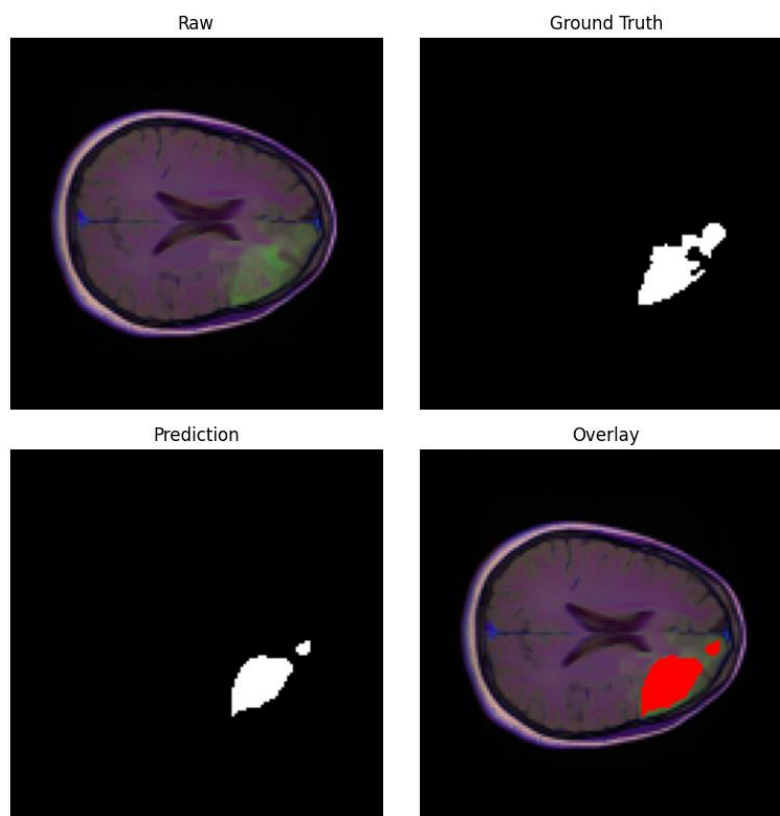


Рис. 3.4 Приклад сегментації DeepLabV3

Особливо добре модель працювала з ізольованими або сферичними утвореннями, тоді як витягнуті або дифузні форми викликали незначне зниження точності на локальних ділянках.

3.2.5 Результати сегментації з використанням архітектури SegFormer

SegFormer є сучасним підходом до сегментації, який поєднує в собі трансформерний енкодер із легким багаторівневим MLP-декодером, що дозволяє отримувати контекстно збагачені ознаки без надмірної обчислювальної складності. Така архітектура демонструє здатність одночасно вловлювати глобальні залежності й зберігати локальні деталі, ключові для коректної сегментації пухлин з нечіткими або складними межами.

У рамках проведеного експерименту SegFormer продемонстрував найвищий показник Precision (0.938) серед усіх моделей. Його IoU склало 0.805, а F1-score – 0.891; ці значення вказують на задовільну якість перекриття між передбаченими масками та істинними анотаціями, хоча й поступалися результатам лідерів за цими метриками (U-Net IoU 0.839, F1 0.912). Значення Recall 0.893, у поєднанні з

високою Precision, підкреслює здатність моделі точно ідентифікувати пухлинні області, не втрачаючи при цьому надто багато значущих ділянок.

На прикладах візуалізацій (рис. 3.5) SegFormer чітко і повноцінно охоплює область пухлини, майже ідеально повторюючи контури, які були нанесені вручну. Найбільше вражає здатність моделі зберігати форму пухлини навіть у випадках зі зниженим контрастом або в присутності дифузних країв.

