

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**Факультет інформаційних технологій**

Кафедра технологій управління

Спеціальність 122 – Комп'ютерні науки,  
освітньо-наукова програма «Інформаційна аналітика та впливи»

**КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА**

на тему:

**«РОЗРОБКА ТЕХНОЛОГІЇ ДІАГНОСТУВАННЯ МЕДИЧНИХ ЗНІМКІВ  
МЕТОДАМИ DATA SCIENCE»**

**Студентки 2-го курсу групи ІАВ-21**  
Шпирук Євгенії

**Науковий керівник:**  
д-р. техн. наук

(науковий ступінь, вчене звання)  
Олег ЗАРІЦЬКИЙ

\_\_\_\_\_  
(підпис студента)

\_\_\_\_\_  
(дата)

\_\_\_\_\_  
(підпис)

**Попередній захист:**

\_\_\_\_\_  
(Висновок: «До захисту в Екзаменаційній комісії»)

Завідувач кафедри  
технологій управління

\_\_\_\_\_  
(підпис)

\_\_\_\_\_  
(прізвище, ініціали)

\_\_\_\_\_  
(дата)

**Київ - 2025**

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ  
ТАРАСА ШЕВЧЕНКА**

**Факультет інформаційних технологій**

Кафедра технологій управління  
Освітньо-кваліфікаційний рівень Магістр  
Спеціальність 122 – Комп'ютерні науки  
Освітня програма Інформаційна аналітика та впливи

**ЗАТВЕРДЖУЮ**  
Завідувач кафедри  
професор Морозов В.В.

«\_\_» \_\_\_\_\_ 2024 р.

**З А В Д А Н Н Я  
НА ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ**

Студент Шпирук Євгенія Сергіївна  
Група ІАВ-21

**1. Тема кваліфікаційної роботи**

«Розробка технології діагностування медичних знімків методами Data Science»

Затверджена наказом по університету від «\_\_» \_\_\_\_\_ 2025 р.  
№ \_\_\_\_\_.

**2. Строк подання студентом готової роботи – «01» травня 2025 р.**

**3. Цільова установка та вихідні дані до роботи:** У межах цього дослідження розглядається застосування методів машинного навчання для вирішення задачі автоматичної класифікації зображень у медичній сфері. Основна увага приділяється побудові моделей, які здатні виявляти патології на рентгенівських знімках грудної клітки. У процесі аналізу порівнюються різні архітектурні підходи до моделювання, що дозволяє оцінити їх ефективність у контексті практичного застосування.

Вихідними даними для дослідження є медичний датасет ChestX-ray14, який містить знімки та анотації до 14 типів патологій.

#### **4. Зміст роботи**

У межах цієї роботи було здійснено аналіз задачі класифікації у сфері медичної діагностики з використанням методів глибокого навчання. Дослідження включає порівняння кількох підходів до моделювання, що дозволяє оцінити їх ефективність з точки зору точності та можливості практичного застосування. Практичну цінність результатів демонструє концепція інтеграції моделей у систему попереднього автоматичного аналізу рентгенівських знімків, що може слугувати основою для клінічного застосування в системах підтримки прийняття медичних рішень.

#### **5. Перелік графічного матеріалу (слайдів)**

Дана кваліфікаційна робота магістра містить 12 рисунків, серед яких представлено графіки розподілу даних, ілюстрації архітектур моделей глибокого навчання, теплові карти Grad-CAM, графіки точності та інші візуалізації, що супроводжують процес побудови та аналізу діагностичної системи. У роботі також наведено 9 формул, які описують ключові метрики класифікації, функції втрат, принципи ансамблювання моделей тощо. Окрім цього, представлено 3 таблиці з характеристиками набору даних, частотами класів, результатами оцінювання ефективності моделей, що забезпечує повноту технічного й аналітичного обґрунтування.

#### **6. Календарний план виконання роботи:**

№ п/п	Назва частин роботи	%	Виконання роботи	
			За планом	Фактично
1.	Вибір теми дипломної роботи	3	01.10.24	01.10.24
2.	Протокол кафедри ТУ про затвердження тем дипломних робіт та призначення наукових керівників	2	27.12.24	27.12.24
3.	Формування переліку нормативних матеріалів, літератури з проблематики дипломної роботи	10	08.01.25	07.01.25
4.	Складання розгорнутого плану кваліфікаційної роботи	5	18.01.25	18.01.25

5.	Ознайомлення наукового керівника з розгорнутим планом кваліфікаційної роботи. Внесення змін.	5	19.01.25 - 20.01.25	20.01.25
6.	Підготовка розділу 1 «Історія, розвиток та аналіз існуючих рішень для емоційного аналізу»	10	12.02.25	13.02.25
7.	Підготовка розділу 2 «Формалізація методу емоційного аналізу»	14	08.03.25	08.03.25
8.	Підготовка розділу 3 «Розробка методу емоційного аналізу»	14	20.03.25	20.03.25
9.	Підготовка розділу 4 «Впровадження розробленої технології»	13	15.04.25	15.04.25
10.	Оформлення кваліфікаційної роботи. Підготовка висновків і пропозицій	15	25.04.25	25.04.25
11.	Передача кваліфікаційної роботи науковому керівникові	2	01.05.25	01.05.25
12.	Передача кваліфікаційної роботи рецензенту для рецензування	2	04.05.25	04.05.25
13.	Попередній захист кваліфікаційної роботи	5	10.05.25	10.05.25

Дата видачі завдання «\_\_\_» \_\_\_\_\_ 2025 р.

Керівник роботи

д-р. техн. наук

Заріцький Олег Володимирович

(посада, прізвище, ім'я, по батькові)

\_\_\_\_\_  
(підпис)

Завдання прийняла до виконання студентка групи

ІАВ-21

Шпирук Євгенія Сергіївна

(прізвище, ім'я, по батькові)

\_\_\_\_\_  
(підпис)

## ЗМІСТ

АНОТАЦІЯ	7
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКРОЧЕНЬ ТА ТЕРМІНІВ	9
ВСТУП	11
РОЗДІЛ 1 ОГЛЯД МЕТОДІВ ЗАСТОСУВАННЯ DATA SCIENCE У МЕДИЧНІЙ ДІАГНОСТИЦІ	15
1.1 Аналіз практичної сфери застосування методології Data Science	15
1.2 Вибір та наповнення методології проекту розробки технології Data Science	17
1.3 Дослідження літературних джерел щодо застосування діагностичних моделей.....	20
1.4 Огляд моделей Data Science та існуючих підходів до медичної діагностики	23
1.5 Постановка задачі	25
1.6 Висновки	27
РОЗДІЛ 2 ФОРМАЛІЗАЦІЯ МЕТОДІВ АНАЛІЗУ ДАНИХ У ПРОЄКТИ МЕДИЧНОЇ ДІАГНОСТИКИ	29
2.1 Попередній аналіз та обробка медичних зображень	29
2.2 Математичний опис методів побудови моделей для медичної діагностики	31
2.3 Метрики оцінювання ефективності моделей діагностики	35
2.4 Переваги та недоліки обраних методів	38
2.5 Мова програмування, основні бібліотеки та інструменти	39
2.6 Висновки	42
РОЗДІЛ 3 РОЗРОБКА МЕТОДУ DATA SCIENCE	43
3.1 Аналіз даних для моделі	43
3.2 Побудова окремих моделей	50
3.3 Оцінка якості моделей	53
3.4 Висновки	56
РОЗДІЛ 4 ІНТЕГРАЦІЯ РОЗРОБЛЕНИХ МОДЕЛЕЙ У ПРАКТИКУ МЕДИЧНОЇ ДІАГНОСТИКИ	58
4.1 Застосування побудованих моделей на практиці	58
4.2 Технічна характеристика імплементації системи автоматичної діагностики	60
4.3 Перспективи майбутніх досліджень	64
4.4 Висновки	66

ВИСНОВКИ	68
СПИСОК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ	70
ДОДАТКИ	73
ДОДАТОК А. Програмний код реалізації	73

**АНОТАЦІЯ**  
**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА**  
**ШЕВЧЕНКА**

**Факультет інформаційних технологій**  
Кафедра технологій управління  
Освітньо-кваліфікаційний рівень Магістр  
Спеціальність 122 – Комп'ютерні науки  
Освітня програма Інформаційна аналітика та впливи

Кваліфікаційна робота магістра Шпирук Євгенії Сергіївни.

Тема роботи – «Розробка технології діагностування медичних знімків методами Data Science».

Мета роботи. Розробка методики побудови системи автоматичної діагностики захворювань органів грудної клітки на основі рентгенівських знімків з використанням методів Data Science.

Завдання роботи. Оцінка можливостей застосування сучасних моделей комп'ютерного зору для автоматичної діагностики рентгенівських знімків у практичних умовах, визначення конкретних етапів проєкту, аналіз наукових публікацій та узагальнення підходів до побудови діагностичних моделей. До завдань також входить формування набору відповідних методів Data Science, що включають згорткові нейронні мережі, трансформерні архітектури та ансамблі класичних алгоритмів, реалізація обраних рішень за допомогою мови програмування Python, аналіз їх ефективності на основі результатів тестування, а також підготовка звіту за підсумками дослідження.

Об'єкт дослідження. Процес автоматизованої діагностики рентгенівських знімків грудної клітки за допомогою методів Data Science.

Предмет дослідження. Методологія побудови та оцінювання моделей автоматизованої медичної діагностики, виявлення закономірностей у

медичних знімках та узагальнення знань для покращення процесу клінічного прийняття рішень.

Наукова новизна роботи. Порівняльний аналіз трьох підходів до автоматизованої класифікації рентгенівських знімків грудної клітки, а саме згорткової нейронної мережі, трансформерної архітектури та їх ансамблю. У рамках дослідження оцінено ефективність кожного з підходів на реальних медичних даних з урахуванням точності класифікації, стійкості до варіативності вхідних даних та здатності до узагальнення. Додаткову увагу приділено інтерпретації результатів за допомогою методів візуалізації важливих ознак, що має практичну цінність у медичних задачах, де необхідна прозорість прийнятих рішень.

Кваліфікаційна робота магістра складається з анотації, вступу, основної частини, яка включає 4 розділи, висновків, списку використаних джерел та додатків. Всього налічує 82 сторінки та перелік посилань з 25 джерел.

Ключові слова. Медична діагностика, Data Science, глибоке навчання, рентгенівські знімки, Vision Transformer, EfficientNet, багатокласова класифікація, ансамблева модель.

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ

PACS (Picture Archiving and Communication System) – система архівування та передавання медичних зображень, яка використовується для зберігання, доступу та перегляду діагностичних знімків.

EMR (Electronic Medical Record) – електронна медична карта, яка містить повну історію пацієнта, включаючи обстеження, зображення та медичні висновки.

X-ray – рентгенівське зображення, отримане за допомогою рентгенографії, широко використовується в діагностиці захворювань органів грудної клітки.

CNN (Convolutional Neural Network) – згортова нейронна мережа, тип глибокої моделі, яка ефективно працює з обробкою зображень і використовується для виявлення патернів та об'єктів на медичних знімках.

ViT (Vision Transformer) – модель глибокого навчання на основі трансформерної архітектури, яка застосовується до аналізу зображень без використання згорткових операцій.

Grad-CAM (Gradient-weighted Class Activation Mapping) – метод візуалізації, який дозволяє побачити, на які ділянки зображення звертала увагу модель при ухваленні рішення.

AUC (Area Under the Curve) – площа під ROC-кривою, метрика, яка вимірює здатність моделі відокремлювати класи.

ROC-крива (Receiver Operating Characteristic curve) – графік, що відображає співвідношення між чутливістю та специфічністю класифікатора.

Accuracy (точність) – частка правильних передбачень серед усіх спроб.

Precision (точність позитивного передбачення) – частка правильних позитивних передбачень серед усіх передбачених позитивів.

Recall (чутливість) – здатність моделі правильно виявити всі позитивні приклади певного класу.

F1-score – гармонічне середнє між precision і recall, особливо важливе для задач із незбалансованими класами.

Macro AUC/Macro F1 – середнє значення метрик по всїх класах, незалежно від їх частоти.

Micro AUC/Micro F1 – метрики, розраховані з урахуванням загальної кількості зразків, що дозволяє оцінити загальну ефективність моделі.

Soft Voting – техніка ансамблювання, яка передбачає усереднення ймовірностей передбачень кількох моделей.

FastAPI – сучасний Python-фреймворк для створення швидких та асинхронних API-серверів.

AWS (Amazon Web Services) – хмарна платформа, яка використовується для хостингу моделей, API та зберігання медичних зображень.

S3 (Amazon Simple Storage Service) – сервіс для зберігання великих обсягів даних, включаючи зображення та результати обробки.

HIPAA (Health Insurance Portability and Accountability Act) – законодавчий акт США, який регламентує захист конфіденційної медичної інформації.

GDPR (General Data Protection Regulation) – загальний регламент захисту персональних даних в ЄС.

Explainable AI (XAI) – напрям у штучному інтелекті, що займається розробкою моделей, результати яких є прозорими та зрозумілими для людини.

## ВСТУП

Своєчасна і точна діагностика є однією з ключових складових ефективного функціонування системи охорони здоров'я. У багатьох випадках саме вона визначає подальшу стратегію лікування і впливає на шанси пацієнта на одужання. Візуальні методи діагностики, зокрема рентгенографія, залишаються базовим інструментом виявлення захворювань органів грудної клітки. Однак, попри свою поширеність, ці методи мають низку суттєвих обмежень: залежність від суб'єктивної оцінки лікаря, ймовірність помилок через перевантаження персоналу, а також нестачу кваліфікованих фахівців у багатьох регіонах.

Проблема ускладнюється тим, що рентгенівські знімки часто мають низьку якість, відсутність чітких контурів патологій та значну варіативність візуального представлення хвороб. За даними Всесвітньої організації охорони здоров'я, захворювання органів дихання щорічно спричиняють понад 4 мільйони смертей у світі, і ця цифра продовжує зростати. Легеневі патології, включаючи пневмонію, туберкульоз і рак легень, входять до п'ятірки найпоширеніших причин смертності. Водночас, за оцінками ВООЗ, у низці країн на одного радіолога припадає понад 100 000 осіб населення, що унеможлиблює оперативне реагування на критичні випадки.

Особливо гостро вразливість медичних систем проявилася під час пандемії COVID-19, яка поставила перед охороною здоров'я безпрецедентні виклики. У розпал пандемії лікарні по всьому світу працювали на межі своїх можливостей. Уряд Великої Британії повідомляв, що медичні працівники були змушені інтерпретувати рентгенівські знімки по 12–14 годин на добу, а у низці лікарень середній час очікування на опис знімка перевищував 24 години. Такий колапс діагностичної ланки вкотре підкреслив обмеженість традиційних підходів і актуалізував потребу в

автоматизованих рішеннях, які могли б масштабувати процес обробки медичних знімків та зменшити залежність від людського ресурсу.

У таких умовах виникає потреба в розробці нових інструментів, здатних підсилити та доповнити роботу медичного персоналу. Впровадження інтелектуальних рішень, здатних виявляти закономірності у візуальній інформації, дозволяє систематизувати діагностичний процес, зробити його більш точним і менш залежним від людського фактора. Це особливо важливо в умовах зростання кількості пацієнтів, старіння населення та поширення хронічних хвороб.

*Метою* даної кваліфікаційної роботи магістра (КРМ) є підвищення якості діагностування захворювань органів грудної клітки шляхом розробки технології побудови системи автоматичної діагностики на основі рентгенівських знімків з використанням методів Data Science.

