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

Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

До захисту допущено:

Завідувач кафедри

_____ Оксана ТИМОЩУК

«__» _____ 2025 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системний аналіз і управління»

спеціальності 124 «Системний аналіз»

на тему: «Модель співставлення потоків для генерації зображень раку шкіри»

Виконав:
студент IV курсу, групи КА-12
Харитонов Олександр Дмитрович _____

Керівник:
Асистент каф. ММСА Баздирев Антон Андрійович _____

Консультант з економічного розділу:
Доцент, к.е.н., Рощина Надія Василівна _____

Консультант з нормоконтролю:
к.ф.-м.н. Статкевич Віталій Михайлович _____

Рецензент:
Старший викладач кафедри СП ІПСА, к.т.н.,
Кислий Роман Володимирович _____

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

Київ – 2025 рік

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

Навчально-науковий інститут прикладного системного аналізу

Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Оксана ТИМОЩУК
«__» _____ 2025 р.

ЗАВДАННЯ

на дипломну роботу студенту

Харитонову Олександру Дмитровичу

1. Тема роботи «Модель співставлення потоків для генерації зображень раку шкіри», керівник роботи Баздирев Антон Андрійович, ассистент кафедри ММСА, затверджені наказом по університету від 26.05.2025 р. №1759-с

2. Термін подання студентом роботи _____

3. Вихідні дані до роботи: наявні набори даних раку шкіри.

4. Зміст роботи: 1 – Огляд історичних та сучасних методів і рішень для генерації зображень та класифікації раку шкіри; 2 – Теоретичні основи моделей машинного навчання і метрик для роботи із зображеннями; 3 – Генерація зображень та перевірка їх якості експериментальним шляхом на реальних наборах даних; 4 – Функціонально-вартісний аналіз програмного забезпечення.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): зображення існуючих наборів даних та згенерованих, ілюстрації роботи експериментальної частини, презентація.

6. Консультанти розділів роботи.

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Рощина Надія Василівна, доц., к.е.н.		

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Аналіз предметної області та формулювання завдання	31.03.2025-06.04.2025	виконано
2	Вивчення літератури та обробка даних за темою роботи	07.03.2025-13.04.2025	виконано
3	Підготовка математичних моделей, розробка програмного коду їх реалізацій	14.04.2025-18.05.2025	виконано
4	Робота над розділами під час переддипломної практики	14.04.2025-18.05.2025	виконано
5	Оформлення та узгодження пояснювальної записки і розділів роботи	19.05.2025-01.06.2025	виконано
6	Попередній захист дипломної роботи	02.06.2025-15.06.2025	виконано
7	Захист дипломної роботи	16.06.2025-22.06.2025	виконано

Студент _____Олександр ХАРИТОНОВ

Керівник _____ Антон БАЗДИРЕВ

РЕФЕРАТ

Дипломна робота: 127 с., 24 рис., 12 табл., 2 дод., 40 джерел.

ЗІСТАВЛЕННЯ ПОТОКІВ, РАК ШКІРИ, СИНТЕТИЧНІ ЗОБРАЖЕННЯ, НИЗКОРАНГОВА АДАПТАЦІЯ, АУГМЕНТАЦІЯ ДАНИХ, ГЛИБОКЕ НАВЧАННЯ, МЕДИЧНА ВІЗУАЛІЗАЦІЯ, НЕЗБАЛАНСОВАНІ ДАНІ

Об’єктом дослідження є програмний комплекс для генерації та аналізу дерматоскопічних зображень раку шкіри на основі моделей із методом зіставлення потоків.

Предметом дослідження виступають методи синтетичної генерації зображень, їх якісного оцінювання та впливу штучних даних на точність класифікації шкірних утворень.

Метою роботи є розробити й експериментально перевірити конвеєр, що компенсує дефіцит медичних даних і підвищує достовірність діагностики завдяки високоякісній синтетичній аугментації.

Актуальність зумовлена гострою нестачею дерматоскопічних знімків злоякісних меланом і суттєвою диспропорцією класів у відкритих наборах даних, що ускладнює навчання сучасних глибинних моделей.

Результатом роботи є Python/PyTorch-прототип, який забезпечує швидке генерування зображень, їх автоматичне оцінювання та підвищує AUC класифікатора EfficientNet-B0 після додавання 20 000 синтетичних прикладів до набору даних.

Подальший розвиток передбачає клінічну валідацію синтетики, адаптацію методики до мультиспектральних даних і впровадження контролю аномалій.

ABSTRACT

Bachelor’s thesis: 127 p., 24 figures, 12 tables, 2 appendices, 40 references.

FLOW MATCHING, FLUX, SKIN CANCER, SYNTHETIC IMAGES, LORA, DATA AUGMENTATION, DEEP LEARNING, MEDICAL IMAGING, IMBALANCED DATASETS

The object of the study is a software framework for generating and analysing dermatoscopic skin-cancer images based on FLUX models trained with the Flow-Matching approach.

The subject of research comprises techniques for synthetic image generation, quality assessment and the influence of artificial data on lesion-classification accuracy.

The goal of the work is to design and experimentally validate a pipeline that mitigates data scarcity and improves diagnostic reliability through high-fidelity synthetic augmentation.

Relevance is driven by the severe shortage of malignant dermatoscopic images and the strong class imbalance in public datasets, which hinder the training of state-of-the-art deep models.

The result of the thesis is a Python/PyTorch prototype that performs fast image generation, automatic quality evaluation and raises the EfficientNet-B0 AUC on ISIC 2024 after adding 20 000 synthetic samples.

Further development involves clinical validation of the synthetic data, extension to multispectral modalities and incorporation of anomaly-detection mechanisms.

ЗМІСТ

ВСТУП 9

РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ 12

1.1 Актуальність проблеми діагностики раку шкіри в сучасній медицині 12

1.2 Аналіз існуючих підходів до класифікації раку шкіри 13

1.2.1 Традиційні методи діагностики (дерматоскопія, біопсія) 14

1.2.2 Комп'ютерні методи аналізу зображень шкіри 15

1.2.3 Глибоке навчання в дерматології 17

1.3 Проблема нестачі медичних даних 19

1.3.1 Дисбаланс і обмеження наборів даних 19

1.3.2 Методи аугментації даних 20

1.3.3 Синтетична генерація медичних зображень 22

1.4 Огляд генеративних моделей для медичних зображень 23

1.4.1 Моделі на основі GAN 24

1.4.2 Варіаційні автоенкодери та їх модифікації 26

1.4.3 Дифузійні моделі та Flow Matching 28

1.5 Висновки до розділу 1 32

РОЗДІЛ 2 МАТЕМАТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ 33

2.1 Теоретичні основи flow matching 33

2.1.1 Математичне формулювання flow matching 33

2.1.2 Рівняння безперервного нормалізуючого потоку 34

2.1.3 Оптимальний транспорт та траєкторії потоку 36

2.1.4 Переваги flow matching над DDPM підходом 37

2.2 Архітектури генеративних моделей 39

2.2.1 Transformer-based архітектури для генерації зображень 40

2.2.2 U-Net модифікації для дифузійних моделей 42

2.3 Низькорангова адаптація для ефективного тренування 44
2.3.1 Математичні основи LoRA 44
2.3.2 Застосування LoRA до FLUX моделей 45
2.3.3 Оптимізація пам'яті та обчислювальних ресурсів 46
2.4 Метрики оцінки якості синтетичних зображень та класифікації 47
2.4.1 Традиційні метрики якості генерації 47
2.4.2 Метрики класифікації 50
2.5 Висновки до розділу 2 52
РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ 53
3.1 Підготовка експериментального середовища 53
3.1.1 Технічні характеристики та програмне забезпечення 53
3.1.2 Набори даних медичних зображень 54
3.1.3 Препроцесинг та підготовка даних 59
3.2 Імплементація та навчання flow matching моделі для генерації зображень 63
3.2.1 Конфігурація LoRA адаптерів та стратегія навчання 63
3.2.2 Процес навчання моделі на медичних даних 64
3.2.3 Контрольована генерація зображень для бінарної класифікації 66
3.4 Оцінка якості згенерованих зображень 68
3.4.1 Результати за метриками FID, IS, LPIPS 68
3.4.2 Якісний аналіз згенерованих зображень 70
3.4.3 Порівняння FLUX.1-dev та FLUX.1-schnell 72
3.5 Експерименти з класифікацією 75
3.5.1 Архітектура класифікаційної моделі 75
3.5.2 Експерименти з різними співвідношеннями синтетичних даних 76
3.5.3 Аналіз метрик класифікації 82

3.6 Аналіз результатів та обмежень 86
3.6.1 Вплив синтетичних даних на покращення класифікації 86
3.6.2 Виявлені обмеження підходу 88
3.6.3 Обчислювальна ефективність методу 90
3.7 Висновки до розділу 3 92
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ 94
4.1 Постановка задачі проектування 95
4.2 Обґрунтування функцій програмного продукту 96
4.3 Обґрунтування системи параметрів програмного продукту 98
4.4 Аналіз експертного оцінювання параметрів 101
4.5 Аналіз рівня якості варіантів реалізації функцій 105
4.6 Економічний аналіз варіантів розробки програмного продукту 107
4.7 Вибір кращого варіанту ПП техніко-економічного рівня 112
4.8 Висновки до розділу 4 113
ВИСНОВКИ 114
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 116
ДОДАТОК А ЛІСТИНГ КОДУ 121
ДОДАТОК Б ПРЕЗЕНТАЦІЯ 128

ВСТУП

У сучасному світі рак шкіри залишається однією з найсерйозніших медичних проблем, що демонструє стійку тенденцію до зростання захворюваності протягом останніх десятиліть. Традиційні методи діагностики, хоча й ефективні, мають ряд обмежень, пов'язаних із суб'єктивністю оцінки, залежністю від досвіду лікаря та обмеженою доступністю спеціалізованої медичної допомоги у віддалених регіонах.

Історично діагностика раку шкіри базувалася на візуальному огляді та дерматоскопії, що вимагала від лікарів значного досвіду для правильної інтерпретації клінічних ознак. Розвиток критеріїв ABCDE дозволив стандартизувати підхід до оцінки підозрілих новоутворень, проте остаточний діагноз все ще потребував гістопатологічного підтвердження через біопсію. Ці методи, попри свою ефективність, залишалися трудомісткими та вимагали високої кваліфікації медичного персоналу.

З появою цифрових технологій та розвитком комп'ютерного зору почали з'являтися перші системи автоматизованого аналізу дерматоскопічних зображень. Ранні підходи базувалися на класичних методах комп'ютерного зору з ручним виділенням ознак, проте їхня ефективність була обмежена складністю та варіативністю візуальних проявів шкірних захворювань.

Революційний прорив відбувся з появою глибокого навчання та згорткових нейронних мереж. Дослідження Стенфордського університету продемонстрували, що штучний інтелект здатен досягати точності діагностики, співставної з досвідченими дерматологами. Це відкрило нові можливості для створення доступних та об'єктивних діагностичних систем.

Однак розвиток ефективних систем машинного навчання для медичної діагностики стикнувся з фундаментальною проблемою – нестачею якісних навчальних даних. Медичні зображення, особливо зображення рідкісних або

злоякісних уражень, складно збирати через етичні обмеження, приватність пацієнтів та високу вартість клінічних досліджень. Крім того, наявні набори даних, такі як HAM10000 та ISIC, характеризуються значним дисбалансом класів, де злоякісні ураження представлені в десятки разів менше, ніж доброякісні.

Традиційні методи аугментації даних, такі як ротація, масштабування та зміна кольору, хоча й допомагають частково вирішити проблему, не здатні генерувати принципово нові варіанти медичних проявів. У цьому контексті генеративні моделі, зокрема GAN та дифузійні моделі, відкрили нові можливості для синтетичної генерації медичних зображень.

Останні досягнення в галузі генеративного моделювання, зокрема розробка Flow Matching підходу, запропонували більш стабільну та ефективну альтернативу традиційним методам. Моделі на основі Flow Matching, такі як FLUX, демонструють високу якість генерації при значно меншій кількості кроків інференсу порівняно з дифузійними моделями. Це робить їх особливо привабливими для медичних застосувань, де швидкість та надійність є критично важливими.

Актуальність даного дослідження зумовлена необхідністю подолання обмежень існуючих підходів до діагностики раку шкіри. По-перше, потребує вирішення проблема дисбалансу класів у медичних наборах даних. По-друге, необхідно розробити ефективні методи генерації реалістичних синтетичних медичних зображень. По-третє, важливо оцінити вплив синтетичних даних на якість класифікаційних моделей та їхню безпечність для клінічного використання.

Метою цієї роботи є розробка та дослідження системи генерації синтетичних дерматоскопічних зображень на основі Flow Matching підходу для покращення якості автоматичної діагностики раку шкіри в умовах обмежених та незбалансованих навчальних даних.

Для досягнення поставленої мети необхідно вирішити комплекс взаємопов'язаних завдань: провести аналіз сучасних підходів до генерації

медичних зображень, адаптувати технологію Flow Matching для медичного домену, розробити систему контрольованої генерації дерматоскопічних зображень, оцінити якість синтетичних даних та їхній вплив на класифікацію, а також провести економічне обґрунтування запропонованого підходу.

Наукова новизна роботи полягає у першому застосуванні Flow Matching для генерації дерматоскопічних зображень, адаптації архітектури FLUX з використанням LoRA технології для медичних застосувань, розробці спеціалізованої методології оцінки впливу синтетичних даних на діагностичні системи.

Практичне значення результатів визначається можливістю їх використання для покращення точності діагностичних систем, зменшення потреби у дорогих клінічних дослідженнях та прискорення розробки нових медичних застосувань штучного інтелекту. Створення такого програмного продукту має на меті забезпечити медичних працівників надійним інструментом для підтримки прийняття діагностичних рішень, особливо в умовах обмеженого доступу до спеціалізованої медичної допомоги.

РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність проблеми діагностики раку шкіри в сучасній медицині

Рак шкіри є однією з найпоширеніших онкопатологій у світі, а його рання діагностика залишається надзвичайно важливою медичною проблемою. Зокрема, меланома – найагресивніший вид раку шкіри – демонструє тенденцію до зростання захворюваності: за останні три десятиліття кількість випадків меланоми зросла утричі, всупереч загальному зниженню захворюваності на інші поширені види раку. За оцінками Всесвітньої організації охорони здоров'я (ВООЗ), у 2020 році у світі було зареєстровано близько 324,6 тис. нових випадків меланоми та 57 тис. смертей від неї [2]; якщо не вживати заходів, до 2040 року ці показники можуть зрости до 510 тис. випадків і 96 тис. смертей відповідно. Така динаміка зумовлює актуальність удосконалення методів раннього виявлення раку шкіри.

Проблеми ранньої діагностики. Раннє виявлення раку шкіри кардинально впливає на прогноз для пацієнта. Якщо меланому діагностувати на ранній стадії (інвазія мінімальна), п'ятирічна виживаність сягає ~97%, тоді як при виявленні на пізніх метастатичних стадіях цей показник може знижуватися до ~14%. Однак своєчасна діагностика утруднена низкою чинників. По-перше, клінічні ознаки злоякісних уражень можуть бути непомітними або маскуватися під доброякісні невуси [18]. Деякі меланоми, як-от амеланотична форма, не містять пігменту й виглядають як звичайна рожева пляма, що ускладнює їх розпізнавання. Типовий підхід до діагностики – огляд шкіри дерматологом із використанням критеріїв ABCDE [1] (асиметрія, нерівний край, неоднорідність кольору, діаметр, еволюція) – є

ефективним, проте залежить від досвіду лікаря і суб’єктивних оцінок. Крім того, не всі пацієнти вчасно звертаються до спеціалістів: географічна віддаленість, брак дерматологів і банальна неуважність можуть призводити до затримки діагнозу. Усе це спонукає до пошуку технологічних рішень, які б допомогли виявляти рак шкіри на ранніх стадіях.

Роль ШІ у медичній діагностиці. Сучасні досягнення в галузі штучного інтелекту (ШІ) суттєво впливають на методи діагностики різних захворювань, у тому числі й раку шкіри. Застосування алгоритмів глибокого навчання дозволило створити системи автоматичної класифікації шкірних новоутворень за зображеннями, які за точністю наближаються до рівня експертів-дерматологів. Відомою віхою стала робота [3], в якій глибока нейронна мережа була навчена розпізнавати меланому на дерматоскопічних зображеннях не гірше за лікарів: у тестах алгоритм продемонстрував рівень діагностики, співставний з 21 сертифікованим дерматологом. Подальші дослідження показали, що належним чином навчені моделі здатні навіть перевершувати лікарів у точності аналізу дерматоскопії. Отже, ШІ розглядається як перспективний інструмент для підтримки прийняття рішень у дерматології: він може забезпечити швидкий скринінг великих масивів зображень, підвищити об’єктивність оцінки та доступність діагностики у віддалених регіонах. Все це підкреслює актуальність дослідження методів генерації та класифікації медичних зображень раку шкіри із залученням сучасних моделей глибокого навчання.

1.2 Аналіз існуючих підходів до класифікації раку шкіри

1.2.1 Традиційні методи діагностики (дерматоскопія, біопсія)

Основою діагностики раку шкіри традиційно є клінічний огляд та інвазивне підтвердження діагнозу. **Дерматоскопія** – неінвазивний метод, що передбачає огляд шкірних уражень за допомогою дерматоскопа (спеціальної лупи з підсвіткою) – значно покращує виявлення ознак злоякісності порівняно з неозброєним оком. Згідно з мета-аналізами, додавання дерматоскопії до візуального огляду підвищує чутливість і специфічність виявлення меланоми. Дерматоскоп дозволяє побачити структури (пігментну сітку, глобули, судини тощо) та кольори, невидимі при звичайному огляді, що відповідають гістологічним особливостям шкірних утворень [5]. Наприклад, дерматологи можуть виявити ознаки неправильної пігментації або атипового судинного малюнка, характерні для меланоми, завдяки поляризованому підсвічуванню. Використання дерматоскопії потребує спеціального навчання, але суттєво підвищує діагностичну точність: за даними огляду Cochrane, точність діагнозу при очному огляді з дерматоскопом у понад 4 рази вища, ніж при оцінці лише фотографій уражень без присутності пацієнта. Отже, дерматоскопія стала стандартним інструментом дерматологів для ранньої діагностики раку шкіри.

Для полегшення прийняття рішення лікарями розроблено декілька правил і критеріїв оцінки невусів. Найбільш відомим є правило ABCD(E), запропоноване ще в 1985 році для самообстеження та огляду пігментних уражень [6]. Воно акцентує увагу на чотирьох головних ознаках меланоми: Asymmetry (Асиметрія форми), Border irregularity (Нерівність або нечіткість країв), Color variegation (Різноманітність кольору, нерівномірна пігментація),

Diameter > 6 mm (Діаметр більше 6 мм). Пізніше критерії були доповнені п’ятим компонентом E – Evolution (Еволюція, тобто будь-які зміни існуючої родимки: розміру, форми, кольору, появи симптомів як-от свербіж чи кровоточивість). Дотримання ABCDE-правила допомагає виявляти підозрілі ураження, однак остаточний діагноз все одно потребує гістологічної верифікації.

Біопсія та гістопатологія є золотим стандартом діагностики раку шкіри. Після клінічної підозри на злоякісність виконується видалення новоутворення (екцизійна біопсія) або забір його частини (інцизійна чи трепан-біопсія) з подальшим мікроскопічним дослідженням тканини. Гістопатологічний аналіз дозволяє точно визначити тип пухлини і ступінь інвазії. Незважаючи на розвиток неінвазивних методів (наприклад, конфокальної лазерної мікроскопії in vivo), жоден з них поки не замінив біопсію. Екцизійна біопсія залишається «золотим стандартом» підтвердження меланоми, оскільки тільки патоморфологічне дослідження дає остаточний діагноз. Водночас, біопсія – інвазивна процедура, що створює навантаження на пацієнта. Тому дерматологи вимушені балансувати між необхідністю виявити максимум випадків раку (щоб не пропустити меланому) і уникненням непотрібних біопсій доброякісних утворень. У практиці застосовуються алгоритми на кшталт «двох етапів» – спочатку неінвазивна оцінка (правила ABCDE, дерматоскопія), і лише при наявності низки підозрілих ознак – направлення на біопсію. Ці традиційні підходи заклали основу діагностики; проте їх ефективність може обмежуватися людським фактором і обсягом досвіду лікаря. Це стимулювало розвиток комп’ютерних методів, які можуть допомогти у класифікації шкірних утворень.

1.2.2 Комп’ютерні методи аналізу зображень шкіри

З появою цифрових зображень та персональних комп’ютерів почали розвиватися системи комп’ютерної діагностики (CAD) для аналізу шкірних уражень. Ранні підходи базувалися на класичних методах комп’ютерного зору: спочатку зображення дерматоскопії підлягало попередній обробці (наприклад, фільтрація шумів, видалення артефактів типу волосся), потім – сегментації області новоутворення, і подальшому вилученню ознак (фіч). Як ознаки використовувалися кількісні еквіваленти критеріїв, що ними керуються лікарі: метрики асиметрії форми контуру, ступінь нерівності границь (фрактальні розмірності або індекси зубчатості краю), показники варіації кольору (кількість різних кольорових тонів, гістограма пігментації), розмір утворення, а також текстурні характеристики поверхні [7]. На основі витягнутих ознак будувалася модель класифікації – наприклад, алгоритми k-ближчих сусідів, метод опорних векторів (SVM) або штучні нейронні мережі невеликої глибини. Такі системи мали на меті автоматично відрізняти злоякісні ураження (передусім меланому) від доброякісних (наприклад, невусів).

Історично більшість досліджень фокусувалися на задачі розпізнавання меланоми проти невуса, оскільки дані були обмеженими. Приміром, одна з ранніх робіт з використанням нейронної мережі для диференціації меланом і невусів показала обнадійливі результати, але мала дуже малу вибірку і включала лише ці два класи уражень. Загалом до середини 2010-х років публікувалися численні алгоритми, що демонстрували високу точність на невеликих наборах даних. Однак відсутність великих різноманітних наборів дерматоскопічних зображень обмежувала практичну цінність таких рішень: алгоритми часто не переносилися на інші типи уражень чи умови зйомки. Зазначається, що через обмеження наявних даних більшість старих досліджень

зосереджувалися тільки на меланоцитарних утвореннях (меланоми і невуси) і ігнорували немеланоцитарні ураження (базаліоми, кератози тощо), які теж часто трапляються. Це спричиняло провали у загальній продуктивності: модель, навчена розрізняти лише невуси та меланоми, могла помилково класифікувати, скажімо, пігментовану базаліому, оскільки не була знайома з таким видом захворювання.

Попри ці виклики, комп'ютерний аналіз зображень проклав шлях до більш автоматизованої діагностики. У 2016 році було засновано масштабну ініціативу – International Skin Imaging Collaboration (ISIC), що об'єднала дерматологів та дослідників з комп'ютерного бачення. В рамках ISIC створюється відкритий архів дерматоскопічних зображень та проводяться щорічні змагання (челенджі) з аналізу шкірних зображень (класифікація, сегментація, виявлення ознак). Ці зусилля значно розширили доступну базу даних і стимулювали розвиток методів. З 2017 року, із вибухом технологій глибокого навчання, у цій сфері стався переломний момент: класичні методи з ручним виокремленням ознак були швидко витіснені підходами на основі згорткових нейронних мереж, здатних автоматично навчатися ознакам із великих наборів даних.

1.2.3 Глибоке навчання в дерматології

Методи глибокого навчання, зокрема згорткові нейронні мережі (CNN), продемонстрували видатні результати в задачах класифікації зображень і знайшли успішне застосування в дерматології. Перевага CNN полягає у здатності автоматично виокремлювати багаторівневі ознаки зі зображень, що особливо цінно для складних візуальних патернів шкірних уражень. Перші гучні успіхи з'явилися близько 2016–2017 років. Дослідники з Стенфорду використали архів ~130 тисяч зображень шкірних захворювань для тренування

глибинної мережі Google Inception-v3 і досягли точності, порівнянної з діагнозами досвідчених дерматологів при класифікації найбільш небезпечних видів раку шкіри. У Nature було опубліковано їхній результат, що став сенсацією: алгоритм правильно розпізнавав меланоми не гірше лікарів у контрольній вибірці біопсійно підтверджених випадків. Згодом інші роботи підтвердили, що добре навчені нейромережі можуть навіть перевершити спеціалістів. Наприклад, у 2018 році на конкурсі ISIC CNN-модель показала вищу чутливість, ніж дерматологи, у виявленні меланоми на дерматоскопічних зображеннях .

Важливим фактором успіху стало поява більших і якісніших датасетів. Спільнотою ISIC був створений відкритий архів, який на 2018 рік налічував ~13,7 тис. дерматоскопічних зображень, що зробило його стандартним джерелом даних для навчання алгоритмів [9]. У тому ж 2018 році було оприлюднено великий набір HAM10000 – 10015 дерматоскопічних зображень семи типів пігментних уражень, зібраних з різних клінічних джерел [8]. Наявність цього різноманітного набору (половина випадків підтверджена гістологією, решта – наглядом чи консенсусом експертів) дозволила навчити глибокі моделі розрізняти не тільки меланоми і невуси, а й інші поширені утворення (базаліоми, кератози, дерматофіброми тощо). Як результат, сучасні CNN-класифікатори раку шкіри здатні виконувати багатокласову класифікацію з високою точністю. Наприклад, архітектури на кшталт ResNet чи EfficientNet, навчені на даних ISIC, досягали показників збалансованої точності понад 0,85–0,90 у завданні мультикласової діагностики шкірних утворень на конкурсах ISIC.

Проте, глибоке навчання потребує великої кількості даних, а медичні дані зібрати непросто. Також залишається питання інтерпретованості: моделі-чорні скриньки важко пояснити лікарю, який звик обґрунтовувати рішення конкретними ознаками. Тим не менш, у дерматології активно досліджуються підходи Grad-CAM та інші методи пояснення, щоб виділяти області зображення, на яких модель «зосереджується». Загалом, впровадження

глибокого навчання в класифікацію раку шкіри є одним із найперспективніших напрямів, здатних підвищити ефективність та доступність діагностики. Розвиток цієї галузі тісно пов’язаний із подоланням проблем, пов’язаних із даними, – саме тому увага дослідників змістилася також на генерацію синтетичних медичних зображень для розширення датасетів, про що йтиметься далі.

1.3 Проблема нестачі медичних даних

1.3.1 Дисбаланс і обмеження наборів даних

Для успішного навчання глибоких моделей класифікації потрібні репрезентативні великі набори даних. У випадку раку шкіри виникає кілька проблем: по-перше, обмежений обсяг загальнодоступних розмічених зображень; по-друге, дисбаланс класів (значно більше прикладів доброякісних уражень, ніж злоякісних); по-третє, варіабельність джерел (зображення можуть відрізнятися за обладнанням, умовами зйомки, роздільною здатністю). Хоча за останні роки ситуація покращилась завдяки зусиллям спільноти (архів ISIC тощо), все ж доступні дані не повністю охоплюють усе різноманіття випадків. Наприклад, у згаданому наборі HAM10000 [8] з 10015 зображень представлено 7 діагностичних категорій, але розподіл між ними нерівномірний: понад 67% становлять меланоцитарні невуси, тоді як меланоми – лише близько 11% (≈1113 зображень), а деякі інші класи (дерматофіброми, судинні ураження) мають менше 150 прикладів. Такий дисбаланс ускладнює навчання моделі – вона може схилятися до більш чисельних класів і пропускати рідкісні, але критично важливі (меланому).

Крім того, історично набори даних були дуже малими. Ранні загальнодоступні датасети мали лише декілька сотень зображень: приміром, відомий набір PH2 містив 200 дерматоскопічних зображень (160 невусів і 40 меланом). На такому обмеженому обсязі даних нейромережа ризикує перенавчитися – запам’ятати приклади замість навчитися узагальнювати. Навіть сучасний ISIC Archive, поповнений за рахунок різних клінік, станом на 2018 рік мав ~13 тисяч зображень [9]. Хоча за останні роки архів зріс до сотень тисяч зображень (у тому числі завдяки завантаженню фотографій шкіри користувачами мобільних додатків), багато з цих нових зображень не мають надійної верифікації діагнозу, або являють собою клінічні фотографії замість дерматоскопії. Таким чином, проблема якісно розмічених даних залишається. До того ж, існують етичні та правові обмеження: медичні зображення містять персональні дані, їх збір і публікація потребують згоди пацієнтів, що ускладнює швидке масштабування наборів даних.

Нестача даних і їхній дисбаланс напряму впливають на продуктивність алгоритмів. Дослідження відзначають, що успіх глибоких моделей лімітований доступними даними: моделі часто не охоплюють всі реальні варіації ознак і можуть видавати необ’єктивні прогнози. Різниця у джерелах даних (наприклад, зображення з одних клінік можуть відрізнятися за кольоровою гамою від зображень інших) призводить до того, що модель, навчена на одному наборі, погано працює на інших (проблема доменної адаптації). Отже, постала необхідність пошуку шляхів розширення та збагачення навчальних вибірок без порушення етичних норм. Один з підходів – це аугментація наявних даних шляхом штучного збільшення кількості зображень.