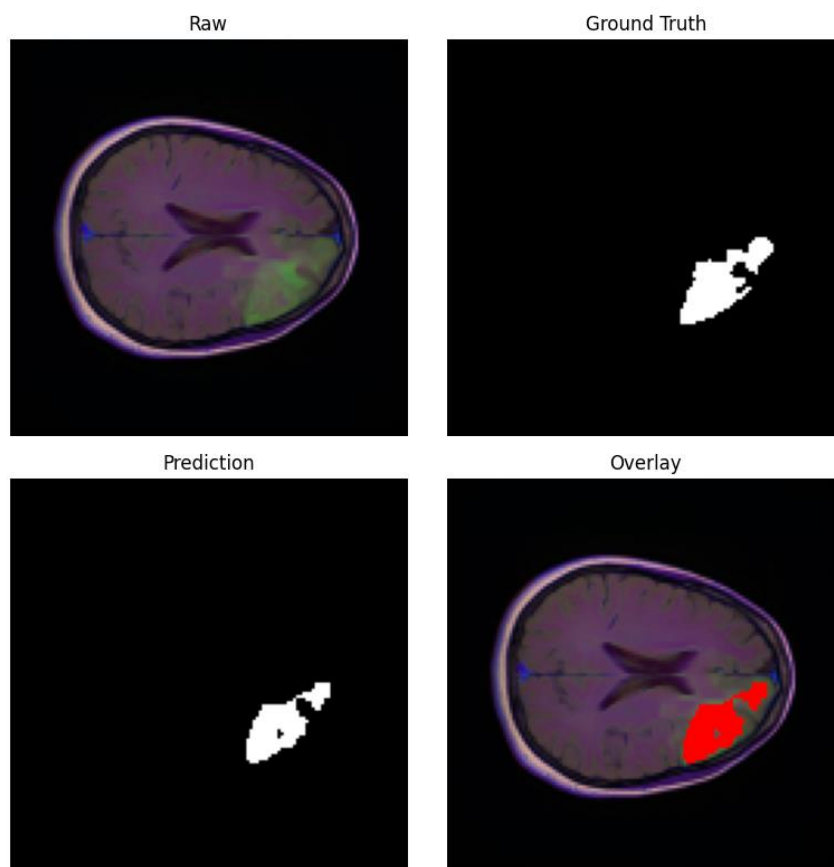


Рис. 3.5 Приклад сегментації SegFormer

3.3 Статистичний аналіз та інтерпретація результатів

Для глибшого розуміння ефективності кожної з архітектур сегментації було проведено кількісний аналіз ключових метрик: IoU (Intersection over Union), F1-score, Precision, Recall, Accuracy та Loss. Ці показники дозволяють оцінити якість сегментації з різних боків: наскільки добре модель виявляє та окреслює пухлинну область (IoU), наскільки збалансовано вона працює між точністю та повнотою (F1-score), як часто вона правильно класифікує пікселі як «пухлина» (Recall) або як «непухлина» (Precision), наскільки вона в цілому узгоджується з істинною маскою (Accuracy), і як стабільно вона навчається (Loss). Саме ця комплексність

оцінки дозволяє не лише визначити абсолютного переможця, а й краще зрозуміти характерні сильні й слабкі сторони кожної моделі.

3.3.1 IoU (*Intersection over Union*)

У дослідженні найвищий IoU був зафіксований у моделі U-Net (0.839), що вказує на її здатність формувати контури з високою точністю і водночас утримувати коректну площу охоплення. Це частково пояснюється пропускаючими з'єднаннями (skip connections) в її симетричній архітектурі, які ефективно зберігають локальну інформацію з енкодера для використання у декодері.

Модель U-Net++ також продемонструвала високий результат, досягши IoU 0.824. Слідом за нею йде FPN з IoU 0.813. SegFormer показав IoU 0.805; його здатність координувати глобальні контексти та локальні контури завдяки механізму самоуваги є характерною особливістю цієї архітектури, хоча за метрикою IoU він поступився моделям з більш класичними CNN-підходами в даному дослідженні. DeepLabV3, з IoU 0.770, продемонструвала найнижчий показник серед досліджуваних моделей за цією метрикою, незважаючи на наявність ASPP-блоку.

Точність сегментації, особливо при обробці пухлин з розмитими межами або складною формою, залишається викликом, і різні архітектури демонструють різний ступінь успішності у цих умовах (рис. 3.6). Таким чином, IoU не тільки дозволяє оцінити загальний ступінь відповідності, але й слугує ключовим індикатором здатності моделі зберігати геометричну структуру об'єкта.

