

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Навчально-науковий інститут прикладного системного аналізу

Кафедра штучного інтелекту

До захисту допущено:

В. о. завідувачки кафедри

_____ Ірина ДЖИГИРЕЙ

«__» _____ 20__ р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системи і методи штучного інтелекту»

спеціальності 122 «Комп'ютерні науки»

**на тему: «Інтелектуальна система стадіювання раку молочної залози за
гістологічними зображеннями цілого зрізу в умовах слабкої розмітки»**

Виконав:

студент IV курсу, групи KI-21

Богатиренко Віктор Андрійович _____

Керівник:

асистент кафедри штучного інтелекту,

Сидорський Володимир Сергійович _____

Консультант з нормоконтролю:

фахівець першої категорії кафедри штучного інтелекту, к.т.н., доцент,

Комариста Богдана Миколаївна _____

Рецензент:

асистент кафедри математичних методів системного аналізу, к.т.н.,

Канцедал Георгій Олегович _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2026 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Навчально-науковий інститут прикладного системного аналізу
Кафедра штучного інтелекту

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 122 «Комп'ютерні науки»

Освітньо-професійна програма «Системи і методи штучного інтелекту»

ЗАТВЕРДЖУЮ

В. о. завідувачки кафедри

_____ Ірина ДЖИГИРЕЙ

«15» січня 2026 р.

ЗАВДАННЯ
на дипломну роботу студенту
Богатиренку Віктору Андрійовичу

1. Тема роботи: «Інтелектуальна система стадіювання раку молочної залози за гістологічними зображеннями цілого зрізу в умовах слабкої розмітки», керівник роботи Сидорський Володимир Сергійович, асистент, затверджені наказом по НН ІПСА від «27» травня 2026 р. № 1944-с.
2. Термін подання студентом роботи «05» червня 2026 року.
3. Вихідні дані до роботи: передбачені класи слайдів, відповідні ймовірності, та розподіл ваг уваги по патчах для інтерпретації рішень моделі.
4. Зміст роботи: розробка інтелектуальної системи автоматичного стадіювання раку молочної залози за гістологічними зображеннями цілого зрізу (WSI) в умовах слабкої розмітки на основі методів Multiple Instance Learning, фундаментальних моделей комп'ютерного зору та механізмів уваги
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): схеми пайплайну обробки WSI, архітектури CLAM та інших MIL-моделей, процесу сегментації тканини й відбору патчів, результати класифікації та стадіювання, матриці невідповідностей, ROC-криві, карти уваги (attention maps), візуалізації простору ознак та презентація результатів дослідження.

6. Дата видачі завдання «02» лютого 2026 року.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Вивчення літератури за темою роботи	20.04.2026	Виконано
2	Підготовка першого розділу.	27.04.2026	Виконано
3	Проведення дослідження за темою роботи	04.05.2026	Виконано
4	Розробка методу	11.05.2026	Виконано
5	Підготовка другого розділу	18.05.2026	Виконано
6	Проведення експериментів	25.05.2026	Виконано
7	Підготовка третього розділу	01.06.2026	Виконано
8	Оформлення дипломної роботи	05.06.2026	Виконано

Студент

Віктор БОГАТИРЕНКО

Керівник

Володимир СИДОРСЬКИЙ

РЕФЕРАТ

Дипломна робота: 111 с., 25 рис., 7 табл., 45 посилань, 1 додаток.

РАК МОЛОЧНОЇ ЗАЛОЗИ, ГІСТОПАТОЛОГІЧНИЙ АНАЛІЗ, ЦИФРОВА ПАТОЛОГІЯ, WHOLE SLIDE IMAGE, MULTIPLE INSTANCE LEARNING, СЛАБКА РОЗМІТКА, CLAM, МЕХАНІЗМ УВАГИ, ФУНДАМЕНТАЛЬНІ МОДЕЛІ, UNI2-H, СТАДІЮВАННЯ.

Об'єктом дослідження є процес автоматизованої гістопатологічної діагностики раку молочної залози на основі Whole Slide Images (WSI) в умовах слабкої розмітки.

Предметом дослідження є методи Multiple Instance Learning, архітектури глибокого навчання з механізмами уваги та пайплайни обробки WSI для автоматичного стадіювання раку молочної залози.

Метою роботи є розробка інтелектуальної системи автоматичного стадіювання раку молочної залози за гістологічними WSI в умовах слабкої розмітки даних.

Оглянуто сучасні архітектури глибокого навчання та методи самонавчання; як екстрактор ознак застосовано UNI2-H (ViT-Huge, 1536-D), для агрегації — CLAM з attention-pooling. Розроблено пайплайн обробки WSI з сегментацією тканини, нарізкою на патчі 224×224 пікселів та semi-hard patch mining.

Порівняння ABMIL, CLAM, TransMIL і DSMIL на CAMELYON16+17 показало: DSMIL найкращий для бінарної класифікації ($AUC = 0,977$, $F1 = 0,986$), CLAM — для мультикласового стадіювання ($QWK = 0,961$); на TCGA-SKCM усі моделі досягли $Recall = 1,0$ для класу меланоми.

ABSTRACT

Thesis: 111 pp., 25 fig., 7 tables, 45 references, 1 appendix.

BREAST CANCER, HISTOPATHOLOGICAL ANALYSIS, DIGITAL PATHOLOGY, WHOLE SLIDE IMAGE, MULTIPLE INSTANCE LEARNING, WEAK LABELLING, CLAM, ATTENTION MECHANISM, FUNDAMENTAL MODELS, UNI2-H, STAGING.

The object of research is the process of automated histopathological diagnosis of breast cancer based on Whole Slide Images (WSI) under weak annotation conditions.

The subject of research is Multiple Instance Learning methods, deep learning architectures with attention mechanisms, and WSI processing pipelines for automatic breast cancer staging.

The aim of the work is to develop an intelligent system for automatic breast cancer staging based on histological WSI under weak labelling conditions.

Modern deep learning architectures and self-supervised methods were reviewed; UNI2-H (ViT-Huge, 1536-D) was used as the feature extractor and CLAM with attention-pooling for aggregation. A WSI pipeline with tissue segmentation, 224×224 px patching, and semi-hard patch mining was developed.

Comparison of ABMIL, CLAM, TransMIL, and DSMIL on CAMELYON16+17 showed: DSMIL performed best for binary classification ($AUC = 0.977$, $F1 = 0.986$), CLAM for multi-class staging ($QWK = 0.961$); on TCGA-SKCM all models reached $Recall = 1.0$ for the melanoma class.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ	9
ВСТУП	11
РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ	
АНАЛІЗУ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ	14
1.1 Рак молочної залози як медична проблема	14
1.1.1 Опис задачі діагностики раку молочної залози	15
1.1.2 Основні виклики й проблеми	16
1.2 ІСППР для діагностики раку молочної залози	17
1.2.1 Опис представлення даних (WSI)	18
1.2.2 Огляд існуючих ІСППР для діагностики раку мо- лочної залози	20
1.2.3 Типи розмітки й виклики	21
1.3 Моделі й методи комп'ютерного зору в задачах аналізу медичних зображень	22
1.3.1 Математичні основи машинного навчання	22
1.3.2 Задача класифікації зображень	23
1.4 Сучасні архітектури згорткових мереж	25
1.4.1 ResNet (Residual Networks)	26
1.4.2 EfficientNet	28
1.4.3 Vision Transformer (ViT)	30
1.4.4 Задача навчання без вчителя (Self-Supervised Learning)	33
1.4.5 Задача Multi Instance Learning (MIL)	35

1.5	Патологічні фундаментальні моделі	38
1.6	Постановка задачі	39
РОЗДІЛ 2	АНАЛІЗ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ СТАДІЮВАННЯ	44
2.1	Навчання на множинах екземплярів на основі механізму уваги	44
2.2	Архітектура CLAM	46
2.2.1	Об'єднання на основі уваги	46
2.2.2	Контроль на рівні екземплярів	47
2.2.3	Загальна функція втрат	48
2.2.4	Мультикласова версія	48
2.3	Гістологічні фундаментальні моделі	50
2.3.1	Роль векторних представлень у пайплайні MIL	50
2.3.2	UNI2-H	51
2.3.3	CONCH	53
2.3.4	Virchow2	54
2.3.5	Вибір енкодера	55
2.4	Огляд наборів даних	55
2.4.1	CAMELYON16 та CAMELYON17	55
2.4.2	Melanoma-WSI (CMB-MEL + CPTAC-CM)	56
РОЗДІЛ 3	ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ	59
3.1	Пайплайн попередньої обробки даних	60
3.1.1	Сегментація тканини та нарізка на фрагменти	61
3.1.2	Отримання векторних представлень UNI2-H	62
3.1.3	Відбір інформативних фрагментів	64
3.2	Порівнювані методи MIL	65
3.2.1	ABMIL	66
3.2.2	CLAM	66

3.2.3	TransMIL	67
3.2.4	DSMIL	67
3.3	Налаштування експерименту	68
3.3.1	Гіперпараметри навчання	68
3.3.2	Метрики оцінювання	70
3.4	Результати на CAMELYON16+17	72
3.4.1	Бінарна класифікація	72
3.4.2	Мультикласове стадіювання	74
3.4.3	Порівняльний аналіз методів	78
3.4.4	Стадіювання на рівні пацієнта	79
3.4.5	Порівняння з аналогічними системами на CAMELYON	82
3.5	Архітектура та інфраструктура розгортання вебдодатка ін- телектуальної системи	83
3.6	Результати на Melanoma-WSI	88
3.6.1	Порівняння методів MIL на Melanoma-WSI	88
3.6.2	Якість класифікації та ROC-аналіз	90
3.6.3	Аналіз карт уваги та візуалізація простору ознак	91
3.6.4	Порівняння з аналогічними системами на Melanoma-WSI	94
	ВИСНОВКИ	98
	ПЕРЕЛІК ПОСИЛАНЬ	101
	ДОДАТОК А ЛІСТИНГ ПРОГРАМИ	107

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

ІСППР	– інтелектуальна система підтримки прийняття рішень;
ІІІ	– штучний інтелект;
WSI	– Whole Slide Image (цифрове зображення цілого гістологічного зрізу);
MIL	– Multiple Instance Learning (навчання на основі множинних екземплярів);
CLAM	– Clustering-constrained Attention Multiple Instance Learning (навчання на основі уваги з кластерними обмеженнями);
ROI	– Region of Interest (область інтересу);
CNN	– Convolutional Neural Network (згорткова нейронна мережа);
ViT	– Vision Transformer (трансформер для обробки зображень);
FM	– Foundation Model (базова попередньо навчена модель);
H&E	– Hematoxylin and Eosin (гематоксилін-еозинове забарвлення);
AUC	– Area Under the Curve (площа під ROC-кривою);
ROC	– Receiver Operating Characteristic (робоча характеристика класифікатора);
TNM	– Tumor, Node, Metastasis (система стадіювання злоякісних новоутворень за ознаками пухлини, лімфовузлів та метастазів);
CAMELYON	– Cancer Metastases in Lymph Nodes Challenge (змагання з виявлення метастазів раку в лімфатичних вузлах);
TCGA	– The Cancer Genome Atlas (атлас геному раку);
GPU	– Graphics Processing Unit (графічний процесор);

- RAM** – Random Access Memory (оперативна пам'ять);
- JSON** – JavaScript Object Notation (текстовий формат обміну даними).

ВСТУП

Актуальність теми. Розвиток¹ методів автоматизованого аналізу медичних зображень є одним із пріоритетних напрямів сучасної медичної інформатики та цифрової патології. Серед різноманіття клінічних задач особливе місце займає онкологічна діагностика, де точність і своєчасність виявлення патології безпосередньо визначають виживаність пацієнта.

Рак молочної залози залишається одним із найпоширеніших злоякісних новоутворень серед жіночого населення світу та однією з провідних причин онкологічної смертності. Ефективність лікування суттєво залежить від стадії захворювання на момент встановлення діагнозу, що обумовлює критичну роль точного та відтворюваного стадіювання у клінічній практиці.

Гістопатологічний аналіз є основним методом морфологічної верифікації онкологічних захворювань. Перехід до цифрової патології та широке впровадження Whole Slide Images (WSI) — гігапіксельних цифрових копій гістологічних препаратів — відкривають нові можливості для автоматизації діагностичного процесу. Водночас надвеликий обсяг даних, висока морфологічна гетерогенність тканин та міжекспертна варіативність інтерпретації препаратів формують суттєві виклики для побудови ефективних інтелектуальних систем аналізу.

У цьому контексті методи глибокого навчання та комп'ютерного зору набувають особливого значення. Серед них виділяється парадигма Multiple Instance Learning (MIL) — підхід, орієнтований на сценарії, де детальна локальна розмітка

¹Тут і нижче використано такий інструмент штучного інтелекту як чат-бот з генеративним штучним інтелектом ChatGPT, виключно для корегування та редагування тексту, створеного автором цієї дипломної роботи, на основі автоматизованої перевірки граматики, структури та стилю, що відповідає Політиці використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського (протокол №11 Вченої ради КПІ ім. Ігоря Сікорського від 11 грудня 2023 р.).

даних відсутня або економічно недоцільна. У рамках MIL гістологічний слайд розглядається як «мішок» (bag) із набору патчів (instances), а модель навчається приймати рішення на рівні всього препарату, агрегуючи інформацію з його фрагментів. Це дозволяє працювати в режимі слабкої розмітки (weak supervision), коли доступна лише глобальна slide-level мітка без локальних анотацій.

Сучасні архітектури на основі трансформерів та механізмів уваги дозволяють не лише класифікувати слайди, а й локалізувати найінформативніші ділянки тканини. Така інтерпретованість є принциповою вимогою для клінічного застосування: лікар отримує не лише результат класифікації, а й обґрунтування прийнятого рішення у вигляді теплових карт уваги.

Окремою проблемою залишається варіативність реальних клінічних даних: відмінності у протоколах фарбування між лабораторіями, артефакти сканування та доменного зсуву між різними медичними центрами суттєво ускладнюють узагальнення моделей. Стійкість до шуму та здатність до крос-доменного переносу є, таким чином, не другорядними, а принциповими вимогами до систем автоматизованої патоморфологічної діагностики.

Зазначені чинники визначають актуальність розробки інтелектуальної системи стадіювання раку молочної залози за гістологічними зображеннями цілого зрізу в умовах слабкої розмітки — задачі, що знаходиться на перетині штучного інтелекту, медичної інформатики та цифрової патології.

Об’єкт дослідження — процес автоматизованої гістопатологічної діагностики раку молочної залози на основі цифрових зображень цілого зрізу (Whole Slide Images) в умовах слабкої розмітки даних.

Предмет дослідження — методи та моделі Multiple Instance Learning, архітектури глибокого навчання з механізмами уваги, а також пайплайни обробки WSI, що застосовуються для автоматичного стадіювання раку молочної залози за гістологічними зображеннями.

Мета роботи — розробка та експериментальна перевірка системи автоматичного стадіювання раку молочної залози за гістологічними зображеннями

цілого зрізу (WSI), яка функціонує в умовах слабкої розмітки та спирається на підхід Multiple Instance Learning.

Задля досягнення поставленої мети визначено наступні задачі дослідження.

1. Проаналізувати сучасний стан методів автоматизованого аналізу гістопатологічних зображень, зокрема підходів на основі MIL, архітектур глибокого навчання та attention-механізмів.
2. Вивчити доступні набори гістологічних даних та оцінити їх відповідність вимогам навчання в режимі слабкої розмітки.
3. Запропонувати архітектуру MIL-моделі на основі CLAM, здатну ефективно агрегувати інформацію з окремих патчів WSI.
4. Побудувати повний пайплайн попередньої обробки даних (сегментація, нарізка на патчі, отримання ембедінгів).
5. Навчити модель та провести оптимізацію гіперпараметрів.
6. Порівняти запропонований підхід із базовими методами MIL.
7. Дослідити інтерпретованість моделі за допомогою теплових карт уваги (attention maps).

Результати роботи. У ході дослідження розвинуто архітектуру Clustering-constrained Attention Multiple Instance Learning (CLAM) стосовно задачі багаторівневого стадіювання раку молочної залози за гістологічними знімками повного зрізу (WSI). Адаптований підхід враховує обмежену кількість навчальних прикладів та високу варіативність мікрооточення пухлини, що характерні для реальних клінічних наборів даних. Розроблена система функціонує в парадигмі слабкерованого навчання, що знижує залежність від трудомісткої детальної піксельної розмітки та розширює можливості її застосування в умовах наявної клінічної інфраструктури. Отримане рішення може слугувати інструментом підтримки прийняття рішень для лікарів-патоморфологів та онкологів, забезпечуючи попередній скринінг, пріоритезацію складних випадків і стандартизацію діагностичного процесу, а також сприяючи зменшенню міжекспертної варіативності при оцінці стадії захворювання.

РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ АНАЛІЗУ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ

1.1 Рак молочної залози як медична проблема

Рак молочної залози є одним із найбільш поширених злоякісних онкологічних захворювань серед жіночого населення у світі та становить одну з ключових медичних, соціальних і економічних проблем сучасної системи охорони здоров'я [1; 2]. Висока захворюваність, значний рівень смертності, а також критична залежність прогнозу від своєчасності виявлення патології визначають особливу актуальність розробки ефективних методів ранньої діагностики, прогнозування та стадіювання цього захворювання [1; 3]. Незважаючи на суттєвий прогрес у розвитку скринінгових програм, молекулярної діагностики та персоналізованих підходів до лікування [3], рак молочної залози продовжує залишатися однією з провідних причин онкологічної смертності серед жінок [1; 2; 4].

Клінічна складність захворювання зумовлена його високою біологічною гетерогенністю, варіативністю морфологічних проявів, різноманіттям молекулярних підтипів та складністю раннього виявлення окремих патологічних змін [3; 5]. Прогноз перебігу хвороби значною мірою залежить від стадії, на якій патологію було виявлено, а також від точності визначення локального поширення пухлини, стану регіонарних лімфатичних вузлів та наявності віддалених метастазів [6]. Саме тому задачі діагностики, гістопатологічного аналізу та стадіювання раку молочної залози мають фундаментальне значення для вибору терапевтичної тактики та підвищення виживаності пацієнтів [3; 6].

1.1.1 Опис задачі діагностики раку молочної залози

Діагностика раку молочної залози є багаторівневим комплексним процесом, що охоплює клінічне обстеження, візуалізаційні методи, морфологічний аналіз тканин та остаточну патогістологічну верифікацію [3]. У сучасній клінічній практиці особливе значення мають гістопатологічні дослідження, оскільки саме вони забезпечують найбільш точне підтвердження злоякісності, дозволяють визначити тип пухлини, ступінь диференціації, інвазивність, а також оцінити наявність метастатичного ураження [7; 8].

У межах цифрової патології ключовим джерелом даних виступають цифрові гігапіксельні копії гістологічних препаратів (whole slide images, WSI), отримані шляхом сканування тканинних зрізів з високою роздільною здатністю [8; 9]. Аналіз таких зображень дозволяє досліджувати тканинну архітектуру, клітинну морфологію та просторовий розподіл патологічних структур [7]. Основною задачею діагностичної системи є виявлення патологічно значущих ознак, класифікація зразків за клінічно релевантними категоріями та підтримка визначення стадії захворювання відповідно до міжнародних систем, зокрема TNM [6; 10].

У контексті інтелектуальних систем аналізу гістологічних даних задача ускладнюється надзвичайно великим розміром WSI, високою морфологічною різноманітністю тканин та частою відсутністю детальної локальної розмітки [10; 11]. Це перетворює класичну задачу класифікації зображень на більш складну проблему аналізу на рівні препарату зі слабким контролем, де рішення приймається на основі множини локальних фрагментів тканини [10—12].

1.1.2 Основні виклики й проблеми

Попри значний розвиток сучасних медичних технологій, діагностика раку молочної залози залишається пов'язаною з низкою критичних викликів. Однією з основних проблем є міжекспертна варіативність інтерпретації гістопатологічних препаратів, оскільки навіть висококваліфіковані фахівці можуть демонструвати відмінності в оцінці морфологічних ознак, особливо у складних або межових випадках [7]. Це безпосередньо впливає на точність постановки діагнозу та вибір клінічної стратегії.

Додаткову складність становить надзвичайно великий обсяг даних, що генеруються у форматі WSI. Один цифровий гістологічний зріз може містити гігапиксельне зображення, аналіз якого потребує значних часових витрат та високого рівня концентрації спеціаліста [8]. У таких умовах зростає ризик пропуску дрібних патологічних осередків, зокрема мікрометастазів або мінімальних інвазивних змін [13].

Ще однією суттєвою проблемою є дефіцит детально анотованих медичних даних. Формування високоякісної розмітки на рівні фрагментів або пікселів потребує значних ресурсів патоморфологів, що ускладнює побудову масштабованих систем із повним контролем [14]. Унаслідок цього особливого значення набувають підходи зі слабким контролем та навчання на множинах екземплярів, здатні працювати з мітками на рівні препарату [10; 12].

Також важливими викликами є:

- висока морфологічна гетерогенність пухлинних структур;
- варіативність гістологічного забарвлення між лабораторіями;
- зсув розподілу між різними сканерами та медичними центрами;
- складність інтеграції алгоритмічних рішень у реальні клінічні процеси;
- необхідність інтерпретованості результатів для медичних фахівців.

Таким чином, сучасна задача діагностики раку молочної залози виходить далеко за межі традиційного виявлення патології та потребує створення інтелектуальних систем, здатних поєднувати високу точність, масштабованість, інтерпретованість та адаптивність до умов реальної клінічної практики [7; 10].

1.2 ІСППР для діагностики раку молочної залози

Зростання обсягів медичних даних, складність сучасної онкологічної діагностики та необхідність підвищення точності клінічних рішень зумовили активний розвиток інтелектуальних систем підтримки прийняття рішень (ІСППР) у медицині. У контексті діагностики раку молочної залози ІСППР розглядаються як програмно-аналітичні комплекси, що поєднують алгоритми машинного навчання, комп'ютерного зору та аналізу медичних даних для підтримки лікаря на етапах виявлення патології, морфологічної інтерпретації, стратифікації ризику та стадіювання захворювання [7].

Основною метою ІСППР у цій предметній області є зниження діагностичної невизначеності, підвищення відтворюваності результатів, мінімізація міжекспертної варіативності та оптимізація навантаження на патоморфологів. Такі системи не замінюють клінічного спеціаліста, а виступають як інструмент аналітичної підтримки, здатний автоматизувати рутинні етапи аналізу, виділяти потенційно патологічні ділянки та формувати додаткові кількісні критерії для прийняття клінічних рішень [8; 15].

У задачах гістопатологічної діагностики особливу цінність ІСППР мають у сценаріях аналізу великих цифрових гістологічних препаратів, де обсяг даних суттєво перевищує можливості швидкого ручного аналізу. Саме тому сучасні ІСППР дедалі частіше інтегрують методи слабокерованого навчання, навчання

на множинах екземплярів та фундаментальні моделі для ефективної роботи з WSI [10].

1.2.1 Опис представлення даних (WSI)

Ключовим форматом даних у сучасному цифровому гістопатологічному аналізі є WSI – цифрове зображення цілого гістологічного зрізу надвисокої роздільної здатності, отримане шляхом сканування тканинного препарату [9]. WSI є гігапіксельними зображеннями, що можуть містити мільярди пікселів та відображають повну морфологічну структуру досліджуваного зразка.

На відміну від традиційних медичних зображень фіксованого розміру, WSI мають багаторівневу пірамідальну структуру масштабів (рис. 1.1), що дозволяє аналізувати тканину як на макрорівні архітектури, так і на мікрорівні клітинних деталей [16]. Такий формат забезпечує максимально повне цифрове представлення гістологічного препарату, однак створює суттєві обчислювальні труднощі через надвеликий обсяг даних [7].

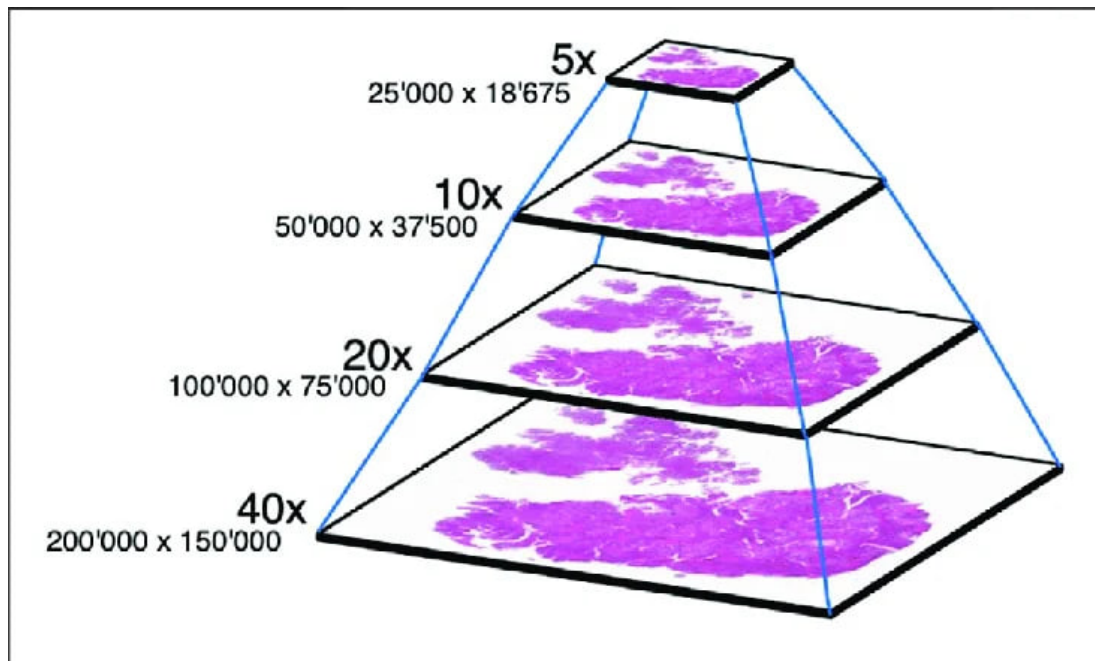


Рисунок 1.1 – Пірамідальна багаторівнева структура WSI: від низькороздільного загального огляду до повного рівня клітинної деталізації. Основа піраміди відповідає максимальній роздільній здатності [16]

Для практичного використання в системах машинного навчання WSI зазвичай розбиваються на окремі фрагменти фіксованого розміру, які виступають базовими одиницями аналізу. Кожен фрагмент може містити локальні морфологічні ознаки тканини, тоді як сукупність фрагментів формує представлення всього препарату на рівні слайду. Такий підхід лежить в основі більшості сучасних пайплайнів навчання на множинах екземплярів [10—12].