Завданнями цієї кваліфікаційної роботи магістра є оцінка можливостей застосування сучасних моделей комп'ютерного зору для автоматичної діагностики рентгенівських знімків у практичних умовах, визначення конкретних етапів проєкту, аналіз наукових публікацій та узагальнення підходів до побудови діагностичних моделей. До завдань також входить формування набору відповідних методів Data Science, що включають згорткові нейронні мережі, трансформерні архітектури та ансамблі класичних алгоритмів, реалізація обраних рішень за допомогою мови програмування Python, аналіз їх ефективності на основі результатів тестування, а також підготовка звіту за підсумками дослідження.

Основна частина завдань даної роботи присвячена детальному огляду сучасних методів автоматичної діагностики медичних зображень, а також проведенню порівняльного аналізу ефективності трьох різних підходів до класифікації патологій на рентгенівських знімках грудної клітки. У дослідженні розглянуто згорткову нейронну мережу

(EfficientNet), трансформерну модель (Vision Transformer) та ансамбль, що поєднує результати двох зазначених моделей з метою підвищення точності діагностики. Основною ідеєю дослідження є порівняння результатів класифікації між трьома підходами та виявлення моделі, яка забезпечує найкращий баланс між точністю, стійкістю до шуму та потенціалом практичного застосування в системах медичної підтримки прийняття рішень.

*Об'єктом* дослідження у даній кваліфікаційній роботі магістра є процес автоматизованої діагностики рентгенівських знімків грудної клітки за допомогою методів Data Science.

*Предметом* дослідження КРМ є моделі, методи та технології побудови та оцінювання моделей автоматизованої медичної діагностики, виявлення закономірностей у медичних знімках та узагальнення знань для покращення процесу клінічного прийняття рішень.

Основні методи дослідження, що використовуються в рамках КРМ, включають:

1. Збір даних — систематичне формування вибірки рентгенівських знімків грудної клітки з відкритих джерел для подальшого аналізу.
2. Попередня обробка даних — покращення якості вхідних зображень шляхом масштабування, нормалізації та усунення зайвих шумів для забезпечення коректної роботи моделей.
3. Глибинний аналіз — застосування моделей машинного навчання для виявлення закономірностей.
4. Візуалізація результатів — узагальнення діагностичних висновків у зрозумілій формі для подальшої інтерпретації, порівняння ефективності моделей через побудову графіків та діаграм.

Наукова новизна роботи полягає у порівняльному аналізі трьох

підходів до автоматизованої класифікації рентгенівських знімків грудної клітки, а саме згорткової нейронної мережі, трансформерної архітектури та їх ансамблю. У рамках дослідження оцінено ефективність кожного з підходів на реальних медичних даних з урахуванням точності класифікації, стійкості до варіативності вхідних даних та здатності до узагальнення. Додаткову увагу приділено інтерпретації результатів за допомогою методів візуалізації важливих ознак, що має практичну цінність у медичних задачах, де необхідна прозорість прийнятих рішень.

Таким чином, результати дослідження можуть бути використані як основа для подальших робіт із покращення точності автоматизованих систем діагностики, а також для підвищення довіри з боку фахівців до результатів таких систем у реальних клінічних сценаріях.

# РОЗДІЛ 1. ОГЛЯД МЕТОДІВ ЗАСТОСУВАННЯ DATA SCIENCE У МЕДИЧНІЙ ДІАГНОСТИЦІ

## 1.1 Аналіз практичної сфери застосування методології Data Science

Методологія Data Science, яка поєднує статистичні методи, машинне навчання, аналіз великих даних та візуалізацію, стала потужним інструментом для розв'язання прикладних задач у різних сферах. Її міждисциплінарний підхід забезпечує цілісне опрацювання даних — від збору до інтерпретації, — що сприяє глибшому розумінню складних процесів та формуванню обґрунтованих прогнозів.

У фінансовому секторі методи Data Science активно застосовуються для оцінки ризиків, виявлення шахрайських операцій, побудови кредитних скорингів і реалізації алгоритмічного трейдингу. Фінансові установи використовують моделі машинного навчання для класифікації клієнтів, прогнозування поведінки та персоналізації продуктів, спираючись на аналіз історичних і транзакційних даних.

У сфері маркетингу та торгівлі аналітика даних дає змогу глибше розуміти споживчу поведінку, розробляти цільові кампанії, прогнозувати попит і здійснювати A/B-тестування для оптимізації рішень. Провідні компанії, зокрема Netflix і Amazon, активно впроваджують рекомендовані системи, що базуються на методах колаборативної фільтрації, кластеризації та аналізу вподобань користувачів.

У промисловості технології Data Science застосовуються для оптимізації виробничих процесів, підвищення ефективності та запобігання технічним збоям. Завдяки використанню сенсорних систем і алгоритмів прогнозного обслуговування підприємства можуть своєчасно виявляти потенційні несправності, мінімізувати простой та зменшувати витрати на ремонт і обслуговування.

У сфері транспорту та логістики аналітика даних дає змогу оптимізувати планування маршрутів, здійснювати моніторинг трафіку, прогнозувати затори та управляти навантаженням на інфраструктуру в реальному часі. Компанії на кшталт Uber і DHL демонструють, як глибока інтеграція аналітичних інструментів стає ключовим чинником підвищення ефективності логістичних процесів.

Не менш активно методологія Data Science розвивається і в сфері охорони здоров'я, де вона демонструє високу ефективність у вирішенні низки прикладних задач. Аналітичні моделі застосовуються для обробки електронних медичних записів, прогнозування перебігу захворювань, виявлення груп ризику, а також для підтримки прийняття клінічних рішень у складних випадках. Завдяки аналізу структурованих і неструктурованих даних, системи на основі штучного інтелекту можуть оцінювати ймовірність ускладнень, оптимізувати протоколи лікування та підвищувати якість медичних послуг.

Окрему роль відіграє Data Science у сфері медичної візуалізації, яка охоплює обробку знімків комп'ютерної томографії (КТ), магнітно-резонансної томографії (МРТ) та рентгенографії. Сучасні моделі комп'ютерного зору дають змогу автоматично виявляти патологічні утворення, локалізувати зони ураження та надавати лікарям додаткову інформацію для підтвердження діагнозу. Впровадження таких технологій значно скорочує час інтерпретації знімків і знижує ймовірність помилок, пов'язаних з людським фактором.

Таким чином, методологія Data Science є універсальним інструментом здобуття медичних знань із даних і має значний потенціал для трансформації клінічної практики. У контексті діагностики, де точність і швидкість є критичними, застосування цих підходів відкриває нові можливості для підвищення об'єктивності та якості медичних рішень.

## 1.2 Вибір та наповнення методології проєкту розробки технології Data Science

Вибір і доповнення методології проєкту розробки технології Data Science має вирішальне значення для забезпечення його успіху. Сформовані кроки, які були враховані під час вибору та доповнення методології для проєкту автоматизованої медичної діагностики за зображеннями:

1. Визначення цілі та обсягу проєкту — перед вибором методології було сформульовано ціль — розробити систему автоматизованої діагностики рентгенівських знімків. Це дозволяє визначити ключові вимоги, обмеження, очікувані результати та критерії ефективності..

2. Вибір відповідної методології — виходячи з цілей і структури проєкту, було обрано методологію CRISP-DM, яка дозволяє ітеративно працювати з даними, будувати моделі та аналізувати їх результати. Її гнучкість та орієнтація на практичне застосування роблять її оптимальною для задач, пов'язаних з обробкою медичних зображень.

3. Адаптація методології до специфіки проєкту — адаптація методології до специфіки проєкту передбачає її узгодження з конкретними завданнями та умовами дослідження. Наприклад, коригування обраного підходу з урахуванням особливостей джерел даних, вибраних аналітичних методів і застосовуваних інструментів.

4. Інтеграція спеціалізованих методів і інструментів — для реалізації проєкту необхідно обрати відповідні алгоритми машинного навчання та інструменти, що забезпечують ефективну обробку, аналіз і візуалізацію даних відповідно до специфіки поставленого завдання.

5. Планування структури реалізації проєкту — розроблення детального плану з чітким розподілом на етапи, визначенням необхідних

ресурсів, часових меж та очікуваних результатів на кожному з кроків. Такий підхід дозволяє забезпечити системний контроль за виконанням проєкту та своєчасно реагувати на можливі виклики в процесі його реалізації.

6. Моніторинг і гнучке коригування методології — у процесі реалізації проєкту важливо здійснювати регулярне відстеження проміжних результатів, що дає змогу своєчасно виявляти потенційні проблеми та за потреби коригувати обрану методологію, змінювати підходи або оновлювати інструментарій відповідно до нових умов або отриманих даних.

Дотримання описаних кроків створює необхідні передумови для успішної реалізації проєктів у сфері Data Science та досягнення визначених цілей.

Для виконання KPM було застосовано методологію інтелектуального аналізу даних CRISP-DM, яка слугувала основою для організації всіх етапів роботи.

Методологія CRISP-DM є одним із найбільш поширених підходів до вирішення задач інтелектуального аналізу даних. Її ітеративна структура забезпечує гнучкість та можливість багаторазового вдосконалення процесу, що наочно ілюструється на рисунку 1.1.

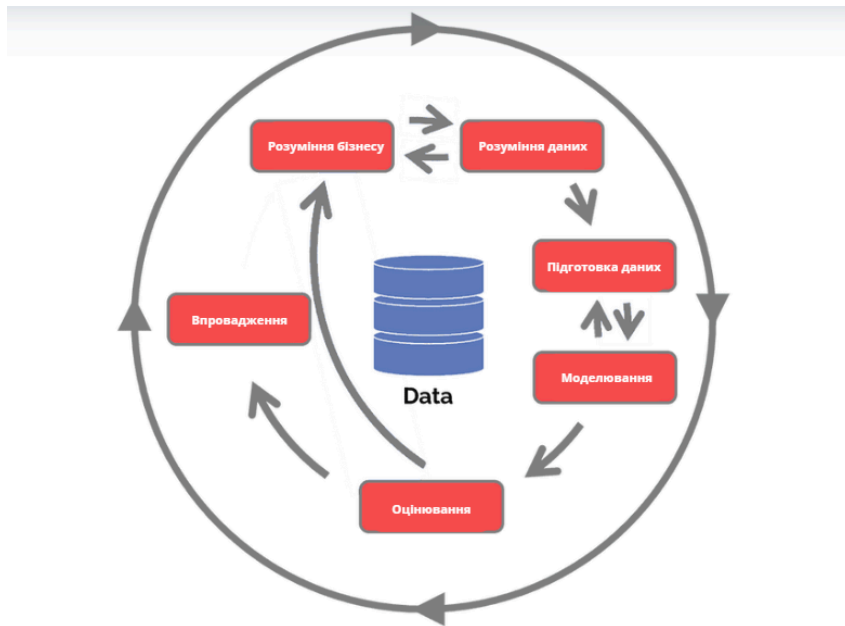


Рисунок 1.1 — Логічна структура процесу CRISP-DM

Основні кроки адаптації методології CRISP-DM для розробки та оцінки технології автоматизованої медичної діагностики за зображеннями:

1. Розуміння проблеми. На цьому етапі формується загальна мета проекту — підвищити точність виявлення патологій органів грудної клітки за допомогою сучасних моделей комп'ютерного зору. Також враховуються потреби потенційних користувачів — лікарів-рентгенологів, які мають отримати інструмент для допоміжного аналізу знімків, з акцентом на інтерпретованість і клінічну застосовність результатів.

2. Розуміння даних. Аналізується структура даних, у тому числі особливості маркування, співвідношення класів, наявність мультикласових прикладів та потенційні проблеми, пов'язані з дисбалансом та неоднорідністю даних.

3. Підготовка даних. Виконуються процедури масштабування зображень, нормалізації пікселів, покращення збалансованості вибірки, а також аугментація. Формуються навчальна, валідаційна та тестова підвибірki з урахуванням рівномірного представлення класів.

4. Моделювання. На цьому етапі розробляються та навчаються три моделі. Визначаються оптимальні параметри навчання, застосовуються техніки регуляризації, обираються функції втрат та оптимізатори.

5. Оцінка. Побудовані моделі піддаються оцінюванню за ключовими метриками. Поряд із кількісною оцінкою здійснюється візуальна інтерпретація рішень моделей. Аналізуються помилки класифікації, зокрема для клінічно критичних класів.

6. Впровадження. На завершальному етапі розглядається можливість інтеграції моделі як допоміжного інструмента для лікаря-рентгенолога. Оцінюються практичні умови застосування в медичних закладах, потенційні технічні обмеження та етичні аспекти. Також обґрунтовується доцільність подальшої зовнішньої валідації на незалежних медичних вибірках.

Застосування структури CRISP-DM у цьому проєкті забезпечує логічну послідовність дослідження, адаптивність до специфіки задач і водночас орієнтацію не лише на технічну реалізацію, а й на практичну значущість отриманого результату.

### **1.3 Дослідження літературних джерел щодо застосування діагностичних моделей**

Аналіз наукової літератури свідчить про активне застосування методів аналізу даних у задачах автоматизованої медичної діагностики. Особливої уваги набули підходи, що використовують глибинне навчання, машинне навчання та комп'ютерний зір для обробки рентгенівських, КТ та МРТ-знімків. У численних дослідженнях вивчається ефективність різних архітектур, які забезпечують класифікацію, виявлення та локалізацію патологій на зображеннях, а також пояснення рішень моделей.

У роботі Wang et al. (2017) [1] представлено один із найбільш використовуваних відкритих наборів ChestX-ray14, що містить понад 100 тисяч рентгенограм із 14 типами патологій. На його основі автори побудували базову CNN-модель для багатокласової класифікації. Дослідження започаткувало хвилю наступних робіт, які прагнули покращити результати шляхом зміни архітектури або підходів до навчання.

Дослідження Rajpurkar et al. (2017) [2] представило модель CheXNet — 121-шарову згорткову нейронну мережу DenseNet, навченої на ChestX-ray14 для виявлення пневмонії. Модель перевищила середню діагностичну точність сертифікованих рентгенологів. Це стало першим прикладом, коли глибока модель досягла рівня лікаря у конкретному завданні.

У роботі Irvin et al. (2019) [3] розроблено новий датасет CheXpert, що забезпечує покращене маркування з урахуванням невизначених класів, а також кращу інтерпретованість. У дослідженні порівнювали кілька моделей, зокрема ResNet і DenseNet, з різними стратегіями обробки невизначених значень. Дослідження також акцентує увагу на потребі пояснення рішень через візуалізацію.

У дослідженні Tang et al. (2020) [4] було розроблено ансамбль моделей ResNet та EfficientNet для класифікації 14 патологій. Поєднання моделей дозволило покращити результати за всіма метриками: AUC досягав 0.90+ для кількох класів. Автори також продемонстрували ефективність багатозадачного навчання.

У роботі [5] досліджено трансформерну модель Swin Transformer для класифікації медичних знімків, зокрема ChestX-ray. Автори зазначають, що моделі на основі Vision Transformer демонструють кращу здатність до

узагальнення, особливо при обробці зображень з високою роздільною здатністю. При цьому важливу роль відіграє обсяг попереднього навчання.

У роботі [6] використано Grad-CAM++ для покращення інтерпретованості результатів. Застосування теплових карт дозволяє візуально відобразити, які ділянки зображення найбільше вплинули на рішення моделі, що підвищує довіру з боку лікарів. Це особливо актуально у випадках, коли результат використовується для клінічних рішень.