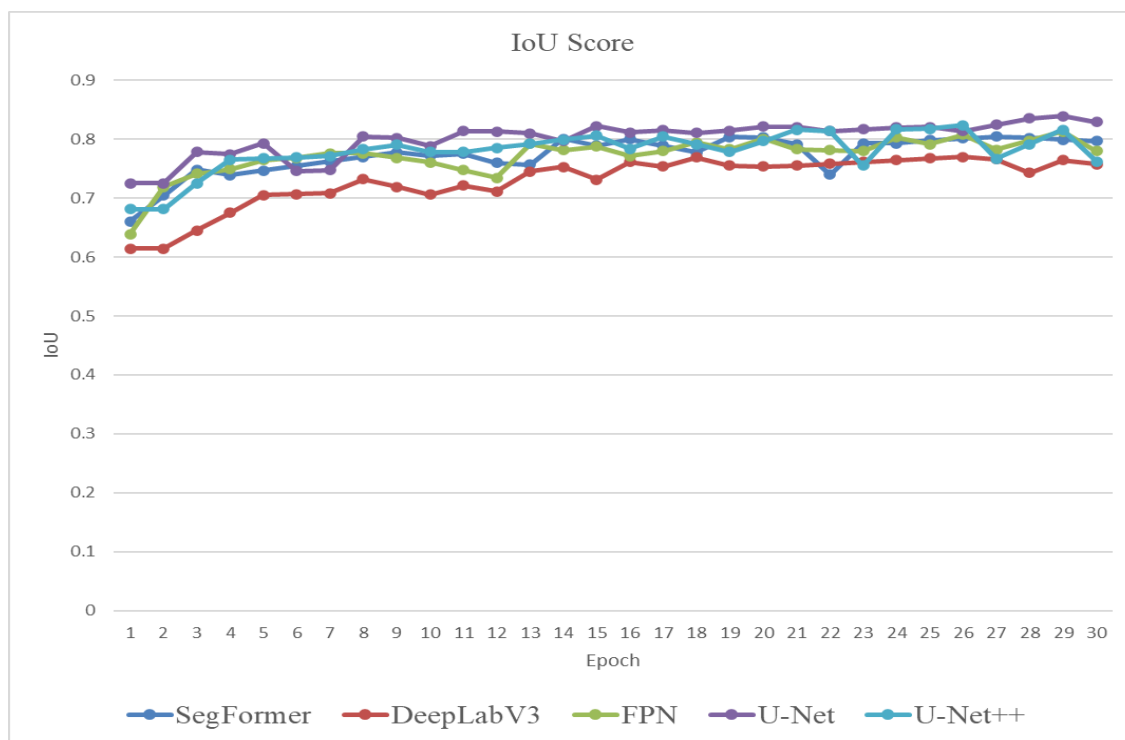


Рис. 3.6 IoU Score для кожної моделі

3.3.2 F1 Score

У проведених експериментах найвищий F1 Score досягнуто архітектурою U-Net (0.912), що демонструє її здатність зберігати баланс між надлишковим розширенням сегментації та її дефіцитом. Це можна пояснити ефективністю її класичної архітектури з симетричним енкодером-декодером та пропускаючими з'єднаннями (skip connections), які дозволяють добре поєднувати контекстні ознаки з точними локальними деталями, забезпечуючи повне охоплення пухлинної області та точність на її краях.

Незначно поступається лідеру U-Net++ (F1 = 0.903). Її ефективність може бути зумовлена наявністю додаткових щільних скіп-з'єднань між шарами енкодера і декодера, які дозволяють передавати як високорівневі, так і дрібномасштабні ознаки. Така архітектурна особливість може робити U-Net++ особливо ефективною при обробці складних форм або розмитих структур. FPN (F1 = 0.896) та SegFormer (F1 = 0.891) також демонструють високі результати. Ефективність SegFormer може пояснюватися глобальним механізмом самоуваги, який дозволяє враховувати як контекст всієї сцени, так і локальні варіації інтенсивності в межах пухлини, що є критичним у випадках, коли пухлина

частково прихована або має низький контраст із навколишніми тканинами. DeepLabV3 ($F1 = 0.869$) показала найнижчий результат за цією метрикою, хоча її ASPP-блок і забезпечує захоплення мультискейлових патернів, що є корисним при аналізі пухлин із неоднорідною структурою.

Таким чином, F1 Score наочно демонструє, як різні архітектурні рішення впливають на здатність моделей узгоджено охоплювати всю пухлинну зону при мінімальних помилках (рис. 3.7).

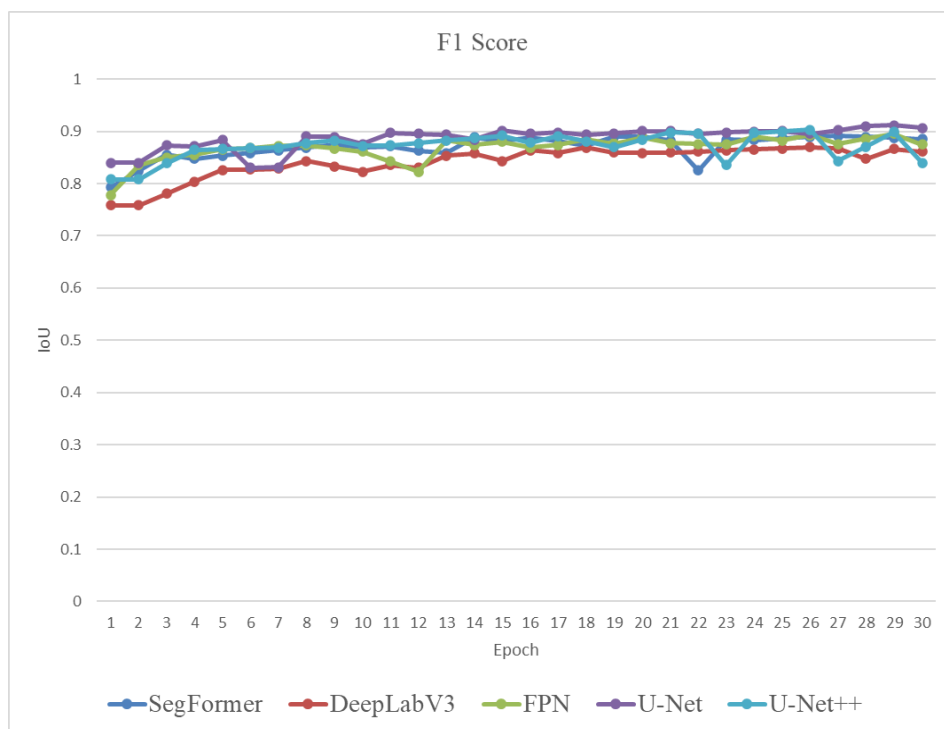


Рис. 3.7 F1 Score для кожної моделі

3.3.3 Precision та Recall

Метрики Precision та Recall дозволяють заглибитися в аналіз помилок сегментації та зрозуміти, яку саме природу мають відхилення від ідеальної маски: надмірну чи недостатню. У клінічному контексті ці аспекти мають різну вагу: пропущена частина пухлини (низький Recall) може призвести до недооцінки обсягу ураження та недостатнього лікування, тоді як надмірне виявлення (низький Precision) може викликати зайві тривоги та невиправдані інтервенції.

