

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Методи використання систем штучного інтелекту для підтримки
прийняття рішень в задачах медичної діагностики
(тема)

Виконав:
здобувач другого року навчання,
групи СШМ-24-1

Антон ЖУК
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва освітньої програми)

Керівник проф. Марина КУДРЯВЦЕВА
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ

Лариса ЧАЛА
(підпис) (власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
Кафедра _____ Штучного інтелекту _____
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системи штучного інтелекту _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«_____» _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Жуку Антону Владиславовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи Методи використання систем штучного інтелекту для підтримки прийняття рішень в задачах медичної діагностики

затверджена наказом університету від 6 квітня 2026 р. № 269Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 19 травня 2026 р.

3. Вихідні дані до роботи наукові статті та публікації з області комп'ютерного зору та медичної діагностики, зокрема дослідження, присвячені застосуванню згорткових нейронних мереж, архітектури ResNet та Vision Transformer для класифікації медичних зображень; роботи, що описують підходи до багатоміткової класифікації та методи попереднього навчання моделей; документацію та матеріали щодо використання бібліотек глибокого навчання; відкритий набір даних NIH Chest X-ray та супровідні ресурси для його обробки і використання у задачах машинного навчання.

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі та постановка задачі _____

2) Теоретичні дослідження _____

3) Програмна реалізація _____

4) Експериментальні дослідження _____

КАЛЕНДАРНИЙ ПЛАН

[illegible]

Дата видачі завдання 6 квітня 20 26 р.

Здобувач _____
(підпис)

Керівник роботи _____
(підпис)

проф. Марина Кудрявцева
(посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 130 с., 51 рис., 2 табл., 1 дод., 21 джерело.

БАГАТОМІТКОВА КЛАСИФІКАЦІЯ, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, КОМП'ЮТЕРНИЙ ЗІР, МЕДИЧНА ДІАГНОСТИКА, РЕНТГЕНОГРАФІЯ ГРУДНОЇ КЛІТКИ, ШТУЧНИЙ ІНТЕЛЕКТ, AUC, CONVOLUTIONAL NEURAL NETWORK, MULTI-HEAD ATTENTION, RESIDUAL NETWORK, ROC CURVE, VISION TRANSFORMER.

Об'єктом дослідження є системи медичної діагностики, орієнтовані на автоматизований аналіз рентгенівських знімків грудної клітки із застосуванням методів штучного інтелекту.

Предметом дослідження є моделі глибокого навчання, зокрема нейромережеві архітектури CNN, ResNet та Vision Transformer.

Мета роботи – підвищення точності підтримки прийняття рішень у задачах медичної діагностики шляхом дослідження та порівняння сучасних архітектур глибокого навчання для класифікації рентгенівських знімків.

Методами дослідження є методи глибокого навчання та методи обробки та аналізу медичних зображень.

У результаті роботи проведено дослідження ефективності сучасних архітектур глибокого навчання у задачі багатоміткової класифікації рентгенівських знімків грудної клітки. Встановлено, що гібридна архітектура, яка поєднує згорткові мережі та механізм самоуваги.

Результати роботи можуть бути використані для створення систем підтримки прийняття рішень у медичній діагностиці, зокрема для автоматизованого аналізу рентгенівських знімків. Сфера застосування включає медичні інформаційні системи, телемедицину та інтелектуальні системи аналізу медичних зображень.

ABSTRACT

Master's thesis contains: 130 pp., 51 fig., 2 tabl., 1 ann., 21 references.

ARTIFICIAL INTELLIGENCE, AUC, CHEST X-RAY IMAGING, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORK, MEDICAL DIAGNOSIS, MULTI-HEAD ATTENTION, MULTI-LABEL CLASSIFICATION, RESIDUAL NETWORK, ROC CURVE, VISION TRANSFORMER.

The object of the study is medical diagnostic systems focused on the automated analysis of chest X-ray images using artificial intelligence methods.

The subject of the study is deep learning models, in particular neural network architectures such as CNN, ResNet, and Vision Transformer.

The aim of the work is to improve the accuracy of decision support in medical diagnostic tasks by studying and comparing modern deep learning architectures for the classification of X-ray images.

The research methods include deep learning techniques and methods for processing and analyzing medical images.

As a result of the study, an evaluation of the effectiveness of modern deep learning architectures in the task of multi-label classification of chest X-ray images was conducted. It was established that a hybrid architecture combining convolutional neural networks and a self-attention mechanism is effective.

The results of the work can be used to develop decision support systems in medical diagnostics, particularly for automated analysis of X-ray images. The field of application includes medical information systems, telemedicine, and intelligent systems for medical image analysis.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	10
Вступ.....	11
1 Аналіз предметної галузі та постановка задачі.....	13
1.1 Роль штучного інтелекту в сучасній медичній діагностиці	13
1.1.1 Процес медичної діагностики та підтримка прийняття рішень	14
1.1.2 Системи штучного інтелекту в клінічній практиці	16
1.1.3 Переваги і недоліки штучного інтелекту в медицині	18
1.2 Медичні дані як основа інтелектуальної діагностики.....	19
1.2.1 Основні типи медичних даних у діагностичних системах	21
1.2.2 Особливості обробки медичних зображень	23
1.2.3 Проблеми якості, повноти та доступності медичних даних	25
1.3 Методи штучного інтелекту для аналізу медичних даних	26
1.3.1 Методи машинного навчання для медичної класифікації.....	27
1.3.2 Глибоке навчання для аналізу медичних зображень	28
1.3.3 Використання нейронних мереж у системах підтримки медичної діагностики	30
1.4 Мульти模альні підходи у медичній діагностиці.....	31
1.4.1 Поняття мульти模альних даних у медицині	33
1.4.2 Поєднання різних типів медичної інформації у системах ШІ ..	34
1.4.3 Мульти模альні моделі для підвищення точності діагностики	35
1.5 Постановка задачі.....	36
2 Теоретичні дослідження	37
2.1 Багатоміткова класифікація рентгенівських знімків грудної клітки .	37
2.1.1 Сутність багатоміткової класифікації медичних зображень.....	38
2.1.2 Особливості представлення класів у рентгенівських знімках грудної клітки	40
2.1.3 Специфіка застосування багатоміткової класифікації у задачах медичної діагностики	42

2.2 Класичні архітектури згорткових нейронних мереж	44
2.2.1 Архітектура CNN	45
2.2.2 Основні шари та принцип формування ознак у CNN	47
2.2.3 Особливості застосування CNN у задачах класифікації рентгенівських знімків	49
2.3 Архітектура ResNet як розвиток згорткових нейронних мереж	50
2.3.1 Проблема зникнення градієнта в глибоких мережах	52
2.3.2 Архітектура ResNet і механізм residual learning	53
2.3.3 Особливості застосування ResNet у задачах медичної класифікації зображень	56
2.4 Трансформерний підхід у комп'ютерному зорі.....	57
2.4.1 Передумови переходу від CNN до трансформерів.....	59
2.4.2 Переваги трансформерного підходу порівняно зі згортковими мережами.....	60
2.4.3 Базова архітектура трансформера	62
2.5 Архітектура Vision Transformer	64
2.5.1 Поділ зображення на патчі у Vision Transformer	66
2.5.2 Основні компоненти Vision Transformer	68
2.5.3 Особливості класифікації зображень у Vision Transformer	69
2.6 Механізм самоуваги як теоретична основа Vision Transformer	71
2.6.1 Призначення self-attention у задачах аналізу зображень	72
2.6.2 Формування запитів, ключів і значень	74
2.6.3 Multi-head attention та його роль у виділенні ознак.....	75
2.7 Вплив попереднього навчання на ефективність моделі.....	77
2.7.1 Сутність підходу pre-training у глибокому навчанні.....	78
2.7.2 Роль попереднього навчання для Vision Transformer	79
2.8 Гібридна архітектура на основі ResNet та Vision Transformer	81
2.8.1 Аналіз обмежень класичних та трансформерних архітектур....	81
2.8.2 Обґрунтування доцільності використання гібридного підходу	84
2.8.3 Формування структури гібридної архітектури	86

2.8.4	Опис основних модулів запропонованої архітектури.....	88
2.8.5	Формалізація алгоритму навчання гібридної моделі	89
3	Програмна реалізація.....	92
3.1	Опис програмного середовища реалізації дослідження	92
3.1.1	Мова програмування та середовище розробки.....	92
3.1.2	Бібліотеки для обробки даних і навчання моделей.....	92
3.2	Програмна підготовка набору даних до навчання моделей	93
3.2.1	Завантаження та структура даних	94
3.2.2	Попередня обробка рентгенівських знімків.....	95
3.2.3	Формування цільових міток для багатоміткової класифікації..	95
3.3	Програмна реалізація моделей класифікації.....	96
3.3.1	Реалізація моделі CNN	97
3.3.2	Реалізація моделі ResNet	98
3.3.3	Реалізація моделей ViT-v1 та ViT-v2.....	99
3.3.4	Реалізація гібридної моделі ViT-ResNet.....	102
3.4	Реалізація процесу навчання моделей	103
3.5	Реалізація засобів оцінювання результатів	106
3.5.1	Обчислення показників Accuracy та AUC.....	106
3.5.2	Побудова ROC-кривих і кривих навчання	107
3.5.3	Формування attention maps для трансформерних моделей.....	108
4	Експериментальні дослідження.....	109
4.1	План експериментальних досліджень.....	109
4.1.1	Мета та гіпотеза експерименту.....	109
4.1.2	Критерії порівняння моделей	110
4.2	Характеристика набору даних NIH Chest X-ray	111
4.3	Результати дослідження базових моделей	112
4.3.1	Результати моделі CNN.....	113
4.3.2	Результати моделі ResNet	115
4.4	Результати дослідження трансформерних моделей	116
4.4.1	Результати моделі ViT-v1	117

4.4.2 Результати моделі ViT-v2	118
4.4.3 Результати моделі ViT-ResNet.....	119
4.5 Аналіз впливу попереднього навчання на результати класифікації	121
4.6 Порівняльний аналіз отриманих результатів	122
Висновки	126
Перелік джерел посилання	128
Додаток А Відомість кваліфікаційної роботи	130

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AUC – Area Under Curve – площа під ROC-кривою, метрика оцінювання якості класифікації;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

Cross-Entropy – функція втрат, що використовується для задач класифікації;

Deep Learning – підхід машинного навчання, що базується на багатошарових нейронних мережах;

Image Classification – задача автоматичного визначення класу зображення;

Multi-head Attention – механізм самоуваги, що дозволяє моделі одночасно враховувати кілька представлень даних;

Multi-label Classification – багатоміткова класифікація, при якій об'єкт може належати до кількох класів одночасно;

Pre-training – попереднє навчання моделі на великому наборі даних перед донавчанням;

ResNet – Residual Neural Network – глибока нейронна мережа з залишковими зв'язками;

ROC – Receiver Operating Characteristic – характеристична крива, що відображає співвідношення між чутливістю та специфічністю моделі;

Self-Attention – механізм самоуваги, що дозволяє моделі враховувати залежності між різними частинами вхідних даних;

ViT – Vision Transformer – трансформерна архітектура для обробки зображень;

ViT-ResNet – гібридна модель, що поєднує Vision Transformer і ResNet.

ВСТУП

В умовах стрімкого розвитку інформаційних технологій значна увага приділяється використанню методів штучного інтелекту для розв'язання складних задач аналізу даних. Однією з найбільш перспективних сфер застосування таких технологій є медицина, де сучасні методи машинного навчання дозволяють автоматизувати процес аналізу великих обсягів медичної інформації. Використання інтелектуальних систем дає можливість підвищити точність діагностики, прискорити обробку клінічних даних та забезпечити підтримку прийняття рішень для медичних фахівців. Особливо актуальним є застосування штучного інтелекту в умовах постійного зростання обсягів медичних даних, які включають результати лабораторних досліджень, медичні зображення, електронні медичні записи та фізіологічні сигнали.

Останніми роками активно розвиваються підходи, що базуються на використанні машинного та глибокого навчання для аналізу медичних даних. Зокрема, значне поширення отримали моделі глибоких нейронних мереж, які здатні ефективно обробляти складні структуровані та неструктуровані дані. Такі моделі дозволяють виявляти приховані закономірності у великих наборах даних, що є важливим для раннього виявлення захворювань та прогнозування стану пацієнтів. Одним із перспективних напрямів розвитку є використання мультимодальних підходів, які передбачають поєднання декількох типів медичної інформації для отримання більш повної та достовірної картини стану пацієнта.

Використання мультимодальних моделей дозволяє інтегрувати різні джерела інформації, зокрема медичні зображення, клінічні дані та фізіологічні показники, що сприяє підвищенню точності та надійності діагностичних систем. Однак створення таких систем пов'язане з низкою складних науково-технічних задач, серед яких визначення ефективних архітектур моделей машинного навчання, вибір оптимальних параметрів

навчання, а також забезпечення балансу між точністю моделей, швидкістю їх роботи та обчислювальними витратами. Це зумовлює необхідність проведення комплексних досліджень, спрямованих на оцінювання ефективності різних підходів до побудови інтелектуальних діагностичних систем.

У межах даної роботи розглядається застосування методів машинного навчання для аналізу медичних даних з метою підвищення ефективності медичної діагностики. Дослідження спрямоване на визначення оптимальних архітектур моделей для розв'язання діагностичних задач, проведення експериментального оцінювання їх якості, аналіз впливу параметрів навчання на результати роботи моделей, а також формування рекомендацій щодо інтеграції таких моделей у реальні системи медичної діагностики. Очікується, що результати дослідження сприятимуть підвищенню ефективності використання штучного інтелекту в медичній практиці та розвитку інтелектуальних систем підтримки прийняття медичних рішень.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Роль штучного інтелекту в сучасній медичній діагностиці

Стрімкий розвиток цифрових технологій та накопичення значних обсягів медичних даних зумовили активне впровадження методів штучного інтелекту у сферу охорони здоров'я. Традиційні підходи до медичної діагностики ґрунтуються на досвіді лікаря, аналізі клінічних симптомів та результатів лабораторних і інструментальних досліджень.

Штучний інтелект дозволяє аналізувати великі масиви різномірних даних [1], виявляти приховані закономірності та підвищувати точність діагнозу. Системи штучного інтелекту здатні ефективно працювати з різними типами медичної інформації (рисунок 1.1), зокрема медичними зображеннями, клінічними записами, лабораторними показниками та генетичними даними. Використання алгоритмів машинного навчання і глибокого навчання дозволяє автоматизувати процеси аналізу зображень, виявлення патологічних змін та прогнозування розвитку захворювань.



Рисунок 1.1 – Напрями використання штучного інтелекту в медичній діагностиці

Важливою перевагою використання штучного інтелекту є можливість підвищення точності діагностики та зменшення ймовірності помилок. Алгоритми здатні обробляти значні обсяги даних та виявляти навіть незначні зміни, які можуть бути непомітними для людини. Крім того, системи штучного інтелекту сприяють підвищенню доступності медичних послуг, оскільки можуть використовуватися для дистанційного аналізу даних, моніторингу стану пацієнтів та автоматизації рутинних процедур у медичних закладах.

Разом з тим застосування штучного інтелекту в медицині пов'язане з низкою викликів. До них належать питання етичного використання медичних даних, забезпечення конфіденційності інформації, а також необхідність високої якості навчальних наборів даних для побудови надійних моделей. Незважаючи на ці обмеження, сучасні дослідження демонструють значний потенціал використання штучного інтелекту для підвищення ефективності медичної діагностики, оптимізації клінічних процесів та розвитку персоналізованої медицини.

1.1.1 Процес медичної діагностики та підтримка прийняття рішень

Процес медичної діагностики [2] є складною багатоступеневою процедурою, що передбачає аналіз клінічних симптомів, результатів лабораторних досліджень, медичних зображень та анамнезу пацієнта. Основною метою цього процесу є встановлення точного діагнозу, який дозволяє визначити оптимальну стратегію лікування. У традиційній практиці прийняття діагностичних рішень значною мірою залежить від професійного досвіду лікаря та його здатності інтерпретувати наявну інформацію. Однак сучасна медицина характеризується значним зростанням обсягів медичних даних, що ускладнює їх повноцінний аналіз без використання спеціалізованих інформаційних систем.

У зв'язку з цим у медичній практиці активно застосовуються системи підтримки прийняття клінічних рішень (рисунок 1.2). Такі системи призначені для допомоги лікарю у процесі аналізу медичних даних та формування рекомендацій щодо діагностики або лікування. Вони можуть використовувати різні підходи до обробки інформації, зокрема бази медичних знань, статистичні методи або алгоритми машинного навчання.

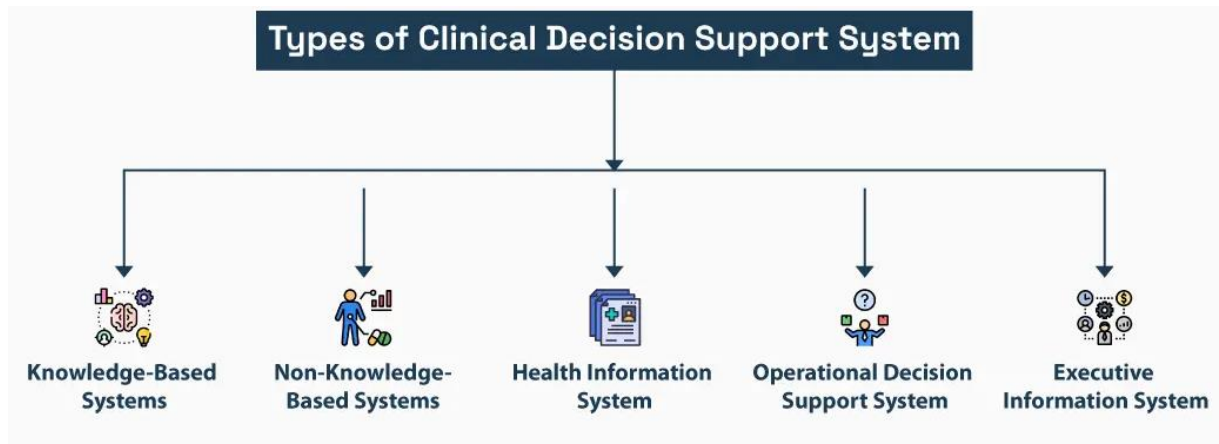


Рисунок 1.2 – Типи систем підтримки прийняття клінічних рішень

Залежно від принципів побудови такі системи можуть бути засновані на правилах, що формуються на основі експертних знань, або використовувати підходи машинного навчання для автоматичного виявлення закономірностей у медичних даних. Крім того, до структури систем підтримки прийняття рішень можуть входити інформаційні системи управління медичними даними, операційні системи підтримки рішень та аналітичні системи для керівництва медичних установ.

Типова архітектура (рисунок 1.3) системи підтримки клінічних рішень включає декілька послідовних етапів обробки інформації. Спочатку відбувається отримання медичних даних із електронних медичних записів пацієнта, після чого виконується їх попередня обробка та підготовка до аналізу.

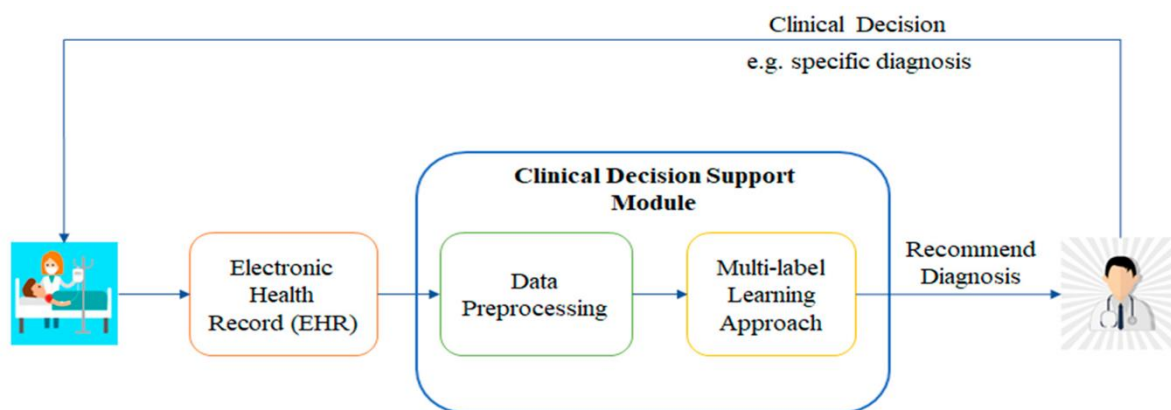


Рисунок 1.3 – Архітектура системи підтримки клінічних рішень на основі медичних даних

Далі застосовуються алгоритми машинного навчання або інші аналітичні методи для виявлення закономірностей у даних та формування рекомендацій щодо можливого діагнозу. Використання систем підтримки прийняття рішень дозволяє підвищити якість медичної допомоги, зменшити ризик діагностичних помилок та оптимізувати процес прийняття клінічних рішень. У сучасних умовах такі системи стають важливим інструментом цифрової медицини, оскільки вони сприяють інтеграції великих обсягів медичних даних, підвищують ефективність роботи лікарів та забезпечують більш обґрунтований підхід до постановки діагнозу.

1.1.2 Системи штучного інтелекту в клінічній практиці

У сучасній медицині системи штучного інтелекту поступово інтегруються у клінічну практику [3] та стають важливим інструментом підтримки лікарів у процесі діагностики, лікування та моніторингу стану пацієнтів (рисунок 1.4). Завдяки здатності обробляти великі обсяги медичних даних та виявляти складні закономірності, алгоритми штучного інтелекту можуть значно підвищити ефективність клінічних процесів.

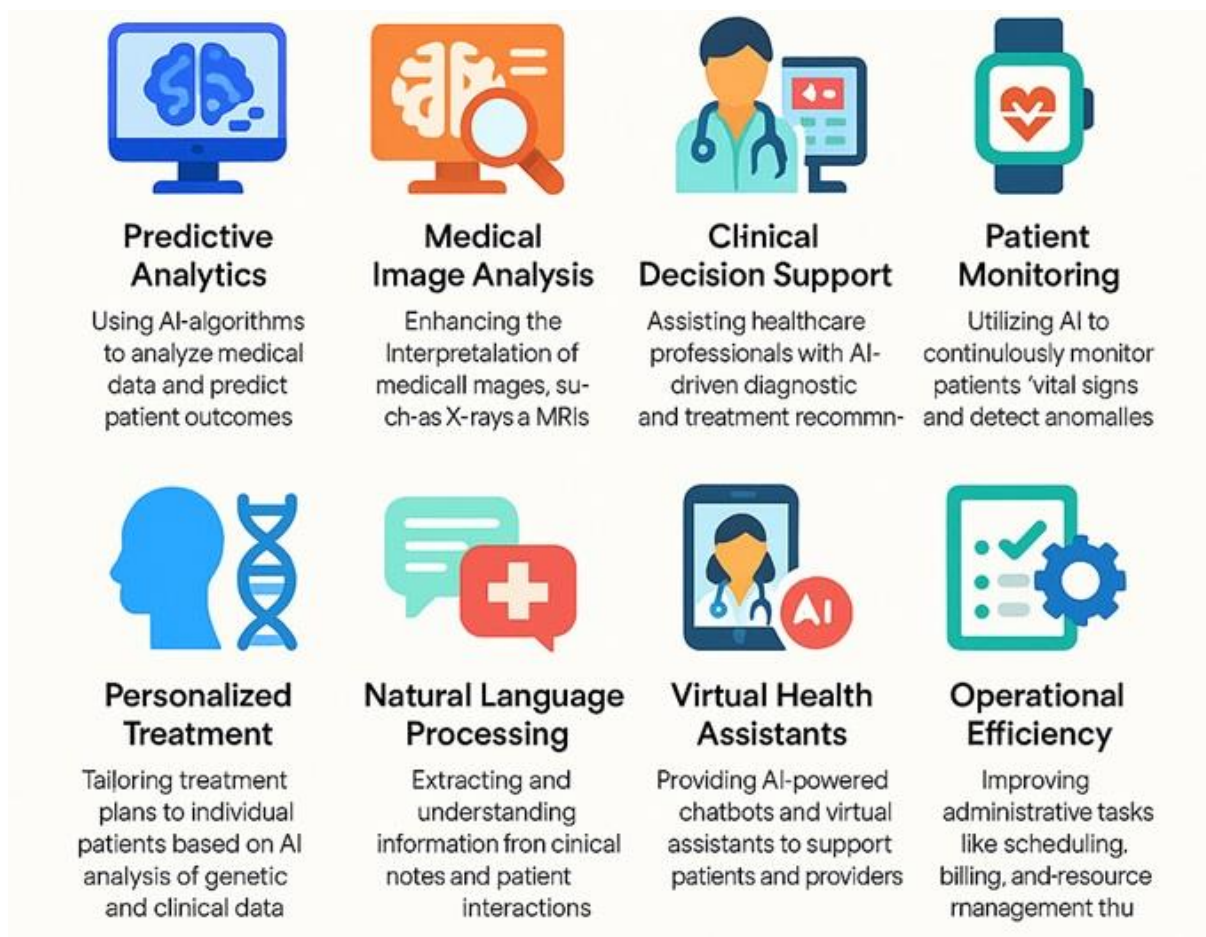


Рисунок 1.4 – Напрями використання систем штучного інтелекту в клінічній практиці

Одним із найпоширеніших напрямів застосування штучного інтелекту є аналіз медичних зображень. Алгоритми глибокого навчання дозволяють автоматично виявляти патологічні зміни на рентгенівських знімках, комп'ютерній томографії та магнітно-резонансних зображеннях. Це значно полегшує роботу лікарів-радіологів та підвищує точність діагностики, оскільки система може звертати увагу на деталі, які можуть бути пропущені під час візуального аналізу.

Іншим важливим напрямом є використання систем підтримки клінічних рішень, що допомагають лікарям формувати рекомендації щодо діагностики та лікування. Такі системи аналізують клінічні дані пацієнта, результати лабораторних досліджень та історію хвороби, після чого

пропонують можливі варіанти діагнозу або лікувальної тактики. Це дозволяє зменшити ризик помилок та забезпечити більш обґрунтований підхід до лікування.

Крім того, системи штучного інтелекту активно використовуються для моніторингу стану пацієнтів та прогнозування розвитку захворювань. Алгоритми можуть аналізувати показники життєдіяльності пацієнтів у режимі реального часу, виявляти аномальні зміни та попереджати медичний персонал про можливі ускладнення. У поєднанні з методами аналізу природної мови такі системи також здатні обробляти медичні записи та клінічні нотатки, що дозволяє отримувати додаткову інформацію для прийняття клінічних рішень.

Важливим напрямом розвитку є персоналізована медицина, у межах якої алгоритми штучного інтелекту використовуються для адаптації лікування до індивідуальних особливостей пацієнта. Аналіз генетичних, клінічних та поведінкових даних дозволяє формувати індивідуальні плани лікування та підвищувати ефективність медичних втручань. Таким чином, використання систем штучного інтелекту у клінічній практиці сприяє підвищенню якості медичної допомоги, оптимізації роботи медичних установ та розвитку нових підходів до діагностики та лікування.