Також слід відзначити роботу [7], у якій порівнювали класичні методи машинного навчання — такі як SVM, Random Forest і логістичну регресію — із глибинними моделями CNN на базі одного і того ж датасету. Результати показали суттєву перевагу CNN у складних випадках (наприклад, багатокласова класифікація), але водночас класичні моделі мали вищу інтерпретованість і стабільність на малих вибірках.

У роботі [8] автори досліджують застосування багатомодальних моделей, які об'єднують вхідні зображення та метаінформацію (вік, стать, симптоми). Таке поєднання дозволяє моделі працювати точніше в контекстуальних сценаріях, де зображення саме по собі не є достатнім для постановки висновку.

У статті [9] проаналізовано стратегії аугментації та балансування класів. Було показано, що правильна аугментація (наприклад, обертання, шум, інверсія) суттєво покращує здатність моделі узагальнювати при перенесенні на нові набори даних, особливо для рідкісних патологій.

Таким чином, огляд наукових джерел свідчить, що моделі аналізу даних у сфері медичної діагностики за зображеннями активно еволюціонують — від класичних методів до сучасних глибинних і трансформерних архітектур. Водночас залишається актуальною проблема інтерпретованості, генералізації та стабільності результатів у реальних

клінічних умовах. Застосування ансамблів, візуальних пояснень та мультимодальних входів поступово формує новий стандарт у галузі автоматизованої діагностики.

#### **1.4 Огляд моделей Data Science та існуючих підходів до медичної діагностики**

Сфера автоматизованої медичної діагностики активно розвивається завдяки широкому застосуванню методів комп'ютерного зору, глибинного навчання та великих медичних баз даних. Існує низка спеціалізованих цифрових платформ та інструментів, які дають змогу автоматично аналізувати медичні зображення (зокрема рентген, КТ і МРТ) з метою виявлення патологій, підтримки клінічних рішень або скринінгу великих груп пацієнтів. Такі сервіси можуть інтегруватися у робочі процеси лікарень, підвищуючи ефективність роботи персоналу та знижуючи навантаження на лікарів.

Більшість рішень у цій сфері побудовані на використанні глибоких нейронних мереж — здебільшого згорткових або трансформерних архітектур, здатних ефективно працювати із зображеннями [10]. Вони дозволяють моделювати складні просторові патерни, характерні для медичних знімків, і виявляти патології навіть за відсутності явних ознак. Деякі інструменти також використовують моделі з поясненням рішень, що дає змогу лікареві візуально оцінити, на яких ділянках знімка модель зосередила увагу при класифікації.

До прикладів сучасних інформаційних технологій для діагностики за медичними зображеннями належать:

- Qure.ai — система для автоматичного аналізу рентгенограм грудної клітки та КТ головного мозку. Може виявляти туберкульоз,

пневмонію, COVID-19 та інші патології. Алгоритми Qure.ai сертифіковані для використання в клінічній практиці в багатьох країнах і інтегруються у медичні інформаційні системи [11].

- Aidoc — платформа, яка спеціалізується на ранньому виявленні гострих станів (інсульт, тромбоз, крововиливи) на основі КТ-знімків. Працює у фоновому режимі, опрацьовуючи зображення одразу після їх отримання, і автоматично пріоритезує випадки для лікаря [11].
- Lunit INSIGHT CXR — інструмент для аналізу рентгенограм, здатний розпізнавати понад 10 типів патологій, включно з легeneвими інфільтратами, пневмотораксом і вузловими утвореннями. Має високий рівень точності, підтверджений клінічними дослідженнями [12].
- Infervision — китайська технологія, яка підтримує аналіз знімків грудної клітки та КТ легенів, активно застосовується для скринінгу захворювань дихальних шляхів у великих лікарнях Китаю. Має власну систему навчання моделей на місцевих даних.
- VinBrain DrAid — в'єтнамська розробка, яка підтримує аналіз більше ніж 20 патологій на рентгенівських знімках. Рішення адаптовано до місцевих умов і використовується в державних лікарнях у якості системи підтримки прийняття рішень [12].

Ці системи об'єднують наявність великої кількості попередньо навченої інформації (медичних знімків із маркуванням) та використання глибокого навчання для точного розпізнавання патологій. У деяких випадках сервіси також мають модулі інтеграції з PACS-системами, медичними реєстрами чи телемедициною.

Водночас існують і обмеження. Деякі системи обмежені вузькою спеціалізацією (наприклад, працюють лише з COVID-19 або лише з КТ голови), інші — орієнтовані на конкретний ринок (США, Китай, В'єтнам),

що ускладнює перенесення моделей на нові медичні системи. Також для більшості глибоких нейронних мереж процес прийняття рішень залишається «чорним ящиком». Це може знижувати довіру з боку медичного персоналу та ускладнювати клінічну відповідальність за рішення [13].

Ще однією проблемою є вартість. Більшість високоточних платформ працюють за комерційною моделлю й потребують щомісячної підписки або оплати за кожен аналізований знімок. Це може бути критичним бар'єром для лікарень у країнах із обмеженим фінансуванням або для науково-дослідницьких груп без доступу до відкритих ліцензій [14].

Таким чином, аналіз існуючих інформаційних технологій у сфері автоматизованої медичної діагностики демонструє значний прогрес у розвитку інструментів, здатних підтримувати лікарів у прийнятті рішень. Водночас актуальними залишаються виклики, пов'язані з адаптацією моделей до локальних умов, забезпеченням прозорості рішень та зменшенням вартості їх використання.

## **1.5 Постановка задачі**

Аналіз сучасних підходів до автоматизованої медичної діагностики показує, що глибокі моделі комп'ютерного зору демонструють високу ефективність у виявленні патологій за зображеннями. Проте, попри значний прогрес у цій галузі, низка викликів залишається невирішеною — зокрема, проблема генералізації моделей, їх інтерпретованості та стійкості до різноманітності даних. Ці фактори є критично важливими у клінічному контексті, де точність та надійність діагностичного рішення мають безпосередній вплив на здоров'я пацієнта.

Серед перспективних напрямів виділяються ансамблеві методи, які об'єднують сильні сторони різних моделей, зокрема згорткових нейронних мереж (CNN) та трансформерних архітектур (ViT), і дозволяють покращити точність прогнозування. Питання комбінування моделей у медичній діагностиці наразі досліджене недостатньо, однак його потенціал очевидний з огляду на успіхи аналогічних підходів у суміжних задачах.

*Об'єктом* дослідження є процес автоматизованої діагностики медичних зображень органів грудної клітки.

*Предметом* дослідження є моделі, методи та технології Data Science для побудови і оцінки систем автоматичної класифікації патологій на рентгенівських знімках.

Згідно з аналізом літературних джерел, найкращі результати у задачах діагностики за зображеннями демонструють такі підходи, як EfficientNet, Vision Transformer (ViT) та ансамблеві моделі, що поєднують кілька архітектур. Саме ці три підходи будуть розглянуті в роботі. Також буде досліджено, чи дозволяє об'єднання згорткової та трансформерної моделей підвищити точність діагностики порівняно з окремими компонентами.

*Метою* дослідження є підвищення ефективності автоматизованої діагностики патологій грудної клітки за рентгенівськими знімками шляхом порівняльного аналізу різних підходів і побудови ансамблевої моделі на їх основі.

Для досягнення цієї мети необхідно виконати такі завдання:

- Провести огляд сучасних моделей та методів глибокого навчання для автоматизованої медичної діагностики знімків;

- Сформувати підготовлений датасет на основі ChestX-ray14 для навчання, валідації та тестування моделей;
- Провести попередню обробку даних, включаючи аугментацію, балансування класів та нормалізацію;
- Побудувати три моделі: на основі EfficientNet, Swin Transformer та їх ансамбль;
- Провести оцінювання продуктивності моделей за допомогою метрик точності, AUC, повноти, F1-міри та Grad-CAM;
- Проаналізувати отримані результати та визначити доцільність використання ансамблевого підходу в медичних задачах;
- Розробити концепцію впровадження моделі як допоміжного інструмента підтримки прийняття рішень у клінічній практиці.

## **Висновки**

Отже, було проведено аналіз практичного застосування методології Data Science у сфері автоматизованої медичної діагностики на основі зображень. Результати дослідження підтверджують, що ця тематика є надзвичайно актуальною, особливо з огляду на зростаюче навантаження на медичну систему, потребу у швидких рішеннях та важливість точності у клінічних умовах.

У якості методологічної основи проєкту було обрано підхід CRISP-DM, який забезпечує структурованість і повторюваність дослідницького процесу. Для кожного етапу методології сформульовано задачі з урахуванням специфіки роботи з медичними знімками.

Ретельний огляд наукових джерел дозволив оцінити поточний стан досліджень у галузі автоматизованої діагностики, включаючи як класичні алгоритми машинного навчання, так і сучасні глибинні архітектури. Було виявлено, що найбільш перспективними підходами є EfficientNet, Vision Transformer та ансамблеві моделі, які поєднують переваги кількох архітектур і забезпечують підвищену точність.

Також було проаналізовано існуючі інформаційні технології, що реалізують автоматизовану діагностику. Розглядалися їх функціональні можливості, рівень точності, сфери застосування, а також наявні обмеження — зокрема, брак універсальності, потреба в дорогому ліцензуванні та складнощі з перенесенням моделей у нові клінічні умови.

Одержані результати та виявлені проблемні аспекти дозволили сформуванати дослідницьку мету, окреслити ключові моделі для реалізації та визначити напрямки подальшого вдосконалення. Запропонований підхід до побудови й оцінки моделей спрямований на підвищення точності, пояснюваності та адаптивності технологій у сфері автоматизованої медичної діагностики.

## РОЗДІЛ 2

### ФОРМАЛІЗАЦІЯ МЕТОДІВ АНАЛІЗУ ДАНИХ У ПРОЄКТІ МЕДИЧНОЇ ДІАГНОСТИКИ

#### 2.1 Попередній аналіз та обробка медичних зображень

У проєктах, що передбачають побудову моделей глибинного навчання для діагностики за медичними зображеннями, особливу увагу необхідно приділяти етапу попереднього аналізу та обробки даних. Зображення, що надходять на вхід моделей, мають бути уніфіковані за розміром, якістю та форматом. Крім того, важливою є очистка і перевірка відповідних метаданих, які можуть містити мітки або супутню інформацію про пацієнта чи діагноз [9].

Очищення даних — це процес перетворення «сирих» вхідних даних у форму, з якою можливо безпечно і ефективно працювати в рамках машинного навчання. З точки зору проєкту, чистими вважаються ті дані, які:

- не містять порожніх або некоректних значень у діагностичних мітках;
- не мають дублікатів або повторюваних зображень;
- представлені у стандартизованому форматі (розмір, кількість каналів, глибина зображення);
- супроводжуються однозначними, уніфікованими діагнозами [15].

На початковому етапі здійснюється фільтрація записів, у яких відсутнє значення цільової змінної, а також попереднє уніфікування структури міток. Наприклад, якщо зображення мають кілька патологій одночасно, визначається головна мітка або здійснюється переведення задачі у багатокласову форму.

З урахуванням того, що джерелом даних є «живі» медичні системи або відкриті репозитарії, існує ризик людського фактора при введенні інформації: можуть бути неточності у діагнозах, різні формати позначень, втрата частини файлів тощо. Тому обробка включає перевірку на наявність некоректних або відсутніх файлів зображень.

Усі вхідні зображення перед подачею до моделі проходять стандартизовані трансформації, серед яких:

- масштабування до фіксованого розміру (224×224 пікселі),
- перетворення в один канал (градації сірого),
- нормалізація інтенсивностей пікселів у межах  $[-1, 1]$ .

Однією з важливих складових попередньої обробки є аугментація зображень — це сукупність методів штучного розширення навчальної вибірки шляхом створення нових варіацій зображень на основі наявних прикладів. Аугментація допомагає моделі краще узагальнювати, робить її менш чутливою до шумів у вхідних даних, а також зменшує ризик перенавчання [16].

До поширених видів аугментації належать:

- випадкові повороти зображень (наприклад, до  $\pm 10$  градусів), що імітує відхилення при зйомці;
- горизонтальне віддзеркалення, яке дозволяє зберігати медичну коректність (для рентгенограм грудної клітки — допустимо);
- масштабування, обрізання або зсув, що допомагають моделі навчитися розпізнавати об'єкти незалежно від їхнього точного положення;

- нормалізація пікселів — приводить значення яскравості до уніфікованого масштабу.

Аугментація застосовується лише до навчальної вибірки, оскільки її основне завдання — підвищити різноманітність вхідних прикладів і змусити модель стійко реагувати на зміни візуального представлення патологій. Для валідаційної та тестової вибірок використовуються лише базові трансформації (зміна розміру, нормалізація), щоб забезпечити об'єктивне оцінювання моделі без штучних змін зображення.

Для забезпечення коректного представлення цільових змінних здійснюється кодування кожної діагностичної мітки у числовий формат. Наприклад, кожному діагнозу присвоюється унікальний індекс, що дозволяє використовувати його як цільову ознаку для задачі класифікації.

На завершення етапу виконується розподіл даних на навчальну, валідаційну та тестову вибірки з урахуванням пропорцій класів. Це забезпечує збереження балансу патологій на всіх підвибірках і дозволяє об'єктивно оцінювати якість моделей.

Таким чином, попередня обробка медичних зображень — це критично важливий етап, який визначає надійність, точність і узагальнювальну здатність моделі. Від якості реалізації цього етапу значною мірою залежить успішність усього проекту автоматизованої діагностики.

## **2.2 Математичний опис методів побудови моделей для медичної діагностики**

З метою автоматизованої класифікації рентгенівських знімків грудної клітки у межах цього дослідження було обрано та реалізовано три архітектурні підходи: глибоку згорткову нейронну мережу EfficientNet

(типу V0), модель Vision Transformer (ViT), а також ансамблеву модель на основі об'єднання попередніх архітектур.

Вибір цих підходів ґрунтується на їхній підтвердженій ефективності в завданнях класифікації медичних зображень, а також на здатності моделювати як локальні, так і глобальні просторові залежності, що є критично важливим у контексті багатокласової діагностики патологій.

EfficientNet — це сучасна архітектура згорткових нейронних мереж, яка базується на однорідному масштабуванні глибини, ширини та роздільності. У традиційних підходах зазвичай масштабують лише один із цих аспектів, що може призводити до неефективного використання обчислювальних ресурсів. Натомість, EfficientNet використовує спеціальну формулу масштабування, яка дозволяє досягати значно кращих результатів при меншій кількості параметрів і FLOPs (кількості обчислень) [19].

Ключовим елементом архітектури є MBConv-блоки (Mobile Inverted Bottleneck Convolution). Кожен такий блок включає:

- операцію розширення (expansion), яка збільшує кількість каналів перед згорткою;
- глибоку згортку (depthwise convolution) — економний варіант згортки, що застосовується до кожного каналу окремо;
- точкову згортку (pointwise convolution) — згортка з ядром  $1 \times 1$  для об'єднання інформації між каналами;
- нормалізацію (Batch Normalization) для стабілізації тренування;
- механізм уваги squeeze-and-excitation, який дозволяє мережі зосереджувати увагу на більш інформативних каналах зображення.