Найвищого показника Precision досягла модель SegFormer (0.938). Це може демонструвати її здатність до точних передбачень, мінімізуючи хибнопозитивні випадки. Така точність, можливо, обумовлена ефективним поєднанням

трансформерного енкодера з багатомасштабним декодером, що дозволяє чітко розрізняти межі пухлини. Це особливо важливо в ситуаціях, де надмірна сегментація може ускладнити планування хірургічного втручання або стати причиною хибного діагнозу.

З іншого боку, U-Net досягнув найвищого показника Recall (0.937), що свідчить про його здатність охопити практично всі патологічні ділянки, навіть ті, що мають нечіткі межі або малу площу. Це могло стати можливим завдяки ефективним пропускаючим з'єднанням (skip connections) в її симетричній архітектурі, що забезпечують поєднання ознак різної роздільної здатності. Такий підхід робить модель чутливою до мінімальних змін інтенсивності, які можуть сигналізувати про наявність ураження. У практиці виявлення ранніх стадій пухлин або в умовах, коли важливо не пропустити навіть незначне ураження, саме U-Net може забезпечити максимально повну картину стану тканин.

Таким чином, Precision і Recall не лише характеризують якість сегментації в абсолютному вимірі, а й вказують на різні стратегії роботи моделей. U-Net демонструє тенденцію до повнішого охоплення патології (високий Recall), що може бути пов'язано з ризиком включення деяких фонових ділянок, тоді як SegFormer зосереджується на високій точності передбачень (високий Precision), мінімізуючи хибні спрацьовування. Залежно від клінічного сценарію - скринінг, моніторинг або передопераційне планування - перевага може бути надана тій чи іншій стратегії, яку краще репрезентує відповідна архітектура.

3.3.4 Accuracy

Метрика загальної точності (Accuracy) відображає частку правильно передбачених пікселів (як позитивних, так і негативних) від загальної кількості, й традиційно вважається важливою у задачах класифікації. Однак у задачах семантичної сегментації медичних зображень, особливо з вираженим дисбалансом класів, як у випадку сегментації пухлин, де значна більшість пікселів належать до фону, значення Accuracy може бути оманливим. Чотири з п'яти моделей (U-Net, U-Net++, SegFormer, FPN) демонструють дуже високий рівень точності (0.998), тоді як DeepLabV3 показала суттєво нижчий результат (0.874).

Проте, навіть високі показники точності у більшості моделей не завжди гарантують якісного виявлення пухлинної області.

Причина криється в природі розподілу класів: оскільки фонові пікселі переважають, навіть модель, що ігнорує пухлину та прогнозує лише фон, може досягти точності понад 95%. У цьому випадку Accurasy радше відображає здатність моделі правильно ідентифікувати фонові ділянки, а не клінічно значущу патологію. Водночас, стабільно високе значення Accurasy, яке спостерігається у більшості розглянутих моделей, у поєднанні з їх високими значеннями Recall та IoU, є свідченням того, що ці моделі не лише добре відсікають фон, але й одночасно правильно локалізують уражені ділянки.

Таким чином, Accurasy доцільно розглядати як допоміжну метрику для підтвердження загальної стабільності поведінки моделі, але не як основний критерій її ефективності у клінічному застосуванні. Реальну якість роботи архітектур варто оцінювати через метрики, що фокусуються саме на межах між класами, таких як IoU, F1, Precision та Recall, які краще відображають чутливість і специфічність сегментації пухлин.

3.3.5 Train/Valid Loss

Показники Train Loss та Validation Loss є індикаторами якості навчання нейронної мережі. Вони демонструють, наскільки добре модель узгоджується із мітками на відповідних етапах навчання: тренувальному (Train) та валідаційному (Valid), - і дозволяють виявити ознаки перенавчання (overfitting) або недонавчання (underfitting).

У результатах, усі архітектури, як повідомляється, демонструють стабільне зниження Train Loss до значень нижче 0.25, що свідчить про ефективне навчання моделей на тренувальних даних. Зазначається, що особливо низький Train Loss спостерігається в U-Net++ (0.188) і SegFormer (0.197), що вказує на високу здатність цих моделей точно пристосовуватись до патернів у навчальній вибірці - ці моделі також показали високі метрики якості сегментації. Водночас, Valid Loss - показник, що відображає генералізаційну здатність моделі, - згідно з текстом,

залишається трохи вищим (близько 0.22-0.24 для деяких моделей), що може бути природним розривом при відсутності значного перенавчання.