1.1.3 Переваги і недоліки штучного інтелекту в медицині

Використання систем штучного інтелекту в медицині відкриває значні можливості для підвищення ефективності діагностики та лікування. Однією з головних переваг таких систем є здатність аналізувати великі обсяги медичних даних за короткий час та виявляти складні закономірності, які можуть бути непомітними для людини. Це дозволяє підвищити точність діагностики, зменшити ймовірність медичних помилок та забезпечити більш обґрунтоване прийняття клінічних рішень. Крім того, застосування

штучного інтелекту сприяє автоматизації рутинних процесів, що дозволяє медичним працівникам зосередитися на безпосередній роботі з пацієнтами.

Разом з тим використання штучного інтелекту в медицині має певні обмеження. Однією з основних проблем є залежність ефективності алгоритмів від якості та обсягу навчальних даних. Недостатня репрезентативність або наявність помилок у даних можуть призводити до некоректних результатів моделювання. Також важливими залишаються питання конфіденційності медичної інформації, етичного використання даних пацієнтів та прозорості роботи алгоритмів. У зв'язку з цим системи штучного інтелекту розглядаються не як повна заміна лікаря, а як інструмент підтримки прийняття рішень, що доповнює професійні знання медичних спеціалістів.

1.2 Медичні дані як основа інтелектуальної діагностики

Сучасні системи штучного інтелекту [3] в медицині ґрунтуються на аналізі великих обсягів медичних даних, які формуються в процесі діагностики, лікування та моніторингу пацієнтів. Медичні дані можуть включати результати лабораторних досліджень, електронні медичні записи, медичні зображення, генетичну інформацію та показники, отримані з різних сенсорних пристроїв. Саме ці дані (рисунк 1.5) є основним джерелом інформації для побудови моделей машинного навчання, які використовуються для автоматизованого аналізу стану здоров'я пацієнтів.

Завдяки розвитку цифрових технологій у медичних установах активно впроваджуються електронні системи зберігання медичних записів та різноманітні інформаційні платформи, що дозволяють накопичувати значні обсяги клінічної інформації. Це створює передумови для використання методів аналізу великих даних, які дозволяють об'єднувати інформацію з різних джерел та формувати комплексне уявлення про стан пацієнта.

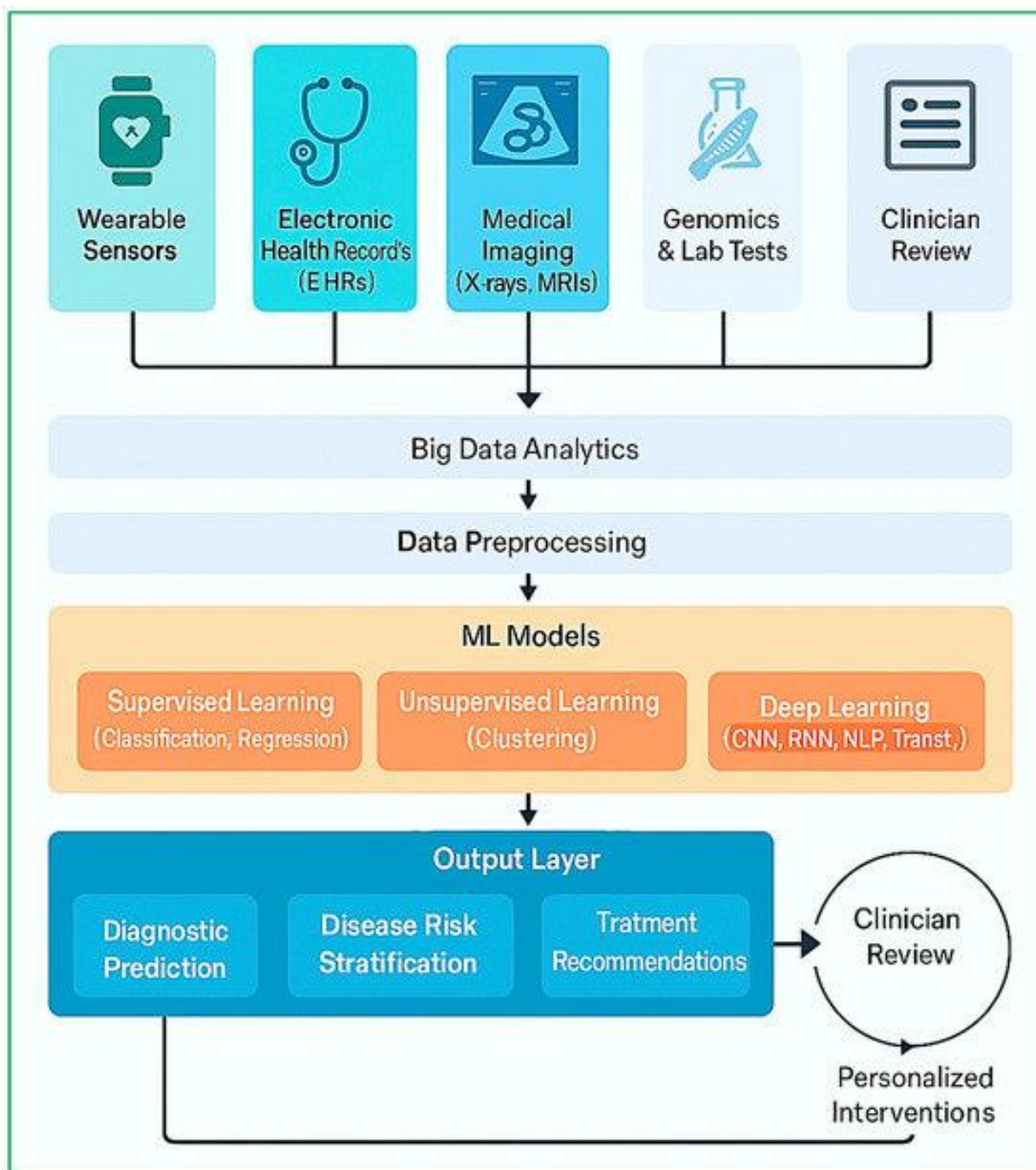


Рисунок 1.5 – Основні джерела медичних даних та етапи їх обробки в системах інтелектуальної медичної діагностики

У більшості інтелектуальних систем діагностики процес роботи з медичними даними складається з кількох послідовних етапів. Спочатку відбувається збір інформації з різних джерел, після чого виконуються процедури попередньої обробки даних, зокрема очищення, нормалізація та структуризація інформації. Після цього підготовлені дані використовуються

для навчання моделей машинного навчання, таких як методи класифікації, кластеризації або глибокі нейронні мережі.

Результатом роботи таких моделей є формування аналітичних висновків, які можуть використовуватися для прогнозування розвитку захворювань, визначення рівня ризику або формування рекомендацій щодо лікування. При цьому отримані результати аналізу зазвичай проходять додаткову перевірку лікарем, що дозволяє поєднати можливості автоматизованих систем з клінічним досвідом медичних спеціалістів. Таким чином, медичні дані виступають ключовим ресурсом для функціонування систем інтелектуальної діагностики та забезпечують основу для створення ефективних систем підтримки прийняття медичних рішень.

1.2.1 Основні типи медичних даних у діагностичних системах

У сучасних системах інтелектуальної медичної діагностики використовуються різноманітні типи медичних даних (рисунк 1.6), які формуються в процесі клінічного обстеження пацієнтів, проведення лабораторних досліджень та медичних процедур. Такі дані можуть мати різну структуру, формат і рівень деталізації, що зумовлює необхідність застосування спеціалізованих методів їх обробки та аналізу. Поєднання різних типів медичної інформації дозволяє отримати більш повне уявлення про стан пацієнта та підвищити точність діагностики.

Одним із найбільш поширених типів медичних даних є структуровані текстові дані. До цієї категорії належать електронні медичні записи пацієнтів, демографічна інформація, результати лабораторних аналізів, показники життєдіяльності, а також коди діагнозів і медичні призначення. Такі дані зазвичай зберігаються у стандартизованому вигляді та можуть ефективно оброблятися інформаційними системами для подальшого аналізу та використання у системах підтримки прийняття клінічних рішень.

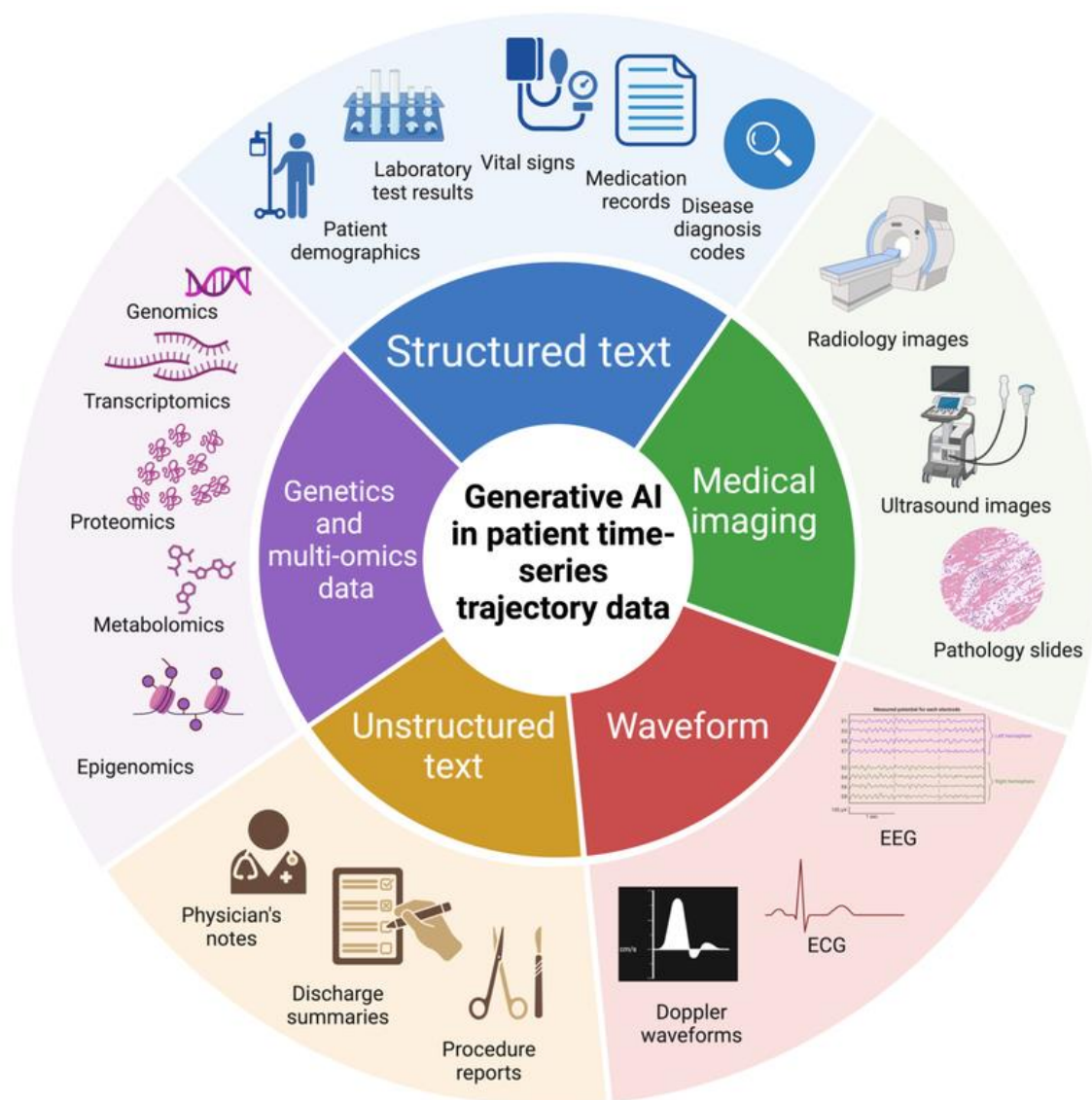


Рисунок 1.6 – Основні типи медичних даних, що використовуються у системах інтелектуальної медичної діагностики

Іншим важливим джерелом інформації є медичні зображення, до яких належать рентгенівські знімки, зображення комп'ютерної томографії, ультразвукові дослідження та гістологічні зразки. Аналіз таких зображень дозволяє виявляти патологічні зміни у тканинах та органах, що є важливим елементом сучасної діагностики. Для обробки медичних зображень активно застосовуються методи глибокого навчання, які дозволяють автоматично розпізнавати характерні ознаки захворювань.

Крім того, у діагностичних системах використовуються текстові дані неструктурованого типу, такі як лікарські нотатки, виписки з історії хвороби та звіти про медичні процедури. Для аналізу таких даних застосовуються методи обробки природної мови, що дозволяють автоматично виділяти ключову інформацію з текстових документів. Важливу роль також відіграють сигналові дані, зокрема електрокардіограми, електроенцефалограми та інші біомедичні сигнали, які характеризують фізіологічний стан організму.

Окрему категорію становлять генетичні та мультиомні дані, що включають геноміку, протеоміку, метаболоміку та інші типи молекулярної інформації. Аналіз таких даних дозволяє виявляти генетичні фактори розвитку захворювань та формувати персоналізовані підходи до лікування. Таким чином, використання різних типів медичних даних у поєднанні з методами штучного інтелекту створює передумови для розвитку комплексних систем інтелектуальної діагностики.

1.2.2 Особливості обробки медичних зображень

Медичні зображення є одним із найважливіших джерел інформації для сучасної діагностики, оскільки дозволяють візуально оцінювати стан органів і тканин пацієнта.

До таких зображень (рисунок 1.7) належать рентгенівські знімки, комп'ютерна томографія, магнітно-резонансна томографія, ультразвукові дослідження та гістологічні зображення. Вони широко застосовуються для виявлення патологій, оцінювання прогресування захворювань та контролю ефективності лікування. У сучасних системах медичної діагностики обробка таких зображень дедалі частіше виконується за допомогою методів штучного інтелекту.

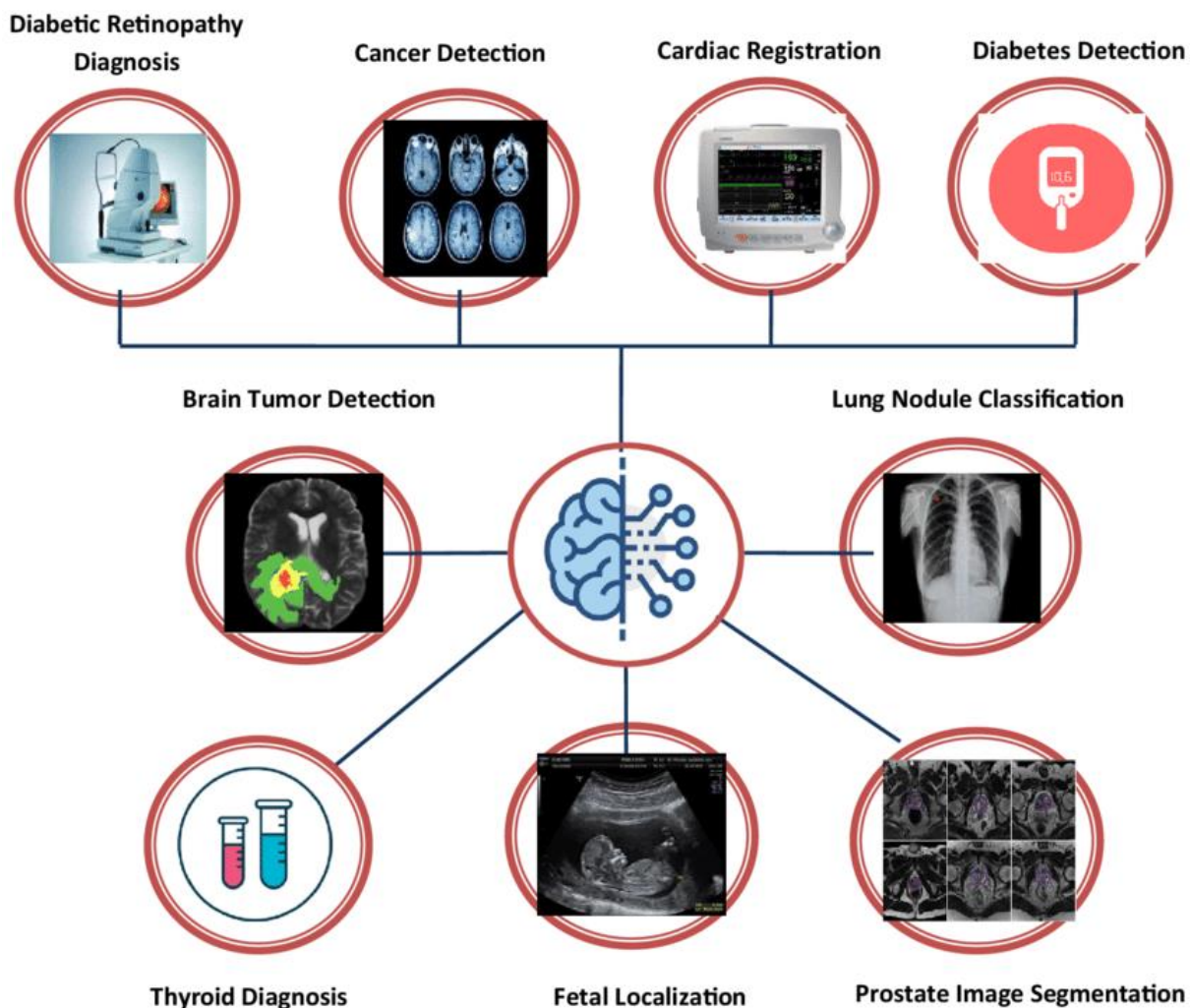


Рисунок 1.7 – Приклади застосування аналізу медичних зображень у задачах медичної діагностики

Однією з ключових особливостей обробки медичних зображень є необхідність попередньої підготовки даних. Перед застосуванням алгоритмів машинного навчання зображення проходять етапи нормалізації, фільтрації шуму, масштабування та виділення області інтересу. Такі процедури дозволяють підвищити якість вхідних даних і забезпечити більш точну роботу алгоритмів аналізу. У деяких випадках також виконується сегментація зображень, що передбачає виділення конкретних анатомічних структур або патологічних областей.

Для аналізу медичних зображень широко використовуються методи глибокого навчання, зокрема згорткові нейронні мережі. Такі моделі здатні

автоматично виділяти характерні ознаки зображень та використовувати їх для класифікації або виявлення патологій. Завдяки цьому системи штучного інтелекту можуть застосовуватися для діагностики різних захворювань, зокрема пухлин головного мозку, раку легенів, діабетичної ретинопатії, патологій щитоподібної залози та інших медичних станів.

Важливою особливістю використання медичних зображень у системах штучного інтелекту є необхідність високої точності та надійності результатів. Медичні рішення мають безпосередній вплив на здоров'я пацієнтів, тому алгоритми повинні проходити ретельну перевірку та клінічну валідацію. Поєднання автоматизованого аналізу з експертною оцінкою лікаря дозволяє підвищити ефективність діагностики та забезпечити більш точне виявлення патологічних змін.

1.2.3 Проблеми якості, повноти та доступності медичних даних

Ефективність систем штучного інтелекту у медичній діагностиці значною мірою залежить від якості медичних даних, що використовуються для навчання та функціонування алгоритмів. Медичні дані часто формуються з різних джерел, зокрема електронних медичних записів, лабораторних досліджень, медичних зображень та результатів моніторингу пацієнтів. Через різноманітність форматів і способів збирання інформації виникають проблеми неоднорідності та несумісності даних, що може ускладнювати їх подальшу обробку та аналіз.

Однією з основних проблем є недостатня якість медичних даних. У клінічній практиці можуть виникати помилки під час введення інформації, неточності у вимірюваннях або неповнота записів у медичній документації. Такі фактори можуть негативно впливати на процес навчання моделей машинного навчання та призводити до зниження точності прогнозування. Крім того, медичні дані можуть містити шум або артефакти, що потребує

застосування спеціальних методів попередньої обробки та очищення інформації.

Іншою важливою проблемою є неповнота медичних даних. У багатьох випадках клінічна інформація про пацієнта може бути частково відсутньою або неповною, що ускладнює формування повної картини стану здоров'я. Відсутність певних показників або історичних даних може обмежувати можливості алгоритмів штучного інтелекту щодо формування точних діагностичних висновків або прогнозів.

Крім того, суттєвим обмеженням є доступність медичних даних. Через вимоги щодо конфіденційності та захисту персональної інформації доступ до медичних даних часто є обмеженим. Це створює труднощі для формування великих навчальних наборів даних, необхідних для побудови ефективних моделей машинного навчання. У зв'язку з цим особливої актуальності набувають питання анонімізації даних, забезпечення інформаційної безпеки та розробки механізмів безпечного обміну медичною інформацією між різними медичними установами.

1.3 Методи штучного інтелекту для аналізу медичних даних

Застосування штучного інтелекту у медицині [4] базується на використанні різноманітних методів аналізу даних, що дозволяють виявляти закономірності у складних медичних наборах інформації. Медичні дані можуть включати структуровані записи, текстову інформацію, біомедичні сигнали та медичні зображення, що потребує використання спеціалізованих алгоритмів обробки та аналізу. Методи штучного інтелекту дозволяють автоматизувати процес інтерпретації таких даних та формувати рекомендації для підтримки клінічних рішень [21].

Однією з основних груп методів є алгоритми машинного навчання, які використовуються для класифікації, прогнозування та виявлення закономірностей у медичних даних. Такі алгоритми можуть навчатися на

основі історичних даних пацієнтів та використовувати отримані знання для прогнозування розвитку захворювань або оцінювання ризиків. У медичних задачах часто застосовуються методи класифікації для визначення наявності певного захворювання, а також методи регресії для прогнозування числових показників стану здоров'я.

Важливу роль у сучасних дослідженнях відіграють методи глибокого навчання, які дозволяють ефективно аналізувати складні типи медичних даних, зокрема медичні зображення та біомедичні сигнали. Глибокі нейронні мережі здатні автоматично виділяти значущі ознаки з даних без необхідності ручного формування характеристик. Завдяки цьому такі моделі широко використовуються для автоматичної діагностики, сегментації медичних зображень та виявлення патологічних змін.

Крім того, у медичних інформаційних системах активно застосовуються методи обробки природної мови, що дозволяють аналізувати текстові медичні записи, клінічні звіти та лікарські нотатки. Використання таких методів дає змогу витягувати важливу інформацію з великих обсягів текстових документів та інтегрувати її з іншими типами медичних даних. Таким чином, поєднання різних методів штучного інтелекту створює основу для побудови комплексних систем аналізу медичних даних та підтримки прийняття клінічних рішень.

1.3.1 Методи машинного навчання для медичної класифікації

Методи машинного навчання відіграють важливу роль у задачах медичної діагностики, оскільки дозволяють автоматично аналізувати клінічні дані та класифікувати їх відповідно до певних медичних станів або захворювань. У задачах медичної класифікації алгоритми навчаються на основі попередньо зібраних медичних даних, що містять приклади з відомими діагнозами. Після завершення навчання модель може

використовуватися для визначення ймовірної категорії захворювання або стану пацієнта на основі нових даних.

Одними з найбільш поширених підходів є методи керованого навчання, які використовують розмічені навчальні набори даних. До таких методів належать логістична регресія, дерева рішень, метод опорних векторів та ансамблеві алгоритми, зокрема випадкові ліси та градієнтний бустинг. Ці алгоритми широко застосовуються для класифікації медичних записів, прогнозування розвитку захворювань та виявлення факторів ризику. Завдяки можливості працювати з різними типами даних вони можуть використовуватися для аналізу лабораторних показників, демографічної інформації та клінічних параметрів пацієнтів.

Важливою перевагою методів машинного навчання є їх здатність виявляти складні залежності між різними характеристиками медичних даних. Це дозволяє використовувати такі алгоритми для раннього виявлення захворювань, оцінювання ймовірності розвитку патологій та підтримки прийняття клінічних рішень. У багатьох випадках моделі машинного навчання можуть демонструвати точність, співставну з результатами експертного аналізу лікарів, що підвищує їх практичну цінність у медичних інформаційних системах. Разом з тим ефективність застосування методів машинного навчання у медичній класифікації значною мірою залежить від якості та обсягу навчальних даних. Наявність шуму, пропущених значень або нерепрезентативних вибірок може знижувати точність моделей.

1.3.2 Глибоке навчання для аналізу медичних зображень

Глибоке навчання є одним із найбільш ефективних підходів до аналізу медичних зображень [4] у сучасних системах інтелектуальної діагностики. На відміну від класичних методів машинного навчання, які потребують ручного формування ознак, глибокі нейронні мережі здатні автоматично виділяти інформативні характеристики зображень у процесі навчання.

Найбільш поширеним типом моделей для обробки зображень є згорткові нейронні мережі. Такі моделі складаються з кількох послідовних шарів, які виконують операції згортки, підвибірки та нелінійної активації, що дозволяє поступово виділяти все більш складні ознаки зображення. У задачах медичної діагностики ці мережі можуть використовуватися для класифікації медичних зображень, виявлення патологічних змін або сегментації анатомічних структур.

У практичних дослідженнях часто застосовується підхід Transfer Learning (рисунок 1.8), який передбачає використання попередньо навчених моделей глибокого навчання. Такі моделі спочатку навчаються на великих наборах загальних зображень, після чого адаптуються до конкретних медичних задач шляхом додаткового навчання на спеціалізованих медичних даних.

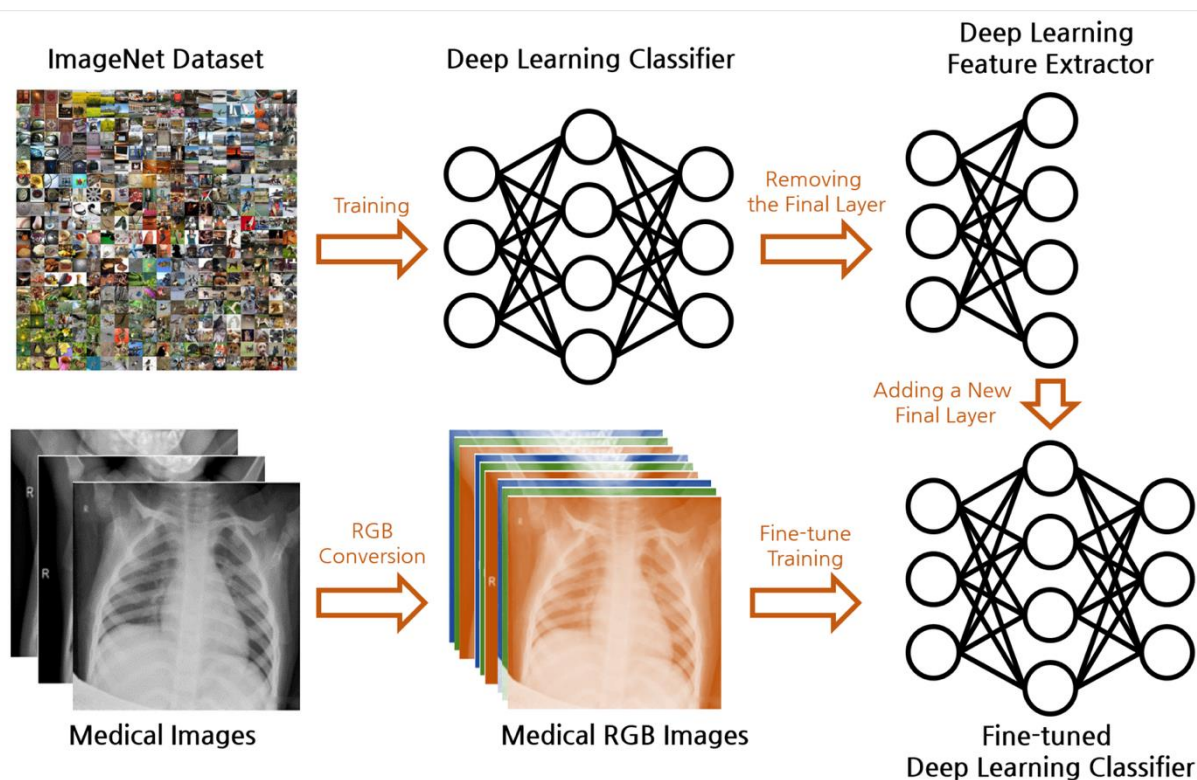


Рисунок 1.8 – Використання Transfer Learning для аналізу медичних зображень у моделях глибокого навчання

Застосування Transfer Learning дозволяє суттєво зменшити обсяг необхідних навчальних даних та скоротити час навчання моделей, що є особливо важливим для медичних задач, де доступність великих розмічених наборів даних часто обмежена. Завдяки цьому підхід широко використовується у задачах автоматичної діагностики, зокрема для виявлення патологій легенів, пухлин головного мозку, офтальмологічних захворювань та інших медичних станів.