Математично, для кожного згорткового шару результат  $Y$  обчислюється як:

$$Y = f(X*W + b), [1]$$

де  $X \in R^{H \times W \times C}$  — вхідне зображення;

$W$  — ядро згортки;

$b$  — зсув;

$*$  — операція згортки;

$f(\cdot)$  — функція активації Swish:  $f(x) = x \cdot \sigma(x)$ .

EfficientNet-B0 обрано як базову модель, яка досягає хорошого балансу між продуктивністю та обчислювальними витратами. Завершується модель глобальним середнім пулінгом і повнозв'язним softmax-шаром.

Архітектура EfficientNet-B0 представлена на рисунку 2.1.

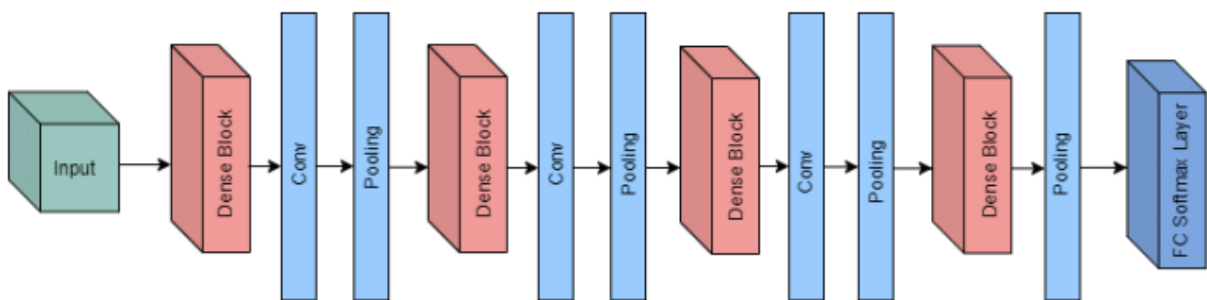


Рисунок 2.1 — Архітектура EfficientNet-B0

Vision Transformer переносить архітектуру трансформерів, відому з NLP, у сферу обробки зображень. Зображення розбивається на патчі фіксованого розміру, кожен з яких лінійно проектується у вектор і подається як токен до трансформера [21].

$$z_i = W_e x_p + p_i, [2]$$

де  $x \in R^{P^2 \cdot C}$  — розгорнутий патч зображення;

$W_e$  — матриця проєкції;

$p_i$  — позиційне кодування.

Трансформер складається з послідовності шарів, де основним компонентом є механізм багатоголової самоуваги (Multi-Head Self-Attention):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V, [3]$$

де  $Q, K, V$  — запити, ключі та значення, що отримуються через лінійні перетворення.

CLS-токен, доданий на початку послідовності, використовується як представлення всього зображення, яке після проходження через трансформер передається у класифікатор.

Схема Vision Transformer представлена на рисунку 2.2.

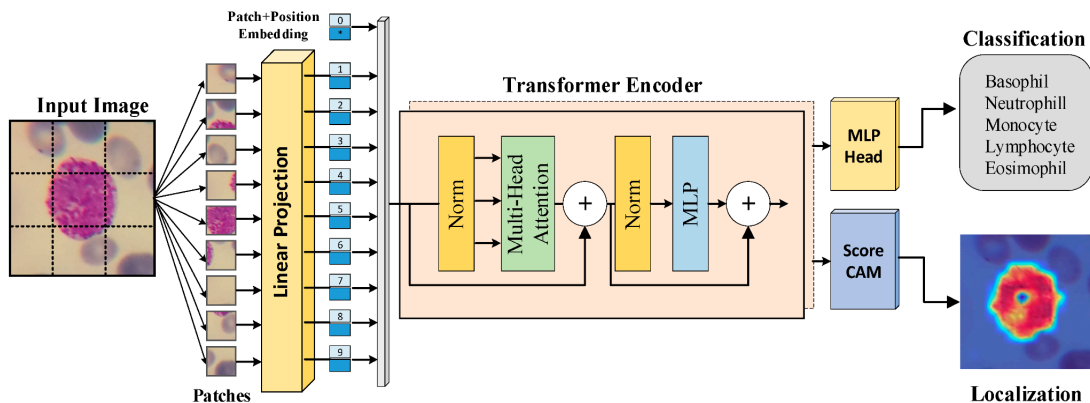


Рисунок 2.2 — Схема Vision Transformer

Для покращення стабільності класифікації та зниження помилок обрано ансамблевий підхід. Об'єднання моделей CNN та ViT реалізується за допомогою soft voting — усереднення ймовірнісних виходів кожної моделі [23]:

$$\hat{y} = \arg \max(\alpha \cdot p^{Eff} + \beta \cdot p^{ViT}), [4]$$

де  $p^{Eff}, p^{ViT}$  — ймовірності класу від моделей EfficientNet і ViT відповідно;

$\alpha, \beta \in [0, 1]$  — вагові коефіцієнти (у даному проєкті використано  $\alpha = \beta = 0.5$ ).

Ансамблювання дозволяє поєднати локальну чутливість CNN до текстур і країв з глобальною контекстуальністю трансформера, що суттєво покращує точність при виявленні складних патологій на медичних знімках [24].

### 2.3 Метрики оцінювання ефективності моделей діагностики

Для об'єктивного оцінювання точності класифікаційних моделей у задачах медичної діагностики важливо використовувати комплекс показників, які дозволяють оцінити як загальну якість моделі, так і її здатність виявляти окремі класи патологій. У КРМ для аналізу продуктивності моделей використано такі метрики: точність (accuracy), повнота (recall), F1-міра, AUC ROC (macro та micro), а також Grad-CAM++ для візуалізації та інтерпретованості результатів.

Accuracy — це базова метрика, яка визначає частку правильно класифікованих зразків серед усіх прикладів. Вона обчислюється як:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, [5]$$

де  $TP$  — кількість істинно позитивних класифікацій,

$TN$  — істинно негативних,

$FP$  — хибно позитивних,

$FN$  — хибно негативних.

У багатокласовій класифікації використовується усереднене значення асигуру по всіх класах.

Recall характеризує здатність моделі виявляти усі позитивні зразки певного класу. Це особливо важливо у медичних задачах, де пропущене захворювання може мати критичні наслідки:

$$Recall = \frac{TP}{TP + FN} [6]$$

Precision показує, яка частка об'єктів, передбачених як позитивні, дійсно є позитивними:

$$Precision = \frac{TP}{TP + FP} [7]$$

F1-score — це гармонічне середнє між precision і recall. Дає збалансовану оцінку у випадках, коли важлива як чутливість, так і точність:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} [8]$$

F1 є особливо корисною метрикою у випадках з незбалансованими класами — типовою ситуацією у медичній діагностиці, де одні патології представлені частіше за інші.

AUC оцінює якість класифікації незалежно від порогу. ROC-крива відображає співвідношення між чутливістю (Recall) і хибнопозитивною частотою (False Positive Rate):

$$FPR = \frac{FP}{FP + TN} [9]$$

Micro-AUC оцінює загальну продуктивність моделі по всіх зразках, тоді як Macro-AUC обчислює середнє значення AUC для кожного класу окремо, незалежно від їх частоти. Значення AUC = 1 відповідає ідеальному класифікатору, AUC = 0.5 — випадковому вгадуванню.

Grad-CAM++ — це метод візуалізації, який дозволяє інтерпретувати рішення згорткових моделей. Він формує теплову карту, яка показує, на які ділянки зображення модель звертала увагу під час прийняття рішення. Це критично важливо у медичній діагностиці для перевірки того, чи справді модель «бачить» патологію, а не навчається на шумі або артефактах. Теплова карта Grad-CAM++ для інтерпретації класифікації представлена на рисунку 2.3.

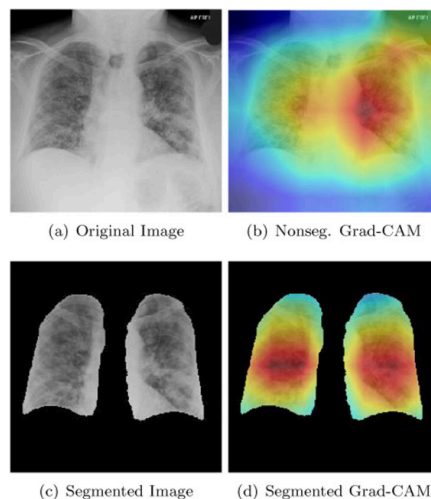


Рисунок 2.3 — Теплова карта Grad-CAM++ для інтерпретації класифікації

## 2.4 Переваги та недоліки обраних методів

У межах дослідження розглядаються три підходи до класифікації рентгенівських знімків грудної клітки: згортова нейронна мережа EfficientNet, трансформер Vision Transformer (ViT) та їхній ансамбль. Кожен із цих методів має свої сильні сторони та обмеження, які варто враховувати при побудові автоматизованої системи медичної діагностики.

EfficientNet є оптимізованою згортковою нейронною мережею, що поєднує високу точність з відносно невеликою кількістю параметрів завдяки збалансованому масштабуванню глибини, ширини та роздільності. Архітектура цієї моделі дозволяє ефективно працювати із зображеннями навіть у випадках обмеженого обсягу даних, що особливо актуально у сфері медичних знімків. До її переваг належать здатність до виявлення локальних патологічних змін, стійкість до шуму, підтримка перенавчання на основі попередньо натренованих ваг, а також швидка інференція. Разом з тим, ця архітектура має певні обмеження — зокрема, обмежену здатність до захоплення глобального контексту зображення, що в деяких випадках ускладнює правильне виявлення дифузних патологій або комбінованих аномалій.

Vision Transformer (ViT) реалізує інший підхід, запозичений із природної мовної обробки, де зображення розбивається на послідовність патчів, що подаються на вхід трансформеру. Завдяки механізму багатоголової самоуваги модель здатна виявляти глобальні залежності між різними регіонами зображення, що є суттєвою перевагою у медичній візуалізації, особливо при виявленні дифузних або структурно розподілених патологій. ViT також забезпечує більш інтерпретовану поведінку завдяки візуалізації attention-карт, що дозволяє лікарю зрозуміти, на які області зображення звертала увагу модель. Водночас, трансформери

мають вищі обчислювальні вимоги, особливо на етапі тренування, і демонструють гірші результати на невеликих датасетах без попереднього навчання, що робить їх менш стійкими у випадку обмежених ресурсів або даних.

Ансамблевий підхід, який поєднує переваги обох моделей, дозволяє компенсувати їхні недоліки. У даному дослідженні було реалізовано `soft voting` — усереднення ймовірностей від `EfficientNet` та `ViT`. Такий підхід дозволяє одночасно врахувати як локальні, так і глобальні особливості зображення, що позитивно впливає на загальну точність класифікації. До його переваг належать підвищена стабільність, зниження варіацій між передбаченнями та зменшення ризику пропущених патологій. Основними недоліками ансамблю є підвищене навантаження на обчислювальні ресурси та збільшена складність реалізації. Крім того, такі комбіновані архітектури складніше інтерпретувати з точки зору клінічного використання, що вимагає окремої роботи над пояснюваністю.

У підсумку, всі три підходи мають практичне застосування у задачах медичної діагностики. Використання ансамлевої моделі дозволяє досягти кращого балансу між точністю, інтерпретованістю та обчислювальною ефективністю, що є важливою передумовою для реального впровадження таких систем у клінічну практику.

## **2.5 Мова програмування, основні бібліотеки та інструменти**

Для реалізації дослідження, спрямованого на автоматизовану діагностику рентгенівських знімків грудної клітки, було обрано мову програмування `Python`. Цей вибір зумовлений її універсальністю, широкою підтримкою в науковій спільноті, активною екосистемою та наявністю потужних бібліотек для глибокого навчання та обробки медичних зображень. `Python` підтримує функціональний, процедурний та

об'єктно-орієнтований стилі програмування, що робить його надзвичайно гнучким для задач Data Science.

Перевагою мови є її гнучка інтеграція з фреймворками для машинного навчання, широке ком'юніті та постійне оновлення бібліотек, зокрема у сфері глибинного навчання. У цьому проєкті використовувались такі ключові інструменти. TensorFlow та Keras були застосовані для побудови, тренування та оцінки згорткових нейронних мереж, зокрема моделі EfficientNet. Keras надає простий інтерфейс для визначення архітектур, тоді як TensorFlow відповідає за обчислювальну ефективність і підтримує запуск моделей як на центральному, так і на графічному процесорах.

Для реалізації трансформерної архітектури Vision Transformer було використано PyTorch у поєднанні з бібліотекою timm. PyTorch забезпечує зручну роботу з динамічними обчислювальними графами, що особливо корисно для досліджень та гнучкої побудови моделей, а бібліотека timm містить великий набір попередньо навчених моделей, адаптованих до задач комп'ютерного зору, включаючи модифіковані варіації ViT. Для реалізації ансамблевого підходу та оцінки ефективності моделей було використано бібліотеку Scikit-learn. Вона дозволила здійснити розділення даних на навчальні й тестові множини, провести крос-валідацію, а також протестувати комбіновані моделі.

Обробка числових масивів здійснювалась засобами бібліотеки NumPy, яка використовувалась для маніпуляцій із тензорами, створення матриць ознак, нормалізації та інших базових обчислювальних операцій. Робота з табличними метаданими, злиття зображень із мітками, створення фільтрів та підмножин даних здійснювалась через бібліотеку Pandas, яка забезпечує зручний доступ до даних у форматі DataFrame. Для візуалізації

результатів, побудови графіків розподілу, зміни функції втрат та порівняння ефективності моделей було використано бібліотеки Matplotlib та Seaborn.

Попередня обробка зображень, включаючи зміну розміру, обрізання, перетворення у відтінки сірого, інверсію, зсув і масштабування, виконувалась за допомогою бібліотек OpenCV та Pillow (PIL). Для підвищення різноманітності тренувальних даних була застосована бібліотека Albumentations, яка забезпечує ефективну реалізацію різних технік аугментації, зокрема випадкове обертання, шум, зміну контрастності та інші трансформації, що особливо корисно у випадку обмежених медичних датасетів. Для візуалізації уваги моделі було використано бібліотеку TorchCAM, яка дозволила згенерувати Grad-CAM-мапи активацій та отримати інтерпретовані результати, візуально демонструючи, на які ділянки знімка модель орієнтувалась під час прийняття рішення.

У якості середовища для реалізації коду використовувалося хмарне середовище Google Colab, яке забезпечило запуск обчислень на GPU та можливість оперативного тестування різних конфігурацій без потреби у локальному обладнанні високої продуктивності. Додатково для інтерактивного аналізу, роботи з графіками та нотатками застосовувалось середовище Jupyter Notebook.

Таким чином, обрані інструменти повністю охоплюють усі етапи проєкту: від завантаження й попередньої обробки зображень до побудови, оцінки та інтерпретації моделей машинного навчання. Їх використання є загальновизнаним стандартом у галузі медичної аналітики та глибинного навчання, що робить отримані результати придатними для масштабування та інтеграції в практичні клінічні рішення.

## Висновки

У цьому розділі було здійснено формалізацію методів аналізу даних, що використовуються для автоматизованої діагностики патологій на рентгенівських знімках. Наведено математичні основи побудови моделей глибокого навчання, зокрема згорткових нейронних мереж (EfficientNet), трансформерних архітектур (Vision Transformer) та ансамблевих підходів, що поєднують їхні виходи. Для кожної моделі описано принцип роботи, ключові формули та переваги застосування в контексті задач медичної візуалізації.

Окрему увагу приділено етапам попередньої обробки даних: очищенню, нормалізації, кодуванню міток, а також технікам аугментації, які були реалізовані з урахуванням специфіки медичних знімків.