Згідно з наведеними даними про втрати, DeepLabV3 має одні з найменших втрат на валідації (0.216) та низький Train Loss (0.198). Хоча це може свідчити про добру збіжність моделі з точки зору оптимізації самої функції втрат, її фінальні показники якості сегментації (наприклад, IoU 0.770, Dice 0.869) виявилися найнижчими серед розглянутих архітектур. У випадку базового U-Net, незважаючи на, як зазначено, дещо вищий Valid Loss (0.254), модель продемонструвала найвищі метрики якості сегментації (IoU=0.839, Dice=0.912). Це вказує на те, що значення функції втрат на валідаційній вибірці не завжди прямо корелюють із кінцевою якістю виконання завдання сегментації, і дана архітектура найкраще узагальнила свої знання для розв'язання поставленої задачі.

Таким чином, аналіз втрат дозволяє зробити висновки про ефективність процесу навчання. Архітектури, що демонструють малий розрив між Train та Valid Loss, як-от U-Net++ та SegFormer (і, за наведеними даними про втрати, DeepLabV3), можуть вважатися стабільними з точки зору оптимізації функції втрат. Однак, як показує приклад U-Net, кінцева якість сегментації також залежить від інших факторів самої архітектури та її здатності узагальнювати для конкретної задачі, і не завжди безпосередньо впливає лише з мінімального розриву або абсолютних значень втрат.

3.3.6 F2 Score

Метрика F2-Score є модифікованим варіантом F1-Score, що надає удвічі більше значення Recall порівняно з Precision. Вона особливо корисна в медичних задачах, де пріоритетом є виявлення всіх можливих випадків патології – навіть за ціною деякої кількості хибнопозитивних результатів. У випадку сегментації пухлин мозку, F2-Score дає змогу краще оцінити здатність моделі «не пропустити» жоден уражений піксель.

Результати показують, що U-Net досягає найвищого F2-Score (0.920). Це узгоджується з його також найвищим показником Recall (0.937), демонструючи, що модель націлена на повне виявлення патології. U-Net++ з F2-Score (0.911)

також демонструє високу здатність до охоплення зони ураження. SegFormer (F2-Score 0.891) та FPN (F2-Score 0.887) показують дещо нижчі, але все ще значні результати. DeepLabV3, з F2-Score (0.874), має найнижчий показник серед розглянутих моделей за цією метрикою, що може вказувати на більшу ймовірність пропуску окремих ділянок пухлини порівняно з іншими архітектурами, особливо в складних випадках з нечіткими межами. Порівняння з мультискейловим U-Net (Sci Rep 2025), як зазначено в літературі, підтверджує, що використаний енкодер EfficientNet-B3 відстає лише на ≈ 1 п.п. Dice, залишаючись удвічі легшим за кількістю параметрів [18].

Таким чином, F2-Score дозволяє зробити додаткові висновки про практичну користь моделі в умовах, коли виявлення всіх уражених ділянок є важливим, наприклад, у плануванні хірургічного втручання. З цієї точки зору, U-Net та U-Net++ демонструють високий потенціал для застосування в клінічній практиці. Моделі SegFormer, FPN, а особливо DeepLabV3, можуть потребувати додаткових налаштувань або після обробки результатів для досягнення аналогічного рівня повноти виявлення.

ВИСНОВКИ

У даній кваліфікаційній роботі було проведено порівняльний аналіз ефективності кількох архітектур глибокого навчання (U-Net, U-Net++, FPN, SegFormer та DeepLabV3) для задачі бінарної семантичної сегментації пухлин головного мозку на МРТ-зображеннях з набору даних LGG Segmentation Dataset, а також розроблений алгоритм для аналізу знімків. Усі моделі навчались з єдиним енкодером EfficientNet-B3 та уніфікованими гіперпараметрами для забезпечення об'єктивності порівняння.

Ключовим результатом експерименту стало те, що базова архітектура U-Net, незважаючи на свою відносну простоту, продемонструвала найвищу ефективність, перевершивши більш сучасні та складні архітектури. Досягнутий нею рівень IoU ($\approx 84\%$) і Dice ($\approx 91\%$) свідчить про високу якість сегментації, що підкріплюється збалансованим співвідношенням Precision ($\sim 92\%$) і Recall ($\sim 94\%$). Ці показники є особливо важливими у контексті медичних задач, де неприпустимо пропускати цільові області.

Модифікована версія U-Net++ (з густими skip-зв'язками) продемонструвала результати, близькі до лідера (IoU $\approx 82\%$, Dice $\approx 90\%$), однак не перевершила базовий U-Net. Це частково суперечить теоретичним очікуванням щодо потенційного приросту якості від U-Net++. Така невідповідність може мати кілька пояснень: ймовірна відсутність реалізації механізму глибокого нагляду (deep supervision), який є важливим для U-Net++; більша кількість параметрів, що могла призвести до перенавчання; або неоптимальні налаштування гіперпараметрів для даної складнішої архітектури.