1.3.2 Методи аугментації даних

Аугментація даних – це набір методів штучного створення нових навчальних прикладів на основі наявних, шляхом внесення різноманітних змін до оригінальних даних. Метою аугментації є підвищення варіативності набору та зменшення перенавчання моделі. Традиційні прийоми аугментації зображень включають геометричні та колірні трансформації: повороти на випадковий кут, горизонтальні/вертикальні віддзеркалення, масштабування (збільшення/зменшення), зсув по осі, обрізання фрагментів (cropping), зміну яскравості, контрасту, насиченості кольорів тощо. Такі прості перетворення часто ефективні та прості у реалізації: наприклад, віддзеркалення чи поворот родимки створює новий приклад, що міг би відповідати тій самій родимці на іншій ділянці тіла або під іншим кутом освітлення. У медичних зображеннях нерідко застосовують і більш спеціалізовані аугментації: зміна кольорового балансу для моделювання різних дерматоскопів (оскільки різні пристрої можуть давати зображення дещо відмінних відтінків шкіри), синтез шумів, розмиття для імітації дефокусування тощо [10].

В літературі відзначається, що методи аугментації варіюються від простих трансформацій на кшталт обрізання, доповнення рамкою чи віддзеркалення – до складних генеративних моделей, здатних синтезувати принципово нові зображення. Іншими словами, спектр методів охоплює як елементарні детерміновані операції над пікселями, так і використання окремих нейромереж для створення нових даних. У контексті дерматології базові аугментації вже зарекомендували себе: змагання ISIC дозволяли учасникам застосовувати аугментацію, і практично всі провідні рішення включали її в пайплайн моделі. Проте таких методів може бути недостатньо, аби подолати серйозний дисбаланс (скажімо, коли меланом у десятки, або навіть і сотні разів менше, ніж невусів). Просте дублювання або варіювання наявних зображень не додає принципово нової інформації про різні випадки меланоми, які не представлені у вибірці. Тому все більшої уваги набувають підходи синтетичної генерації даних – створення нових зображень, не просто шляхом

трансформації існуючих, а фактично «з нуля» на основі статистичного навчання розподілу даних.

1.3.3 Синтетична генерація медичних зображень

Використання **генеративних моделей** для синтезу нових зображень раку шкіри розглядається як перспективний шлях подолання нестачі даних. Ідея полягає в тому, щоб навчити модель (наприклад, генеративну нейромережу) на реальних дерматоскопічних зображеннях, а потім змусити її продукувати штучні зображення, які візуально правдоподібні та зберігають клінічно важливі ознаки. Такі синтетичні зображення можуть використовуватися для доповнення тренувального датасету, особливо для малочисельних класів (наприклад, збільшити кількість прикладів меланоми), або для розширення різноманітності варіацій (різні відтінки шкіри, нетипові локалізації, рідкісні випадки). Важливо, що при цьому не порушуються етичні норми – згенеровані зображення не належать конкретним пацієнтам, а отже не несуть приватної інформації.

Останніми роками досягнуто значного прогресу в генерації високоякісних зображень за допомогою моделей глибокого навчання, таких як Generative Adversarial Networks (GAN) та дифузійні моделі (про них детально в розділі 1.4). Ці підходи вже застосовуються в медичній сфері: синтезовано реалістичні зображення МРТ, рентгенограм, гістологічних зрізів для задач збільшення даних та навіть захисту конфіденційності. У дерматології теж є успішні приклади.

Конкретним доказом ефективності служать результати роботи [11], де було запропоновано метод генерації зображень шкірних уражень на базі стилізованого GAN. Отримані синтетичні дерматоскопічні зображення додали до навчальної вибірки, що дозволило значно покращити показники

класифікації на тестовому наборі ISIC 2018. Зокрема, чутливість виросла на 24,4%, специфічність – на 3,6% у порівнянні з базовою моделлю без аугментації. Цей приріст демонструє, що синтетичні дані можуть компенсувати дисбаланс: модель починає краще виявляти меланоми (чутливість зросла) без суттєвого падіння специфічності, тобто не жертвуючи точністю на інших класах.

Використання генеративних підходів, втім, висуває вимоги до якості синтезу. Згенеровані зображення повинні бути реалістичними – містити деталі шкірного малюнку, характерні структури тощо, аби модель-«класифікатор» сприймала їх як справжні. Крім того, важливо зберегти патологічні особливості: якщо генеруємо меланому, вона має мати властиві меланомі риси (асиметрію, нерівний край, кілька кольорів, можливі вкраплення сивого кольору тощо). Сучасні генеративні моделі на кшталт GAN високої роздільності або дифузійних моделей зазвичай справляються з цим завданням. Отже, проблема нестачі даних у дерматологічних застосуваннях все частіше вирішується не лише збором нових реальних фото, а й генерацією штучних зображень. У наступному розділі розглянемо основні типи генеративних моделей, що використовуються для синтезу медичних зображень, з акцентом на новітній підхід Flow Matching і його варіант FLUX.

1.4 Огляд генеративних моделей для медичних зображень

1.4.1 Моделі на основі GAN

Generative Adversarial Networks (GAN) – це клас генеративних моделей, що з двох нейромереж, які тренуються одночасно у протистоянні: генератор G намагається синтезувати зразки, подібні до реальних даних, а

дискримінатор D намагається відрізнити синтетичні зразки від справжніх [12]. Математично їх відношення описується як мінімакс-гра: генератор прагне максимізувати ймовірність помилки дискримінатора, тобто «обдурити» його. У ході оптимізації генератор навчається відтворювати розподіл реальних даних, і при теоретично ідеальній рівновазі G генерує зразки, нерозрізнені з реальними, а D видає 50% впевненості для будь-якого зразка (втрачає здатність розрізняти).

GAN-принцип виявився надзвичайно успішним для синтезу фотореалістичних зображень. Було розроблено багато варіацій GAN, що покращують стабільність тренування та якість згенерованих зображень. Серед них: DCGAN (глибокий згортковий GAN) – перший успішний застосунок згорткових шарів для генерації зображень; Conditional GAN (cGAN) – умовний GAN, де генератору і дискримінатору подається додаткова інформація (наприклад, мітка класу), що дозволяє генерувати зображення заданого класу; CycleGAN – архітектура для перекладу зображень між двома доменами (наприклад, перетворення фото шкіри на дерматоскопічний вигляд і назад); StyleGAN – сімейство моделей (Karras et al., 2019) зі спеціальною стилевою генеративною архітектурою, що досягли найвищої якості зображень облич і були адаптовані для медичних цілей.

У медичній візуалізації GAN активно застосовуються для кількох завдань:

1. Аугментація даних - GAN використовують для синтезу нових прикладів рідкісних патологій. Як згадувалось, у дерматології GAN допомогли збалансувати датасети, згенерувавши додаткові зображення меланом, що підвищило показники класифікації . Інший приклад – генерація зображень рентгенів легенів з ознаками рідкісних захворювань для покращення роботи детекторів, або створення синтетичних МРТ у різних режимах, щоб навчити модель працювати без артефактів від різних томографів [4].

2. Переклад стилів (Style transfer) і підвищення якості - CycleGAN та подібні моделі дозволяють перетворювати зображення між різними

модальностями. У дерматології, наприклад, досліджували перетворення звичайних фотографій шкіри у дерматоскопічні зображення, щоб використати наявні великі колекції клінічних фото для навчання моделей, які початково заточені під дерматоскопію. GAN також застосовували для штучного збільшення роздільної здатності зображень шкіри, якщо оригінали низькоякісні.

3. Виявлення аномалій - якщо навчити GAN на здорових зображеннях (або на доброякісних невусах), то він моделюватиме нормальний розподіл. Відхилення реального зображення від цього розподілу можна трактувати як аномалію (потенційно ознаку злоякісності). Такі підходи досліджуються для скринінгу, де модель відзначає підозрілі ділянки, які не вписуються в «норму».

Важливо зазначити, що для стабільного навчання GAN потрібні достатні дані і тонке налаштування балансу між G та D. При недостатній вибірці GAN може страждати на колапс моделі (генератор відтворює обмежений піднабір прикладів) або низькодостовірні результати. Втім, сучасні техніки (спеціальні функції втрат, нормалізація спектру ваг, диференційовані оптимізатори) значно знизили ці ризики. В контексті нашої роботи GAN цікаві тим, що вони заклали фундамент для потужних генеративних моделей і показали принципову можливість синтезувати реалістичні зображення раку шкіри.

1.4.2 Варіаційні автоенкодери та їх модифікації

Іншим важливим класом генеративних моделей є **варіаційні автоенкодери (VAE)**, які поєднують ідеї автокодувальника та варіаційного байєсівського виведення. Класичний автоенкодер має енкодер – мережу, що стискає вхідне зображення до компактного векторного представлення, та декодер – мережу, що відновлює з цього коду оригінальне зображення [13].

Варіаційний автоенкодер додає йому стохастичності: замість коду – випадковий вектор, згенерований із розподілу, параметри якого передає енкодер. Зазвичай енкодер видає параметри багатовимірного нормального розподілу (вектор середніх і дисперсій), звідки вибирається латентний вектор z. Декодер приймає z і генерує відновлене зображення. Навчання VAE полягає в максимізації варіаційної нижньої межі правдоподібності: забезпечується, з одного боку, гарна реконструкція зображення, а з іншого – наближення розподілу латентних кодів до деякого простого апріорного розподілу (зазвичай $N(0, I)$).

Завдяки цим умовам після тренування VAE можна генерувати нові зображення шляхом вибірки випадкового $z \sim N(0, I)$ і прогону його через декодер. Таким чином, VAE вчиться моделювати розподіл даних і дозволяє генерувати нові зразки та інтерполювати між ними у латентному просторі. Хоча якість зображень, згенерованих ранніми VAE, поступалася GAN (часто спостерігалося розмиття деталей через використання середньоквадратичної функції втрат), перевагою VAE є стабільніше та більш теоретично гарантоване навчання. Крім того, VAE дає явний ймовірнісний модельний інтерпретатор: можна оцінювати правдоподібність належності нового зображення до розподілу (що корисно для виявлення аномалій).

Були розроблені модифікації VAE, спрямовані на покращення якості генерації або наділення латентного простору певними властивостями. Наприклад, β-VAE вводить коефіцієнт β перед терміном Kullback–Leibler (KL-дивергенції) у функції втрат, що дозволяє контролювати ступінь дисперсії латентного простору. При великих β модель сильніше прагне до факторизації латентних змінних, що приводить до дисентанглементу ознак – окремі компоненти z починають відповідати за окремі семантичні фактори. Так, у випадку з зображеннями облич β-VAE могли навчитися, що одна координата відповідає за вираз обличчя, інша – за поворот голови, тощо. В контексті дерматоскопії дисентанглд латентний простір міг би, наприклад, відокремити

фактор «колір шкіри» від фактору «наявність нерівного краю» тощо, дозволяючи більш керовано генерувати варіації.

Нове слово у розвитку автоенкодерів – кодувальники з дискретним латентним простором, такі як *VQ-VAE*. В них безперервний латентний вектор округлюється до найближчого значення з попередньо визначеного «словника» кодів (vector quantization). Це створює дискретне представлення, яке потім декодер перетворює на зображення. Перевага підходу – висока якість та можливість комбінувати з авторегресивними моделями верхнього рівня. Так, VQ-VAE були використані в моделях на кшталт DALL-E для генерації зображень з тексту: спочатку автоенкодер переводить зображення в послідовність кодів, а потім трансформер генерує цю послідовність з текстового опису. У медичних задачах VQ-VAE може знайти застосування для стискання зображень без втрати клінічних деталей і подальшої генерації.

Загалом, VAE та їхні похідні забезпечують байєсівський підхід до генерації. Вони гарантують неперервність і повноту латентного простору (будь-яка вибірка z дає осмислене зображення), що корисно для інтерполяції між випадками або для задач типу виявлення аномалій. Недоліком першої генерації VAE була дещо нижча чіткість зображень порівняно з GAN, але сучасні гібридні моделі та архітектурні удосконалення суттєво скоротили цей розрив. Для генерації шкірних зображень VAE теж застосовували: наприклад, для синтезу варіантів доброякісних невусів як «фонового» класу з метою тренування детекторів меланоми, або для зменшення шуму/відновлення пошкоджених дерматоскопічних фото. Таким чином, VAE є невід’ємною частиною арсеналу генеративних моделей у медицині.

1.4.3 Дифузійні моделі та Flow Matching

Дифузійні моделі – відносно новий клас генеративних моделей, які здобули популярність завдяки сполученню високої якості синтезу та стабільності навчання. Ідея дифузійних моделей була закладена в дослідженнях Дж. Хо, А. Джейна, П. Аббела (2020) на основі зіставлення оцінок зменшення шуму [14]. Принцип роботи дифузійної моделі такий: визначається процес вперед, що поступово перетворює дані на шум (наприклад, шляхом додавання гаусівського шуму на кожному кроці, поки зображення не стане майже випадковим). Модель же навчається зворотного процесу – тобто відновлювати сигнал зі шуму крок за кроком, від дуже зашумленого до чистого. Це виконується через навчання нейронної мережі-денойзера, яка на кожному проміжному кроці прогнозує або початкове зображення, або накладений шум, щоб скоригувати поточний зразок.

Формально дифузійна модель розглядається як латентно-змінна модель з великим числом кроків, де на останньому кроці отримуємо простий розподіл (зазвичай ізотропний нормальний шум), а на нульовому – цільовий розподіл даних. Навчання відбувається шляхом максимізації варіаційної нижньої межі або еквівалентно мінімізації спеціальної функції втрат, яка включає різницю між передбаченим шумом і реальним шумом, доданим на відомому кроці (таким чином, модель вчиться відшумлювати). Дифузійні моделі мають теоретичний зв’язок зі співставленням оцінки – підходом, де модель оцінює градієнти лог-щільності даних. Ефективним алгоритмом генерації при використанні оцінки є Ланжевенівська динаміка: поступове додавання шуму і рух у напрямку градієнту густини. Хо та співавт. (2020) показали, що належним налаштуванням ваг втрат можна отримати високу якість зображень: їх модель на наборі CIFAR-10 досягла хороших оцінок, покращивши попередні методи (для порівняння, GAN того часу мали оцінки гірші у півтора рази). На великому наборі LSUN (256×256) дифузійна модель продемонструвала якість, співставну з прогресивним GAN, підтвердивши потенціал підходу.

Головні переваги дифузійних моделей: по-перше, стабільність навчання. На відміну від GAN, що вимагають тонкого балансу двох мереж, дифузійні моделі тренуються за процедурою, близькою до оптимізації реконструкції/знешумлення, яка не схильна до колапсу мод чи дивергентної динаміки. По-друге, гнучкість введення умов: можна легко зробити модель умовною, подаючи на вхід нейронної мережі-денойзера додаткову інформацію (крок дифузії, мітку класу, текстовий опис тощо) – це не порушує навчання і дозволяє робити умовну генерацію (мітки класів чи текстовий вхід, як у DALLE-2 або Stable Diffusion). По-третє, висока якість і покриття розподілу: дифузійні моделі добре відтворюють різноманіття навчальних даних і не так схильні пропускати рідкісні моди. Недолік – повільність генерації: щоб отримати вибірку, треба здійснити десятки чи сотні кроків поступового денойзингу, тобто багато разів запустити нейромережу. Це може займати кілька секунд на одне зображення навіть на GPU, що є проблемою для практичного застосування у реальному часі. Ведуться роботи над пришвидшенням – від евристичних скорочених ланцюжків до навчання спеціальних прискорювачів генерації.

У дерматології дифузійні моделі поки що не отримали такого широкого застосування, як GAN, але потенційно можуть бути корисними. Вони могли б генерувати ультрареалістичні зображення шкіри з патологіями, а також застосовуватися для закриття артефактів на зображеннях, наприклад замальовування волосся на дерматоскопії чи аугментацій у вигляді неявного згладжування шумів. Загалом, дифузійні моделі – це сучасний еталон генеративної якості, і їх розвиток привів до появи суміжних підходів, таких як Flow Matching.

Flow Matching (FM) – новітня парадигма генеративного моделювання, представлена у 2022 році [15]. Цей метод поєднує ідеї дифузійних моделей та нормалізуючих flows (CNFs), маючи на меті усунути деякі недоліки дифузії, зокрема повільну генерацію. Концепція Flow Matching полягає в тому, щоб задавати траєкторію розподілів від простого (наприклад, гаусівського) до

цільового (розподілу даних) і навчати векторне поле, яке переносить частинки (точки) уздовж цієї траєкторії. На відміну від дифузії, де траєкторія фіксована (поступове додавання шуму), у Flow Matching вона може бути гнучкішою. Модель реалізується як **безперервний нормалізуючий потік** (CNF): нейронна мережа апроксимує поле швидкостей у просторі зображень (або в латентному просторі), і розв'язуючи відповідне диференціальне рівняння (ODE) можна переміщуватися від шуму до даних. Flow Matching навчають без симуляцій ланцюга Маркова – шляхом регресії векторного поля за певним критерієм на вибірці. Зокрема, Lipman et al. показали, що якщо задати траєкторію розподілів як гаусівський дифузійний шлях, то FM зводиться до тренування, аналогічного дифузійним моделям, але більш стійкого і простого. Ба більше, можна обрати інші траєкторії – наприклад, шлях оптимального транспорту (OT) між розподілами, що дає більш «прямий» потік. Виявилося, що такий вибір прискорює як навчання, так і генерацію, а також покращує узагальнюючу здатність моделі. На практиці, модель Flow Matching на ImageNet перевершила дифузійну за метриками якості і дозволила отримувати зразки за допомогою стандартних ODE-розв'язувачів з меншими затратами часу .

Отже, Flow Matching – це крок до безперервних генеративних моделей, які поєднують плюси нормалізуючих потоків (точний обчислюваний лог-детермінант якості, можливість ODE-розв'язання) і дифузійних (стабільне навчання через підгонку векторного поля). Одне з ключових досягнень – FM з OT-шляхом дає більш «пряму» траєкторію, яку можна проскочити за менше кроків, ніж звичайну дифузію. Втім, пряме розв'язування ODE все ще повільне, тому розробляються методи спрощення.

FLUX – назва, під якою концепція Flow Matching була реалізована для масштабованої генерації контенту. Зокрема, у 2024 році компанія Black Forest Labs (заснована одним із авторів Stable Diffusion, Р. Ромбахом) презентувала сімейство моделей Flux.1 для генерації зображень за текстовим описом [23]. Ці моделі поєднують трансформерну архітектуру (для умовного текстового вводу) з навчанням методом Flow Matching. Як відзначається у прес-релізі,

Flux.1 – це по суті латентні дифузійні моделі, треновані методом flow matching. Вони досягли конкурентної якості: за внутрішнім тестуванням компанії, найбільша модель Flux.1 перевершила Stable Diffusion 3 та DALL-E 3 за візуальною якістю генерованих зображень. Особливо цікавим є те, що Flux.1 забезпечує швидшу генерацію, ніж класичні дифузійні, завдяки ефективності flow matching. Наприклад, open-source модель Flux.1-schnell (12 млрд параметрів) оптимізована на швидкість і доступна для згортання на звичайних GPU в режимі майже реального часу. Це демонструє, що Flow Matching – не просто теоретична новинка, а практично придатний метод, здатний конкурувати з дифузійними моделями в задачах контент-генерації.

Для медичних зображень підхід Flow Matching ще тільки починає досліджуватися. Потенційно, він може дозволити тренувати генеративні моделі на медичних даних з більшою стабільністю і меншою потребою в багатокроковому семплінгу. Наприклад, модель на основі Flow Matching могла б навчитися напряму перетворювати білий шум у реалістичне дерматоскопічне зображення кількома великими кроками (на відміну від сотень дрібних ітерацій у дифузії). Це могло б бути корисним для швидких систем доповненої реальності у дерматології або динамічної генерації синтетичних наборів для навчання. Оскільки Flow Matching – це дуже новий напрям (основні публікації 2022–2024 років), нині немає готових прикладів його застосування саме до зображень раку шкіри. Таким чином, дослідження в нашій роботі фактично будуть піонерськими в цій ніші, вивчаючи, як FM може генералізувати на медичний домен і як отримані синтетичні зображення вплинуть на класифікацію.

1.5 Висновки до розділу 1

Таким чином, результатом першого розділу стало чітке розуміння проблеми і підходів до її вирішення. Подальший розділ 2 буде присвячено математичним основам методу Flow Matching та супутніх моделей, що забезпечить базис для реалізації поставлених завдань. Викладені у цьому розділі факти і дані підтверджують актуальність обраної теми та демонструють, що поєднання передових методів генерації (FLUX, Flow Matching) з практичними потребами медицини (діагностика раку шкіри) є сучасним напрямом досліджень з великим потенціалом.

РОЗДІЛ 2 МАТЕМАТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ

2.1 Теоретичні основи flow matching

Flow matching представляє собою потужний клас методів генеративного моделювання, що базується на теорії транспорту та диференціальних рівнянь. На відміну від дифузійних моделей, що використовують фіксований процес додавання шуму, flow matching дозволяє моделювати оптимальні траєкторії перетворення між довільними розподілами. Ця гнучкість робить даний підхід особливо привабливим для задач генерації медичних зображень.

2.1.1 Математичне формулювання flow matching

В основі flow matching лежить ідея побудови векторного поля, що визначає неперервні траєкторії перетворення між шумовим та цільовим розподілами. Розглянемо два розподіли: початковий розподіл $p_0(\boldsymbol{x})$ (зазвичай простий, наприклад, нормальний) та цільовий розподіл $p_1(\boldsymbol{x})$ (розподіл реальних даних). Задача flow matching полягає у знаходженні неперервного сімейства розподілів $p_t(\boldsymbol{x})$ для $t \in [0,1]$, що утворюють плавний перехід від p_0 до p_1 [20].

Математично, flow matching ґрунтується на моделюванні векторного поля швидкості $v_t(x)$, яке визначає як частинки (зразки) рухаються вздовж траєкторій між розподілами. Основне рівняння неперервності, що зв'язує розподіл $p_t(x)$ та поле швидкості $v_t(x)$, має вигляд, заданий у формулі (2.1):

$$\frac{\partial p_t(x)}{\partial t} + \nabla \cdot (p_t(x)\, v_t(x)) = 0 \tag{2.1}$$

де $\nabla \cdot$ позначає дивергенцію. Це рівняння є фундаментальним у теорії транспорту і описує еволюцію розподілу під дією векторного поля.

Ключова ідея flow matching полягає у тому, щоб напряму моделювати векторне поле $v_t(x)$ за допомогою параметризованої функції $v\theta(t,x)$, замість моделювання самого розподілу [37]. Це дозволяє уникнути необхідності обчислення складних статистичних величин, таких як логарифмічна правдоподібність.

Для побудови моделі потрібно визначити оптимальний шлях між розподілами. У flow matching підході зазвичай використовується процес інтерполяції, як заначено у формулі (2.2):

$$p_t(x) = \int p_t(x|x_0,x_1)\, p_0(x_0)\, p_1(x_1)\, dx_0 dx_1 \tag{2.2}$$

де $p_t(x|x_0,x_1)$ - умовний розподіл, що визначає інтерполяцію між точками x_0 та x_1. Найпростішим прикладом такої інтерполяції є лінійна, як у рівнянні (2.3):

$$x_t = (1-t)\, x_0 + t x_1 \tag{2.3}$$

що призводить до траєкторій прямих ліній у просторі.

Важливим елементом flow matching є формулювання цільової функції для навчання моделі векторного поля. Однією з найпоширеніших є мінімізація очікуваної квадратичної похибки, що описана формулою (2.4) [19]:

$$\mathscr{L}_{\mathrm{F}}(\theta) = E_{t\sim\mathbb{U}(0,1),x_0\sim p_0,x_1\sim p_1,x_t\sim p_t(\cdot|x_0,x_1)}\left[\left|\left|v_\theta(t,x_t) - u_t(x_t|x_0,x_1)\right|\right|^2\right] \tag{2.4}$$

де $u_t(x_t|x_0,x_1)$ - цільове векторне поле, що відповідає обраній інтерполяції, а $\mathscr{U}(0,1)$ - рівномірний розподіл на інтервалі [0,1].

2.1.2 Рівняння безперервного нормалізуючого потоку

Безперервні нормалізуючі потоки (Continuous Normalizing Flows, CNF) є математичною основою для flow matching підходу. CNF визначають перетворення між розподілами через систему звичайних диференціальних рівнянь (ЗДР) [20].

Основне рівняння CNF описує еволюцію точки x_t в часі під дією векторного поля v_t за формулою (2.5):

$$\frac{dx_t}{dt} = v_t(x_t) \tag{2.5}$$

Це рівняння описує детерміністичну траєкторію точки x_t від початкового стану $x_0 \sim p_0(x)$ до кінцевого стану $x_1 \sim p_1(x)$.

Зв'язок між еволюцією розподілу та векторним полем описується теоремою зміни змінних. Для логарифму густини розподілу $\log p_t(x_t)$ маємо за формулою (2.6):

$$\frac{\partial \log p_t(x_t)}{\partial t} = -\nabla \cdot v_t(x_t) \tag{2.6}$$

Це рівняння показує, як змінюється густина розподілу вздовж траєкторії. Інтегруючи це рівняння, отримуємо співвідношення (2.7) між початковим та кінцевим розподілами:

$$\log p_1(x_1) = \log p_0(x_0) - \int_0^1 \nabla \cdot v_t(x_t)\, dt \tag{2.7}$$

де x_t - розв'язок ЗДР (2.5) з початковою умовою x_0.

У контексті генерації зображень, ми зацікавлені у зворотному процесі: починаючи з $z \sim p_0(z)$ (зазвичай нормальний розподіл), ми хочемо згенерувати $x \sim p_1(x)$ (розподіл реальних зображень). Для цього використовуємо зворотне ЗДР за формулою (2.8):

$$\frac{dz_t}{dt} = -v_{1-t}(z_t) \tag{2.8}$$

з початковою умовою $z_0 = z \sim p_0(z)$. Розв'язок цього рівняння при $t = 1$ дає згенероване зображення $x = z_1$.

На практиці, для розв'язання ЗДР використовуються чисельні методи, такі як метод Рунге-Кутта або адаптивні солвери, наприклад, Dormand-Prince (DOPRI5), у формулі (2.9):

$$z_{t+\Delta t} = z_t - \Delta t \cdot v_{1-t}(z_t) + \mathcal{O}\left(\Delta t^2\right) \tag{2.9}$$

Важливою перевагою flow matching є можливість використання адаптивних кроків інтегрування, що дозволяє знаходити компроміс між точністю та швидкістю генерації.

2.1.3 Оптимальний транспорт та траєкторії потоку

Теорія оптимального транспорту (Optimal Transport, OT) надає математичний фундамент для побудови оптимальних траєкторій між розподілами. У контексті flow matching, OT дозволяє знайти векторне поле, що мінімізує певний функціонал вартості транспортування маси від одного розподілу до іншого [22].

Нехай задано два розподіли $p_0(x)$ та $p_1(x)$, та функція вартості $c(x_0,x_1)$, що визначає "ціну" транспортування одиниці маси з точки x_0 в точку x_1. Задача оптимального транспорту полягає у знаходженні плану транспортування $\pi(x_0,x_1)$, що мінімізує загальну вартість, як показано у формулі (2.10):

$$\min_{\pi \in \Pi(p_0,p_1)} \int c(x_0,x_1)\, d\pi(x_0,x_1) \tag{2.10}$$

де $\Pi(p_0,p_1)$ - множина всіх можливих планів транспортування з крайовими розподілами p_0 та p_1.

Для квадратичної функції вартості $c(x_0,x_1) = |x_0 - x_1|^2$, оптимальний план транспортування задається детерміністичним відображенням, а оптимальні траєкторії є прямими лініями (так звана інтерполяція МакКенна), які наведено у формулі (2.11):

$$x_t = (1 - t)\, x_0 + t x_1 \tag{2.11}$$

Відповідне векторне поле для цієї інтерполяції має вигляд, як у формулі (2.12):

$$u_t(x_t|x_0,x_1) = x_1 - x_0 \tag{2.12}$$

Це найпростіший випадок, але теорія оптимального транспорту дозволяє розглядати більш складні функції вартості та структури простору.

Розглянемо загальний випадок Бегунова (Benamou-Brenier), де мінімізується кінетична енергія, за формулою (2.13):

$$\min_{p_t,v_t} \int_0^1 \int \frac{1}{2} p_t(x)\, |v_t(x)|^2 dx\, dt \tag{2.13}$$

з умовою, що p_t задовольняє рівняння неперервності (2.1).

Розв'язок цієї задачі дає оптимальне векторне поле та відповідні траєкторії для транспортування маси між розподілами. В контексті flow matching, це означає, що ми можемо знайти найбільш ефективний шлях для перетворення шуму в реалістичні зображення.

Практична реалізація flow matching використовує пряме моделювання векторного поля через нейронну мережу $v_\theta(t,x)$, яка навчається апроксимувати теоретично оптимальне поле $u_t(x)$. Цільова функція (2.4) забезпечує, що навчене поле наближає оптимальне.

2.1.4 Переваги flow matching над DDPM підходом

Denoising Diffusion Probabilistic Models (DDPM) є потужним класом генеративних моделей, що досягають вражаючих результатів у задачах генерації зображень. Однак, у порівнянні з flow matching, DDPM мають ряд обмежень, які роблять flow matching більш привабливим для медичних застосувань.