Основними характеристиками WSI як формату даних є:

- надвелика просторова роздільна здатність;
- багатомасштабна пірамідальна структура;
- висока морфологічна складність та гетерогенність тканин;
- значна частка фонових або діагностично нерелевантних ділянок;
- великий обсяг даних, що ускладнює зберігання та обробку [7].

Таким чином, WSI є не лише джерелом високодеталізованої морфологічної інформації, але й формують специфічні виклики для побудови ефективних інтелектуальних систем підтримки прийняття рішень [7; 10].

1.2.2 Огляд існуючих ІСППР для діагностики раку молочної залози

Сучасні інтелектуальні системи для діагностики раку молочної залози охоплюють широкий спектр підходів – від класичних систем автоматизованого виявлення до складних архітектур глибокого навчання, орієнтованих на аналіз гістологічних WSI [7]. Ранні системи переважно фокусувалися на автоматизованому виявленні підозрілих структур у мамографії або локальній класифікації окремих фрагментів, однак сучасний розвиток обчислювальної патології змістив акцент у бік аналізу на рівні препарату.

Значний внесок у розвиток галузі зробили міжнародні конкурсні платформи, зокрема CAMELYON16 [17] та CAMELYON17 [13], які стимулювали створення систем автоматичного виявлення метастазів у лімфатичних вузлах. На їх основі були розроблені численні моделі, що використовують згорткові нейронні мережі, навчання на множинах екземплярів з механізмом уваги, CLAM [10] та трансформерні архітектури [18].

Серед сучасних напрямів розвитку ІСППР можна виділити:

- системи виявлення метастатичних осередків;
- класифікацію на рівні слайду;
- слабокероване стадіювання;
- пайплайни векторних представлень на основі фундаментальних моделей;
- інтерпретований ШІ для гістопатології.

Попри значний прогрес, більшість існуючих систем орієнтовані переважно на виявлення або бінарну класифікацію, тоді як повноцінне автоматизоване стадіювання раку молочної залози залишається значно менш дослідженим напрямом.

1.2.3 Типи розмітки й виклики

Якість та гранулярність розмітки медичних даних є одним із ключових факторів, що визначають можливості побудови інтелектуальних діагностичних систем. У задачах аналізу гістологічних WSI можна виділити кілька основних типів розмітки:

- pixel-level (піксельна сегментація);
- region-level (локалізація патологічних областей);
- patch-level (класифікація окремих фрагментів);
- slide-level (мітка на рівні всього препарату).

Найбільш інформативною є детальна pixel-level розмітка, однак її отримання потребує надзвичайно великих часових витрат експертів. У реальних клінічних умовах значно частіше доступною є slide-level розмітка, де відома лише загальна діагностична категорія препарату без локалізації патології.

Саме така обмеженість даних формує ключову проблему weak supervision: система повинна навчатися виявляти релевантні патологічні патерни без прямої локальної анотації. Це ускладнюється тим, що:

- патологічно значущі патчі можуть становити лише малу частину WSI;
- присутній високий рівень шуму у мітках на рівні набору екземплярів;
- існує зсув розподілу між різними лабораторіями;
- морфологічна гетерогенність ускладнює узагальнення;
- висока вартість експертної розмітки обмежує масштабування supervised learning.

У зв'язку з цим сучасні ІСППР дедалі більше орієнтуються на weakly supervised paradigms, self-supervised pretraining та Multiple Instance Learning як найбільш практичні підходи для роботи з реальними клінічними гістопатологічними даними.

1.3 Моделі й методи комп'ютерного зору в задачах аналізу медичних зображень

1.3.1 Математичні основи машинного навчання

Математичний апарат машинного навчання базується на концепціях теорії ймовірностей, лінійної алгебри та математичного аналізу [19]. В задачах медичної діагностики навчання моделі розглядається як апроксимація невідомої функції відображення $f : \mathcal{X} \rightarrow \mathcal{Y}$, де $\mathcal{X}$ — простір високовимірних ознак зображення, а $\mathcal{Y}$ — простір міток (наприклад, тип тканини або стадія патології).

Процес оптимізації параметрів моделі θ полягає у мінімізації емпіричного ризику:

$$\arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(f(x_i; \theta), y_i) + \lambda \mathcal{R}(\theta),$$

де $\mathcal{L}$ — функція втрат (наприклад, бінарна крос-ентропія), а $\mathcal{R}(\theta)$ — регуляризуючий член (L1/L2), що запобігає перенавчанню моделі на специфічних артефактах вибірки. Для запобігання перенавчанню додатково застосовуються стратегії відсіювання вузлів [20] та крос-валідація для надійної оцінки узагальнювальної здатності на обмежених медичних вибірках.

1.3.2 Задача класифікації зображень

Класифікація медичних зображень є фундаментальним завданням комп'ютерного зору, що полягає у присвоєнні вхідному зображенню певної категорії з дискретної множини класів [21]. Формально задача визначається як побудова функції:

$$f_{\theta} : \mathcal{X} \rightarrow \mathcal{Y},$$

де $\mathcal{X}$ – простір вхідних зображень, $\mathcal{Y} = \{1, \dots, C\}$ – множина класів, а θ – параметри моделі, що оптимізуються на навчальній вибірці шляхом мінімізації функції втрат (як правило, крос-ентропії). Починаючи з появи глибоких згорткових мереж, цей підхід став домінуючим у задачах аналізу медичних зображень [21; 22].

У задачах аналізу whole slide images (WSI) пряме застосування класичних підходів до класифікації є неможливим через надвисоку роздільну здатність даних. Типове гістологічне зображення може містити понад $100\,000 \times 100\,000$ пікселів, що унеможлиблює його обробку як єдиного тензора в пам'яті [9; 14].

Як показано на рисунку 1.2, це зумовлює необхідність декомпозиції зображення на локальні фрагменти (патчі) з подальшим незалежним аналізом кожного з них перед агрегацією на рівні всього препарату.

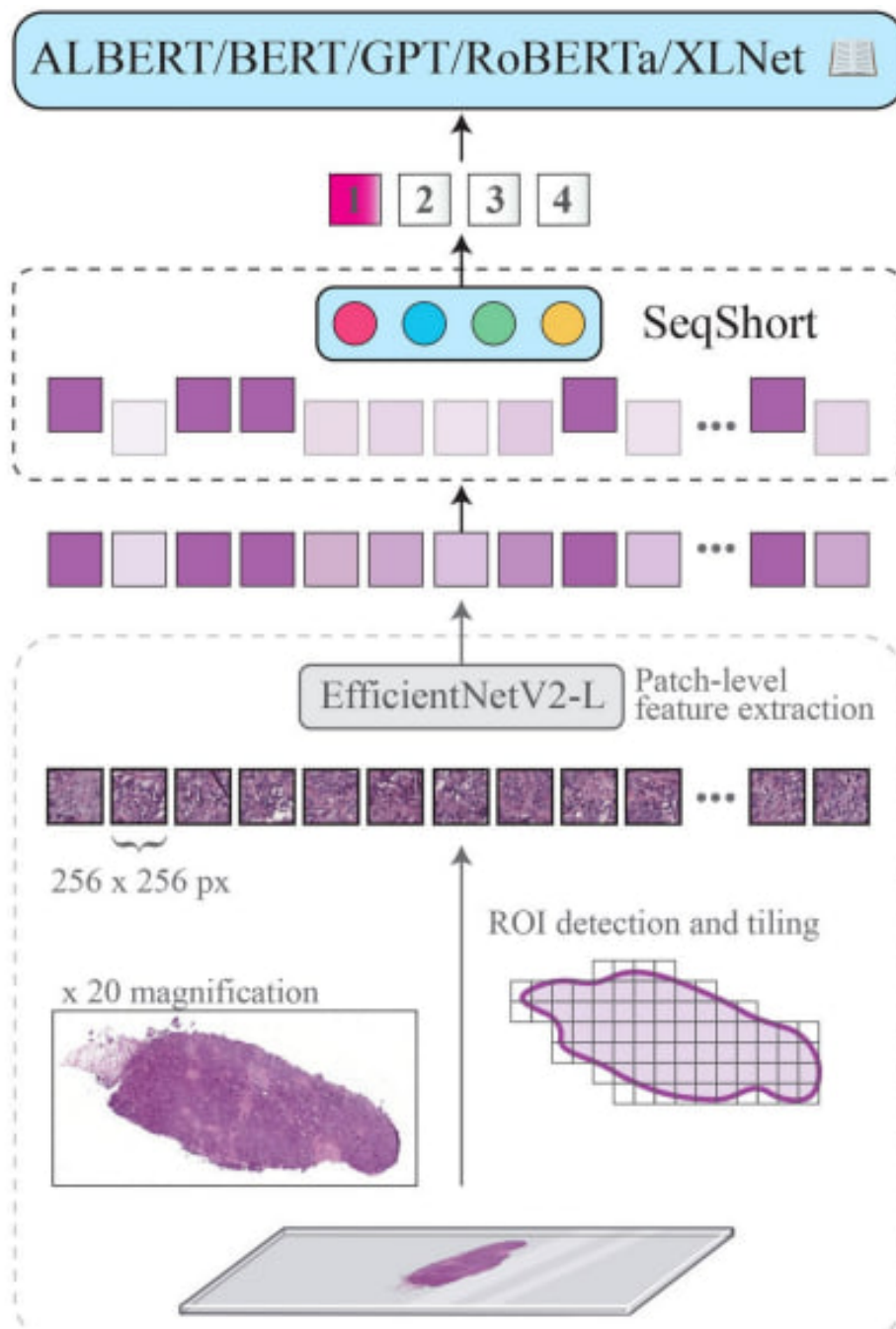


Рисунок 1.2 – Схематичне представлення процесу аналізу whole slide image: поділ зображення на локальні фрагменти (патчі) та подальша класифікація на рівні препарату. Адаптовано з [23]

Сучасні підходи вирішують цю проблему через ієрархічну двохрівневу структуру аналізу [7; 10].

1. Рівень екземпляра – визначення морфологічних ознак локальних ділянок тканини. Кожен фрагмент x_{ij} розміром, як правило, 256×256 або 512×512

пікселів, обробляється кодувальником незалежно, утворюючи вектор ознак $h_{ij} \in \mathbb{R}^d$. Навчання на цьому рівні може вимагати детальної розмітки або слабкого перенесення міток з рівня препарату [8; 14].

2. Рівень препарату – формування остаточного діагнозу шляхом агрегації множини ознак $\{h_{ij}\}$ в єдине представлення $z_i \in \mathbb{R}^d$ з подальшою класифікацією [10; 12]. Саме цей рівень відповідає клінічній постановці задачі, оскільки мітки доступні лише для всього препарату.

Ключова складність полягає в семантичній невідповідності між рівнями: глобальна мітка препарату не відображає локальний розподіл патологічних змін, які можуть бути присутні лише в невеликій частині зображення [10]. Наприклад, при метастатичному ураженні лімфатичних вузлів пухлинна тканина може займати менше 10% площі всього препарату [13].

Це робить задачу класифікації WSI принципово відмінною від стандартної класифікації зображень і вимагає спеціалізованих підходів, зокрема навчання на множинах екземплярів (multiple instance learning), для коректного зв'язку між локальними морфологічними ознаками та глобальним діагнозом [11; 12].

1.4 Сучасні архітектури згорткових мереж

Еволюція архітектур глибокого навчання стала одним із ключових чинників трансформації сучасного комп'ютерного зору, зокрема в задачах аналізу медичних зображень. Перехід від класичних підходів, заснованих на ручному конструюванні ознак і використанні наперед визначених дескрипторів текстури, форми чи кольору, до багаторівневих нейронних мереж дав змогу автоматизувати процес вилучення релевантних представлень безпосередньо з даних [21; 24]. Якщо ранні системи комп'ютерного зору значною мірою залежали від якості

експертно спроектованих фільтрів, то сучасні глибокі архітектури здатні ієрархічно формувати ознаки різного рівня абстракції – від базових контурів і текстур до складних морфологічних патернів тканин [24; 25]. Це забезпечило суттєве підвищення точності, масштабованості та узагальнювальної здатності моделей, зробивши глибоке навчання фундаментальною основою більшості сучасних систем інтелектуального аналізу зображень [21].

1.4.1 ResNet (Residual Networks)

Однією з ключових проблем, що виникали при навчанні глибоких нейронних мереж до появи ResNet, була деградація точності зі збільшенням кількості шарів. Парадоксально, але додавання нових шарів не лише не покращувало результат, а часто призводило до його погіршення навіть на тренувальній вибірці. Причиною цього є проблема зникаючих градієнтів: під час зворотного поширення похибки (backpropagation) градієнт послідовно множиться через велику кількість шарів, що призводить до його експоненційного затухання:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial \mathcal{L}}{\partial x_L} \prod_{l=1}^{L-1} \frac{\partial x_{l+1}}{\partial x_l}$$

При малих значеннях похідних градієнт практично зникає до моменту досягнення перших шарів мережі, що унеможлиблює їх ефективне навчання.

В Resnet було використано концепцію залишкового навчання (residual learning) як рішення цієї проблеми [22]. Замість того, щоб блок навчався прямому відображенню $H(x)$, він навчається залишковій функції:

$$F(x) = H(x) - x,$$

звідки фінальний вихід блоку:

$$H(x) = F(x) + x.$$

Практично це реалізується через пряме з'єднання (shortcut connection), що додає вхід блоку до його виходу, обходячи проміжні шари (рис. 1.3):

$$x_{l+1} = F(x_l, W_l) + x_l.$$

Ключова перевага такої конструкції полягає в тому, що градієнт тепер може поширюватись через пряме з'єднання безпосередньо, не проходячи через нелінійності блоку:

$$\frac{\partial \mathcal{L}}{\partial x_l} = \frac{\partial \mathcal{L}}{\partial x_{l+1}} \left(\frac{\partial F}{\partial x_l} + 1 \right),$$

де доданок $+1$ суттєво зменшує ризик повного згасання градієнта навіть у дуже глибоких мережах. Завдяки цьому стало можливим стабільне навчання архітектур ResNet-50, ResNet-101 та ResNet-152 без деградації точності.

У контексті цифрової патології ResNet тривалий час залишався стандартним енкодером для отримання ознак на рівні фрагментів у MIL-пайплайнах. Попри те, що сучасні фундаментальні моделі на основі Vision Transformer суттєво перевершують ResNet за якістю векторних представлень, архітектурний принцип залишкових зв'язків (residual connections) залишається фундаментальним і присутній у більшості сучасних глибоких нейронних мереж.

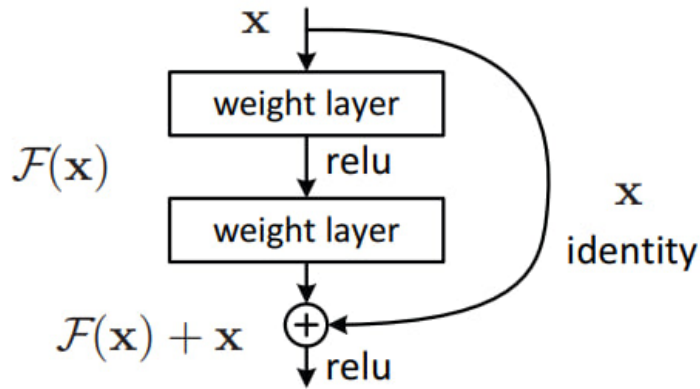


Рисунок 1.3 – Схема залишкового блоку ResNet [22]

1.4.2 *EfficientNet*

Попри успіх ResNet та подібних архітектур, питання ефективного масштабування нейронних мереж залишалось відкритим. Традиційно продуктивність моделей покращували трьома незалежними способами: збільшенням глибини мережі (кількості шарів), збільшенням її ширини (кількості каналів) або підвищенням роздільної здатності вхідного зображення. Однак кожен із цих підходів окремо швидко виходив на плато точності при суттєвому зростанні обчислювальних витрат.

Метод узгодженого масштабування [26] передбачає одночасне масштабування глибини, ширини та роздільної здатності за допомогою єдиного коефіцієнта φ :

$$\text{depth} = \alpha^\varphi, \quad \text{width} = \beta^\varphi, \quad \text{resolution} = \gamma^\varphi,$$

за умови:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \quad \alpha \geq 1, \beta \geq 1, \gamma \geq 1,$$

де α , β , γ — емпірично знайдені константи, а φ — глобальний коефіцієнт масштабування що визначає доступний обчислювальний бюджет.

Базова архітектура EfficientNet-B0 була отримана за допомогою пошуку архітектури нейронної мережі (neural architecture search, NAS), а подальші варіанти B1–B7 — шляхом застосування компаундного масштабування (compound scaling) з різними значеннями φ (рис. 1.4). Це дозволило отримати сімейство моделей різного розміру, де кожна наступна версія є систематично масштабованою, а не довільно модифікованою.

Результати виявились переконливими: EfficientNet-B7 при значно меншій кількості параметрів порівняно з ResNet-152 або DenseNet-201 демонстрував вищу точність на ImageNet. Це зробило архітектуру популярним вибором у задачах де важливий баланс між точністю та обчислювальною ефективністю.

У цифровій патології EfficientNet застосовувався як енкодер для отримання патч-рівневих ознак у MIL-пайплайнах до появи спеціалізованих патологічних фундаментальних моделей. Його компактність та висока точність робили його особливо привабливим для сценаріїв з обмеженими обчислювальними ресурсами.

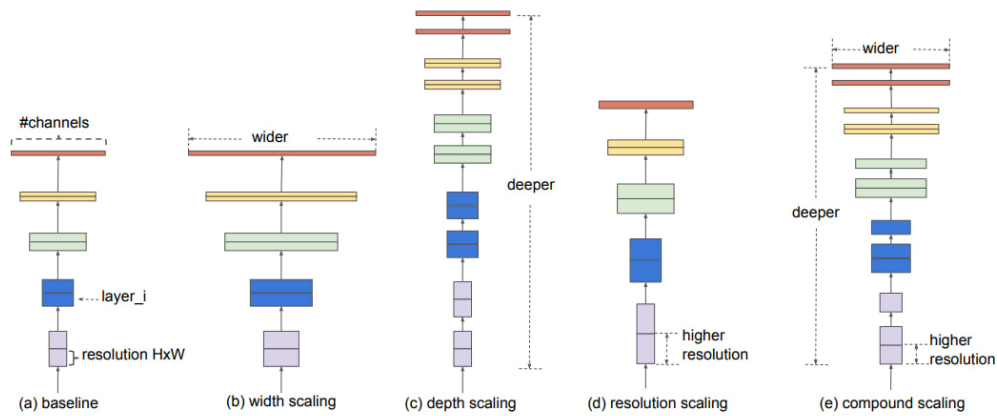


Рисунок 1.4 – Порівняння підходів до масштабування нейронних мереж: збільшення ширини, глибини, роздільної здатності та compound scaling [26]

1.4.3 Vision Transformer (ViT)

Попри високу ефективність згорткових архітектур, їх фундаментальне обмеження полягає в локальності рецептивного поля — кожен згортковий фільтр аналізує лише невелику ділянку зображення. Глобальний контекст формується поступово через стекінг шарів, що є непрямим і потенційно неефективним механізмом для задач де важливі далекосяжні просторові залежності. У цифровій патології це особливо актуально: патоморфолог при аналізі тканини враховує не лише локальну клітинну морфологію, але й ширший тканинний контекст.

У роботі [27] запропоновано Vision Transformer (ViT) – адаптацію архітектури трансформерів до задач комп'ютерного зору. Ключова ідея полягає в тому, щоб розглядати зображення не як суцільну матрицю пікселів, а як послідовність патчів-токенів, аналогічно до слів у текстовій послідовності.

Формально вхідне зображення $x \in \mathbb{R}^{H \times W \times C}$ розбивається на N патчів розміру $P \times P$:

$$N = \frac{H \cdot W}{P^2}.$$

Кожен патч лінеаризується та проєктується у вектор фіксованої розмірності D через навчану матрицю E :

$$z_0 = [x_{\text{class}}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{\text{pos}},$$

де x_{class} — спеціальний CLS-токен що агрегує глобальну інформацію для класифікації, а E_{pos} — positional encoding що зберігає просторову інформацію про розташування патчів.

Центральним компонентом ViT є механізм самоуваги (self-attention). Для кожного токена обчислюються три вектори — запит (Query), ключ (Key) та значення (Value):

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V,$$

де W^Q, W^K, W^V — навчані матриці проєкцій.

Self-attention визначається як:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V,$$

де d_k — розмірність простору ключів, а масштабування через $\sqrt{d_k}$ стабілізує градієнти та запобігає насиченню softmax.

Матриця $QK^T/\sqrt{d_k}$ відображає ступінь релевантності між усіма парами tokenів одночасно. Таким чином кожен патч може напряму взаємодіяти з будь-яким іншим патчем незалежно від просторової відстані між ними — на відміну від

CNN де глобальний контекст формується лише через стекінг багатьох локальних операцій.

Для підвищення репрезентаційної здатності використовується Multi-Head Self-Attention (MHSA) де кілька незалежних голів з механізмом уваги працюють паралельно:

$$\text{MHSA}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O,$$

де $\text{head}_i = \text{Attention}(Q_i, K_i, V_i)$. Кожна голова може фокусуватись на різних аспектах структури зображення — локальній клітинній морфології, глобальній тканинній архітектурі або текстурних паттернах.

Стандартний блок трансформерного кодувальника складається з багатоголової самоуваги (MHSA) та мережі прямого поширення із залишковими з'єднаннями та нормалізацією шарів:

$$z'_l = \text{MHSA}(\text{LN}(z_{l-1})) + z_{l-1},$$

$$z_l = \text{MLP}(\text{LN}(z'_l)) + z'_l.$$

Залишкові зв'язки тут відіграють аналогічну до ResNet роль — стабілізують проходження градієнта та полегшують оптимізацію глибоких моделей.

Принципова відмінність ViT від CNN полягає в тому, що індуктивні упередження (inductive biases) — локальність та трансляційна інваріантність — у CNN закладені архітектурно, тоді як ViT має значно слабшу структурну апріорність. Це означає, що ViT потребує значно більше даних для навчання з нуля. Саме тому прорив ViT став можливим переважно після масштабного попереднього навчання (pretraining) на великих наборах даних, таких як ImageNet-21k.

У цифровій патології трансформерні підходи виявились особливо перспективними через природну сумісність із структурою, що базується на патчах WSI.

Патчі слайду можуть розглядатись як аналог токенів, а механізм уваги (attention) дозволяє моделювати взаємозв'язки між віддаленими ділянками тканини. Це стало основою для появи спеціалізованих патологічних фундаментальних моделей, таких як UNI2-H, що базуються саме на ViT-архітектурі (рис. 1.5), натренованій через self-supervised підхід на мільйонах патологічних патчів.