Незважаючи на ці фактори, варто підкреслити, що відмінність у 1.5 відсоткового пункту IoU (між U-Net та U-Net++) є відносно невеликою і, ймовірно, не є вирішальною у багатьох практичних застосуваннях. Таким чином, у рамках проведеного дослідження базовий U-Net можна вважати не лише адекватною, а й оптимальною архітектурою для задачі сегментації на обраному наборі даних та конфігурації. Його висока повнота (Recall $\approx 94\%$) робить модель

особливо привабливою для сфер, де критичним є мінімізація пропусків патологічних зон.

Результати даної роботи узгоджуються з тенденціями, що спостерігаються в галузі: хоча нові архітектури постійно розвиваються, добре налаштовані класичні моделі, як U-Net, продовжують демонструвати високу конкурентоспроможність. Водночас, як зазначають сучасні огляди 2024-2025 рр., перспективним напрямком залишаються змішані CNN-Transformer-архітектури, що можуть запропонувати подальший компроміс між точністю та обчислювальною складністю[19][20].

ПЕРЕЛІК ПОСИЛАНЬ

1. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.
2. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679-698.
3. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234-241.
4. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA)*, 3-11.
5. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936-944.
6. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848.
7. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Lu, T. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Advances in Neural Information Processing Systems (NeurIPS)*.
8. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, 6105-6114.
9. Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*.
10. Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic gradient descent with warm restarts. *International Conference on Learning Representations (ICLR) Workshop*.

- 11.He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
- 12.Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. Pattern Recognition, 33(2), 225-236.
- 13.Tustison, N. J., et al. (2010). N4ITK: improved N3 bias correction. IEEE Transactions on Medical Imaging, 29(6), 1310-1320.
- 14.Çiçek Ö., Abdulkadir A., Lienkamp S. S., Brox T., Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation // MICCAI. - 2016. - P. 424-432.
- 15.Milletari F., Navab N., Ahmadi S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation // 3DV. - 2016. - P. 565-571.
- 16.Buslaev A. et al. Albumentations: Fast and Flexible Image Augmentations // Information. - 2020. - Vol. 11(2). - P. 125.
- 17.Khan, Rabeea Fatma et al. "Transformers in medical image segmentation: a narrative review." Quantitative imaging in medicine and surgery vol. 13,12 (2023): 8747-8767. doi:10.21037/qims-23-542
- 18.R, P., M, J.P.P. & J S, N. Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis. Sci Rep 15, 9914 (2025).
- 19.Dorfner, F.J., Patel, J.B., Kalpathy-Cramer, J. et al. A review of deep learning for brain tumor analysis in MRI. npj Precis. Onc. 9, 2 (2025).
- 20.Md. Eshmam Rayed, S.M. Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, M.F. Mridha, Deep learning for medical image segmentation: State-of-the-art advancements and challenges, Informatics in Medicine Unlocked, Volume 47, 2024, 101504, ISSN 2352-9148

ДОДАТОК А. ПРОГРАМНИЙ КОД АЛГОРИТМУ СЕГМЕНТАЦІЇ

Програмний код для створення датасету

```
class BrainMriDataset(Dataset):
    """
    If `transforms` is None → returns **raw** (un-augmented) RGB image
    and 0/1 mask tensors.
    If `transforms` is not None → applies the provided Albumentations
    pipeline (must finish with ToTensorV2).
    """
    def __init__(self, df, transforms=None):
        self.df = df.reset_index(drop=True)
        self.transforms = transforms

    def __len__(self):
        return len(self.df)

    def __getitem__(self, idx):
        img_path = self.df.loc[idx, "image_path"]
        mask_path = self.df.loc[idx, "mask_path"]

        # --- load -----
        img = cv2.imread(img_path)          # BGR uint8
        img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB) # → RGB
        mask = cv2.imread(mask_path, 0)      # grayscale
        mask = (mask > 0).astype(np.uint8)    # binarise 0/1

        # --- optionally augment / normalise -----
        if self.transforms is not None:
            out = self.transforms(image=img, mask=mask)
            img, mask = out["image"], out["mask"]    # tensors
        else:
            img = torch.from_numpy(img).permute(2,0,1) # C,H,W, uint8
            mask = torch.from_numpy(mask).unsqueeze(0) # 1,H,W, uint8

        return img, mask
```