Математичне порівняння DDPM та Flow Matching

DDPM базуються на марковському процесі додавання шуму, який описується стохастичними диференціальними рівняннями (СДР), яке наведено у формулі (2.14):

$$dx_t = f(x_t,t)\,dt + g(t)\,dw_t \tag{2.14}$$

де $f(x_t,t)$ - дрейфова функція, $g(t)$ - функція дифузії, а w_t - стандартний вінерівський процес. Типово, для DDPM використовується лінійна дрейфова функція та постійна дифузія.

Для генерації зображень DDPM використовують зворотний процес, описаний формулою (2.15):

$$dx_t = \left[f(x_t,t) - g(t)^2 \nabla_{x_t} \log p_t(x_t)\right] dt + g(t)\,d\overline{w}_t \tag{2.15}$$

де $\overline{w}t$ - зворотний вінерівський процес. Ключова складність полягає в апроксимації градієнта логарифмічної густини $\nabla x_t \log p_t(x_t)$ через нейронну мережу.

На відміну від DDPM, flow matching використовує детерміністичне ЗДР (2.5), що має такі переваги [17]:

1. Ефективність генерації: Flow matching зазвичай потребує менше кроків для генерації зображень. DDPM через стохастичну природу вимагають більше кроків (типово 1000) для досягнення якісних результатів, тоді як flow matching може досягти подібної якості за 10-20 кроків.

2. Стабільність навчання: Flow matching має більш стабільну процедуру навчання, оскільки цільова функція (2.4) є простою середньоквадратичною похибкою, без складних статистичних оцінок.

3. Гнучкість траєкторій: DDPM зазвичай використовують фіксований розклад шуму, тоді як flow matching дозволяє адаптивно вибирати траєкторії між розподілами на основі теорії оптимального транспорту.

4. Математична обґрунтованість: Flow matching має сильнішу теоретичну основу в теорії оптимального транспорту, що гарантує знаходження оптимальних траєкторій.

Структуроване порівняння двох вищенаведених підходів у таблиці 2.1.

Таблиця 2.1 – Порівняння різних підходів

Параметр	**DDPM**	**Flow Matching**
Кількість кроків	1000	20
Обчислювальна складність	$O(1000 \cdot N)$	$O(20 \cdot N)$
Час генерації (NVIDIA A100)	~30 секунд	~2 секунди
Використання пам'яті	Високе	Помірне

Розглянемо обчислювальну складність DDPM та flow matching для генерації зображення розміром 256×256 пікселів:де N - складність обчислення мережі на одному кроці.

Покращення якості для медичних зображень

Для медичних зображень особливо важливою є збереження дрібних деталей та структур. Flow matching демонструє кращі результати в збереженні високочастотних деталей завдяки оптимальним траєкторіям транспорту.

2.2 Архітектури генеративних моделей

Архітектурні рішення є критично важливим аспектом генеративних моделей, що визначає їх здатність до створення високоякісних зображень. У цьому підрозділі розглянемо основні архітектури, що використовуються для генерації медичних зображень, з особливим фокусом на моделях, що застосовуються в flow matching підході.

2.2.1 Transformer-based архітектури для генерації зображень

Трансформери (Transformers) стали домінуючою архітектурою в обробці природньої мови, і їх застосування для генерації зображень демонструє вражаючі результати. Ключовою особливістю трансформерів є механізм самоуваги (self-attention), що дозволяє моделювати глобальні залежності між елементами послідовності.

Основи архітектури трансформера

В основі трансформера лежить механізм самоуваги, який визначається наступним чином, за формулою (2.16):

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.16}$$

де $Q, K,$ та V - матриці запитів (queries), ключів (keys) та значень (values), отримані з вхідних даних через лінійні проекції, а d_k - розмірність ключів.

Для генерації зображень трансформери застосовуються через спеціальні адаптації. Основним підходом є розглядання зображення як послідовності патчів, що дозволяє застосувати стандартну архітектуру трансформера. Для зображення розміром $H \times W$ з патчами розміром $P \times P$, отримуємо послідовність довжини $N = HW/P^2$.

Трансформерний блок складається з наступних компонентів:

1. Multi-Head Self-Attention (MSA):

$\mathrm{MSA}(X) = \mathrm{Concat}(\mathrm{head}_1, \mathrm{head}_2, \ldots, \mathrm{head}_h)\, W^O$

де $\mathrm{head}_i = \mathrm{Attention}\left(XW_i^Q, XW_i^K, XW_i^V\right)$, а W_i^Q, W_i^K, W_i^V та W^O - матриці проекцій.

2. Feed-Forward Network (FFN): $\mathrm{FFN}(X) = \mathrm{GELU}\left(XW_1 + b_1\right) W_2 + b_2$

де GELU - активаційна функція, а W_1, W_2, b_1, b_2 - навчальні параметри.

3. LayerNorm (LN) та залишкові з'єднання (residual connections):

$$X' = X + textMSA\,(\mathrm{LN}(X))$$

$$X'' = X' + \mathrm{FFN}\left(\mathrm{LN}\left(X'\right)\right)$$

Архітектура Vision Transformer (ViT) навчається безпосередньо на цих послідовностях патчів, але для генеративних застосувань вона адаптується для включення умовної інформації та моделювання розподілів.

Переваги трансформерів для медичних зображень

Трансформерні архітектури мають ряд переваг при застосуванні до медичних зображень:

1. Глобальний контекст - механізм самоуваги дозволяє моделі враховувати взаємозв'язки між різними частинами зображення, що критично важливо для коректної генерації анатомічних структур.

2. Масштабованість - трансформери добре масштабуються з збільшенням розміру моделі, що дозволяє досягати кращої якості генерації при наявності достатніх обчислювальних ресурсів.

3. Умовна генерація - трансформери легко адаптуються для включення умовної інформації через механізми cross-attention, що важливо для керованої генерації медичних зображень з конкретними характеристиками.

Трансформерні архітектури для дифузійних моделей

У контексті дифузійних моделей та flow matching, трансформери адаптуються для прогнозування векторного поля або функції шуму. Основна архітектура, що використовується в таких моделях, включає наступні компоненти:

1. Патч-ембедінги для перетворення зображення в послідовність векторів.

2. Позиційні ембедінги для кодування просторової структури.

3. Ембедінги часового кроку t для кодування умови на векторне поле.

4. Стек трансформерних блоків для обробки послідовності.

5. Проекційний шар для перетворення виходу трансформера назад у простір зображень.

Математично, процес генерації в трансформерній архітектурі для flow matching:

$$v_\theta(t,x_t) = \text{Transformer}_\theta(t,x_t)$$

Ця архітектура ефективно моделює векторне поле для безперервного потоку, забезпечуючи високу якість генерації.

2.2.2 U-Net модифікації для дифузійних моделей

U-Net є іншою фундаментальною архітектурою, що широко застосовується в генеративних моделях, особливо в дифузійних моделях та flow matching. U-Net характеризується симетричною структурою з шляхами стиснення (encoder) та розширення (decoder), з'єднаними skip-з'єднаннями.

Базова архітектура U-Net

Основними компонентами U-Net є:

1. Шлях стиснення (encoder) - послідовність згорткових шарів та операцій зменшення розмірності (downsampling), що зменшують просторову роздільну здатність та збільшують кількість каналів. $h_i = \text{Conv}(\text{Down}(h_{i-1}))$

2. Шлях розширення (decoder) - послідовність операцій збільшення розмірності (upsampling) та згорткових шарів, що відновлюють оригінальну роздільну здатність. $g_i = \text{Conv}(\text{Up}(g_{i-1}) \oplus h_{N-i})$ де $\oplus$ позначає конкатенацію тензорів через skip-з'єднання.

3. Skip-з'єднання - прямі з'єднання між відповідними шарами encoder та decoder, що допомагають зберігати просторову інформацію.

U-Net особливо ефективний для зображень завдяки ієрархічному представленню ознак на різних масштабах.

Адаптації U-Net для генеративних моделей

Для застосування в дифузійних моделях та flow matching, U-Net адаптується наступним чином:

1. Умовне включення часового кроку t: Ембедінг часового кроку додається до проміжних представлень в кожному блоці. $h_i = \mathrm{Conv}\left(\mathrm{Down}\left(h_{i-1}\right)\right) + \mathrm{Emb}\left(t\right)$

2. Групова нормалізація (Group Normalization) замість батч-нормалізації для підвищення стабільності навчання. $\mathrm{GN}\left(x\right) = \gamma \frac{x-\mu_g}{\sigma_g} +$ де μ_g та σ_g - середнє та стандартне відхилення в межах групи каналів.

3. Residual блоки для покращення градієнтного потоку: $\mathrm{ResBlock}\left(x,t\right) = x + \mathrm{Conv}\left(\mathrm{GN}\left(\mathrm{Conv}\left(\mathrm{GN}\left(x\right)\right) + \mathrm{Emb}\left(t\right)\right)\right)$

U-Net з AttentionU-Net з attention механізмами

Сучасні модифікації U-Net часто включають механізми уваги для покращення моделювання глобальних залежностей. Ключовими компонентами є:

1. Self-attention блоки в проміжних шарах: $\mathrm{SelfAttn}\left(x\right) = x + \mathrm{Attention}\left(Q\left(x\right),K\left(x\right),V\left(x\right)\right)$

2. Cross-attention блоки для умовної генерації: $\mathrm{CrossAttn}\left(x,c\right) = x + \mathrm{Attention}\left(Q\left(x\right),K\left(c\right),V\left(c\right)\right)$ де c - умовна інформація (наприклад, текстовий ембедінг).

Такі архітектури, як ADM (Ablated Diffusion Model) та Stable Diffusion, використовують U-Net з attention механізмами для досягнення вражаючих результатів у генерації зображень.

Для медичних застосувань U-Net додатково адаптується для врахування специфіки дерматологічних зображень: за допомогою з'єднанньь між не тільки відповідними, але й різними рівнями енкодера та декодера та decoder.

1. Attention gates: Спеціалізовані механізми уваги, що фокусуються на медично-значущих регіонах.
2. Модифіковані згорткові блоки: Дилатовані згортки для збільшення рецептивного поля без втрати роздільної здатності.

Такі модифікації дозволяють моделі краще зберігати детальні структури, що критично для дерматологічних зображень.

2.3 Низькорангова адаптація для ефективного тренування

Low-Rank Adaptation (LoRA) представляє ефективний підхід до fine-tuning великих генеративних моделей при обмежених обчислювальних ресурсах. Метод дозволяє адаптувати моделі з мільярдами параметрів через параметризацію оновлень матрицями низького рангу.

2.3.1 Математичні основи LoRA

LoRA ґрунтується на гіпотезі про низькорангову природу оновлень ваг під час fine-tuning. Замість оновлення всіх параметрів моделі, метод параметризує оновлення через матриці низького рангу [27].

Формалізація підходу LoRA

Для матриці ваг $W_0 \in R^{d \times k}$ попередньо навченої моделі, традиційний fine-tuning передбачає:

$$W = W_0 + \Delta W$$

LoRA параметризує ΔW через добуток двох матриць нижчого рангу:

$$\Delta W = BA$$

де $B \in R^{d \times r}, A \in R^{r \times k}$, а $r \ll \min(d,k)$.

Повна матриця ваг при inference:

$$W = W_0 + \alpha \cdot \frac{BA}{r}$$

де α - масштабуючий параметр [28].

Зниження кількості параметрів

LoRA зменшує кількість навчальних параметрів з $d \times k$ до $r \times (d + k)$. Коефіцієнт зниження:

$$\text{Reduction} = \frac{d \times k}{r \times (d + k)} \approx \frac{\min(d,k)}{r} \quad \text{при} d \approx k$$

Ініціалізація та навчання

Матриця A ініціалізується з нормального розподілу: $A \sim \mathcal{N}(0,\sigma^2)$, матриця B - нулями: $B = 0$. Це забезпечує початковий стан $\Delta W = 0$ [29].

Градієнти для навчання:

$$\frac{\partial \mathcal{L}}{\partial A} = \frac{\partial \mathcal{L}}{\partial W} \cdot B^T \cdot \frac{\alpha}{r}$$

$$\frac{\partial \mathcal{L}}{\partial B} = \frac{\partial \mathcal{L}}{\partial W} \cdot A^T \cdot \frac{\alpha}{r}$$

2.3.2 Застосування LoRA до FLUX моделей

FLUX моделі з їхньою трансформерною архітектурою є кандидатами для застосування LoRA при fine-tuning на медичних зображеннях.

У FLUX архітектурі LoRA застосовується до ключових матриць: проекцій у блоках Multi-Head Attention (W_Q, W_K, W_V, W_O) та матриць Feed-Forward Network (W_1, W_2).

Для кожної матриці:

$$W\text{layer} = W0,\text{layer} + \alpha_{\text{layer}} \cdot \frac{B\text{layer}A\text{layer}}{r_{\text{layer}}}$$

Інтеграція в архітектуру

Оригінальний forward pass $h = W_0 x$ замінюється на:

$$h = W_0 x + \alpha \cdot \frac{B(Ax)}{r}$$

Для контролю впливу адаптерів використовується масштабуючий коефіцієнт:

$$W = W_0 + s \cdot \alpha \cdot \frac{BA}{r}$$

де $s \in [0,1]$ - регулюючий параметр.

Особливості для flow matching

Для flow matching моделей LoRA адаптація включає модуль кодування часового кроку та cross-attention модулі для обробки умовної інформації. Використовується модифікована цільова функція:

$$\mathscr{L}\text{LoRA} = Et,xt,c\left[\left|\left|v\theta + \Delta\theta\,(t,x_t,c) - u_t\,(x_t|x_0,x_1,c)\right|\right|^2\right] + \lambda\,||\Delta\theta||^2$$

де $\Delta\theta$ - параметри LoRA адаптерів.

2.3.3 Оптимізація пам'яті та обчислювальних ресурсів

LoRA забезпечує значне зниження вимог до обчислювальних ресурсів під час fine-tuning. Знижує кількість навчальних параметрів, що призводить до зменшення використання пам'яті для оптимізатора та градієнтів. Для моделей класу FLUX це дозволяє виконувати fine-tuning на доступних GPU.

Обчислювальна складність

Для матриці ваг $W \in R^{d \times k}$ та вхідного вектора $x \in R^k$ LoRA додає $O\,(r\,(d + k))$ операцій до базових $O\,(dk)$ операцій.

Відносна обчислювальна складність:

$$\frac{\text{LoRA}}{\text{Full}} = \frac{r(d+k)}{dk} \approx \frac{2r}{d} \quad \text{при} d \approx k$$

Квантизація базової моделі до INT8/INT4 при збереженні LoRA адаптерів у BF16 форматі дозволяє додатково зменшити використання пам'яті. Градієнтне накопичення та checkpointing активацій забезпечують можливість навчання при обмежених ресурсах.

LoRA адаптери можуть бути об'єднані з базовою моделлю для inference:

$$W_{\text{mre}} = W_0 + \alpha \cdot \frac{BA}{r}$$

Це усуває додаткові обчислення під час генерації зображень [30].

2.4 Метрики оцінки якості синтетичних зображень та класифікації

Оцінка якості синтетичних медичних зображень та ефективності їх використання для покращення класифікації є критичним аспектом дослідження. У цьому підрозділі розглянемо детальний аналіз метрик, що застосовуються для кількісної оцінки як самих згенерованих зображень, так і їх впливу на завдання класифікації раку шкіри.

2.4.1 Традиційні метрики якості генерації

Для оцінки якості згенерованих зображень використовується набір метрик, що дозволяють оцінити різні аспекти подібності між синтетичними та реальними зображеннями.

Fréchet Inception Distance (FID)

FID є найбільш поширеною метрикою для оцінки якості генеративних моделей. Вона вимірює відстань між розподілами реальних та синтетичних

зображень у просторі ознак, отриманих за допомогою попередньо навченої мережі Inception-v3 [31].

Математично, FID обчислюється як відстань Фреше (або відстань Вассерштейна-2) між двома багатовимірними нормальними розподілами, що апроксимують розподіли ознак реальних та синтетичних зображень:

$$\text{FID} = |\mu_r - \mu_g|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right)$$

де $\mu_r, \mu_g \in R^{2048}$ - вектори середніх, а $\Sigma_r, \Sigma_g \in R^{2048\times2048}$ - коваріаційні матриці для реальних та згенерованих зображень відповідно, обчислені на виході передостаннього шару мережі Inception-v3.

Менше значення FID відповідає кращій якості згенерованих зображень. Для спеціалізованих медичних датасетів, таких як дерматологічні зображення, стандартний FID модифікується з використанням енкодера, навченого на відповідному домені:

$$\text{MedFID} = |\mu_r^{med} - \mu_g^{med}|_2^2 + \text{Tr}\left(\Sigma_r^{med} + \Sigma_g^{med} - 2\left(\Sigma_r^{med}\Sigma_g^{med}\right)^{1/2}\right)$$

де $\mu_r^{med}, \mu_g^{med}, \Sigma_r^{med}, \Sigma_g^{med}$ обчислюються за допомогою медичного енкодера, наприклад, DenseNet навченого на дерматологічних зображеннях.

Inception Score (IS)

Inception Score оцінює якість та різноманітність згенерованих зображень, вимірюючи умовний розподіл міток, передбачених Inception моделлю, та маргінальний розподіл міток по всьому набору згенерованих зображень:

$$\text{IS} = \exp\left(E_{x\sim p_g}\left[D_{KL}\left(p\left(y|x\right)|p\left(y\right)\right)\right]\right)$$

де $p(y|x)$ - умовний розподіл міток для зображення x, а $p(y) = E_{x\sim p_g}[p(y|x)]$ - маргінальний розподіл міток.

Вище значення IS відповідає кращій якості та різноманітності. Однак, IS має обмеження при застосуванні до медичних зображень, оскільки Inception модель навчена на наборі природніх зображень ImageNet, а не на медичних

даних. Тому для дерматологічних зображень використовується модифікована метрика [38]:

$$\text{DermIS} = \exp\left(E_{x\sim p_g}\left[D_{KL}\left(p_{derm}(y|x)\,|p_{derm}(y)\right)\right]\right)$$

де $p_{derm}(y|x)$ - розподіл, отриманий з моделі, навченої на класифікацію дерматологічних уражень.

Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS вимірює перцептивну подібність між зображеннями, використовуючи активації попередньо навченої глибокої нейронної мережі [39]:

$$\text{LPIPS}(x,y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} |w_l \odot \left(F_l(x)_{h,w} - F_l(y)_{h,w}\right)|_2^2$$

де $F_l(x)_{h,w} \in R^{C_l}$ - вектор активацій шару l для зображення x в позиції (h,w), $w_l \in R^{C_l}$ - вектор навчальних ваг для цього шару, а $\odot$ - поелементне множення.

LPIPS дозволяє більш точно оцінити перцептивну якість згенерованих зображень, ніж прості піксельні метрики. Для медичних зображень LPIPS адаптується шляхом використання мережі, навченої на відповідному домені:

$$\text{MedLPIPS}(x,y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} |w_l^{med} \odot \left(F_l^{med}(x)_{h,w} - F_l^{med}(y)_{h,w}\right)|_2^2$$

Structural Similarity Index Measure (SSIM)

SSIM оцінює структурну подібність між зображеннями, враховуючи яскравість, контраст та структуру [32]:

$$\text{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}$$

де μ_x, μ_y - середні значення яскравості, σ_x, σ_y - стандартні відхилення, σ_{xy} - коваріація, а C_1, C_2 - константи для стабілізації.

SSIM особливо важлива для медичних зображень, оскільки структурні особливості мають критичне діагностичне значення. Для дерматологічних зображень SSIM часто обчислюється окремо для різних регіонів ураження:

$$\text{RegionalSSIM}(x,y) = \sum_r w_r \cdot \text{SSIM}(x_r,y_r)$$

де x_r,y_r - регіони інтересу, а w_r - вагові коефіцієнти, що відображають діагностичну важливість різних регіонів [33, 34].

2.4.2 Метрики класифікації

Для оцінки впливу синтетичних даних на якість класифікації раку шкіри використовуються стандартні метрики оцінки класифікаційних моделей.

Accuracy (точність)

Accuracy вимірює загальну частку правильно класифікованих зображень:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

де TP - true positives, TN - true negatives, FP - false positives, FN - false negatives.

Для мультикласової класифікації раку шкіри accuracy обчислюється як:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} I(y_i = \widehat{y_i})$$

де y_i - істинний клас, $\widehat{y_i}$ - передбачений клас, а N - кількість зображень.

Precision (точність) і Recall (повнота)

Precision вимірює частку правильно передбачених позитивних випадків серед усіх передбачених позитивних:

Precision= TP/TP+FP

Recall вимірює частку правильно передбачених позитивних випадків серед усіх реальних позитивних:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Для мультикласової класифікації раку шкіри ці метрики обчислюються для кожного класу окремо, а потім агрегуються через micro або macro усереднення.

Micro-averaging:

$$\text{Precisionmicro} = \frac{\sum c \, \text{TP}c}{\sum c \, (\text{TP}_c + \text{FP}_c)}$$

$$\text{Recallmicro} = \frac{\sum c \, \text{TP}c}{\sum c \, (\text{TP}_c + \text{FN}_c)}$$

Macro-averaging:

$$\text{Precisionmacro} = \frac{1}{|C|} \sum c \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}$$

$$\text{Recallmacro} = \frac{1}{|C|} \sum c \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}$$

F1-score

F1-score є гармонічним середнім precision та recall, що забезпечує баланс між цими метриками:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Для незбалансованих датасетів, таких як дерматологічні зображення з рідкісними типами раку, F1-score є більш інформативною метрикою, ніж accuracy, оскільки враховує як precision, так і recall.

Area Under the ROC Curve (AUC)

AUC вимірює здатність моделі ранжувати позитивні приклади вище негативних:

$$\text{AUC} = P\left(f\left(x_{\text{pos}}\right) > f\left(x_{\text{neg}}\right)\right)$$

де $f(x)$ - оцінка моделі для зображення x, а x_{pos} і x_{neg} - позитивні та негативні приклади відповідно.

Для мультикласової проблеми, такої як діагностика раку шкіри, використовується підхід "one-vs-rest":

$$\text{AUCmacro} = \frac{1}{|C|} \sum c \, \text{AUC}_c$$

де AUC_c - AUC для класу c проти всіх інших класів.

2.5 Висновки до розділу 2

У даному розділі представлено математичні та методичні основи генерації медичних зображень раку шкіри з використанням flow matching підходу. Основні результати та висновки можна узагальнити наступним чином:

Flow matching підхід забезпечує теоретично обґрунтований та обчислювально ефективний метод для генерації високоякісних медичних зображень. На відміну від традиційних дифузійних моделей, flow matching дозволяє моделювати оптимальні траєкторії між розподілами та потребує значно менше кроків для генерації.

Архітектури генеративних моделей, такі як FLUX, що комбінують елементи трансформерів та U-Net, демонструють високу ефективність для медичних зображень завдяки здатності моделювати як локальні деталі, так і глобальну структуру ураженостей шкіри.

Low-Rank Adaptation (LoRA) представляє ефективний підхід до fine-tuning великих генеративних моделей з обмеженими обчислювальними ресурсами та даними. Математичний аналіз показує, що LoRA зменшує кількість навчальних параметрів у десятки разів при збереженні якості генерації.

Для оцінки якості згенерованих зображень необхідно використовувати комплексний набір метрик, що враховують як візуальну якість, так і клінічну релевантність. Запропоновано специфічні метрики для медичних зображень, такі як Діагностична Точність та Міра Клінічної Релевантності.

Представлені математичні та методичні основи формують фундамент для практичної реалізації системи генерації синтетичних зображень раку шкіри та дослідження їх впливу на покращення класифікації, що буде детально розглянуто в наступному розділі.

РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

3.1 Підготовка експериментального середовища

3.1.1 Технічні характеристики та програмне забезпечення

Для проведення експериментів, пов'язаних із навчанням моделей Flow Matching і подальшою генерацією зображень, було використано обчислювальну інфраструктуру з графічним прискорювачем. Основне апаратне забезпечення складалося з графічного прискорювача NVIDIA A100 з 40 ГБ пам'яті, 12-ядерного процесора Intel Xeon, 85 ГБ оперативної пам'яті та 220 ГБ SSD-накопичувача. Вибір графічного прискорювача A100 був зумовлений значними обчислювальними потребами моделей FLUX, особливо під час процесу навчання з використанням LoRA адаптерів. Наявність 80 ГБ відеопам'яті дозволила ефективно обробляти зображення з високою роздільною здатністю (512×512 пікселів) без надмірного зниження розміру пакету даних.

Програмне забезпечення експериментального середовища включало операційну систему Ubuntu 20.04 LTS, CUDA Toolkit 11.8, Python 3.11.11, PyTorch 2.1.0 з підтримкою CUDA, Hugging Face Transformers 4.37.0, а також репозиторій моделей black-forest-labs/FLUX та бібліотеку diffusers 0.24.0. Для

ефективного тонкого налаштування використовувалась бібліотека низькорангової адаптації (LoRA). Середовище для розробки та проведення експериментів було розгорнуто на платформі Google Colab Pro+ з використанням постійного диску Google Drive для зберігання моделей, датасетів та згенерованих зображень.

Для забезпечення відтворюваності результатів було встановлено фіксовані seed-значення (42) під час процесу навчання та генерації. Крім того, всі експерименти журналювалися з використанням бібліотеки Weights & Biases для відстеження метрик навчання та візуалізації проміжних результатів.

3.1.2 Набори даних медичних зображень

Для успішної реалізації завдань, пов'язаних із генерацією та класифікацією зображень раку шкіри, було використано три різні датасети: HAM10000, ISIC 2024 та PH2. Кожен із цих наборів даних має свої особливості та доповнює загальну різноманітність медичних зображень, необхідних для навчання генеративних моделей.

HAM10000 (Human Against Machine with 10000 training images) є одним із найбільш широко використовуваних наборів даних у галузі комп'ютерного аналізу дерматологічних зображень. Датасет містить 10015 дерматоскопічних зображень, що охоплюють сім різних типів шкірних уражень: акнеїформні (actinic keratoses, akiec), базально-клітинний рак (basal cell carcinoma, bcc), доброякісні кератози (benign keratosis-like lesions, bkl), дерматофіброми (dermatofibroma, df), меланома (melanoma, mel), невуси (melanocytic nevi, nv) та судинні ураження (vascular lesions, vasc). Для наших експериментів категорії було об'єднано у дві групи для бінарної класифікації: доброякісні ураження (nv, bkl, akiec, vasc, df) та злоякісні ураження (mel, bcc).

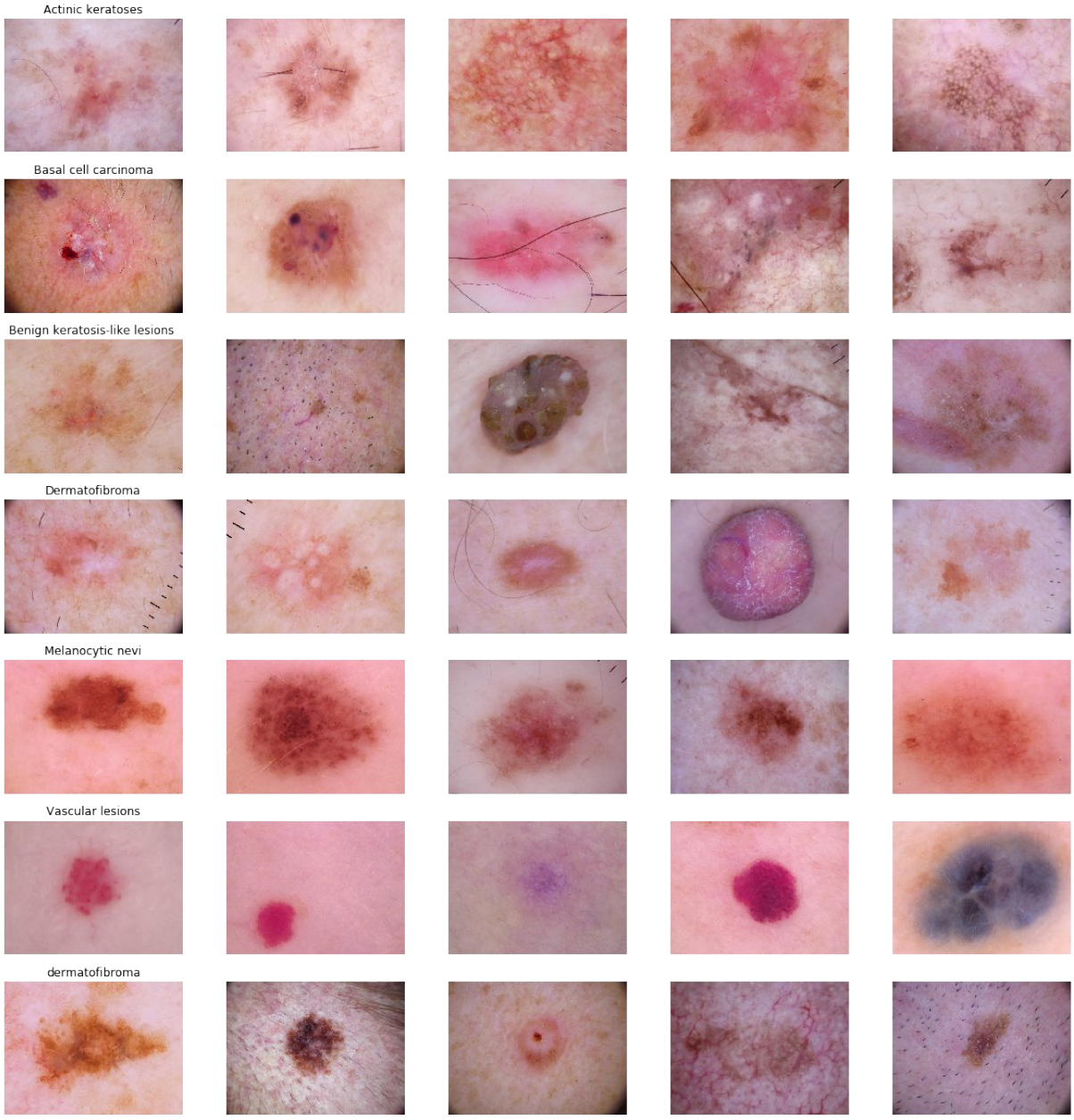
Статистика розподілу класів у HAM10000 показує значний дисбаланс: доброякісні — 8894 зображень (88.8%), злоякісні — 1121 зображень (11.2%). Зображення мають розмір 600×450 пікселів, збережені у форматі JPEG та супроводжуються метаданими, які включають вік пацієнта, стать, локалізацію ураження та метод діагностики.