1.3.3 Використання нейронних мереж у системах підтримки медичної діагностики

Нейронні мережі є одним із ключових інструментів штучного інтелекту, що широко застосовується у системах підтримки медичної діагностики. Завдяки здатності моделювати складні нелінійні залежності між ознаками даних, нейронні мережі дозволяють аналізувати різноманітні медичні параметри та виявляти закономірності, що можуть бути пов'язані з розвитком певних захворювань. У медичних інформаційних системах такі моделі використовуються для аналізу клінічних показників, результатів лабораторних досліджень, медичних зображень та іншого.

Типова структура нейронної мережі (рисунок 1.9) складається з вхідного шару, одного або кількох прихованих шарів та вихідного шару. На вхід моделі подаються медичні дані пацієнта, зокрема результати обстежень або показники життєдіяльності організму. У прихованих шарах виконується обробка інформації та формування внутрішніх представлень ознак, що дозволяє моделі визначати складні взаємозв'язки між параметрами. У системах підтримки медичної діагностики нейронні мережі можуть використовуватися для різних задач, зокрема для класифікації захворювань, прогнозування розвитку патологій та оцінювання ризику виникнення ускладнень.

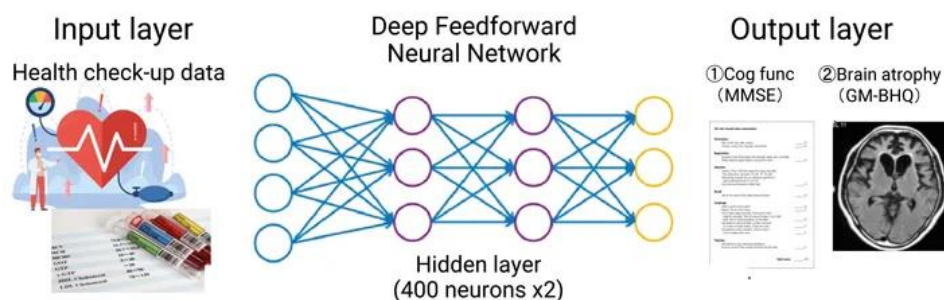


Рисунок 1.9 – Приклад структури нейронної мережі для аналізу клінічних даних у системах підтримки медичної діагностики

Наприклад, на основі клінічних даних пацієнта модель може визначати ймовірність наявності певного захворювання або прогнозувати можливий перебіг хвороби. Це дозволяє лікарям отримувати додаткову аналітичну інформацію, яка може бути використана під час прийняття клінічних рішень.

Важливою перевагою нейронних мереж є їх здатність працювати з великими та складними наборами даних, що характерно для сучасних медичних інформаційних систем. Разом з тим використання таких моделей потребує ретельної перевірки результатів та їх клінічної інтерпретації, оскільки медичні рішення мають безпосередній вплив на здоров'я пацієнтів. Тому нейронні мережі у системах медичної діагностики розглядаються як інструмент підтримки лікаря, що допомагає підвищити точність і обґрунтованість прийнятих рішень.

1.4 Мультимодальні підходи у медичній діагностиці

У сучасній медичній практиці дедалі більшого значення набуває використання мультимодальних підходів до аналізу медичних даних. Під мультимодальністю розуміють поєднання різних типів інформації, що характеризують стан пацієнта, з метою отримання більш повної та точної

діагностичної картини. У медичних системах такими джерелами даних можуть бути медичні зображення, електронні медичні записи, фізіологічні сигнали, результати лабораторних досліджень та інші клінічні показники. Поєднання різних типів інформації дозволяє отримати більш комплексне уявлення про стан здоров'я пацієнта.

Мультимодальні системи штучного інтелекту (рисунок 1.10) здатні об'єднувати дані з різних джерел у єдину модель аналізу. Це досягається за допомогою спеціальних алгоритмів, які виконують інтеграцію різнорідних даних на етапі підготовки інформації або безпосередньо у процесі навчання моделей. У результаті формується узагальнене представлення даних, що враховує різні аспекти медичної інформації та дозволяє підвищити точність прогнозування.

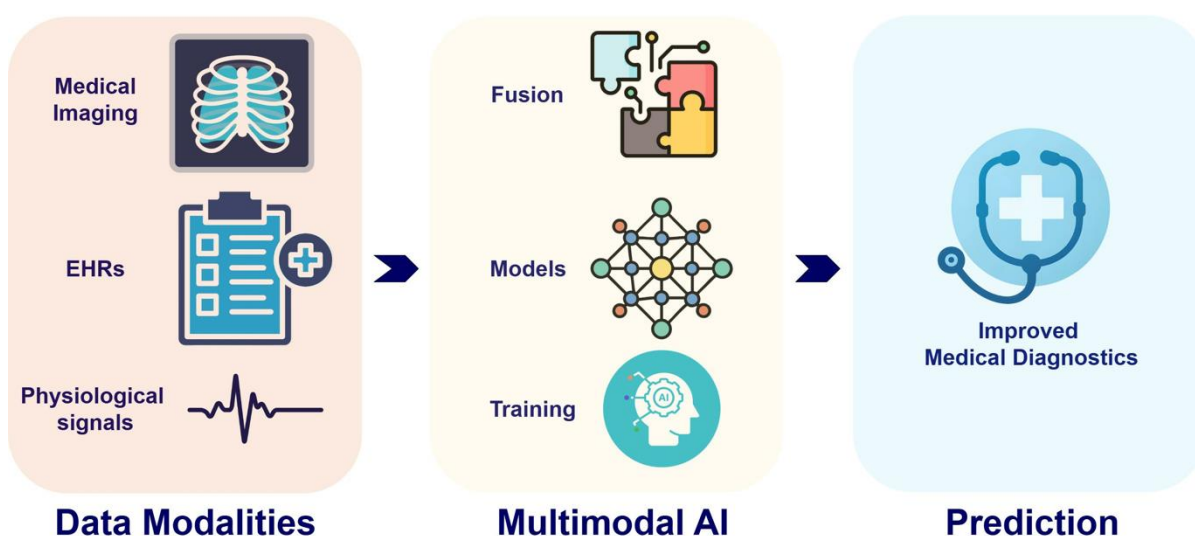


Рисунок 1.10 – Використання мультимодального підходу для інтеграції різних типів медичних даних у системах інтелектуальної діагностики

Однією з основних переваг мультимодальних моделей є їх здатність враховувати взаємозв'язки між різними типами медичних даних. Наприклад, поєднання результатів медичної візуалізації з клінічними показниками та фізіологічними сигналами дозволяє отримати більш точну

оцінку стану пацієнта. Це особливо важливо у складних діагностичних задачах, де використання лише одного типу даних може бути недостатнім для формування точного діагнозу.

Застосування мультимодальних підходів сприяє підвищенню ефективності систем підтримки прийняття клінічних рішень. Інтеграція різних джерел медичної інформації дозволяє створювати більш точні моделі прогнозування захворювань, визначати ризики розвитку патологій та формувати індивідуальні рекомендації щодо лікування. Таким чином, мультимодальні методи штучного інтелекту відкривають нові можливості для розвитку сучасних систем інтелектуальної медичної діагностики.

1.4.1 Поняття мультимодальних даних у медицині

У сучасній медицині інформація про стан пацієнта формується з великої кількості різних джерел, що відрізняються за форматом, структурою та способом отримання. Такі різноманітні джерела інформації прийнято називати мультимодальними даними. У медичному контексті до мультимодальних даних належать медичні зображення, електронні медичні записи, результати лабораторних аналізів, фізіологічні сигнали, генетична інформація, а також текстові клінічні записи лікарів. Використання різних типів даних дозволяє сформувати більш повну характеристику стану здоров'я пацієнта.

Особливістю мультимодальних даних є те, що кожен тип інформації відображає різні аспекти медичного стану пацієнта. Наприклад, медичні зображення дозволяють оцінювати анатомічні структури органів, фізіологічні сигнали відображають функціональний стан організму, а текстові медичні записи містять клінічні висновки та спостереження лікарів. Поєднання таких даних дозволяє отримати більш комплексне уявлення про перебіг захворювання та підвищити точність медичної діагностики.

У системах штучного інтелекту мультимодальні дані використовуються для побудови моделей, здатних одночасно аналізувати декілька типів медичної інформації. Для цього застосовуються спеціальні методи інтеграції даних, які дозволяють об'єднати різні джерела інформації у спільному аналітичному просторі. Це створює можливість виявлення складних взаємозв'язків між різними характеристиками медичних даних.

Таким чином, мультимодальні дані відіграють важливу роль у розвитку сучасних систем інтелектуальної діагностики. Їх використання дозволяє підвищити точність аналізу медичної інформації, покращити процес прийняття клінічних рішень та створити передумови для розвитку персоналізованої медицини.

1.4.2 Поєднання різних типів медичної інформації у системах ШІ

У системах штучного інтелекту для медичної діагностики важливе значення має інтеграція різних типів медичних даних, що дозволяє формувати більш повну модель стану пацієнта. Оскільки кожен тип даних відображає лише окремий аспект медичної інформації, їх поєднання дозволяє отримати більш точні результати аналізу. Наприклад, медичні зображення можуть надавати інформацію про анатомічні зміни, тоді як лабораторні показники або клінічні записи дозволяють оцінити функціональний стан організму.

Для поєднання різних типів медичної інформації у системах штучного інтелекту застосовуються спеціальні методи інтеграції даних. Один із підходів полягає у попередньому об'єднанні різних характеристик у спільному наборі ознак перед навчанням моделі. Інший підхід передбачає використання окремих моделей для аналізу різних типів даних з подальшим об'єднанням результатів їх роботи на етапі прийняття рішення. Такі методи дозволяють враховувати специфіку кожного типу даних та підвищувати ефективність побудованих моделей.

У сучасних дослідженнях також активно застосовуються глибокі нейронні мережі, які можуть обробляти різні типи даних у межах єдиної архітектури. Наприклад, одна частина моделі може аналізувати медичні зображення, тоді як інша частина обробляє текстову або числову інформацію з медичних записів. У подальшому ці дані інтегруються у спільному просторі ознак, що дозволяє моделі враховувати взаємозв'язки між різними характеристиками медичної інформації.

Поєднання різних типів медичних даних у системах штучного інтелекту дозволяє підвищити точність діагностичних моделей та забезпечити більш комплексний аналіз стану пацієнта. Завдяки цьому такі системи можуть використовуватися для раннього виявлення захворювань, оцінювання ризику розвитку патологій та формування рекомендацій щодо лікування.

1.4.3 Мультиmodalьні моделі для підвищення точності діагностики

Мультиmodalьні моделі штучного інтелекту мають суттєві переваги у задачах медичної діагностики, оскільки дозволяють одночасно аналізувати декілька типів медичних даних. У традиційних підходах діагностика часто базується на одному джерелі інформації, наприклад на медичних зображеннях або лабораторних показниках. Однак такий підхід може бути обмеженим, оскільки окремий тип даних не завжди відображає повну клінічну картину захворювання. Мультиmodalьні моделі дозволяють об'єднувати різні джерела інформації та формувати більш повний набір ознак для аналізу.

Однією з головних переваг мультиmodalьних моделей є підвищення точності діагностики. Поєднання різних типів даних дозволяє враховувати взаємозв'язки між анатомічними, функціональними та клінічними характеристиками організму.

1.5 Постановка задачі

У межах даної роботи розглядається задача підвищення ефективності автоматизованої багатоміткової класифікації рентгенівських знімків грудної клітки для підтримки прийняття рішень у медичній діагностиці. Особливістю задачі є необхідність одночасного виявлення кількох можливих патологій на одному зображенні, що потребує моделі, здатної враховувати як локальні візуальні ознаки, так і глобальні залежності між різними ділянками рентгенівського знімка.

Основною науковою задачею роботи є розробка гібридної архітектури класифікації рентгенівських знімків, яка поєднує згортковий компонент для виділення локальних ознак зображення та трансформерний компонент для аналізу глобальних взаємозв'язків між цими ознаками. Такий підхід має забезпечити більш повне представлення медичного зображення порівняно з окремим використанням класичних CNN, ResNet або Vision Transformer.

Для обґрунтування ефективності запропонованої гібридної архітектури необхідно провести порівняльний аналіз її результатів із базовими моделями CNN, ResNet та Vision Transformer. Оцінювання доцільно виконати за показниками точності класифікації, AUC, ROC-кривими та стабільністю навчання, що дозволить визначити, чи забезпечує запропонована архітектура покращення якості класифікації та чи може вона бути використана як основа для систем підтримки прийняття рішень у медичній діагностиці.

2 ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ

2.1 Багатоміткова класифікація рентгенівських знімків грудної клітки

Багатоміткова класифікація [5] є одним із ключових підходів у задачах аналізу медичних зображень, зокрема рентгенівських знімків грудної клітки. На відміну від класичної постановки задачі класифікації, де кожному зображенню відповідає лише один клас, у багатомітковому підході передбачається можливість одночасної наявності кількох патологій. Рентгенівські знімки грудної клітки можуть містити ознаки різних захворювань, таких як пневмонія, туберкульоз, онкологічні ураження та інші патологічні стани легень (рисунок 2.1).

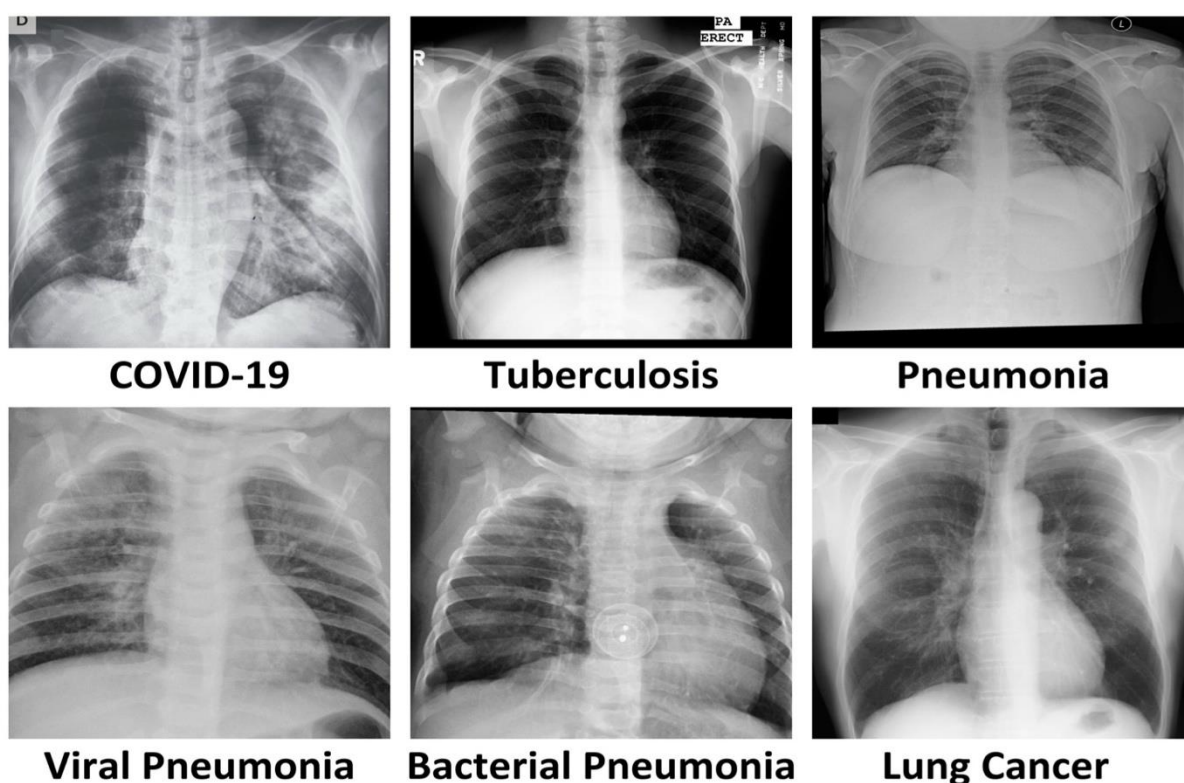


Рисунок 2.1 – Приклади рентгенівських знімків грудної клітки з різними патологіями легень

При цьому один і той самий знімок може демонструвати кілька патологічних процесів одночасно, що обумовлює необхідність застосування багатоміткових моделей класифікації. Це дозволяє більш точно моделювати процес медичної діагностики.

У сучасних дослідженнях значна увага приділяється задачам класифікації медичних зображень, зокрема виявленню патологічних станів на рентгенівських знімках. При цьому традиційні підходи до класифікації не завжди є достатніми для коректного відображення реальних умов, що обумовлює необхідність використання більш гнучких моделей представлення результатів.

У межах даного підрозділу розглядається загальна характеристика задачі багатоміткової класифікації рентгенівських знімків грудної клітки, яка дозволяє враховувати складність і багатофакторність медичних даних.

2.1.1 Сутність багатоміткової класифікації медичних зображень

Задача багатоміткової класифікації [5] є узагальненням класичних підходів до класифікації, що дозволяє одному об'єкту відповідати одразу декільком класам. У контексті медичних зображень це означає, що для одного рентгенівського знімка необхідно визначити наявність або відсутність кількох патологічних станів одночасно. Такий підхід є більш адекватним відображенням реальних умов клінічної практики, де захворювання часто мають комбінований характер.

На відміну від бінарної класифікації, яка передбачає лише два можливих стани, та багатокласової класифікації, де об'єкт належить лише до одного класу, багатоміткова класифікація дозволяє формувати множину міток для кожного зображення (рисунки 2.2). Це означає, що класи не є взаємовиключними, а можуть співіснувати в межах одного зразка.

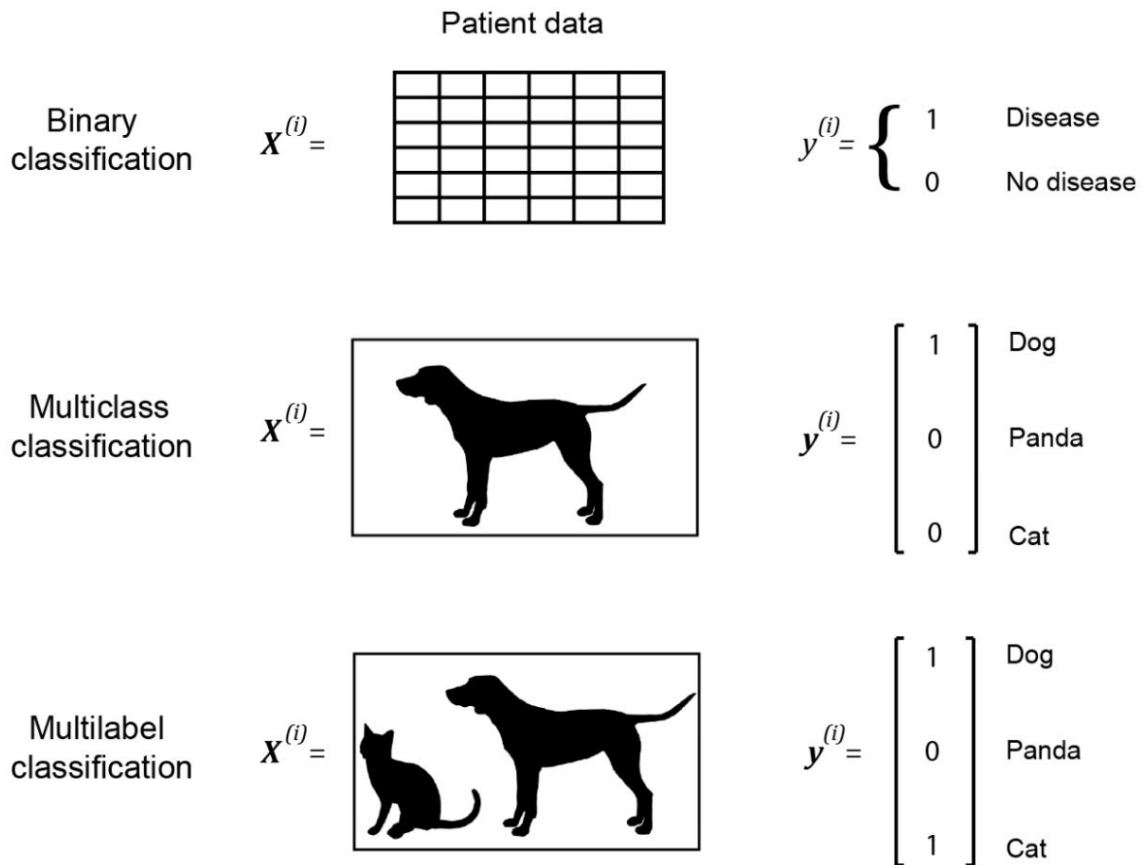


Рисунок 2.2 – Порівняння підходів до класифікації

Формально задача багатоміткової класифікації полягає у відображенні вхідного простору зображень у простір бінарних векторів міток. Кожному зображенню ставиться у відповідність вектор, розмірність якого дорівнює кількості можливих класів, а значення кожного елемента визначає наявність відповідної патології. Таким чином, модель повинна незалежно оцінювати кожен клас, що суттєво ускладнює процес навчання порівняно з традиційними підходами.

У задачах медичної діагностики це призводить до необхідності використання спеціалізованих функцій втрат та архітектур моделей. Зокрема, замість функції softmax, яка застосовується у багатокласовій класифікації, використовується сигмоїдна функція активації для кожного вихідного нейрона. Це дозволяє інтерпретувати вихід моделі як незалежні ймовірності належності до кожного класу.

Складність багатоміткової класифікації також обумовлена тим, що між класами можуть існувати приховані залежності. Наприклад, наявність однієї патології може підвищувати ймовірність іншої, або навпаки – ускладнювати її виявлення. Проте більшість моделей розглядають класи незалежно, що створює додаткові виклики для підвищення точності класифікації.

Таким чином, сутність задачі багатоміткової класифікації медичних зображень полягає у визначенні множини патологічних станів для кожного зображення на основі аналізу його візуальних ознак. Ця задача є більш складною порівняно з традиційними підходами до класифікації, однак саме вона дозволяє створювати більш точні та практично значущі системи підтримки прийняття рішень у медицині.

2.1.2 Особливості представлення класів у рентгенівських знімках грудної клітки

Представлення класів у задачах аналізу рентгенівських знімків грудної клітки має ряд специфічних особливостей, що відрізняють їх від класичних задач класифікації зображень. Це зумовлено як природою медичних даних, так і складністю візуальних проявів патологій, які часто не мають чітких меж і можуть частково перекриватися.

Однією з основних характеристик є невзаємовиключність класів, що означає можливість одночасної наявності кількох захворювань на одному зображенні (рисунк 2.3). У такому випадку кожному знімку відповідає не один клас, а набір міток, які відображають різні патологічні стани. Це принципово відрізняє багатоміткову класифікацію від багатокласової, де кожному об'єкту відповідає лише одна категорія. Крім того, це ускладнює процес навчання, оскільки модель повинна враховувати можливі комбінації патологій та їх спільні прояви.

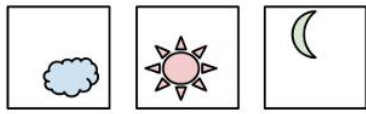
	Multi-Class	Multi-Label
	C = 3 Samples  Labels (t) [0 0 1] [1 0 0] [0 1 0]	Samples  Labels (t) [1 0 1] [0 1 0] [1 1 1]

Рисунок 2.3 – Порівняння прикладів представлення міток у багатокласовій та багатомітковій класифікації

Додатково особливості представлення класів проявляються у способі формування вихідного простору моделі. У багатомітковому підході кожен клас розглядається незалежно, а результат класифікації формується як набір бінарних значень або ймовірностей (рисунок 2.4). Такий підхід дозволяє моделі оцінювати наявність кожного захворювання окремо, що є необхідним для коректного аналізу медичних зображень.

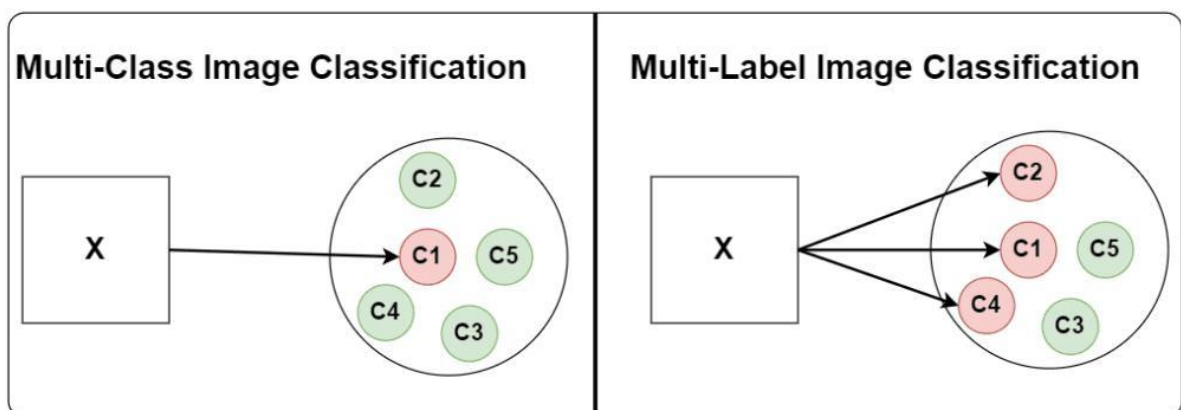


Рисунок 2.4 – Схема відповідності вхідного зображення множині класів у багатокласовому та багатомітковому підходах

Важливою особливістю є також складність інтерпретації візуальних ознак. Багато патологій мають подібні рентгенологічні прояви, такі як затемнення або зміни структури легеневої тканини. Це ускладнює як процес анотації даних, так і побудову моделей, здатних чітко розмежовувати класи.

Окрім цього, класи у медичних наборах даних можуть мати корельований характер. Деякі захворювання часто виникають разом або мають спільні причини, що створює залежності між мітками. Це накладає додаткові вимоги на моделі, які повинні враховувати не лише окремі ознаки, але й взаємозв'язки між класами.