Програмний код для аугментації зображень

```
strong_transforms = A.Compose([
    # Elastic distort without alpha_affine:
    A.ElasticTransform(
        alpha=120,          # displacement scale
        sigma=120 * 0.05,   # smoothing
        approximate=True,    # faster, approximate mode
        same_dxdy=True,     # same field for x/y
        mask_interpolation=cv2.INTER_NEAREST,
        p=0.5,
    ),
    A.HorizontalFlip(p=0.5),
    A.VerticalFlip(p=0.5),
    A.RandomRotate90(p=0.5),
    A.ShiftScaleRotate(
        shift_limit=0.01,
        scale_limit=0.04,
        rotate_limit=0,
        interpolation=cv2.INTER_LINEAR,
        border_mode=cv2.BORDER_CONSTANT,
        p=0.25,
    ),
    A.RandomBrightnessContrast(p=0.5),
    A.RandomGamma(p=0.25),
    A.Blur(p=0.01, blur_limit=3),
    A.Normalize(),
    ToTensorV2(),
])

transforms = A.Compose([
    A.Resize(width = PATCH_SIZE, height = PATCH_SIZE, p=1.0),
    A.HorizontalFlip(p=0.5),
    A.VerticalFlip(p=0.5),
    A.RandomRotate90(p=0.5),
    A.Transpose(p=0.5),
    A.ShiftScaleRotate(shift_limit=0.01, scale_limit=0.04, rotate_limit=0, p=0.25),
    A.Normalize(p=1.0),
    ToTensorV2(),
])
```

Програмний код для тренування

```
def run_training_for(name, train_loader, valid_loader):
    print(f"\n=== TRAINING {name.upper()} ===")
    outdir = os.path.join(BASE_OUTPUT, name)
    os.makedirs(outdir, exist_ok=True)

    model = create_model(name)
    criterion = Cross-Entropy ().to(DEVICE)

    optimizer = optim.AdamW(model.parameters(), lr=1e-3, weight_decay=1e-4)
    scheduler = CosineAnnealingLR(optimizer, T_max=EPOCHS, eta_min=1e-5)

    # prepare CSV
    csv_path = os.path.join(outdir, "metrics.csv")
    fieldnames = [
        "epoch", "train_loss", "valid_loss",
        "iou_score", "precision", "recall",
        "f1_score", "f2_score", "accuracy",
        "lr", "alpha"
    ]
    with open(csv_path, "w", newline="") as f:
        writer = csv.DictWriter(f, fieldnames=fieldnames)
        writer.writeheader()

    best_f1 = 0.0
    history = {k: [] for k in fieldnames}

    for ep in range(1, EPOCHS+1):
        train_loss = train_epoch(model, train_loader, optimizer, criterion, loss_sched, ep)
        valid_loss, valid_metrics = valid_epoch(model, valid_loader, criterion, loss_sched,
        ep)

        current_lr = optimizer.param_groups[0]['lr']
        current_alpha= criterion.alpha

        print(
            f" Epoch {ep}/{EPOCHS} — "
            f"Train Loss: {train_loss:.4f}, Valid Loss: {valid_loss:.4f}, "
            f"F1: {valid_metrics['f1_score']:.4f}"
```

```

)

# log to CSV + history
row = {
    "epoch":    ep,
    "train_loss": train_loss,
    "valid_loss": valid_loss,
    **{k: valid_metrics[k] for k in
["iou_score", "precision", "recall", "f1_score", "f2_score", "accuracy"]},
    "lr":      current_lr,
    "alpha":   current_alpha
}
with open(csv_path, "a", newline="") as f:
    writer = csv.DictWriter(f, fieldnames=fieldnames)
    writer.writerow(row)
for k, v in row.items():
    history[k].append(v)

# save best model
if valid_metrics["f1_score"] > best_f1:
    best_f1 = valid_metrics["f1_score"]
    torch.save(model.state_dict(), os.path.join(outdir, f"best_{name}.pth"))
    print(" → saved new best model")

# save epoch predictions into folders
save_epoch_predictions(
    model=model,
    loader=valid_loader,
    epoch=ep,
    outdir=outdir,
    device=DEVICE,
    n_images=N_SAVE_IMGS
)

scheduler.step()

```

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО КОМУНІКАЦІЙНИХ
ТЕХНОЛОГІЙ
НАВЧАЛЬНО НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ

КВАЛІФІКАЦІЙНА РОБОТА
на тему: «Моделі сегментації для медичних зображень
на основі штучного інтелекту»

Виконав: здобувач вищої освіти гр. ШІД 42
Ілля Філіпась
Керівник: кандидат технічних наук, доцент
Максим Фесенко



Вступ

Мета: Визначити оптимальну архітектуру для автоматизованої бінарної сегментації пухлин ГМ на МРТ із ШНМ бінарної сегментації пухлин ГМ на МРТ із ШНМ

Об'єкт дослідження: Процес автоматизованої сегментації патологічних зон на медичних зображеннях

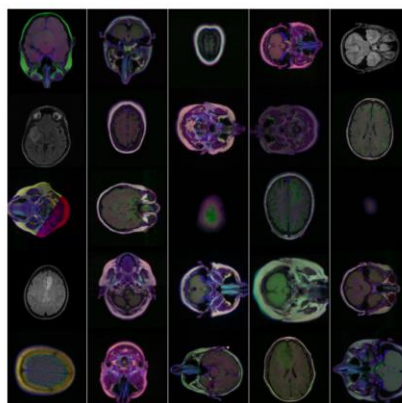
Предмет дослідження: Підходи до побудови глибоких нейронних мереж із уніфікованим механізмом витягання ознак