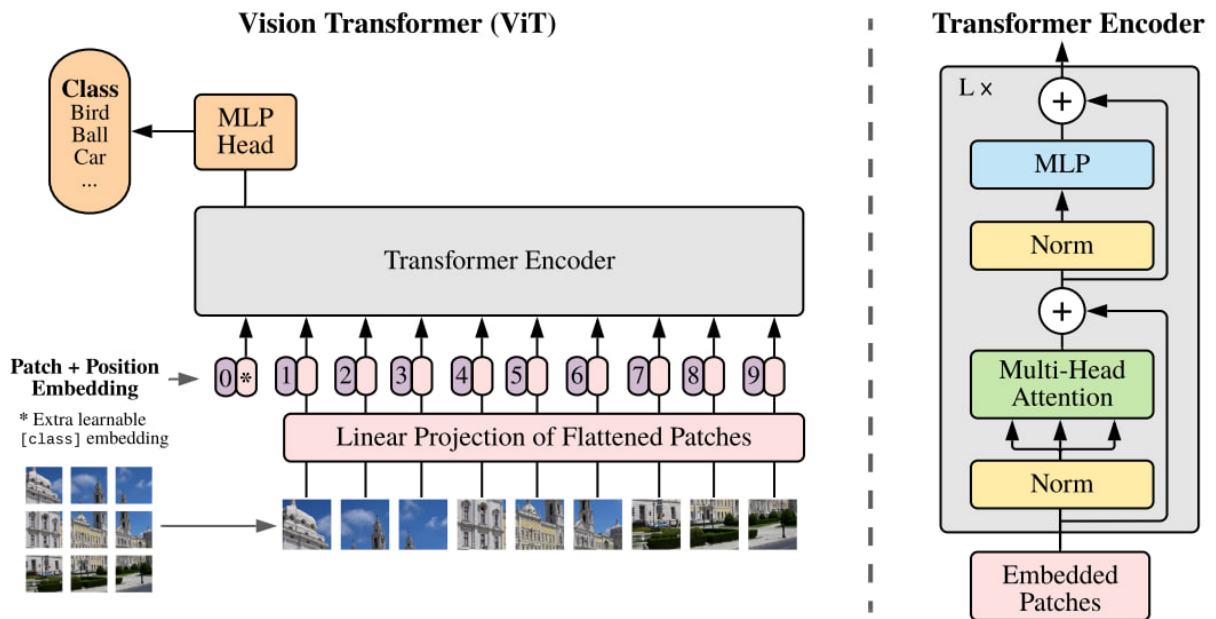


Рисунок 1.5 – Архітектура Vision Transformer: розбиття зображення на патчі-токени та обробка transformer encoder [27]

1.4.4 Задача навчання без вчителя (Self-Supervised Learning)

Однією з ключових проблем у цифровій патології є дефіцит розмічених даних. Отримання якісних анотацій потребує залучення досвідчених патоморфологів, що є дорогим і повільним процесом. Self-supervised learning (SSL) пропонує альтернативу: навчати модель на нерозмічених даних, формулюючи допоміжні

задачі (pretext tasks), розв’язання яких змушує мережу вивчати корисні представлення.

Головна ідея SSL полягає в тому, що модель отримує власний навчальний сигнал безпосередньо з даних — без будь-яких міток від людини. Після такого попереднього навчання отримані векторні представлення переносяться на цільові задачі (класифікація, сегментація, MIL) і, як правило, демонструють значно кращі результати порівняно з випадковою ініціалізацією або навіть з моделями, попередньо навченими на ImageNet, у вузькоспеціалізованих предметних областях.

Серед найбільш впливових підходів варто виділити три напрями.

SimCLR [28] базується на контрастивному навчанні. Для кожного зображення генеруються два аугментовані варіанти — два різних ”погляди” на один об’єкт. Модель навчається максимізувати схожість між представленнями одного зображення та мінімізувати схожість між представленнями різних зображень:

$$\mathcal{L} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)},$$

де $\text{sim}(u, v) = u^T v / \|u\| \|v\|$ — косинусна подібність, τ — температурний параметр, а N — розмір батчу.

DINO [29] використовує підхід самодистиляції (self-distillation): дві копії моделі — student та teacher отримують різні аугментації одного зображення. Teacher оновлюється не за допомогою градієнтного спуску, а через експоненційне ковзне середнє wag student. Модель навчається узгоджувати виходи student з виходами teacher (рис. 1.6):

$$\mathcal{L} = H(P_t(x), P_s(x)) = -P_t(x) \log P_s(x),$$

де P_s та P_t — розподіли ймовірностей student і teacher відповідно. DINO особливо ефективний для ViT-архітектур і демонструє, що карти уваги таких моделей без

жодної розмітки виявляють семантично значущі ділянки зображення. Саме на основі DINOv2 навчена фундаментальна модель UNI2-H, що використовується в цій роботі [30].

MAE (Masked Autoencoder) [31] застосовує інший принцип: частина патчів зображення випадково маскується (зазвичай 75%), і модель навчається відновлювати замасковані ділянки за контекстом видимих патчів. Це змушує encoder формувати глибоке розуміння просторової структури та семантики зображення.

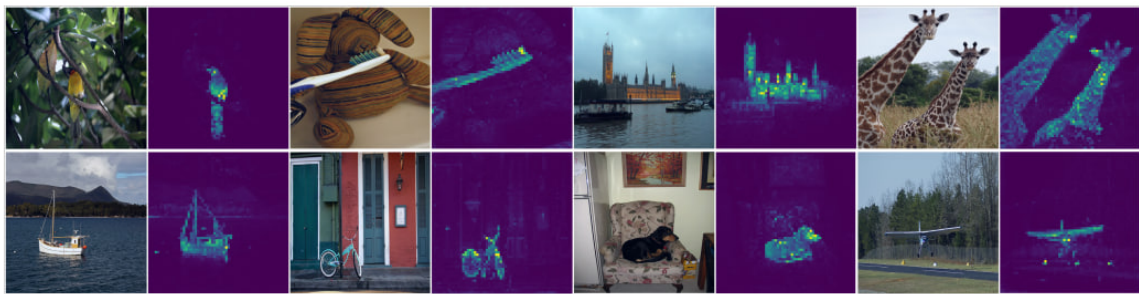


Рисунок 1.6 – Схема self-supervised навчання DINO: student та teacher отримують різні аугментації одного зображення [29]

У контексті цифрової патології SSL відіграє особливо важливу роль, оскільки дозволяє використовувати величезні архіви неанотованих гістологічних слайдів для попередньо натренованих потужних енкодерів. Саме це стало основою для розробки патологічних фундаментальних моделей, що розглядаються далі.

1.4.5 Задача Multi Instance Learning (MIL)

Класичне supervised learning передбачає що кожен об'єкт вибірки має власну мітку. У задачах аналізу WSI це означало б необхідність розмічати кожен патч окремо — що є практично нереальним через обсяги даних та вартість експертної

анотації. Multiple Instance Learning (MIL) вирішує цю проблему, формулюючи задачу навчання на рівні колекцій об'єктів, а не окремих екземплярів [12].

У класичній MIL постановці окремий об'єкт вибірки — це bag, тобто набір instance-елементів:

$$B_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}, \quad (1.1)$$

де B_i — i -й bag (whole slide image), x_{ij} — окремий екземпляр (патч або його ембедінг), n_i — кількість екземплярів у наборі.

Мітка відома лише для всього набору:

$$Y_i \in \{0, 1\},$$

тоді як індивідуальні мітки y_{ij} для екземпляру залишаються невідомими. У базовій бінарній постановці:

$$Y_i = \begin{cases} 1, & \exists j : y_{ij} = 1 \\ 0, & \forall j : y_{ij} = 0 \end{cases}$$

Тобто набір вважається позитивним, якщо хоча б один екземпляр є позитивним. Для гістопатології це добре узгоджується з практикою: слайд класифікується як метастатичний навіть якщо пухлинна тканина займає лише невелику його частину.

Архітектурно MIL складається з двох компонентів. По-перше, екстрактор ознак на рівні екземплярів перетворює кожен фрагмент на вектор ознак: $h_{ij} = f_\theta(x_{ij})$, де f_θ — енкодер (CNN або ViT). По-друге, функція агрегації об'єднує представлення екземплярів у ембедінг на рівні препарату: $z_i = g(\{h_{i1}, \dots, h_{in_i}\})$.

Найпростіші варіанти агрегації — max-pooling та mean-pooling — мають суттєві обмеження. Max-pooling ігнорує контекст більшості тканини, тоді як mean-pooling розмиває рідкісні але критично важливі патологічні ознаки. Саме тому сучасним стандартом є MIL на базі механізму уваги [12], де кожному екземпляру призначається вага уваги:

$$z = \sum_{k=1}^K a_k h_k, \quad \sum_{k=1}^K a_k = 1$$

Агрегація ознак екземплярів виконується згідно з рівнянням (1.4.5).

$$a_k = \frac{\exp\{w^T \tanh(V h_k^T)\}}{\sum_{j=1}^K \exp\{w^T \tanh(V h_j^T)\}},$$

де V та w — параметри моделі. Таким чином модель навчається виділяти найбільш діагностично релевантні патчі без прямої patch-level розмітки, що забезпечує також інтерпретованість через теплові карти уваги (рис. 1.7).

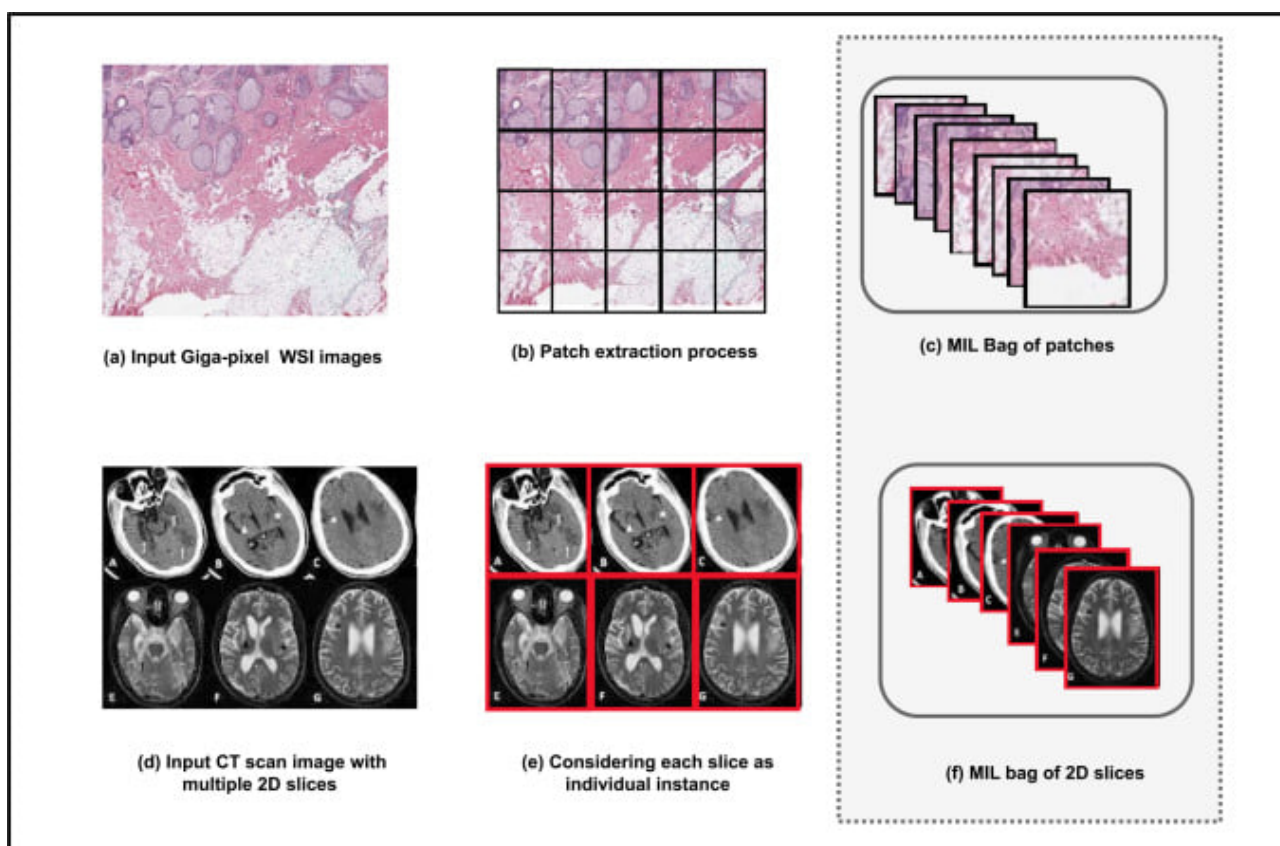


Рисунок 1.7 – Основні підходи Multiple Instance Learning [11]

1.5 Патологічні фундаментальні моделі

Традиційний підхід до побудови моделей у цифровій патології передбачав навчання окремої системи під кожну конкретну задачу [8]. Такі моделі, як правило, добре працювали в межах тренувального домену, проте погано переносились на дані з інших інституцій, протоколів фарбування або типів сканерів [7]. Фундаментальні моделі запропонували принципово інший підхід: спочатку навчити велику модель на масштабному корпусі неанотованих гістологічних даних методами самонавчання, а потім використовувати отримані представлення як універсальну основу для широкого спектру цільових задач [30; 32].

Ключова ідея полягає в тому, що модель, навчена на мільйонах фрагментів із тисяч препаратів різних типів тканин без жодних міток, формує векторні представлення, які кодують морфологічну інформацію на значно глибшому рівні, ніж моделі, навчені на загальних фотографічних наборах даних на кшталт ImageNet [30]. Наприклад, модель UNI навчена на понад 100 мільйонах зображень із більш ніж 100 000 гістологічних препаратів двадцяти типів тканин і перевершує попередні підходи на широкому спектрі задач обчислювальної патології [30]. Такі представлення демонструють кращу узагальнюваність між різними лабораторіями, сканерами та типами патологій [32; 33].

У контексті слабкерованого аналізу WSI фундаментальні моделі відіграють роль замороженого кодувальника: кожен фрагмент препарату перетворюється на вектор ознак, який далі використовується як вхід для MIL-агрегатора [10]. Це дозволяє відокремити задачу вилучення морфологічних ознак від задачі класифікації на рівні препарату, що є особливо важливим при обмеженій кількості розмічених препаратів [11].

1.6 Постановка задачі

У межах цієї роботи розглядається задача автоматизованого стадіювання раку молочної залози за цифровими гістологічними зображеннями цілого зрізу (Whole Slide Images, WSI) в умовах слабкої розмітки даних. На відміну від класичних задач контрольованого навчання, де доступна детальна локальна анотація кожного патологічного регіону, у практичних клінічних умовах здебільшого наявна лише глобальна діагностична інформація на рівні всього препарату або випадку пацієнта. Це зумовлює необхідність побудови моделей, здатних ефективно працювати за відсутності patch-level або pixel-level розмітки.

Кожен цифровий гістологічний зріз являє собою надвелике зображення гігапіксельного масштабу, яке неможливо безпосередньо використовувати як вхід до сучасних моделей глибокого навчання через обмеження пам'яті та обчислювальних ресурсів. Тому WSI подається у вигляді множини локальних фрагментів, отриманих після попередньої сегментації тканини та нарізки зображення (див. (1.1)), де B_i – i -й набір (WSI), x_{ij} – окремий екземпляр (фрагмент або його векторне представлення), n_i – кількість фрагментів у межах конкретного препарату.

Для кожного WSI відома лише глобальна мітка:

$$y_i \in Y,$$

де Y – множина можливих класів або стадій захворювання (наприклад, відсутність пухлини, рання стадія, прогресуюча форма, метастатичне ураження тощо).

Таким чином, основна задача полягає у побудові функції:

$$F : B_i \rightarrow y_i,$$

яка на основі множини екземплярних представлень забезпечує коректне прогнозування глобального класу для всього гістологічного зрізу.

Особливістю цієї задачі є те, що не всі фрагменти всередині одного WSI є однаково інформативними. Значна частина фрагментів може містити нормальну тканину, фон, артефакти сканування або морфологічно незначущі структури. Водночас лише обмежена підмножина екземплярів може містити критично важливі ознаки пухлинного процесу. Отже, модель повинна не лише агрегувати інформацію з великої кількості фрагментів, але й навчитися визначати відносну важливість кожного екземпляра:

$$a_{ij} = \text{Attention}(x_{ij}),$$

де a_{ij} – вага уваги, що відображає внесок конкретного фрагмента у фінальне рішення. З урахуванням цього агреговане представлення WSI визначається згідно з виразом (1.4.5), де h_{ij} – вектор ознак окремого фрагмента після екстрактора ознак, а z_i – глобальне представлення препарату, що використовується для остаточної класифікації.

Кінцевою клінічною метою системи є стадіювання на рівні пацієнта відповідно до класифікації pN системи TNM. Оскільки один пацієнт може мати кілька досліджених лімфатичних вузлів, передбачення стадії на рівні препарату агрегуються до рівня пацієнта через детерміновані правила TNM, що формують другий рівень ієрархічного рішення.

До ключових ускладнень цієї задачі належать такі виклики.

1. Надвеликий розмір WSI та висока обчислювальна складність обробки.
2. Відсутність детальної локальної розмітки.
3. Морфологічна неоднорідність пухлинного мікрооточення.
4. Міжлабораторна варіативність фарбування та сканування.
5. Дисбаланс класів.
6. Необхідність інтерпретованості результатів для клінічного використання.

У зв'язку з переліченими обмеженнями традиційні підходи з повним контролем, що потребують детальної експертної розмітки, є маломасштабованими та економічно затратними. Саме тому доцільним є використання Multiple Instance Learning як парадигми слабкерованого навчання, де навчання відбувається на рівні наборів із глобальними мітками. Отже, розроблювана інтелектуальна система автоматичного стадіювання раку молочної залози виконує такі функції.

1. Виконує попередню обробку WSI (сегментацію тканини, нарізку на патчі, формування embedding-простору).
2. Реалізує MIL-архітектуру на основі attention-механізму та CLAM-підходу.
3. Забезпечує класифікацію WSI за стадіями захворювання в умовах weak supervision.

4. Генерує інтерпретовані attention maps для локалізації найбільш діагностично значущих регіонів.
5. Демонструє стійкість до шуму, варіативності гістологічних даних та обмежень реальної клінічної практики.

Таким чином, ця задача охоплює не лише проблему класифікації великих гістологічних зображень, але й питання масштабованості, інтерпретованості та практичної придатності системи підтримки прийняття рішень у цифровій патології.

Висновки до розділу 1

У першому розділі систематизовано теоретичне підґрунтя задачі автоматизованого стадіювання раку молочної залози за WSI. Показано, що перехід до цифрової патології формує нові можливості для автоматизації, однак надвеликий розмір зображень та дефіцит детальних анотацій визначають weak supervision як єдиний практично масштабований підхід у реальних клінічних умовах.

Показано що слабка розмітка є типовим сценарієм реальної клінічної практики — більшість датасетів містять лише slide-level мітки з медичних карт пацієнтів. Це зумовлює необхідність використання підходів слабокерованого навчання, серед яких Multiple Instance Learning є найбільш практичним та ефективним для аналізу WSI.

Проведено огляд архітектур глибокого навчання — від CNN (ResNet, EfficientNet) до Vision Transformer — та методів самонавчання (SimCLR, DINO, MAE), які стали основою для розробки сучасних патологічних фундаментальних моделей. Встановлено що UNI2-H, навчена на понад 100 тисячах слайдів через

DINOv2, забезпечує найбільш інформативні векторні представлення для цільових задач MIL.

Сформульовано постановку задачі: побудова функції відображення набору патч-ембедінгів у передбачення стадії на рівні слайду в умовах відсутності локальних анотацій. Визначено ключові ускладнення — надвеликий розмір WSI, морфологічна гетерогенність, зсув розподілу та дисбаланс класів — які обґрунтовують вибір MIL-архітектури на основі CLAM з механізмом уваги як основного підходу цього дослідження.

РОЗДІЛ 2 АНАЛІЗ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ СТАДІЮВАННЯ

2.1 Навчання на множинах екземплярів на основі механізму уваги

Класичні підходи до агрегації в MIL – max-pooling та mean-pooling – мають фундаментальне обмеження: вони або повністю ігнорують контекст більшості фрагментів (max), або однаково зважують всі екземпляри незалежно від їх діагностичної цінності (mean). В умовах гістопатологічного аналізу це є критичним недоліком, оскільки патологічно значущі ділянки можуть займати лише невелику частку всього препарату.

У роботі [12] запропоновано принципово інший підхід – MIL на основі механізму уваги, де модель явно навчається визначати відносну важливість кожного екземпляра. Агреговане представлення набору формується як зважена сума векторних представлень екземплярів:

$$z = \sum_{k=1}^K a_k h_k, \quad (2.1)$$

де ваги уваги a_k задовольняють умову $\sum_{k=1}^K a_k = 1$ та обчислюються через навчувану мережу уваги:

$$a_k = \frac{\exp \{w^T \tanh(V h_k^T)\}}{\sum_{j=1}^K \exp \{w^T \tanh(V h_j^T)\}},$$

де $V \in \mathbb{R}^{L \times M}$ та $w \in \mathbb{R}^{L \times 1}$ – навчувані параметри, $h_k \in \mathbb{R}^M$ – векторне представлення k -го екземпляра, L – розмірність прихованого простору уваги.

Функція $\tanh$ вводить нелінійність, що дозволяє моделі вловлювати складні залежності між ознаками фрагментів. Без неї механізм уваги зводився б до лінійної комбінації і мав би обмежену виразність.

Ключова властивість такого механізму – інтерпретованість. Ваги a_k безпосередньо відображають внесок кожного фрагмента у фінальне рішення. Після навчання їх можна візуалізувати у вигляді теплової карти уваги поверх оригінального препарату, що дозволяє патоморфологу бачити які саме ділянки тканини модель вважає діагностично значущими [34].

Також запропоновано розширення – механізм уваги з вентилем (gated attention), де додається сигмоїдний вентиль для більш гнучкого контролю потоку інформації:

$$a_k = \frac{\exp \{w^T (\tanh (V h_k^T) \odot \sigma (U h_k^T))\}}{\sum_{j=1}^K \exp \{w^T (\tanh (V h_j^T) \odot \sigma (U h_j^T))\}},$$

де $U \in \mathbb{R}^{L \times M}$ – додаткова навчувана матриця, $\sigma(\cdot)$ – сигмоїдна функція, $\odot$ – поелементне множення. Сигмоїдний вентиль фактично фільтрує які ознаки впливають на вагу уваги, що підвищує стабільність навчання при великих та різномірних наборах.

Порівняно з $\max$ та mean pooling , MIL на основі механізму уваги демонструє суттєво кращі результати на задачах класифікації WSI [12], оскільки дозволяє моделі зосередитись на клінічно релевантних ділянках тканини. Саме цей підхід став основою для подальших архітектур, зокрема CLAM, яка розглядається в наступному підрозділі.

2.2 Архітектура CLAM

Попри ефективність MIL на основі механізму уваги, класичний підхід має суттєве обмеження: механізм уваги навчається виключно на основі сигналу на рівні набору і може бути нестабільним або фокусуватись на шумових фрагментах. В роботі [10] запропоновано архітектуру CLAM (Clustering-constrained Attention Multiple Instance Learning), яка вирішує цю проблему через додатковий контроль на рівні екземплярів.

Загальна архітектура CLAM складається з чотирьох компонентів.

1. Кодувальник ознак f_θ – перетворює фрагмент на векторне представлення.
2. Мережа уваги – обчислює ваги уваги для кожного екземпляра.
3. Класифікатор на рівні набору – приймає фінальне рішення про клас препарату.
4. Гілка кластеризації на рівні екземплярів – додаткова гілка для стабілізації навчання.

2.2.1 Об'єднання на основі уваги

Агрегація в CLAM виконується аналогічно до MIL з механізмом уваги з вентилем [12]:

$$z = \sum_{k=1}^K a_k h_k, \quad a_k = \frac{\exp \{w^T (\tanh(Vh_k^T) \odot \sigma(Uh_k^T))\}}{\sum_{j=1}^K \exp \{w^T (\tanh(Vh_j^T) \odot \sigma(Uh_j^T))\}},$$

де V , U , w – навчувані параметри мережі уваги. Отримане представлення препарату z подається на класифікатор рівня набору для фінальної класифікації.

2.2.2 Контроль на рівні екземплярів

Ключова інновація CLAM – використання оцінок уваги для генерації псевдоміток на рівні екземплярів без будь-якої ручної розмітки фрагментів [10]. Для кожного препарату вибираються k з найбільшим та k з найменшим значенням уваги фрагментів:

$$\mathcal{P}^+ = \text{TopK}(\{a_k\}_{k=1}^K), \quad \mathcal{P}^- = \text{BottomK}(\{a_k\}_{k=1}^K).$$

Фрагменти з $\mathcal{P}^+$ отримують псевдомітку відповідного класу препарату, фрагменти з $\mathcal{P}^-$ – протилежного. Це відповідає інтуїції: фрагменти з найбільшим значенням уваги мають бути найбільш інформативними для класу препарату, а з найменшим – навпаки.

На основі цих псевдоміток навчається класифікатор екземплярів через додаткову функцію втрат:

$$\mathcal{L}_{\text{inst}} = -\frac{1}{2k} \sum_{j \in \mathcal{P}^+ \cup \mathcal{P}^-} \log p(y_j | h_j),$$

де $p(y_j | h_j)$ – ймовірність псевдомітки для j -го фрагмента.

2.2.3 Загальна функція втрат

Повна функція втрат CLAM поєднує компоненти рівня набору та рівня екземплярів [10]:

$$\mathcal{L} = \mathcal{L}_{\text{bag}} + \lambda \mathcal{L}_{\text{inst}},$$

де $\mathcal{L}_{\text{bag}}$ – функція втрат бінарної крос-ентропії на рівні набору, λ – коефіцієнт балансу між двома компонентами.

Контроль на рівні екземплярів виконує одразу кілька функцій. По-перше, він стабілізує механізм уваги – модель не може просто присвоїти довільні ваги, бо вони безпосередньо впливають на функцію втрат екземплярів. По-друге, він покращує інтерпретованість – карти уваги стають більш чіткими і локалізованими [34]. По-третє, він зменшує ризик фокусування на артефактах або фоновій тканині.