Таким чином, представлення класів у рентгенівських знімках грудної клітки характеризується невзаємовиключністю, незалежною оцінкою кожного класу та складністю візуального розмежування патологій. Ці особливості визначають специфіку побудови моделей багатоміткової класифікації та підходів до їх навчання.

2.1.3 Специфіка застосування багатоміткової класифікації у задачах медичної діагностики

Застосування багатоміткової класифікації у задачах медичної діагностики має ряд специфічних особливостей, що обумовлені високою складністю медичних даних та критичною важливістю точності результатів. На відміну від загальних задач комп'ютерного зору, помилки у медичних системах можуть безпосередньо впливати на прийняття клінічних рішень, що висуває підвищені вимоги до якості моделей.

Однією з ключових особливостей є необхідність одночасного виявлення кількох патологій, які можуть мати різну локалізацію та ступінь прояву на зображенні (рисунок 2.5). Це означає, що модель повинна не лише визначити наявність захворювання, але й враховувати складну просторову структуру зображення.

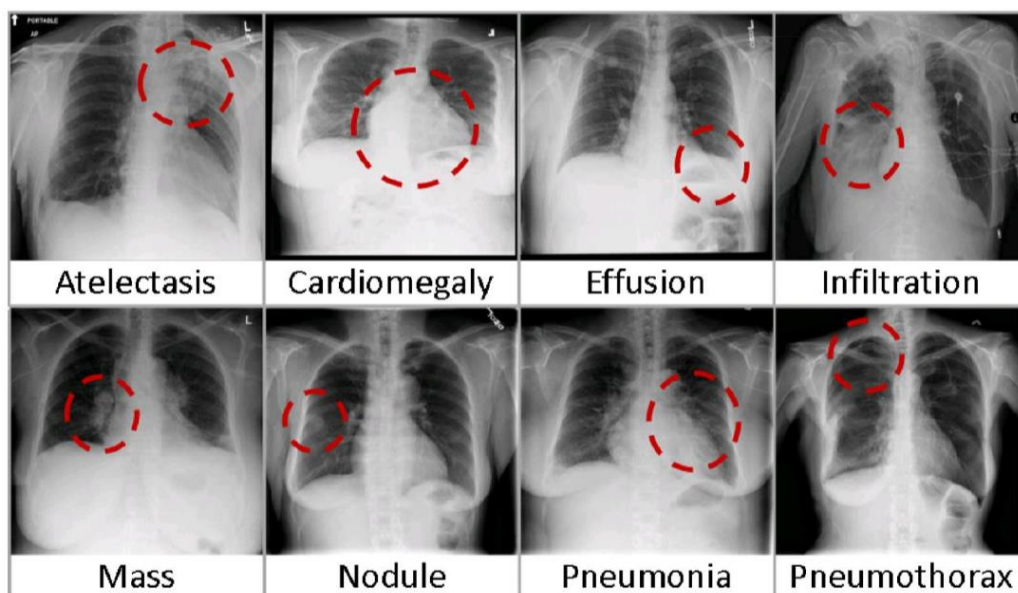


Рисунок 2.5 – Приклади локалізації різних патологій на рентгенівських знімках грудної клітки

Важливим аспектом є також висока варіативність проявів захворювань. Одна і та сама патологія може виглядати по-різному залежно від стадії розвитку, індивідуальних особливостей пацієнта та якості зображення. Це ускладнює узагальнення моделей і вимагає використання великих та різноманітних навчальних вибірок.

Окрім цього, у медичних задачах важливу роль відіграє проблема інтерпретованості результатів. Лікарі повинні розуміти, на основі яких ознак модель зробила той чи інший висновок. Тому сучасні підходи часто доповнюються методами візуалізації уваги моделі, які дозволяють виділяти області зображення, що вплинули на класифікацію.

Ще однією специфічною рисою є наявність дисбалансу даних та обмеженого доступу до якісно розмічених медичних зображень. Збір і анотація таких даних потребують значних ресурсів і участі кваліфікованих фахівців, що обмежує обсяг доступних наборів даних для навчання моделей.

Таким чином, застосування багатоміткової класифікації у задачах медичної діагностики характеризується необхідністю врахування

просторової складності зображень, варіативності патологій, вимог до інтерпретованості та обмеженості даних. Усі ці фактори визначають особливості побудови ефективних систем підтримки прийняття рішень у медицині.

2.2 Класичні архітектури згорткових нейронних мереж

Згорткові нейронні мережі (рисунок 2.6) є одним із базових та найбільш ефективних підходів у задачах комп'ютерного зору, зокрема при обробці та аналізі зображень. Їх широке застосування обумовлене здатністю автоматично виділяти інформативні ознаки з вхідних даних без необхідності ручного конструювання характеристик.

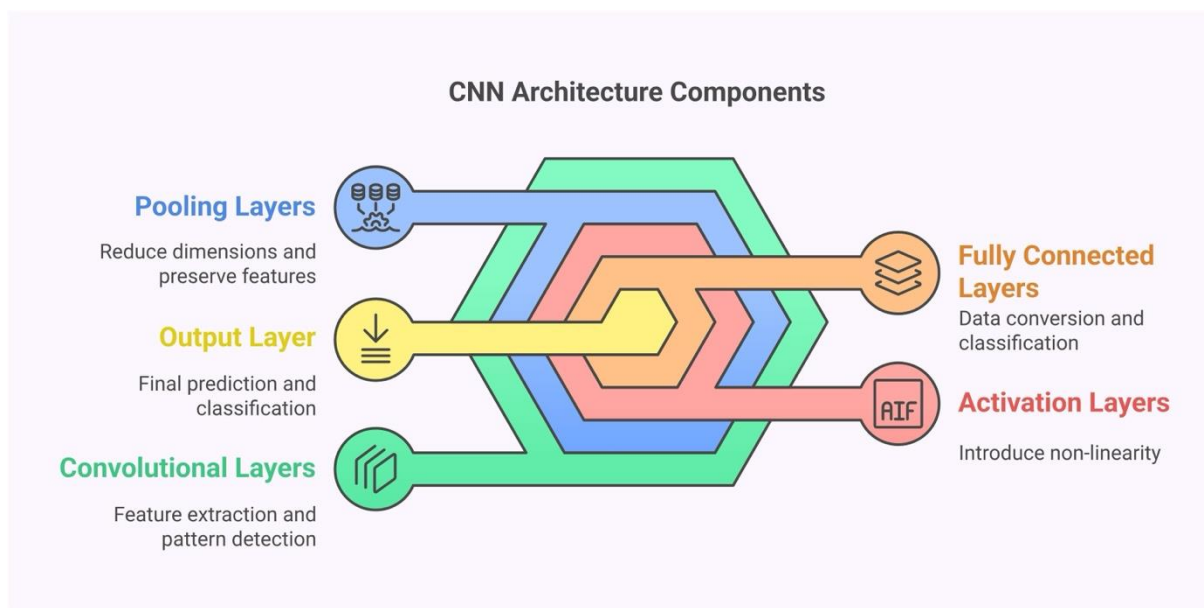


Рисунок 2.6 – Основні компоненти архітектури згорткової нейронної мережі

Основною ідеєю згорткових нейронних мереж є послідовне перетворення вхідного зображення у набір більш абстрактних представлень, які зберігають важливу інформацію про структуру об'єкта. Такий підхід

дозволяє ефективно працювати з високовимірними даними, зменшуючи їх розмірність і одночасно підвищуючи інформативність ознак.

Архітектура згорткових нейронних мереж складається з кількох типів шарів, кожен з яких виконує специфічну функцію у процесі обробки даних. Комбінація цих шарів дозволяє моделі поступово переходити від низькорівневих ознак до високорівневих абстракцій, що є ключовим фактором ефективності даного підходу.

Завдяки своїй структурі, згорткові нейронні мережі здатні враховувати просторові залежності у зображеннях, що є критично важливим для задач класифікації. Використання локальних фільтрів дозволяє виявляти характерні патерни незалежно від їх розташування, що підвищує стійкість моделі до варіацій у даних.

Таким чином, класичні архітектури згорткових нейронних мереж формують основу сучасних методів аналізу зображень і широко застосовуються як базові моделі у багатьох прикладних задачах, включаючи медичну діагностику.

2.2.1 Архітектура CNN

Архітектура згорткової нейронної мережі [6] побудована за принципом послідовного перетворення вхідного зображення у більш компактне та інформативне представлення, що дозволяє виконувати подальшу класифікацію.

Типова архітектура CNN складається з декількох основних блоків, що включають вхідний шар, згорткові шари, підвибіркові шари та повнозв'язні шари (рисунок 2.7). Вхідний шар приймає зображення у вигляді матриці пікселів, після чого відбувається його поступова обробка за допомогою згорткових операцій. У результаті формується набір карт ознак, які відображають локальні патерни зображення.

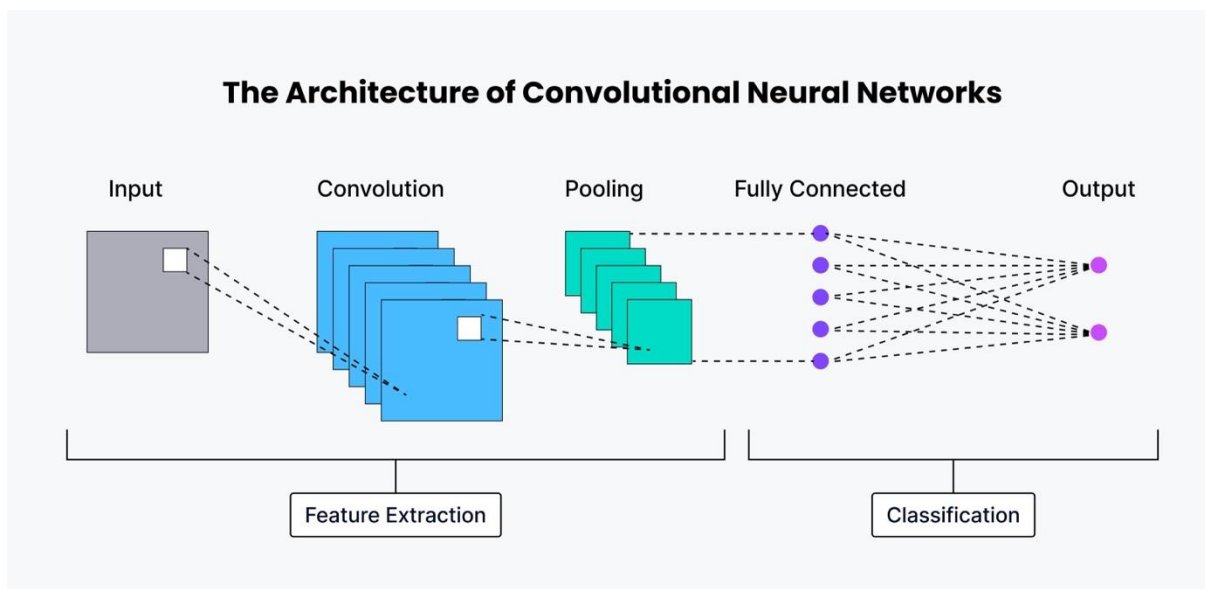


Рисунок 2.7 – Узагальнена архітектура згорткової нейронної мережі

Згорткові шари є ключовим елементом архітектури, оскільки саме вони відповідають за виділення ознак. У цих шарах застосовуються фільтри, які послідовно проходять по зображенню, виявляючи характерні структури, такі як контури, текстури або більш складні елементи. Кожен наступний шар дозволяє формувати більш абстрактні представлення, що є важливим для розпізнавання складних об'єктів.

Після згорткових шарів зазвичай використовуються шари підвибірки, які зменшують розмірність даних і дозволяють виділити найбільш значущі ознаки. Це сприяє зниженню обчислювальної складності та підвищенню стійкості моделі до незначних змін у зображенні.

На завершальному етапі архітектури використовуються повнозв'язні шари, які виконують функцію класифікації. Вони приймають сформовані ознаки та перетворюють їх у вихідний вектор, що відповідає ймовірностям належності до певних класів (рисунок 2.8). Таким чином, відбувається перехід від просторового представлення даних до прийняття рішення.

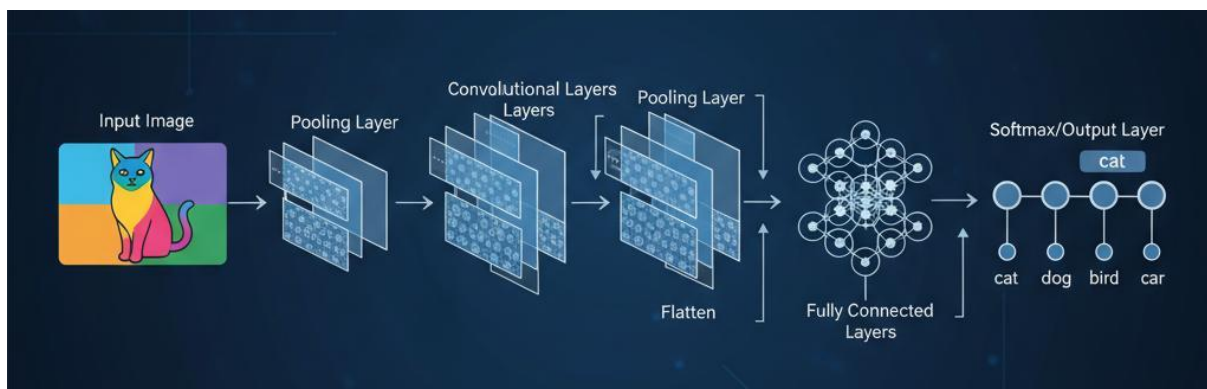


Рисунок 2.8 – Приклад побудови CNN для задачі класифікації зображень

Загалом архітектура згорткової нейронної мережі поєднує етапи виділення ознак та класифікації в єдину систему, що дозволяє ефективно обробляти зображення різної складності. Така структура є основою більшості сучасних моделей комп'ютерного зору.

2.2.2 Основні шари та принцип формування ознак у CNN

Основу згорткових нейронних мереж становлять спеціалізовані шари, кожен з яких виконує окрему функцію у процесі обробки зображення. Сукупність цих шарів забезпечує поступове перетворення вхідних даних у набір інформативних ознак, придатних для подальшої класифікації.

Ключовим елементом є згортковий шар, який реалізує операцію згортки між вхідним зображенням та набором фільтрів. Кожен фільтр дозволяє виділяти певний тип локальних ознак, таких як краї, текстури або інші характерні структури. У результаті застосування фільтрів формується набір карт ознак, що відображають різні аспекти вхідного зображення.

Після згорткових шарів зазвичай застосовуються функції активації, які вводять нелінійність у модель. Це дозволяє нейронній мережі моделювати складні залежності між ознаками та підвищує її здатність до узагальнення. Найчастіше використовується функція ReLU, яка забезпечує ефективне навчання глибоких мереж.

Наступним важливим компонентом є шар підвибірки, або pooling, який використовується для зменшення розмірності карт ознак. Це дозволяє знизити обчислювальні витрати та зробити модель більш стійкою до незначних змін у вхідних даних. Найпоширенішим варіантом є операція max pooling, яка вибирає максимальне значення у межах локального вікна.

Принцип формування ознак у CNN полягає у їх ієрархічному ускладненні. На початкових шарах виділяються прості ознаки, такі як контури та границі об'єктів. У середніх шарах формується представлення текстур і локальних структур, а на глибших рівнях – складніші абстрактні ознаки, що відповідають цілим об'єктам або їх частинам (рисунок 2.9). Такий підхід дозволяє моделі ефективно аналізувати зображення різної складності.

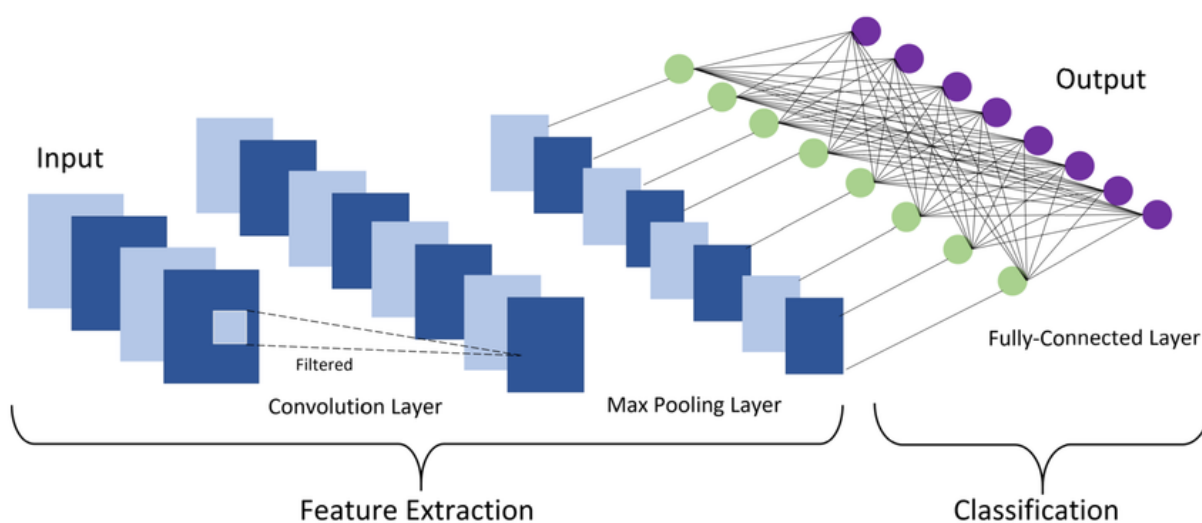


Рисунок 2.9 – Послідовність формування ознак у згортковій нейронній мережі

На завершальному етапі сформовані ознаки передаються до повнозв'язних шарів, де відбувається їх узагальнення та прийняття рішення щодо належності до певного класу. Таким чином, кожен тип шару відіграє важливу роль у побудові ефективної системи класифікації.

2.2.3 Особливості застосування CNN у задачах класифікації рентгенівських знімків

Застосування згорткових нейронних мереж у задачах класифікації рентгенівських знімків грудної клітки має ряд специфічних особливостей, що обумовлені як природою медичних зображень, так і вимогами до точності діагностики. CNN ефективно працюють із просторовими даними, що дозволяє використовувати їх для автоматичного виявлення патологічних змін у легенях.

Однією з ключових особливостей є здатність CNN виділяти локальні ознаки, які відповідають патологічним змінам у тканинах. У рентгенівських знімках такі зміни можуть проявлятися у вигляді затемнень, ущільнень або деформацій структури легенів, які часто мають нечіткі межі (рисунок 2.10). Завдяки використанню згорткових фільтрів модель може ефективно виявляти ці патерни незалежно від їх розташування на зображенні.

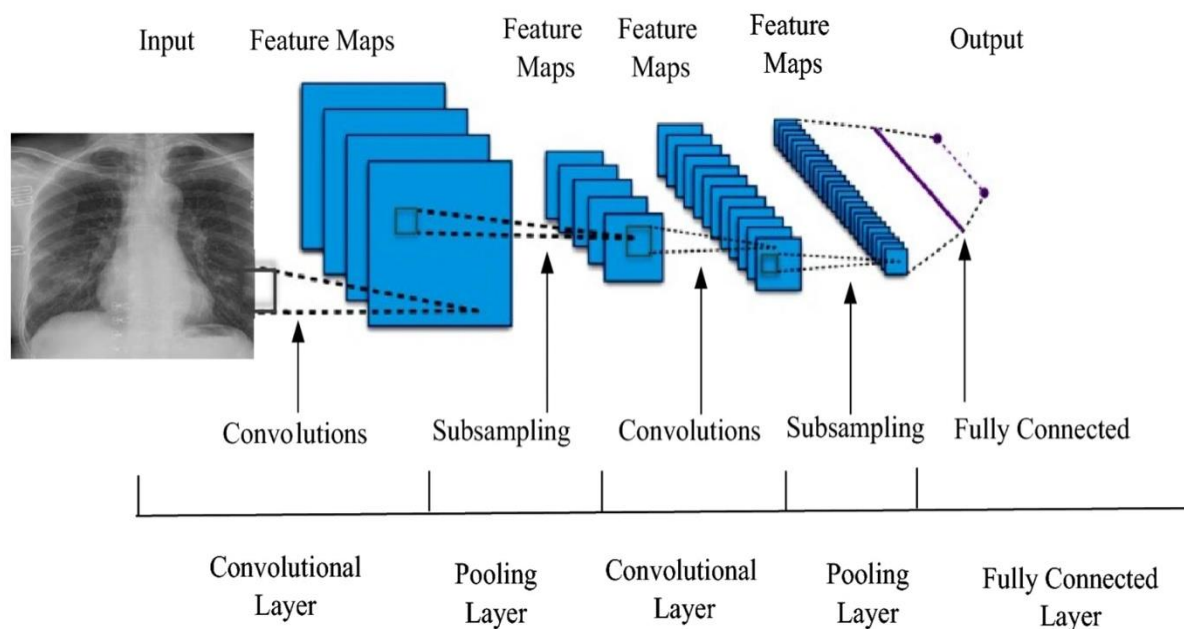


Рисунок 2.10 – Приклад застосування CNN для класифікації рентгенівських знімків грудної клітки

Важливим аспектом є також необхідність роботи з високим рівнем шуму та варіативністю даних. Рентгенівські знімки можуть відрізнятися за якістю, контрастністю та умовами отримання, що ускладнює процес навчання моделей. У зв'язку з цим широко застосовуються методи попередньої обробки зображень та аугментації даних, які дозволяють підвищити стійкість моделей до змін у вхідних даних.

Ще однією особливістю є необхідність врахування дисбалансу класів, характерного для медичних наборів даних. Деякі патології представлені значно меншою кількістю прикладів, що може призводити до зниження точності їх розпізнавання. Для вирішення цієї проблеми використовуються спеціальні методи балансування даних або модифікації функцій втрат.

Крім того, у задачах медичної діагностики важливу роль відіграє інтерпретованість результатів. Хоча CNN забезпечують високу точність, їх рішення можуть бути складними для пояснення. Тому часто застосовуються додаткові методи візуалізації, які дозволяють визначити області зображення, що найбільше вплинули на результат класифікації.

Таким чином, використання згорткових нейронних мереж у задачах класифікації рентгенівських знімків грудної клітки характеризується здатністю ефективно працювати зі складними візуальними даними, але водночас потребує врахування специфічних факторів, таких як варіативність зображень, дисбаланс класів та вимоги до пояснюваності результатів.

2.3 Архітектура ResNet як розвиток згорткових нейронних мереж

Зі збільшенням глибини згорткових нейронних мереж виникає проблема зниження ефективності їх навчання, що обмежує можливість подальшого підвищення точності моделей. Традиційні CNN при значному збільшенні кількості шарів можуть демонструвати деградацію якості, що пов'язано з труднощами оптимізації глибоких архітектур.

Архітектура ResNet була запропонована як підхід до подолання цих обмежень шляхом модифікації структури нейронної мережі. Основною ідеєю є використання спеціальних з'єднань між шарами, які дозволяють передавати інформацію безпосередньо між різними рівнями мережі, мінаючи проміжні перетворення. Це забезпечує більш стабільний процес навчання та дозволяє будувати значно глибші моделі.

У структурі ResNet використовуються так звані residual-блоки, які формують основу архітектури та забезпечують ефективну передачу ознак між шарами (рисунк 2.11). Такі блоки дозволяють мережі не лише вивчати нові ознаки, але й зберігати вже отриману інформацію, що є важливим для побудови глибоких моделей.

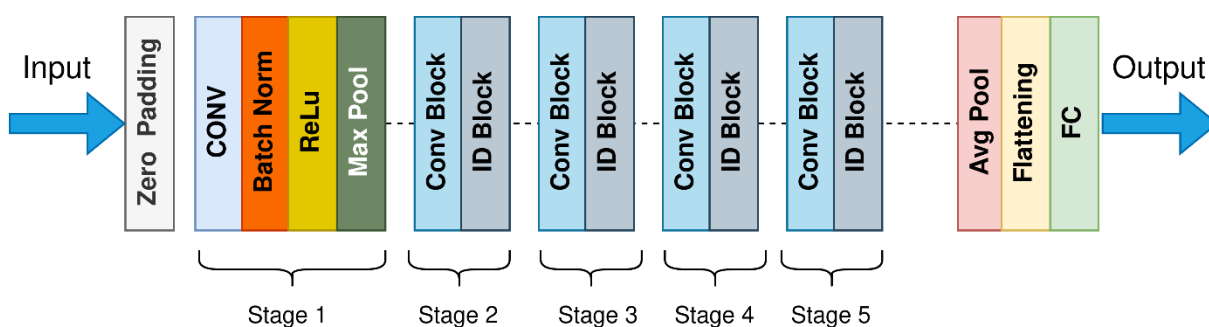


Рисунок 2.11 – Узагальнена архітектура ResNet з використанням блоків residual learning

Ще однією особливістю є модульна побудова архітектури, яка складається з послідовності однакових блоків. Це спрощує масштабування моделі та дозволяє адаптувати її до задач різної складності шляхом зміни кількості блоків. Завдяки цьому ResNet може ефективно працювати як у базових, так і у складних задачах комп'ютерного зору. Таким чином, архітектура ResNet є розвитком класичних згорткових нейронних мереж, що дозволяє подолати обмеження глибоких моделей.

2.3.1 Проблема зникнення градієнта в глибоких мережах

Однією з основних проблем навчання глибоких нейронних мереж є явище зникнення градієнта, яке виникає під час поширення помилки від вихідного шару до початкових шарів мережі. У процесі зворотного розповсюдження похибки значення градієнтів можуть поступово зменшуватися, що призводить до незначного оновлення ваг у початкових шарах.

Суть цієї проблеми полягає в тому, що при проходженні через велику кількість шарів градієнт множиться на похідні функцій активації, які часто мають значення менші за одиницю. У результаті цього відбувається експоненційне зменшення величини градієнта, що ускладнює ефективне навчання глибоких мереж. Особливо це проявляється у випадку використання сигмоїдальних функцій активації.

Наслідком зникнення градієнта є те, що початкові шари мережі практично не навчаються, оскільки їх ваги оновлюються вкрай повільно. Це призводить до того, що мережа не здатна ефективно виділяти низькорівневі ознаки, які є важливими для подальшого формування складних представлень.

На практиці це означає, що збільшення глибини мережі не завжди призводить до покращення результатів, а інколи навіть погіршує їх. Такий ефект відомий як проблема деградації, коли більш глибока модель демонструє гіршу точність порівняно з менш глибокою.

Ілюстрація цієї проблеми показує, що у звичайних глибоких мережах градієнт слабшає при проходженні до ранніх шарів, тоді як використання спеціальних механізмів дозволяє забезпечити більш стабільний потік градієнта (рисунок 2.12). Це стало передумовою для розробки нових архітектур, зокрема ResNet, які спрямовані на вирішення даної проблеми.