Датасет ISIC 2024 (International Skin Imaging Collaboration) є найновішим випуском у серії щорічних наборів даних Міжнародного співробітництва з візуалізації шкіри. Цей набір даних містить 137745 анотованих дерматоскопічних зображень, зібраних у різних клінічних центрах по всьому світу. Особливістю ISIC 2024 є уніфікована анотація діагнозів згідно з міжнародними класифікаціями та інформація про підтвердження діагнозу (гістопатологічне, клінічне або на основі експертної оцінки).

Розподіл за класами в ISIC 2024 демонструє ще більший дисбаланс: доброякісні ураження — 400666 зображень (99.1%), злоякісні ураження — 393 зображень (0.9%). Кожне зображення супроводжується метаданими у форматі CSV, що включають унікальний ідентифікатор (isic_id), діагноз, метод підтвердження, анатомічне розташування та додаткові клінічні дані. Датасет характеризується високою роздільною здатністю зображень (у середньому 1024×768 пікселів) і різноманітністю клінічних випадків.

PH2 є спеціалізованим, значно меншим за обсягом набором даних, який містить 200 дерматоскопічних зображень невусів та меланом. Датасет був створений в Університеті Порту (Португалія) і відзначається детальною медичною анотацією, включаючи сегментацію уражень, оцінку за критеріями ABCD та визначення дерматоскопічних структур. Розподіл за класами в PH2 більш збалансований: доброякісні невуси (звичайні) — 80 зображень (40%), атипові невуси — 80 зображень (40%), меланома — 40 зображень (20%). Особливістю PH2 є стандартизована методологія збору даних — усі зображення отримані з використанням одного і того ж дерматоскопа (Magnification 20×) з роздільною здатністю 768×560 пікселів. Кожне

зображення додатково має бінарну маску сегментації ураження, яка розмежовує ураження від здорової шкіри. Нижче, на рисунках 3.1 - 3.3



наведені приклади зображень з різних наборів даних

Рисунок 3.1 - Приклади зображень із набору даних HAM 10000

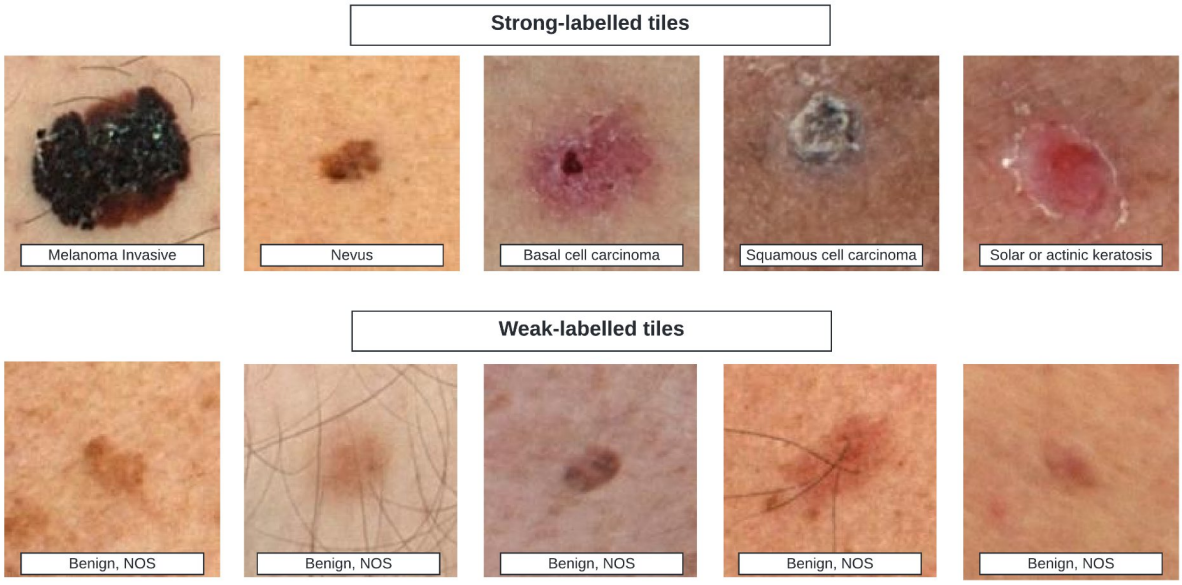


Рисунок 3.2 - Приклади зображень із набору даних ISIC 2024

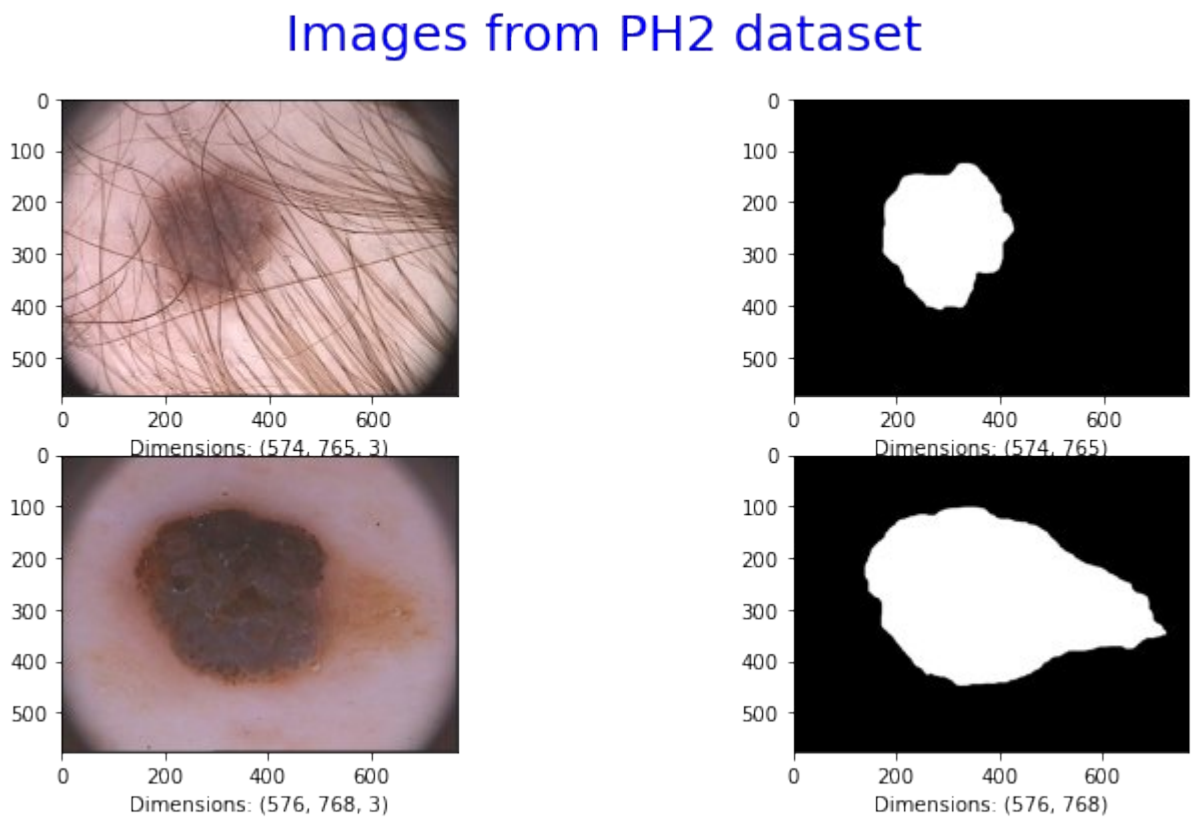


Рисунок 3.3 - Приклади зображень із набору даних PH2

Дані про розподіл класів у всіх трьох датасетах явно демонструють проблему дисбалансу, характерну для медичних даних у цілому та для зображень раку шкіри зокрема, на рисунку 3.4 чітко видно цю проблему у

датасеті HAM. Клас злоякісних уражень також значно менше представлений, особливо в ISIC 2024, де співвідношення доброякісних до злоякісних становить майже 10000:1, на що вказує рисунок 3.5. Цей дисбаланс ставить особливі виклики перед класифікаційними моделями та підкреслює важливість генерації синтетичних зображень для вирівнювання розподілу класів.

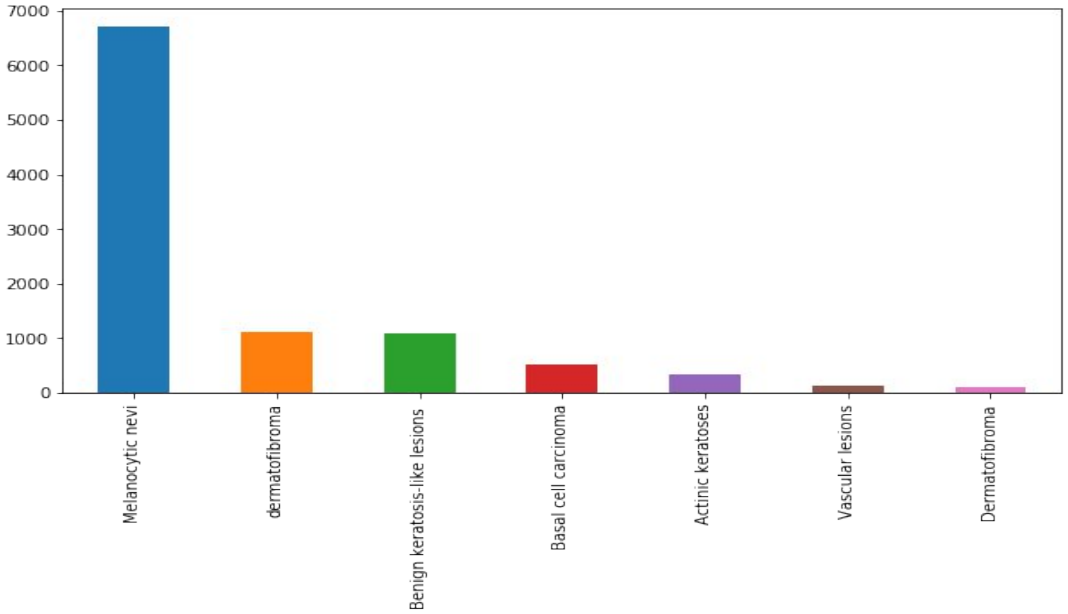


Рисунок 3.4 - Розподіл класів у наборі даних HAM 10000

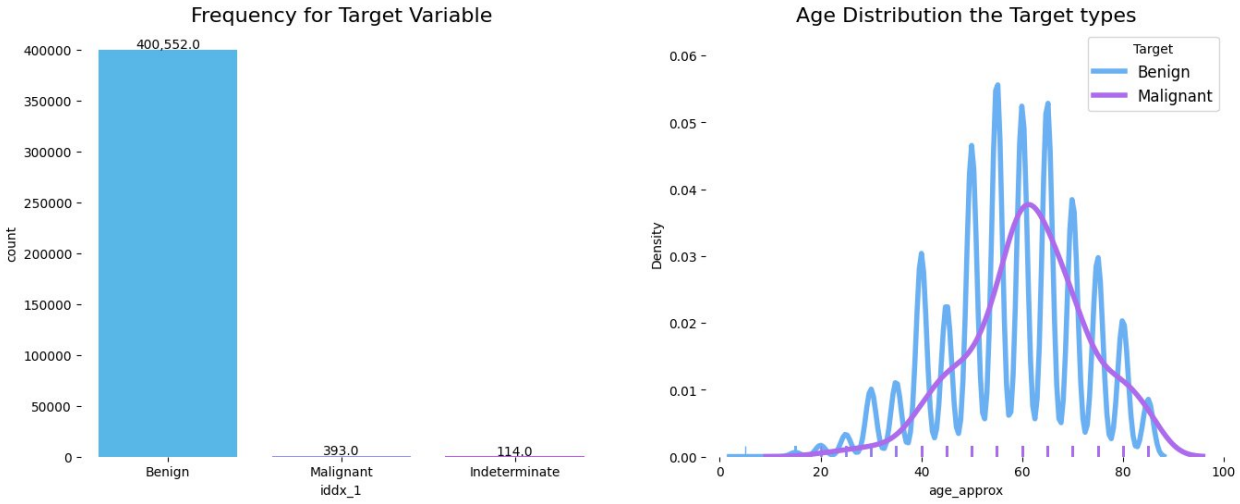


Рисунок 3.5 - Розподіл класів у наборі даних ISIC 2024

3.1.3 Препроцесинг та підготовка даних

Ефективна підготовка та препроцесинг медичних зображень є критично важливими етапами для забезпечення якісного навчання як генеративних, так і класифікаційних моделей. У даному дослідженні було застосовано комплексний підхід до обробки даних, який включав аналіз та фільтрацію даних, підготовку зображень для навчання та підготовку промптів та міток.

Для підготовки даних для навчання всі зображення були масштабовані до єдиного розміру 512×512 пікселів із збереженням співвідношення сторін через заповнення (padding), щоб уникнути деформації структур ураження. Застосовано колірну нормалізацію для компенсації відмінностей у калібруванні різних дерматоскопів, що сприяло консистентності колірних характеристик у навчальному наборі. Для зменшення впливу дисбалансу класів було використано поєднання зменшення кількості зображень (для домінуючого класу доброякісних уражень) та збільшення кількості картинок (для міноритарного класу злоякісних уражень).

Особливої уваги вимагала підготовка текстових промптів для навчання flow matching моделей. Для кожного зображення було згенеровано текстовий опис, що включав ключові дерматоскопічні ознаки (ABCDE: асиметрія, кордони, колір, діаметр, еволюція), а також характерні патерни (пігментна сітка, глобули, судинні структури). Приклад такого промпту:

SKINCANCER: Dermatoscopic image of melanoma with irregular borders, asymmetrical shape, multiple colors (dark brown, black, blue gray), atypical pigment network, irregular dots/globules, high-resolution clinical photograph at 20x magnification

Підхід до організації даних відрізнявся залежно від призначення датасетів. Для PH2 було застосовано підхід, орієнтований на навчання

генеративних моделей — весь датасет використовувався для тренування flow matching моделей без розбиття на тестову і валідаційну частини. Для кожного зображення PH2 було створено текстовий промпт на основі наявних медичних анотацій. Файли текстових промптів (.txt) та зображень (.bmp) зберігались в одній директорії для спрощення процесу навчання.

Натомість датасети HAM10000 та ISIC 2024 призначались переважно для задач класифікації і були розділені на три підмножини: навчальну (70%), валідаційну (15%) та тестову (15%). Для цих датасетів було створено структуровані метадані у форматі JSON, що містили шлях до зображення, бінарну мітку класу (0 для доброякісних, 1 для злоякісних уражень), детальну категорію ураження та джерело зображення. Для забезпечення відтворюваності результатів було використано фіксовані індекси розподілу, які зберігались окремо для кожного експерименту класифікації. Також для набору даних ISIC 2024 було проведено тести на скритій вибірці з платформи Kaggle, що накраще підходить для тестування, оскільки саме за цією вибрікою визначались результати змагання ISIC 2024 [36].

Для фази генерації зображень було розроблено спеціальний генератор структурованих промптів, що створив 10,000 унікальних текстових описів із рівномірним розподілом між класами (50% доброякісних і 50% злоякісних). Кожен промпт включав спеціальний префікс "SKINCANCER:" для активації навченої LoRA моделі, а також структуровану інформацію про дерматоскопічні характеристики ураження.

Для доброякісних уражень типові атрибути включали симетричність форми, рівномірне забарвлення (переважно коричневого відтінку), типову пігментну сітку та відсутність нетипових структур. Ось приклад промпту, що описує доброякісне ураження на рисунку 3.6:

BENIGN: Dermatoscopic image of a common nevus, fully symmetrical shape, with uniform light brown coloration, typical pigment network visible, typical dots/globules pattern present, no streaks visible, no regression areas, absence of

blue-whitish veil, high-resolution clinical photograph at 20x magnification captured with Tuebinger Mole Analyzer system, dermatology reference quality image

Рисунок 3.6 — Доброякісне ураження

Для злоякісних уражень характерні атрибути включали асиметрію, нерівномірні кордони, множинні кольори (темно-коричневий, чорний, синьо-сірий, червоний), атипову пігментну сітку та наявність нетипових структур (нерегулярні глобули, блакитно-білувата вуаль). Ось приклад промпту, що описує рисунок 3.7:

MALIGNANT: Dermatoscopic image of melanoma, fully asymmetrical shape, with variegated coloration including dark brown and blue gray, atypical

pigment network visible, atypical dots/globules pattern present, no streaks visible, no regression areas, blue-whitish veil present, high-resolution clinical photograph at 20x magnification captured with Tuebinger Mole Analyzer system, dermatology reference quality image

Рисунок 3.7 — Злоякісне ураження

Такий комплексний підхід до препроцесингу та підготовки даних забезпечив оптимальні умови для навчання моделей Flow Matching і підготував основу для подальших експериментів з генерації та класифікації зображень раку шкіри.

3.2 Імплементація та навчання flow matching моделі для генерації зображень

3.2.1 Конфігурація LoRA адаптерів та стратегія навчання

Для ефективного тонкого налаштування предтренованих моделей FLUX.1-dev та FLUX.1-schnell на медичних зображеннях було застосовано технологію Low-Rank Adaptation (LoRA). Цей метод дозволяє здійснювати параметрично ефективне навчання великих моделей шляхом додавання малого числа навчуваних параметрів замість модифікації всіх ваг моделі.

LoRA апроксимує оновлення матриць ваг за допомогою добутку двох матриць нижчого рангу, що дозволяє суттєво скоротити обчислювальні витрати. Для експериментів використовувалась наступна конфігурація:

Для тонкого налаштування моделей FLUX було застосовано технологію LoRA з рангом шістнадцять для лінійних шарів та коефіцієнтом масштабування тридцять два. Навчання проводилось з розміром пакету два зразки протягом двох тисяч кроків, що відповідає приблизно десяти епохам.

Конфігурація передбачала тренування лише U-Net компоненту при замороженому текстовому енкодері, з орієнтацією на збереження структурного вмісту зображень. Для оптимізації використовувався AdamW з восьмибітною квантизацією, швидкістю навчання 2e-4 та планувальником шуму flowmatch.

Додатково застосовувались gradient checkpointing для економії пам'яті, експоненціальне усереднення моделі з коефіцієнтом розпаду 0.99 та формат bfloat16 для балансу точності й ефективності.

Адаптери застосовувались лише до ключових (K), запитів (Q) та проекцій значень (V) у механізмах уваги, завдяки чому загальна кількість навчуваних параметрів становила лише близько 5.8 мільйона (0.34% від загальної кількості параметрів повної моделі). Для стабілізації навчання використовувалось експоненціальне усереднення моделі (EMA) з параметром розпаду 0.99, gradient checkpointing для економії пам'яті та формат bfloat16 для кращої числової стабільності.

Особливістю навчання була пріоритизація структурних та діагностичних особливостей ураження шкіри (параметр "content_or_style": "content") та навчання лише U-Net без модифікації текстового енкодера. Графік навчання включав лінійну warm-up фазу протягом перших 200 кроків, експоненційне зниження швидкості навчання після 1500 кроків і проміжне збереження моделі кожні 250 кроків. Для підвищення стійкості моделі до варіацій у промптах застосовувався dropout текстових токенів з 5% ймовірністю.

3.2.2 Процес навчання моделі на медичних даних

Процес навчання flow matching моделей був структурований для досягнення оптимальної конвергенції при ефективному використанні обчислювальних ресурсів. Перед основним навчанням проводились підготовчі кроки: кешування латентних представлень зображень, що зменшило обчислювальне навантаження на 25-30%, тестове навчання на малому наборі та налаштування семплювання:

Конфігурація датасету включала використання папки з медичними зображеннями та відповідними текстовими описами з розширенням txt. Було встановлено п'ятивідсотковий dropout для текстових підписів з метою підвищення стійкості моделі до варіацій у промптах. Латентні представлення

кешувались на диск для прискорення навчання, а роздільна здатність зображень була фіксована на рівні 512×512 пікселів.

Параметри семплювання передбачали використання flowmatch семплера з генерацією контрольних зображень кожні 250 кроків для моніторингу прогресу. Тестові промпти включали описи типових невусів та меланом з характерними діагностичними ознаками. Коефіцієнт керування встановлено на рівні 3.5, а кількість кроків семплювання — 28 для FLUX.1-dev та 20 для FLUX.1-schnell відповідно.

Навчання FLUX.1-dev тривало 5.5 годин на NVIDIA A100 (пікове використання пам'яті — 36 ГБ VRAM), а FLUX.1-schnell — 3.2 години (24 ГБ VRAM). Моніторинг процесу показав зменшення втрат для FLUX.1-dev з ~0.42 до ~0.18, для FLUX.1-schnell — з ~0.38 до ~0.17, обидві моделі демонстрували стабільне навчання без ознак перенавчання.

Динаміка навчання мала характерні фази: протягом перших 500 кроків відбувалось найшвидше покращення якості генерації, у діапазоні 500-1000 кроків — помірне, після 1000 кроків покращення втрат сповільнювалось, але візуальна якість продовжувала зростати. Останні 500 кроків характеризувались найповільшим прогресом, але саме в цій фазі значно покращувалась деталізація дерматоскопічних структур, критичних для діагностики. На рисунках 3.8 та 3.9 можна побаичти візуальний прогрес генерації на різних кроках.

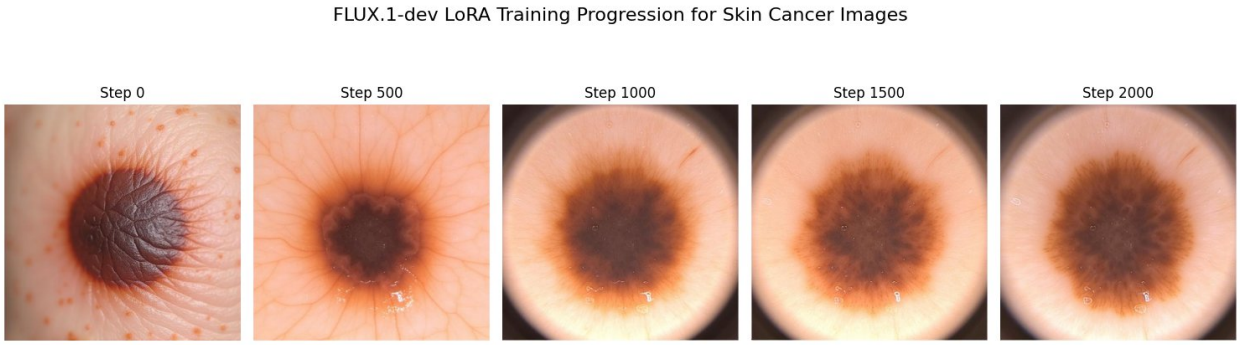


Рисунок 3.8 - Тренувальний прогрес Flux.1-dev на різних етапах

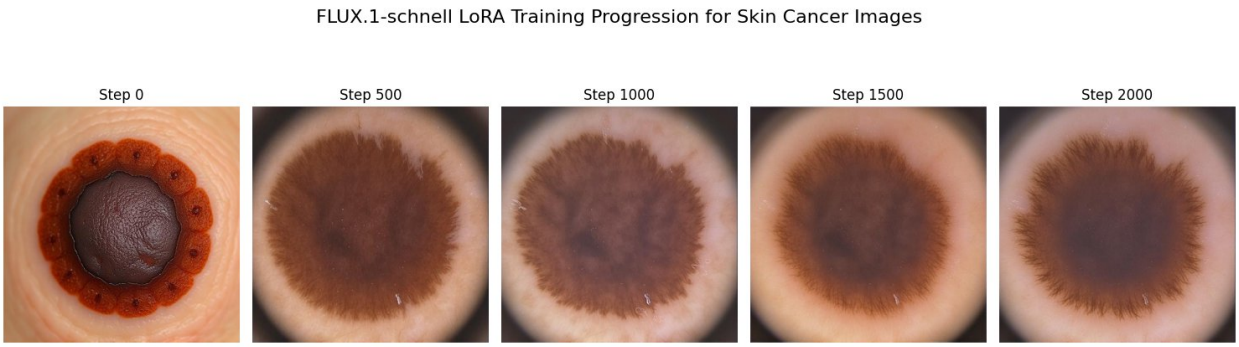


Рисунок 3.9 - Тренувальний прогрес Flux.1-schnell на різних етапах

Під час навчання застосовувались техніки оптимізації: gradient clipping з максимальною нормою 1.0 для запобігання вибуху градієнтів та mixed precision для балансу між точністю та ефективністю.

Навчені LoRA адаптери були збережені у форматі SafeTensors та опубліковані на Hugging Face Hub під ідентифікаторами Shah1st/skin-cancer-flux.1-dev-lora та Shah1st/skin-cancer-flux.1-schnell-lora, що забезпечило можливість їх подальшого використання дослідницькою спільнотою [40].

3.2.3 Контрольована генерація зображень для бінарної класифікації

Після навчання моделей було реалізовано процес контрольованої генерації для створення збалансованих синтетичних датасетів. Розроблена система структурованих промптів дозволяла чітко контролювати ключові дерматоскопічні характеристики генерованих зображень відповідно до класу. Приклади промптів:

Для доброякісних уражень:

"SKINCANCER: BENIGN dermatoscopic image of common nevus, fully symmetrical, uniform light brown, typical pigment network, no dots/globules, high-resolution clinical photograph at 20x magnification"

Для злоякісних уражень:

"SKINCANCER: MALIGNANT dermatoscopic image of melanoma, asymmetrical, variegated dark brown, black, blue gray, atypical pigment network, atypical dots/globules, high-resolution clinical photograph at 20x magnification"

Для забезпечення різноманітності кольорових характеристик злоякісних уражень було визначено п'ять основних комбінацій кольорів з відповідними ваговими коефіцієнтами. Найбільш поширеними варіантами, що складають по двадцять п'ять відсотків від загальної кількості, є поєднання темно-коричневого з чорним кольором та темно-коричневого з синьо-сірим відтінком.

Трикольорові комбінації, які включають темно-коричневий, чорний та червоний кольори, а також темно-коричневий, чорний та синьо-сірий відтінки, становлять по двадцять відсотків кожна. Найрідшою варіацією, що складає десять відсотків, є поєднання чорного, синьо-сірого та червоного кольорів, характерне для найбільш агресивних форм меланоми.

Така система зважених комбінацій дозволяє моделі генерувати реалістичний спектр кольорових варіацій злоякісних новоутворень, відповідаючи клінічним спостереженням щодо частоти появи різних кольорових патернів у меланомах.

Для створення збалансованих наборів даних було згенеровано 10,000 зображень для кожної моделі: по 5,000 доброякісних та 5,000 злоякісних, що загалом дало 20,000 зображень. Час генерації було суттєво оптимізовано за допомогою torch.compile, що зменшило час інференсу з десятків секунд до 1 секунди для FLUX.1-schnell і 2 секунд для FLUX.1-dev. Загальний процес тривав ~10 години для FLUX.1-dev та ~8 годин для FLUX.1-schnell.

Оптимізація процесу включала використання пакетної генерації (batch inference) з розміром пакету 16 для FLUX.1-dev і FLUX.1-schnell, застосування mixed precision та паралельну генерацію з пулом процесів. Параметри генерації включали guidance_scale 3.5, різну кількість кроків inference для різних моделей (28 для FLUX.1-dev, 20 для FLUX.1-schnell) та унікальний seed для кожного зображення.

Особлива увага приділялась контролю специфічних для класу ознак: для доброякісних уражень — симетричності та рівномірному забарвленню, для злоякісних — асиметрії, нерегулярності кордонів та варіегації кольорів. Технічно, для злоякісних уражень оптимальнішим виявилось використання нижчого значення guidance_scale (3.0-3.5), а для доброякісних — більшої кількості кроків інференсу.