2.2.4 Мультикласова версія

В роботі [10] також запропоновано мультикласову версію CLAM-MB (multi-branch), де для кожного класу існує окрема гілка уваги:

$$a_k^{(c)} = \frac{\exp \{w_c^T (\tanh(Vh_k^T) \odot \sigma(Uh_k^T))\}}{\sum_j \exp \{w_c^T (\tanh(Vh_j^T) \odot \sigma(Uh_j^T))\}},$$

де w_c – вектор уваги специфічний для класу c . Це дозволяє моделі виявляти різні морфологічні патерни для різних класів, що особливо важливо при стадіюванні, де кожна стадія має свої характерні гістологічні ознаки.

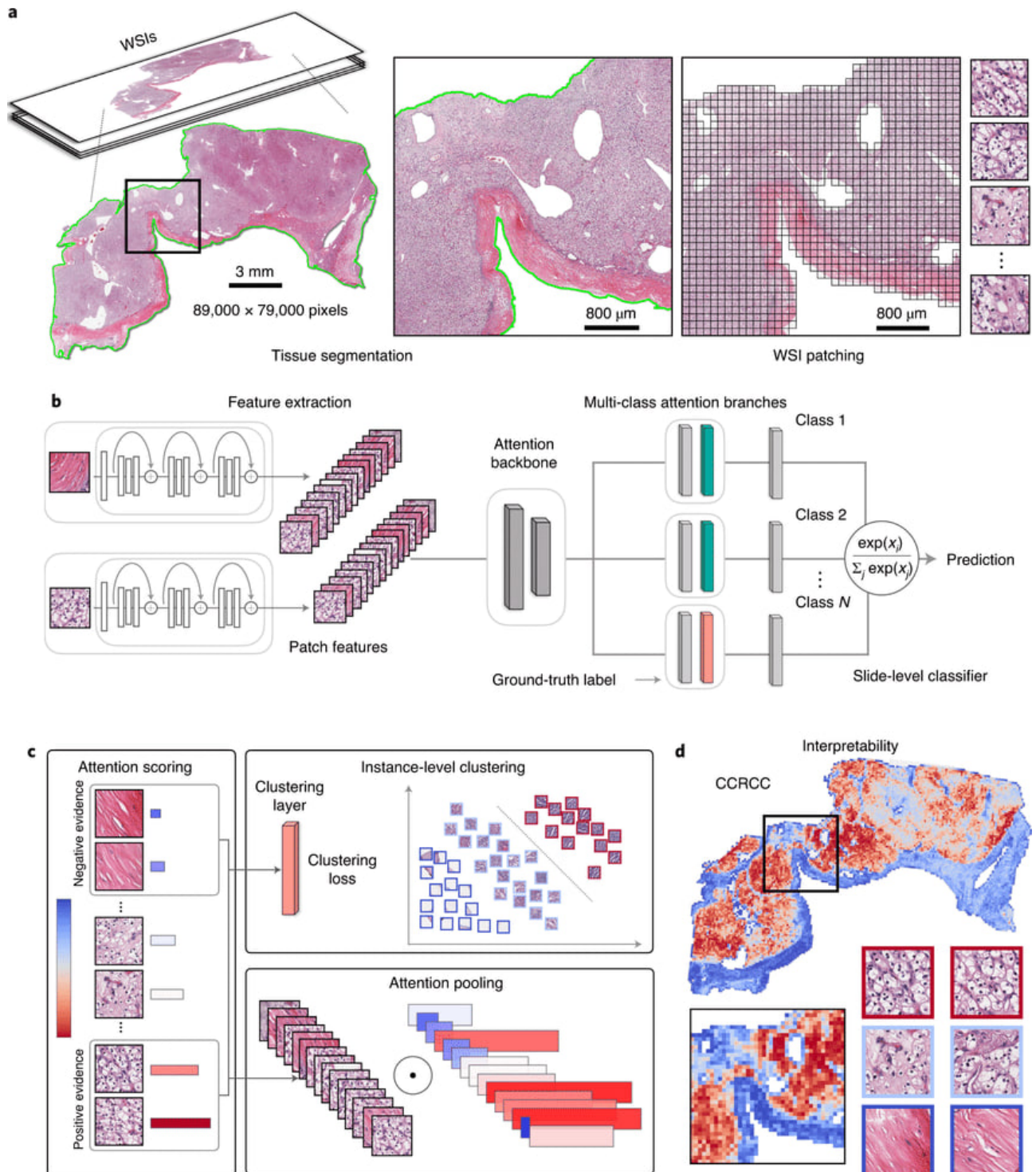


Рисунок 2.1 – Архітектура CLAM: пайплайн обробки WSI, агрегація на основі механізму уваги та інтерпретованість через теплові карти [10]

Завдяки поєднанню об'єднання на основі уваги з контролем на рівні екземплярів CLAM став одним із найбільш поширених базових рішень в обчислювальній патології [10]. Він демонструє стабільні результати на задачах виявлення метастазів, класифікації підтипів та прогнозування виживаності, що і зумовило його вибір як основної архітектури в даній роботі.

2.3 Гістологічні фундаментальні моделі

2.3.1 Роль векторних представлень у пайплайні MIL

У типовому пайплайні зі слабким контролем для аналізу WSI фундаментальна модель виконує роль замороженого кодувальника фрагментів [10]. Після сегментації тканини та нарізки препарату на фрагменти кожен з них незалежно проходить через кодувальник і перетворюється на вектор фіксованої розмірності. Саме ці вектори, а не самі зображення, подаються на вхід MIL-агрегатора. Такий підхід дозволяє ефективно розділити два принципово різних завдання: вилучення морфологічних ознак із фрагментів та агрегацію цих ознак у передбачення на рівні препарату.

Заморожування ваг кодувальника є принциповим архітектурним рішенням. По-перше, типові тренувальні вибірки в задачах стадіювання містять від кількох сотень до кількох тисяч препаратів, що є недостатнім для стабільного фінтунінгу моделей розміру ViT-Huge (понад 600 млн параметрів) без ризику катастрофічного перенавчання [30]. По-друге, сучасні фундаментальні моделі вже містять багате морфологічне представлення тканини, сформоване на мільйонах фрагментів. Завдання MIL-компонента в такому разі зводиться до навчання специфічної клінічної агрегації цих представлень.

Якість векторних представлень безпосередньо визначає верхню межу продуктивності всього пайплайну [11]. Нижче розглянуто три представницькі моделі, що охоплюють різні підходи до попереднього навчання у патологічному домені. Узагальнені характеристики моделей показано у табл. 2.1.

Таблиця 2.1 — Порівняльна характеристика гістологічних фундаментальних моделей

<i>Модель</i>	<i>Архітектура</i>	<i>Обсяг даних</i>	<i>Dim</i>	<i>Метод навчання</i>
UNI2-H	ViT-Huge	100k WSI	1536	DINOv2 (Vision-only)
CONCH	ViT-B/16	1.17M пар	512	CLIP (Vision-Language)
Virchow2	ViT-Huge	3.1M WSI	1280	DINOv2 + ECT + KDE

2.3.2 UNI2-H

UNI2-H є однією з найбільш потужних сучасних фундаментальних моделей для обчислювальної патології [30]. Модель базується на архітектурі ViT-Huge і навчена на корпусі понад 100 тисяч гістологічних препаратів із більш ніж 20 інституцій, що охоплює широкий спектр типів тканин і забезпечує високу узагальнюваність отриманих представлень.

Попереднє навчання базується на методі самодистиляції DINOv2 [35] з модифікаціями для патологічного домену. Модель демонструє стійкість до зсуву розподілу, зумовленого відмінностями у протоколах фарбування H&E між лабораторіями. Завдяки масштабу попереднього навчання векторні представлення UNI2-H кодують морфологічну інформацію на глибшому рівні порівняно з моделями попередніх поколінь, що критично для задач стадіювання [30]. Схему архітектури UNI2-H представлено на рис. 2.2.

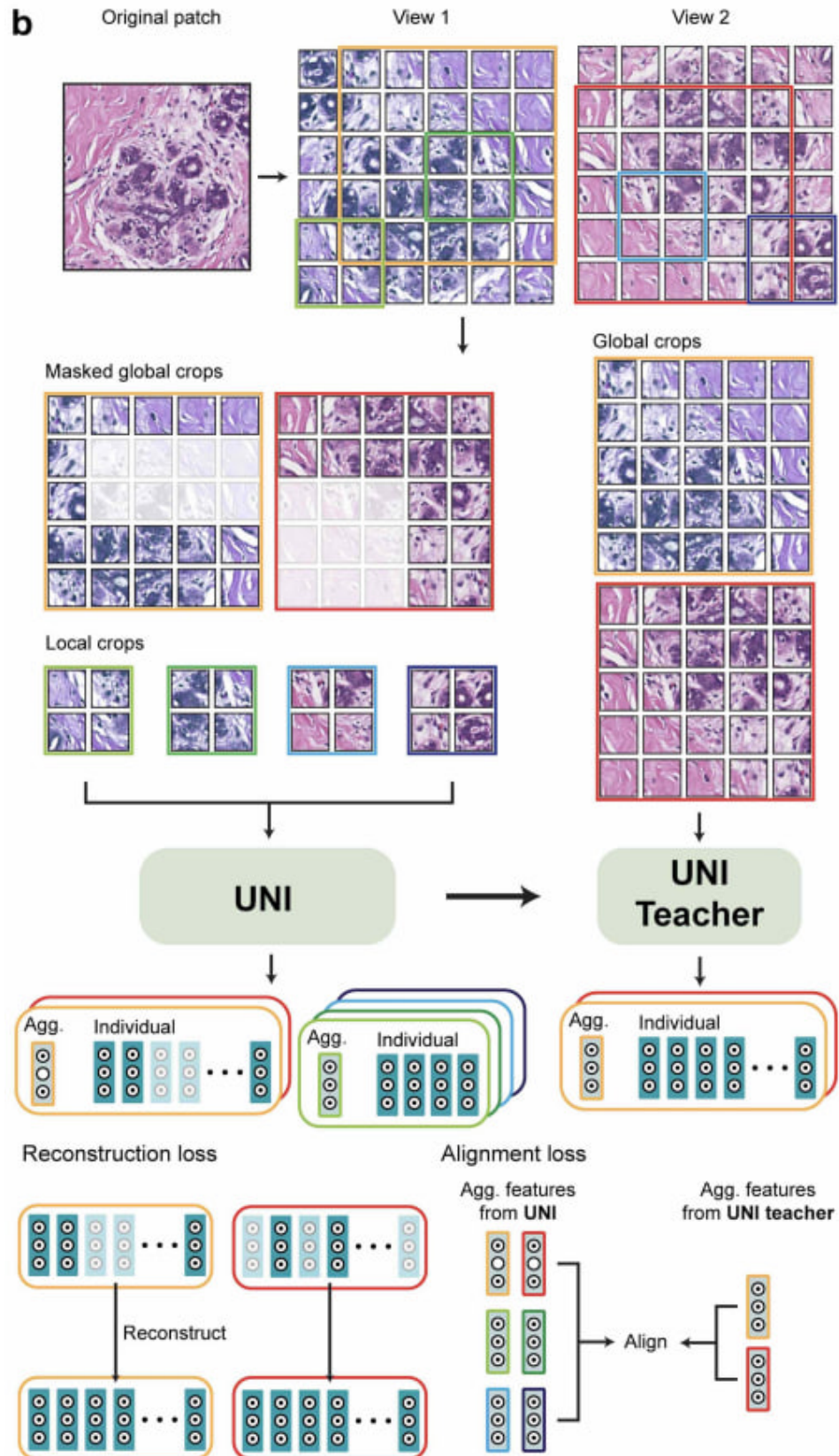


Рисунок 2.2 – Архітектура та пайплайн попереднього навчання UNI2-H [30]

2.3.3 CONCH

CONCH – це мультимодальна фундаментальна модель, що поєднує візуальні ознаки з текстовими патоморфологічними описами через контрастивне мовно-візуальне попереднє навчання [36]. На відміну від суто візуальних моделей, CONCH формує спільний простір представлень, де візуальні фрагменти та їхні текстові описи є близькими.

Така мультимодальна природа дозволяє моделі формувати семантично структуровані векторні представлення, безпосередньо пов'язані з клінічною термінологією. Хоча для суто візуальних задач деталізація може бути нижчою, ніж у UNI2-H, CONCH пропонує унікальну здатність до класифікації без додаткового навчання [36]. Архітектуру мультимодального навчання представлено на рис. 2.3.

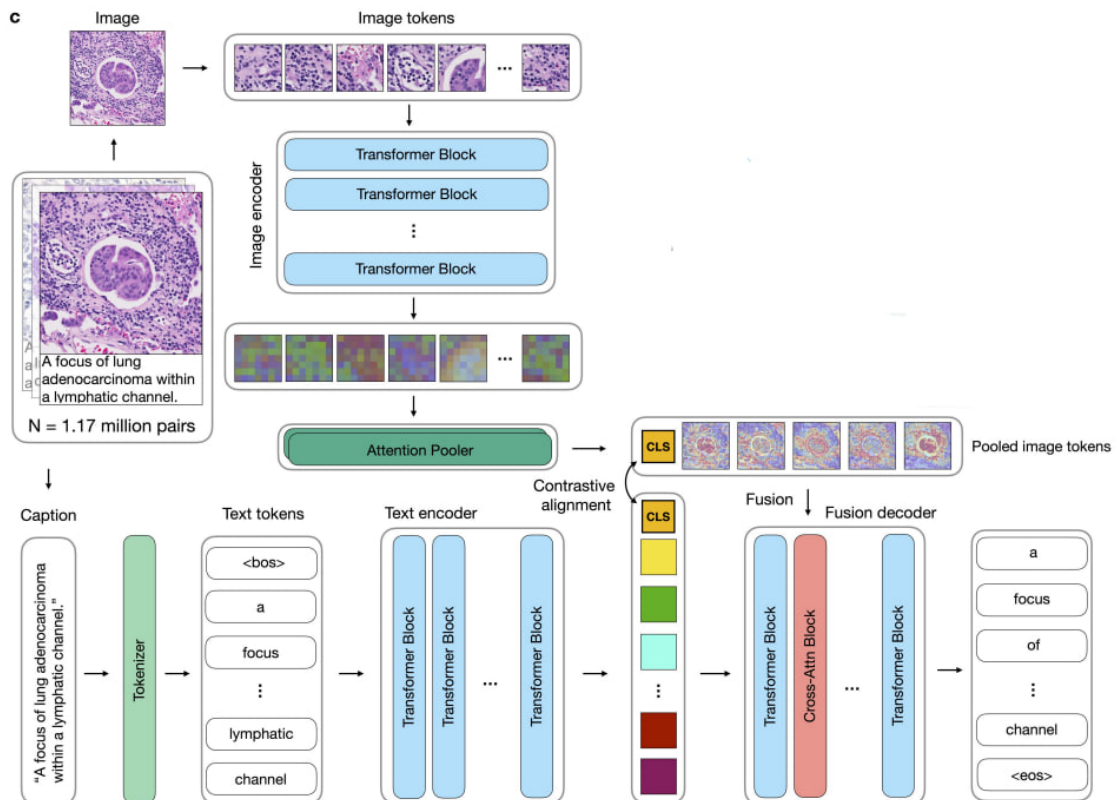


Рисунок 2.3 – Архітектура мультимодального контрастивного попереднього навчання CONCH [36]

2.3.4 Virchow2

Virchow2 є однією з найбільш масштабних моделей, навченою на корпусі понад 3.1 млн препаратів [33]. Її ключовою особливістю є використання мультимасштабного навчання та аугментації з розширеним контекстним перетворенням (Extended-Context Translation, ECT). На відміну від стандартних підходів, Virchow2 навчається на фрагментах різних масштабів (від 5x до 40x), що змушує модель формувати інваріантні до масштабу представлення [33]. ECT-аугментація дозволяє зберігати оригінальну клітинну морфологію без деформацій, характерних для операції зміни масштабу (рис. 2.4).

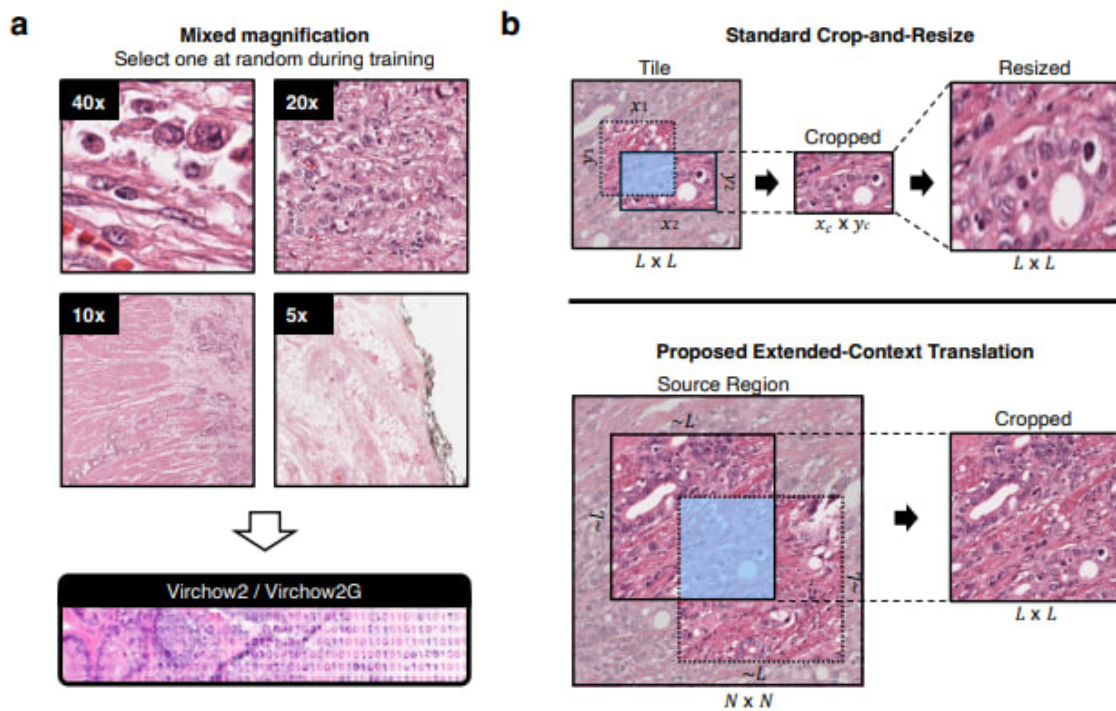


Рисунок 2.4 – Ключові модифікації Virchow2: мультимасштабне навчання (a) та ECT-аугментація (b) [33]

2.3.5 Вибір енкодера

З-поміж розглянутих моделей як основний енкодер у даній роботі обрано UNI2-H. Тренувальний корпус моделі (понад 100 000 WSI із понад 20 типів тканин, включаючи лімфовузли) є найбільш доменно релевантним до цільової задачі стадіювання метастазів у лімфатичних вузлах. Порівняльна оцінка альтернативних енкодерів виходить за межі фокусу цього дослідження і може бути предметом подальших робіт.

2.4 Огляд наборів даних

2.4.1 *CAMELYON16 та CAMELYON17*

CAMELYON16 і CAMELYON17 є двома найбільш поширеними публічними наборами даних для задач виявлення метастазів раку молочної залози у лімфатичних вузлах за гістологічними WSI [13; 17]. Обидва набори були зібрані в рамках однойменних міжнародних конкурсних змагань і стали еталонним орієнтиром для оцінки MIL-методів в обчислювальній патології.

CAMELYON16 містить 400 WSI гістологічних зрізів сигнальних лімфатичних вузлів, отриманих від пацієнток із раком молочної залози [17]. Препарати відскановані з роздільною здатністю 0.23–0.25 мікрон на піксель і поділені на тренувальну (270 препаратів) та тестову (130 препаратів) вибірки. Мітки включають два класи: негативні (відсутність метастазів) та позитивні (наявність метастатичної тканини). Для тренувальної вибірки додатково надані детальні піксельні анотації меж метастатичних регіонів, що робить набір придатним як для навчання зі слабким, так і з повним контролем.

CAMELYON17 розширює CAMELYON16 до 1000 WSI від п'яти різних медичних центрів із різними протоколами сканування та фарбування [13]. Тренувальна вибірка містить 500 препаратів із класифікацією на чотири категорії відповідно до клінічної класифікації метастазів: негативні (відсутність пухлинних клітин), ІТС (ізольовані пухлинні клітини, кластер менше 0.2 мм), мікрометастази (0.2–2.0 мм) та макрометастази (понад 2.0 мм). Така багатокласова структура безпосередньо відповідає клінічній системі pN-стадіювання [37] і робить набір особливо релевантним для задачі стадіювання.

Мультицентрова природа CAMELYON17 є ключовою характеристикою з точки зору побудови стійких моделей: варіативність між центрами охоплює різні типи сканерів, протоколи фарбування H&E та демографічні характеристики пацієнтів. Це дозволяє оцінити здатність моделі до узагальнення в умовах реального зсуву розподілу між медичними центрами.

У цій роботі обидва набори використовуються спільно для навчання та оцінки CLAM-моделі на задачі класифікації препаратів за ступенем метастатичного ураження лімфатичних вузлів в умовах слабкої розмітки – тобто без використання піксельних анотацій, лише на основі міток на рівні препарату.

2.4.2 *Melanoma-WSI (CMB-MEL + CPTAC-CM)*

Для оцінки здатності пайплайну до міждоменного узагальнення на задачі класифікації меланоми використовується об'єднаний датасет Melanoma-WSI, сформований із двох публічних колекцій: CMB-MEL (Cancer Moonshot Biobank Melanoma Collection) [38] та CPTAC-CM (Clinical Proteomic Tumor Analysis Consortium – Cutaneous Melanoma) [39]. Обидва набори містять H&E-зabarвлені WSI пацієнтів із шкірною меланою та доступні через портал GDC (Genomic Data Commons). CPTAC-CM є повторним дослідженням на перетинаючихся зраз-

ках із акцентом на протеомний аналіз, тому обидві колекції охоплюють подібний клінічний розподіл пацієнтів.

Об'єднаний датасет містить 505 H&E-забарвлених WSI: 409 слайдів із CPTAC-CM та 96 слайдів із CMB-MEL. Мітки є бінарними на рівні препарату (melanoma / benign). Розбиття на тренувальну та валідаційну вибірки виконувалось у співвідношенні 80/20 зі стратифікацією за класом, що забезпечує збереження розподілу класів в обох вибірках. Відсутність анотацій на рівні фрагментів робить датасет природним еталоном для підходів зі слабким контролем [11].

Особливістю Melanoma-WSI з точки зору задачі класифікації є висока морфологічна гетерогенність меланоми: пухлинні клітини можуть демонструвати різноманітні патерни – від епітеліоїдних до веретеноподібних форм – що суттєво ускладнює автоматичну класифікацію [38]. Висока мутаційна навантаженість, характерна для меланоми, додатково зумовлює значну варіативність між пацієнтами навіть у межах одного діагностичного класу.

У цій роботі датасет використовується для оцінки міждоменної узагальнюваності пайплайну на задачі бінарної класифікації меланоми в режимі слабкої розмітки – тобто без використання патч-рівневих анотацій, лише на основі міток на рівні препарату.

Висновки до розділу 2

У другому розділі проведено комплексний аналіз сучасних методів інтелектуального аналізу гістопатологічних зображень великого розміру та обґрунтовано вибір технологічного стеку для побудови системи стадіювання. На основі проведеного дослідження зроблено такі висновки.

1. Обґрунтовано вибір парадигми навчання на множинах екземплярів (MIL) як найбільш адекватної для обробки гігапіксельних зображень в умовах слабкої роз-

мітки [11; 12]. Встановлено, що класичні підходи до агрегації (max/mean pooling) не здатні ефективно виділяти поодинокі пухлинні кластери, тоді як механізм уваги дозволяє моделі фокусуватися на діагностично значущих фрагментах та забезпечує інтерпретованість результатів через візуалізацію теплових карт.

2. Визначено архітектуру CLAM як базову для реалізації системи [10]. Ключовою перевагою є інтеграція додаткової гілки кластеризації екземплярів, яка стабілізує навчання механізму уваги та дозволяє виявляти морфологічні ознаки без потреби у детальній піксельній анотації. Мультикласова модифікація CLAM-MB визначена як оптимальна для задач стадіювання, де необхідно диференціювати кілька ступенів патологічного процесу.

3. Проаналізовано клас гістологічних фундаментальних моделей (UNI2-H, CONCH, Virchow2) та їх роль у формуванні якісних векторних представлень. Показано, що використання заморожених кодувальників, попередньо навчених на надвеликих патоморфологічних архівах методами самонавчання (DINOv2) [35], дозволяє уникнути перенавчання на малих клінічних вибірках та забезпечує стійкість системи до варіативності лабораторного забарвлення. На основі аналізу доменної релевантності тренувального корпусу як основний кодувальник обрано UNI2-H [30], чий корпус попереднього навчання є найбільш близьким до цільової задачі стадіювання метастазів у лімфатичних вузлах.

4. Сформовано експериментальну базу, що включає набори даних CAMELYON16/17 [13; 17] та Melanoma-WSI (CMB-MEL + CPTAC-CM) [38; 39]. Цей набір охоплює як задачі класифікації метастазів у лімфатичних вузлах (pN-стадіювання), так і складні випадки морфологічно гетерогенної меланоми, що дозволяє всебічно оцінити якість розроблених алгоритмів у реальних клінічних сценаріях.