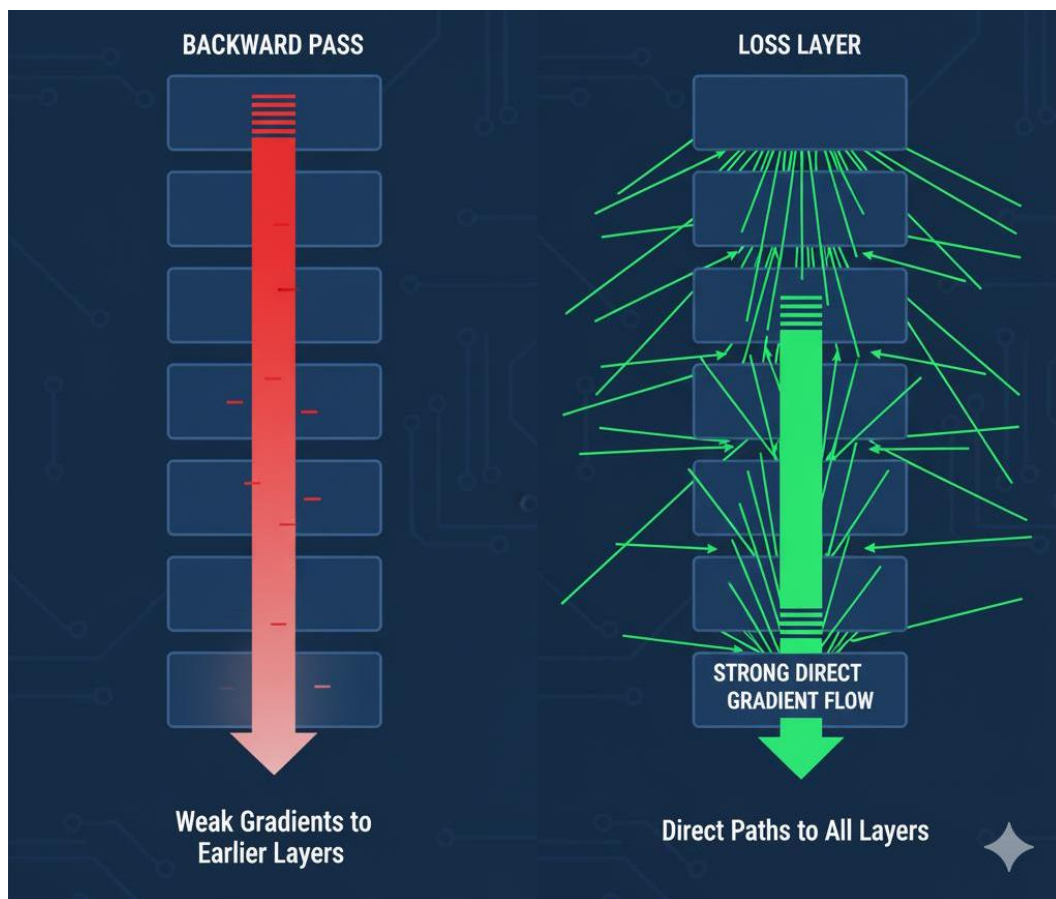


Рисунок 2.12 – Проблема зникнення градієнта та її подолання за рахунок прямих шляхів передачі градієнта

Таким чином, проблема зникнення градієнта є одним із ключових обмежень традиційних глибоких нейронних мереж, що ускладнює їх навчання та знижує ефективність. Її подолання стало важливим кроком у розвитку сучасних архітектур глибокого навчання.

2.3.2 Архітектура ResNet і механізм residual learning

Архітектура ResNet є розвитком класичних згорткових нейронних мереж і спрямована на підвищення ефективності навчання глибоких моделей. Вона базується на використанні стандартних згорткових шарів, однак доповнюється спеціальними структурними елементами, що дозволяють подолати обмеження традиційних підходів.

Загальна структура ResNet включає послідовність згорткових шарів, об'єднаних у блоки, які формують глибоку мережу для виділення ознак різного рівня складності (рисунок 2.13). На відміну від звичайних CNN, де інформація проходить строго послідовно, у ResNet реалізовано додаткові з'єднання між шарами, що забезпечують альтернативні шляхи передачі сигналу.

Основою архітектури є механізм residual learning, який полягає у навчанні не прямого відображення, а залишкової функції. Замість того щоб апроксимувати складну залежність, модель навчається визначати різницю між вхідними даними та бажаним результатом. У результаті вихід блоку визначається як сума вхідного сигналу та залишкової функції.

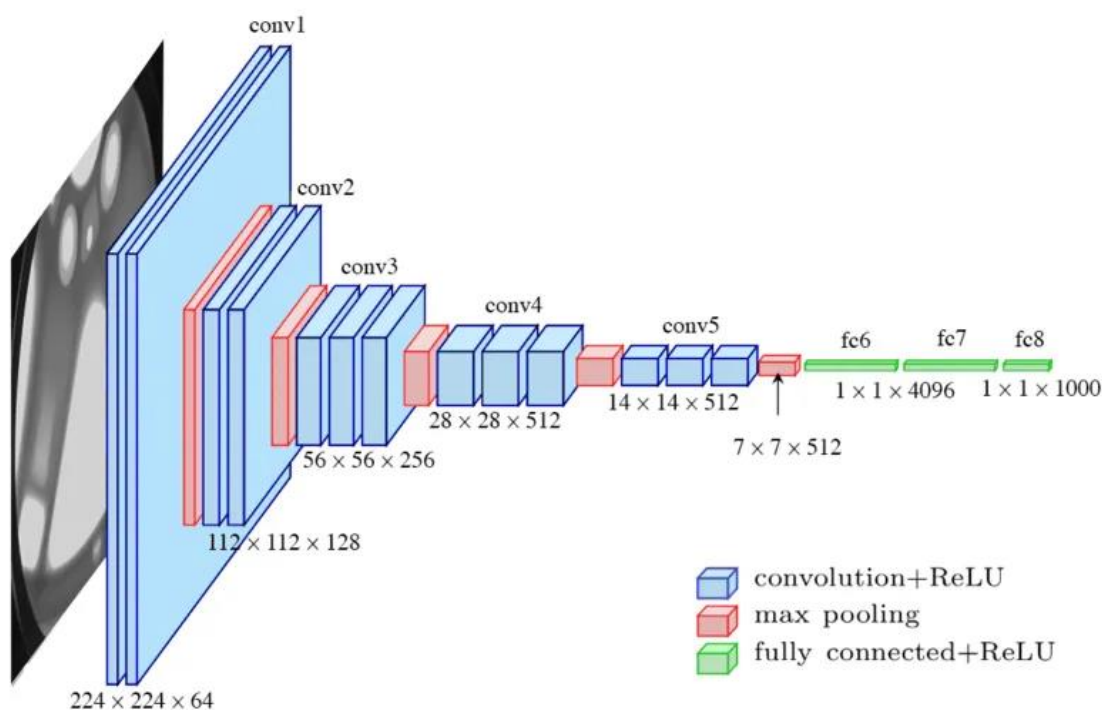


Рисунок 2.13 – Архітектура глибокої згорткової мережі як основа для побудови ResNet

Практична реалізація цього підходу здійснюється за допомогою residual-блоків, які містять основну гілку з кількома згортковими шарами та

обхідне з'єднання, що передає вхідні дані без змін (рисунк 2.14). На виході відбувається операція додавання, яка поєднує результати обох шляхів. Така структура дозволяє зберігати інформацію та забезпечує більш стабільне навчання.

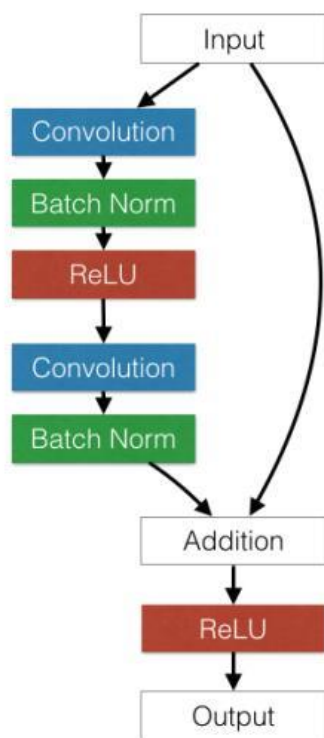


Figure 1. A ResNet basic block

Рисунок 2.14 – Базовий residual-блок із механізмом skip-з'єднання

Використання skip-з'єднань забезпечує ефективний потік градієнта під час зворотного розповсюдження помилки, що дозволяє уникнути проблеми його зникнення. Завдяки цьому стає можливим навчання значно глибших мереж порівняно з класичними архітектурами. Таким чином, архітектура ResNet поєднує класичні підходи до обробки зображень із новим принципом навчання, що базується на використанні залишкових зв'язків.

2.3.3 Особливості застосування ResNet у задачах медичної класифікації зображень

Застосування архітектури ResNet у задачах медичної класифікації зображень характеризується рядом специфічних особливостей, пов'язаних із природою медичних даних та вимогами до точності діагностики. Завдяки використанню механізму residual learning такі моделі демонструють високу ефективність при обробці складних зображень, зокрема рентгенівських знімків грудної клітки.

Однією з ключових переваг є здатність ResNet працювати з глибокими архітектурами без втрати якості навчання. Це дозволяє моделі виділяти як низькорівневі, так і високорівневі ознаки, що є важливим для аналізу медичних зображень, де патології можуть мати складну структуру та різний масштаб прояву.

Важливою особливістю є можливість інтеграції додаткових джерел інформації поряд із зображеннями. У задачах медичної діагностики часто використовуються не лише візуальні дані, але й інформація про пацієнта, така як вік, стать або умови отримання знімка. Це дозволяє підвищити точність класифікації за рахунок врахування контексту (рисунк 2.15).

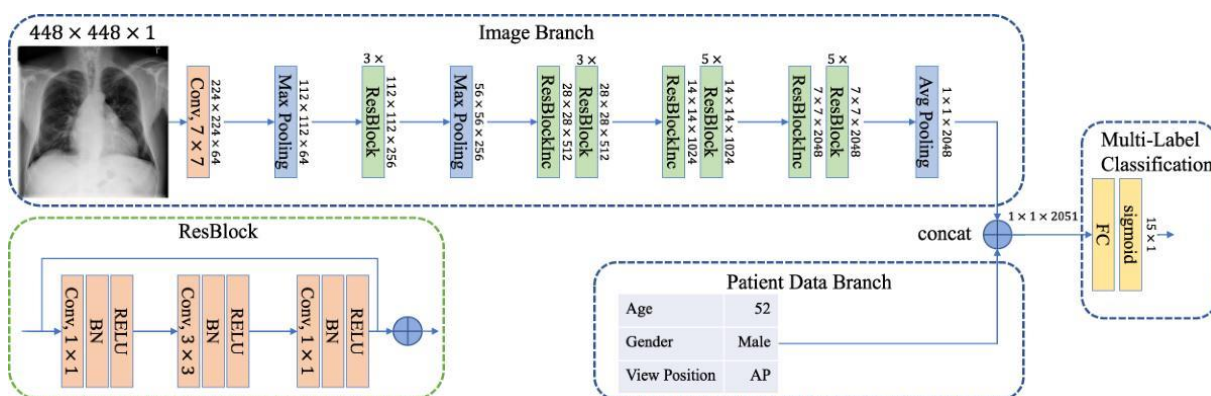


Рисунок 2.15 – Приклад застосування ResNet для багатоміткової класифікації медичних зображень

Також слід відзначити стійкість ResNet до варіативності вхідних даних. Рентгенівські знімки можуть суттєво відрізнятися за якістю, контрастністю та положенням пацієнта, проте глибокі residual-мережі здатні ефективно узагальнювати такі відмінності та зберігати високу точність класифікації.

Окрім цього, архітектура ResNet добре поєднується з методами трансферного навчання, що є важливим у медичних задачах, де доступ до великих розмічених наборів даних обмежений. Використання попередньо навчених моделей дозволяє значно скоротити час навчання та підвищити якість результатів.

Таким чином, застосування ResNet у задачах медичної класифікації зображень забезпечує високу точність, гнучкість у використанні різних типів даних та ефективність при роботі з глибокими архітектурами. Це робить даний підхід одним із найбільш перспективних для побудови сучасних систем підтримки прийняття рішень у медицині.

2.4 Трансформерний підхід у комп'ютерному зорі

У сучасних дослідженнях у галузі комп'ютерного зору спостерігається перехід від класичних згорткових нейронних мереж до трансформерних архітектур. Це пов'язано з обмеженнями CNN у моделюванні глобальних залежностей між різними частинами зображення та необхідністю більш гнучкого підходу до обробки візуальної інформації.

Трансформерний підхід, який спочатку був розроблений для задач обробки природної мови, продемонстрував високу ефективність завдяки використанню механізму самоуваги. Цей механізм дозволяє моделі аналізувати взаємозв'язки між різними елементами вхідних даних незалежно від їх розташування. У задачах комп'ютерного зору трансформери реалізуються через представлення зображення у вигляді послідовності фрагментів або патчів, які розглядаються як окремі елементи

для подальшої обробки (рисунок 2.16). Кожен такий фрагмент перетворюється у векторне представлення, після чого обробляється за допомогою блоків трансформера.

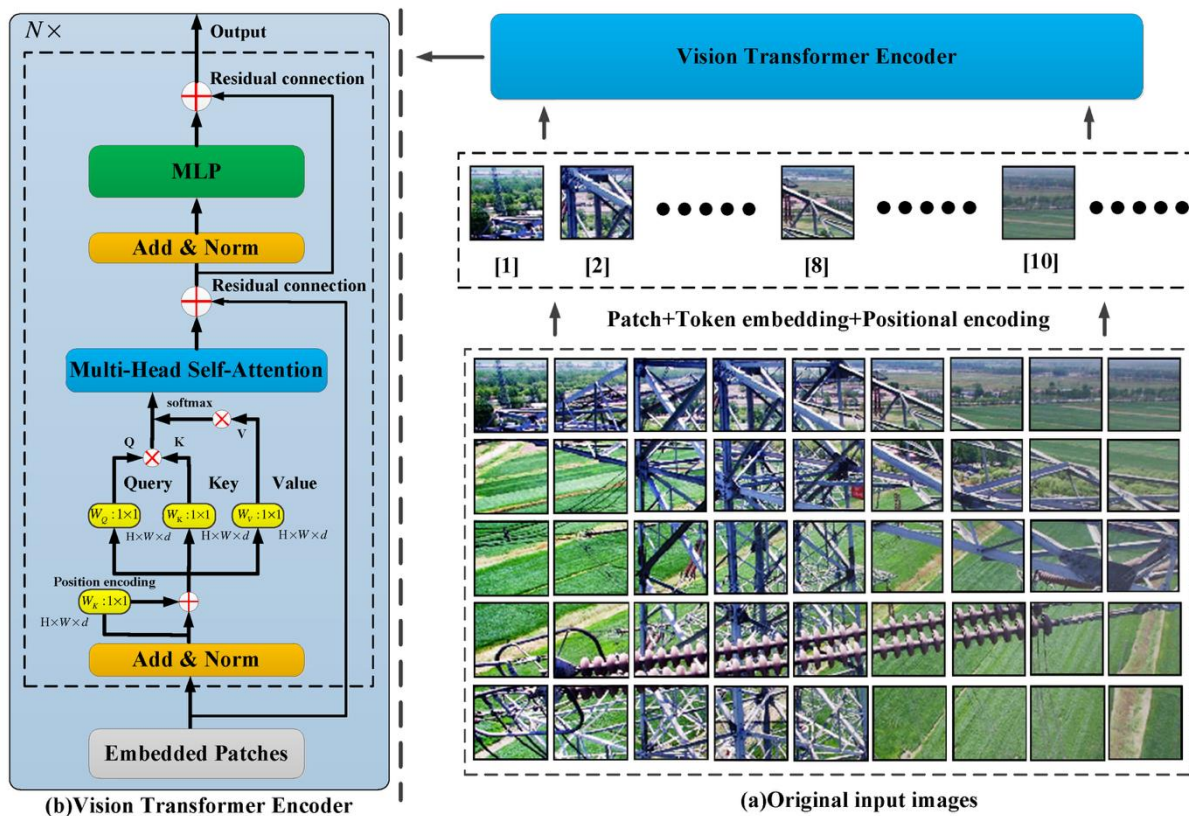


Рисунок 2.16 – Застосування трансформерного підходу до обробки зображень

На відміну від згорткових мереж, де обробка зображення відбувається локально через фільтри, трансформери дозволяють враховувати глобальні залежності між усіма частинами зображення. Це забезпечує більш повне розуміння структури об'єкта та підвищує ефективність класифікації. Таким чином, трансформерний підхід у комп'ютерному зорі є перспективним напрямом розвитку, який дозволяє подолати обмеження традиційних методів і забезпечити більш точний аналіз зображень. Подальші підрозділи присвячені детальному розгляду передумов переходу до трансформерів.

2.4.1 Передумови переходу від CNN до трансформерів

Розвиток методів комп'ютерного зору супроводжувався еволюцією архітектур нейронних мереж, серед яких згорткові нейронні мережі тривалий час залишалися домінуючим підходом. Проте з часом було виявлено ряд обмежень, які знижують ефективність CNN у складних задачах аналізу зображень. Наприклад, трансформерні архітектури дозволяють встановлювати зв'язки між усіма елементами вхідної послідовності незалежно від їх позиції (рисунок 2.17).

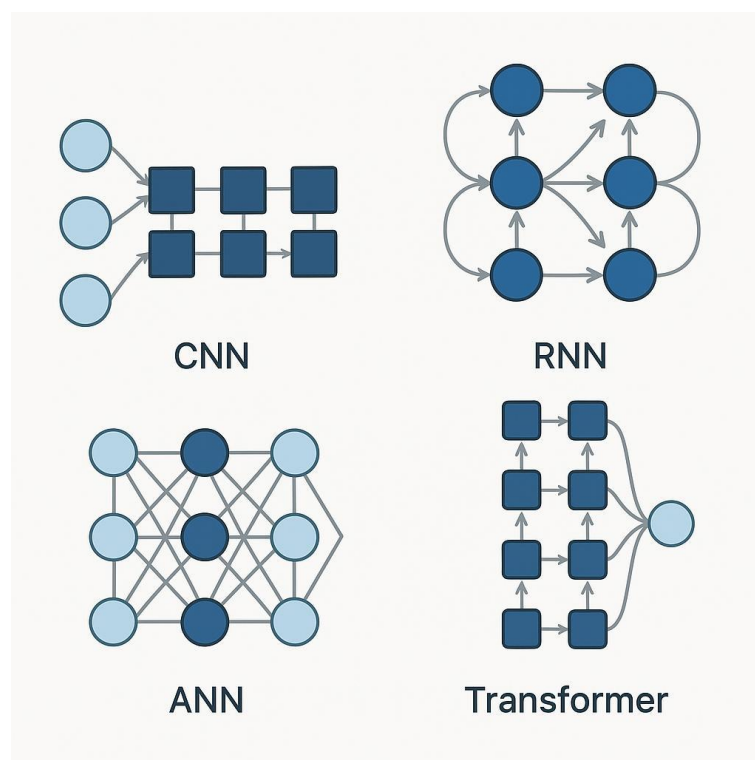


Рисунок 2.17 – Порівняння різних архітектур нейронних мереж та їх структурних особливостей

Однією з основних передумов переходу до трансформерних моделей є обмежена здатність CNN враховувати глобальні залежності між різними частинами зображення. Згорткові мережі працюють на основі локальних

фільтрів, які обробляють невеликі області зображення, що ускладнює встановлення довготривалих взаємозв'язків між віддаленими елементами.

Ще одним фактором є залежність ефективності CNN від глибини мережі. Для розширення області сприйняття необхідно збільшувати кількість шарів, що призводить до зростання обчислювальної складності та ускладнення процесу навчання. Це створює обмеження при побудові моделей, здатних ефективно працювати з високорівневими ознаками.

Важливою передумовою стало також зростання обчислювальних ресурсів та доступність великих наборів даних, що зробило можливим ефективне навчання більш складних моделей. Трансформери, які мають вищу обчислювальну складність, потребують значних ресурсів, однак забезпечують кращу здатність до узагальнення.

Таким чином, обмеження згорткових нейронних мереж у моделюванні глобальних залежностей, складність масштабування та необхідність більш гнучких підходів до аналізу даних стали ключовими передумовами переходу до трансформерного підходу у комп'ютерному зорі.

2.4.2 Переваги трансформерного підходу порівняно зі згортковими мережами

Трансформерні моделі у задачах комп'ютерного зору мають ряд суттєвих переваг порівняно зі згортковими нейронними мережами, що обумовлює їх активне використання у сучасних дослідженнях. Основною відмінністю є підхід до обробки вхідних даних, який дозволяє більш ефективно враховувати складні взаємозв'язки між елементами зображення.

Однією з ключових переваг є здатність трансформерів моделювати глобальні залежності між усіма частинами зображення. У згорткових нейронних мережах обробка виконується локально через застосування фільтрів до обмежених областей зображення (рисунок 2.18), що ускладнює

врахування взаємозв'язків між віддаленими ділянками. Натомість трансформерні моделі аналізують усі фрагменти зображення одночасно, що забезпечує більш повне врахування контексту (рисунок 2.19).

Ще однією важливою перевагою є гнучкість архітектури. Трансформери не мають жорсткої просторової прив'язки, характерної для згорткових мереж, що дозволяє їм ефективно працювати з різними типами даних. У той же час CNN базуються на фіксованих фільтрах і локальних операціях, що обмежує їх здатність до узагальнення.

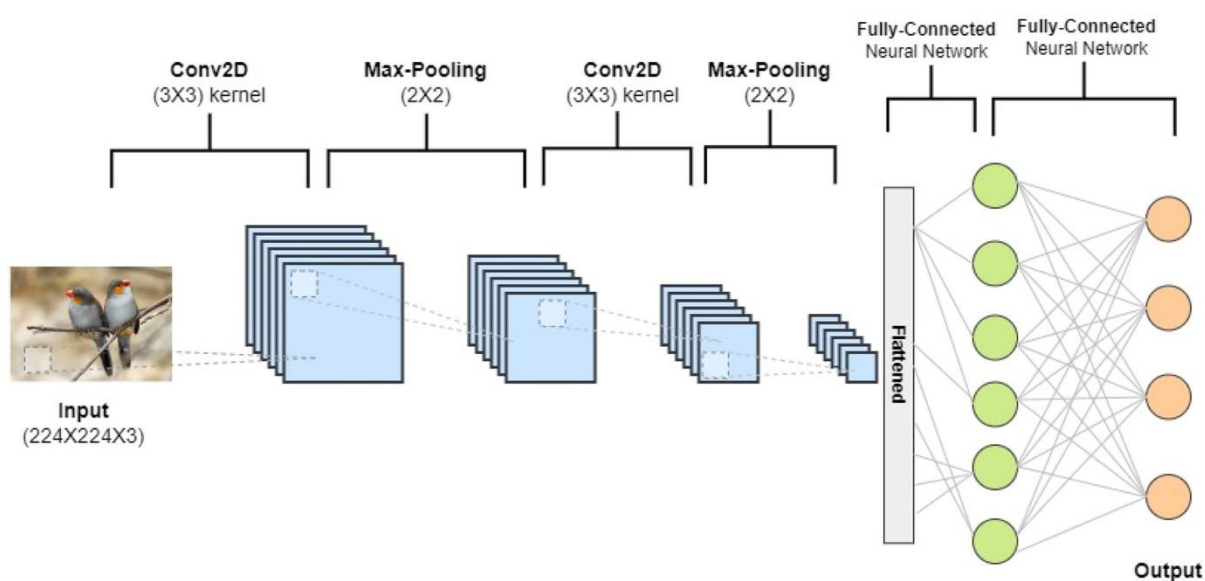


Рисунок 2.18 – Архітектура CNN для класифікації зображень

Також слід відзначити здатність трансформерів краще працювати з великими обсягами даних (рисунок 2.19). При наявності достатньої кількості навчальних прикладів вони демонструють вищу точність порівняно з класичними підходами. Це пов'язано з їх здатністю ефективно використовувати інформацію про взаємозв'язки між ознаками. Крім того, трансформерні моделі забезпечують більш ефективну паралелізацію обчислень. Обробка всіх елементів вхідної послідовності може виконуватися одночасно, що дозволяє скоротити час навчання.

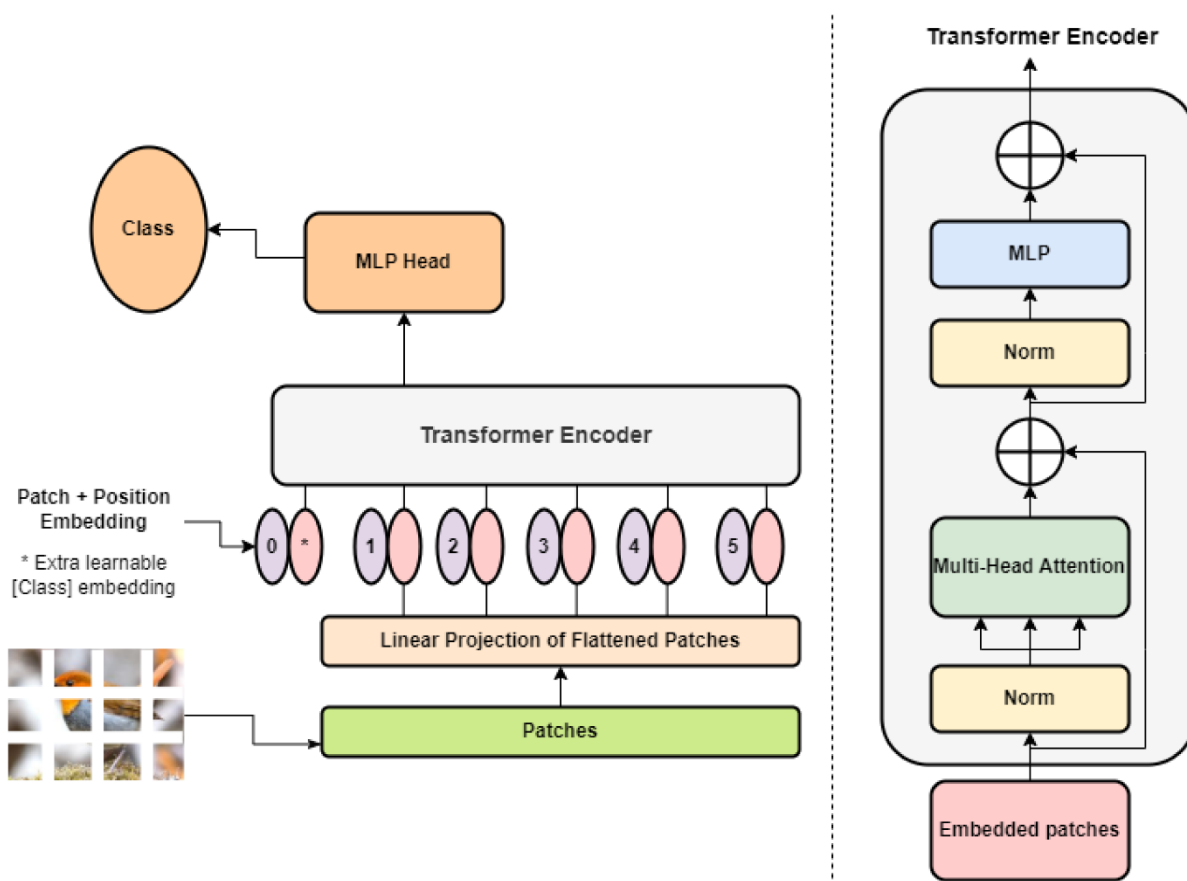


Рисунок 2.19 – Архітектура ViT для обробки зображень

Таким чином, трансформерний підхід перевершує згорткові нейронні мережі за рахунок здатності враховувати глобальні залежності, більшої гнучкості архітектури та ефективності при роботі з великими даними. Це робить його перспективним напрямом розвитку систем комп'ютерного зору.