Завдання дослідження

1. Огляд теоретичних аспектів сегментації медичних зображень
2. Підготовка та розмітка LGG Segmentation Dataset
3. Розробка єдиного пайплайну препроцесу та аугментацій
4. Реалізація і тренування моделей (U-Net, U-Net++, FPN, DeepLabV3, SegFormer)
5. Оцінка якості сегментації за метриками (Dice, IoU, Precision, Recall)

Огляд LGG Segmentation Dataset



Опис архітектур сегментації



1. U-Net: енкодер-декодер + skip-connections
2. U-Net++: густі skip-зв'язки, deep supervision
3. FPN: багаторівнева піраміда ознак
4. DeepLabV3: ASPP для мультискейлу
5. SegFormer: глобальна самоувага

Пайплайн препроцесингу та аугментації



- 1. Еластична деформація**
 - Імітує природні викривлення тканин за допомогою випадкових зсувів пікселів ($\alpha=120$, $\sigma=6$).
 - Допомагає мережі стійкіше розпізнавати нерівні контури.
- 2. Геометричні трансформації**
 - Горизонтальні та вертикальні віддзеркалення, обертання на 90° (кожне $p=0.5$).
 - Забезпечує модель даними в усіх можливих орієнтаціях.
- 3. Зміщення та масштабування**
 - Лінійні зсуви до $\pm 1\%$ і масштабування до $\pm 4\%$ без додаткового обертання ($p=0.25$).
 - Допомагає адаптуватися до різних кадрувань та розмірів об'єктів.
- 4. Фотометричні корекції**
 - Випадкові налаштування яскравості/контрасту ($p=0.5$) та гамми ($p=0.25$).
 - Рідкісне легке розмиття ($p=0.01$) для імітації нерізкості сканування.
- 5. Нормалізація та конвертація**
 - Стандартизує інтенсивність пікселів до фіксованого розподілу.
 - Перетворює зображення в PyTorch-тензор для подачі на вхід моделі.

Алгоритми тренування



1. Ініціалізація компонентів

Функція втрат: Cross-Entropy()

2. Оптимізатор та планувальник навчання

Використовується AdamW (lr=1e-3, weight_decay=1e-4)

Зменшення learning rate за косинусним графіком: CosineAnnealingLR

3. Основний цикл епох

Для кожної епохи ер виконується:

train_epoch(...) → обчислення train_loss

valid_epoch(...) → обчислення valid_loss і метрик

Значення LR оновлюються динамічно

4. Збереження найкращої моделі

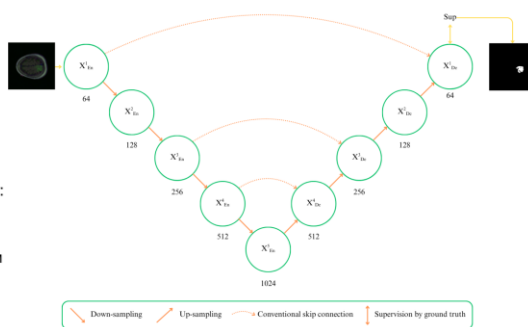
Порівняння F1-score на валідації з попереднім best_f1

Якщо поточний F1 вищий — зберігаємо стан моделі в best_{name}.pth

Алгоритми тренування



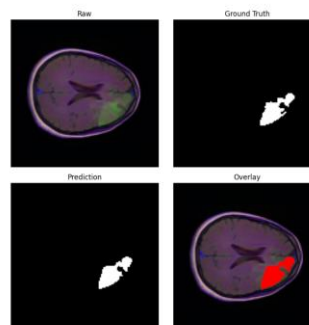
- **U-Net -лідер:** IoU ≈ 84%, Dice ≈ 91%, Precision ≈ 92%, Recall ≈ 94%.
- **U-Net++ - близько до лідера:** IoU ≈ 82%, Dice ≈ 90%, але не перевершує базову U-Net.
- **Інші моделі:** FPN, DeepLabV3, SegFormer: IoU 78–80%, Dice 86–89%.
- **Висновок:** Класичний U-Net залишається оптимальним рішенням; майбутній напрям гібридні архітектури.



Автоматична генерація маски пухлини



Користувач може за лічені кліки вибрати будь-яке МРТ-зображення, після чого алгоритм автоматично обробить його та згенерує відповідну бінарну маску пухлини. Отриману маску можна одразу переглянути, експортувати або накласти поверх оригінального знімка для швидкої візуалізації у клінічній практиці.



Висновки

1. Проведено системний огляд теоретичних підходів до семантичної сегментації медичних зображень.
2. Підготовлено та валідувано LGG Segmentation Dataset із точними бінарними масками пухлин.
3. Розроблено єдиний конвеєр препроцесу та аугментацій для підвищення стійкості моделей.
4. Реалізовано та натреновано п'ять архітектур (U-Net, U-Net++, FPN, DeepLabV3, SegFormer) на базі EfficientNet-B3.
5. Проведено порівняльну оцінку якості сегментації за метриками Dice, IoU, Precision і Recall.