Таким чином, обрана комбінація архітектури CLAM та енкодера UNI2-H створює теоретичне та технологічне підґрунтя для переходу до практичної реалізації та експериментального дослідження системи інтелектуального стадіювання.

РОЗДІЛ 3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

Розроблена інтелектуальна система стадіювання реалізована як наскрізний пайплайн, що охоплює чотири функціональні підсистеми: введення та підготовки даних, екстракції ознак на основі UNI2-H, агрегації знань засобами MIL та пояснення рішення з формуванням діагностичного звіту (рис. 3.1).

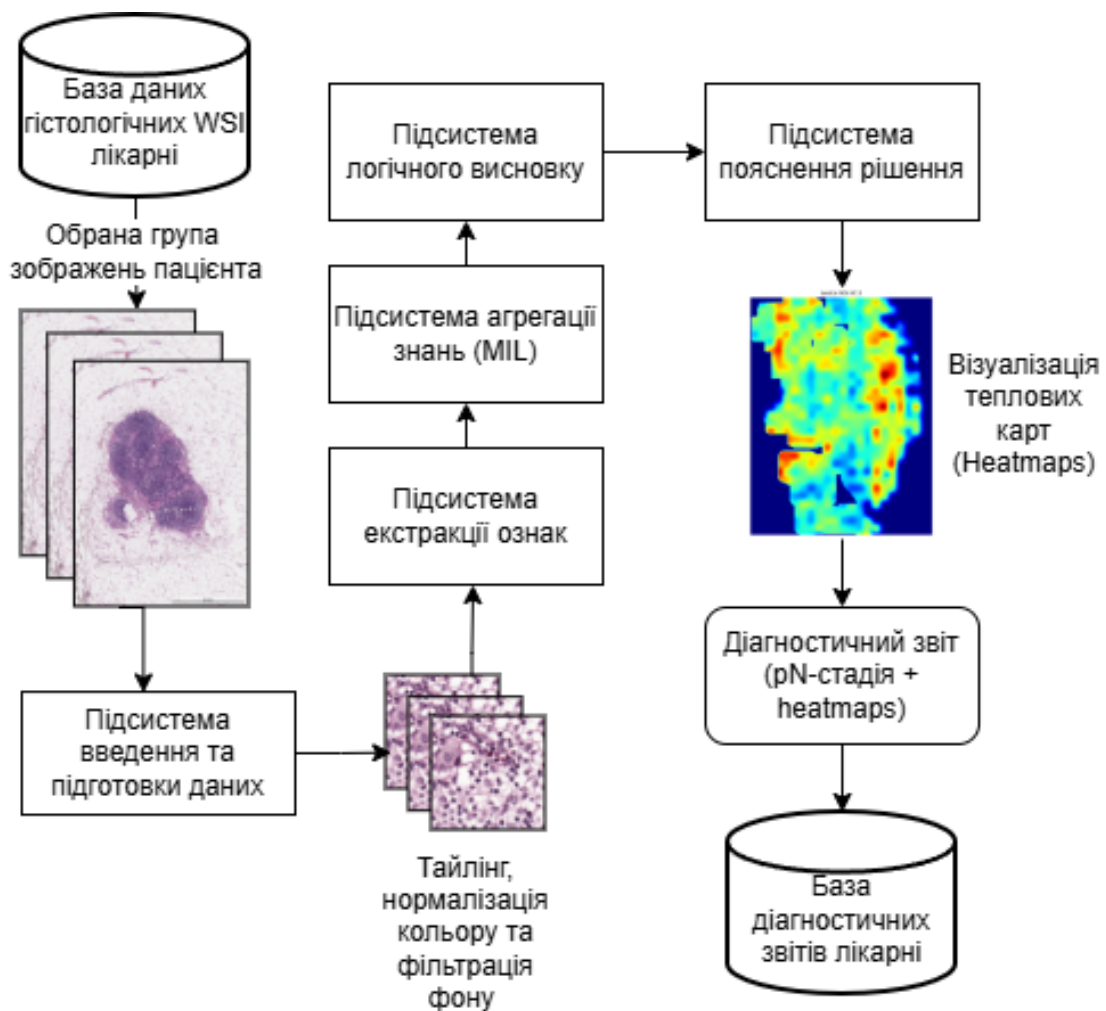


Рисунок 3.1 – Загальна архітектура інтелектуальної системи стадіювання: від бази гістологічних WSI до діагностичного звіту з картою уваги

3.1 Пайплайн попередньої обробки даних

Детальну схему обчислювального пайплайну продемонстровано на рис. 3.2. Пайплайн складається з чотирьох послідовних етапів: нарізки WSI на фрагменти з фільтрацією тканини, вилучення векторних представлень енкодером UNI2-h, навчання CLAM MIL-агрегатора та rule-based агрегації передбачень до рівня пацієнта відповідно до правил TNM.

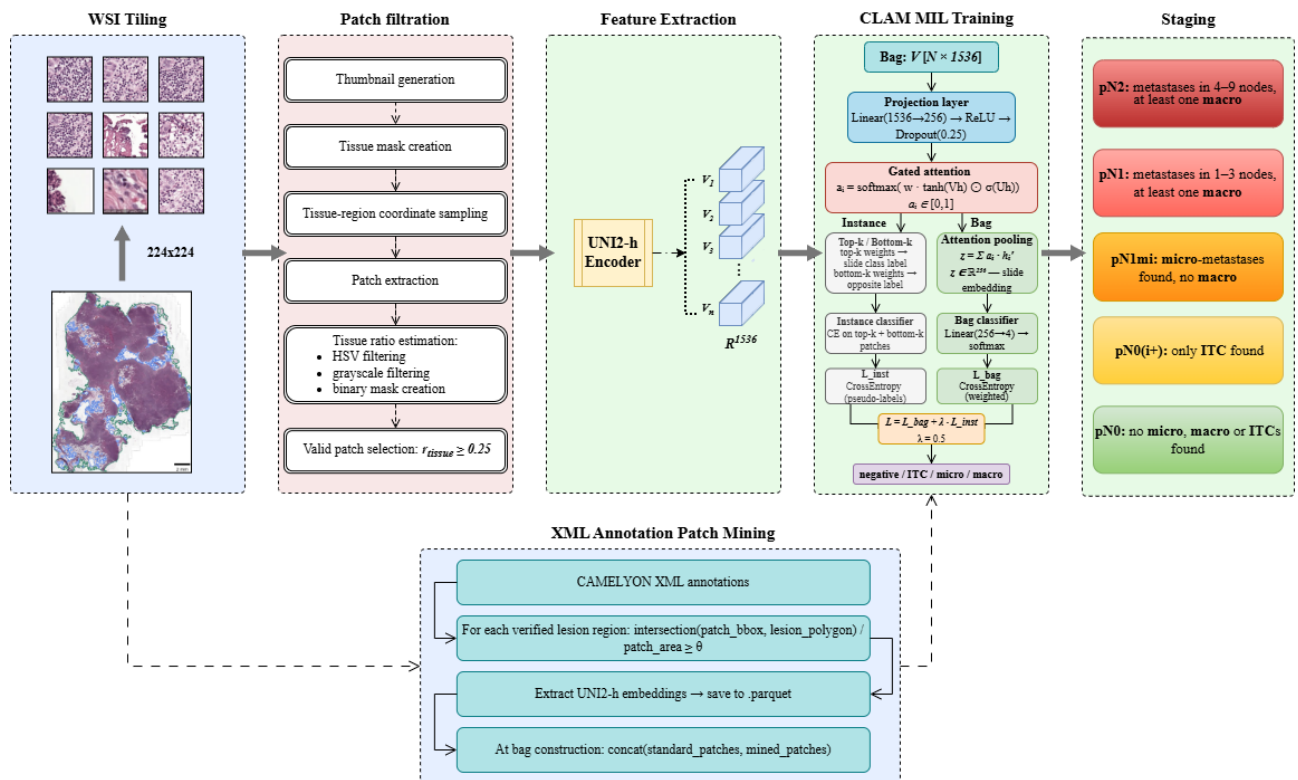


Рисунок 3.2 – Загальна архітектура пайплайну стадіювання раку молочної залози: від WSI до pN-стадії пацієнта

3.1.1 Сегментація тканини та нарізка на фрагменти

Попередня обробка гістологічних препаратів є критично важливим етапом пайплайну, оскільки значна частина WSI містить діагностично нерелевантні ділянки – фон, артефакти сканування та порожній простір між тканинними фрагментами. Включення таких ділянок до набору негативно впливає на якість векторних представлень та ускладнює навчання MIL-агрегатора [10].

У даній роботі сегментація тканини реалізована через порогову фільтрацію за яскравістю фрагментів [40]. Кожен препарат нарізається на фрагменти фіксованого розміру 224×224 пікселів без перекриття. Для кожного фрагмента обчислюється середня яскравість пікселів, і фрагмент включається до набору лише за умови (3.1):

$$20 \leq \bar{b} \leq 230, \quad (3.1)$$

де $\bar{b}$ – середня яскравість фрагмента. Фрагменти з $\bar{b} > 230$ відповідають фону (скло або порожній простір), тоді як фрагменти з $\bar{b} < 20$ є артефактами темного забарвлення. Таким чином до подальшої обробки потрапляють лише фрагменти, що містять тканину.

Розмір фрагмента 224×224 пікселів обраний відповідно до вхідного розміру моделі UNI2-H [30], що дозволяє уникнути зміни роздільної здатності та зберегти морфологічні деталі тканини.

3.1.2 Отримання векторних представлень UNI2-H

Для вилучення ознак на рівні фрагментів використовується заморожена базова модель UNI2-H, завантажена через HuggingFace Hub (MahmoodLab/UNI2-h) [30]. Модель базується на архітектурі ViT-Giant з параметрами: patch_size=14, embed_dim=1536, depth=24, num_heads=24, активація SwiGLU FFN.

Перед подачею до кодувальника кожен фрагмент проходить стандартну трансформацію:

$$\begin{aligned} x' &= \text{Normalize}(\text{Resize}(x, 224 \times 224), \\ &\quad \mu = (0.485, 0.456, 0.406), \\ &\quad \sigma = (0.229, 0.224, 0.225)), \end{aligned}$$

де значення μ та σ відповідають статистиці ImageNet. Обчислення виконується у режимі половинної точності (FP16) для підвищення ефективності на GPU [41].

Результатом є вектор розмірності 1536 для кожного фрагмента. Векторні представлення всіх фрагментів одного препарату зберігаються у файл формату .pt і утворюють набір для MIL-агрегатора згідно з виразом (1.1), де $h_{ij} \in \mathbb{R}^{1536}$. Для кожного препарату використовується до 512 фрагментів, що забезпечує баланс між повнотою покриття тканини та обчислювальною ефективністю навчання.

Для якісної оцінки просторового розподілу патч-рівневих оцінок проведено візуалізацію теплових карт активності UNI2-h на препараті CAMELYON17 (patient_010_node_4). На рис. 3.3 показано загальну теплову карту норми векторних представлень $\|h_j\|_2$ для всіх чотирьох тканинних фрагментів препарату.

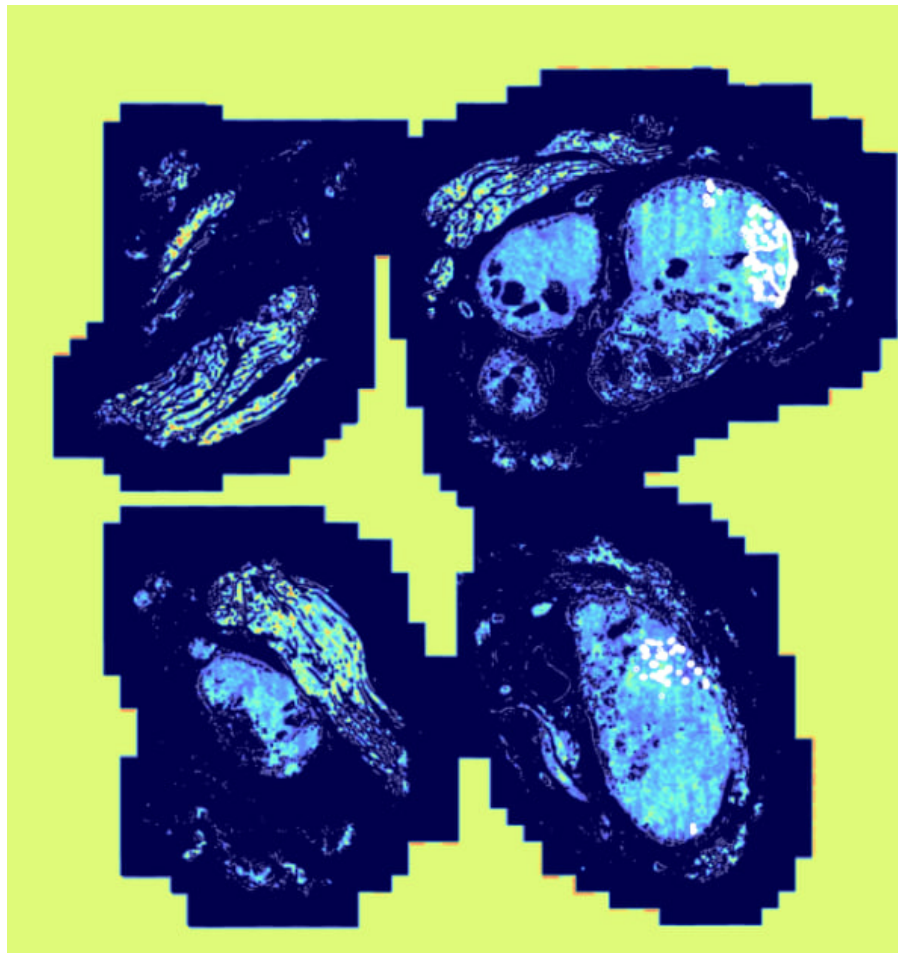


Рисунок 3.3 – Загальна теплова карта патч-рівневих оцінок UNI2-h для препарату CAMELYON17 patient_010_node_4. Білим контуром позначено межі верифікованих анотацій метастазів. Жовто-зелені ділянки відповідають підвищеній нормі векторного представлення $\|h_j\|_2$

Для детальнішого аналізу на рис. 3.4 показано порівняння тканинного кропу з накладеним контуром XML-анотації та відповідної теплової карти для репрезентативної ділянки з верифікованим метастазом (Lesion #13). Ділянки з підвищеною активністю просторово узгоджуються з областю верифікованого метастазу, що підтверджує здатність заморожених ознак UNI2-h відображати патоморфологічно значущі структури без додаткового донавчання енкодера.

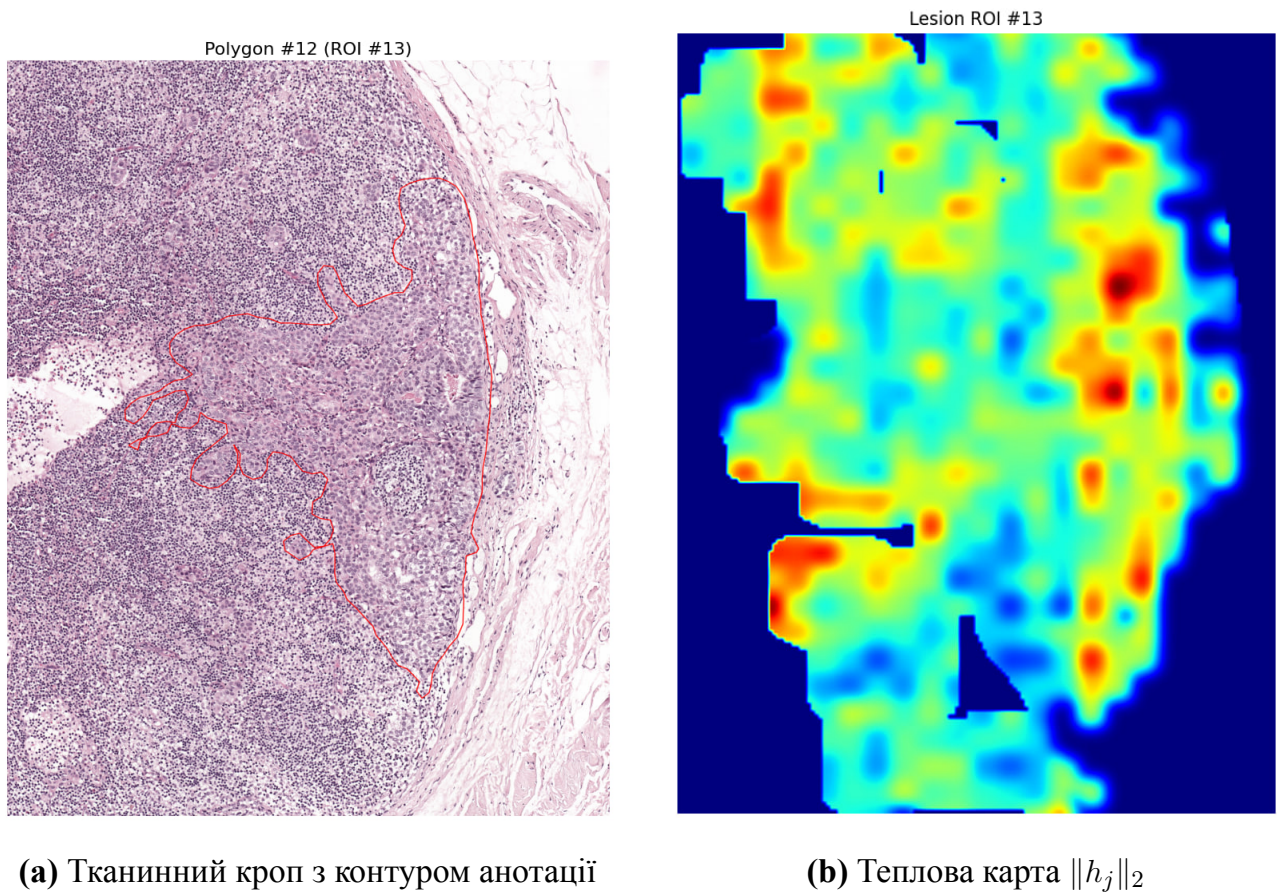


Рисунок 3.4 – Просторова візуалізація патч-рівневих оцінок UNI2-h для ділянки з верифікованим метастазом (Lesion #13, CAMELYON17, patient_010_node_4): тканинний кроп із червоним контуром XML-анотації (ліворуч) та відповідна теплова карта норми векторного представлення $\|h_j\|_2$ (праворуч)

3.1.3 Відбір інформативних фрагментів

У даній роботі застосовується дворівневий механізм відбору інформативних фрагментів. Перший рівень — статичне збагачення bag: для препаратів із доступними XML-анотаціями CAMELYON попередньо відібрані фрагменти, що перетинаються з верифікованими регіонами метастазів, зберігаються як UNI2-h

ембедінги у parquet-файлі та конкатенуються до стандартного набору екземплярів на всіх етапах навчання.

Другий рівень — динамічний напівжорсткий відбір (semi-hard patch mining). Починаючи з епохи `mining_start_epoch`, для кожного препарату обчислюються оцінки екземплярів $\{s_j\}_{j=1}^M$ на основі поточних ваг моделі, і фрагменти впорядковуються за їх значенням. Далі відбираються K фрагментів за правилом:

$$\mathcal{P}_{\text{main}} \sim \text{Uniform}(\{s_{[K/2]}, \dots, s_{[K \cdot \gamma]}\}),$$

де $\gamma = 3.0$ — коефіцієнт пулу, а квадратні дужки позначають ранг у впорядкованій послідовності. Додатково випадково додається $\lfloor K \cdot r \rfloor$ фрагментів із решти, де $r = 0.2$ — коефіцієнт випадковості. Перший рівень гарантує присутність діагностично верифікованих ділянок у bag, тоді як другий зосереджує навчання на складних але не аномальних фрагментах, що разом знижує ризик перенавчання на шумових артефактах.

3.2 Порівнювані методи MIL

Для оцінки ефективності різних підходів до агрегації векторних представлень фрагментів у даній роботі порівнюються чотири методи MIL. Всі методи використовують однаковий заморожений кодувальник UNI2-H та однакові тренувальні дані, що забезпечує коректне порівняння архітектур агрегації. Гіперпараметри навчання уніфіковані: прихований розмір $hid = 256$, відсіювання $= 0.25$, швидкість навчання 2×10^{-4} , регуляризація вагового затухання 10^{-4} , косинусний планувальник з 2 епохами розігріву, раннє зупинення з терпінням 6 [42].

3.2.1 ABMIL

ABMIL (Attention-Based Multiple Instance Learning) є класичним методом агрегації на основі механізму уваги з вентилем [12]. На відміну від звичайного softmax, механізм з вентилем використовує дві паралельні гілки для більш вибіркового контролю потоку інформації:

$$a_k = \frac{\exp \left\{ w^T \left(\tanh \left(V h_k^T \right) \odot \sigma \left(U h_k^T \right) \right) \right\}}{\sum_{j=1}^K \exp \left\{ w^T \left(\tanh \left(V h_j^T \right) \odot \sigma \left(U h_j^T \right) \right) \right\}},$$

де $\tanh$ -гілка визначає що є важливим, а σ -гілка виступає вентилем що регулює скільки інформації пропускається. Добуток двох гілок забезпечує більш вибіркову агрегацію порівняно зі звичайним softmax. Архітектурно модель складається з проєкційного шару $\mathbb{R}^{1536} \rightarrow \mathbb{R}^{256}$, мережі уваги з прихованим розміром $L = 128$ та класифікаційної головки.

3.2.2 CLAM

CLAM розширює ABMIL додатковим контролем на рівні екземплярів через псевдомітки [10]. Детальний опис архітектури представлено у підрозділі 2.2. У порівняльному експерименті використовується бінарна версія CLAM-SB з механізмом уваги з вентилем та функцією втрат екземплярів на $k = 8$ найбільш та найменш уважних фрагментах, коефіцієнт $\lambda = 0.5$ для балансу між компонентами функції втрат.

3.2.3 TransMIL

TransMIL [18] замінює класичний механізм уваги трансформерним кодувальником, що дозволяє моделювати взаємозв'язки між фрагментами через самоувагу. Ключовою відмінністю є використання апроксимації Hiström [43] для ефективного обчислення самоуваги на великих наборах:

$$\text{NystromAttention}(Q, K, V) \approx \hat{A} \cdot \tilde{A}^{-1} \cdot \bar{A} \cdot V,$$

де $\hat{A}$, $\tilde{A}$, $\bar{A}$ – матриці що апроксимують повну самоувагу через m опорних точок замість усіх N фрагментів, що знижує складність з $O(N^2)$ до $O(Nm)$.

Архітектура включає проєкційний шар $\mathbb{R}^{1536} \rightarrow \mathbb{R}^{256}$, навчуваний класифікаційний токен що агрегує глобальну інформацію, та $depth = 2$ блоки TransMIL з $heads = 8$ головами уваги та $landmarks = 256$. Фінальна класифікація виконується на основі класифікаційного токена після проходження через усі блоки.

3.2.4 DSMIL

DSMIL (Dual-Stream Multiple Instance Learning) [44] використовує принципово інший підхід: замість єдиного агрегатора модель має два паралельні потоки – потік екземплярів та потік набору.

Потік екземплярів обробляє кожен фрагмент незалежно та обчислює оцінку екземпляра:

$$s_j = f_{\text{inst}}(h_j), \quad s_j \in \mathbb{R}.$$

Критичний фрагмент з максимальною оцінкою виступає як запит для потоку набору:

$$j^* = \arg \max_j s_j, \quad q = h_{j^*}.$$

Потік набору обчислює увагу між запитом та всіма фрагментами:

$$\alpha_j = \text{softmax} \left(\frac{q \cdot h_j^T}{\sqrt{d}} \right), \quad z = \sum_j \alpha_j h_j.$$

Фінальний результат є зваженою сумою передбачень екземплярів та набору:

$$\hat{y} = \frac{1}{N} \sum_j s_j + f_{\text{bag}}(z),$$

де вага між двома потоками регулюється гіперпараметром $\lambda_{\text{dsml}} = 0.5$. Двопоточковий підхід дозволяє одночасно враховувати як локальні ознаки екземплярів, так і глобальний контекст набору [44].

3.3 Налаштування експерименту

3.3.1 Гіперпараметри навчання

Для забезпечення коректного порівняння всі методи multiple instance learning (MIL) навчались з однаковими гіперпараметрами оптимізації. Навчан-

ня виконувалось на графічному прискорювачі NVIDIA Tesla T4. Оптимізацію параметрів моделі здійснювали за допомогою методу AdamW [45], а основні гіперпараметри показано в таблиці 3.1.

Таблиця 3.1 — Гіперпараметри навчання MIL-моделей

<i>Параметр</i>	<i>Значення</i>
Оптимізатор	AdamW
Learning rate	2×10^{-4}
Weight decay	1×10^{-4}
Scheduler	косинусне згасання learning rate
Warmup epochs	2
Максимальна кількість епох	20
Early stopping patience	6
Gradient clipping	5.0
Прихований розмір (h_{dim})	256
Dropout	0.25
Розмір batch	1 слайд
Кількість патчів на слайд	512
Random seed	42

Розмір batch дорівнює одному слайду, що є стандартною практикою для MIL у задачах аналізу whole slide images – кожен слайд розглядається як окремий bag і обробляється повністю за один forward pass. Використання косинусного scheduler забезпечує плавне зменшення learning rate від початкового значення до $0.1 \cdot lr_{base}$ протягом навчання, що сприяє стабільній збіжності моделі.