2.4.3 Базова архітектура трансформера

Базова архітектура трансформера побудована на принципі обробки вхідних даних у вигляді послідовності елементів із використанням механізму самоуваги. Вона складається з двох основних компонентів: енкодера та декодера, кожен з яких виконує окрему функцію у процесі перетворення вхідної інформації у вихідне представлення.

Енкодер призначений для формування узагальненого представлення вхідних даних. Він складається з послідовності однакових блоків, кожен з яких містить два основні підшари: багатоголову самоувагу та повнозв'язну нейронну мережу. Перед подачею на вхід енкодера дані перетворюються у векторні представлення, до яких додається позиційне кодування для врахування порядку елементів (рисунок 2.20).

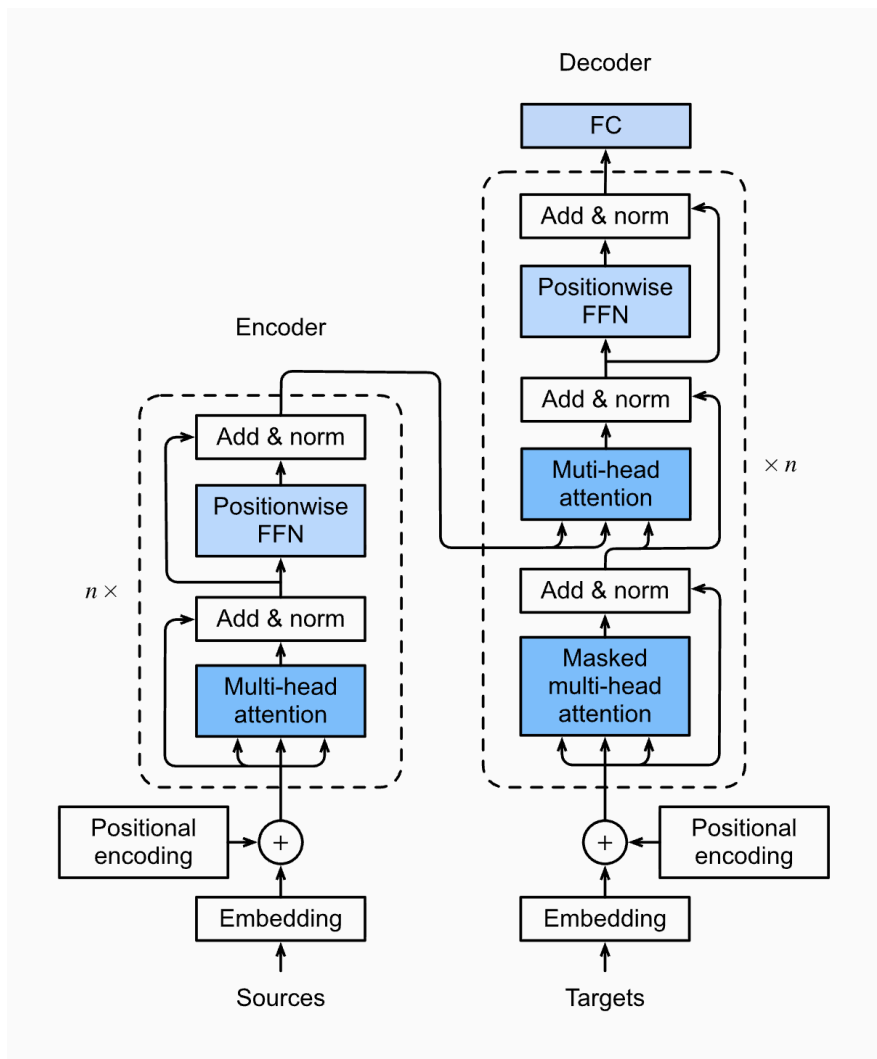


Рисунок 2.20 – Базова архітектура трансформера з енкодером та декодером

Механізм багатоголової самоуваги дозволяє моделі одночасно аналізувати різні аспекти взаємозв'язків між елементами послідовності. Це досягається шляхом формування запитів, ключів і значень, які

використовуються для обчислення вагових коефіцієнтів уваги. У результаті кожен елемент отримує інформацію від усіх інших елементів, що забезпечує глобальне врахування контексту.

Після шару самоуваги застосовується повнозв'язний підшар, який виконує нелінійне перетворення отриманих ознак. Для стабілізації процесу навчання використовуються операції нормалізації та залишкові з'єднання, які дозволяють уникнути деградації якості при збільшенні глибини моделі.

Декодер має подібну структуру, однак додатково включає механізм маскованої самоуваги, який обмежує доступ до майбутніх елементів послідовності. Це є важливим для задач генерації, де необхідно забезпечити послідовне формування вихідних даних. Крім того, декодер використовує інформацію, отриману від енкодера, що дозволяє узгоджувати вхідні та вихідні представлення. Таким чином, базова архітектура трансформера забезпечує ефективне моделювання складних залежностей між елементами даних за рахунок використання механізму самоуваги, позиційного кодування та багаторівневої обробки інформації.

2.5 Архітектура Vision Transformer

Vision Transformer є сучасною архітектурою глибокого навчання, що адаптує трансформерний підхід до задач комп'ютерного зору. На відміну від згорткових нейронних мереж, які використовують локальні операції згортки, Vision Transformer обробляє зображення як послідовність елементів, що дозволяє ефективно враховувати глобальні залежності.

Основною ідеєю цієї архітектури є перетворення вхідного зображення у набір фрагментів фіксованого розміру, які надалі обробляються як окремі елементи послідовності. Кожен фрагмент перетворюється у векторне представлення, до якого додається позиційна інформація, що дозволяє зберегти просторову структуру зображення (рисунок 2.21).

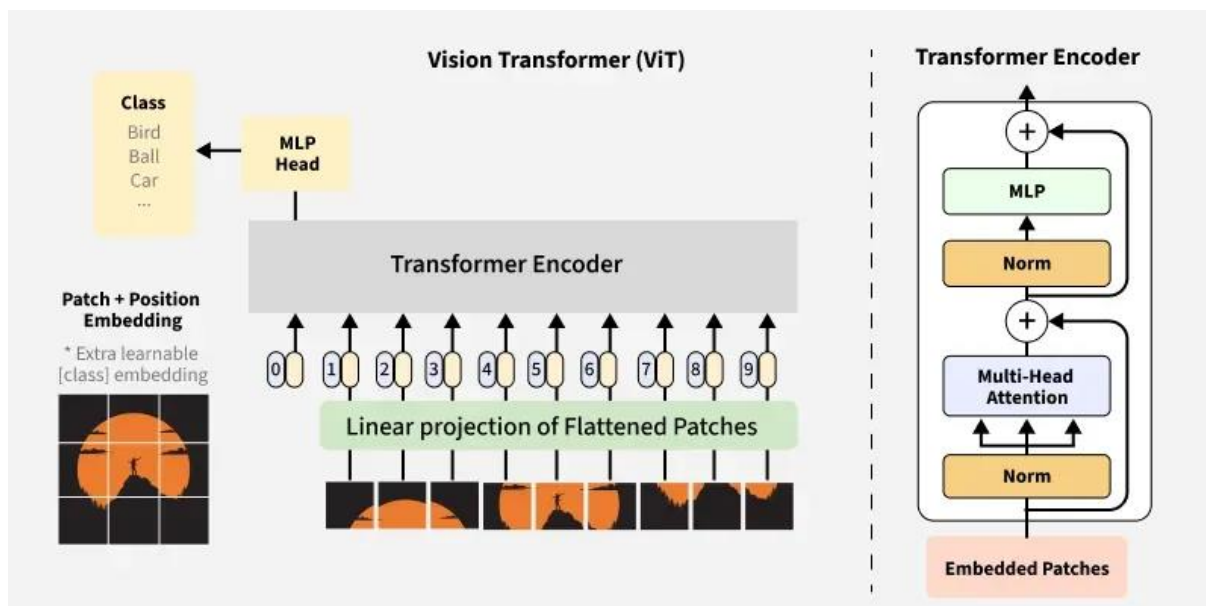


Рисунок 2.21 – Архітектура Vision Transformer із використанням патчів та трансформерного енкодера

Обробка сформованої послідовності здійснюється за допомогою трансформерного енкодера, який складається з декількох однакових блоків. Кожен блок включає механізм багатоголової самоуваги та повнозв'язну мережу, що дозволяє моделі враховувати взаємозв'язки між усіма частинами зображення.

Для задач класифікації використовується спеціальний вектор, який додається до послідовності та накопичує інформацію про все зображення. На основі цього вектора формується фінальний результат класифікації за допомогою повнозв'язного шару.

Таким чином, архітектура Vision Transformer забезпечує ефективну обробку зображень за рахунок поєднання механізму самоуваги та представлення даних у вигляді послідовності. Це дозволяє досягати високої точності у задачах класифікації та робить даний підхід перспективною альтернативою класичним згортковим мережам.

2.5.1 Поділ зображення на патчі у Vision Transformer

Однією з ключових особливостей архітектури Vision Transformer є підхід до представлення вхідного зображення, який відрізняється від традиційних згорткових методів. Замість обробки всього зображення як єдиної структури, воно розбивається на набір фрагментів фіксованого розміру, що дозволяє застосовувати трансформерні моделі до візуальних даних.

Після поділу кожен патч перетворюється у вектор шляхом операції лінеаризації, тобто розгортання двовимірної структури у одновимірний масив (рисунок 2.22). Отримані вектори формують послідовність, яка подається на вхід трансформерного енкодера. Таким чином, зображення перетворюється у формат, аналогічний текстовій послідовності у задачах обробки природної мови.

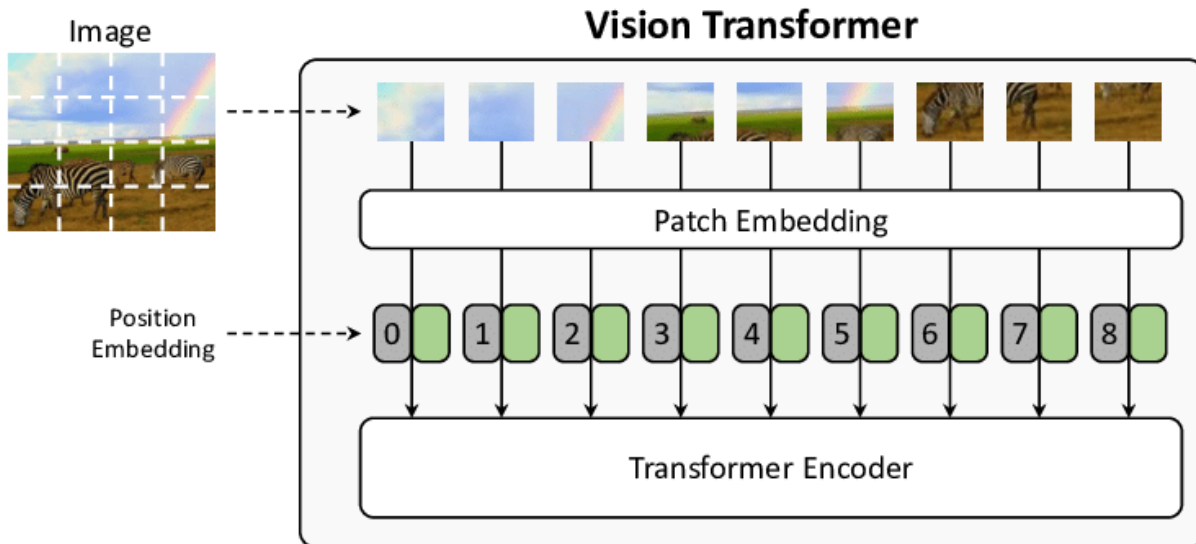


Рисунок 2.22 – Приклад поділу зображення на патчі у Vision Transformer

Процес поділу зображення полягає у розбитті його на невеликі області однакового розміру, які називаються патчами (рисунок 2.23). Кожен патч є підматрицею пікселів і містить локальну інформацію про певну частину

зображення. Розмір патчів зазвичай обирається фіксованим, наприклад 16×16 або 32×32 пікселів, що визначає кількість елементів у подальшій послідовності.

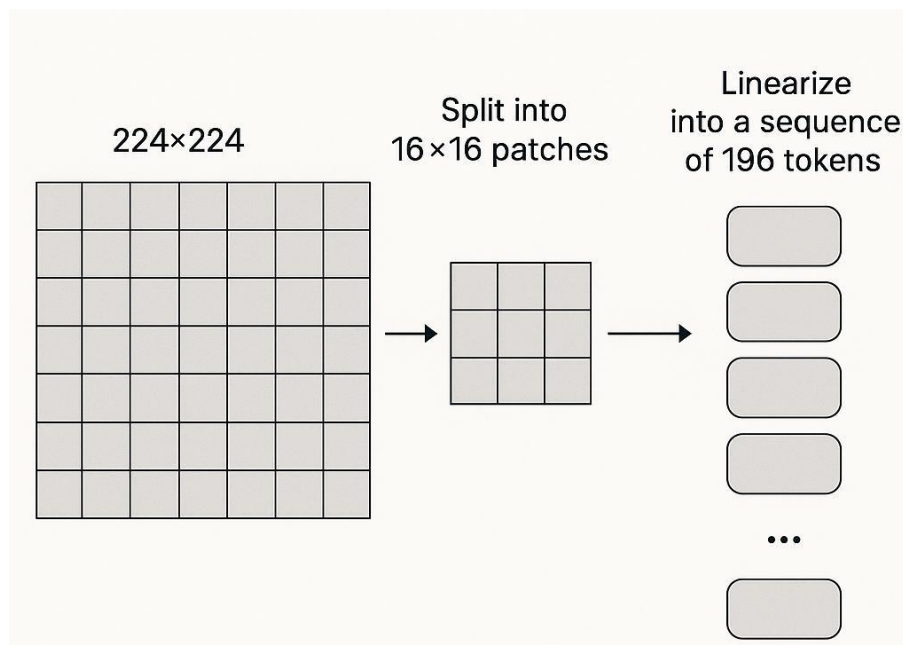


Рисунок 2.23 – Процес поділу зображення на патчі та їх перетворення у векторну послідовність

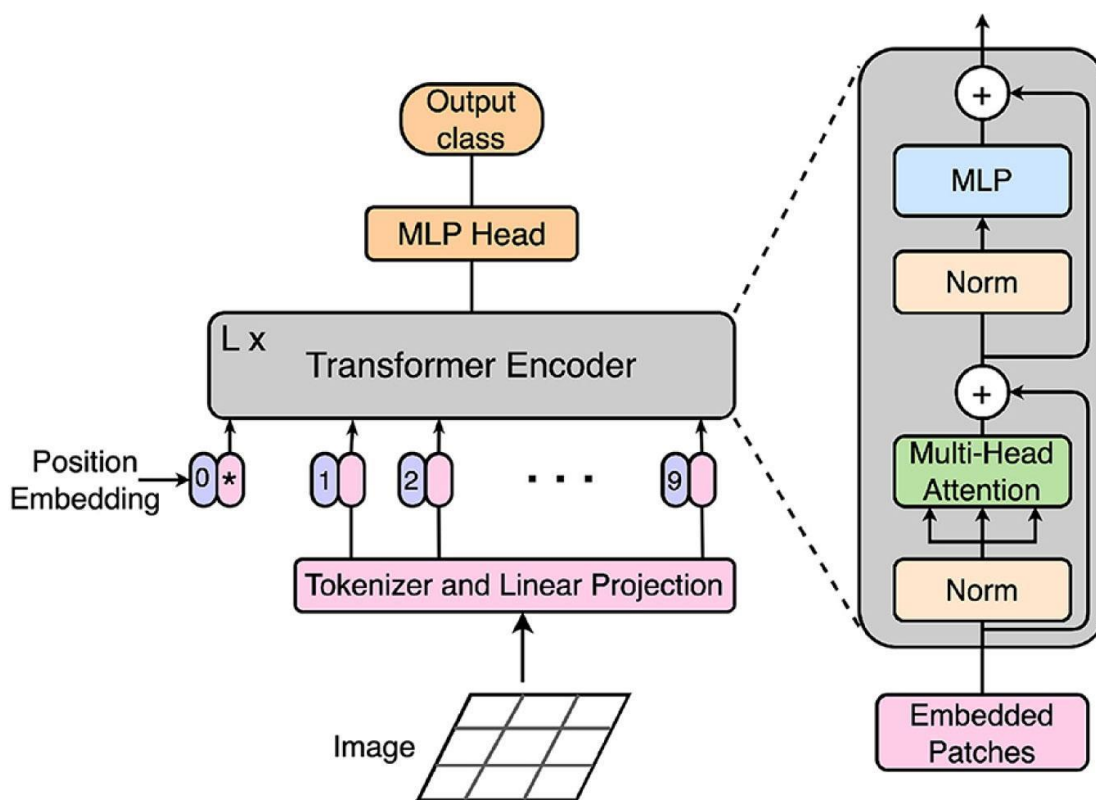
Важливою складовою цього процесу є додавання позиційного кодування, яке дозволяє моделі враховувати просторове розташування патчів. Оскільки трансформери не мають вбудованого механізму врахування порядку елементів, позиційна інформація є необхідною для збереження структури зображення.

Такий підхід дозволяє ефективно поєднати методи комп'ютерного зору та трансформерні архітектури. Подання зображення у вигляді послідовності патчів створює основу для використання механізму самоуваги, що забезпечує врахування взаємозв'язків між різними частинами зображення.

2.5.2 Основні компоненти Vision Transformer

Архітектура Vision Transformer [7] складається з набору ключових компонентів, які забезпечують перетворення зображення у формат, придатний для обробки трансформерною моделлю, а також виконують подальший аналіз і класифікацію даних. Кожен із цих компонентів відіграє важливу роль у формуванні ефективного представлення зображення.

Першим етапом є токенизація зображення, яка передбачає поділ на патчі та їх подальше перетворення у векторне представлення. Цей процес реалізується через лінійну проєкцію, що дозволяє отримати компактні вектори фіксованої розмірності. Отримані вектори називаються токенами і формують послідовність, яка подається на вхід моделі (рисунок 2.24).



Source: *frontiersin*

Рисунок 2.24 – Основні компоненти архітектури Vision Transformer

Наступним важливим компонентом є позиційне кодування, яке додається до токенів для збереження інформації про їх просторове розташування у зображенні. Оскільки трансформер не враховує порядок елементів природним чином, ця інформація є необхідною для коректної обробки візуальних даних.

Основу моделі становить трансформерний енкодер, який складається з декількох однакових блоків. Кожен блок включає механізм багатоголової самоуваги, що дозволяє встановлювати взаємозв'язки між усіма токенами, а також повнозв'язну нейронну мережу, яка виконує нелінійне перетворення ознак. Для стабілізації навчання використовуються операції нормалізації та залишкові з'єднання.

Окрему роль відіграє спеціальний токен класифікації, який додається до вхідної послідовності. У процесі обробки він накопичує інформацію про все зображення та використовується для формування фінального результату. Після проходження через енкодер цей токен подається на класифікаційну голову, що реалізується у вигляді повнозв'язного шару.

Таким чином, основні компоненти Vision Transformer включають етапи токенізації, позиційного кодування, обробки за допомогою трансформерного енкодера та класифікації. Їх узгоджена робота забезпечує ефективний аналіз зображень та високу точність у задачах комп'ютерного зору.

2.5.3 Особливості класифікації зображень у Vision Transformer

Класифікація зображень у Vision Transformer має ряд специфічних особливостей, що відрізняють цей підхід від класичних згорткових нейронних мереж. Основною відмінністю є спосіб формування представлення зображення та прийняття рішення на основі отриманих ознак.

На відміну від CNN, де класифікація здійснюється на основі агрегованих карт ознак, у ViT використовується спеціальний токен класифікації, який додається до послідовності входних даних (рисунок 2.25). У процесі проходження через трансформерний енкодер цей токен взаємодіє з усіма іншими токенами, що дозволяє йому накопичувати узагальнену інформацію про все зображення.

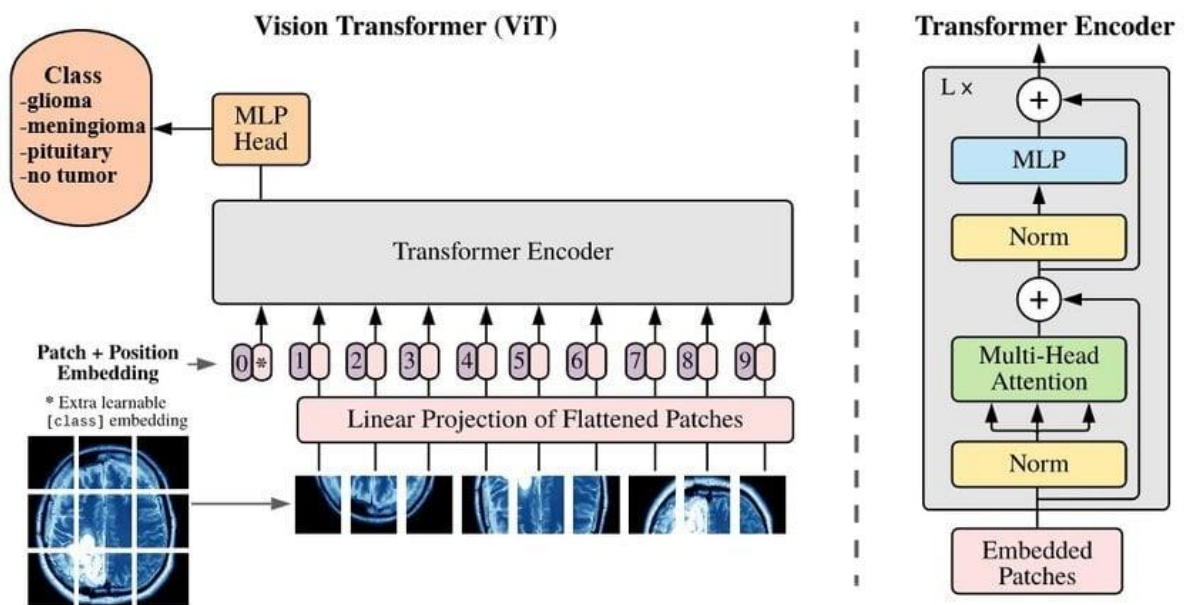


Рисунок 2.25 – Процес класифікації зображень у Vision Transformer із використанням токена класифікації

Важливою особливістю є те, що класифікація базується на глобальному аналізі всього зображення. Завдяки механізму самоуваги модель враховує взаємозв'язки між усіма патчами, що дозволяє ефективно аналізувати складні структури та залежності, характерні для візуальних даних. Ще одним аспектом є відсутність індуктивних обмежень, характерних для згорткових мереж. Vision Transformer не використовує локальні фільтри, тому не має вбудованого припущення про просторову

локальність ознак. Це робить модель більш гнучкою, але водночас підвищує вимоги до обсягу навчальних даних.

Крім того, класифікаційна частина моделі реалізується у вигляді повнозв'язного шару, який отримує на вхід представлення токена класифікації. Це дозволяє перетворити узагальнену інформацію про зображення у ймовірності належності до відповідних класів.

Таким чином, класифікація у Vision Transformer базується на використанні токена класифікації, глобальному аналізі взаємозв'язків між патчами та застосуванні трансформерного енкодера. Це забезпечує високу ефективність моделі у задачах класифікації зображень складної структури.

2.6 Механізм самоуваги як теоретична основа Vision Transformer

Механізм самоуваги є ключовим елементом трансформерних архітектур і становить теоретичну основу моделі Vision Transformer. Його використання дозволяє ефективно моделювати взаємозв'язки між різними частинами вхідних даних, що є особливо важливим у задачах аналізу зображень.

На відміну від згорткових нейронних мереж, де обробка виконується локально, механізм самоуваги забезпечує можливість одночасного врахування інформації з усіх елементів вхідної послідовності. Це дозволяє моделі встановлювати глобальні залежності між різними частинами зображення, що підвищує точність аналізу складних структур.

Суть механізму полягає у визначенні важливості кожного елемента відносно інших. Для цього кожен вхідний елемент перетворюється у три вектори: запит, ключ і значення. На основі цих векторів обчислюються коефіцієнти уваги, які визначають, яку інформацію слід враховувати при формуванні вихідного представлення (рисунок 2.26).

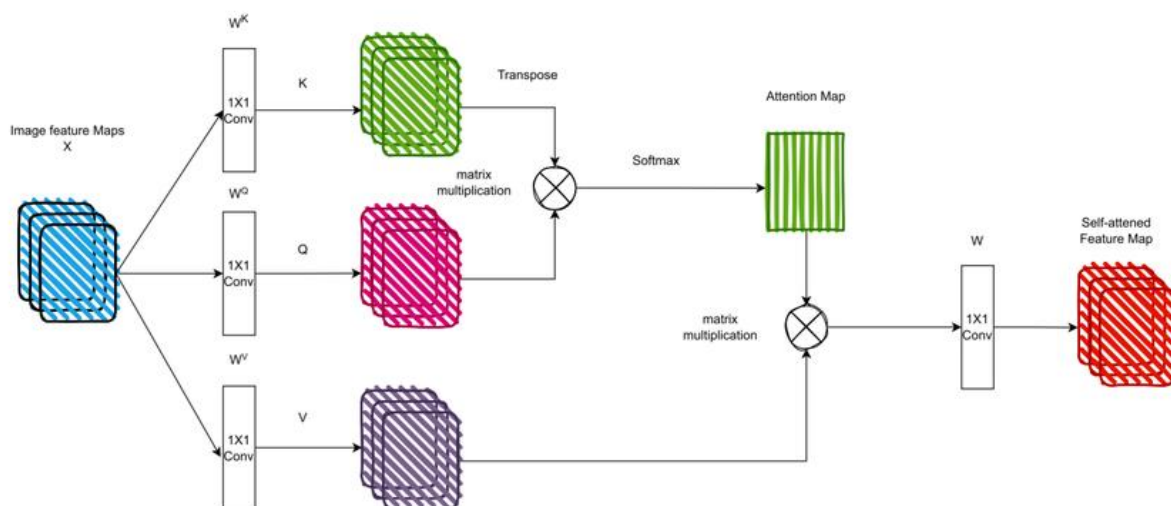


Рисунок 2.26 – Схема роботи механізму самоуваги з використанням запитів, ключів і значень

У результаті кожен елемент отримує нове представлення, яке є зваженою сумою значень інших елементів. Це дозволяє моделі адаптивно фокусуватися на найбільш релевантних частинах даних, що є важливим для задач комп'ютерного зору.

Таким чином, механізм самоуваги забезпечує ефективний спосіб моделювання складних залежностей у даних та дозволяє обробляти інформацію на глобальному рівні. Його використання є ключовим фактором успішності трансформерних моделей, зокрема Vision Transformer.