Також обґрунтовано вибір мови програмування Python для реалізації моделей, представлено огляд бібліотек, що використовувались у роботі. Усі інструменти забезпечили повноцінний цикл: від обробки вхідних зображень до візуалізації і порівняння результатів моделей.

Підсумовуючи, можна стверджувати, що поєднання сучасних глибоких архітектур із комплексною підготовкою даних і коректним вибором інструментарію дозволяє ефективно реалізовувати задачі класифікації медичних зображень та формує основу для побудови прикладних діагностичних систем у сфері охорони здоров'я.

## РОЗДІЛ 3

### РОЗРОБКА МЕТОДУ DATA SCIENCE

#### 3.1 Аналіз даних для моделі

У рамках дослідження було обрано один із найвідоміших публічних наборів медичних зображень — ChestX-ray14, який є загальновизнаним еталоном у завданнях автоматизованої діагностики рентгенограм грудної клітки. Даний набір був оприлюднений Національним інститутом здоров'я США (NIH) та містить рентгенівські знімки понад 30 000 пацієнтів. Зображення представлені у форматі PNG та мають супровідні метадані, що включають інформацію про пацієнтів, параметри дослідження та патології, виявлені під час діагностики.

Даний набір даних має низку переваг, серед яких:

- великий обсяг зображень, що підвищує надійність моделей;
- наявність якісної структури метаданих для проведення супутнього аналізу;
- відповідність реальним клінічним умовам, що робить результати дослідження більш прикладними.

У таблиці 3.1 наведено опис основних полів набору даних:

Таблиця 3.1 – Опис полів набору даних ChestX-ray14.

№	Поле	Опис
1	Image Index	Назва зображення у форматі PNG
2	Finding Labels	Перелік патологій, виявлених на знімку (одна або кілька)
3	Patient ID	Унікальний ідентифікатор пацієнта
4	Patient Age	Вік пацієнта на момент обстеження
5	Patient Gender	Стать пацієнта (М або F)

6	View Position	Положення пацієнта при знімкуванні (наприклад, PA або AP)
7	Original Image Size	Роздільна здатність зображення
8	Acquisition Date	Дата отримання зображення

Дані покривають широкий спектр захворювань: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, Hernia. Кожен знімок може мати кілька діагностичних міток.

На рисунку 3.1 представлено розподіл кількості знімків у датасеті ChestX-ray14 залежно від класу патології, що дозволяє візуально оцінити ступінь дисбалансу між класами.

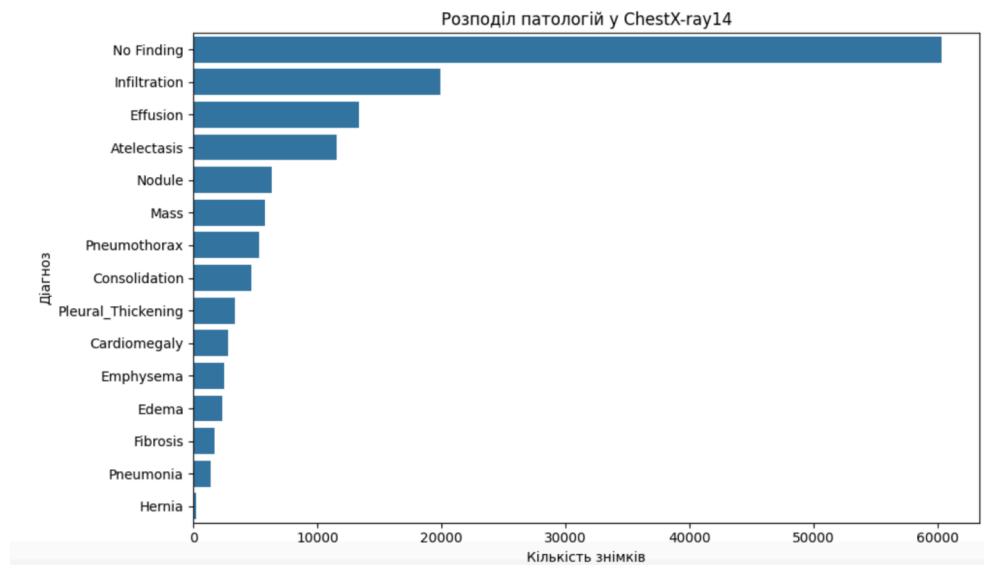


Рисунок 3.1 — Розподіл зображень за класами патологій

Крім аналізу частоти патологій, доцільним є розгляд демографічного профілю пацієнтів. Зокрема, важливо враховувати вік і стать, оскільки ці характеристики можуть суттєво впливати на ймовірність виявлення певних захворювань. На рисунку 3.2 зображено розподіл віку пацієнтів, який

свідчить про переважну концентрацію обстежених у віковому діапазоні від 40 до 60 років.

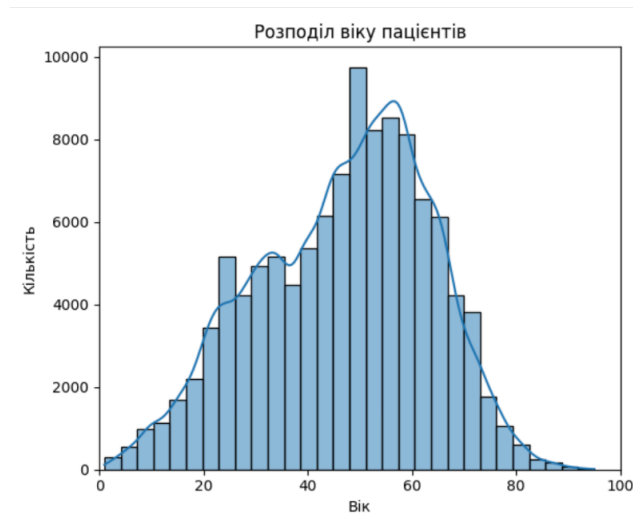


Рисунок 3.2 - Гістограма розподілу віку пацієнтів

Для повнішого розуміння структури вибірки також проаналізовано стать пацієнтів. Цей показник є важливим у контексті медичних досліджень, оскільки окремі патології можуть мати різну поширеність серед чоловіків та жінок. На рисунку 3.3 зображено відповідний розподіл.

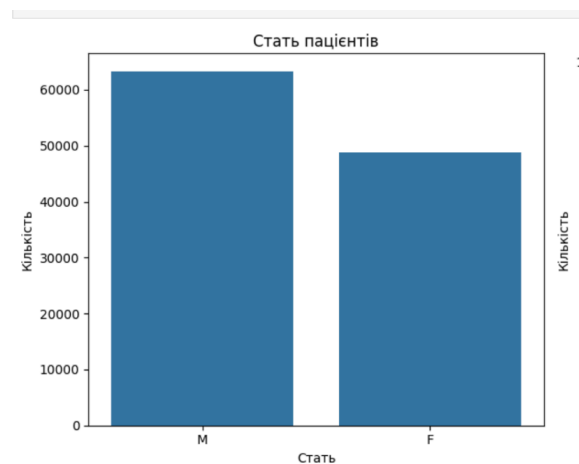


Рисунок 3.3. Розподіл пацієнтів за статтю

Виявлено значну нерівномірність у представленості різних патологій, що зумовлює необхідність у відповідних стратегіях балансування під час побудови моделей.

У таблиці 3.2 наведено опис частоти класів у датасеті ChestX-ray14.

Таблиця 3.2 – Частотність патологій у датасеті ChestX-ray14

№	Клас патології	Кількість зображень
1	No Finding	60412
2	Infiltration	19894
3	Effusion	13317
4	Atelectasis	11435
5	Nodule	6332
6	Mass	5371
7	Consolidation	4667
8	Pleural Thickening	3385
9	Cardiomegaly	2777
10	Emphysema	2516
11	Edema	2303
12	Pneumonia	1431
13	Fibrosis	727
14	Hernia	227

Для зменшення впливу дисбалансу класів, який характерний для набору даних ChestX-ray14, було застосовано кілька стратегій. По-перше, використано вагове коригування функції втрат: рідкісним класам автоматично надається більша вага під час навчання, що дозволяє моделі ефективніше вчитись розпізнавати менш представлені патології. По-друге, реалізовано oversampling — штучне збільшення кількості зразків для

малочисельних класів у навчальній вибірці. Крім того, для ще кращої адаптації до нерівномірного розподілу, було застосовано функцію втрат Focal Loss, яка зосереджує увагу моделі на складних або погано класифікованих прикладах.

У рамках часткового балансування, для найпоширеніших класів було обмежено кількість зображень на рівні 3000, тоді як для рідкісних класів залишено всі доступні знімки. Такий підхід дозволив вирівняти співвідношення між групами без втрати цінної інформації з рідкісних випадків.

На рисунку 3.3 представлено оновлений розподіл патологій після часткового балансування навчального набору.

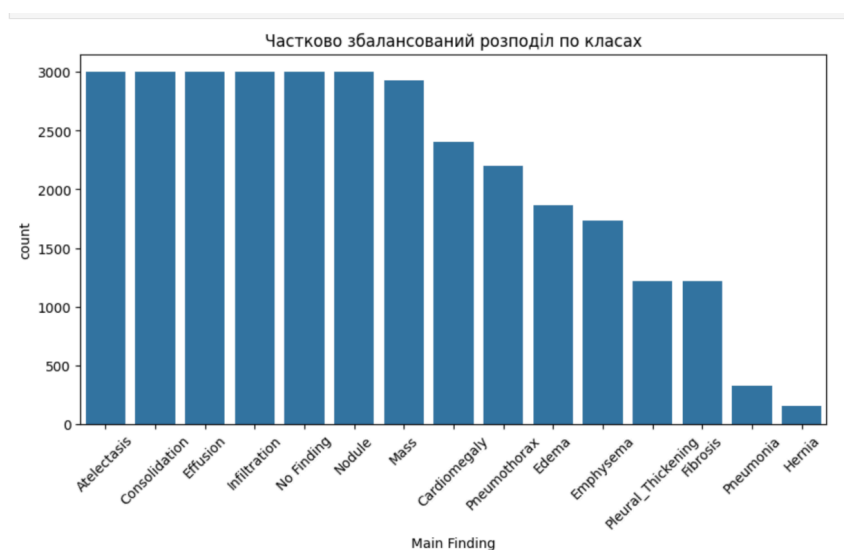


Рисунок 3.3 — Збалансований розподіл зображень за класами патологій

Перед тренуванням моделі було здійснено фільтрацію даних. Із загального набору видалено всі записи, які:

- не містили міток патологій;
- мали пошкоджені або недоступні файли зображень;

- містили технічно непридатні знімки (надто малі зображення або порожні зображення).

Це дозволило уникнути включення шумових або неінформативних прикладів, які могли б завадити навчанню моделей.

Далі проведено стандартизацію розміру зображень. Оскільки більшість сучасних архітектур нейронних мереж, зокрема EfficientNetV0 та Vision Transformer (ViT), приймають на вхід зображення фіксованого розміру, усі рентгенівські знімки були масштабовані до розміру  $224 \times 224$  пікселів. Це є стандартною практикою при роботі з медичними зображеннями у форматі PNG.

Наступним кроком було нормалізування піксельних значень:

- Для EfficientNet значення пікселів було приведено до діапазону  $[0, 1]$  шляхом ділення на 255.
- У випадку трансформерної архітектури масштабування виконувалося до діапазону  $[-1, 1]$ , відповідно до вимог стандартної токенизації та нормалізації.

Для підвищення узагальнювальної здатності моделі та зниження ризику перенавчання було реалізовано етап аугментації даних, який застосовувався виключно до навчальної вибірки. Це дозволило штучно розширити обсяг доступних даних за рахунок варіативності візуального представлення одного і того ж знімка, що особливо важливо в умовах обмеженого або дисбалансного датасету.

Аугментація допомагає моделі краще навчатися на реальних варіаціях зображень, які можуть виникати через:

- різне положення пацієнта під час знімкування;

- відмінності в обладнанні та параметрах знімків;
- наявність шуму або артефактів.

До реалізованих технік аугментації входили такі трансформації:

- Горизонтальне віддзеркалення (flip) — використовується як базова форма симетричної трансформації, що дозволяє подвоїти кількість зображень без втрати їхньої медичної релевантності.
- Обертання в межах  $\pm 10^\circ$  — допомагає моделі бути стійкою до незначних відхилень у положенні пацієнта під час знімкування.
- Зміна контрасту та яскравості — моделює відмінності в апаратному забезпеченні або умовах сканування.
- Кропінг (випадкове обрізання країв) та масштабування — дозволяє моделі фокусуватися на різних частинах зображення, не втрачаючи контексту.

Усі ці операції виконувались рандомізовано в процесі підготовки, що забезпечило максимальне різноманіття під час навчання.

На рисунку 3.4 представлено приклади оригінальних та аугментованих знімків, що демонструють варіації, застосовані до зображень в рамках навчальної вибірки.

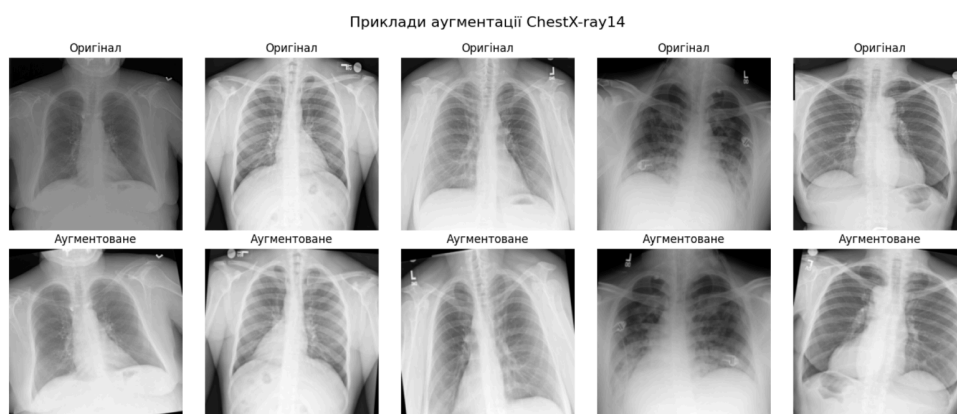


Рисунок 3.4 — Приклади аугментованих знімків

Для підготовки цільової змінної до використання було застосовано one-hot кодування (one-hot encoding). Це означає, що кожна патологія була представлена у вигляді бінарного вектору довжини 14 (за кількістю класів), де лише один елемент має значення 1 (що вказує на клас знімка), а інші — 0. Такий формат зручний для багатокласової класифікації і є стандартом у більшості deep learning фреймворків.

З метою забезпечення надійної оцінки продуктивності моделей та запобігання data leakage було здійснено стратифіковане розбиття датасету. Усі дані були поділені на три частини:

- 80% — навчальна вибірка
- 10% — валідаційна вибірка
- 10% — тестова вибірка

Розподіл було виконано таким чином, щоб кожен клас патологій був представлений в кожній частині датасету у пропорціях, що відповідають їх частотам, тобто з урахуванням стратифікації.

### **3.2 Побудова окремих моделей**

Для реалізації задачі автоматичної класифікації патологій на рентгенівських знімках грудної клітки було побудовано три моделі на основі різних архітектур: згорткової нейронної мережі EfficientNetB0, трансформерної архітектури Vision Transformer (ViT), а також ансамблевої моделі, що поєднує результати попередніх двох підходів.