Розбиття даних на тренувальну та валідаційну вибірки виконувалось на рівні пацієнтів для набору CAMELYON17 [46] – тобто всі слайди одного пацієнта потрапляли або до train, або до validation, що запобігає витоку інформації між вибірками. Валідаційна вибірка містить 95 слайдів виключно з CAMELYON17, тоді як тренувальна вибірка складається з 500 слайдів із об'єднаних наборів CAMELYON16 та CAMELYON17.

3.3.2 Метрики оцінювання

Для всебічної оцінки якості класифікації використовується комбінація метрик, що охоплюють різні аспекти продуктивності моделі.

AUC (Area Under the ROC Curve) використовується для бінарної задачі (наявність/відсутність метастазів). Метрика оцінює здатність моделі розрізняти класи незалежно від порогу класифікації та є стійкою до дисбалансу класів.

Формально AUC визначається як площа під ROC-кривою:

$$\text{AUC} = \int_0^1 \text{TPR}(t) d\text{FPR}(t),$$

де TPR (true positive rate) та FPR (false positive rate) визначаються як функції порогу класифікації. Значення $\text{AUC} = 1$ відповідає ідеальній класифікації, тоді як $\text{AUC} = 0.5$ відповідає випадковому вгадуванню.

F1-score використовується як додаткова метрика для бінарної задачі та є гармонічним середнім між precision та recall:

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

де $\text{Precision} = \frac{TP}{TP+FP}$ — частка правильно передбачених позитивних прикладів, $\text{Recall} = \frac{TP}{TP+FN}$ — повнота виявлення позитивного класу.

QWK (Quadratic Weighted Kappa) використовується для мультикласової задачі стадіювання (negative / ITC / micro / macro). Метрика враховує порядкову структуру класів, тобто помилки між близькими класами штрафуються менше, ніж між віддаленими.

Коефіцієнт визначається як:

$$\kappa = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}}, \quad w_{ij} = \frac{(i - j)^2}{(N - 1)^2},$$

де O_{ij} — матриця спостережуваних збігів, E_{ij} — матриця очікуваних збігів за випадкової моделі, а w_{ij} — квадратичні ваги штрафу. Значення $\kappa = 1$ відповідає ідеальній згоді, тоді як $\kappa = 0$ відповідає випадковому узгодженню. Дана метрика використовується як основна в CAMELYON17 challenge [17] для оцінки якості pN-стадіювання. Для оцінки якості на рівні пацієнта використовується той самий показник QWK, що обчислюється між передбаченим та істинним pN-класом пацієнта після агрегації node-level передбачень.

Для аналізу структури помилок класифікації використовується **матриця невідповідностей (confusion matrix)**, яка відображає розподіл передбачених класів відносно істинних. Вона є особливо інформативною для мультикласових задач, оскільки дозволяє виявити систематичні помилки між суміжними класами — наприклад, плутанину між ITC та negative або між micro та macro метастазами.

Додатково для мультикласової задачі звітується **accuracy** на рівні слайду (slide-level accuracy) — частка правильно класифікованих зразків серед усіх прикладів.

3.4 Результати на CAMELYON16+17

3.4.1 Бінарна класифікація

Для переходу від неперервних ймовірнісних оцінок, що генеруються MIL-моделями, до дискретних бінарних міток (negative/positive) необхідно зафіксувати робочий поріг класифікації. Зважаючи на високу ціну хибнонегативного пропуску мікрометастазів та поодиноких пухлинних клітин в задачах онкодіагностики, стандартний поріг $t = 0,5$ є неефективним, оскільки призводить до зниження чутливості системи.

Для компенсації цього ефекту та максимізації балансу між чутливістю (повнотою виявлення) та специфічністю на валідаційній вибірці було експериментально визначено та зафіксовано оптимальний поріг бінарної класифікації $t_{\text{opt}} = 0,29$. Будь-які передбачені моделлю ймовірності, що дорівнюють або перевищують це значення, класифікуються як позитивні. Саме за умови фіксації порогу на рівні 0,29 обчислено всі наведені далі точкові метрики (F_1 -score) та відповідні матриці невідповідностей.

Результати бінарної класифікації (наявність/відсутність метастазів) на валідаційній вибірці з 95 слайдів показано у таблиці 3.2. Валідаційна вибірка містить 60 негативних та 35 позитивних слайдів, що відповідає помірному дисбалансу класів, характерному для клінічних даних.

Таблиця 3.2 — Результати бінарної класифікації на CAMELYON16+17

<i>Метод</i>	<i>AUC</i>	<i>F1</i>	<i>QWK</i>
CLAM	0.9571	0.9394	0.9073
TransMIL	0.9719	0.9577	0.9325
DSMIL	0.9767	0.9855	0.9772
ABMIL	0.9690	0.9706	0.9542

Усі розглянуті методи демонструють високу якість бінарної класифікації з AUC у діапазоні 0.957–0.977, що підтверджує ефективність UNI2-h як енкодера ознак у MIL-пайплайні. Водночас спостерігаються відмінності між методами за стабільністю збіжності та швидкістю навчання.

Найвищі результати у бінарній задачі досягаються методом DSMIL (AUC = 0.9767, F1 = 0.9855). Дана модель досягає оптимального значення вже на ранніх епохах навчання з подальшим застосуванням ранньої зупинки на сьомій епосі, що свідчить про швидку збіжність. Dual-stream архітектура ефективно реалізує MIL-припущення, згідно з яким наявність хоча б одного патологічного патча є достатньою для позитивного прогнозу.

Метод TransMIL посідає друге місце за AUC (0.9719). Застосування трансформерного механізму самоуваги з Nyström-апроксимацією дозволяє моделювати взаємозв'язки між патчами, зокрема просторову узгодженість морфологічно схожих ділянок. Водночас спостерігається менш стабільна динаміка навчання, що проявляється у коливаннях валідаційної функції втрат.

Метод ABMIL із gated attention показує AUC = 0.9690 та F1 = 0.9706. Його перевагою є проста та стабільна структура агрегації ознак. Використання gated attention, що поєднує tanh- та σ -гілки, забезпечує ефективне виділення інформативних патчів без значного ускладнення моделі.

Метод CLAM демонструє найнижчі результати у бінарній задачі (AUC = 0.9571, F1 = 0.9394). Попри це, він досягає найкращого значення вже на другій

епосі, після чого спостерігається деградація якості через перенавчання. Це може бути пов'язано з нестабільністю псевдоміток, які формуються на основі attention-score на ранніх етапах навчання.

3.4.2 Мультикласове стадіювання

Результати мультикласової класифікації за стадіями метастазування (negative / ITC / micro / macro) представлено у таблиці 3.3. Задача є суттєво складнішою порівняно з бінарною класифікацією через тонкі морфологічні відмінності між суміжними класами та виражений дисбаланс розподілу класів у валідаційній вибірці (рис. 3.5).

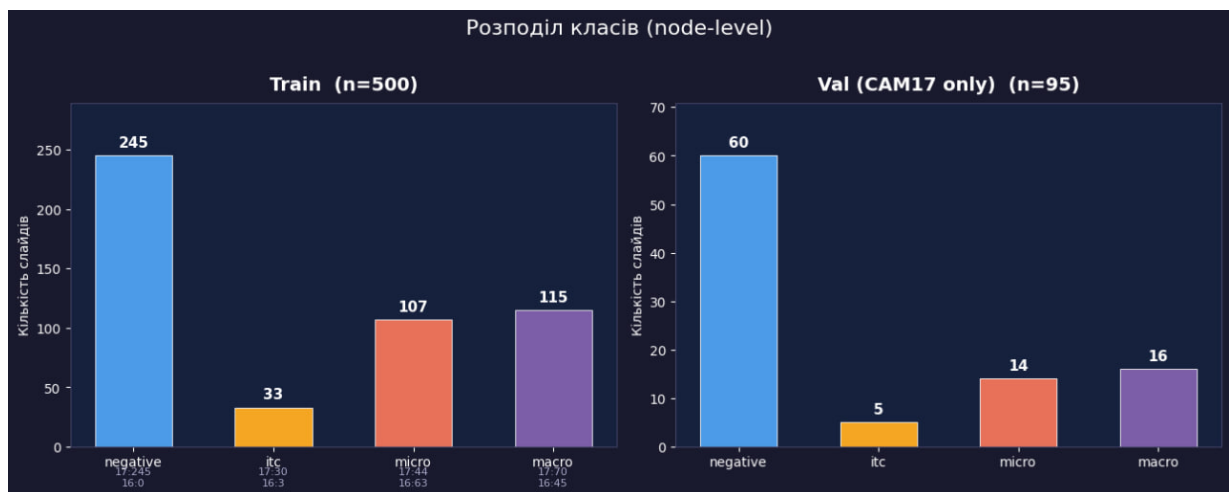


Рисунок 3.5 – Розподіл класів на рівні вузла: тренувальна вибірка (n=500, ліворуч) та валідаційна вибірка з CAMELYON17 (n=95, праворуч). Виражений дисбаланс на користь класу negative ускладнює навчання на рідкісних класах ITC та мікрометастазів

Таблиця 3.3 — Результати мультикласового стадіювання на
CAMELYON16+17

<i>Метод</i>	<i>QWK</i>	<i>Node Accuracy</i>
CLAM	0.9609	0.8421
TransMIL	0.9443	0.8737
DSMIL	0.9429	0.8737
ABMIL	0.9413	0.8526

У мультикласовій задачі стадіювання спостерігається істотна зміна відносної ефективності методів порівняно з бінарною постановкою. Зокрема, CLAM демонструє найвищий показник $QWK = 0.9609$ при node accuracy 0.8421, тоді як у задачі виявлення метастазів він поступався іншим підходам. Це є важливим результатом, що підкреслює принципову різницю між задачами детекції та стадіювання. Матриці невідповідностей для бінарної та мультикласової задач показано на рис. 3.6.

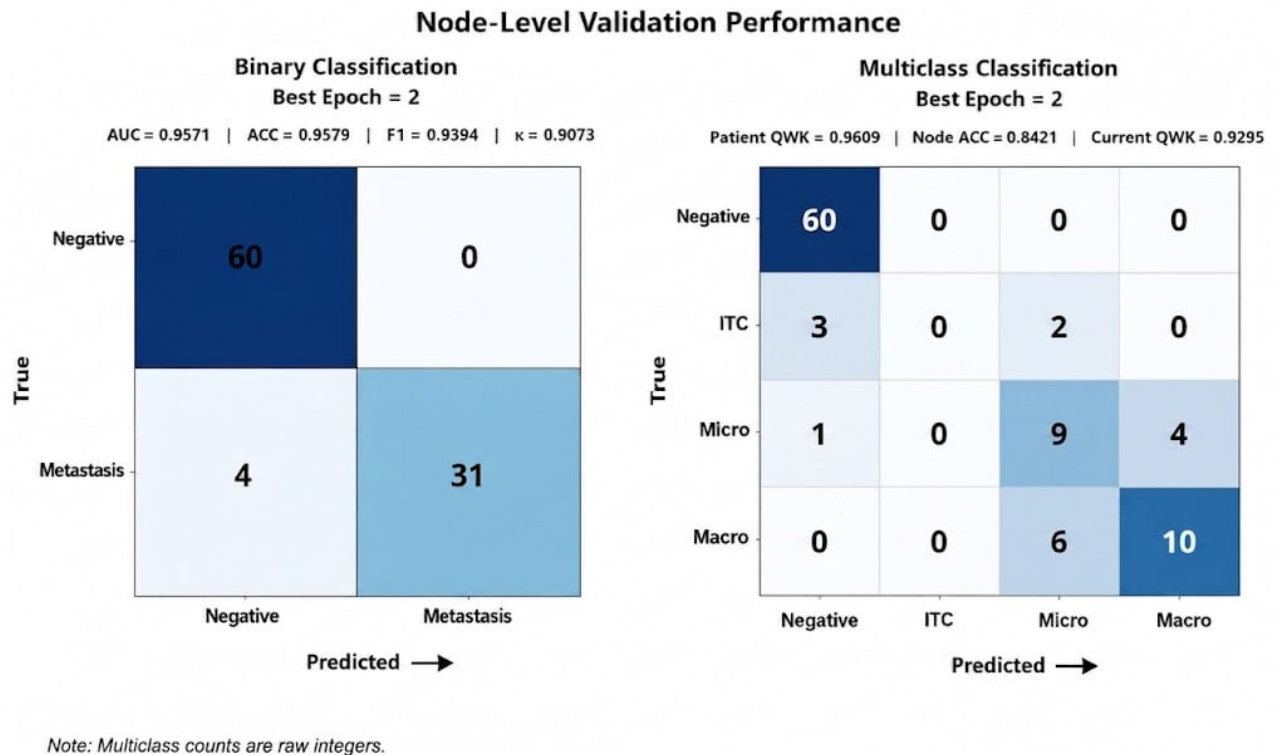


Рисунок 3.6 – Матриці невідповідностей на валідаційній вибірці для CLAM MIL: бінарна класифікація (AUC=0.9571, ліворуч) та мультикласове стадіювання (ACC=0.8421, QWK=0.9609, праворуч). Основні помилки у мультикласовій задачі зосереджені між класами мікро та макро, що відповідає їх морфологічній схожості

Instance-level supervision у CLAM, що базується на псевдомітках, згенерованих із оцінок уваги, виявляється особливо ефективним саме для порядкової класифікації. На відміну від бінарної задачі, де достатньо визначити наявність патологічної тканини, стадіювання вимагає розрізнення між ITC (ізольовані пухлинні клітини, < 0.2 мм), мікрометастазами (0.2–2.0 мм) та макрометастазами (> 2.0 мм). У цьому контексті механізм контролю на рівні екземплярів дозволяє моделі виділяти більш тонкі морфологічні ознаки, що відповідають різним стадіям прогресії захворювання, формуючи більш структуровані карти уваги.

CLAM досягає найкращого значення QWK вже на другій епі (QWK = 0.9609), після чого спостерігається поступове погіршення якості через перена-

вчання. Це вказує на те, що псевдомітки екземплярів забезпечують сильний навчальний сигнал на ранніх етапах, однак можуть призводити до деградації узагальнювальної здатності при тривалому навчанні.

Методи TransMIL та DSMIL демонструють однакову *node accuracy* (0.8737), однак поступаються CLAM за QWK (0.9443 та 0.9429 відповідно). Така різниця між метриками свідчить про характер помилок: ці моделі частіше помиляються між віддаленими класами (наприклад, *negative vs macro*), тоді як CLAM краще зберігає порядкову структуру між суміжними стадіями, що є більш важливим з клінічної точки зору.

Метод ABMIL демонструє найнижчий показник QWK = 0.9413 серед усіх розглянутих підходів, однак досягає стабільного результату лише на 10-й епосі, що свідчить про повільнішу збіжність у порівнянні з іншими архітектурами в умовах мультикласової задачі.

Варто зазначити, що всі методи демонструють суттєво вищу *node accuracy* на негативному класі порівняно з рідкісними класами ІТС та мікрометастазів, що є типовою проблемою дисбалансу класів у задачах гістопатологічного аналізу [14]. Зокрема, клас ІТС, який містить лише 5 прикладів у валідаційній вибірці, є найбільш складним для коректної ідентифікації через мінімальну площу патологічних змін та слабо виражені морфологічні ознаки.

Для додаткової оцінки якості вивчених представлень побудовано UMAP-проекцію векторних представлень фрагментів тренувальної вибірки (рис. 3.7). Проекція демонструє часткове відокремлення класів у просторі ознак: фрагменти класу *negative* утворюють компактні кластери, тоді як патологічні класи (ІТС, *micro*, *macro*) займають суміжні зони з частковим перекриттям, що відповідає їх морфологічній схожості.

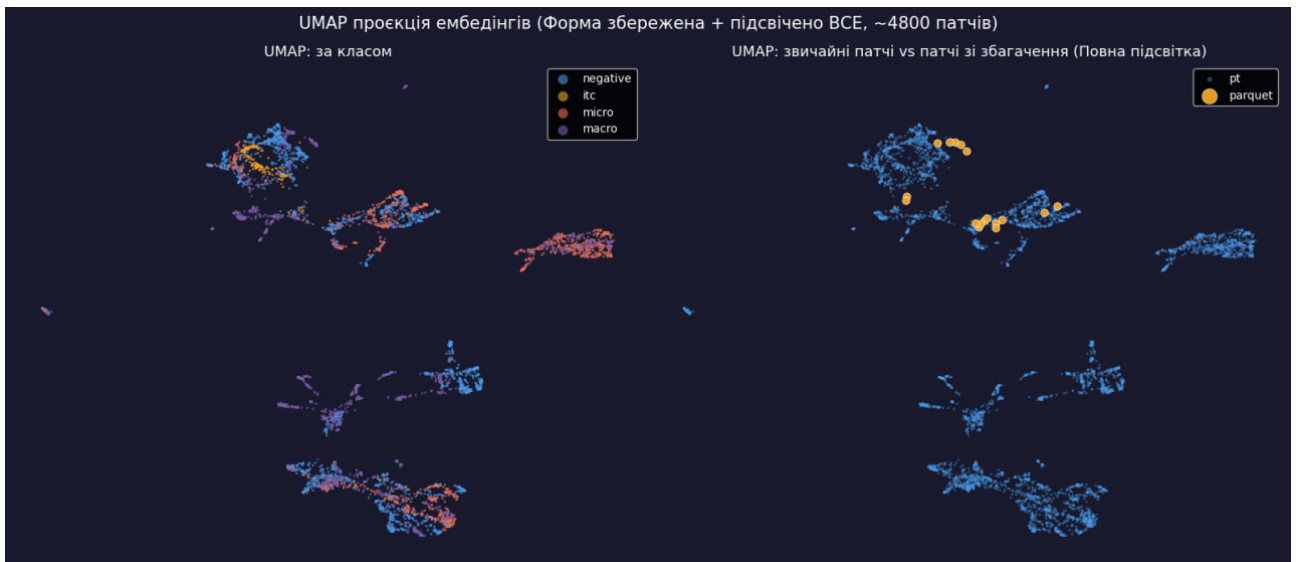


Рисунок 3.7 – UMAP-проекція векторних представлень фрагментів тренувальної вибірки (≈ 4800 фрагментів): за класом метастазування (ліворуч) та за джерелом фрагментів – стандартні .pt проти збагачених parquet-фрагментів (праворуч) [47]

3.4.3 Порівняльний аналіз методів

Аналіз результатів обох задач у сукупності дозволяє сформулювати декілька ключових спостережень щодо поведінки MIL-методів у різних постановках задачі.

По-перше, жоден із розглянутих методів не є універсально найкращим одночасно для бінарної та мультикласової класифікації. Метод DSMIL демонструє найвищу ефективність у задачі виявлення метастазів, тоді як CLAM забезпечує найкращі результати у задачі стадіювання. Цей результат підтверджує, що вибір MIL-агрегатора має безпосередньо залежати від клінічної постановки задачі та характеру необхідної дискримінації ознак.

По-друге, усі методи демонструють значно більшу варіативність результатів у мультикласовій задачі (QWK у діапазоні 0.941–0.961) порівняно з бінарною

постановкою (AUC у діапазоні 0.957–0.977). Таке розширення розкиду значень відображає об’єктивно вищу складність задачі стадіювання, яка вимагає не лише виявлення патології, але й її тонкої морфологічної диференціації між суміжними класами.

По-третє, якість представлень, отриманих за допомогою UNI2-h, забезпечує високий базовий рівень для всіх досліджених методів. Навіть найпростіший підхід АВМІЛ демонструє конкурентні результати у порівнянні з більш складними архітектурами, що підтверджує визначальну роль якості енкодера ознак у МІЛ-пайплайні та його вплив на кінцеву ефективність класифікації.

3.4.4 Стадіювання на рівні пацієнта

Задача стадіювання на рівні вузла є проміжним кроком у клінічному пайплайні. Кінцевою діагностичною метою є визначення рN-стадії пацієнта відповідно до системи TNM [37], що потребує агрегації передбачень по всіх досліджених лімфатичних вузлах одного пацієнта.

Розглянемо правила агрегації за вищезгаданою системою. Нехай для i -го пацієнта отримано множину node-level передбачень $\{\hat{y}_{ij}\}_{j=1}^{n_i}$, де $\hat{y}_{ij} \in \{\text{negative, itc, micro, macro}\}$. рN-стадія пацієнта визначається детермінованими правилами TNM:

$$\hat{s}_i = \begin{cases} \text{pN2} & \text{якщо } |\{j : \hat{y}_{ij} = \text{macro}\}| \geq 4, \\ \text{pN1} & \text{якщо } |\{j : \hat{y}_{ij} = \text{macro}\}| \geq 1, \\ \text{pN1mi} & \text{якщо } |\{j : \hat{y}_{ij} = \text{micro}\}| \geq 1, \\ \text{pN0(i+)} & \text{якщо } |\{j : \hat{y}_{ij} = \text{itc}\}| \geq 1, \\ \text{pN0} & \text{інакше.} \end{cases} \quad (3.2)$$

Правила (3.2) не потребують додаткового навчання та є повністю детермінованими: кожне рішення на рівні пацієнта безпосередньо виводиться з передбачень окремих вузлів, що забезпечує повну інтерпретованість пайплайну.

Оцінювання проводилось на 19 пацієнтах валідаційної вибірки CAMELYON17 з розбиттям на рівні пацієнтів (підрозділ 3.3.1). Кількісні результати представлено у таблиці 3.4, розподіл pN-стадій — на рис. 3.8, матрицю невідповідностей — на рис. 3.9.

Таблиця 3.4 — Результати стадіювання на рівні пацієнта (CAMELYON17, $n = 19$)

Метрика	Значення
Patient-level QWK	0,9577
Patient-level Accuracy	0,8421
Кількість пацієнтів	19

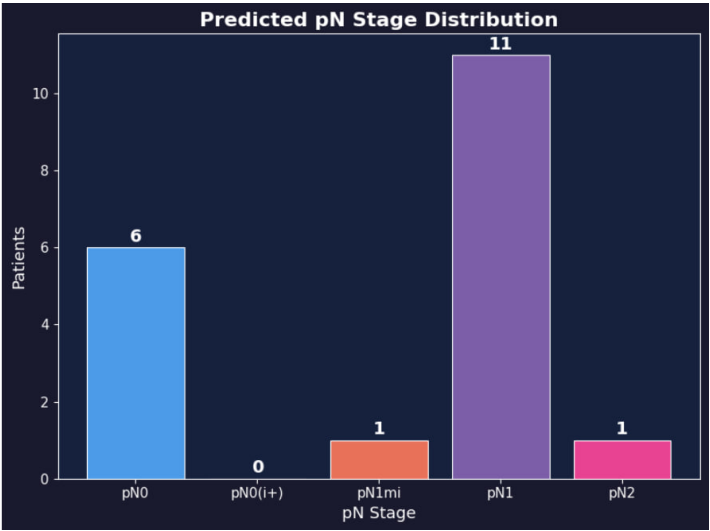


Рисунок 3.8 – Розподіл передбачених pN-стадій на рівні пацієнта (CAMELYON17, $n = 19$)

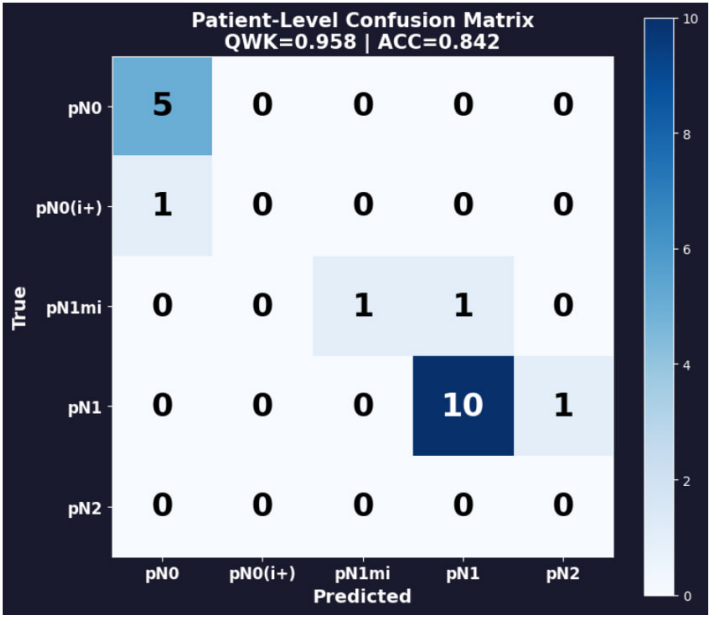


Рисунок 3.9 – Матриця невідповідностей на рівні пацієнта (CAMELYON17, $n = 19$, QWK=0,9577, ACC=0,8421)

Матриця невідповідностей (рис. 3.9) демонструє, що основні помилки зосереджені на межі класів pN0 та pN1mi, що відповідає клінічно складним випадкам мікрометастазів із малою кількістю уражених вузлів. Окремо варто відзначити

присутність у тестовій вибірці одного пацієнта з класом pN0(i+): ізольовані пухлинні клітини без клінічно значущого метастазування є граничним випадком між pN0 та pN1mi і є джерелом систематичної похибки в умовах слабкої розмітки на рівні слайду.