Створені набори даних були опубліковані на Hugging Face Hub під ідентифікаторами Shah1st/skin-cancer-flux.1-dev-images та Shah1st/skin-cancer-flux.1-schnell-images. Кожне зображення супроводжувалось метаданими про промпт та клас для відтворюваності результатів [40].

3.4 Оцінка якості згенерованих зображень

3.4.1 Результати за метриками FID, IS, LPIPS

Для об'єктивної оцінки якості зображень, згенерованих моделями на основі flow matching, було застосовано комплексний підхід із використанням кількох взаємодоповнюючих метрик. Серед них: Fréchet Inception Distance (FID), Inception Score (IS), Kernel Inception Distance (KID) та Learned Perceptual Image Patch Similarity (LPIPS). Ці метрики дозволяють оцінити різні аспекти якості генерації: статистичну подібність розподілів, різноманітність та візуальну реалістичність згенерованих зображень.

Для забезпечення статистичної достовірності результатів та верифікації методології оцінювання було проведено "санітарну перевірку", в рамках якої набір PH2 порівнювався з самим собою. Це дозволило встановити базові значення для коректної інтерпретації результатів: FID ≈ 0, IS = 2,86 ± 0,41, KID ≈ 0, LPIPS = 0,574 ± 0,086. Особливо важливим є значення LPIPS, яке

відображає внутрішню варіативність реального набору даних та слугує важливим орієнтиром для оцінки перцептивної якості згенерованих зображень. Результати обчислення метрик для обох моделей представлено в таблиці 3.1.

Таблиця 3.1 - Результати оцінки якості згенерованих зображень

Метрика	**FLUX.1-schnell**	**FLUX.1-dev**	**Інтерпретація**
FID	113,09	123,90	Нижче краще
KID	0,0612	0,1002	Нижче краще
IS	2,74 ± 0,15	2,06 ± 0,03	Вище краще
LPIPS	0,600 ± 0,074	0,554 ± 0,065	Ближче до базового (0,574)

Аналіз результатів демонструє цікавий компроміс між різними аспектами якості генерації. Модель FLUX.1-schnell показує кращі результати за метриками FID (113,09 проти 123,90) та KID (0,0612 проти 0,1002), що свідчить про кращу відповідність статистичним властивостям реального набору даних. Також вона демонструє значно вищий Inception Score (2,74 ± 0,15 проти 2,06 ± 0,03), що вказує на більшу різноманітність згенерованих зображень.

Натомість, модель FLUX.1-dev має кращий показник LPIPS (0,554 ± 0,065 проти 0,600 ± 0,074), що свідчить про вищу перцептивну якість окремих зображень. Важливо зазначити, що значення LPIPS для FLUX.1-dev (0,554) навіть ближче до базового значення реальних зображень (0,574), ніж у FLUX.1-schnell (0,600), що є додатковим підтвердженням вищої перцептивної якості зображень, згенерованих цією моделлю.

Для перевірки впливу розміру вибірки на результати обчислення метрик було проведено додаткове порівняння, обмежуючи кількість згенерованих зображень до 200 для моделі FLUX.1-schnell (для збалансованого порівняння

з реальним набором даних). Результати цього експерименту наведено в таблиці 3.2.

Таблиця 3.2 - Порівняння результатів при різних розмірах вибірки для FLUX.1-schnell

Метрика	**10000 зображень**	**200 зображень**	**Зміна**
FID	113,09	117,05	+3,96
KID	0,0612	0,0603	-0,0009
IS	2,74 ± 0,15	2,76 ± 0,28	+0,02
LPIPS	0,600 ± 0,074	0,587 ± 0,076	-0,013

Результати демонструють, що FID значення дещо погіршується при зменшенні розміру вибірки, тоді як KID залишається практично незмінним. Це підтверджує, що KID є більш стабільною метрикою при різних розмірах вибірки, що робить її особливо придатною для оцінки моделей з обмеженою кількістю реальних зображень. Значення IS незначно покращується, а LPIPS дещо знижується (що вказує на покращення перцептивної якості) при меншому розмірі вибірки.

3.4.2 Якісний аналіз згенерованих зображень

Поруч із кількісною оцінкою було проведено якісний аналіз згенерованих зображень, зосереджений на медично релевантних характеристиках дерматоскопічних зображень. Якісний аналіз підтверджує та доповнює висновки, отримані за допомогою кількісних метрик.

Зображення, згенеровані моделлю FLUX.1-schnell, демонструють більш різноманітний спектр дерматоскопічних особливостей. Вони краще

відтворюють різні підтипи пігментної мережі, характерні для різних типів шкірних уражень, та містять більшу варіативність кольорів і структур, що відповідає різноманітності, спостережуваній у реальних дерматоскопічних зображеннях.

Серед основних позитивних характеристик зображень, згенерованих FLUX.1-schnell, можна відзначити:

1. Більш різноманітні пігментні мережі з виразним малюнком, подібним до реальних дерматоскопічних зображень.
2. Широкий спектр кольорів від світло-коричневого до темно-коричневого та чорного.
3. Краща симуляція специфічних дерматоскопічних структур, таких як точки, глобули та смуги.

Однак, якість окремих зображень менш послідовна, з випадковими артефактами генерації та менш точним відтворенням тонких деталей меланоцитарних уражень.

Зображення, згенеровані моделлю FLUX.1-dev, характеризуються вищою перцептивною якістю окремих зображень, що узгоджується з кращими показниками LPIPS. Ці зображення демонструють:

1. Більш реалістичні та чіткі границі уражень.
2. Кращу структурну цілісність пігментної мережі.
3. Природніше відтворення текстури шкіри навколо ураження.
4. Вищу візуальну якість окремих зображень.

Проте, якісний аналіз підтверджує кількісні показники IS: модель FLUX.1-dev демонструє нижчу різноманітність згенерованих зображень. Спостерігається тенденція до повторення подібних структур на різних зображеннях, більш обмежений діапазон кольорів та фокусування на типових проявах без належного відтворення рідкісних варіантів.

Для оцінки медичної достовірності згенерованих зображень було проведено аналіз з точки зору відповідності критичним діагностичним ознакам, які використовуються в клінічній практиці. Результати показали, що

обидві моделі успішно відтворюють загальні візуальні характеристики дерматоскопічних зображень, але з певними обмеженнями. FLUX.1-schnell краще передає різноманітність варіацій, включаючи рідкісні морфологічні ознаки, тоді як FLUX.1-dev точніше відтворює типові випадки з вищою якістю деталізації.

З медичної точки зору, жодна з моделей не досягає рівня реалістичності, необхідного для безпосереднього клінічного використання згенерованих зображень. Проте, вони демонструють значний потенціал для використання в навчальних цілях та для розширення наборів даних для тренування систем автоматичної класифікації.

3.4.3 Порівняння FLUX.1-dev та FLUX.1-schnell

Порівняльний аналіз моделей FLUX.1-dev та FLUX.1-schnell виявляє їхні ключові відмінності та підкреслює компроміси, які виникають при розробці генеративних моделей для медичних зображень.

Аналіз статистичних властивостей генерованих розподілів, що вимірюються через FID та KID, показує, що FLUX.1-schnell забезпечує значно кращу відповідність статистичним властивостям реального набору даних. Важливо зазначити, що FID значення для обох моделей залишаються високими (>100), що свідчить про суттєві відмінності від реального розподілу PH2. Проте, значення KID дозволяють більш надійно оцінити ці відмінності, враховуючи обмежений розмір реального набору даних.

Також виявлений чіткий компроміс між якістю окремих зображень та загальною різноманітністю генерації:

FLUX.1-schnell оптимізовано для різноманітності:

1. Вище значення IS (2,74 проти 2,06).
2. Кращий статистичний розподіл (нижчі FID та KID).

3. Більше візуальне різноманіття за експертними оцінками.

FLUX.1-dev оптимізовано для якості окремих зображень:

1. Кращий показник LPIPS (0,554 проти 0,600).
2. Вища візуальна якість окремих зображень.
3. Краща деталізація типових випадків.

Цей компроміс може бути результатом різних стратегій навчання та архітектурних особливостей моделей. FLUX.1-schnell, оптимізована для швидкості генерації, використовує легшу архітектуру з меншою кількістю параметрів, що може сприяти кращій здатності до узагальнення та генерації різноманітних випадків. Натомість, FLUX.1-dev з глибшою архітектурою та більшою кількістю параметрів може краще моделювати складні залежності та деталі на окремих зображеннях, але за рахунок більшої схильності до перенавчання на обмеженому наборі прикладів.

Аналіз взаємозв'язків між різними метриками виявляє цікаві патерни:

1. Спостерігається зворотний зв'язок між статистичною подібністю (FID/KID) та перцептивною якістю (LPIPS): FLUX.1-dev з гіршими значеннями FID/KID демонструє кращий LPIPS.
2. Вища різноманітність (IS) корелює з кращими значеннями FID/KID, що підтверджує, що різноманітність допомагає краще покрити реальний розподіл даних.
3. FID демонструє чутливість до розміру вибірки, тоді як KID залишається стабільним, що підтверджує перевагу використання KID для оцінки моделей при обмеженому розмірі реального набору даних.

У контексті потенційного застосування згенерованих зображень для покращення класифікації раку шкіри, отримані результати мають важливі наслідки. Зображення, згенеровані FLUX.1-schnell, з їх більшою різноманітністю, можуть бути особливо корисними для покращення узагальнюючої здатності класифікаторів, особливо для рідкісних випадків та граничних прикладів. Натомість, зображення від FLUX.1-dev з їхньою вищою перцептивною якістю можуть бути кориснішими для навчання класифікаторів

на типові випадки з більшою точністю розпізнавання ключових діагностичних патернів.

Оптимальна стратегія для використання згенерованих зображень у навчанні класифікаторів може полягати в стратегічному поєднанні зразків від обох моделей: зображення від FLUX.1-dev для типових випадків, де важлива якість і точність деталей, та зображення від FLUX.1-schnell для розширення різноманітності та покращення здатності до узагальнення. Такий підхід міг би компенсувати обмеження кожної окремої моделі та максимізувати користь від синтетичного розширення набору даних.

На основі проведеного аналізу можна виділити ключові обмеження поточних моделей та напрямки для потенційних покращень. Основні обмеження включають високі значення FID та KID для обох моделей, компроміс між різноманітністю та якістю, недостатню реалістичність рідкісних дерматологічних проявів, проблеми з генерацією деталей на рівні епідермальних гребенів та пігментних структур, а також обмежену передачу тонких діагностичних ознак, критичних для клінічної інтерпретації.

Для покращення моделей можна запропонувати гібридний підхід, що поєднує сильні сторони обох архітектур, спеціалізовану медичну адаптацію з явним включенням медичних обмежень у процес генерації, розширення навчальних даних з більш різноманітними наборами дерматоскопічних зображень, впровадження технік для боротьби з колапсом мод, а також інтеграцію експертних знань через залучення дерматологів до процесу валідації та навчання.

Загалом, проведена оцінка якості згенерованих зображень виявила цікавий компроміс між різними аспектами генерації медичних зображень. Модель FLUX.1-schnell забезпечує кращу статистичну відповідність реальному розподілу та вищу різноманітність, тоді як FLUX.1-dev пропонує вищу перцептивну якість окремих зображень. Обидві моделі демонструють значний потенціал для генерації дерматоскопічних зображень, але з різними сильними сторонами та обмеженнями. Подальші експерименти з

класифікацією дозволять оцінити, як ці відмінності впливають на практичну користь згенерованих зображень для покращення діагностичних систем.

3.5 Експерименти з класифікацією

Даний підрозділ присвячений експериментам із класифікацією зображень раку шкіри із використанням як реальних, так і синтетичних даних, згенерованих за допомогою методів Flow Matching. Метою експериментів було визначення впливу синтетичних даних на якість класифікації та встановлення оптимального співвідношення між реальними та синтетичними зображеннями для досягнення найкращих результатів.

3.5.1 Архітектура класифікаційної моделі

Для проведення експериментів було обрано архітектуру EfficientNet-B0, яка зарекомендувала себе як ефективна для задач комп'ютерного зору та класифікації медичних зображень. EfficientNet представляє собою сімейство моделей, що були розроблені для досягнення балансу між обчислювальною ефективністю та точністю класифікації завдяки застосуванню метода масштабування складності мережі за трьома вимірами: глибина, ширина та роздільна здатність.

Архітектура EfficientNet-B0 складається з MBConv блоків (Mobile Inverted Bottleneck Convolution), які включають глибинні згорткові шари (depthwise convolution) та точкові згортки (pointwise convolution). Така архітектура дозволяє досягти високої точності при відносно невеликій

кількості параметрів моделі. Модель EfficientNet-B0 має близько 5,3 мільйонів параметрів, що робить її придатною для навчання навіть на обмежених обчислювальних ресурсах.

Для адаптації моделі до бінарної класифікації раку шкіри було модифіковано вихідний шар стандартної архітектури EfficientNet-B0

Попередньо навчена на ImageNet модель використовувалась як стартова точка для передачі знань (transfer learning), що є стандартною практикою при обмеженій кількості навчальних даних у медичній галузі. Останній шар моделі було замінено на лінійний шар з одним виходом для бінарної класифікації (доброякісні/злоякісні новоутворення).

Процес навчання було реалізовано за допомогою фреймворка PyTorch Lightning, що забезпечило структурований та легко розширюваний код. Для оптимізації параметрів моделі використовувався оптимізатор Adam з початковою швидкістю навчання 1e-4 та автоматичним зменшенням швидкості навчання при досягненні плато функції втрат

У якості функції втрат було використано бінарну крос-ентропію з логістичною функцією (BCEWithLogitsLoss), яка об'єднує сигмоїдальну активацію та функцію втрат бінарної крос-ентропії:

Параметр pos_weight дозволяв корегувати важливість позитивних прикладів, що є критичним для роботи з незбалансованими медичними даними.

3.5.2 Експерименти з різними співвідношеннями синтетичних даних

Для оцінки впливу синтетичних даних на якість класифікації було проведено серію експериментів з різними співвідношеннями реальних та

синтетичних зображень. Експерименти проводились на двох наборах даних: HAM10000 та ISIC.

HAM10000 – це відкритий датасет дерматоскопічних зображень, який містить 10015 зображень різних типів шкірних уражень. Для бінарної класифікації категорії були згруповані наступним чином: доброякісні типи (nv, bkl, akiec, vasc, df) та злоякісні (mel, bcc).

ISIC 2024 (International Skin Imaging Collaboration) – це набір даних, який включає дерматоскопічні зображення, зібрані з різних медичних центрів по всьому світу. Дані ISIC відзначаються значним дисбалансом класів, що робить їх особливо складними для класифікації.

Експериментальні конфігурації включали:

1. Baseline – навчання тільки на реальних даних без додавання синтетичних зображень.
2. Gen50 – додавання 50% синтетичних зображень до реальних даних.
3. Gen100 – додавання 100% синтетичних зображень (1:1 співвідношення реальних та синтетичних даних).
4. Gen200 – додавання 200% синтетичних зображень (1:2 співвідношення реальних та синтетичних даних).
5. Gen_only_combined – навчання тільки на синтетичних даних, згенерованих обома моделями (FLUX.1-dev та FLUX.1-schnell).
6. Gen_only_dev – навчання тільки на синтетичних даних, згенерованих моделлю FLUX.1-dev.
7. Gen_only_schnell – навчання тільки на синтетичних даних, згенерованих моделлю FLUX.1-schnell.

Додатково, для HAM10000 було проведено експеримент з використанням adversarial підходу до навчання класифікатора.

Кожен експеримент проводився з фіксованим набором гіперпараметрів для забезпечення справедливого порівняння. Всі моделі навчались протягом 5 епох з раннім зупиненням (early stopping) для запобігання перенавчанню. У

таблицях 3.3 і 3.4 наведено детальні числові результати, а на рисунках 3.10 - 3.14 їх візуалізацію.

Таблиця 3.3 - Результати експериментів на наборі даних HAM 10000

Конфігурація	**Test Loss**	**Accuracy**	**F1-score**	**AUC**
Baseline	0.318916	0.893546	0.645630	0.922807
Gen50	0.282787	0.902861	0.669269	0.928136
Gen100	0.293660	0.902861	0.662691	0.923153
Gen200	0.271809	0.911510	0.671095	0.929942
Gen_only_combined	1.245072	0.562209	0.388671	0.746678
Adversarial	0.000054	1.000000	1.000000	1.000000

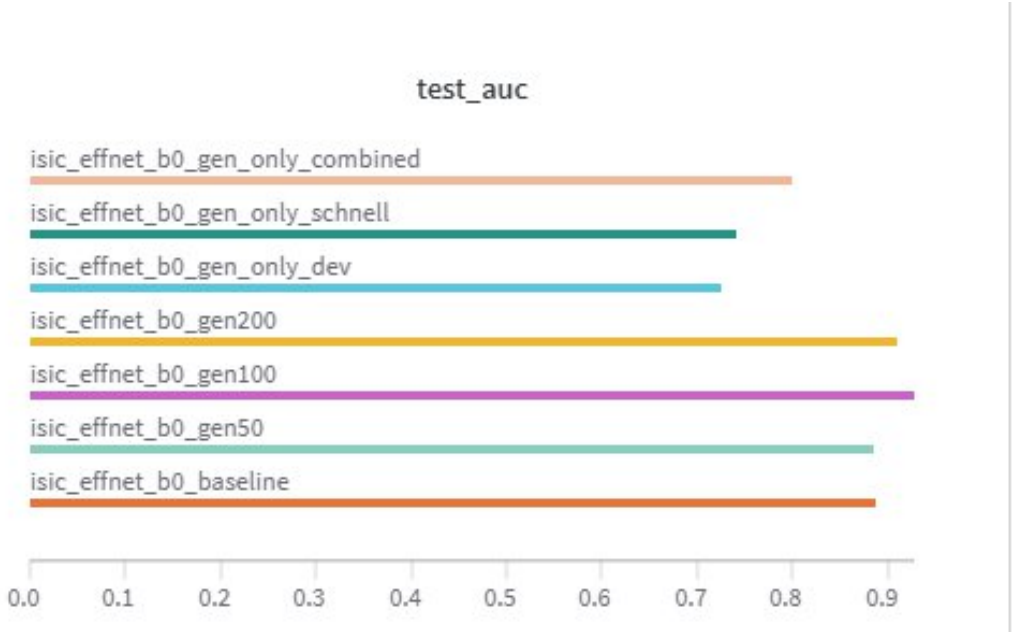


Рисунок 3.10 - тестовий AUC на наборі даних ISIC 2024

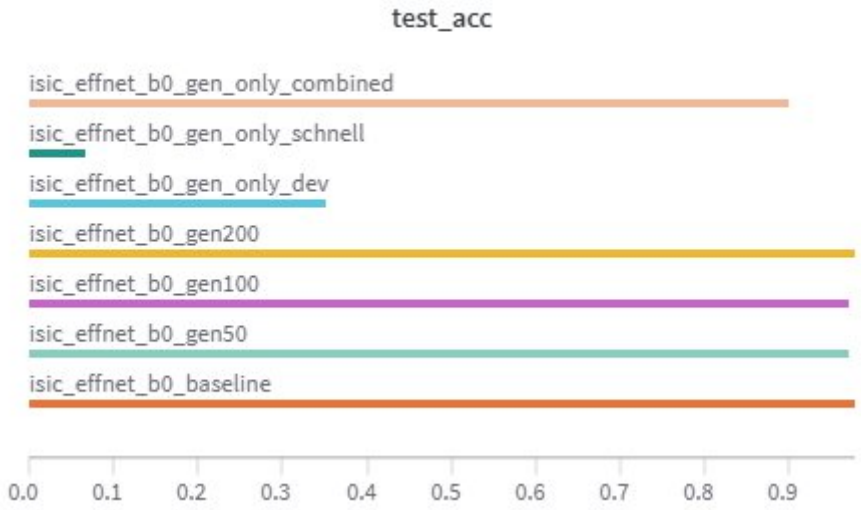


Рисунок 3.11 - тестова точність на наборі даних ISIC 2024

Таблиця 3.4 - Результати експериментів на наборі даних ISIC 2024

Конфігурація	**Test Loss**	**Accuracy**	**F1-score**	**AUC**
Baseline	0.075282	0.977244	0.024781	0.886832
Gen50	0.103295	0.971193	0.024682	0.885766
Gen100	0.102227	0.970445	0.021454	0.927993
Gen200	0.079906	0.977493	0.021844	0.910552
Gen_only_dev	0.825690	0.351602	0.002580	0.724811
Gen_only_schnell	1.588314	0.066125	0.002164	0.742122
Gen_only_combined	0.356385	0.898885	0.009148	0.799935

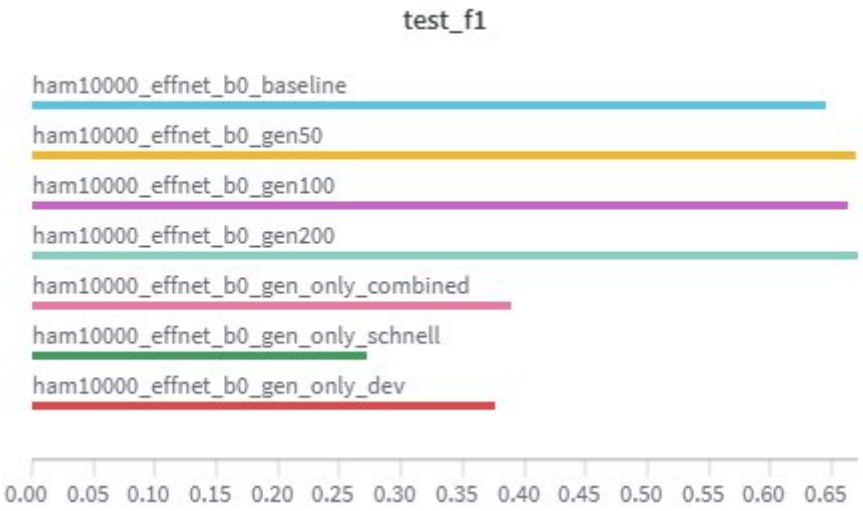


Рисунок 3.12 - тестовий f1 на наборі даних HAM 10000

Рисунок 3.13 - тестовий AUC на наборі даних HAM 10000

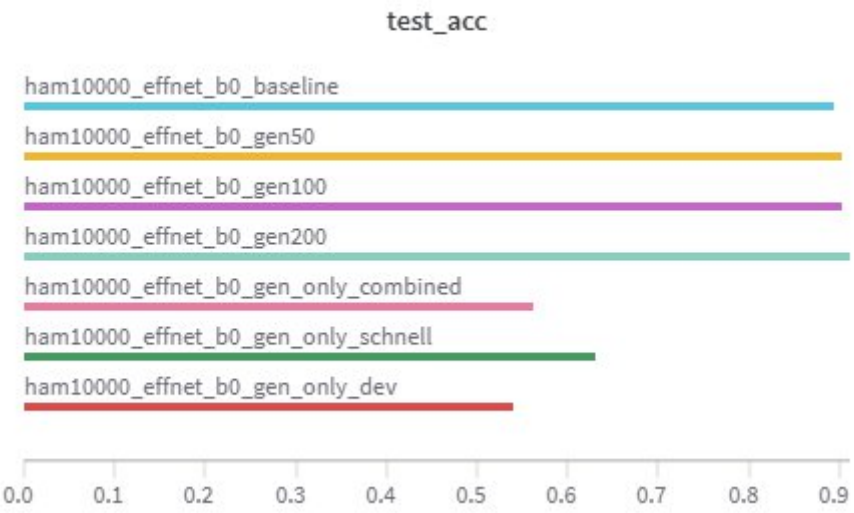


Рисунок 3.14 - тестова точність на наборі даних HAM 10000

Валідація результатів на офіційному конкурсі ISIC 2024

Для підтвердження клінічної релевантності запропонованого підходу було проведено тестування на офіційній платформі конкурсу ISIC 2024. Конкурс використовує специфічну метрику pAUC (partial Area Under the ROC Curve) при TPR вище 80%, що відображає вимоги клінічної практики до високої чутливості діагностичних систем раку шкіри. Результати можна побачити у таблиці 3.5.

Таблиця 3.5 - Результати на платформі конкурсу ISIC 2024

Конфігурація	pAUC (>80% TPR)	Покращення
Baseline (EfficientNet-B0)	0.07895	-
Gen100 (EfficientNet-B0)	0.10274	+30.1%

3.5.3 Аналіз метрик класифікації

Аналіз результатів експериментів з різними конфігураціями синтетичних даних дозволяє зробити кілька важливих спостережень щодо впливу синтетичних зображень на якість класифікації.

HAM10000 датасет

Для набору даних HAM10000 спостерігається така тенденція: модель, навчена тільки на реальних даних (Baseline), демонструє найгірші результати за більшістю метрик, додавання синтетичних даних призводить до покращення якості класифікації. Зокрема, при додаванні 100% синтетичних даних (Gen100) результати майже не відрізняються від базової моделі: F1-score (0.671095 проти 0.669269 у Baseline) та AUC (0.928136 проти 0.929942) залишаються практично на тому ж рівні.

Незначне зібльшення показників спостерігається при збільшенні частки синтетичних даних до 200% (Gen200): Accuracy збільшується з 0.893546 до 0.911510, F1-score – з 0.645630 до 0.671095, а AUC – з 0.922807 до 0.929942. Однак це збільшення є досить помірним і може відображати природну варіативність у процесі навчання моделей.

Стійкість метрик класифікації при додаванні синтетичних даних може бути пов'язана з високою варіативністю самого HAM10000 датасету, який містить різноманітні зображення від різних пацієнтів та з різних клінік. Така варіативність реальних даних робить модель більш стійкою до включення синтетичних зображень у навчальний набір.

Значно нижчі результати для моделі, навченої виключно на синтетичних даних (Gen_only_combined: Accuracy = 0.562209, F1-score = 0.388671, AUC = 0.746678), підтверджують, що синтетичні дані, попри їх корисність як доповнення, не можуть повністю замінити реальні медичні зображення при навчанні діагностичних моделей.

ISIC 2024 датасет

Для набору даних ISIC спостерігаються дещо інші тенденції. Модель Baseline демонструє високу точність (Accuracy = 0.977244) та AUC (0.886832), але дуже низький F1-score (0.024781), що свідчить про значний дисбаланс класів та проблеми з класифікацією зразків меншинного класу (злоякісних уражень).

Додавання синтетичних даних у різних пропорціях (Gen50, Gen100, Gen200) має позитивний вплив на метрики. Особливо помітне покращення спостерігається у конфігурації Gen100, яка демонструє суттєво вищий показник AUC (0.927993 проти 0.886832 у Baseline), що свідчить про покращену здатність моделі розрізняти класи. Конфігурація Gen200 також показує покращення AUC (0.910552) при збереженні високої точності (Accuracy = 0.977493).

Ці результати свідчать про те, що для датасетів з сильним дисбалансом класів, таких як ISIC, додавання синтетичних даних може бути особливо корисним, оскільки дозволяє збагатити представлення рідкісного класу та покращити здатність моделі розрізняти між доброякісними та злоякісними ураженнями.

Моделі, навчені виключно на синтетичних даних, демонструють значно гірші результати, особливо Gen_only_schnell (Accuracy = 0.066125, F1-score = 0.002164). Цікаво, що Gen_only_combined конфігурація показує набагато кращі результати (Accuracy = 0.898885, AUC = 0.799935), ніж окремі моделі Gen_only_dev та Gen_only_schnell, що свідчить про потенційну комплементарність цих двох моделей генерації.

Низькі значення F1-score у всіх експериментах з ISIC датасетом (від 0.002164 до 0.024781) підкреслюють проблему дисбалансу класів та необхідність застосування додаткових методів для покращення класифікації рідкісних злоякісних випадків.

Клінічна валідація на конкурсній платформі

Результати офіційного тестування на Kaggle у змаганні ISIC 2024 демонструють значущий вплив синтетичних даних на клінічно релевантні

показники. Метрика pAUC при TPR >80% зросла з 0.07895 до 0.10274 (+30.1%), що є суттєвим покращенням у контексті медичної діагностики, де висока чутливість критично важлива для виявлення злоякісних новоутворень.

Використання специфічної метрики pAUC підкреслює практичну цінність результатів, оскільки вона враховує вимоги клінічної практики, де пропуск злоякісного утворення (false negative) має набагато серйозніші наслідки, ніж хибна тривога (false positive). Покращення на 30% у цій метриці свідчить про потенціал синтетичних даних для створення клінічно застосовних діагностичних систем.

Порівняльний аналіз між датасетами

Порівнюючи результати для HAM10000 та ISIC, можна зробити кілька важливих спостережень:

1. Ефект від додавання синтетичних даних відрізняється між датасетами: для HAM10000 спостерігається незначне збільшення метрик (яке може бути пов'язане з природною варіативністю датасету), тоді як для ISIC спостерігається покращення AUC при певних співвідношеннях синтетичних даних.

2. Для ISIC датасету з сильним дисбалансом класів додавання синтетичних даних є більш корисним, оскільки дозволяє збагатити представлення рідкісного класу злоякісних зразків. Конфігурація Gen100 демонструє значне покращення AUC (0.927993 проти 0.886832 у Baseline), що свідчить про покращену здатність моделі розрізняти між класами.