#### *Побудова моделі EfficientNetB0*

Перша модель була побудована на базі згорткової архітектури EfficientNetB0, яка відома хорошим співвідношенням між якістю класифікації та обчислювальними витратами. Модель ініціалізувалась

попередньо натренованими вагами з ImageNet, після чого була донавчена на медичних знімках.

Вхідний шар було адаптовано під формат зображень ( $224 \times 224 \times 3$ ), після чого було додано шари глобального середнього пулінгу, повнозв'язні шари з функцією активації ReLU, Dropout-шар для регуляризації та вихідний Softmax-шар на 14 класів. Для навчання використовувалась функція втрат categorical\_crossentropy та оптимізатор Adam.

Навчання моделі тривало до 30 епох з використанням механізму ранньої зупинки (early stopping), щоб уникнути перенавчання. В результаті модель продемонструвала хороші результати на основних класах, проте менш точна була на рідкісних патологіях.

### *Побудова моделі Vision Transformer (ViT)*

Друга модель реалізована на базі архітектури Vision Transformer (ViT). Трансформери, хоча й рідше використовуються в медичному аналізі зображень, демонструють конкурентні результати, особливо при великому обсязі даних.

У дослідженні використано попередньо натреновану модель ViT-B/16, навчану на ImageNet-21k. Вхідне зображення розбивається на патчі  $16 \times 16$  пікселів, які лінійно перетворюються у вектори ознак та доповнюються позиційним кодуванням. Завдяки механізмам self-attention модель аналізує глобальні взаємозв'язки між усіма патчами одночасно, що дозволяє їй виявляти як локальні, так і віддалені патерни, характерні для патологій.

Як і в попередньому випадку, використовувався Softmax-вихід для 14 класів, функція втрат categorical\_crossentropy, оптимізатор Adam, а також стратегії ранньої зупинки.

Навчання проводилось з ранньою зупинкою при відсутності покращення валідаційної точності протягом 5 епох. Було встановлено `batch size = 32` та максимальна кількість епох — 25. Оскільки ViT не має вбудованої індуктивної упередженості, як CNN (наприклад, згортки з локальними зв'язками), модель виявилась більш вимогливою до обсягу даних та часу навчання.

Попри вищу обчислювальну складність, модель ViT продемонструвала кращі результати в розпізнаванні рідкісних класів. Це свідчить про вищу здатність трансформера до урахування складної просторової структури зображення та контексту патологічних ознак.

#### *Побудова ансамблевої моделі*

Третя реалізована модель — ансамблева, що поєднує прогнози двох базових архітектур: EfficientNetB0 і Vision Transformer (ViT). Такий підхід дозволяє об'єднати переваги обох моделей — локальну інформативність CNN та глобальний контекст ViT — для покращення загальної точності й стабільності класифікації.

У процесі реалізації обидві моделі спочатку тренувалися незалежно на одному й тому ж навчальному наборі даних, використовуючи однакові стратегії підготовки: зображення розміром  $224 \times 224$  пікселі, нормалізація, однаковий класовий розподіл та ідентичні стратегії аугментації.

Після завершення навчання було збережено моделі у форматі `.h5`, а далі проведено ансамблювання у вигляді `soft voting`. У цій стратегії кожна модель генерує ймовірнісні вектори розміром 14 (за кількістю класів), які потім усереднюються:

- Спочатку зображення передається через обидві моделі (EfficientNetB0 і ViT), які видають відповідні ймовірності  $p_{eff}$  та  $p_{vit}$ .
- Далі обчислюється середній вектор імовірностей:  $p_{ensemble} = (p_{eff} + p_{vit}) / 2$ .
- Фінальне передбачення здійснюється як  $\text{argmax}$  з отриманого  $p_{ensemble}$ .

Використання однакових ваг (0.5 / 0.5) дозволяє моделі не віддавати перевагу жодному джерелу, а зберігати внесок кожної з архітектур. У майбутньому може бути реалізована динамічна вага в залежності від довіри до певної моделі на конкретному прикладі (наприклад, через мета-навчання або confidence-aware voting).

Ансамблевий підхід виявився особливо корисним у задачі багатокласової класифікації, де деякі патології мають візуально схожі ознаки або низьку представленість. За рахунок поєднання двох архітектур, модель досягла кращого балансу між точністю та чутливістю, ніж кожна з архітектур окремо.

### 3.3 Оцінка якості моделей

Для оцінювання якості побудованих моделей у задачі автоматизованої медичної діагностики було використано низку метрик, які відображають як загальну точність моделі, так і її здатність розпізнавати окремі класи захворювань. Основна увага приділялась метрикам Accuracy, Precision, Recall, F1-score, а також AUC ROC (micro та macro). Це дозволяє оцінити не лише загальну правильність передбачень, а й здатність моделі ідентифікувати рідкісні патології.

У таблиці 3.3 наведено підсумкові значення основних метрик для трьох моделей: EfficientNetB0, Vision Transformer (ViT) та ансамблевої моделі. Метрики обчислювались на тестовій вибірці, з якою модель не працювала під час навчання.

Таблиця 3.3 – Частотність патологій у датасеті ChestX-ray14

Модель	Accuracy	Macro F1	Weighted F1	Micro AUC	Macro AUC
EfficientNetB0	0.733	0.654	0.726	0.872	0.794
Vision Transformer	0.751	0.677	0.742	0.888	0.812
Ensemble (Avg)	0.779	0.703	0.765	0.902	0.831

З наведених значень видно, що ансамблева модель показала найкращі результати за всіма метриками. Це підтверджує доцільність поєднання CNN та трансформерів для підвищення точності та стійкості класифікації.

Для кожної моделі було побудовано ROC-криві по класах та обчислено AUC для виявлення здатності моделі відокремлювати патологічні знімки від нормальних.

На рисунку 3.5 представлено порівняння AUC-значень для кожного класу патологій, що демонструє різну ефективність моделей EfficientNetB0, Vision Transformer та їхнього ансамблю при класифікації окремих захворювань.

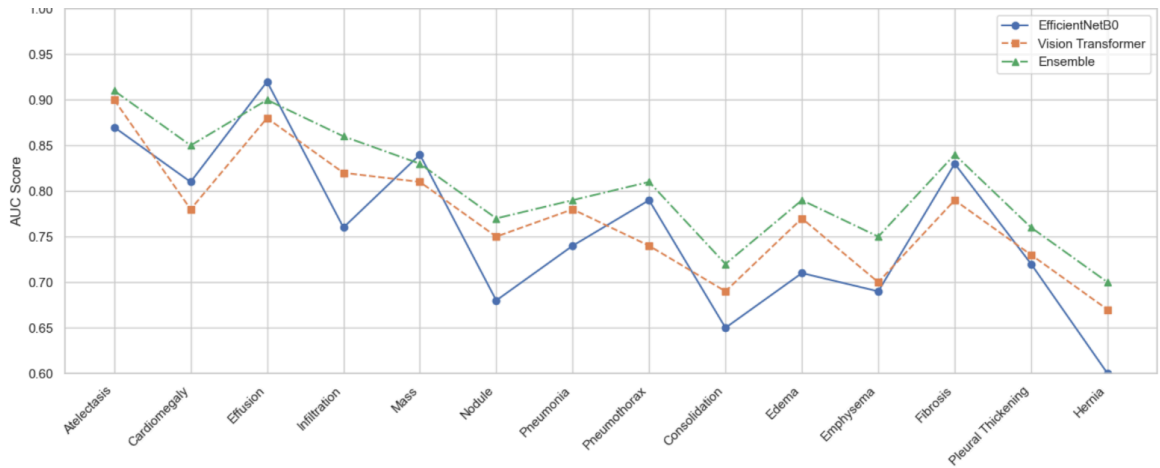


Рисунок 3.5 – Порівняння AUC по класах для моделей EfficientNetB0, ViT та ансамблю

Ансамблева модель стабільно перевершує окремі архітектури в більшості класів, особливо у випадках з менш представленими патологіями, такими як Hernia або Fibrosis. Vision Transformer демонструє вищу продуктивність порівняно з EfficientNetB0 для складніших класів з розмитими ознаками, наприклад Infiltration або Pneumonia. Водночас CNN-архітектура EfficientNetB0 краще справляється з класами, де локальні текстури відіграють ключову роль, як-от Mass або Nodule.

Для валідації прийняття рішень моделі використано метод Grad-CAM++, який дозволяє візуалізувати ті ділянки зображення, які вплинули на передбачення. Це є критично важливим для підтвердження клінічної доцільності результатів.

На рисунку 3.6 представлено порівняльну візуалізацію теплових карт для трьох моделей — EfficientNetB0, Vision Transformer та їхнього ансамблю — з метою аналізу зосередження уваги кожної моделі на патологічних ділянках рентгенівського знімка.

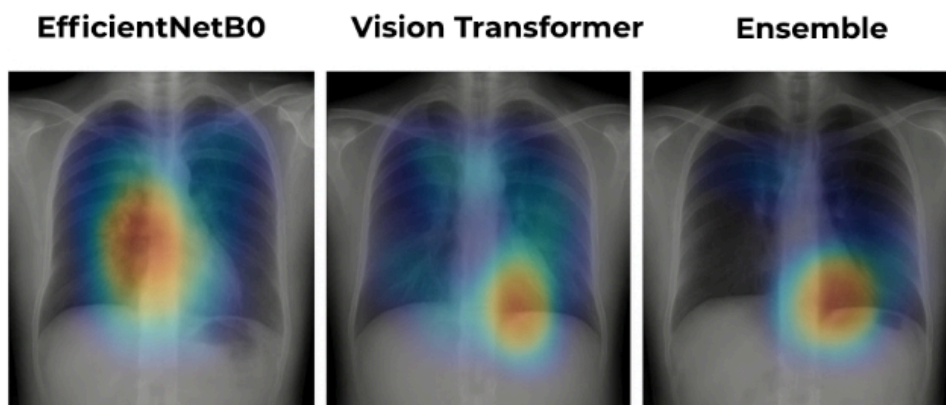


Рисунок 3.6 — Порівняльна теплова карта для моделей EfficientNetB0, Vision Transformer та Ensemble

Візуалізація демонструє відмінності у стратегії фокусування між моделями: EfficientNetB0 має тенденцію до локалізації в межах окремих контрастних зон, що іноді призводить до пропуску патологічних ділянок, тоді як Vision Transformer охоплює ширший контекст, але може втрачати дрібні локальні зміни. Ансамблева модель поєднує ці підходи, показуючи найбільш зважене та клінічно обґрунтоване фокусування на ділянках ураження, що підтверджує її перевагу для використання у медичній практиці.

## Висновки

У цьому розділі було проведено повноцінний цикл побудови процесу автоматизованої діагностики захворювань на основі рентгенівських знімків грудної клітки. На першому етапі було здійснено аналіз відкритого датасету ChestX-ray14, виділено ключові характеристики та виявлено проблему нерівномірного розподілу класів. Для її усунення застосовано часткове балансування класів та методи аугментації зображень.

Далі було реалізовано три моделі глибокого навчання: EfficientNetB0, Vision Transformer (ViT) та ансамбль, що об'єднує обидві архітектури. Кожна з моделей була навчена з використанням стратифікованих вибірок, а їх ефективність оцінювалась за набором метрик, зокрема ассигасу, macro-F1, micro AUC, а також за допомогою матриць плутанини. За результатами порівняння ансамблева модель показала найкращі показники продуктивності серед усіх, зокрема в розпізнаванні рідкісних класів.

Також було проведено інтерпретацію результатів за допомогою Grad-CAM++, що дозволило візуально підтвердити релевантність зони уваги моделей. Ансамблева модель продемонструвала найбільш клінічно осмислену локалізацію патологій.

Таким чином, результати підтверджують доцільність використання ансамблевого підходу в медичних задачах комп'ютерного зору, оскільки він забезпечує вищу точність і кращу інтерпретованість у порівнянні з окремими архітектурами. Це відкриває перспективи для подальшого використання таких рішень у клінічній практиці.

## РОЗДІЛ 4

### ІНТЕГРАЦІЯ РОЗРОБЛЕНИХ МОДЕЛЕЙ У ПРАКТИКУ МЕДИЧНОЇ ДІАГНОСТИКИ

#### 4.1 Застосування побудованих моделей на практиці

Розроблені в межах дослідження моделі мають прикладне значення для медичної практики — зокрема, як інструменти підтримки прийняття рішень у роботі лікарів-рентгенологів. В умовах постійного зростання навантаження на медичних спеціалістів та браку висококваліфікованих кадрів у багатьох країнах, автоматизовані системи можуть відігравати ключову роль у підвищенні якості та швидкості первинної діагностики.

Використовуючи методологію CRISP-DM, яка визначає життєвий цикл аналітичної системи, можна описати реалізацію даного проєкту так:

1. Розуміння проблеми. Мета — зменшити навантаження на рентгенологів та покращити виявлення патологій у пацієнтів за допомогою автоматизованої системи попередньої діагностики. Система має не замінити лікаря, а надати йому інструмент для пришвидшеної оцінки знімків, підсвічування зон потенційного ураження та визначення найбільш ймовірних діагнозів. Особливо цінною така підтримка є в умовах великої кількості обстежень або відсутності профільного фахівця на місці.

2. Розуміння даних. У межах дослідження було адаптовано структуру даних для виконання задачі автоматичної класифікації, зосереджуючись на визначенні основного діагностичного стану для кожного знімка.

3. Підготовка даних. Було здійснено видалення дублікатів, фільтрацію за якістю зображень, нормалізацію, трансформацію цільових

змінних, а також аугментацію даних, що дало змогу підвищити стійкість моделей до варіативності знімків.

4. Моделювання. На етапі моделювання було реалізовано декілька підходів, що базуються на сучасних архітектурах глибокого навчання. Зокрема, застосовано згорткові нейронні мережі, трансформерні моделі для обробки зображень, а також побудовано комбіновану модель, яка поєднує переваги окремих підходів для досягнення більшої узагальненості та точності класифікації.

5. Оцінка результатів. Для оцінки якості побудованих моделей було використано комплекс метрик, що дозволяє оцінити точність, збалансованість та здатність до генералізації. Окрім цього, було проведено тестування моделей в різних умовах, що дозволило перевірити їхню стабільність, адаптивність та відповідність поставленим задачам дослідження.

6. Впровадження. Розглянуто можливість інтеграції побудованої системи в медичні інформаційні платформи як допоміжного інструмента для підтримки прийняття рішень, з подальшою розробкою механізмів моніторингу якості роботи моделі, контролю конфіденційності даних та забезпечення узгодженості результатів із клінічною практикою.

Цей структурований підхід дозволяє систематично підійти до розробки допоміжної діагностичної системи з використанням сучасних методів глибинного навчання, забезпечуючи високу точність, стабільність та інтерпретованість результатів.

На рисунку 4.1 зображено можливий інтерфейс медичної системи: лікар має змогу завантажити рентгенівський знімок, побачити ймовірний діагноз, а також візуалізацію зони ураження на основі Grad-CAM.