Досягнутий показник $QWK = 0,9577$ на рівні пацієнта підтверджує, що дво-етапний пайплайн — attention-based MIL для node-level класифікації з подальшою rule-based TNM-агрегацією — забезпечує клінічно релевантне стадіювання в умовах слабкої розмітки.

3.4.5 Порівняння з аналогічними системами на CAMELYON

Для контекстуалізації отриманих результатів у таблиці 3.5 показано порівняння з опублікованими системами на датасеті CAMELYON17. Пряме порівняння є умовним: методи з лідерборду оцінювались на офіційному тест-сеті ($n = 100$ пацієнтів), тоді як результати даної роботи отримано на валідаційній вибірці ($n = 19$ пацієнтів), сформованій із тренувального розподілу CAMELYON17. Крім того, більшість конкурентних методів використовують детальну pixel-level розмітку для навчання детекторів метастазів, тоді як запропонований пайплайн працює виключно в режимі слабкої розмітки на рівні слайду [10; 12].

Таблиця 3.5 — Порівняння з аналогічними системами на CAMELYON17
(pN-стадіювання на рівні пацієнта, метрика QWK)

<i>Метод</i>	<i>Розмітка</i>	<i>Тест-сет</i>	<i>QWK</i>
Lunit Inc. [13]	pixel-level	офіційний ($n=100$)	0.9203
HMS-MGH-CCDS [13]	pixel-level	офіційний ($n=100$)	0.8958
DeepBio [13]	pixel-level	офіційний ($n=100$)	0.8794
RF-Ensemble [48]	pixel-level	офіційний ($n=100$)	0.9090
Топ лідерборду [49]	—	офіційний ($n=100$)	0.9570
CLAM + UNI2-H (ця робота)	slide-level	валідаційний ($n=19$)	0.9577

Запропонований пайплайн досягає $QWK=0,9577$ на рівні пацієнта, що є порівнянним із найкращим актуальним результатом публічного лідерборду CAMELYON17 ($QWK=0,957$ [49]). Водночас методи лідерборду, як правило, спираються на детальну pixel-level розмітку та спеціалізовані модулі детекції метастазів. Натомість запропонований підхід досягає конкурентного результату виключно на основі міток на рівні слайду, що підкреслює практичну перевагу weakly supervised парадигми з точки зору вартості розмітки та масштабованості на нові клінічні домени.

3.5 Архітектура та інфраструктура розгортання вебдодатка інтелектуальної системи

Для практичної демонстрації розроблених методів та забезпечення можливості інтеграції створеної інтелектуальної системи у реальний клінічний процес

було спроектовано та реалізовано вебдодаток на базі спеціалізованого фреймворку Streamlit [50]. Головною метою розробки цього інтерфейсу є надання користувачеві (лікарю-патоморфологу) інтуїтивно зрозумілого інструментарію для взаємодії з MIP-моделями без необхідності безпосередньої роботи з програмним кодом.

Вибір Streamlit як технологічної основи обумовлений прагматичними міркуваннями: фреймворк дозволяє будувати інтерактивний вебінтерфейс безпосередньо на Python без залучення окремого фронтенд-стеку, що суттєво скорочує час розробки прототипу. Порівняно з альтернативними підходами — зокрема Gradio або повноцінним стеком FastAPI + React — Streamlit забезпечує достатній рівень інтерактивності для демонстраційних цілей без ускладнення архітектури [15]. Такий підхід відповідає сучасній практиці швидкого прототипування клінічних AI-інструментів на початкових стадіях розробки, де пріоритетом є верифікація концепції, а не повноцінне промислове розгортання.

З огляду на значні обчислювальні вимоги патологічної фундаментальної моделі UNI2-H та алгоритмів агрегації CLAM, вебдодаток було розгорнуто за клієнт-серверною архітектурною схемою на базі виділеного сервера. Обчислювальний конвеєр системи оптимізовано під архітектуру графічного процесора NVIDIA GeForce RTX 3090 [51] із 24 ГБ відеопам'яті типу GDDR6X, що дало змогу забезпечити паралельний інференс та стабільну обробку гігапксельних зображень цілого зрізу (WSI) в інтерактивному режимі. Для читання багаторівневих файлів форматів SVS та TIFF використовується бібліотека OpenSlide [9], яка забезпечує vendor-neutral доступ до пірамідальних WSI-форматів різних виробників сканерів. Головне вікно розробленого користувацького інтерфейсу на етапі ініціалізації аналізу представлено на рис. 3.10.

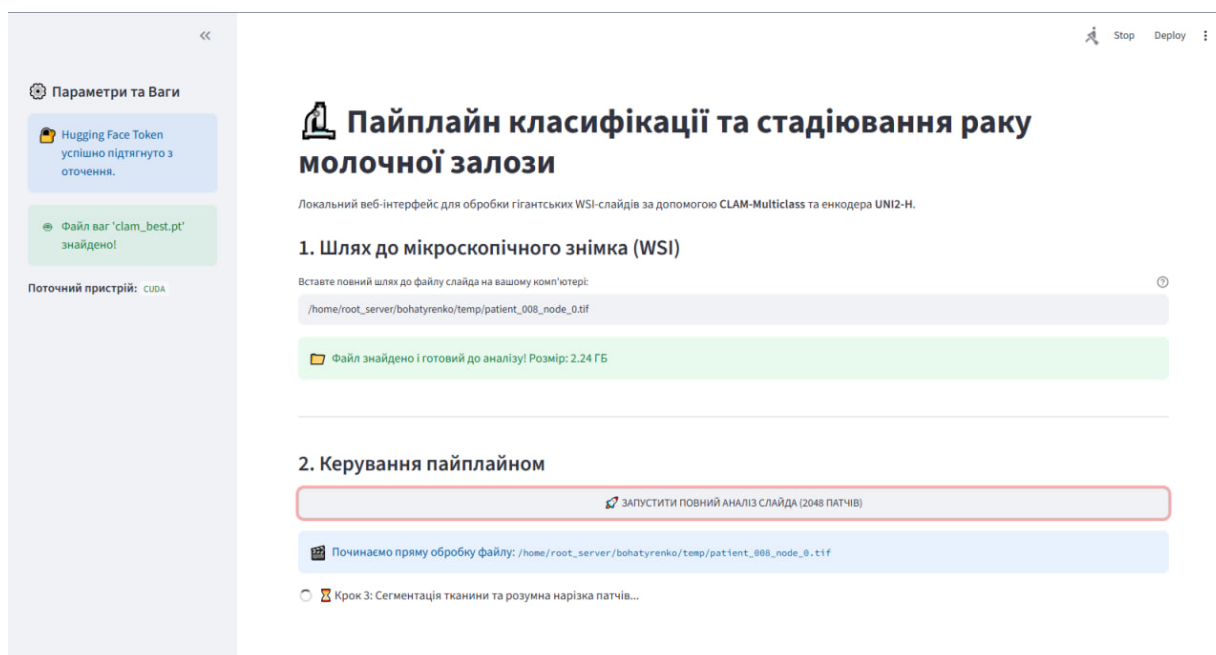


Рисунок 3.10 – Головне вікно інтерфейсу вебдодатка інтелектуальної системи на етапі завантаження файлу WSI та запуску пайплайну обробки

Архітектура розробленого вебдодатка складається з таких ключових взаємозв'язаних компонентів:

- модуль керування сховищем даних, що відповідає за зв'язок із локальною базою даних сайту додатка та структуроване підвантаження файлів форматів SVS або TIFF засобами OpenSlide [9];
- обчислювальне ядро (Inference Engine), яке виконує асинхронну сегментацію тканини, екстракцію ознак за допомогою трансформера UNI2-H та фінальне мультикласове стадіювання [30];
- апаратний прискорювач на базі графічного процесора RTX 3090, що виконує тензорні обчислення з використанням технології CUDA та змішаної точності (Mixed Precision FP16) [41];
- інтерактивний інтерфейс користувача Streamlit, який відображає результати класифікації та побудовані карти уваги (Attention Maps) для інтерпретації прийнятих рішень.

Процес взаємодії з вебдодатком під час проведення інференсу організовано за таким сценарієм. Користувач через графічний інтерфейс задає повний шлях до цифрового слайда, що зберігається на сервері. Програмне забезпечення автоматично валідує файл WSI (як показано на рис. 3.10, на прикладі файлу розміром 2,24 ГБ) та ініціює асинхронний виклик до обчислювального ядра, запускаючи нарізку на фрагменти і передачу сформованих тензорів у пам'ять графічного прискорювача.

Завдяки високій пропускній здатності архітектури Ampere, на якій побудовано RTX 3090, час виконання повного циклу інференсу для одного слайду скоротився до прийняттого для клінічної практики інтервалу. Обчислювальна складність часу обробки одного запиту T_{inf} для всього гігапксельного зображення за умов апаратного прискорення визначається виразом:

$$T_{\text{inf}} \approx t_{\text{seg}} + \frac{N \cdot t_{\text{enc}}}{B} + t_{\text{clam}},$$

де t_{seg} — час порогової сегментації тканини та нарізки слайду; N — загальна кількість відфільтрованих патчів; t_{enc} — час прямого проходу (forward pass) одного патча крізь кодувальник UNI2-H; B — розмір батчу обробки на GPU ($B = 64$ для RTX 3090); t_{clam} — час агрегації ознак та генерації теплової карти моделлю CLAM.

Після завершення обчислень вебдодаток динамічно оновлює сторінку та виводить блок фінальних результатів аналізу, який містить діагностовану стадію раку, впевненість моделі (Confidence), загальний ймовірнісний розподіл за всіма чотирма класами та інтерактивну карту уваги (рис. 3.11).

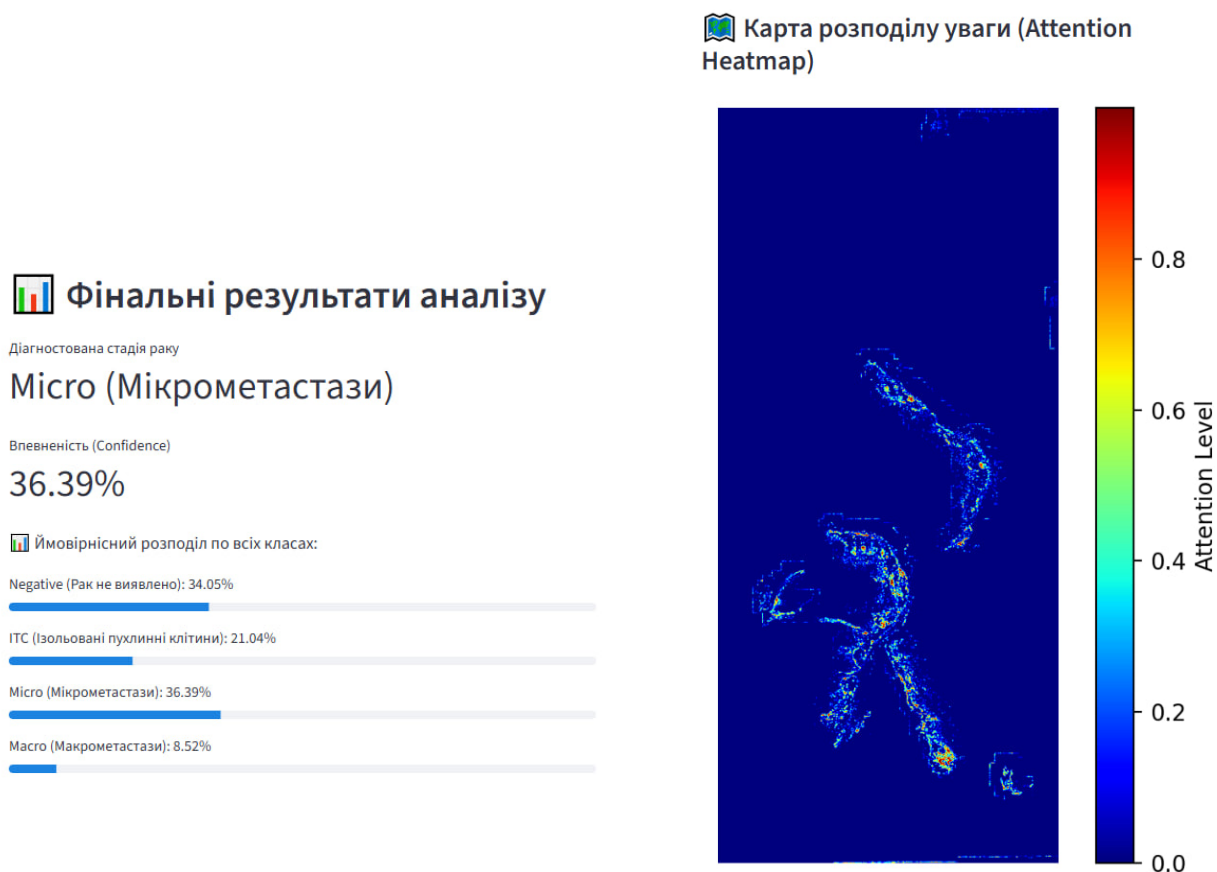


Рисунок 3.11 – Візуалізація вихідних даних вебдодатка: ліворуч — ймовірнісний розподіл стадіювання метастазів у лімфовузлах; праворуч — згенерована моделлю CLAM карта розподілу уваги (Attention Heatmap)

Як демонструє рис. 3.11, для тестового зрізу система зафіксувала категорію «Micro (Мікрометастази)» із рівнем впевненості 36,39%, детально відобразивши внесок інших стадій. Згенерована теплова карта (рис. 3.11, праворуч) наочно підсвічує патологічні ділянки (значення індексу уваги наближається до 1,0), що дозволяє лікарю швидко верифікувати ділянки мікрометастазування під час мікроскопічного аналізу.

Відображення карти уваги є ключовим елементом інтерпретованості системи з точки зору концепції пояснюваного штучного інтелекту (Explainable AI, XAI) [34]. На відміну від підходів на основі градієнтної локалізації [52], механізм уваги CLAM формує просторово інтерпретовані ваги безпосередньо в процесі MIL-агрегації, що усуває необхідність у постобробці та забезпечує узгодженість

між класифікаційним рішенням і відображеними ділянками. Це дозволяє лікарю не лише отримати автоматичний прогноз стадії, а й оцінити морфологічні ділянки, на основі яких цей прогноз сформовано [10], що є важливою передумовою для практичного застосування AI-систем у клінічній патоморфології [14].

3.6 Результати на Melanoma-WSI

3.6.1 Порівняння методів MIL на Melanoma-WSI

Незважаючи на тематичну відмінність Melanoma-WSI від основного датасету дослідження, використання цього набору даних дозволяє оцінити здатність розробленого пайплайну до міждоменного узагальнення на іншій гістологічній задачі, що є важливим показником його практичної застосовності. З цією метою всі чотири розглянуті MIL-архітектури було навчено та протестовано на Melanoma-WSI у задачі бінарної класифікації (melanoma проти benign). Для забезпечення коректності порівняння використовувалися ті самі умови навчання, що й у підрозділі 3.3.1: заморожений енкодер UNI2-h, уніфіковані гіперпараметри оптимізації та однакова стратегія ранньої зупинки. Отримані результати представлено в таблиці 3.6.

Таблиця 3.6 — Результати бінарної класифікації (melanoma / benign) на Melanoma-WSI

<i>Метод</i>	<i>AUC</i>	<i>F1</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>Kappa</i>
CLAM	0.9495	0.9829	0.973	0.9663	1.000	0.9191
ABMIL	0.9642	0.9773	0.964	0.9556	1.000	0.8905
DSMIL	0.9723	0.9829	0.973	0.9663	1.000	0.9191
TransMIL	0.9884	0.9718	0.955	0.9451	1.000	0.8611

Усі методи демонструють $\text{Recall} = 1.0$ на класі melanoma, що є пріоритетним з клінічної точки зору: жоден меланомний препарат не було пропущено. Водночас методи суттєво відрізняються за AUC та точністю передбачення позитивного класу.

Найвищий показник AUC досягає TransMIL (0.9884), що свідчить про кращу дискримінацію класів у просторі оцінок впевненості. Трансформерний механізм самоуваги з Nyström-апроксимацією дозволяє моделювати міжпатчеві залежності, що є корисним при аналізі меланомних препаратів із дифузним розподілом атипових клітин. Проте вищий AUC супроводжується нижчою Accuracy (0.955) та Kappa (0.8611) порівняно з CLAM та DSMIL, що вказує на гірший вибір порогу класифікації при максимізації розділення класів.

DSMIL та CLAM демонструють однакові F1 (0.9829), Accuracy (0.973) та Kappa (0.9191) — найвищі серед усіх методів за цими метриками, однак DSMIL перевищує CLAM за AUC (0.9723 проти 0.9495). Нижчий AUC CLAM на Melanoma-WSI узгоджується з поведінкою, спостереженою на CAMELYON: модель швидко досягає хорошого порогового рішення, однак менш впевнено розподіляє оцінки вздовж усієї ROC-кривої, що є наслідком раннього перенавчання через нестабільність псевдоміток на початкових епохах. ABMIL забезпечує проміжний результат ($\text{AUC} = 0.9642$), зберігаючи перевагу простої та стабільної структури агрегації.

Порівнюючи результати на Melanoma-WSI із результатами на CAMELYON16+17 (табл. 3.2), слід відзначити збереження відносного порядку між методами за AUC: TransMIL залишається лідером за дискримінацією, тоді як CLAM та DSMIL домінують за пороговими метриками. Це підтверджує стійкість виявлених закономірностей між датасетами різної природи та узагальнюваність висновків щодо поведінки MIL-агрегаторів.

Для детальнішого аналізу інтерпретованості у наступних підрозділах розглядається модель CLAM як архітектура з явним механізмом instance-level supervision, що забезпечує найбільш змістовні карти уваги для патоморфологічної інтерпретації.

3.6.2 Якість класифікації та ROC-аналіз

Детальний аналіз якості класифікації проводиться для моделі CLAM як архітектури з найвищими пороговими метриками (Accuracy = 0.973, Kappa = 0.9191) та найбільш інтерпретованим механізмом уваги. На рис. 3.12 показано матрицю невідповідностей та ROC-криву для найкращої моделі, отриманої за результатами валідації.

Поріг класифікації визначався методом повного перебору по сітці з 401 значення у діапазоні $[0, 1]$ з максимізацією загальної точності на валідаційному наборі. Оптимальна робоча точка визначалась за індексом Юдена:

$$J = \text{TPR} - \text{FPR}, \quad (3.3)$$

де TPR — чутливість (частка правильно виявлених меланомних препаратів), FPR — частка хибнопозитивних передбачень серед доброякісних зразків.

Матриця невідповідностей дозволяє оцінити баланс між чутливістю та специфічністю моделі. У медичному контексті хибнонегативний прогноз (пропущена меланома) є значно небезпечнішим за хибнопозитивний, що обґрунтовує вибір порогу з акцентом на максимізацію Recall.

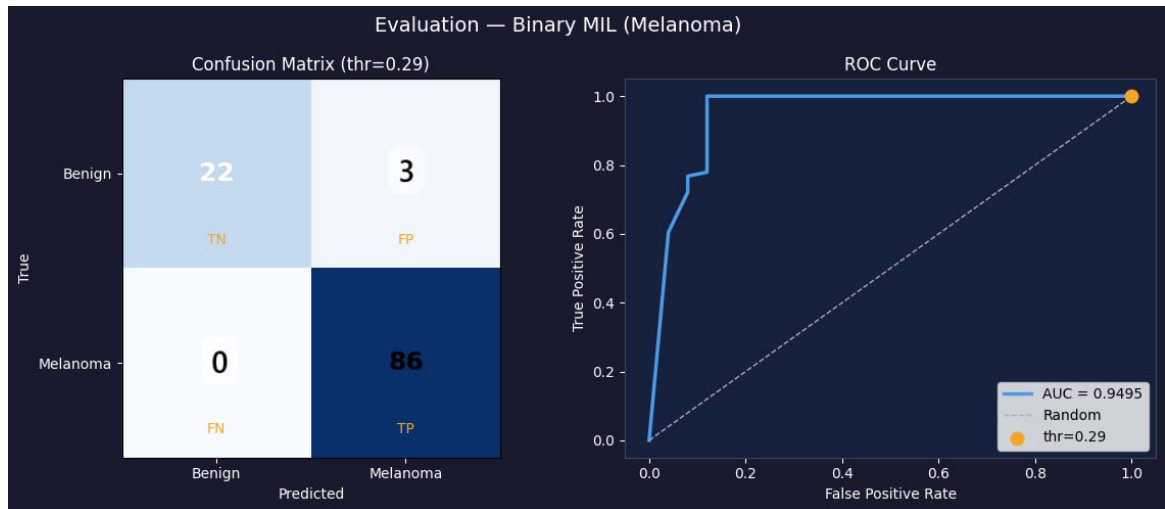


Рисунок 3.12 – Оцінка бінарного MIL-класифікатора CLAM на валідаційній вибірці Melanoma-WSI: (ліворуч) матриця невідповідностей при оптимальному порозі $t = 0.29$; (праворуч) ROC-крива з позначенням оптимальної робочої точки за індексом Юдена (3.3)

3.6.3 Аналіз карт уваги та візуалізація простору ознак

Для інтерпретації рішень моделі проведено аналіз просторового розподілу ваг уваги $\alpha_i = \text{softmax}(a_i)$ на рівні окремих фрагментів згідно з виразом (2.1). На рис. 3.13 відображено карти уваги для репрезентативних препаратів валідаційного набору.



Рисунок 3.13 – Розподіл ваг уваги по фрагментах для репрезентативних препаратів: верхній рядок — меланомні, нижній — доброякісні. Золотим кольором виділено топ-5 фрагментів з найбільшою вагою уваги

Для меланомних препаратів розподіл уваги є концентрованим: невелика частина фрагментів отримує домінуючі ваги, що узгоджується з припущенням МІЛ про локальність патологічного сигналу. Фрагменти з найвищою увагою, як правило, відповідають ділянкам із вираженою ядерною атипією, підвищеним мітотичним індексом та порушенням тканинної архітектури — морфологічними ознаками, що є діагностично значущими при верифікації меланому. У випадку доброякісних зразків розподіл ваг є рівномірнішим, що відображає відсутність локалізованих пухлинних вогнищ.

Для кількісної оцінки якості сформованих представлень побудовано UMAP-проекцію прихованого простору ознак після проєкційного шару $\mathbf{W}_1 \in \mathbb{R}^{D \times 256}$ (рис. 3.14). Для кожного препарату відібрано $K = 30$ фрагментів із найви-

щими значеннями уваги, що дозволяє зосередитись на найбільш інформативних ділянках тканини.

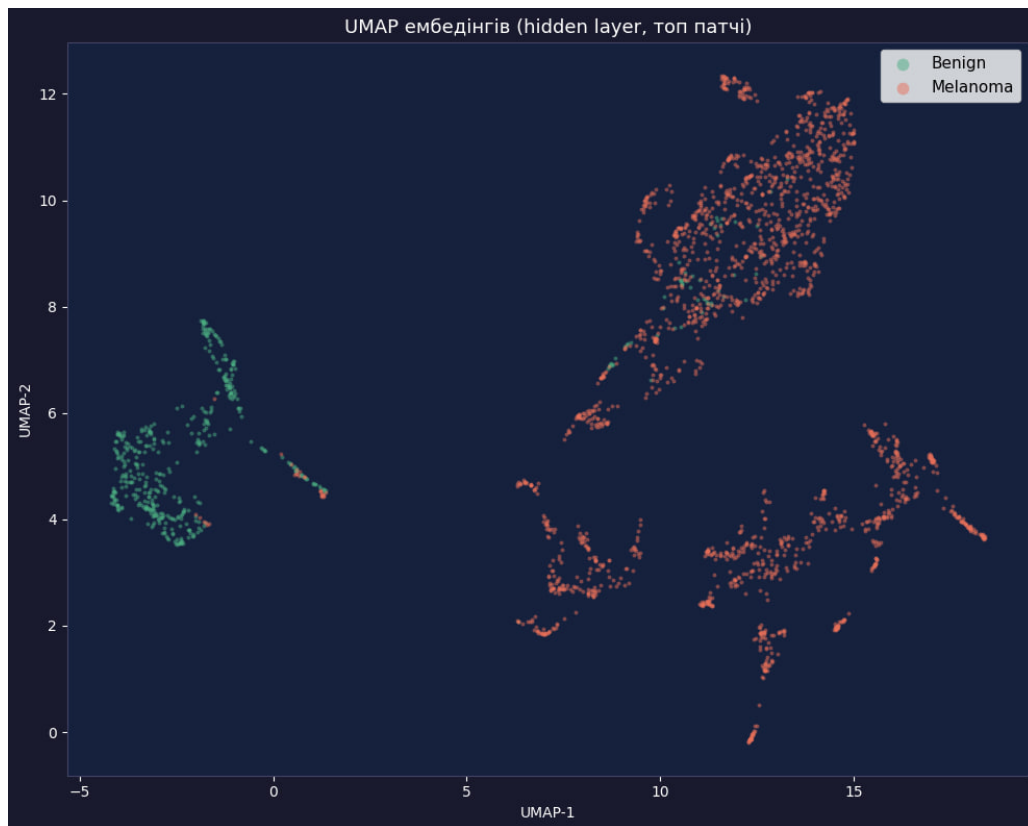


Рисунок 3.14 – UMAP-проекція простору ознак після проєкційного шару моделі CLAM. Кольором позначено діагностичний клас фрагментів (червоний — меланома, зелений — доброякісні); розмір точок корелює з вагою уваги α_i

Проекція демонструє формування двох чітко відокремлених кластерів: фрагменти меланомних препаратів концентруються в одній компактній області простору ознак, тоді як доброякісні утворюють окрему групу. Незначне перекриття на межі кластерів відповідає фрагментам із низькою впевненістю моделі і є характерним для граничних випадків із неоднозначною морфологією. Така структура простору ознак підтверджує, що проєкційний шар CLAM формує лінійно розділювані представлення на основі заморожених ембедінгів UNI2-h, що є необхідною умовою для стабільної роботи attention-агрегатора.