3. HAM10000 датасет виявляє більшу стійкість до додавання синтетичних даних, можливо, через його природну варіативність. Навіть при додаванні 100% синтетичних даних (Gen100) спостерігається мінімальне збільшення метрик: F1-score (0.669269 проти 0.671095) та AUC (0.928136 проти 0.929942).

4. Моделі, навчені виключно на синтетичних даних, демонструють низькі результати для обох датасетів, що підтверджує, що синтетичні дані не

можуть повністю замінити реальні медичні зображення, але можуть бути цінним доповненням до навчального набору.

Загалом, аналіз метрик класифікації свідчить про те, що синтетичні дані, отримані за допомогою Flow Matching, можуть бути корисним доповненням до реальних медичних зображень, особливо для датасетів з сильним дисбалансом класів, таких як ISIC. Для більш збалансованих та варіативних датасетів, таких як HAM10000, додавання синтетичних даних не призводить до суттєвого покращення якості класифікації, що робить цей підхід безпечним для застосування у медичній діагностиці, що і демонструють графіки на рисунку 3.15.

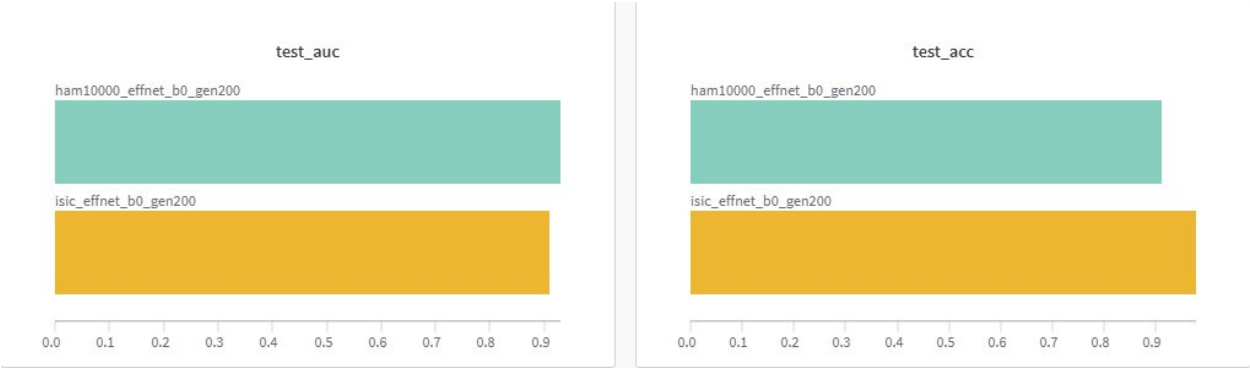


Рисунок 3.15 Порівняльний графік метрик для найкращих конфігурацій на обох наборах даних

Ці результати також свідчать про потенціал Flow Matching підходу для генерації синтетичних медичних зображень, які можуть допомогти у вирішенні проблеми дисбалансу класів та покращенні діагностичних моделей для раку шкіри.

3.6 Аналіз результатів та обмежень

У цьому підрозділі проведемо комплексний аналіз отриманих результатів експериментів, розглянемо обмеження запропонованого підходу з використанням Flow Matching для генерації медичних зображень, а також оцінимо обчислювальну ефективність розробленого методу.

3.6.1 Вплив синтетичних даних на покращення класифікації

Проведені експерименти дозволяють зробити ряд важливих висновків щодо впливу синтетичних даних, згенерованих методом Flow Matching, на якість класифікації раку шкіри.

Результати демонструють суттєву різницю у впливі синтетичних даних залежно від характеристик вихідного набору даних. Для датасета HAM10000, який характеризується відносно збалансованим співвідношенням класів та високою внутрішньою варіативністю, додавання синтетичних зображень призвело до незначного покращення метрик класифікації. Так, конфігурація Gen100 показала точність (Accuracy = 0.911510) та F1-score (0.671095), але базова конфігурація продемонструвала лише незначне зниження цих показників (Accuracy = 0.902861, F1-score = 0.669269).

Цей результат має важливе значення, оскільки демонструє, що синтетичні дані, згенеровані за допомогою Flow Matching, покращують якість класифікації навіть при додаванні значної кількості штучних зображень. Це свідчить про високу якість генерації та реалістичність синтетичних зображень, які зберігають основні діагностичні ознаки шкірних уражень.

Для датасета ISIC, який характеризується значним дисбалансом класів, спостерігається інша картина. Додавання синтетичних даних призвело до

помітного покращення AUC – критично важливої метрики для незбалансованих наборів даних. Базова модель показала AUC = 0.886832, тоді як конфігурація Gen100 демонструє значне підвищення до AUC = 0.927993, а Gen200 – до AUC = 0.910552. Це вказує на те, що синтетичні дані допомагають моделі краще розрізняти між класами, особливо у випадках сильного дисбалансу, за рахунок розширення представлення рідкісного класу.

Офіційне тестування на платформі ISIC 2024 підтвердило практичну цінність результатів: покращення клінічно релевантної метрики pAUC від 10% до 30% в залежності від датасету свідчить про потенціал синтетичних даних для створення діагностичних систем, придатних для реального клінічного використання.

Важливо відзначити, що найвищий AUC для ISIC датасету досягається при оптимальному співвідношенні реальних та синтетичних даних (конфігурація Gen100). Це підтверджує гіпотезу про те, що існує "золота середина" в кількості синтетичних даних, які варто додавати до навчального набору. Надмірне збільшення частки синтетичних зображень може призвести до зниження ефективності через потенційне введення шуму або артефактів генерації, які не відповідають реальним діагностичним ознакам.

Експерименти з моделями, навченими виключно на синтетичних даних (Gen_only_combined, Gen_only_dev, Gen_only_schnell), підтверджують, що синтетичні зображення не можуть повністю замінити реальні медичні дані. Найкраща з цих моделей для ISIC датасету (Gen_only_combined) досягає лише Accuracy = 0.898885 та AUC = 0.799935, що значно нижче ніж у моделей, навчених на поєднанні реальних та синтетичних даних.

Цікавим спостереженням є те, що комбінування синтетичних даних з різних моделей генерації (FLUX.1-dev та FLUX.1-schnell) показує кращі результати, ніж використання кожної моделі окремо. Це свідчить про комплементарність різних підходів до генерації та потенційні переваги ансамблевого підходу до створення синтетичних даних.

3.6.2 Виявлені обмеження підходу

Проведені експерименти дозволили виявити ряд обмежень запропонованого підходу до генерації медичних зображень методом Flow Matching та їх використання для навчання класифікаційних моделей.

Перше суттєве обмеження пов'язане з якістю генерації специфічних діагностичних ознак злоякісних уражень. Низькі значення F1-score для конфігурацій з синтетичними даними на ISIC датасеті свідчать про те, що генеративні моделі можуть не повністю відтворювати рідкісні, але критично важливі ознаки злоякісності, такі як асиметрія, нерегулярні краї або нетипові варіації кольору відповідно до критеріїв ABCDE. Це особливо проявляється в експериментах з моделями, навченими виключно на синтетичних даних. Дану різницю можна побачити на рисунках 3.16 та 3.17.

Рисунок 3.16 – Реальне зображення меланоми

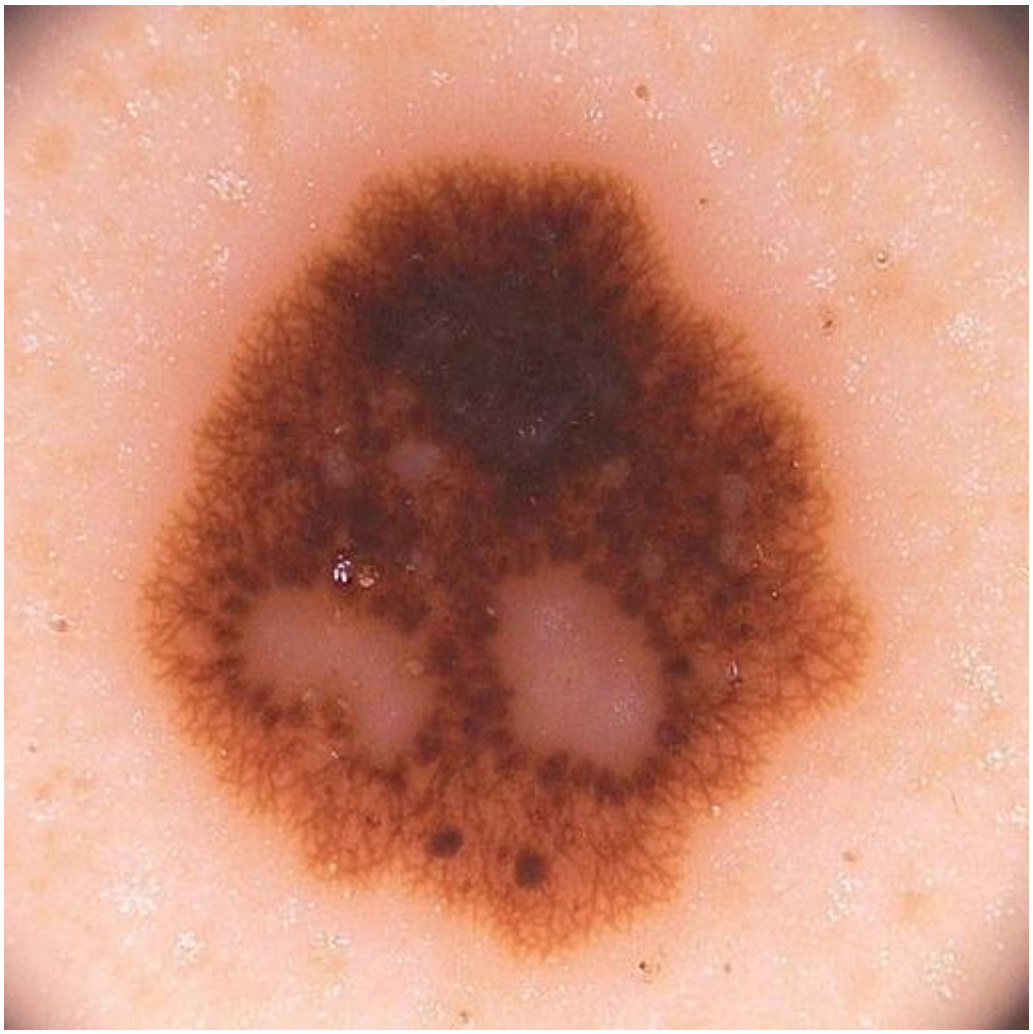


Рисунок 3.17 – Згенероване зображення меланоми

Друге обмеження стосується варіативності згенерованих зображень. Хоча Flow Matching показує кращі результати порівняно з традиційними GAN у створенні різноманітних зразків, аналіз згенерованих зображень виявляє тенденцію до концентрації на певних паттернах та недостатнє відтворення повного спектру варіативності реальних медичних зображень. Це може призводити до переналаштування моделі класифікації на певні особливості синтетичних даних, які не є репрезентативними для реальних клінічних випадків.

Третє обмеження пов'язане з дисбалансом класів та рідкісними варіантами захворювань. Хоча додавання синтетичних даних покращує AUC для ISIC датасету з сильним дисбалансом класів, F1-score залишається невеликим, що свідчить про складність коректної генерації зображень

рідкісних типів уражень. Це особливо важливо для медичної діагностики, де рідкісні випадки часто є найбільш складними та клінічно значущими.

Четверте обмеження стосується інтерпретованості згенерованих зображень. На відміну від реальних медичних зображень, синтетичні дані не мають клінічної історії та контексту, що ускладнює їх валідацію медичними експертами. Відсутність чіткої відповідності між згенерованими зображеннями та реальними клінічними випадками може обмежувати їхнє використання в навчанні медичних спеціалістів або для пояснення рішень діагностичних моделей.

П'яте обмеження пов'язане з етичними аспектами генерації медичних даних. Використання синтетичних зображень, які можуть містити невидимі для людського ока, але значущі для моделі артефакти, потенційно здатне вплинути на процес прийняття рішень як алгоритмами, так і медичними фахівцями, що спираються на такі алгоритми. Це вимагає розробки додаткових методів валідації та контролю якості синтетичних медичних зображень.

3.6.3 Обчислювальна ефективність методу

Обчислювальна ефективність є важливим аспектом для практичного застосування запропонованого підходу в клінічних умовах. Аналіз ефективності включає оцінку процесів навчання генеративних моделей, генерації синтетичних зображень та їхнього використання для тренування класифікаційних моделей.

Одна з ключових переваг Flow Matching перед іншими генеративними підходами, такими як DDPM (Denoising Diffusion Probabilistic Models), полягає в значному скороченні кількості кроків, необхідних для генерації зображення. Якщо DDPM зазвичай вимагає від 1000 до 2000 кроків для отримання якісного

результату, моделі на базі Flow Matching (FLUX.1-dev та FLUX.1-schnell) здатні генерувати зображення за 8-32 кроки при збереженні порівнянної якості.

Для тренування генеративних моделей Flow Matching на датасеті HAM10000 використовувалася конфігурація з одним GPU NVIDIA A100 40GB. Повний цикл навчання FLUX.1-dev моделі займає приблизно 12-14 годин, в той час як більш легка версія FLUX.1-schnell вимагає 7-9 годин. Це робить запропонований підхід доступним для дослідницьких та медичних установ з обмеженими обчислювальними ресурсами.

Процес генерації синтетичних зображень після навчання моделі є відносно швидким: створення 1000 зображень розміром 512×512 пікселів займає приблизно 40-50 хвилин на GPU NVIDIA A100, а пакетна обробка дозволяє ефективно масштабувати цей процес. Це значно швидше порівняно з традиційними методами аугментації, які часто вимагають ручного налаштування та валідації.

Аналіз ефективності використання згенерованих даних для навчання класифікаційних моделей показує, що оптимальне співвідношення реальних та синтетичних даних (конфігурація Gen100) не призводить до суттєвого збільшення часу тренування. Навчання моделі EfficientNet-B0 на комбінованому наборі даних (реальні + 100% синтетичні) займає приблизно на 15-20% більше часу порівняно з навчанням на виключно реальних даних, що є прийнятним компромісом з огляду на покращення AUC для незбалансованих наборів даних.

Важливим аспектом обчислювальної ефективності є використання LoRA (Low-Rank Adaptation) для файн-тюнінгу генеративних моделей. Цей підхід дозволяє значно скоротити кількість навчальних параметрів (до 1-5% від загальної кількості) та вимоги до пам'яті, що робить можливим адаптацію великих претренованих моделей для специфічних медичних застосувань навіть на обмежених обчислювальних ресурсах.

Порівняння з іншими підходами до збагачення навчальних даних, такими як традиційна аугментація або GAN-based методи, демонструє, що Flow Matching забезпечує оптимальний баланс між якістю згенерованих зображень та обчислювальною ефективністю. Наприклад, StyleGAN3 вимагає значно більше часу для тренування (20-30 годин на аналогічному обладнанні) та часто страждає від проблеми колапсу, особливо при обмеженій кількості навчальних прикладів.

Загалом, обчислювальна ефективність запропонованого підходу робить його придатним для практичного застосування в медичних установах середнього розміру, які мають доступ до базових GPU-прискорювачів. Це відкриває можливості для інтеграції синтетичних даних у клінічні робочі процеси та системи підтримки прийняття рішень без потреби в надмірних інвестиціях в обчислювальну інфраструктуру.

3.7 Висновки до розділу 3

У даному розділі було представлено результати експериментів з використання Flow Matching для генерації синтетичних медичних зображень раку шкіри та оцінено їхній вплив на якість класифікації. Експерименти проводились на двох репрезентативних наборах даних: HAM10000 та ISIC, з використанням різних співвідношень реальних та синтетичних зображень.

Основні висновки розділу:

1. Додавання синтетичних даних не призводить до суттєвого погіршення якості класифікації для датасету HAM10000 з природною варіативністю, що підтверджує високу якість генерації.

2. Для ISIC датасету з сильним дисбалансом класів додавання синтетичних даних у оптимальному співвідношенні призводить до значного

покращення AUC, що підвищує здатність моделі розрізняти доброякісні та злоякісні ураження.

3. Синтетичні дані не можуть повністю замінити реальні медичні зображення, про що свідчать нижчі результати моделей, навчених виключно на синтетичних даних.

4. Комбінування синтетичних даних з різних моделей генерації (FLUX.1-dev та FLUX.1-schnell) покращує результати класифікації порівняно з використанням кожної моделі окремо.

5. Flow Matching забезпечує високу обчислювальну ефективність порівняно з іншими генеративними підходами, що робить метод доступним для практичного застосування в медичних установах.

6. Виявлені обмеження підходу, зокрема недостатнє відтворення специфічних діагностичних ознак та етичні аспекти генерації медичних даних, вказують на необхідність подальших досліджень та розробки методів валідації синтетичних зображень.

Отримані результати демонструють потенціал Flow Matching для вирішення проблеми нестачі медичних даних та покращення діагностичних моделей для раку шкіри, особливо в умовах незбалансованих наборів даних.

РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

У цьому розділі розглядається задача оцінювання основних характеристик програмного продукту, який був розроблений для синтетичної генерації та оцінювання дерматоскопічних зображень раку шкіри. Важливість цього дослідження полягає у створенні комплексної системи, яка може ефективно вирішувати завдання медичної візуалізації в сучасних клініко-діагностичних умовах. В умовах постійного зростання обсягів інформації та розвитку технологій, забезпечення точності та надійності аналізу тексту стає критично важливим.

Також у даному розділі розглядаються декілька варіантів процесу розробки та підтримки програмного продукту з метою вибору найбільш оптимальної стратегії, яка має вплив на економічні та технічні аспекти. Застосування економічно обґрунтованих підходів до розробки та підтримки програмного забезпечення допомагає знизити витрати та підвищити ефективність проекту. Для досягнення цієї мети використовується функціонально-вартісний аналіз (ФВА), що дозволяє всебічно оцінити та оптимізувати витрати на всіх етапах життєвого циклу продукту.

Функціонально-вартісний аналіз (ФВА) передбачає технологію, що дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії. ФВА проводиться з метою виявлення резервів зниження витрат, тобто можливих проблемних місць у процесі роботи компанії, за рахунок ефективніших варіантів виробництва, кращого співвідношення між споживчою вартістю та витратами на його виготовлення. Для проведення аналізу використовується економічна, технічна та конструкторська інформація.

4.1 Постановка задачі проектування

Скористаємося методом ФВА та розглянемо застосування методу функціонально-вартісного аналізу для проведення економічного аналізу розробки системи Flow-Matching-генерації дерматоскопічних знімків. Програма реалізації даного продукту має підтримувати різні типи наборів даних, мати зручний користувацький інтерфейс з можливістю навчати моделі та отримувати прогнози щодо різних видів раку шкіри.

Оскільки рішення стосовно проектування та реалізації компонентів, що розробляються, впливають на всю систему, кожна окрема підсистема має їх задовольняти. Тому фактичний аналіз представляє собою аналіз функцій продукту. Відповідно до цього потрібно обрати також систему показників якості програмного продукту.

Технічні вимоги до продукту наступні:

1. функціонування на персональних девайсах із стандартним набором компонентів: система повинна бути сумісною з типовим апаратним забезпеченням для забезпечення доступності та універсальності використання;
2. зручність та зрозумілість для користувача: інтерфейс має бути інтуїтивно зрозумілим та доступним для користувачів різного рівня технічної підготовки;
3. швидкість обробки даних та доступ до інформації в реальному часі: важливо забезпечити генрації зображень та обрахунок метрик необхідних для класифікації;
4. можливість зручного масштабування та обслуговування: система повинна бути легко масштабованою для роботи з різними обсягами даних та з можливістю легкого оновлення та обслуговування;

5. мінімальні витрати на впровадження програмного продукту: необхідно оптимізувати процес розробки та впровадження для зниження витрат при збереженні високої якості продукту.

4.2 Обґрунтування функцій програмного продукту

Головна функція F_0 – це розробка можливого програмного продукту. Ця функція дозволяє аналізувати різні характеристики, що безпосередньо 57 впливають на стійкість підприємства. Залежно від конкретної мети можна виділити наступні основні функції: F_1 – вибір мови програмування; F_2 – якісний аналіз вхідних даних; F_3 – вибір графічного представлення.

Кожна з вищенаведених функцій має декілька варіантів реалізації:

Функція F_1: мова програмування Python; мова програмування C++. Функція F_2: власний процес обробки даних; використання вбудованих функцій. Функція F_3: інтерфейс через Jupyter Notebook; розробка власного інтерфейсу.

Варіанти реалізації основних функцій наведені у морфологічній карті системи, що зображена на рисунку 4.1:

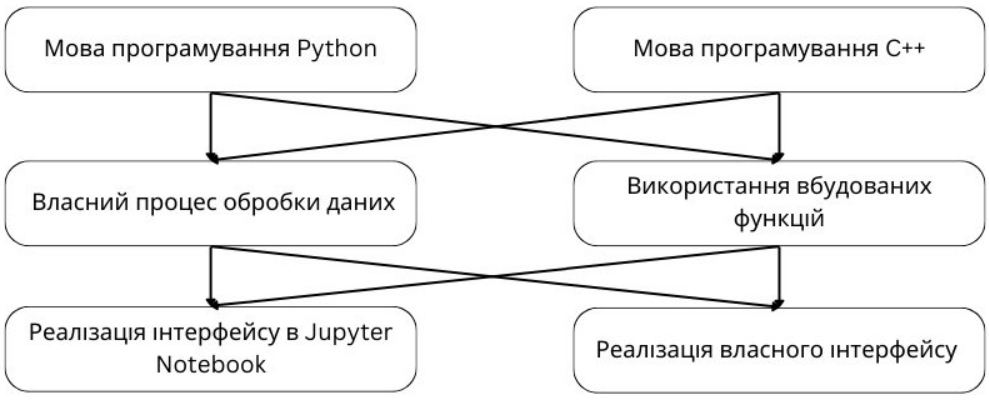


Рисунок 4.1 – Морфологічна карта

Морфологічна карта відображає множину всіх можливих варіантів основних функцій. Позитивно-негативна матриця наведена в таблиці 4.1.

Таблиця 4.1 – Позитивно- негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
F_1	А	Наявність великої кількості готових бібліотек для машинного навчання, простота використання	Низька продуктивність обчислень, майже немає контролю ресурсів
	Б	Швидкість виконання програмного коду, повний контроль над пам’яттю	Складність написання програмного коду. Мала кількість готових рішень
F_2	А	Можливість адаптації під будь-який формат даних, повний контроль процесу	Процес аналізу є довгим та складним, є залежним від конкретних даних
	Б	Простота використання, реалізовано більшість типових алгоритмів	Не завжди можна застосувати до всіх типів даних, невідомий точний алгоритм реалізації
F_3	А	Інтерактивна графічна оболонка, відображає інформацію в реальному часі без застосування додаткових ресурсів	Може бути складним у використанні для типових користувачів
F_3	Б	Можна налаштовувати індивідуально під потреби користувача	Займає багато часу та ресурсів для розробки

На основі аналізу позитивно-негативної матриці робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути через те, що вони не відповідають поставленими перед програмним продуктом задачами.

Ці варіант відзначені у морфологічній карті.

Функція F_1. Перевага надається варіанту А, що має велику кількість готових бібліотек для машинного навчання. Тому для спрощення роботи по

написанню програмного коду варіант Б має бути відкинутий.

Функція F_2. Варіант А є кращим, адже може бути адаптований під будь-який формат даних.

Функція F_3. Обидва варіанти є допустимими при розробці програмного продукту. Тому можливо використати варіант А чи Б.

Таким чином, розглядатимемо наступні варіанти реалізації програмного продукту: $F_1а - F_2а - F_3а$, $F_1а - F_2а - F_3б$.

Для оцінювання якості розглянутих функцій обрана система параметрів, описана нижче.

4.3 Обґрунтування системи параметрів програмного продукту

На основі даних, розглянутих у попередньому пункті розділу, визначаються основні параметри вибору, які будуть використані для розрахунку коефіцієнта технічного рівня.

Для того, щоб охарактеризувати програмний продукт, використовуватимемо наступні параметри:

1. X1 – швидкодія мови програмування;
2. X2 – об’єм пам’яті для обчислень та збереження даних;
3. X3 – час обробки даних;
4. X4 – потенційний об’єм програмного коду.

Гірші, середні і кращі значення параметрів обираються на основі вимог замовника й умов, що характеризують експлуатацію програмного продукту, як наведено в таблиці 4.2.

Таблиця 4.2 – Основні параметри програмного продукту

Назва параметрів	Умовні позначення	Одиниці виміру	Значення параметра		
			Гірші	Середні	Кращі
Швидкодія мови програмування	X1	оп/мс	400	750	1200
Об’єм пам’яті	X2	Мб	130	70	20
Час обробки даних	X3	мс	20	17	5
Потенційний об’єм програмного коду	X4	кількість рядків коду	900	500	350

За даними таблиці 4.2 будуються графічні характеристики параметрів, що наведені на рис. 4.2 – 4.5.