2.6.1 Призначення self-attention у задачах аналізу зображень

Механізм самоуваги у задачах аналізу зображень використовується для визначення важливості різних частин вхідних даних та встановлення взаємозв'язків між ними. Його основне призначення полягає у забезпеченні ефективного виділення інформативних ознак, які є найбільш значущими для виконання конкретної задачі, зокрема класифікації або виявлення патологій.

На відміну від локальних методів обробки, таких як згорткові операції, self-attention дозволяє враховувати глобальний контекст зображення. Це означає, що модель може встановлювати залежності між віддаленими ділянками зображення, що є важливим для аналізу складних структур, характерних для медичних знімків.

У процесі роботи механізм самоуваги формує карту уваги, яка відображає ступінь впливу різних частин зображення на формування вихідного представлення (рисunek 2.27). Це дозволяє моделі фокусуватися на ключових областях, ігноруючи менш значущі елементи, що підвищує ефективність обробки даних.

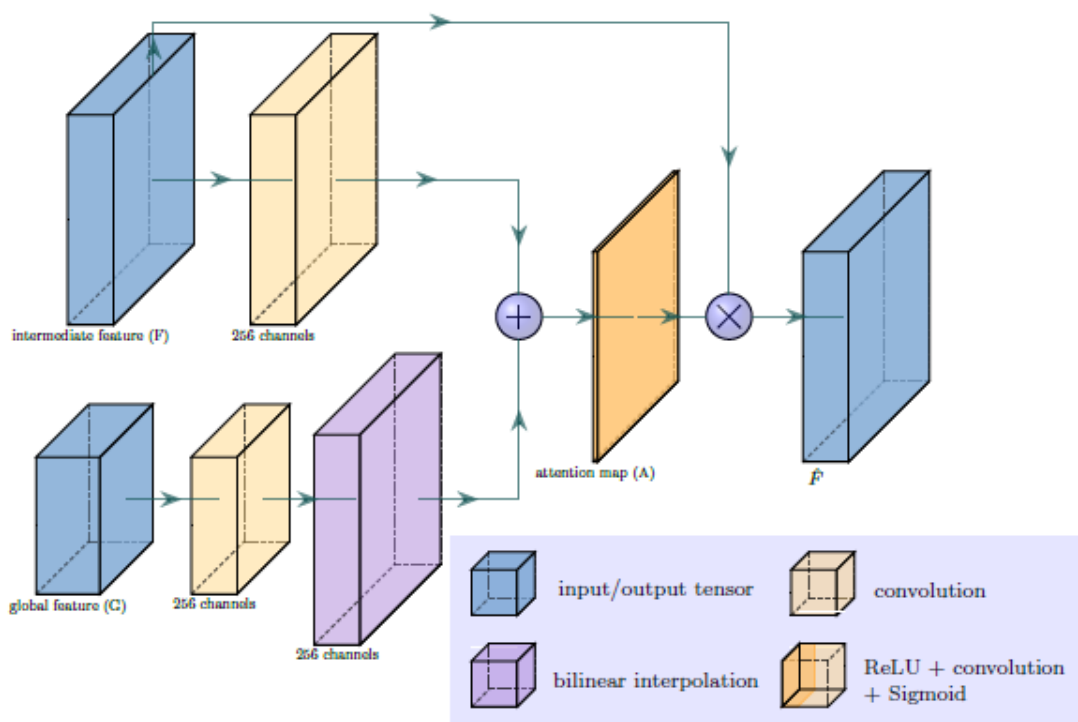


Рисунок 2.27 – Використання механізму самоуваги для виділення релевантних ознак у зображенні

Особливо важливим є використання self-attention у задачах медичної діагностики, де патологічні зміни можуть бути локалізовані у різних частинах зображення та мати різний ступінь вираженості. Завдяки здатності

враховувати глобальні залежності модель може більш точно визначати наявність захворювань.

Таким чином, механізм самоуваги забезпечує адаптивне виділення релевантних ознак та дозволяє враховувати складні взаємозв'язки у зображеннях. Це робить його важливим інструментом у сучасних методах аналізу візуальних даних.

2.6.2 Формування запитів, ключів і значень

Механізм самоуваги [8] базується на формуванні трьох основних компонентів – запитів, ключів і значень, які використовуються для визначення взаємозв'язків між елементами вхідної послідовності. Ці компоненти дозволяють моделі оцінювати важливість кожного елемента відносно інших та формувати узагальнене представлення даних.

Для кожного вхідного елемента обчислюються три векторні представлення шляхом лінійних перетворень: вектор запиту, вектор ключа та вектор значення. Вектор запиту використовується для пошуку релевантної інформації, ключі визначають, наскільки кожен елемент відповідає запиту, а значення містять інформацію, яка буде використана при формуванні виходу.

Процес обчислення уваги полягає у визначенні коефіцієнтів відповідності між запитом та всіма ключами. Для цього виконується операція скалярного добутку, після чого результати нормалізуються за допомогою функції softmax. Отримані коефіцієнти визначають ваги, з якими відповідні значення впливають на підсумковий результат (рисунки 2.28).

У результаті формується вихідний вектор, який є зваженою сумою векторів значень. Це дозволяє моделі інтегрувати інформацію з різних частин послідовності з урахуванням їх важливості для конкретного запиту.

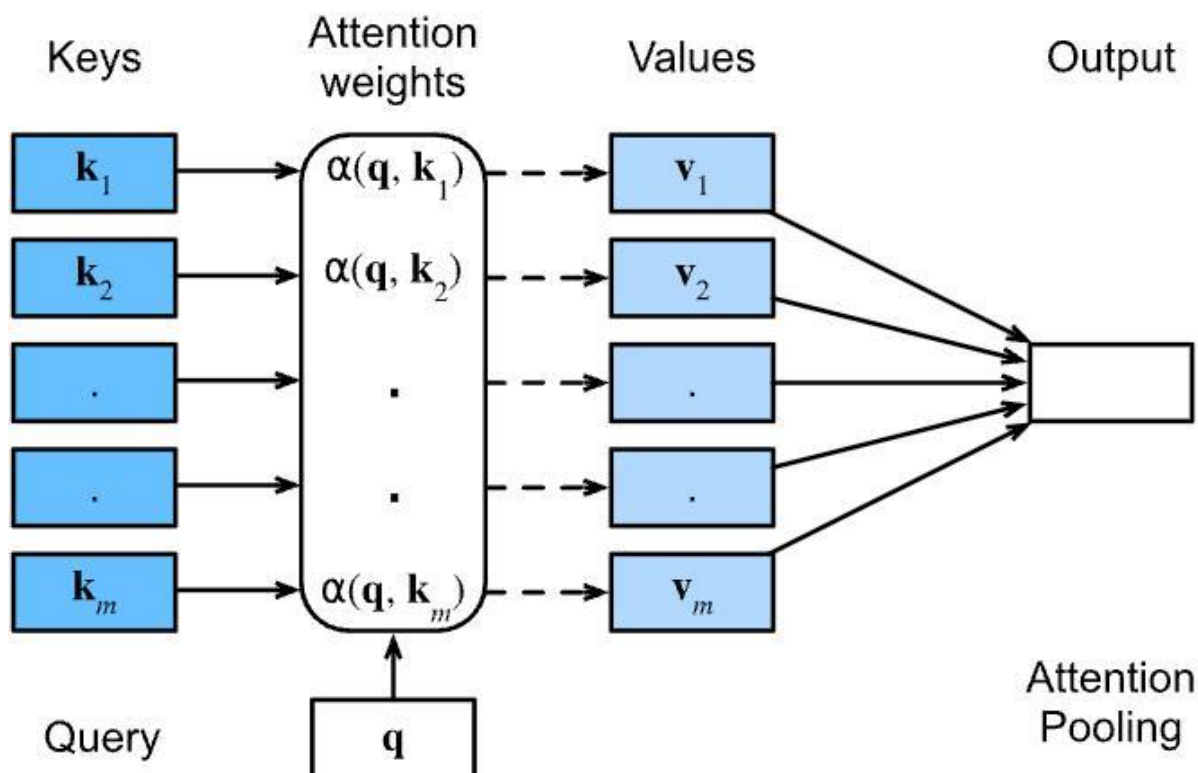


Рисунок 2.28 – Формування запитів, ключів і значень у механізмі самоуваги

Такий підхід забезпечує гнучке та адаптивне моделювання залежностей між елементами даних. На відміну від традиційних методів, де зв'язки між елементами є фіксованими, механізм самоуваги.

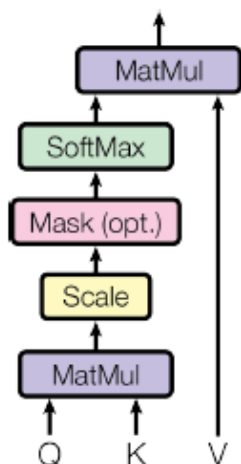
2.6.3 Multi-head attention та його роль у виділенні ознак

Механізм багатоголової самоуваги є розширенням базового механізму self-attention і відіграє ключову роль у підвищенні ефективності трансформерних моделей. Його основна ідея полягає у паралельному використанні декількох незалежних механізмів уваги, що дозволяє моделі одночасно аналізувати різні аспекти вхідних даних.

У межах multi-head attention вхідні дані проєктуються у кілька різних підпросторів, після чого для кожного з них окремо обчислюється механізм

самоуваги. Це означає, що для кожної «голови» формуються власні запити, ключі та значення, які дозволяють виділяти різні типи залежностей між елементами (рисунок 2.29).

Scaled Dot-Product Attention



Multi-Head Attention

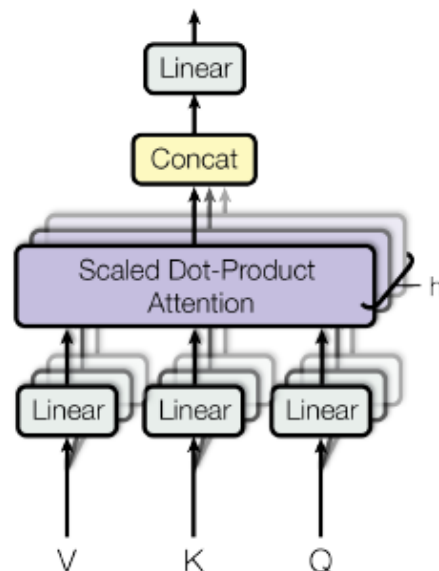


Рисунок 2.29 – Механізм multi-head attention та його структура

Отримані результати з усіх голів об'єднуються шляхом конкатенації та подальшого лінійного перетворення. Такий підхід забезпечує більш повне представлення даних порівняно з використанням одного механізму уваги, оскільки кожна голова може фокусуватися на окремих характеристиках вхідної інформації. У задачах аналізу зображень це дозволяє моделі одночасно враховувати як локальні, так і глобальні ознаки.

Використання багатоголової самоуваги також підвищує здатність моделі до узагальнення, оскільки вона отримує можливість аналізувати дані з різних точок зору. Це особливо важливо у складних задачах, де ознаки мають багаторівневу структуру.

2.7 Вплив попереднього навчання на ефективність моделі

Попереднє навчання [9] є важливим етапом у побудові сучасних моделей глибокого навчання, що суттєво впливає на їх ефективність. Його сутність полягає у навчанні моделі на великому обсязі даних з метою формування узагальнених ознак, які можуть бути використані при розв'язанні інших задач.

Попередньо навчена модель містить ваги (рисунок 2.30), які вже відображають базові закономірності у даних, що дозволяє уникнути необхідності повного навчання з нуля. Це особливо важливо у випадках, коли доступний обсяг даних для конкретної задачі є обмеженим, як це часто спостерігається у медичних застосуваннях.

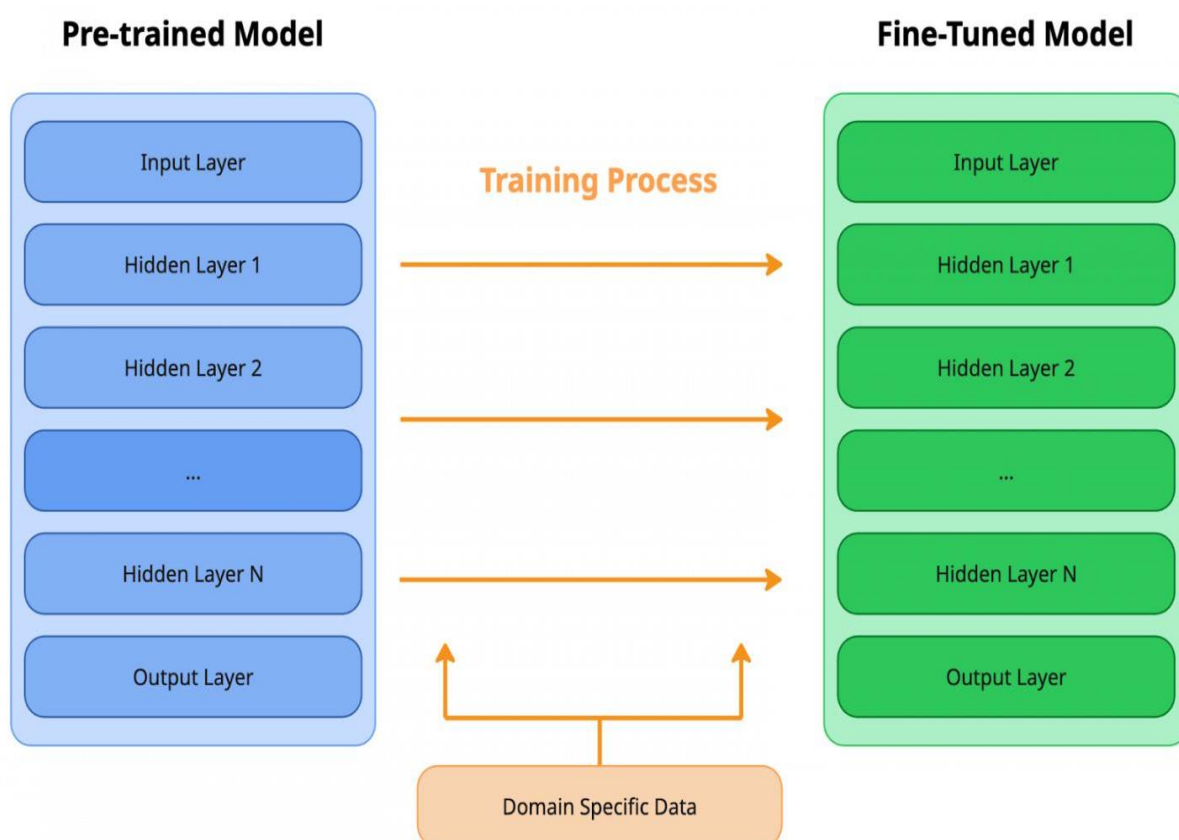


Рисунок 2.30 – Використання попередньо навченої моделі та її адаптація до нових даних

Використання попереднього навчання дозволяє значно прискорити процес навчання моделі, оскільки початкові ваги вже знаходяться у близькому до оптимального стані.

Крім того, попередньо навчена модель демонструє кращу здатність до узагальнення, що дозволяє підвищити точність класифікації на нових даних. Це пояснюється тим, що модель вже навчена виділяти універсальні ознаки, які є корисними для широкого спектра задач (рисунок 2.30).

2.7.1 Сутність підходу pre-training у глибокому навчанні

Підхід попереднього навчання у глибокому навчанні (рисунок 2.31) полягає у формуванні початкових параметрів моделі на основі великого обсягу даних до її застосування у конкретній задачі. Такий підхід дозволяє моделі навчитися виділяти загальні закономірності та базові ознаки, які можуть бути використані надалі без необхідності навчання з нуля.

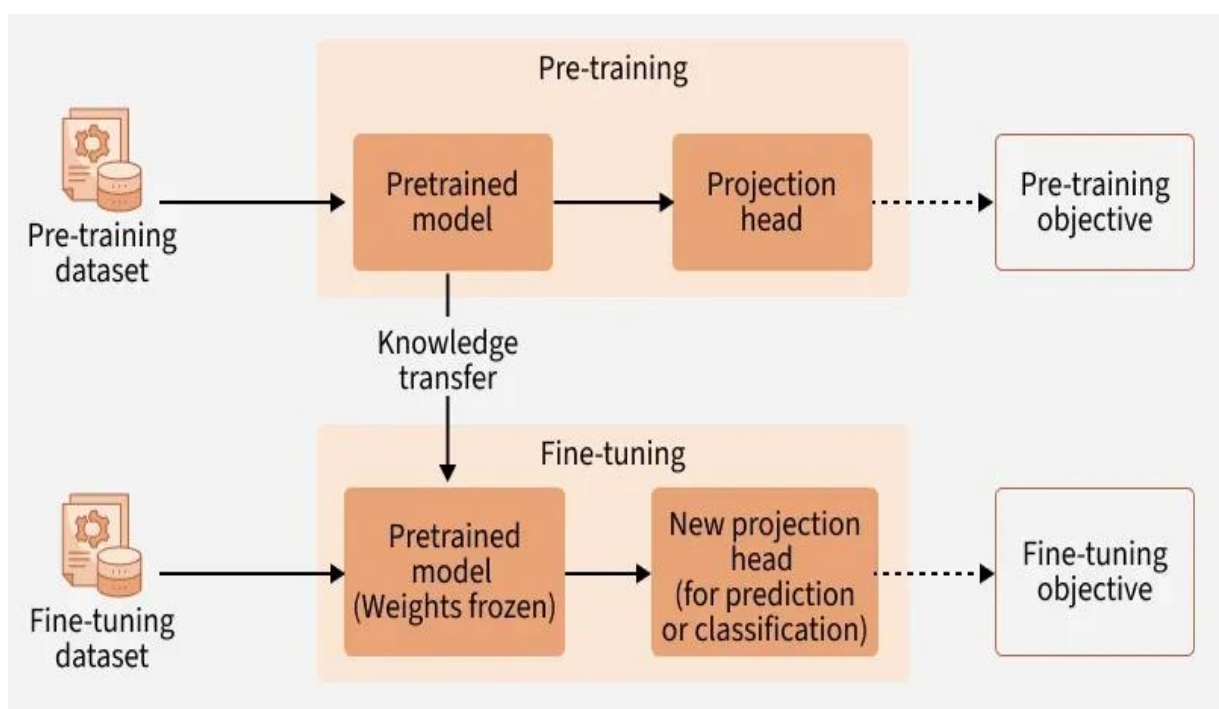


Рисунок 2.31 – Процес попереднього навчання моделі

У процесі попереднього навчання модель оптимізується на великому наборі даних, що дозволяє сформувати універсальні представлення вхідної інформації. Після цього отримані ваги можуть бути використані як початкове наближення для розв’язання нових задач, що значно підвищує ефективність подальшого навчання.

Сутність цього підходу полягає у перенесенні знань, закладених у параметрах моделі, що дозволяє використовувати вже сформовані ознаки без необхідності їх повторного навчання. Це особливо важливо для задач, де обсяг доступних даних є обмеженим, оскільки модель вже має здатність виділяти інформативні характеристики.

Важливою перевагою попереднього навчання є прискорення процесу оптимізації. Оскільки модель починає навчання з уже сформованими вагами, кількість необхідних ітерацій для досягнення високої точності значно зменшується. Крім того, це дозволяє уникнути проблем, пов’язаних із випадковою ініціалізацією параметрів.

2.7.2 Роль попереднього навчання для Vision Transformer

Попереднє навчання відіграє ключову роль у забезпеченні ефективності моделі Vision Transformer, оскільки дана архітектура має значну кількість параметрів і не містить індуктивних обмежень, характерних для згорткових нейронних мереж.

На відміну від CNN, які мають вбудовані механізми локальної обробки та інваріантності до зсувів, Vision Transformer повністю покладається на дані для навчання структури ознак. У зв’язку з цим використання попереднього навчання на великих наборах даних, таких як ImageNet, є необхідною умовою для досягнення високої точності.

Попередньо навчена модель отримує здатність ефективно виділяти загальні візуальні закономірності, що дозволяє значно покращити

результати при подальшому застосуванні. Це особливо важливо у задачах медичної діагностики, де обсяг доступних даних є обмеженим.

Як показано на графіку, моделі, що використовують попереднє навчання на більшому обсязі даних, демонструють вищу точність порівняно з моделями, навченими з нуля (рисунок 2.32). Це підтверджує, що попереднє навчання є критичним фактором для підвищення ефективності Vision Transformer.

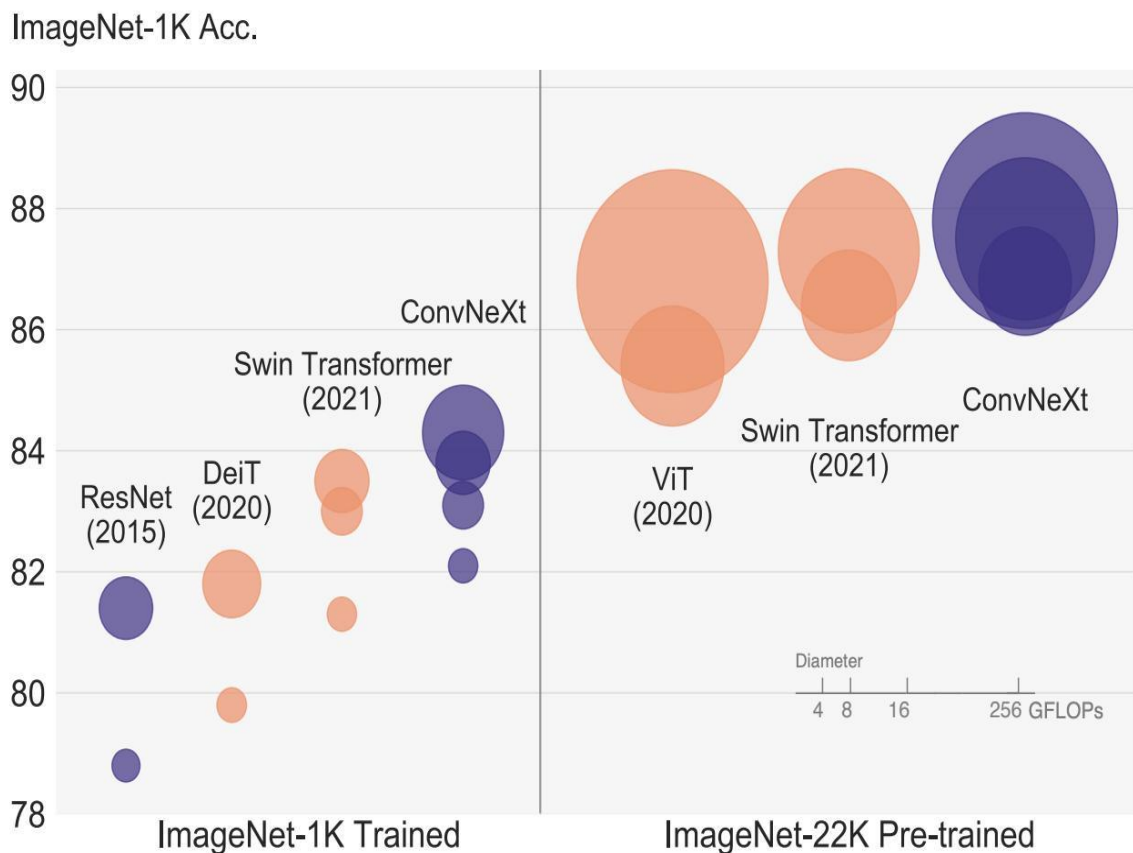


Рисунок 2.32 – Порівняння ефективності моделей залежно від використання попереднього навчання

Таким чином, попереднє навчання є необхідним компонентом успішного застосування Vision Transformer, що забезпечує підвищення точності, стабільності навчання та здатності моделі до узагальнення у задачах класифікації зображень.

2.8 Гібридна архітектура на основі ResNet та Vision Transformer

У сучасних задачах аналізу медичних зображень спостерігається тенденція до поєднання різних архітектур глибокого навчання з метою підвищення ефективності класифікації [10]. Зокрема, згорткові нейронні мережі, такі як ResNet, демонструють високу здатність до виділення локальних ознак, тоді як Vision Transformer забезпечує ефективне моделювання глобальних залежностей між різними частинами зображення. Це обумовлює доцільність використання гібридних підходів, які поєднують переваги обох архітектур у єдиній моделі.

Гібридна архітектура на основі ResNet та Vision Transformer дозволяє одночасно враховувати як локальні структурні особливості медичних зображень [11], так і глобальний контекст їх інтерпретації. Такий підхід є особливо актуальним у задачах медичної діагностики, де патологічні зміни можуть мати як локальний характер, так і проявлятися у складних взаємозв'язках між різними ділянками зображення. Використання гібридних моделей сприяє підвищенню точності класифікації та створює передумови для побудови більш ефективних систем підтримки прийняття рішень у медицині.

2.8.1 Аналіз обмежень класичних та трансформерних архітектур

У задачах класифікації медичних зображень традиційно широко застосовуються згорткові нейронні мережі, які демонструють високу ефективність у виділенні локальних ознак. Проте такі архітектури мають ряд обмежень, пов'язаних із їх природою. Основною характеристикою CNN є локальність операцій згортки, що дозволяє ефективно аналізувати невеликі області зображення, але ускладнює врахування глобальних залежностей між віддаленими ділянками. Це призводить до необхідності збільшення глибини мережі для розширення області сприйняття, що, у свою

чергу, ускладнює процес навчання та підвищує обчислювальні витрати. Як показано це нижче (рисунок 2.33), згорткові мережі зосереджуються на локальних фрагментах зображення, поступово агрегуючи інформацію, що може призводити до втрати глобального контексту при аналізі складних структур.