3.6.4 Порівняння з аналогічними системами на Melanoma-WSI

Датасет Melanoma-WSI (CMB-MEL + CPTAC_CM) є авторською тестовою вибіркою, сформованою з публічних джерел у рамках цієї роботи, тому пряме числове порівняння з опублікованими системами на ідентичному тест-сеті неможливе. Для контекстуалізації результатів у таблиці 3.7 представлено два типи порівнянь: (1) опубліковані методи класифікації меланоми на WSI у режимі слабкої розмітки на зовнішніх датасетах та (2) результати усіх чотирьох MIL-архітектур, оцінених авторами на Melanoma-WSI в однакових умовах із замороженим енкодером UNI2-H.

Серед опублікованих робіт найближчою за постановкою задачі є робота [53], де weakly supervised класифікація меланоми проти не-меланоми на WSI-рівні (EfficientNetB1 + max-pooling агрегація) досягає $AUC = 0,821$ на тест-сеті з 90 слайдів. Робота [54] систематично порівнює шість MIL-фреймворків (ABMIL, CLAM, DSMIL, TransMIL та інші) безпосередньо на меланомних WSI для класифікації імунних підтипів на датасеті Leeds Melanoma Cohort ($n=667$), досягаючи найкращого $AUC = 0,80$ з pathology-specific SSL-енкодером. Більш потужний фреймворк Melan-Dx [55] досягає ROC $AUC = 0,915$ у few-shot режимі для класифікації меланоцитарних уражень, однак спирається на спеціалізовану knowledge-enhanced архітектуру та приватний датасет Університету Пенсильванії. Робота [56] застосовує weakly supervised TransMIL із енкодером Prov-GigaPath на 426 первинних меланомних WSI для прогнозування метастатичного ризику, досягаючи $AUC = 0,883$.

Таблиця 3.7 — Порівняння з аналогічними системами для класифікації меланому на WSI (метрика AUC)

<i>Метод</i>	<i>Encoder</i>	<i>Датасет</i>	<i>Задача</i>	<i>AUC</i>
<i>Опубліковані роботи (зовнішні датасети)</i>				
EfficientNetB1 + max-pool [53]	CNN (ImageNet)	приватний ($n=90$)	бінарна	0.821
DSMIL + SSL [54]	ResNet50 (SSL)	LMC ($n=667$)	імунний підтип	0.800
TransMIL [56]	Prov-GigaPath	приватний ($n=426$)	метастат. ризик	0.883
CONCH + ABMIL [55]	CONCH	приватний (UPenn)	бінарна	0.883
Melan-Dx + MUSK [55]	MUSK	приватний (UPenn)	бінарна	0.915
<i>Ця робота</i>				
ABMIL [12]	UNI2-H	Melanoma-WSI	бінарна	0.9642
DSMIL [44]	UNI2-H	Melanoma-WSI	бінарна	0.9723
CLAM [10]	UNI2-H	Melanoma-WSI	бінарна	0.9495
TransMIL [18]	UNI2-H	Melanoma-WSI	бінарна	0.9884

Усі чотири методи, навчені на Melanoma-WSI із замороженим енкодером UNI2-H, суттєво перевищують опубліковані weakly supervised підходи зі спеціалізованими pathology foundation models на задачах класифікації меланому: TransMIL+Prov-GigaPath (AUC=0,883 [56]), CONCH+ABMIL (AUC=0,883 [55]), а також MIL-методи з SSL-енкодером безпосередньо на

меланомних WSI ($AUC = 0,80$ [54]). Результати є конкурентними щодо Melan-Dx + MUSK ($AUC = 0,915$), який використовує knowledge-enhanced архітектуру з retrieval-augmented базою знань [55]. Найвищий показник дискримінації досягає TransMIL ($AUC = 0,988$), тоді як CLAM та DSMIL забезпечують кращий баланс між чутливістю та специфічністю ($F1 = 0,983$, $Kappa = 0,919$). Це підтверджує, що заморожений енкодер UNI2-H формує достатньо потужні patch-рівневі представлення для ефективної slide-level класифікації меланоми без будь-яких патч-рівневих анотацій.

Висновки до розділу 3

У третьому розділі представлено повний експериментальний пайплайн — від попередньої обробки whole slide images (WSI) до порівняльного аналізу методів MIL на двох незалежних датасетах.

Розроблений пайплайн попередньої обробки забезпечує ефективну фільтрацію неінформативних ділянок за рахунок порогової сегментації тканини та формування bag із патчів розміром 224×224 пікселів. Застосування механізму hard-patch mining дозволяє динамічно відбирати найбільш інформативні патчі під час навчання, що є критично важливим для задач стадіювання, де патологічно значущі області займають малу частку слайду.

Порівняння чотирьох MIL-методів на CAMELYON16+17 виявило принципову різницю між задачами виявлення та стадіювання метастазів. У бінарній класифікації найкращі результати показав DSMIL ($AUC = 0.9767$, $F1 = 0.9855$), тоді як для мультикласового стадіювання найвищий результат досягнуто для CLAM ($QWK = 0.9609$). Отримані результати вказують, що вибір архітектури MIL-агрегатора повинен узгоджуватись із клінічною природою задачі: dual-stream підхід DSMIL краще відповідає задачі виявлення локальних метастати-

чних осередків, тоді як instance-level supervision у CLAM є більш придатним для розрізнення тонких морфологічних відмінностей між стадіями.

На датасеті Melanoma-WSI усі чотири методи підтвердили узагальнюваність підходу: Recall = 1.0 для класу melanoma досягається незалежно від архітектури агрегатора, тоді як за AUC лідирує TransMIL (0.9884), а за пороговими метриками — CLAM та DSMIL ($F1 = 0.9829$). UMAP-проекція прихованого простору ознак моделі CLAM демонструє чітке розділення класів melanoma та benign, що підтверджує лінійну розділюваність представлень UNI2-h після навчання проєкційного шару.

Узагальнюючи, результати підтверджують, що заморожений енкодер UNI2-h у поєднанні з attention-based MIL-агрегацією забезпечує стабільно високу якість класифікації на обох наборах даних, тоді як оптимальний вибір агрегатора визначається специфікою клінічної задачі.

ВИСНОВКИ

Дослідження спрямоване на вирішення практично значущої задачі автоматизованого стадіювання раку молочної залози за гістологічними зображеннями цілого зрізу в умовах обмеженої розмітки даних. Робота успішно продемонструвала, що поєднання сучасних фундаментальних моделей цифрової патології з парадигмою Multiple Instance Learning дозволяє будувати клінічно релевантні системи класифікації без залучення трудомісткої локальної анотації на рівні патчів або пікселів.

Обґрунтовано вибір підходу Multiple Instance Learning як єдиної практично масштабованої парадигми для аналізу гігапіксельних WSI в умовах реальної клінічної практики, де доступні лише slide-level мітки. Розроблений пайплайн попередньої обробки забезпечує ефективну фільтрацію нерелевантних ділянок препарату та формування інформативних bag-представлень через дворівневий механізм відбору патчів: статичне збагачення верифікованими анотаційними регіонами та динамічний semi-hard patch mining починаючи з певної епохи навчання. Така комбінація знижує ризик перенавчання на шумових ділянках і дозволяє моделі зосередитись на діагностично значущих морфологічних структурах навіть при обмеженій кількості препаратів.

Використання замороженого енкодера UNI2-H на базі ViT-Huge, попередньо навченого методом DINOv2 на понад ста тисячах гістологічних препаратів, забезпечило якісну основу векторних представлень для всіх досліджених MIL-агрегаторів. Примітно, що навіть найпростіша архітектура ABMIL демонструє конкурентні результати завдяки високій інформативності ембедінгів UNI2-H, що підтверджує визначальну роль якості енкодера у загальній ефективності пайплайну порівняно з вибором конкретної стратегії агрегації.

Порівняльний аналіз чотирьох MIL-архітектур виявив принципову відмінність між задачами виявлення та стадіювання метастазів, яка не є очевидною апріорі. У бінарній класифікації найкращі результати показав DSMIL з $AUC = 0,977$ та $F1 = 0,986$, тоді як для мультикласового стадіювання за категоріями *negative* / *ITC* / *micro* / *macro* найвищу якість забезпечила модель CLAM з $QWK = 0,961$. Цей результат вказує на те, що *instance-level supervision* через псевдомітки є особливо ефективним саме для задач порядкової класифікації, де необхідно розрізняти тонкі морфологічні відмінності між суміжними стадіями, тоді як *dual-stream* підхід DSMIL краще відповідає задачі локалізації окремих метастатичних осередків.

Двоетапний пайплайн стадіювання на рівні пацієнта, що поєднує *attention-based MIL* для *node-level* класифікації з детерміністичною TNM-агрегацією, досяг $QWK = 0,958$ на валідаційній вибірці CAMELYON17. Основні помилки зосереджені на межі класів *pN0* та *pN1mi*, що відповідає клінічно найбільш неоднозначним випадкам мікрометастазів і залишається складним навіть для досвідчених патоморфологів. Результати на датасеті TCGA-SKCM підтвердили здатність пайплайну до міждоменного узагальнення: усі досліджені архітектури досягли $Recall = 1,0$ для класу меланоми, що є пріоритетним з клінічної точки зору критерієм якості скринінгової системи.

Аналіз теплових карт уваги показав, що фрагменти з найвищими вагами просторово узгоджуються з верифікованими метастатичними регіонами без жодного використання локальних анотацій під час навчання. Така властивість є принципово важливою для клінічного застосування, оскільки система надає не лише результат класифікації, а й локалізацію патологічно значущих ділянок, придатну для верифікації спеціалістом.

Таким чином, розроблена система підтверджує практичну доцільність поєднання фундаментальних патологічних моделей з *attention-based MIL*-агрегацією для гістопатологічної класифікації в умовах слабкої розмітки. Отриманий підхід може слугувати інструментом підтримки прийняття рішень для патоморфологів

при скринінгу гістологічних препаратів, пріоритезації складних випадків та стандартизації стадіювання, знижуючи залежність від трудомісткої детальної анотації та розширюючи можливості застосування таких систем у реальній клінічній практиці.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] Hyuna Sung et al. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries”. В: *CA: A Cancer Journal for Clinicians* 71.3 (2021), с. 209—249. DOI: 10.3322/caac.21660.
- [2] World Health Organization. *Breast cancer [Електронний ресурс]*. Режим доступу: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Дата звернення: 27.04.2026. 2026.
- [3] Nadia Harbeck and Michael Gnant. “Breast cancer”. В: *The Lancet* 389.10074 (2017), с. 1134—1150. DOI: 10.1016/S0140-6736(16)31891-8.
- [4] А. І. Шевченко, О. О. Колеснік et al. “Аналіз захворюваності на рак молочної залози в Україні та місті Києві”. В: *Фармацевтичний журнал* (2022). pharmj.org.ua.
- [5] Charles M. Perou et al. “Molecular portraits of human breast tumours”. В: *Nature* 406.6797 (2000), с. 747—752. DOI: 10.1038/35021093.
- [6] Mahul B. Amin, Stephen B. Edge, Frederick L. Greene et al., ред. *AJCC Cancer Staging Manual*. 8th. New York: Springer, 2017.
- [7] Chetan L. Srinidhi, Ozan Ciga and Anne L. Martel. “Deep neural network models for computational histopathology: A survey”. В: *Medical Image Analysis* 67 (2021), с. 101813. DOI: 10.1016/j.media.2020.101813.
- [8] Daisuke Komura and Shumpei Ishikawa. “Machine Learning Methods for Histopathological Image Analysis”. В: *Computational and Structural Biotechnology Journal* 16 (2018), с. 34—42.
- [9] Adam Goode et al. “OpenSlide: A vendor-neutral software foundation for digital pathology”. В: *Journal of Pathological Informatics* 4.1 (2013), с. 27.

- [10] Ming Y. Lu et al. “Data-efficient and weakly supervised computational pathology on whole-slide images”. B: *Nature Biomedical Engineering* 5 (2021), c. 555—570.
- [11] Muhammad Waqas et al. “Exploring Multiple Instance Learning (MIL): A Brief Survey”. B: *Expert Systems with Applications* (2024).
- [12] Maximilian Ilse, Jakub M. Tomczak and Max Welling. “Attention-based Deep Multiple Instance Learning”. B: *ICML*. 2018.
- [13] Peter Bandi et al. “From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge”. B: *IEEE Transactions on Medical Imaging* 38.2 (2019), c. 550—560.
- [14] Gabriele Campanella et al. “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images”. B: *Nature Medicine* 25.8 (2019), c. 1301—1309. DOI: 10.1038/s41591-019-0508-1.
- [15] Jeroen van der Laak, Geert Litjens and Francesco Ciompi. “Deep learning in histopathology: the path to the clinic”. B: *Nature Medicine* 27 (2021), c. 775—784. DOI: 10.1038/s41591-021-01343-4.
- [16] Niccolò Marini et al. “Multi_Scale_Tools: A Python Library to Exploit Multi-Scale Whole Slide Images”. B: *Frontiers in Computer Science* (2021). License: CC BY 4.0. DOI: 10.3389/fcomp.2021.684521.
- [17] Babak Ehteshami Bejnordi et al. “Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer”. B: *JAMA* 318.22 (2017), c. 2199—2210. DOI: 10.1001/jama.2017.14585.
- [18] Zhuchen Shao et al. “TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification”. B: *NeurIPS*. 2021.
- [19] Юрій Петрович Зайченко. *Теорія прийняття рішень*. Київ: НТУУ «КПІ», 2014, с. 412.

- [20] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. B: *Journal of Machine Learning Research* 15 (2014), c. 1929—1958.
- [21] Geert Litjens et al. “A Survey on Deep Learning in Medical Image Analysis”. B: *Medical Image Analysis* 42 (2017), c. 60—88. DOI: 10.1016/j.media.2017.07.005.
- [22] Kaiming He et al. “Deep Residual Learning for Image Recognition”. B: *CVPR*. 2016.
- [23] “Computational Pathology in the Era of Deep Learning: A Review”. B: *Cancers* 13.13 (2021), c. 3208. URL: <https://www.mdpi.com/2072-6694/13/13/3208>.
- [24] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. “Deep learning”. B: *Nature* 521.7553 (2015), c. 436—444. DOI: 10.1038/nature14539.
- [25] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. B: *Advances in Neural Information Processing Systems*. T. 25. 2012.
- [26] Mingxing Tan and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. B: *ICML*. 2019.
- [27] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. B: *ICLR* (2021).
- [28] Ting Chen et al. “A Simple Framework for Contrastive Learning of Visual Representations”. B: *ICML*. 2020.
- [29] Mathilde Caron et al. “Emerging Properties in Self-Supervised Vision Transformers”. B: *ICCV*. 2021.
- [30] Richard J. Chen et al. “A General-Purpose Self-Supervised Model for Computational Pathology”. B: *Nature Medicine* (2024).
- [31] Kaiming He et al. “Masked Autoencoders Are Scalable Vision Learners”. B: *CVPR*. 2022.

- [32] Hanwen Xu et al. “A whole-slide foundation model for digital pathology from real-world data”. B: *Nature* (2024).
- [33] Eric Zimmermann et al. “Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology”. B: (2024). arXiv:2408.00738.
- [34] Erico Tjoa and Cuntai Guan. “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI”. B: *IEEE Transactions on Neural Networks and Learning Systems* 32.11 (2021), c. 4793—4813. DOI: 10.1109/TNNLS.2020.3027314.
- [35] Maxime Oquab, Timothée Darcet, Théo Moutakanni et al. “DINOv2: Learning Robust Visual Features without Supervision”. B: *Transactions on Machine Learning Research* (2023). arXiv:2304.07193.
- [36] Ming Y. Lu et al. “A Visual-Language Foundation Model for Computational Pathology”. B: *Nature Medicine* (2024).
- [37] Leslie H. Sobin, Mary K. Gospodarowicz and Christian Wittekind, ред. *TNM Classification of Malignant Tumours*. 7th. Oxford: Wiley-Blackwell, 2009.
- [38] Cancer Moonshot Biobank. *Cancer Moonshot Biobank – Melanoma Collection (CMB-MEL)*. Bep. 3. Dataset. The Cancer Imaging Archive, 2022. DOI: 10.7937/GWSP-WH72. URL: <https://doi.org/10.7937/GWSP-WH72>.
- [39] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). *The Clinical Proteomic Tumor Analysis Consortium Cutaneous Melanoma Collection (CPTAC-CM)*. Bep. 10. Dataset. The Cancer Imaging Archive, 2018. DOI: 10.7937/K9/TCIA.2018.ODU24GZE. URL: <https://doi.org/10.7937/K9/TCIA.2018.ODU24GZE>.
- [40] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. B: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), c. 62—66. DOI: 10.1109/TSMC.1979.4310076.

- [41] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. B: *NeurIPS*. 2019.
- [42] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. B: *International Conference on Learning Representations (ICLR)*. 2019. URL: <https://openreview.net/forum?id=Bkg6RiCqY7>.
- [43] Yunyang Xiong et al. “Nyströmformer: A Nyström-Based Self-Attention Mechanism”. B: *AAAI*. 2021.
- [44] Bin Li, Yin Li and Kevin W. Eliceiri. “Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning”. B: *CVPR*. 2021.
- [45] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. B: *International Conference on Learning Representations (ICLR)*. 2019. URL: <https://openreview.net/forum?id=Bkg6RiCqY7>.
- [46] Babak Ehteshami Bejnordi et al. “Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer”. B: *JAMA* 318.22 (2017), c. 2199—2210. DOI: 10.1001/jama.2017.14585.
- [47] Leland McInnes, John Healy and James Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. B: *arXiv preprint arXiv:1802.03426* (2018). DOI: 10.48550/arXiv.1802.03426.
- [48] Bastiaan S. Veeling et al. “Rotation Equivariant CNNs for Digital Pathology”. B: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2018, c. 210—218. DOI: 10.1007/978-3-030-00934-2_24.
- [49] *CAMELYON17 Challenge Leaderboard*. <https://camelyon17.grand-challenge.org/evaluation/challenge/leaderboard/>. Accessed: 07.06.2025.
- [50] Streamlit Inc. *Streamlit — A faster way to build and share data apps*. Режим доступа: <https://streamlit.io>. Дата звернення: 01.06.2026. 2019.

- [51] NVIDIA Corporation. *NVIDIA Ampere GA102 GPU Architecture*. Тех. звіт. Дата звернення: 01.06.2026. NVIDIA Corporation, 2021.
- [52] Ramprasaath R. Selvaraju et al. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”. B: *ICCV*. 2017.
- [53] Meng Li et al. “Deep Learning Approach to Classify Cutaneous Melanoma in a Whole Slide Image”. B: *Cancers* 15.6 (2023), c. 1907. DOI: 10.3390/cancers15061907.
- [54] Lucy Godson et al. “Immune subtyping of melanoma whole slide images using multiple instance learning”. B: *Medical Image Analysis* 93 (2024), c. 103097. DOI: 10.1016/j.media.2024.103097.
- [55] Jialu Yao et al. “Melan-Dx: a knowledge-enhanced vision-language framework improves differential diagnosis of melanocytic neoplasm pathology”. B: *npj Digital Medicine* (2026). DOI: 10.1038/s41746-026-02357-3.
- [56] Erlend S. Holm et al. “Early detection of metastatic risk in primary cutaneous melanoma using weakly supervised learning”. B: *Scientific Reports* (2026). DOI: 10.1038/s41598-026-45588-w.

ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

Повний лістинг коду розміщено за посиланням https://github.com/VictorBohatyrenko/Breast_Cancer_Staging.

```

1 import os
2 import numpy as np
3 import pandas as pd
4 import torch
5 import torch.nn as nn
6 from torch.utils.data import Dataset, DataLoader
7 from sklearn.metrics import cohen_kappa_score, accuracy_score
8
9 # Налаштування пристрою обчислень
10 DEVICE = "cuda" if torch.cuda.is_available() else "cpu"
11 STAGE2ID = {"negative": 0, "itc": 1, "micro": 2, "macro": 3}
12 ID2STAGE = {v: k for k, v in STAGE2ID.items()}
13
14 class SlideEmbDataset(Dataset):
15     """Датасет для завантаження векторних представлень цілих зрізів"""
16     def __init__(self, df: pd.DataFrame, emb_dir: str,
17                 parquet_bags: dict = None):
18         self.df = df.reset_index(drop=True)
19         self.emb_dir = emb_dir
20         self.parquet_bags = parquet_bags or {}
21
22     def __len__(self):
23         return len(self.df)
24
25     def __getitem__(self, idx):
26         row = self.df.iloc[idx]
27         slide_uid = row["slide_uid"]

```

```

27     patient_id = -1 if pd.isna(row["patient_id"]) else int(
        row["patient_id"])
28     y_stage = torch.tensor(int(row["stage_id"]), dtype=
        torch.long)
29
30     # Завантаження попередньо обчислених ознак      UNI2-H
31     pt_path = os.path.join(self.emb_dir, slide_uid + ".pt")
32     data = torch.load(pt_path, map_location="cpu")
33     emb = data["features"].float()
34
35     # Конкатенація додаткових ознак із      parquet файлів – якщо ( наявні )
36     if slide_uid in self.parquet_bags:
37         extra = self.parquet_bags[slide_uid]
38         if extra.shape[1] == emb.shape[1]:
39             emb = torch.cat([emb, extra], dim=0)
40
41     return emb, y_stage, slide_uid, patient_id
42
43 def collate_batch(batch):
44     return batch[0]
45
46 class CLAMMulti(nn.Module):
47     """Архітектура моделі CLAM для багатокласової класифікації"""
48     def __init__(self, d: int, hid: int = 256, dropout: float =
        0.25, n_classes: int = 4):
49         super().__init__()
50         self.feats = nn.Sequential(
51             nn.Linear(d, hid),
52             nn.ReLU(inplace=True),
53             nn.Dropout(dropout)
54         )

```

```

55     self.attn = nn.Linear(hid, 1)
56     self.inst_head = nn.Linear(hid, n_classes)
57     self.bag_head = nn.Linear(hid, n_classes)
58
59     def forward(self, emb: torch.Tensor):
60         h = self.feat(emb)
61         a = self.attn(h).squeeze(1)
62         w = torch.softmax(a, dim=0)
63
64         inst_logits = self.inst_head(h)
65         # Агрегація ознак з урахуванням вагу ваги
66         z = (w.unsqueeze(1) * h).sum(0)
67         bag_logits = self.bag_head(z).unsqueeze(0)
68
69         return bag_logits, inst_logits, w
70
71 def instance_ce_topk(inst_logits: torch.Tensor, y_stage: torch.
Tensor, k: int = 8):
72     """Обчислення Cross-Entropy на top-k найважливіших фрагментах"""
73     cls = int(y_stage.item())
74     M = inst_logits.shape[0]
75     if M == 0: return inst_logits.sum() * 0.0
76     k = min(k, M)
77
78     with torch.no_grad():
79         scores = torch.softmax(inst_logits.detach(), dim=1)[: ,
            cls]
80
81     topk_idx = torch.topk(scores, k=k).indices
82     topk_logits = inst_logits[topk_idx]
83     target = y_stage.expand(k)

```

```

84     return nn.functional.cross_entropy(topk_logits, target)
85
86 def train_one_epoch_mc(model, loader, optimizer, device, epoch,
87     cfg):
88     """Цикл навчання на одній епохі без використання hard mining"""
89     model.train()
90     total = 0.0
91     bag_criterion = nn.CrossEntropyLoss(weight=cfg.get("
92         class_weights"))
93
94     for emb, y_stage, _, _ in loader:
95         emb, y_stage = emb.to(device), y_stage.to(device).view
96             (1)
97
98         bag_logits, inst_logits, _ = model(emb)
99
100         loss_bag = bag_criterion(bag_logits, y_stage)
101         loss_inst = instance_ce_topk(inst_logits, y_stage, k=
102             cfg.get("inst_k", 8))
103         loss = loss_bag + cfg.get("lambda_inst", 0.3) *
104             loss_inst
105
106         optimizer.zero_grad(set_to_none=True)
107         loss.backward()
108         nn.utils.clip_grad_norm_(model.parameters(), cfg["
109             grad_clip"])
110         optimizer.step()
111         total += float(loss.item())
112
113     return total / max(1, len(loader))

```

```
109 # Ініціалізація та запуск навчання
110 model_mc = CLAMMulti(d=1536, hid=256, n_classes=4).to(DEVICE)
111 optimizer_mc = torch.optim.AdamW(model_mc.parameters(), lr=1e
    -4, weight_decay=1e-5)
```

Listing 1: Реалізація багатокласового MIL-пайплайну для класифікації стадій метастазування