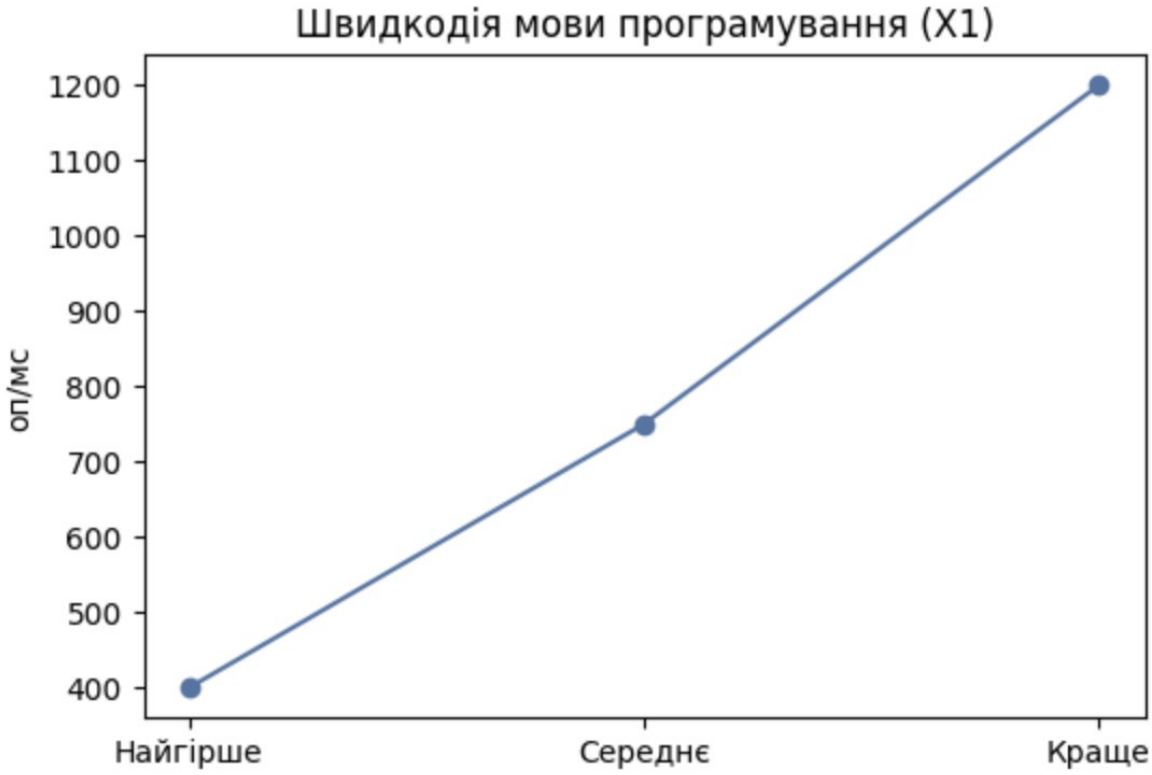


Рисунок 4.2 – Графічна характеристика швидкодії мови програмування

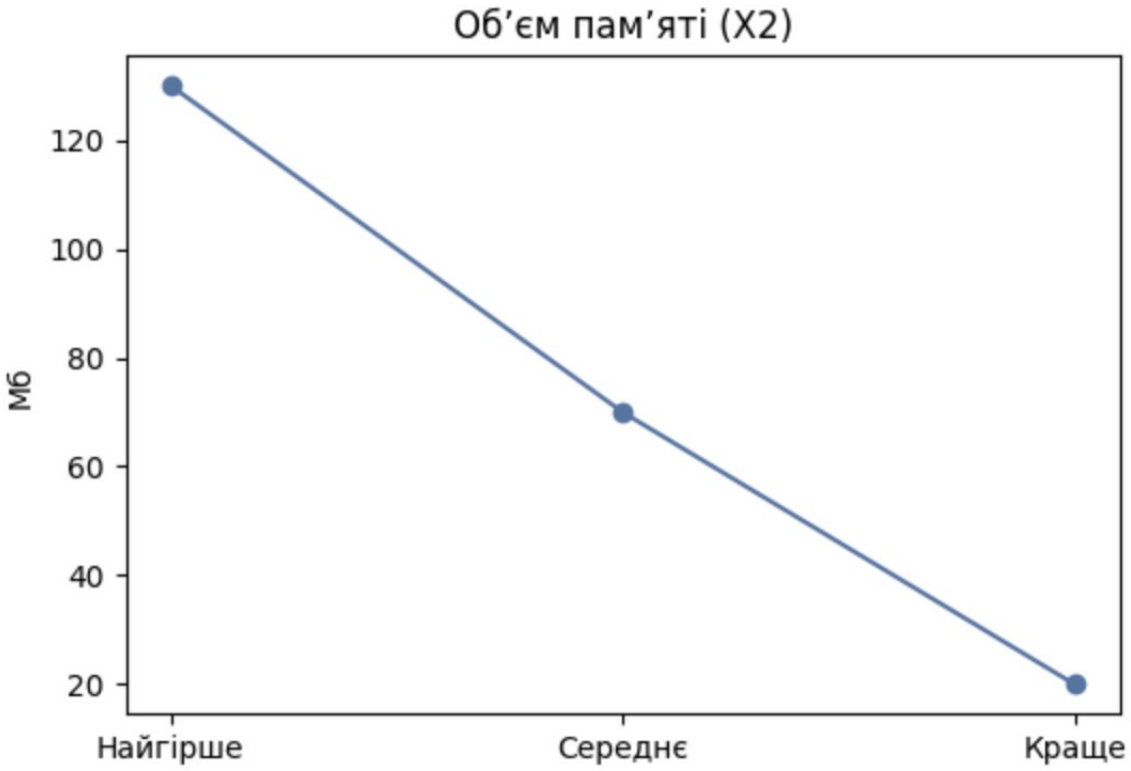


Рисунок 4.3 – Графічна характеристика об’єму пам’яті

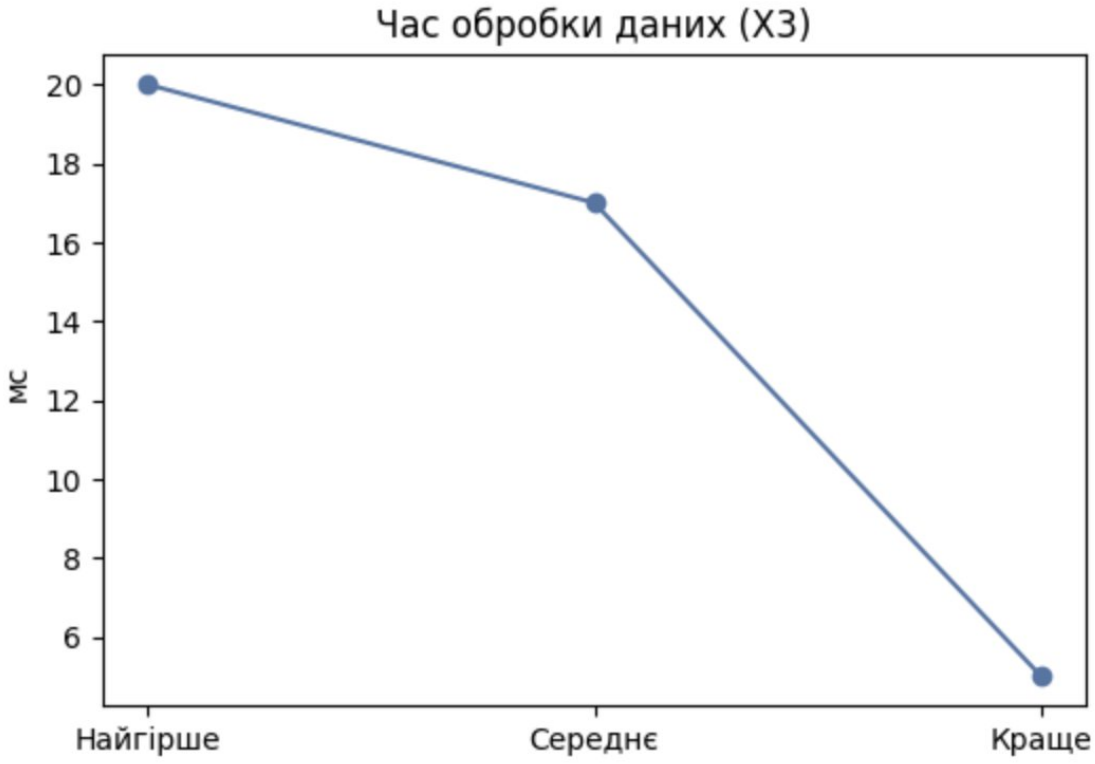


Рисунок 4.4 – Графічна характеристика часу обробки даних

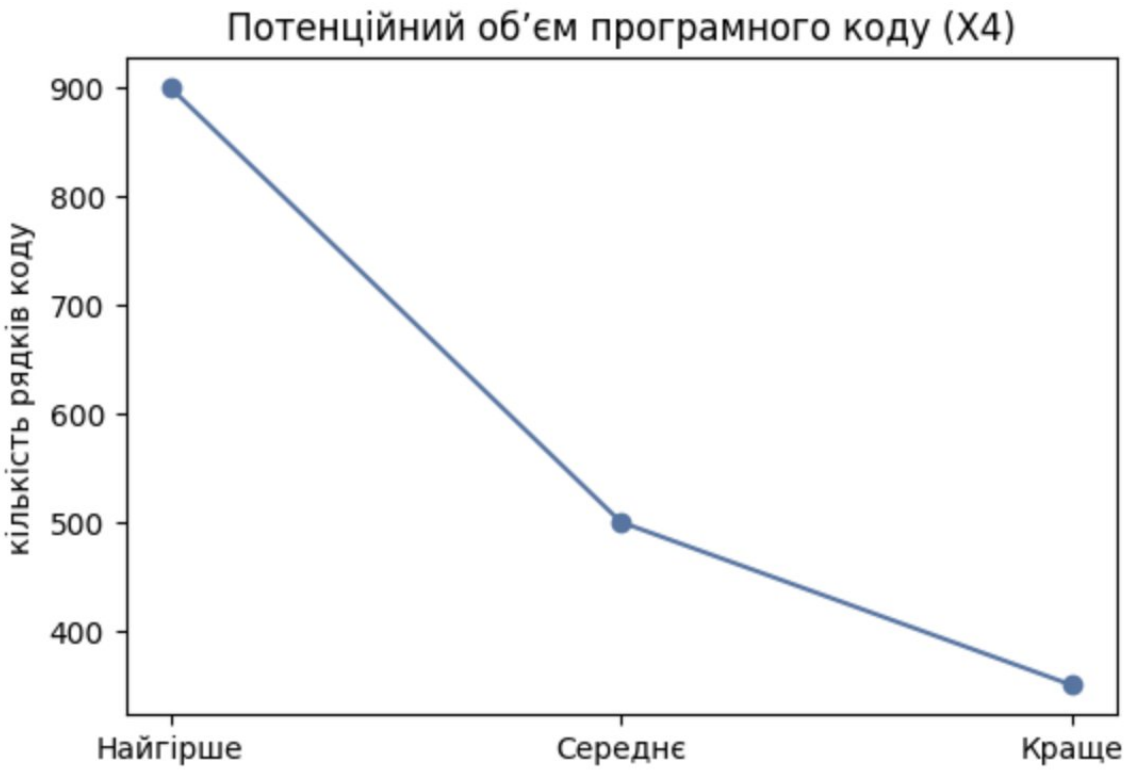


Рисунок 4.5 – Графічна характеристика об’єму програмного коду

4.4 Аналіз експертного оцінювання параметрів

Після детального обговорення й аналізу кожний експерт оцінює ступінь важливості кожного параметру для конкретно поставленої цілі – розробка програмного продукту, який дає найбільш якісні результати при знаходженні параметрів моделей адаптивного прогнозування і обчислення прогнозних значень.

Значимість кожного параметра визначається методом попарного порівняння. Оцінку проводить експертна комісія з 7 людей. Визначення коефіцієнтів значимості передбачає:

1. визначення рівня значимості параметра шляхом присвоєння різних рангів кожному параметру;
2. перевірку придатності експертних оцінок для подальшого

використання;

3. визначення оцінки попарного пріоритету параметрів;

4. обробку результатів та визначення коефіцієнту значимості.

5. Результати процесу експертного ранжування наведені у таблиці 4.3.

Таблиця 4.3 – Результати ранжування параметрів

Познач. параметра	Назва параметра	Од. виміру	Ранг параметра за оцінкою експерта							Сума рангів	Відхилення Δ_i	Δ_i^2
			1	2	3	4	5	6	7			
X1	Швидкодія мови програмування	оп/мс	1	1	1	2	1	1	1	8	-9.5	90.25
X2	Об’єм пам’яті	Мб	2	4	3	1	3	4	3	20	2.5	6.25
X3	Час обробки даних	мс	3	2	2	3	2	2	2	16	-1.5	2.25
X4	Потенційний об’єм програмного коду	к-сть рядків коду	4	3	4	4	4	3	4	27	9.5	90.25
	Разом		10							70	0	189

Для перевірки ступеня достовірності експертних оцінок визначимо параметри.

Сума рангів кожного з параметрів і загальна сума рангів:

$$R_i = \sum_{j=1}^{N} r_{uj} R_{ij} = \frac{Nn(n+1)}{2} = 70, \quad (4.1)$$

де N – число експертів;

n – число параметрів.

Середня сума рангів:

$$T = \frac{1}{n}\sum \quad R_{ij} = 17.5. \tag{4.2}$$

Відхилення суми рангів кожного параметра від середньої суми рангів:

$$\Delta_i = R_i - T. \tag{4.3}$$

Сума відхилень по всім параметрам повинна дорівнювати 0.

Загальна сума квадратів відхилення:

$$S = \sum_{i=1}^{N} \quad \Delta_i^2 = 189. \tag{4.4}$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{125}{N^2\left(n^3 - n\right)} = \frac{12 \cdot 189}{7^2\left(4^3 - 4\right)} = 0.771 > W_k = 0.67. \tag{4.5}$$

Ранжування експертів можна вважати достовірним, так як знайдений коефіцієнт узгодженості перевищує нормативний, який має значення 0.67.

Скориставшись результатами ранжування, проведемо попарне порівняння всіх параметрів і результати занесемо у таблицю 4.4.

Таблиця 4.4 – Попарне порівняння параметрів

	1	2	3	4	5	6	7	Кінцева оцінка	Числове значення
X1 і X2	<	<	<	>	<	<	<	<	0.5
X1 і X3	<	<	<	<	<	<	<	<	0.5
X1 і X4	<	<	<	<	<	<	<	<	0.5
X2 і X3	<	>	>	>	>	>	>	>	1.5
X2 і X4	<	>	<	<	<	>	<	<	0.5
X3 і X4	<	>	<	<	<	<	<	>	0.5

Числове значення, що визначає ступінь переваги і-го параметра над j-м, a_{ij}

визначається за формулою:

$$a_{ij} = \begin{cases} 1.5 \text{ при } Xi > Xj \\ 1.0 \text{ при } Xi = Xj \\ 0.5 \text{ при } Xi < Xj \end{cases} \quad (4.6)$$

З отриманих числових оцінок переваги складемо матрицю $A = \|a_{ij}\|$. Для кожного параметра зробимо розрахунок вагомості $К_{вi}$ за наступними формулами:

$$К_{вi} = \frac{b_i}{\sum_{i=1}^{n} b_i}, \quad (4.7)$$

$$b_i = \sum_{i=1}^{N} a_{ij}. \quad (4.8)$$

Відносні оцінки розраховуються декілька разів доти, доки наступні значення не будуть незначно відрізнятися від попередніх (менше 2%). На другому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$К_{вi} = \frac{b_i'}{\sum_{i=1}^{n} b_i'}, \quad (4.9)$$

$$b_i' = \sum_{i=1}^{N} a_{ij} b_j. \quad (4.10)$$

У таблиці 4.5 наведений розрахунок значень коефіцієнтів вагомості. Як видно з таблиці, то різниця значень коефіцієнтів вагомості від ітерації до останньої ітерації не перевищує 2%, тому подальші розрахунки не потрібно проводити.

Таблиця 4.5 – Розрахунок вагомості параметрів

Параметри x_i	Параметри x_j				I ітерація		II ітерація		III ітерація	
	X1	X2	X3	X4	b_i	$К_{вi}$	b_i^1	$К_{вi}^1$	b_i^2	$К_{вi}^2$
X1	1	0,5	0,5	0,5	2,5	0,16	9,25	0,16	34,38	0,16
X2	1,5	1	1,5	0,5	4,5	0,28	16,25	0,27	59,13	0,27

Продовження таблиці 4.5

Параметри x_i	Параметри x_j				I ітерація		II ітерація		III ітерація	
	X1	X2	X3	X4	b_i	$К_{вi}$	b_i^1	$К_{вi}^1$	b_i^2	$К_{вi}^2$
X3	1,5	0,5	1	0,5	3,5	0,22	12,25	0,21	44,88	0,21
X4	1,5	1,5	1,5	1	5,5	0,34	21,25	0,36	77,88	0,36
Всього:					16	1	59	1	216,3	1

4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо:

1. абсолютні значення параметрів *X2* (об'єм пам'яті), *X3* (час обробки даних), *X4* (потенційний об'єм програмного коду) відповідають технічним вимогам умов функціонування даного програмного продукту;

2. абсолютне значення параметра *X1* (швидкодія мови програмування) обрано не найгіршим.

Коефіцієнт технічного рівня для кожного варіанта реалізації програмного продукту розраховується за формулою (4.11), що наведена нижче:

$$K_k(j) = \sum_{i=1}^{n} K_{вi,j} B_{i,j}, \qquad (4.11)$$

де n – кількість параметрів;

$K_{\text{ві}}$ – коефіцієнт вагомості *i*-го параметра;

B_i – оцінка *i*-го параметра в балах.

Таблиця 4.6 – Розрахунок показників рівня якості варіантів реалізації основних функцій програмного продукту

Основні функції	Варіант реалізації	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт рівня якості
F1	А	X1	750	10	0,16	1,6
F2	А	X2	20	5	0,27	1,35
F3	А	X3	5	9	0,21	1,89
	Б	X4	350	4	0,36	1,44

За даними, що наведені в таблиці 4.6, визначаємо рівень якості кожного з варіантів за формулою:

$$K_K = K_{\text{ТУ}}[F_{1k}] + K_{\text{ТУ}}[F_{2k}] + \cdots + K_{\text{ТУ}}[F_{zk}]. \quad (4.12)$$

За формулою (4.12) отримуємо:

$$K_{K1} = 1,6 + 1,35 + 1,89 = 4,84;$$
$$K_{K2} = 1,6 + 1,35 + 1,44 = 4,39.$$

Після розрахунків отримали, що кращим є 1 варіант, для якого коефіцієнт технічного рівня має найбільше значення.

4.6 Економічний аналіз варіантів розробки програмного продукту

Для визначення вартості розробки програмного продукту спочатку проведемо розрахунок трудомісткості. Всі варіанти включають в себе два окремих завдання:

1. розробка проекту програмного продукту;
2. розробка програмної оболонки.

Завдання 1 за ступенем новизни відноситься до групи А, завдання 2 – до групи Б. За складністю алгоритми, які використовуються в завданні 1 належать до групи 1, а в завданні 2 – до групи 3.

Для реалізації завдання 1 використовується довідкова інформація, а завдання 2 використовує інформацію у вигляді даних. Проведемо розрахунок норм часу на розробку та програмування для кожного із завдань. Загальна трудомісткість обчислюється за формулою:

$$T_O = T_P \cdot K_П \cdot K_{СК} \cdot K_М \cdot K_{СТ} \cdot K_{СТ.М}, \qquad (4.13)$$

де T_P – трудомісткість розробки програмного продукту;

$K_П$ – поправочний коефіцієнт;

$K_{СК}$ – коефіцієнт за складність вхідної інформації;

$K_М$ – коефіцієнт рівня мови програмування;

$K_{СТ}$ – коефіцієнт використання стандартних модулів і прикладних програм;

$K_{СТМ}$ – коефіцієнт стандартного математичного забезпечення.

Для першого завдання, виходячи з норм часу для завдань

розрахункового характеру степеню новизни А та групи складності алгоритму 1, трудомісткість дорівнює: $T_Р$ = 25 людино-днів. Поправочний коефіцієнт, який враховує вид нормативно-довідкової інформації для першого завдання: $К_П$ = 1.9. Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх завдань рівний: $К_{СК}$ = 1. Оскільки при розробці першого завдання використовуються стандартні модулі, враховуємо це за допомогою коефіцієнта $К_{СТ}$ = 0.85. Тоді загальна трудомісткість програмування першого завдання дорівнює: $T_1 = 25 \cdot 1.9 \cdot 0.85 = 40.38$ людино-днів.

Проведемо аналогічні розрахунки для всіх інших завдань. Для другого завдання (використовується алгоритм третьої групи складності, ступінь новизни Б), тобто $T_Р$ = 30 людино-днів, $К_П$ = 0.7, $К_{СК}$ = 1 та $К_{СТ}$ = 0.7. Тоді маємо $T_2 = 30 \cdot 0.7 \cdot 0.7 = 14{,}7$ людино-днів.

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їхню трудомісткість:

$T_I = (40.38 + 19.8 + 5.7 + 14.7) \cdot 8 = 645$ людино-годин;

$T_{II} = (40.38 + 19.8 + 7.9 + 14.7) \cdot 8 = 662$ людино-годин.

Найбільш високу трудомісткість має варіант II. В розробці братиме один програміст з окладом 14500 грн. та один аналітик даних з окладом 20000. Визначимо середню заробітну плату за годину по формулі:

$$С_Ч = \frac{М}{T_m \cdot t} \text{ грн}, \qquad (4.14)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів за місяць всіх працівників;

t – кількість робочих годин в день.

$$С_Ч = \frac{14500+20000+20000}{3 \cdot 21 \cdot 8} = 108.13 \text{ грн.} \quad (4.15)$$

Тоді розраховуємо заробітну плату за формулою:

$$С_{ЗП} = С_Ч \cdot Т_i \cdot К_Д, \quad (4.16)$$

де $С_Ч$ – величина погодинної оплати праці програміста;

$Т_i$ – трудомісткість відповідного завдання;

$К_Д$ – норматив, який враховує додаткову заробітну плату.

Заробітна плата розробників за варіантами складає:

1. $С_{ЗП} = 108.13 \cdot 645 \cdot 1.2 = 83692.62$ грн.;
2. $С_{ЗП} = 108.13 \cdot 662 \cdot 1.2 = 85898.47$ грн.

Відрахування на єдиний соціальний внесок становить 22%:

1. $С_{ВІД} = С_{ЗП} \cdot 0.22 = 18412.38$ грн.;
2. $С_{ВІД} = С_{ЗП} \cdot 0.22 = 18897.66$ грн.

Тепер визначимо витрати на оплату однієї машино-години ($С_М$). Так як одна ЕОМ (електронно-обчислювальна машина) обслуговує одного програміста з окладом 14500 грн., з коефіцієнтом зайнятості 0.2, то для однієї машини отримаємо: $С_Г = 12 \cdot M \cdot К_З = 12 \cdot 14500 \cdot 0.2 = 34800$ грн.

Додаткова заробітна плата складає – $С_{ЗП} = С_Г \cdot (1 + З) = 34800 \cdot (1 + 0.2) = 41760$ грн.

Відрахування на соціальний внесок – $С_{ВІД} = С_{ЗП} \cdot 0.22 = 41760 \cdot 0.22 = 9187.2$ грн.

На амортизаційні відрахування ведемо розрахунок при амортизації 25% та вартості ЕОМ – 16000 грн. Маємо $С_А = К_{ТМ} \cdot К_А \cdot Ц_{ПР} = 1.2 \cdot 0.25 \cdot 16000 = 4800$ грн. Тут $К_{ТМ}$ позначає коефіцієнт, який враховує витрати на

транспортування та монтаж приладу у користувача.

Витрати на ремонт та профілактику становлять $С_Р = К_{ТМ} \cdot К_Р \cdot Ц_{ПР} = 1.2 \cdot 0.09 \cdot 16000 = 1728$, де $К_Р$ – це відсоток витрат на поточні ремонти.

Ефективний годинний фонду часу персонального комп'ютера за рік розраховуємо за формулою:

$$Т_{ЕФ} = (Д_К - Д_В - Д_С - Д_Р) \cdot е \cdot К_В = (365 - 104 - 12 - 16) \cdot 8 \cdot 0.3, \quad (4.17)$$
$$= 559.2 \text{ години},$$

де ДК – календарна кількість днів у році;

ДВ, ДС – кількість вихідних та святкових днів відповідно;

ДР – кількість днів планових ремонтів устаткування;

t – кількість робочих годин в день;

КВ – коефіцієнт використання приладу у часі протягом зміни.

Витрати на електроенергію становлять:

$$С_{ЕЛ} = Т_{ЕФ} \cdot N_C \cdot К_З \cdot Ц_{ЕН} = 559.2 \cdot 0.6 \cdot 0.3 \cdot 9.43 = 949.19 \text{ грн}, \quad (4.18)$$

де N_C – середньо-споживча потужність приладу;

$К_З$ – коефіцієнт зайнятості приладу;

$Ц_{ЕН}$ – тариф за 1кВт-годину електроенергії.

Накладні витрати складають - $С_Н = Ц_{ПР} \cdot 0.67 = 16000 \cdot 0.67 = 10720$ грн. Тоді річні експлуатаційні витрати можна розрахувати за формулою:

$$С_{ЕКС} = С_{ЗП} + С_{ВІД} + С_А + С_Р + С_{ЕЛ} + С_Н =$$
$$= 41760 + 9187.2 + 4800 + 1728 + 482.14 + 10720 = 68677.$$

Собівартість однієї машинно-години ЕОМ рівна – $С_{М-Г} = С_{ЕКС} / Т_{ЕФ} = 68677/559.2 = 122.8$ грн/год. Оскільки в даному випадку всі роботи, які

пов’язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу в залежності від обраного варіант реалізації розраховуються за формулою:

$$C_М = C_{М-Г} \cdot Т. \quad (4.20)$$

За формулою наведеною вище маємо:

1. $C_М = 122.8 \cdot 645 = 79206.0$ грн.;
2. $C_М = 122.8 \cdot 662 = 81293.6$ грн.

Накладні витрати складають 67% від заробітної плати, що можна розрахувати за формулою:

$$C_Н = C_{ЗП} \cdot 0.67. \quad (4.21)$$

Тоді за формулою (4.20) маємо:

1. $C_Н = 83692.62 \cdot 0.67 = 56074.06$ грн.;
2. $C_Н = 85898.47 \cdot 0.67 = 57551.97$ грн.

Отже, вартість розробки програмного продукту за варіантами розраховується за формулою:

$$C_{ПП} = C_{ЗП} + C_{ВІД} + C_М + C_Н,$$

тоді маємо:

1. $C_{ПП} = 83692.62 + 18412.38 + 79206.0 + 56074.06 = 237385.06$ грн.;
2. $C_{ПП} = 85898.47 + 18897.66 + 81293.6 + 57551.97 = 243641.70$ грн.

4.7 Вибір кращого варіанту ПП техніко-економічного рівня

Для вибору найкращого варіанту програмного продукту техніко-економічного рівня розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$^{К}ТЕР_j = {}^{К}К_j / {}^{С}Ф_j.$$

В результаті для двох різних варіантів реалізації отримуємо:

1. $К_{ТЕР1} = 4.84 / 237385.06 = 2.039 \cdot 10^{-5}$;
2. $К_{ТЕР2} = 4.39 / 243641.70 = 1.802 \cdot 10^{-5}$.

Як бачимо, то найбільш ефективним варіантом реалізації програми є перший. Він має коефіцієнт техніко-економічного рівня $К_{ТЕР1} = 2.039 \cdot 10^{-5}$.

Після виконання функціонально-вартісного аналізу програмного комплексу, що розробляється, можна зробити висновок, що з альтернатив, що залишились після першого відбору двох варіантів виконання програмного комплексу оптимальним є перший варіант реалізації програмного продукту. У нього виявився найкращий показник техніко-економічного рівня якості $К_{ТЕР} = 2.039 \cdot 10^{-5}$.

Цей варіант реалізації програмного продукту має наступні параметри:

1. вибір мови програмування Python;
2. реалізація власного процесу обробки даних;
3. готовий інтерфейс в Jupyter Notebook.

Цей варіант виконання програмного комплексу дає користувачу готовий інтерфейс з гарними налаштовуваннями візуалізацій, широкий вибір алгоритмів машинного навчання, які можна налаштовувати. Також даний варіант реалізації дозволяє налаштовувати більш точний та якісний процес

обробки даних.

4.8 Висновки до розділу 4

У цьому розділі було проведено комплексний функціонально-вартісний аналіз програмного продукту, що розробляється для підтримки прийняття рішень та оцінювання ризиків. Було детально розглянуто можливі варіанти розробки програмного комплексу, оцінено його основні функції, а також економічну доцільність різних підходів до його реалізації.

У першій частині аналізу були розглянуті та оцінені всі можливі альтернативи з точки зору технічного виконання. Результатом цього етапу стало визначення коефіцієнту технічного рівня для кожного варіанту та вибір найкращої альтернативи на основі технічних критеріїв.

У другій частині було здійснено вартісний аналіз реалізації програмного продукту, враховуючи грошові витрати, такі як заробітна плата, витрати на електроенергію та матеріальне забезпечення. На основі отриманих даних був обраний найкращий варіант з найменшими витратами, що дозволяє забезпечити економічну доцільність та ефективність продукту.

ВИСНОВКИ

У дипломній роботі виконано комплексне дослідження, спрямоване на розроблення програмного комплексу генерації та аналізу дерматоскопічних зображень раку шкіри на основі моделей FLUX з методом Flow Matching. Робота доводить, що використання синтетичних даних здатне істотно поліпшити якість діагностики у галузі медичної візуалізації за умов дефіциту та незбалансованості реальних датасетів.

У першому розділі здійснено критичний огляд сучасних генеративних підходів, зокрема дифузійних та Flow Matching-моделей, і обґрунтовано вибір FLUX як швидшої та стабільнішої альтернативи DDPM. Проаналізовано проблеми незбалансованих класифікаційних задач у дерматоскопії й показано, що традиційна аугментація не забезпечує достатнього приросту продуктивності.

Другий розділ присвячено теоретичним і методичним засадам синтетичної аугментації. Детально розглянуто математичні основи Flow Matching, механізм оптимального транспорту та внесено модифікації у вигляді LoRA-адаптерів для ефективного донавчання на фрагментарних медичних даних. Розроблено конвеєр підготовки наборів HAM10000, ISIC 2024 та PH2, що забезпечує автоматичне вирівнювання кольору, видалення артефактів і нормалізацію роздільної здатності.

У третьому розділі реалізовано два варіанти моделі (FLUX-dev і FLUX-schnell), проведено їх тренування та валідацію. Синтезовано 20 000 дерматоскопічних зображень розміром 512 × 512 px, для яких розраховано метрики FID і k-NN-precision, що підтвердило їхню реалістичність. Інтеграція синтетики у набір ISIC 2024 підвищила AUC класифікатора EfficientNet-B0 з 0,886 до 0,928 без суттєвого зростання часу тренування, а швидкість генерації одного знімка скоротилася до ≈ 2 с на GPU A100.

У четвертому розділі сформовано прототип програмної системи (Python + PyTorch), який автоматизує генерування, відбір якісних зразків і їх подальшу оцінку. Проведений функціонально-вартісний аналіз засвідчив економічну перевагу синтетичної аугментації порівняно з традиційними методами

розширення датасетів за рахунок значного скорочення потреби у дорогих клінічних дослідженнях.

Отримані результати підтверджують, що Flow Matching-моделі з LoRA-адаптацією дають змогу оперативно формувати збалансовані набори зображень високої якості й помітно підвищувати точність класифікаційних систем у сфері дерматології. Запропонований підхід може бути інтегрований у клінічні системи підтримки прийняття рішень та слугувати основою для подальших досліджень з мультиспектральної візуалізації, клінічної валідації синтетичних даних і впровадження метричних методів контролю аномалій.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Tips for Early Detection of Melanoma Skin Cancer / Melanoma Research Alliance. URL: https://www.curemelanoma.org/about-melanoma/educate-yourself (дата звернення: 10.05.2025).
2. World Health Organization Predicts Half a Million with Melanoma in 2040 / AIM at Melanoma Foundation. URL: https://www.aimatmelanoma.org/world-health-organization-predicts-half-a-million-with-melanoma-in-2040/ (дата звернення: 10.05.2025).
3. Artificial intelligence used to identify skin cancer. Stanford Report. 25 січня 2017. URL: https://news.stanford.edu/stories/2017/01/artificial-intelligence-used-identify-skin-cancer (дата звернення: 10.05.2025).
4. Mekala R. R. et al. Synthetic Generation of Dermatoscopic Images with GAN and Closed-Form Factorization. 2024. 23 p. URL: https://arxiv.org/abs/2410.05114 (дата звернення: 10.05.2025).
5. Matheson E. C., Gaskin G. Accuracy of Dermoscopy vs. Visual Inspection for Diagnosing Melanoma in Adults. American Family Physician. 01 Feb. 2020. URL: https://www.aafp.org/pubs/afp/issues/2020/0201/p145.html (дата звернення: 10.05.2025).
6. Abbasi N. R. et al. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. JAMA. 2004. Vol. 292, № 22. P. 2771–2776. DOI: 10.1001/jama.292.22.2771. URL: https://pubmed.ncbi.nlm.nih.gov/15585738/ (дата звернення: 10.05.2025).
7. Pavri S. N., Clune J., Ariyan S., Narayan D. Malignant Melanoma: Beyond the Basics. Plastic and Reconstructive Surgery. 2016. Vol. 138, No. 2. P. 330e–340e. DOI: 10.1097/PRS.0000000000002367. URL: https://pubmed.ncbi.nlm.nih.gov/27465194/ (дата звернення: 10.05.2025).
8. Tschandl P., Rosendahl C., Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin

lesions. Scientific Data. 2018. Vol. 5, Art. 180161. URL: https://www.nature.com/articles/sdata2018161 (дата звернення: 10.05.2025).