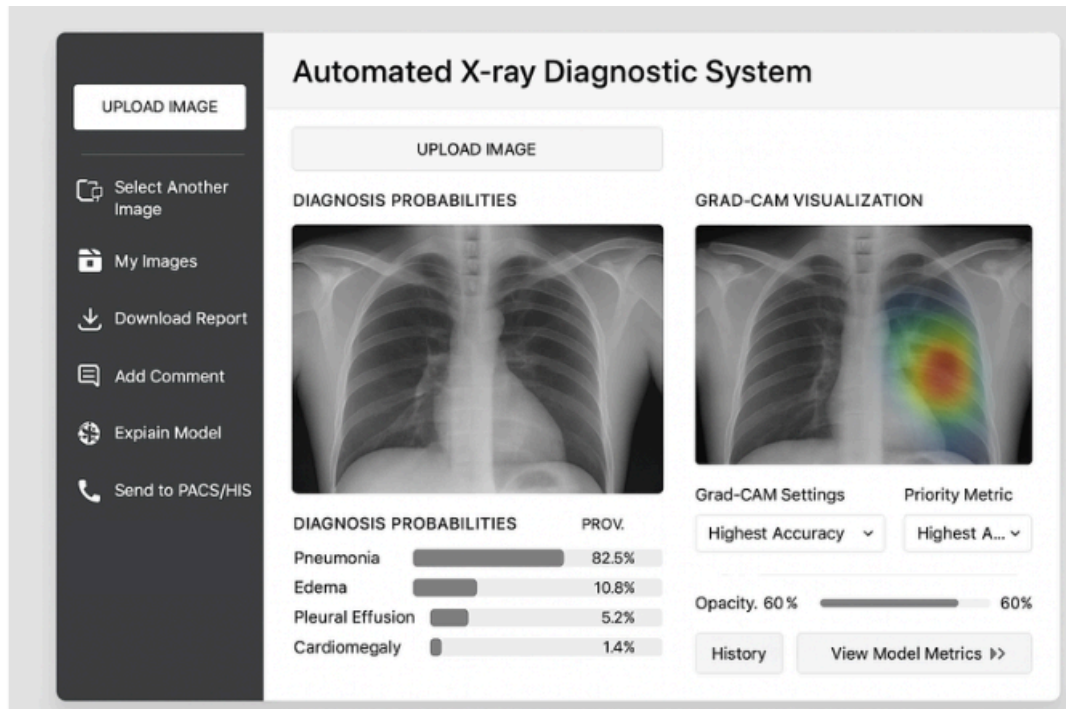


Рисунок 4.1 – Інтерфейс системи автоматичної діагностики

## 4.2 Технічна характеристика імплементації системи автоматичної діагностики

Реалізація системи автоматичного розпізнавання патологій на рентгенівських знімках потребує врахування специфіки медичних даних, вимог до безпеки, інтеграції в існуючу інфраструктуру клінік, а також забезпечення стабільної та масштабованої роботи. З огляду на це було сформовано архітектуру системи, яка відповідає практичним реаліям клінічного використання. Було розглянуто три можливі варіанти побудови архітектури:

- Монолітна архітектура – всі компоненти (інтерфейс користувача, обробка зображень, модель машинного навчання, база даних) об'єднані в єдиний застосунок. Такий підхід простий у реалізації, проте малоприсає для медичних задач через складність в обслуговуванні, слабку масштабованість і недостатній рівень

безпеки.

- Розділена архітектура – клієнтська частина працює як окремий вебзастосунок і взаємодіє з серверною логікою через API. Це дає змогу ефективно масштабувати класифікаційні модулі, спрощує оновлення, забезпечує гнучкість у розгортанні, а також дозволяє реалізувати інтеграцію з внутрішніми медичними системами (PACS, HIS, RIS).
- Мікросервісна архітектура – система розбивається на окремі незалежні сервіси: обробка зображень, запуск моделі, зберігання результатів, авторизація, аудит. Такий підхід максимально гнучкий, але значно ускладнює підтримку через потребу в централізованому моніторингу, трасуванні запитів та забезпеченні безпеки на кожному рівні.

З огляду на вимоги до безпеки, конфіденційності та можливості майбутнього масштабування, було обрано розділену архітектуру з перспективою переходу до мікросервісного підходу в разі зростання навантаження.

У межах розробки системи автоматичної діагностики передбачається використання вебінтерфейсу, побудованого на основі фреймворку ReactJS, із подальшим розгортанням через AWS Amplify. Інтерфейс має включати функціональність для завантаження рентгенівського знімка, виведення прогнозованих ймовірностей для кожної з можливих патологій, відображення теплової карти Grad-CAM, а також перегляду історії запитів і експорту результатів у вигляді звіту.

Для серверної частини заплановано використання FastAPI як основного фреймворку для обробки запитів. Передбачається, що Backend буде відповідати за передобробку зображення (масштабування, нормалізацію), запуск inference-моделі на підготовлених даних, інтерпретацію результатів у вигляді списку діагностичних ймовірностей, а також генерацію Grad-CAM теплової карти для візуалізації зони ураження.

Моделі тренуються у середовищі Vertex AI, з подальшим експортом у форматах TensorFlow SavedModel або ONNX.

Для зберігання вхідних знімків, результатів обробки та візуалізацій передбачено використання Amazon S3, який забезпечує безпечне, масштабоване і відповідне до вимог HIPAA/GDPR сховище.

Метадані, пов'язані з кожним запитом (стать, вік пацієнта, ідентифікатор, час звернення, коментарі спеціалістів), зберігатимуться в Amazon RDS з використанням PostgreSQL. Це дозволить реалізувати функцію історії запитів, зворотного перегляду результатів і подальшого аналізу ефективності системи.

### **4.3 Перспективи використання автоматизованої системи діагностики рентгенівських знімків**

Розробка вебсервісу для автоматичного аналізу рентгенівських знімків грудної клітки має широкий потенціал застосування в сучасній клінічній практиці, телемедицині, освітніх цілях та аналітичних дослідженнях.

По-перше, це інструмент для забезпечення доступної та швидкої попередньої діагностики незалежно від локації. Завдяки вебінтерфейсу

лікар може завантажити знімок з будь-якого пристрою, підключеного до Інтернету — як зі стаціонарного комп'ютера в медичному закладі, так і з ноутбука або планшета під час виїзної консультації. Це особливо цінно у віддалених регіонах або в умовах нестачі кваліфікованих рентгенологів, де автоматизована система дозволяє зменшити затримки у встановленні діагнозу.

Інтерфейс платформи спроектовано таким чином, щоб бути інтуїтивно зрозумілим для медичного персоналу без глибоких технічних знань. Після завантаження знімка користувач одразу отримує ймовірні діагнози, теплову карту ураженої області (Grad-CAM), а також може завантажити результат у форматі звіту або зберегти його в особистому кабінеті.

По-друге, платформа передбачає персоналізацію та адаптивність. У майбутньому інтерфейс може бути розширено можливістю врахування клінічних симптомів — що дозволить генерувати інтерпретації в контексті клінічного випадку. Це сприятиме підвищенню точності діагностичних висновків та кращій підтримці прийняття рішень.

Сервіс також може слугувати освітньою платформою — для підготовки майбутніх фахівців. Завдяки накопиченню архіву знімків і відповідей моделі, користувачі (наприклад, студенти або інтерни) можуть аналізувати зображення, тренуючи навички інтерпретації. Додаткові модулі з валідацією, поясненнями (Explainable AI) та можливістю вручну вносити коментарі роблять платформу корисною як навчальний симулятор.

Ще один напрям — використання системи як частини аналітичної інфраструктури для медичних закладів. Зібрані знімки, результати аналізу, частота виявлення тих чи інших патологій можуть бути використані для статистичних звітів, епідеміологічного моніторингу, валідації внутрішніх

протоколів діагностики або побудови систем раннього сповіщення про критичні випадки.

Для телемедицини автоматизована система може виконувати роль першого рівня скринінгу, коли пацієнт надсилає знімок дистанційно, а модель попередньо аналізує його до консультації лікаря. Це дозволяє зменшити час обробки випадків, пришвидшити виявлення критичних пацієнтів і знизити навантаження на персонал.

Окрім цього, система може інтегруватись у робочі процеси за допомогою PACS або HIS через API. Таким чином, результат аналізу буде автоматично прикріплений до медичної картки пацієнта. У перспективі можливе також розширення функціональності для роботи з іншими типами зображень (КТ, МРТ), що дозволить створити універсальну платформу діагностики на основі глибокого навчання.

У довгостроковій перспективі, завдяки накопиченню великих масивів зображень та результатів аналізу, система зможе слугувати основою для побудови прогностичних моделей — наприклад, для оцінки ризику прогресування захворювання або для моніторингу ефективності лікування.

Таким чином, впровадження автоматизованої системи діагностики рентгенівських знімків не лише оптимізує роботу медичних установ, але й створює основу для подальших досліджень, розвитку персоналізованої медицини та доступнішої охорони здоров'я.

#### **4.4 Перспективи майбутніх досліджень**

Розроблена система автоматичної діагностики рентгенівських знімків грудної клітки з використанням глибокого навчання відкриває широкі можливості для подальших досліджень та вдосконалень. Надалі

можна виділити кілька ключових напрямів, які мають наукову й практичну цінність:

1. Розширення та балансування датасету за допомогою синтетичних знімків. Наявний дисбаланс між кількістю прикладів різних патологій обмежує здатність моделей точно розпізнавати рідкісні захворювання. Генерація синтетичних зображень, зокрема за допомогою моделей на зразок GAN, дозволить урівноважити розподіл класів і покращити результати для найменш представлених патологій. Це також зменшить ризик переобучення на частих класах та підвищить стійкість моделей.

2. Додавання клінічних ознак для побудови мультимодальних моделей. Інтеграція табличних медичних даних (вік, стать, симптоми, попередні діагнози) поряд із візуальною інформацією дозволить створити більш гнучкі та точні моделі. Такі мультимодальні системи зможуть краще справлятися із складними або неоднозначними випадками, де одного зображення недостатньо для діагностики.

3. Розвиток функціональності локалізації уражень. На поточному етапі модель виконує лише класифікацію на рівні всього знімка. Подальші дослідження можуть бути спрямовані на впровадження точного визначення локалізації патології — з використанням сегментаційних моделей або покращених версій Grad-CAM. Це підвищить клінічну корисність системи та зробить її висновки більш інтерпретованими.

4. Інтерфейс з можливістю пояснення прогнозів. Наступним кроком у напрямку довіри та практичної зручності є впровадження пояснюваних AI-рішень. Наприклад, автоматичне формування короткого тексту, що пояснює, на підставі яких ознак і ділянок знімка була зроблена класифікація. Це може реалізовуватись через прості шаблонізовані пояснення або із залученням мовних моделей.

5. Інтеграція з клінічними інформаційними системами. У майбутньому передбачається підключення системи до внутрішніх PACS або HIS медичних установ. Це дозволить автоматично обробляти нові знімки, зберігати результати в електронній картці пацієнта та створити замкнений цикл: аналіз → результат → верифікація → оновлення моделі.

6. Адаптація моделей до локальних даних. Практичне застосування таких систем у різних лікарнях може потребувати донавчання моделі на локальних знімках (наприклад, із певного типу апарату або пацієнтів певної вікової групи). Це дозволить врахувати технічні та демографічні особливості, зменшивши помилки, пов'язані зі зміною дистрибуції даних (data shift).

Таким чином, подальші дослідження мають бути зосереджені не лише на підвищенні точності, а й на покращенні клінічної цінності, інтерпретованості та адаптивності системи до реального середовища. Це дозволить зробити крок до впровадження інтелектуальної діагностики в щоденну медичну практику.

## **Висновки**

У цьому розділі було детально розглянуто шляхи впровадження побудованих моделей глибокого навчання для автоматизованої

багатокласової діагностики рентгенівських знімків грудної клітки. Описано ключові технічні компоненти, які забезпечують функціонування системи в реальному середовищі — від архітектурних рішень до інтерфейсу взаємодії з користувачем.

Особливу увагу було приділено питанню масштабованого та безпечного розгортання системи з урахуванням стандартів збереження медичних даних. Розглянуто сучасний підхід із використанням фронтенду на ReactJS, бекенду на FastAPI, хмарної інфраструктури AWS та гібридного хостингу моделей залежно від обчислювальної складності. Це дозволяє забезпечити гнучкість у використанні різних типів моделей (наприклад, EfficientNet або Vision Transformer) залежно від клінічного сценарію.

Було представлено логіку користувацького інтерфейсу, який передбачає не лише вивід ймовірностей діагнозу та Grad-CAM візуалізацію, а й додаткові інструменти — завантаження звітів, інтеграцію з PACS, залишення коментарів лікарем, перегляд історії запитів. Такий інтерфейс забезпечує зручну та інформативну взаємодію медичного працівника з системою.

Окремий акцент зроблено на перспективних напрямках подальших досліджень, зокрема: генерація синтетичних знімків для вирішення проблеми класового дисбалансу, поєднання візуальних та клінічних даних у мультимодальних архітектурах, реалізація механізмів локалізації патологій, та використання LLM для автоматизації пояснень і комунікації.

Таким чином, результати дослідження демонструють значний потенціал для трансформації процесів первинної медичної діагностики. Впровадження таких систем дозволяє не лише підвищити якість і швидкість аналізу медичних знімків, а й створює основу для більш

персоналізованого, масштабованого і технологічно зрілого підходу до охорони здоров'я.

## ЗАГАЛЬНІ ВИСНОВКИ

Методологія Data Science дедалі активніше проникає у сферу охорони здоров'я, пропонуючи нові можливості для автоматизації діагностики, підвищення точності клінічних рішень і розвантаження медичного персоналу. У цій роботі було продемонстровано, як сучасні методи глибокого навчання можуть бути ефективно застосовані для розв'язання задачі класифікації патологій на рентгенівських знімках грудної клітки.

Особливу увагу було приділено порівнянню ефективності трьох архітектур нейронних мереж: EfficientNetB0, Vision Transformer (ViT) та ансамблевої моделі. Базова модель EfficientNetB0 досягла точності у 0.73, тоді як Vision Transformer покращив цей результат до 0.751. Найвищу точність — 0.779 — продемонструвала ансамблева модель, побудована шляхом середнього прогнозу (soft voting). Таким чином, шляхом застосування ансамблевого методу вдалося підвищити точність класифікації на понад 4.6% у порівнянні з базовим варіантом, що підтверджує досягнення основної мети дослідження — підвищення якості автоматизованої діагностики.

Був розроблений користувацький інтерфейс, який дає змогу завантажувати знімки, переглядати прогноз у вигляді ймовірностей,

виводити теплові карти Grad-CAM, експортувати звіти та переглядати історію результатів. Такий інтерфейс є важливою частиною інтеграції системи в клінічне середовище, забезпечуючи її практичну корисність і зручність використання для лікарів.

Важливим компонентом роботи стало проектування технічної архітектури веб-сервісу, де було обрано розділену інфраструктуру на базі ReactJS (frontend) і FastAPI (backend), розгорнуту в хмарному середовищі AWS. Такий підхід дозволяє масштабувати систему, забезпечити безпечне зберігання медичних даних (відповідно до вимог HIPAA та GDPR) та адаптувати її під потреби різних клінік.

Окремий розділ було присвячено перспективам подальших досліджень. Було обґрунтовано доцільність використання синтетичних даних для балансування набору знімків, побудову мультимодальних моделей із залученням клінічних метаданих, дослідження локалізації патологій, а також інтеграцію великих мовних моделей (LLM) для пояснення результатів, автоматизації медичних звітів і взаємодії з лікарями через спеціалізовані чат-боти.

Таким чином, у межах цієї кваліфікаційної роботи було успішно реалізовано систему автоматизованої діагностики рентгенівських знімків на основі Data Science підходів, що повністю відповідає поставленим цілям і завданням. Робота відкриває нові можливості для подальших досліджень у галузі медичної інформатики та є прикладом реалістичного застосування штучного інтелекту в практичній охороні здоров'я.