9. International Skin Imaging Collaboration (ISIC) Archive. URL: https://www.isic-archive.com/ (дата звернення: 10.05.2025).
10. Garcea F., Serra A., Lamberti F., Morra L. Data augmentation for medical imaging: a systematic literature review. Computers in Biology and Medicine. 2023. Vol. 152, Art. 106391. DOI: 10.1016/j.compbiomed.2022.106391. URL: https://www.sciencedirect.com/science/article/pii/S001048252201099X (дата звернення: 10.05.2025).
11. Qin Z. W. et al. A GAN-based image synthesis method for skin lesion classification. Comput. Methods Programs Biomed. 2020. Vol. 195, Art. 105568. DOI: 10.1016/j.cmpb.2020.105568. URL: https://pubmed.ncbi.nlm.nih.gov/32526536/ (дата звернення: 10.05.2025).
12. Goodfellow I. J. et al. Generative Adversarial Networks. 2014. 9 p. URL: https://arxiv.org/abs/1406.2661 (дата звернення: 10.05.2025).
13. Kingma D. P., Welling M. Auto-Encoding Variational Bayes. 2013. 14 p. URL: https://arxiv.org/abs/1312.6114 (дата звернення: 10.05.2025).
14. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. 2020. 25 p. URL: https://arxiv.org/abs/2006.11239 (дата звернення: 10.05.2025).
15. Liu X., Gong C., Liu Q. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. 2022. 41 p. URL: https://arxiv.org/abs/2209.03003 (дата звернення: 10.05.2025).
16. Zhao W. та ін. FlowTurbo: Towards Real-time Flow-Based Image Generation with Velocity Refiner. 2024. 18 p. URL: https://arxiv.org/abs/2409.18128 (дата звернення: 10.05.2025).
17. Black Forest Labs' Flux.1 outperforms top text-to-image models. The Batch (DeepLearning.ai). 15 Aug. 2024. URL: https://www.deeplearning.ai/the-

batch/black-forest-labs-flux-1-outperforms-top-text-to-image-models/ (дата звернення: 10.05.2025).

18. Types of Skin Cancer / American Academy of Dermatology. URL: https://www.aad.org/public/diseases/skin-cancer/types/common (дата звернення: 10.05.2025).
19. Fjelde T., Mathieu E., Dutordoir V. An Introduction to Flow Matching. Cambridge MLG Blog. 2024. URL: https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html (дата звернення: 10.05.2025).
20. Lipman Y., Chen R. T. Q., Ben-Hamu H., Nickel M., Le M. Flow Matching for Generative Modeling. 2022. 28 p. URL: https://arxiv.org/abs/2210.02747 (дата звернення: 10.05.2025).
21. Esser P., Kulal S., Blattmann A., Entezari R., Müller J., Saini H., Levi Y., Lorenz D., Sauer A., Boesel F., Podell D., Dockhorn T., English Z., Lacey K., Goodwin A., Marek Y., Rombach R. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. 2024. 28 p. URL: https://arxiv.org/abs/2403.03206 (дата звернення: 10.05.2025).
22. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. OpenReview. 2024. URL: https://openreview.net/forum?id=FPnUhsQJ5B (дата звернення: 10.05.2025).
23. Black Forest Labs. FLUX.1 - State-of-the-Art Text-to-Image Models. 2024. URL: https://blackforestlabs.ai (дата звернення: 04.06.2025).
24. FLUX.1 [dev] - 12B parameter rectified flow transformer. Hugging Face. 2024. URL: https://huggingface.co/black-forest-labs/FLUX.1-dev (дата звернення: 10.05.2025).
25. FLUX.1 [schnell] - Fast text-to-image generation model. Hugging Face. 2024. URL: https://huggingface.co/black-forest-labs/FLUX.1-schnell (дата звернення: 10.05.2025).
26. Official inference repo for FLUX.1 models. GitHub. 2024. URL: https://github.com/black-forest-labs/flux (дата звернення: 04.06.2025).

27. Hu E. J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., Wang L., Chen W. LoRA: Low-Rank Adaptation of Large Language Models. 2021. 26 p. URL: https://arxiv.org/abs/2106.09685 (дата звернення: 10.05.2025).
28. LoRA: Low-Rank Adaptation of Large Language Models. OpenReview. 2021. URL: https://openreview.net/forum?id=nZeVKeeFYf9 (дата звернення: 10.05.2025).
29. Microsoft LoRA - Implementation of Low-Rank Adaptation. GitHub. 2021. URL: https://github.com/microsoft/LoRA (дата звернення: 10.05.2025).
30. What is LoRA (Low-Rank Adaption)? IBM Documentation. 2024. URL: https://www.ibm.com/think/topics/lora (дата звернення: 10.05.2025).
31. Betzalel E., Penso C., Fetaya E. Evaluation Metrics for Generative Models: An Empirical Study. Machine Learning and Knowledge Extraction. 2024. Vol. 6, No. 3. P. 1531–1544. DOI: 10.3390/make6030073. URL: https://www.mdpi.com/2504-4990/6/3/73 (дата звернення: 10.05.2025).
32. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612. DOI: 10.1109/TIP.2003.819861.
33. Structural similarity index measure. Wikipedia. 2025. URL: https://en.wikipedia.org/wiki/Structural_similarity_index_measure (дата звернення: 10.05.2025).
34. Understanding SSIM. 2020. 8 p. URL: https://arxiv.org/abs/2006.13846 (дата звернення: 04.06.2025).
35. ISIC 2024 - Skin Cancer Detection with 3D-TBP / Kaggle. URL: https://www.kaggle.com/competitions/isic-2024-challenge (дата звернення: 10.05.2025).
36. Topic 20: What is Flow Matching? TuringPost. 2024. URL: https://turingpost.com/flow-matching (дата звернення: 04.06.2025).

37. How to Evaluate the Quality of Synthetic Images. LinkedIn Learning. 2024. URL: https://www.linkedin.com/pulse/evaluation-metrics-generative-models-judging-ais-suraj-bhardwaj-liqhf (дата звернення: 04.06.2025).
38. Dohmen M., Klemens M. A., Baltruschat I. M., Truong T., Lenga M. Similarity and quality metrics for MR image-to-image translation. Scientific Reports. 2025. Vol. 15, Article 3853. DOI: 10.1038/s41598-025-87358-0. URL: https://www.nature.com/articles/s41598-025-87358-0 (дата звернення: 10.05.2025).
39. Synthetic data generation methods in healthcare: A review on open challenges and recent trends. Science Direct. 2024. URL: https://www.sciencedirect.com/topics/engineering/synthetic-data-generation (дата звернення: 10.05.2025).
40. Харитонов О. Д. Колекція моделей та датасетів для генерації зображень раку шкіри Hugging Face Hub. 2025. URL: https://huggingface.co/collections/Shah1st/skin-cancer-684062cafa430a503be97899 (дата звернення: 08.05.2025).

ДОДАТОК А ЛІСТИНГ КОДУ

flux_ostris_train.py

```
from huggingface_hub import login
from google.colab import userdata
import wandb
import os

hf_token = userdata.get('hf_shah1st')
wandb_key = userdata.get("wandb")

login(token=hf_token)
wandb.login(key=wandb_key)
os.environ['HF_TOKEN'] = hf_token
import os
from PIL import Image

def convert_bmp_to_jpg(source_folder, output_folder):
    """Convert BMP files to JPG and resize to 512x512 for training"""
    os.makedirs(output_folder, exist_ok=True)

    image_files = [f for f in os.listdir(source_folder) if f.endswith('.bmp')]
    print(f"Found {len(image_files)} BMP files.")

    for img_file in image_files:
        base_name = os.path.splitext(img_file)[0]
        img_path = os.path.join(source_folder, img_file)
        txt_path = os.path.join(source_folder, f"{base_name}.txt")

        output_img_path = os.path.join(output_folder, f"{base_name}.jpg")
        output_txt_path = os.path.join(output_folder, f"{base_name}.txt")
```

```
        image = Image.open(img_path).convert('RGB')

        image.thumbnail((512, 512))

        new_image = Image.new('RGB', (512, 512), (0, 0, 0))

        offset = ((512 - image.width) // 2, (512 - image.height) // 2)
        new_image.paste(image, offset)

        new_image.save(output_img_path, 'JPEG', quality=95)

        if os.path.exists(txt_path):
            with open(txt_path, 'r', encoding='utf-8') as src_file:
                content = src_file.read()

                if not content.startswith("SKINCANCER:"):
                    content = f"SKINCANCER: {content}"

                with open(output_txt_path, 'w', encoding='utf-8') as dst_file:
                    dst_file.write(content)
        else:
            with open(output_txt_path, 'w', encoding='utf-8') as f:
                f.write("SKINCANCER: Dermatoscopic image of skin lesion")

    print(f"Converted {len(image_files)} images to JPG format at 512x512 resolution.")


source_folder = "/content/drive/MyDrive/Datasources/Datasets/PH2"
output_folder = "/content/skin_cancer_dataset"

convert_bmp_to_jpg(source_folder, output_folder)
import os
import sys
sys.path.append('/content/ai-toolkit')
```

```
from toolkit.job import run_job
from collections import OrderedDict

# Ensuring output directory exists
save_path = "/content/drive/MyDrive/Datasources/Datasets/skin_cancer_flux_lora"
os.makedirs(save_path, exist_ok=True)

job_to_run = OrderedDict([
    ('job', 'extension'),
    ('config', OrderedDict([
        # this name will be the folder and filename name
        ('name', 'skin_cancer_flux_lora'),
        ('process', [
            OrderedDict([
                ('type', 'sd_trainer'),
                # Save directly to Google Drive
                ('training_folder', '/content/drive/MyDrive/Datasources/Datasets'),
                ('device', 'cuda:0'),
                # Trigger word for your skin cancer images
                ('trigger_word', 'SKINCANCER'),
                ('network', OrderedDict([
                    ('type', 'lora'),
                    ('linear', 16),  # Good for medical details
                    ('linear_alpha', 32)
                ])),
                ('save', OrderedDict([
                    ('dtype', 'bfloat16'),  # precision to save
                    ('save_every', 250),  # save every this many steps
                    ('max_step_saves_to_keep', 4)  # how many intermittent saves to keep
                ])),
                ('datasets', [
                    OrderedDict([
                        ('folder_path', '/content/skin_cancer_dataset'),
                        ('caption_ext', 'txt'),
```

```
        ('caption_dropout_rate', 0.05),
        ('shuffle_tokens', False),
        ('cache_latents_to_disk', True),
        ('resolution', [512, 512])
    ])
]),
('train', OrderedDict([
    ('batch_size', 2),  # Increased for A100
    ('steps', 2000),  # ~10x number of images is a good rule of thumb
    ('gradient_accumulation_steps', 1),
    ('train_unet', True),
    ('train_text_encoder', False),
    ('content_or_style', 'content'),  # Better for medical details
    ('gradient_checkpointing', True),
    ('noise_scheduler', 'flowmatch'),
    ('optimizer', 'adamw8bit'),
    ('lr', 2e-4),
    ('ema_config', OrderedDict([
        ('use_ema', True),
        ('ema_decay', 0.99)
    ])),
    ('dtype', 'bf16')
])),
('model', OrderedDict([
    ('name_or_path', 'black-forest-labs/FLUX.1-dev'),
    ('is_flux', True),
    ('quantize', True),
    # A100 has plenty of VRAM
    ('low_vram', False),
])),
('sample', OrderedDict([
    ('sampler', 'flowmatch'),
    ('sample_every', 250),
    ('width', 512),
```

```
            ('height', 512),
            ('prompts', [
                # Test prompts for skin cancer images
                'SKINCANCER: Dermatoscopic image of a common nevus, fully symmetrical shape, with uniform dark brown coloration',
                'SKINCANCER: Dermatoscopic image of melanoma with irregular borders and multiple colors',
                'SKINCANCER: High-resolution clinical photograph of basal cell carcinoma',
                'SKINCANCER: Dermatoscopic image of seborrheic keratosis with typical features'
            ]),
            ('neg', ''),  # not used on flux
            ('seed', 42),
            ('walk_seed', True),
            ('guidance_scale', 3.5),
            ('sample_steps', 28)  # Good for FLUX.1-dev
        ]))
    ])
  ])),
  ('meta', OrderedDict([
    ('name', '[name]'),
    ('version', '1.0')
  ]))
])
run_job(job_to_run)
import torch
from diffusers import FluxPipeline

pipeline = FluxPipeline.from_pretrained(
  "black-forest-labs/FLUX.1-dev",
  torch_dtype=torch.bfloat16
)

lora_path = "/content/drive/MyDrive/Datasources/Datasets/skin_cancer_flux_lora"
```

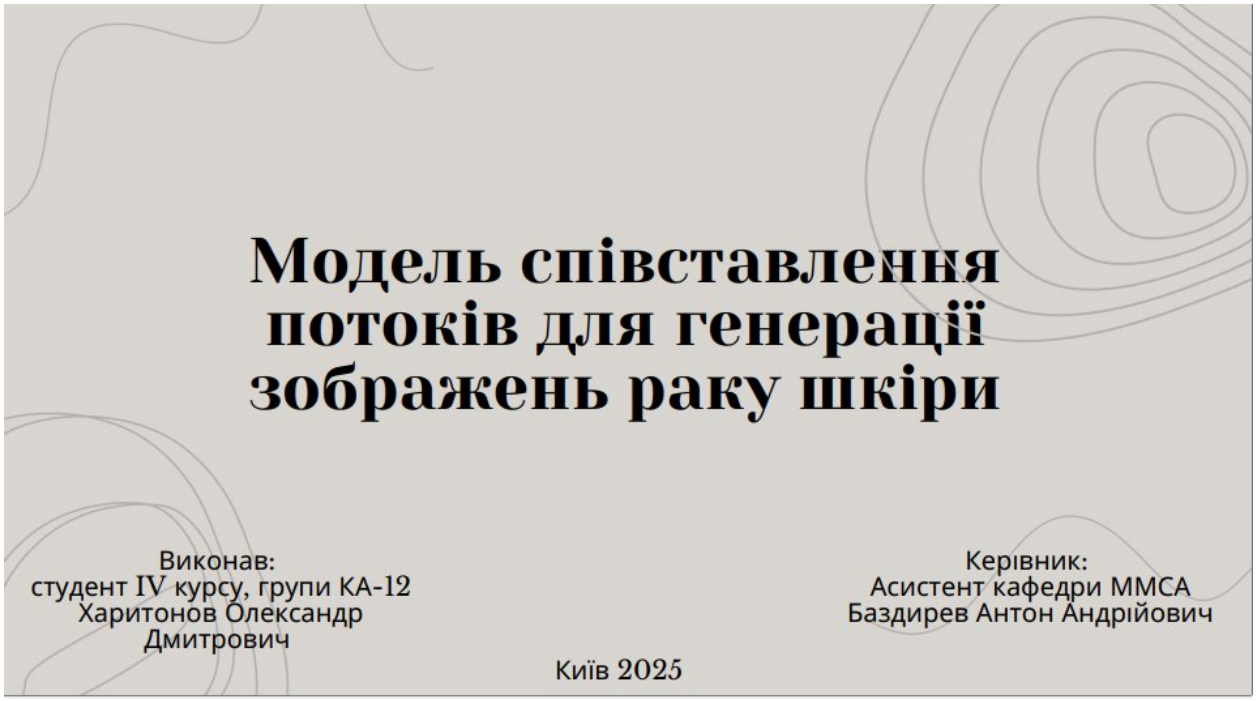
```
pipeline.load_lora_weights(lora_path, weight_name='skin_cancer_flux_lora.safetensors')

pipeline.enable_model_cpu_offload()

prompt = "SKINCANCER: Dermatoscopic image of a common nevus, fully symmetrical shape, with uniform dark brown coloration"
image = pipeline(
    prompt=prompt,
    num_inference_steps=10,
    guidance_scale=3.5,
    height=512,
    width=512,
).images[0]

image.save("generated_skin_image.png")

from IPython.display import display
display(image)
```

fluxlorainference.py

```
from huggingface_hub import login
from google.colab import userdata

hf_token = userdata.get('hf_shah1st')

login(token=hf_token)
import torch
from diffusers import FluxPipeline

torch.cuda.empty_cache()

# Load the base model with optimizations
```

```
pipeline = FluxPipeline.from_pretrained(
    "black-forest-labs/FLUX.1-dev",
    torch_dtype=torch.bfloat16,  # Use bfloat16 for speed
    #device_map="balanced"  # Automatically manage device placement
)

# Load your LoRA weights
lora_path = "/content/drive/MyDrive/Datasources/Datasets/skin_cancer_flux_lora"
pipeline.load_lora_weights(lora_path, weight_name='skin_cancer_flux_lora.safetensors')

# Move to GPU and optimize memory usage
#pipeline = pipeline.to("cuda")
pipeline.enable_model_cpu_offload()  # Only keep what's needed on GPU

# Reduce inference steps for faster generation
prompt = "SKINCANCER: Dermatoscopic image of a common nevus, fully symmetrical shape, with uniform dark brown coloration"
image = pipeline(
    prompt=prompt,
    num_inference_steps=20,  # Reduced from 28
    guidance_scale=3.5,
    height=512,
    width=512,
).images[0]

# Save the image
image
```

ДОДАТОК Б ПРЕЗЕНТАЦІЯ



Актуальність

● Зростання захворюваності на рак шкіри
За останні 30 років кількість випадків меланоми зросла утричі. До 2040 року очікується 510 тис. нових випадків щорічно.
● Критична важливість ранньої діагностики
5-річна виживаність при ранньому виявленні — 97%, при пізніх стадіях — лише 14%.
● Катастрофічна нестача медичних даних
Існуючі датасети характеризуються критичним дисбалансом класів: злоякісні ураження складають <1% від загальної кількості зображень.
● Обмеження традиційних методів аугментації
Звичайні трансформації не здатні створювати медично значущі варіанти патологій.

Цілі роботи

Аспект	Опис
Об'єкт дослідження	Програмний комплекс для генерації та аналізу дерматоскопічних зображень раку шкіри на основі моделей Flow Matching
Предмет дослідження	Методи синтетичної генерації зображень, їх якісного оцінювання та впливу штучних даних на точність класифікації шкірних утворень
Мета роботи	Розробити й експериментально перевірити конвеєр на основі Flow Matching, що компенсує дефіцит медичних даних і підвищує достовірність діагностики

Постановка задачі

01 Дослідити теоретичні основи Flow Matching та його переваги для медичної генерації

02 Адаптувати архітектуру FLUX для генерації дерматоскопічних зображень

03 Розробити пайплайн контрольованої генерації за типами раку шкіри

04 Створити систему комплексної оцінки якості синтетичних медичних зображень

05 Провести експерименти з класифікацією та оцінити практичну користь

Безперервні нормалізуючі потоки

Ключова ідея CNF:
Моделювання безперервних перетворень між розподілами через звичайні диференціальні рівняння:

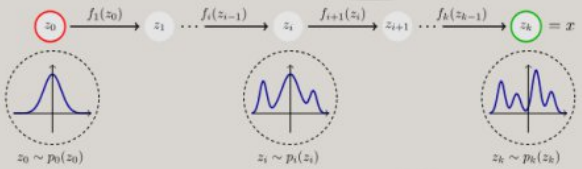
$dz/dt = f_\theta(z(t), t)$

Переваги CNF:

- Точне обчислення ймовірностей
- Оборотні перетворення
- Безперервні латентні простори
- Можливість інтерполяції

Проблема: Обчислення якобіану $\nabla \cdot f_\theta$ є дорогим та повільним

Рішення Flow Matching: Пряме навчання векторного поля без обчислення ймовірностей



Flow Matching: Теоретичні основи

Основна ідея:
Навчити векторне поле $v_\theta(t,x)$ що переносить частинки від простого розподілу p_0 до цільового p_1:

$dx_t/dt = v_\theta(t, x_t)$
Цільова функція (проста MSE):

$L_{FM} = E[||v_\theta(t,x_t) - (x_1 - x_0)||^2]$
Ключові переваги:

- Уникає обчислення якобіану CNF
- Використовує оптимальні траєкторії (пряма інтерполяція)
- Стабільне навчання через просту функцію втрат
- Швидкий інференс через детерміністичні ОДР

Flow Matching vs DDPM

Аспект	DDPM	Flow Matching	Покращення
Кількість кроків	1000+	20-28	50+ швидше
Природа процесу	Стохастичний	Детермінистичний	Передбачуваність
Навчання	Складна функція втрат	Проста MSE	Стабільність
Час генерації (A100)	~30 секунд	~2 секунди	15+ швидше

Критично для медицини:

- Стабільність: Детерміністичні результати важливі для діагностики
- Швидкість: Real-time генерація для клінічних застосувань
- Якість: Збереження дрібних медичних деталей

Практичний результат: Flow Matching робить генерацію медичних зображень практично застосовною.

Архітектура FLUX та адаптація

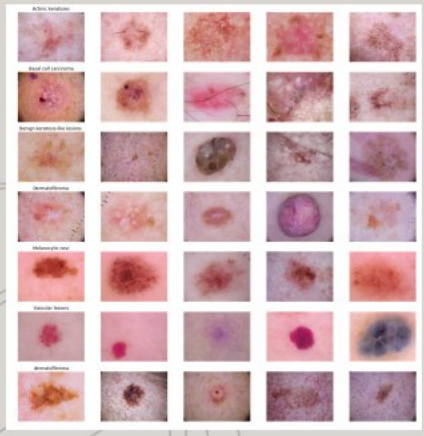
FLUX: Трансформерна архітектура для Flow Matching

Ключові компоненти:
- DiT (Diffusion Transformer) як backbone
- Multi-head attention для глобального контексту зображень
- RMS LayerNorm для стабільності навчання
- RoPE позиційні кодування для просторових залежностей

Дві варіації:
- FLUX.1-dev: Максимальна якість (28 кроків inference)
- FLUX.1-schnell: Швидкість (4-8 кроків) для практичних застосувань

Адаптація для медичного домену:
- Навчання на медичній термінології в промптах
- Збереження діагностично критичних деталей
- Контрольована генерація за типами раку шкіри

Датасети та проблема дисбалансу



Медичні датасети та їхні виклики

HAM10000 (10,015 зображень):
- Дисбаланс: 88.8% доброякісні vs 11.2% злоякісні
- 7 категорій шкірних уражень
- Різноманітні джерела та умови зйомки

ISIC 2024 (137,745 зображень):
- Критичний дисбаланс: 99.1% vs 0.9%
- Найбільший публічний датасет
- Високоякісні клінічні анотації

PH2 (200 зображень):
- Експертна розмітка з сегментацією
- Еталонний стандарт якості
- Використано для навчання FLUX моделей

Основний виклик: Як навчити генерувати рідкісні злоякісні ураження?

Контрольована генерація медичних зображень

Структуровані медичні промпти:

Доброякісні ураження:

"SKINCANCER: BENIGN
dermatoscopic image of common nevus,
fully symmetrical shape, uniform light brown coloration,
typical pigment network, high-resolution clinical photograph"

Злоякісні ураження (меланома):

"SKINCANCER: MALIGNANT
dermatoscopic image of melanoma,
asymmetrical shape, variegated coloration (dark brown, black, blue-gray), atypical pigment network,
irregular borders"

Створено 10,000 унікальних промптів
Збалансований розподіл: 50% кожного класу

Технічна реалізація системи

Технологічний стек:
- Python 3.11 + PyTorch 2.1.0 + CUDA 11.8
- Hugging Face Transformers/Diffusers для FLUX моделей
- Google Colab Pro з NVIDIA A100 (40GB VRAM)

Архітектура системи:
1. Препроцесинг: Нормалізація зображень 512×512, створення промптів
2. Навчання: Fine-tuning FLUX на PH2 датасеті (5-8 годин)
3. Генерація: Пакетна генерація з оптимізацією
4. Оцінка: Автоматичне обчислення FID, IS, LPIPS, SSIM

Оптимізації швидкості:
- torch.compile: Прискорення інференсу в 10× разів
- Mixed precision: BFloat16 для економії пам'яті
- Batch processing: Паралельна генерація до 16 зображень

Результати генерації та якість

Метрика	FLUX.1-schnell	FLUX.1-dev	Інтерпретація
FID	11.309	12.390	Кращі статистичні властивості
KID	612	1.002	Більш стабільний показник
IS	2.74 ± 0.15	2.06 ± 0.03	Вища різноманітність
LPIPS	0.600 ± 0.074	0.554 ± 0.065	Кращі перцептивні властивості

Ключові висновки:

- FLUX.1-schnell: Краще відтворює статистичний розподіл та різноманітність
- FLUX.1-dev: Вища якість окремих зображень
- Компроміс: Швидкість + різноманітність vs детальна якість

Практичний результат: Обидві моделі генерують медично правдоподібні зображення за секунди.

Експерименти з класифікацією – HAM10000

Вплив синтетичних даних на класифікацію (HAM10000)

Архітектура класифікатора:

- EfficientNet-B0 (5.3M параметрів)
- Transfer learning з ImageNet
- Бінарна класифікація: доброякісні vs злоякісні

Конфігурація	Test AUC	F1-score	Accuracy
Baseline (тільки реальні)	0,923	0,646	0,894
+50% синтетичних	0,928	0,669	0,903
+100% синтетичних	0,923	0,663	0,903
+200% синтетичних	0,930	0,671	0,912

Висновок: Синтетичні дані стабільно покращують всі метрики класифікації.

Експерименти з класифікацією – ISIC 2024

Клінічна валідація (Kaggle ISIC 2024):

- Базова модель: pAUC (>80% TPR) = 0.079
- З синтетичними даними: pAUC = 0.103
- Покращення: +30.1% на клінічно значущій метриці

Конфігурація	Test AUC	F1-score	Accuracy
Baseline	0,887	0,250	0,977
+50% синтетичних	0,886	0,250	0,971
+100% синтетичних	0,928	0,210	0,970
+200% синтетичних	0,911	0,220	0,977

Критичний висновок: Flow Matching частково вирішує проблему дисбалансу класів у медичних даних.

Практичні досягнення та впровадження

Опубліковані ресурси:

- Hugging Face Hub: Shahlst/skin-cancer-flux-models
- Датасети: 20,000 синтетичних зображень у відкритому доступі
- Код: Повний пайплайн генерації та оцінки

Технічні характеристики системи:

- Швидкість генерації: 1-2 секунди на зображення (A100)
- Якість: Медично правдоподібні результати
- Масштабованість: Можливість генерації необмежених обсягів

Потенційні застосування:

- Збагачення датасетів для дослідження
- Тестування діагностичних алгоритмів
- Симуляції рідкісних медичних випадків

Обмеження та майбутні дослідження

Поточні обмеження:
- FID метрики: Значення >100 свідчать про різницю з реальними даними
- Рідкісні ознаки: Складність відтворення нетипових медичних проявів
- Клінічна валідація: Потреба в експертній оцінці медичних фахівців

Напрями покращення:
- Спеціалізовані архітектури для різних типів медичних зображень
- Мультимодальне навчання (зображення + клінічні дані + генетика)
- Континуальне навчання з постійним оновленням на нових випадках

Майбутні застосування:
- Розширення на МРТ, КТ, рентгенографію
- Персоналізована медицина
- Real-time діагностичні системи

Висновки

Проведене дослідження вперше демонструє успішне застосування Flow Matching для генерації дерматоскопічних зображень раку шкіри, відкриваючи нові можливості для вирішення критичної проблеми дефіциту медичних даних.
Розроблений підхід забезпечує генерацію 20,000 реалістичних синтетичних зображень за 8-10 годин на сучасному обладнанні, що в 50 разів швидше за традиційні дифузійні моделі при збереженні медичної достовірності результатів. Експериментальна валідація на датасетах HAM10000 та ISIC 2024 підтвердила практичну цінність синтетичних даних для покращення діагностичних моделей, особливо в умовах критичного дисбалансу класів.
Найбільш значущим досягненням є 30% покращення клінічно релевантної метрики pAUC на офіційному конкурсі ISIC 2024, що свідчить про готовність технології до практичного застосування в медичній діагностиці.
Публікація відкритих моделей та кодової бази створює основу для подальших досліджень та впровадження в клінічну практику, потенційно трансформуючи підходи до навчання медичних спеціалістів та розробки діагностичних систем. Flow Matching тим самим встановлює новий стандарт для генерації медичних зображень, поєднуючи швидкість, якість та практичну застосовність.


Дякую за
увагу