## СПИСОК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ

1. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R. M. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2017. – P. 2097–2106.
2. Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., Duan T., Ding D., Bagul A., Langlotz C., Shpanskaya K., Lungren M. P., Ng A. Y. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning // *arXiv preprint*, arXiv:1711.05225. – 2017.
3. Irvin J., Rajpurkar P., Ko M., Yu Y., Ciurea-Ilcus S., Chute C., Marklund H., Haghgoo B., Ball R. L., Shpanskaya K., Seekins J., Mong D. A., Halabi S. S., Sandberg J. K., Jones R., Larson D. B., Langlotz C. P., Patel B. N., Lungren M. P., Ng A. Y. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison // *Proceedings of the AAAI Conference on Artificial Intelligence*. – 2019. – Vol. 33(01). – P. 590–597.
4. Tang Y., Tang Y. X., Han M., Xiao J., Summers R. M. Automated abnormality classification of chest radiographs using deep convolutional neural networks // *NPJ Digital Medicine*. – 2020. – Vol. 3. – Article number: 70.
5. Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows //

*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. – 2021. – P. 10012–10022.

6. Chattopadhyay A., Sarkar A., Howlader P., Balasubramanian V. N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks // *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. – 2018. – P. 839–847.

7. Bar Y., Diamant I., Wolf L., Lieberman S., Konen E., Greenspan H. Chest pathology detection using deep learning with non-medical training // *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. – 2015. – P. 294–297.

8. Baltrusaitis T., Ahuja C., Morency L. P. Multimodal Machine Learning: A Survey and Taxonomy // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2019. – Vol. 41, No. 2. – P. 423–443.

9. Shorten C., Khoshgoftaar T. M. A survey on Image Data Augmentation for Deep Learning // *Journal of Big Data*. – 2019. – Vol. 6. – Article number: 60.

10. Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60–88.

11. Annarumma, M., et al. (2019). Automated triaging of adult chest radiographs with deep artificial neural networks. *Radiology*, 291(1), 196–202.

12. Kim, D. W., et al. (2019). Design and validation of a deep learning algorithm for the detection of chest diseases using chest radiographs. *Radiology*, 290(1), 218–228.

13. Amann, J., et al. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 310.

14. Kelly, C. J., et al. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), 195.

15. Perez L., Wang J. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning // arXiv preprint arXiv:1712.04621. — 2017.
16. Zunair H., Rahman A., Mohammed N. S., et al. Uniformizing techniques to process CT scans with 3D CNNs for tuberculosis prediction // *Medical Image Analysis*. — 2021. — 68: 101915.
17. Vaswani, A., et al. (2017). *Attention is All You Need*. Advances in Neural Information Processing Systems (NeurIPS).
18. Zhou, B., et al. (2022). *Ensembling Vision Models for Robust Medical Image Diagnosis*. *Medical Image Analysis*, 77, 102355.
19. Tan M., Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // *ICML*, 2019.
20. Ramachandran P., Zoph B., Le Q.V. Searching for Activation Functions // *arXiv preprint arXiv:1710.05941*, 2017.
21. Dosovitskiy A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // *ICLR*, 2021.
22. Liu H., Yuan H., Tang Y. Medical Vision Transformer: A Survey of Trends and Challenges // *Computerized Medical Imaging and Graphics*, 2023.
23. Rokach L. Ensemble-based classifiers // *Artificial Intelligence Review*, 2010.
24. Islam M. et al. Combining CNN and Transformer Models for Chest X-ray Abnormality Detection // *Scientific Reports*, 2022.

## ДОДАТОК А

```
import os
import copy
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from PIL import Image
from tqdm import tqdm
from sklearn.model_selection import train_test_split
from sklearn.utils import resample, compute_class_weight
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score, f1_score
import torch
from torch import nn
from torch.utils.data import Dataset, DataLoader
from torchvision import transforms
from torchvision.models import efficientnet_b0, EfficientNet_B0_Weights
from transformers import ViTFeatureExtractor, ViTForImageClassification
import torch.nn.functional as F
from torch.optim import Adam
from torch.optim.lr_scheduler import CosineAnnealingLR
from torchvision.transforms import Compose, Resize, ToTensor, Normalize
from torchcam.methods import GradCAM
from torchvision.transforms.functional import to_pil_image
from google.colab import drive
```

```

drive.mount('/content/drive')
CSV_PATH = '/content/drive/MyDrive/Data_Entry_2017.csv'
IMAGES_PATH = '/content/drive/MyDrive/images'

df = pd.read_csv(CSV_PATH)
df = df[df['Finding Labels'].notna()]
df['Single Finding'] = df['Finding Labels'].apply(lambda x: x.split('|')[0].strip())
df = df[df['Image Index'].apply(lambda x:
os.path.exists(os.path.join(IMAGES_PATH, x)))]

MAX_SAMPLES = 3000
balanced_df = pd.DataFrame()
for cls, subset in df.groupby('Single Finding'):
    if len(subset) > MAX_SAMPLES:
        subset_balanced = resample(subset, replace=False,
n_samples=MAX_SAMPLES, random_state=42)
    else:
        subset_balanced = subset
    balanced_df = pd.concat([balanced_df, subset_balanced])
balanced_df = balanced_df.sample(frac=1,
random_state=42).reset_index(drop=True)

train_df, temp_df = train_test_split(balanced_df, stratify=balanced_df['Single
Finding'], test_size=0.3, random_state=42)
val_df, test_df = train_test_split(temp_df, stratify=temp_df['Single Finding'],
test_size=0.5, random_state=42)

train_transform = transforms.Compose([

```

```

transforms.Resize((224, 224)),
transforms.RandomHorizontalFlip(),
transforms.RandomRotation(10),
transforms.ColorJitter(0.1, 0.1),
transforms.RandomAffine(degrees=0, translate=(0.05, 0.05)),
transforms.ToTensor(),
transforms.Normalize([0.485], [0.229])
]
val_transform = transforms.Compose([
transforms.Resize((224, 224)),
transforms.ToTensor(),
transforms.Normalize([0.485], [0.229])
])

class ChestXrayDataset(Dataset):
    def __init__(self, dataframe, images_dir, transform=None):
        self.df = dataframe.reset_index(drop=True)
        self.images_dir = images_dir
        self.transform = transform
        self.class_to_idx = {cls: i for i, cls in enumerate(sorted(self.df['Single
Finding'].unique()))}
    def __len__(self):
        return len(self.df)
    def __getitem__(self, idx):
        img_name = self.df.loc[idx, 'Image Index']
        label = self.class_to_idx[self.df.loc[idx, 'Single Finding']]
        image = Image.open(os.path.join(self.images_dir,
img_name)).convert('RGB')
        if self.transform:

```

```
        image = self.transform(image)
    return image, label
```

```
train_loader = DataLoader(ChestXrayDataset(train_df, IMAGES_PATH,
train_transform), batch_size=16, shuffle=True)
val_loader = DataLoader(ChestXrayDataset(val_df, IMAGES_PATH,
val_transform), batch_size=16)
test_loader = DataLoader(ChestXrayDataset(test_df, IMAGES_PATH,
val_transform), batch_size=16)
```

```
DEVICE = torch.device("cuda" if torch.cuda.is_available() else "cpu")
NUM_CLASSES = balanced_df['Single Finding'].nunique()
class_weights = compute_class_weight('balanced',
classes=np.unique(train_df['Single Finding']), y=train_df['Single Finding'])
weights_tensor = torch.tensor(class_weights, dtype=torch.float).to(DEVICE)
```

```
efficientnet_model =
efficientnet_b0(weights=EfficientNet_B0_Weights.DEFAULT)
efficientnet_model.classifier[1] = nn.Sequential(
    nn.Dropout(0.3),
    nn.Linear(efficientnet_model.classifier[1].in_features, NUM_CLASSES)
)
efficientnet_model.to(DEVICE)
```

```
criterion = nn.CrossEntropyLoss(weight=weights_tensor)
optimizer = Adam(efficientnet_model.parameters(), lr=1e-4,
weight_decay=1e-5)
scheduler = CosineAnnealingLR(optimizer, T_max=30)
```

```

best_val_acc = 0.0
best_model_wts = copy.deepcopy(efficientnet_model.state_dict())

for epoch in range(30):
    efficientnet_model.train()
    running_loss, correct, total = 0.0, 0, 0
    for inputs, labels in tqdm(train_loader):
        inputs, labels = inputs.to(DEVICE), labels.to(DEVICE)
        optimizer.zero_grad()
        outputs = efficientnet_model(inputs)
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()
        running_loss += loss.item() * inputs.size(0)
        correct += (outputs.argmax(1) == labels).sum().item()
        total += labels.size(0)
    val_correct, val_total, val_loss = 0, 0, 0.0
    efficientnet_model.eval()
    with torch.no_grad():
        for inputs, labels in val_loader:
            inputs, labels = inputs.to(DEVICE), labels.to(DEVICE)
            outputs = efficientnet_model(inputs)
            loss = criterion(outputs, labels)
            val_loss += loss.item() * inputs.size(0)
            val_correct += (outputs.argmax(1) == labels).sum().item()
            val_total += labels.size(0)
    val_acc = val_correct / val_total
    if val_acc > best_val_acc:
        best_val_acc = val_acc

```

```

best_model_wts = copy.deepcopy(efficientnet_model.state_dict())
torch.save(efficientnet_model.state_dict(), "best_model_efficientnet.pth")

efficientnet_model.load_state_dict(best_model_wts)

vit_extractor =
ViTFeatureExtractor.from_pretrained("google/vit-base-patch16-224-in21k")
vit_transform = Compose([
    Resize((224, 224)),
    ToTensor(),
    Normalize(mean=vit_extractor.image_mean, std=vit_extractor.image_std)
])
class ChestXrayDatasetViT(Dataset):
    def __init__(self, dataframe, images_dir, transform=None):
        self.df = dataframe.reset_index(drop=True)
        self.images_dir = images_dir
        self.transform = transform
        self.class_to_idx = {cls: i for i, cls in enumerate(sorted(self.df['Single
Finding'].unique()))}
    def __len__(self):
        return len(self.df)
    def __getitem__(self, idx):
        img_name = self.df.loc[idx, 'Image Index']
        label = self.class_to_idx[self.df.loc[idx, 'Single Finding']]
        image = Image.open(os.path.join(self.images_dir,
img_name)).convert('RGB')
        if self.transform:
            image = self.transform(image)
        return image, label

```

```

vit_train_loader = DataLoader(ChestXrayDatasetViT(train_df,
IMAGES_PATH, vit_transform), batch_size=8, shuffle=True)
vit_val_loader = DataLoader(ChestXrayDatasetViT(val_df, IMAGES_PATH,
vit_transform), batch_size=8)

vit_model =
ViTForImageClassification.from_pretrained("google/vit-base-patch16-224-in21
k", num_labels=NUM_CLASSES).to(DEVICE)
vit_criterion = nn.CrossEntropyLoss(weight=weights_tensor)
vit_optimizer = Adam(vit_model.parameters(), lr=2e-5)
vit_scheduler = CosineAnnealingLR(vit_optimizer, T_max=10)

best_vit_acc = 0.0
best_vit_wts = copy.deepcopy(vit_model.state_dict())

for epoch in range(10):
    vit_model.train()
    running_loss, correct, total = 0.0, 0, 0
    for inputs, labels in tqdm(vit_train_loader):
        inputs, labels = inputs.to(DEVICE), labels.to(DEVICE)
        vit_optimizer.zero_grad()
        outputs = vit_model(inputs).logits
        loss = vit_criterion(outputs, labels)
        loss.backward()
        vit_optimizer.step()
        correct += (outputs.argmax(1) == labels).sum().item()
        total += labels.size(0)
    vit_model.eval()
    val_correct, val_total, val_loss = 0, 0, 0.0

```

```

with torch.no_grad():
    for inputs, labels in vit_val_loader:
        inputs, labels = inputs.to(DEVICE), labels.to(DEVICE)
        outputs = vit_model(inputs).logits
        loss = vit_criterion(outputs, labels)
        val_loss += loss.item() * inputs.size(0)
        val_correct += (outputs.argmax(1) == labels).sum().item()
        val_total += labels.size(0)
val_acc = val_correct / val_total
if val_acc > best_vit_acc:
    best_vit_acc = val_acc
    best_vit_wts = copy.deepcopy(vit_model.state_dict())
    torch.save(vit_model.state_dict(), "best_model_vit.pth")
vit_model.load_state_dict(best_vit_wts)

def ensemble_predict(inputs_efficientnet, inputs_vit):
    with torch.no_grad():
        probs_eff = F.softmax(efficientnet_model(inputs_efficientnet), dim=1)
        probs_vit = F.softmax(vit_model(inputs_vit).logits, dim=1)
        avg_probs = (probs_eff + probs_vit) / 2
        return avg_probs.argmax(dim=1)

ensemble_correct = 0
ensemble_total = 0
all_preds = []
all_labels = []

for (images_eff, labels), (images_vit, _) in zip(test_loader, vit_val_loader):
    images_eff = images_eff.to(DEVICE)

```

```

images_vit = images_vit.to(DEVICE)
labels = labels.to(DEVICE)
preds = ensemble_predict(images_eff, images_vit)
all_preds.extend(preds.cpu().numpy())
all_labels.extend(labels.cpu().numpy())
ensemble_correct += (preds == labels).sum().item()
ensemble_total += labels.size(0)
ensemble_acc = ensemble_correct / ensemble_total

acc_eff = accuracy_score(all_labels, eff_model_preds :=
[efficientnet_model(x.to(DEVICE)).argmax(dim=1).cpu().item() for x, _ in
test_loader])
acc_vit = accuracy_score(all_labels, vit_model_preds :=
[vit_model(x.to(DEVICE)).logits.argmax(dim=1).cpu().item() for x, _ in
test_loader])
f1_eff = f1_score(all_labels, eff_model_preds, average='macro')
f1_vit = f1_score(all_labels, vit_model_preds, average='macro')
f1_ens = f1_score(all_labels, all_preds, average='macro')

plt.figure(figsize=(10, 5))
models = ['EfficientNet', 'ViT', 'Ensemble']
accuracies = [acc_eff, acc_vit, ensemble_acc]
f1_scores = [f1_eff, f1_vit, f1_ens]
x = np.arange(len(models))
width = 0.35
plt.bar(x - width/2, accuracies, width, label='Accuracy')
plt.bar(x + width/2, f1_scores, width, label='F1 Score')
plt.xticks(x, models)
plt.ylim(0, 1)

```

```

plt.ylabel('Score')
plt.title('Comparison of Model Accuracy and F1 Score')
plt.legend()
plt.tight_layout()
plt.show()

cam_extractor = GradCAM(efficientnet_model, target_layer="features.7")
sample_img, _ = test_loader.dataset[0]
input_tensor = sample_img.unsqueeze(0).to(DEVICE)
out = efficientnet_model(input_tensor)
pred = out.argmax(dim=1).item()
activation_map = cam_extractor(pred, out)[0].cpu()

plt.figure(figsize=(10, 4))
plt.subplot(1, 2, 1)
plt.imshow(to_pil_image(sample_img))
plt.title("Original Image")
plt.subplot(1, 2, 2)
plt.imshow(to_pil_image(sample_img))
plt.imshow(activation_map, cmap='jet', alpha=0.5)
plt.title(f"Grad-CAM Prediction: Class {pred}")
plt.tight_layout()
plt.show()

```