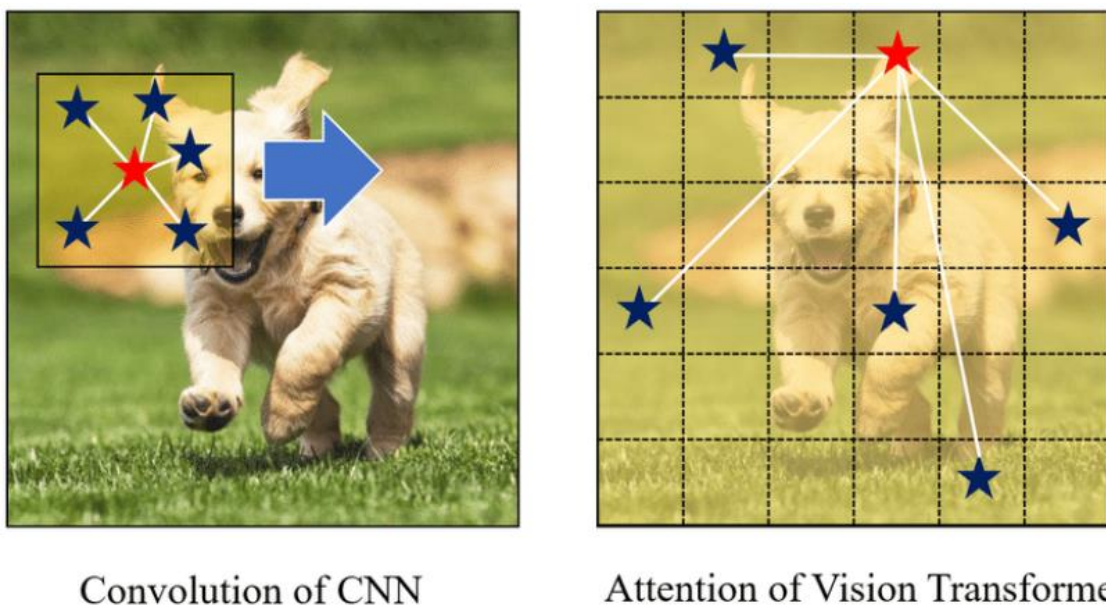
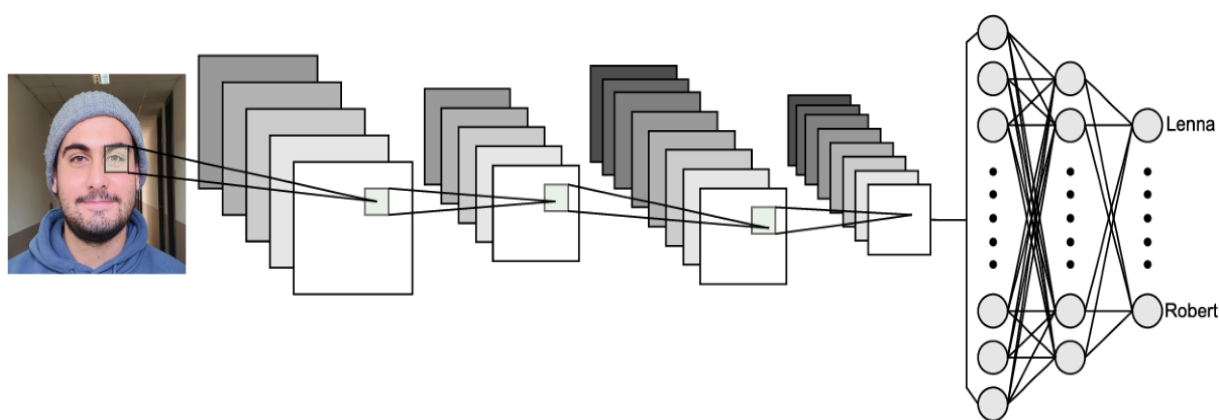


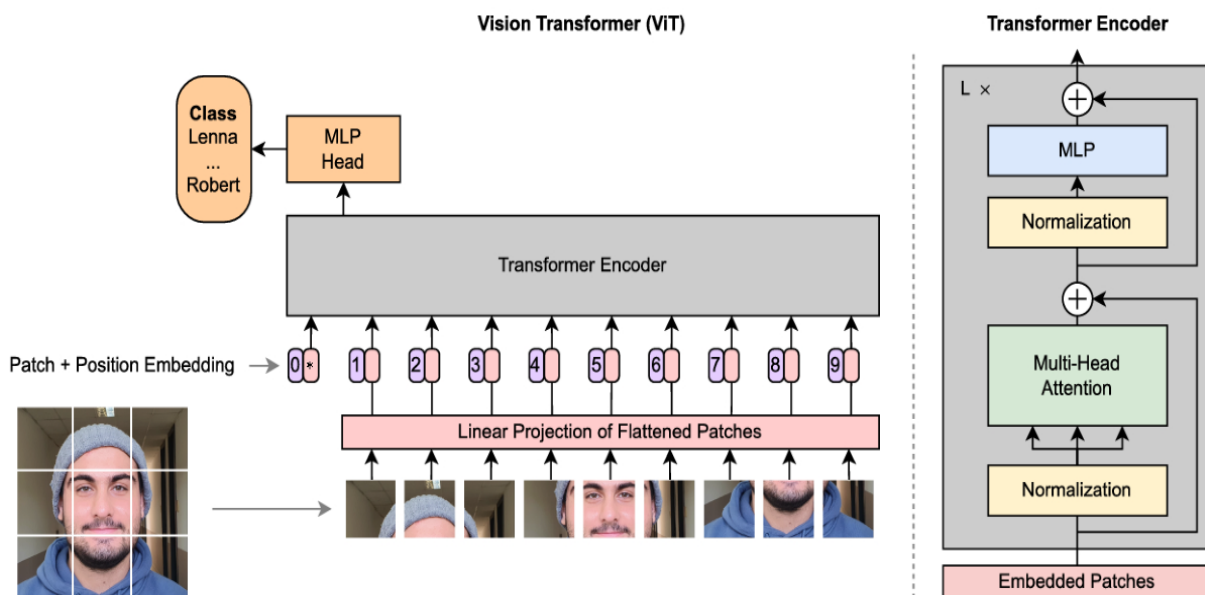
Рисунок 2.33 – Порівняння підходів CNN та Vision Transformer до обробки зображень

Іншим обмеженням CNN є їх залежність від індуктивних припущень, зокрема припущення про просторову локальність і трансляційну інваріантність. Хоча ці властивості є корисними для багатьох задач комп'ютерного зору, у випадку медичних зображень вони можуть обмежувати здатність моделі враховувати складні взаємозв'язки між різними анатомічними структурами. Крім того, для досягнення високої точності CNN часто потребують значної кількості параметрів і великої кількості навчальних даних, що є проблематичним у медичній галузі, де доступ до якісно розмічених даних є обмеженим.

З іншого боку, трансформерні архітектури (рисунок 2.34), зокрема Vision Transformer, дозволяють ефективно моделювати глобальні залежності між усіма частинами зображення за рахунок використання механізму самоуваги. Трансформер аналізує взаємозв'язки між усіма патчами зображення одночасно, що забезпечує більш повне врахування контексту. Це є суттєвою перевагою порівняно з CNN, особливо у задачах, де важливу роль відіграють взаємозалежності між різними ділянками зображення.



(a) Common CNN architecture



(b) Vision Transformer architecture

Рисунок 2.34 – Архітектурне представлення CNN та Vision Transformer

Разом з тим трансформерні моделі також мають ряд обмежень. Одним із ключових є висока вимогливість до обсягу навчальних даних. На відміну від CNN, які мають вбудовані індуктивні обмеження, трансформери не мають таких припущень, що призводить до необхідності використання великих датасетів для ефективного навчання. Крім того, обчислювальна складність механізму самоуваги зростає квадратично відносно кількості вхідних елементів, що ускладнює застосування таких моделей до високороздільних зображень. Також трансформери є менш ефективними у виділенні дрібних локальних ознак без додаткових механізмів, що може знижувати їх ефективність у задачах, де важливі тонкі візуальні деталі.

Таким чином, як згорткові нейронні мережі, так і трансформерні архітектури мають власні обмеження, пов'язані з особливостями їх побудови та принципами обробки даних. Це обумовлює необхідність пошуку комбінованих підходів, які дозволяють поєднати переваги обох архітектур та мінімізувати їх недоліки, що є передумовою для створення гібридних моделей у задачах медичної діагностики.

2.8.2 Обґрунтування доцільності використання гібридного підходу

У сучасних задачах аналізу медичних зображень використання лише одного типу архітектури не завжди забезпечує достатній рівень точності та узагальнення. Згорткові нейронні мережі ефективно виділяють локальні ознаки, однак мають обмеження у врахуванні глобальних залежностей між різними ділянками зображення. У свою чергу трансформерні моделі забезпечують глобальний аналіз за рахунок механізму самоуваги, проте потребують значного обсягу даних та можуть бути менш ефективними при обробці дрібних локальних деталей.

З метою подолання зазначених обмежень доцільним є використання гібридного підходу (рисунок 2.35), який поєднує переваги обох архітектур. Згорткова частина моделі виконує попереднє виділення ознак із вхідного

зображення, формуючи інформативні feature maps, які надалі перетворюються у послідовність векторів. Ця послідовність подається на вхід трансформерного енкодера, який забезпечує моделювання глобальних взаємозв'язків між різними областями зображення.

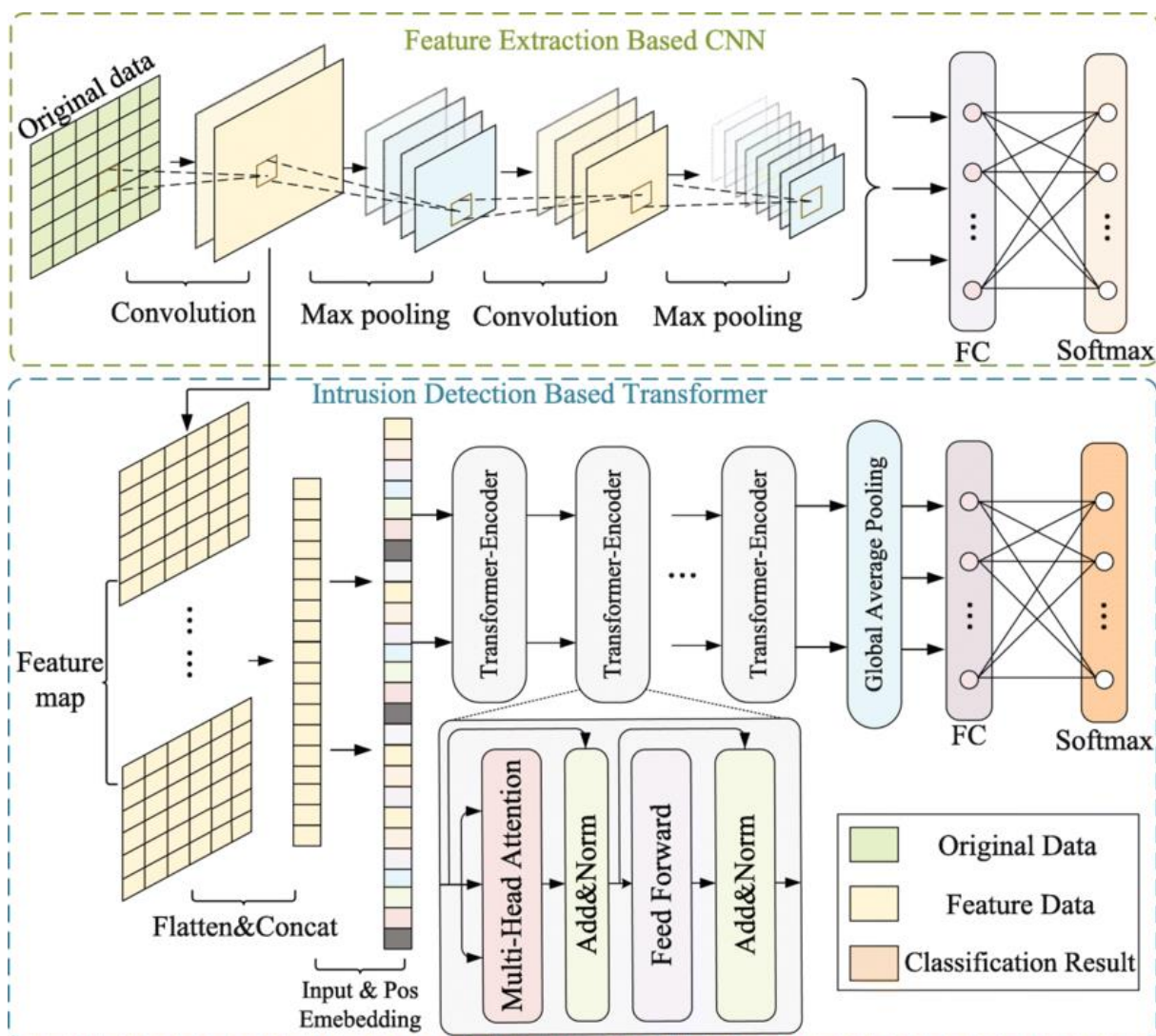


Рисунок 2.35 – Концептуальна схема поєднання згорткової нейронної мережі та трансформерної архітектури у гібридній моделі класифікації зображень

Такий підхід дозволяє ефективно комбінувати локальну та глобальну інформацію, що є особливо важливим у задачах медичної діагностики. Рентгенівські знімки грудної клітки характеризуються складною

структурою, де патологічні зміни можуть проявлятися як у вигляді локальних аномалій, так і у вигляді взаємозалежностей між різними анатомічними зонами. Використання гібридної архітектури дає змогу підвищити якість класифікації за рахунок більш повного врахування таких особливостей.

Таким чином, поєднання згорткових нейронних мереж та трансформерів є обґрунтованим підходом, що дозволяє мінімізувати недоліки кожної з архітектур окремо та створює передумови для підвищення ефективності систем підтримки прийняття рішень у медичній діагностиці.

2.8.3 Формування структури гібридної архітектури

У межах даного дослідження формування структури гібридної архітектури базується на поєднанні сильних сторін згорткових нейронних мереж та трансформерних моделей.

Основною ідеєю є використання ResNet (рисунок 2.36) як ефективного екстрактора локальних ознак із вхідного зображення та подальша передача отриманих представлень до трансформерного енкодера для моделювання глобальних залежностей. Такий підхід дозволяє забезпечити комплексний аналіз медичних зображень, що включає як локальні деталі, так і загальний контекст.

Вхідне зображення подається на згортковий блок, який складається з кількох послідовних стадій обробки. Кожна стадія включає згорткові шари, операції нормалізації та нелінійні функції активації, що дозволяє поступово формувати багаторівневі представлення ознак. Використання residual-зв'язків у структурі ResNet забезпечує стабільність навчання та зменшує проблему зникнення градієнта, що є характерною для глибоких нейронних мереж.

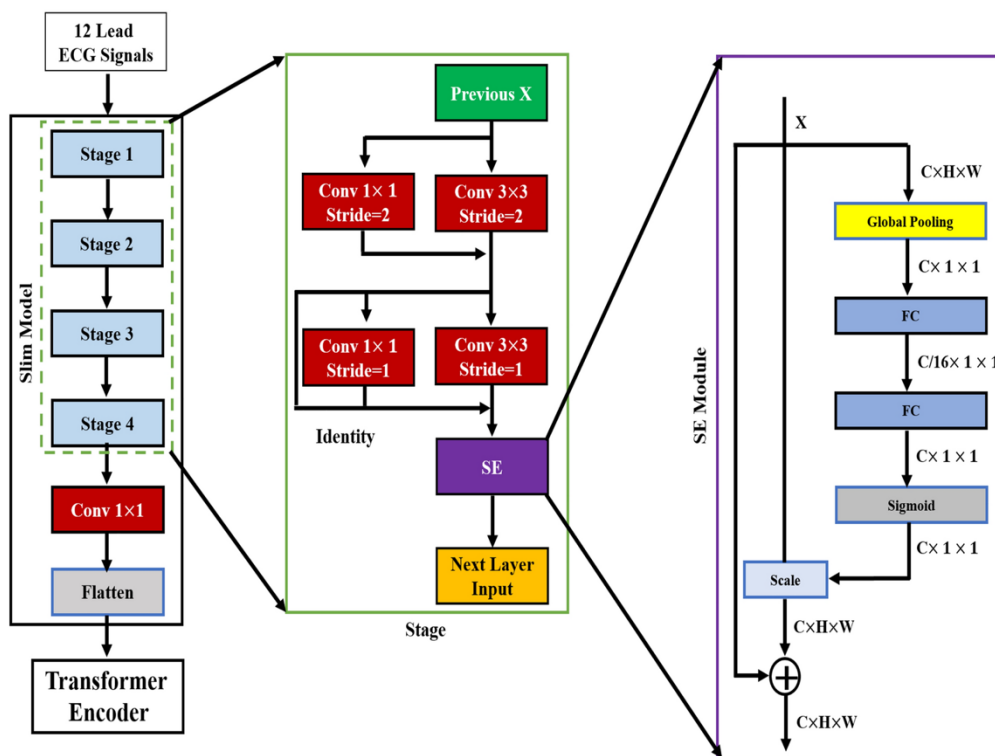


Рисунок 2.36 – Структура гібридної архітектури на основі ResNet та Vision Transformer

Після проходження згорткового блоку формується карта ознак, яка перетворюється у векторизоване представлення та подається на вхід трансформерного енкодера. У трансформерному блоці реалізується механізм самоуваги, що дозволяє моделі враховувати взаємозв'язки між різними частинами зображення незалежно від їх просторового розташування. Це особливо важливо у задачах медичної діагностики, де патологічні зміни можуть мати складний та розподілений характер.

Додатково у структурі моделі можуть використовуватися механізми підсилення значущих ознак, зокрема блоки типу squeeze-and-excitation, що дозволяють адаптивно масштабувати каналні представлення. Завершальним етапом є класифікаційний блок, який включає повнозв'язні шари та функцію активації для отримання ймовірностей належності зображення до кожного з класів у задачі багатоміткової класифікації.

2.8.4 Опис основних модулів запропонованої архітектури

У запропонованій гібридній архітектурі можна виділити кілька ключових модулів, кожен із яких виконує окрему функцію у процесі обробки медичних зображень. Загальна взаємодія цих модулів охоплює повний цикл обробки даних – від попередньої підготовки зображень до отримання результатів класифікації (рисунк 2.37).

Першим етапом є модуль попередньої обробки даних, який забезпечує приведення рентгенівських знімків до єдиного формату. На цьому етапі виконується нормалізація інтенсивності, зміна розміру зображень та їх перетворення у числові масиви, після чого дані розподіляються на тренувальну, валідаційну та тестову вибірки.

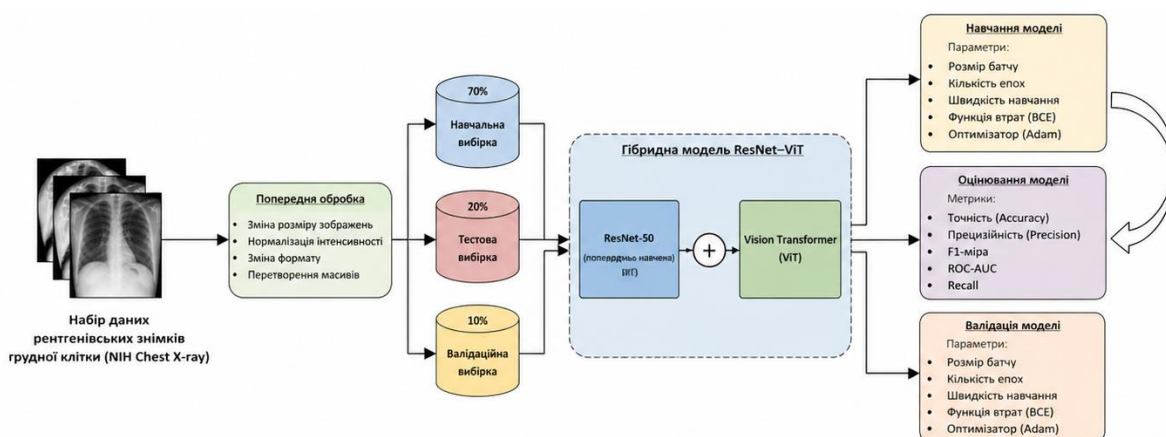


Рисунок 2.37 – Загальна схема функціонування гібридної моделі класифікації рентгенівських знімків

Основним компонентом архітектури є згортковий модуль на базі ResNet-50, який виконує екстракцію локальних ознак із вхідного зображення (рисунк 2.38). Вхідні дані проходять через послідовність згорткових блоків, організованих у вигляді стадій, що дозволяє формувати багаторівневі представлення ознак. Використання residual-зв'язків

забезпечує стабільність навчання та ефективну передачу інформації між шарами (рисунк 2.38).

Отримана карта ознак перетворюється у послідовність векторів шляхом розбиття на патчі та їх лінійного перетворення. Це дозволяє адаптувати вихід згорткової мережі до формату, придатного для подальшої обробки трансформером.

Наступним етапом є трансформерний модуль, реалізований у вигляді Vision Transformer, який забезпечує моделювання глобальних залежностей між різними частинами зображення. У його основі лежить стек енкодерів із механізмом multi-head attention, що дозволяє враховувати взаємозв'язки між усіма патчами зображення. Завершальним етапом є класифікаційний модуль, який включає MLP-голову та функцію активації типу sigmoid для реалізації багатоміткової класифікації.

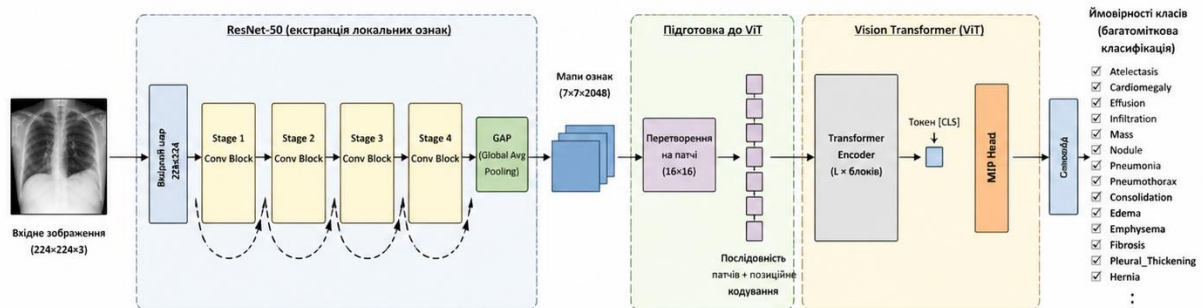


Рисунок 2.38 – Деталізована структура основних модулів гібридної архітектури ViT-ResNet

2.8.5 Формалізація алгоритму навчання гібридної моделі

Алгоритм навчання (рисунк 2.39) запропонованої гібридної моделі базується на класичному підході навчання глибоких нейронних мереж із використанням методу зворотного поширення помилки. Вхідними даними є

рентгенівські знімки грудної клітки та відповідні багатоміткові анотації, що відображають наявність або відсутність патологій.

На першому етапі виконується попередня обробка зображень, яка включає нормалізацію інтенсивності, зміну розміру та приведення до єдиного формату. Після цього зображення подається на вхід згорткового модуля ResNet, який виконує екстракцію локальних ознак. Отримані карти ознак перетворюються у послідовність векторів шляхом розбиття на патчі та лінійного проєктування.

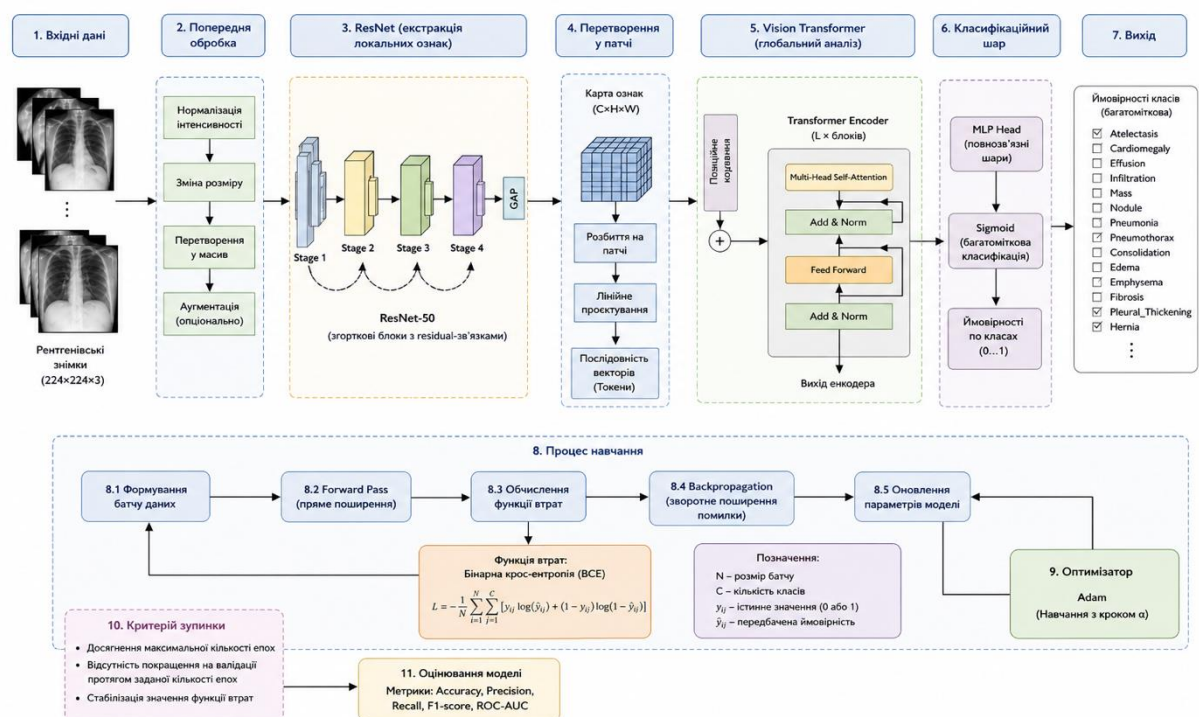


Рисунок 2.39 – Алгоритм навчання гібридної моделі ViT-ResNet

Далі сформована послідовність подається на вхід трансформерного енкодера, який за допомогою механізму self-attention здійснює моделювання глобальних залежностей між елементами зображення. Результатом цього етапу є узагальнене представлення ознак, яке передається до класифікаційного шару.

На етапі класифікації формується вектор ймовірностей належності зображення до кожного з класів, що реалізується за допомогою сигмоїдної функції активації. Для оцінювання якості прогнозу використовується функція втрат бінарної крос-ентропії, що є стандартною для задач багатоміткової класифікації.

Оновлення параметрів моделі здійснюється за допомогою оптимізаційного алгоритму, зокрема Adam, який забезпечує ефективну адаптацію ваг моделі під час навчання. Процес навчання повторюється протягом заданої кількості епох до досягнення стабілізації значення функції втрат на валідаційній вибірці.

Таким чином, алгоритм навчання гібридної моделі поєднує етапи локального виділення ознак, глобального аналізу залежностей та оптимізації параметрів, що забезпечує ефективне вирішення задачі багатоміткової класифікації медичних зображень.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Опис програмного середовища реалізації дослідження

Програмна реалізація дослідження виконувалась із використанням сучасних інструментів машинного навчання, що забезпечують ефективну обробку медичних зображень [14] та навчання глибоких нейронних мереж. Основна увага приділялася створенню відтворюваного експериментального середовища, яке дозволяє реалізувати архітектури CNN, ResNet та Vision Transformer.

3.1.1 Мова програмування та середовище розробки

Для реалізації дослідження обрано мову програмування Python, яка є стандартом де-факто у задачах машинного навчання та аналізу зображень. Вибір обумовлений наявністю широкого спектру бібліотек для роботи з нейронними мережами, а також підтримкою GPU-прискорення.

Розробка програмного забезпечення здійснювалась у середовищі Jupyter Notebook та Visual Studio Code, що забезпечують зручність інтерактивної роботи з даними та поступове тестування моделей. Такий підхід дозволяє поетапно реалізовувати алгоритми, аналізувати результати навчання та візуалізувати метрики якості.

3.1.2 Бібліотеки для обробки даних і навчання моделей

Для обробки даних та реалізації моделей використано набір спеціалізованих бібліотек Python. Основною бібліотекою для побудови та навчання нейронних мереж є TensorFlow [9] та його високорівневий інтерфейс Keras, що дозволяє ефективно реалізовувати архітектури CNN і ResNet.

Для реалізації трансформерних моделей застосовано бібліотеки, що підтримують механізм self-attention та роботу з патчами зображень, що є ключовим для Vision Transformer [15]. Обробка зображень здійснюється за допомогою бібліотек OpenCV та PIL. Для роботи з масивами даних використовується NumPy [10], а для організації датасетів – Pandas. Візуалізація результатів навчання та оцінювання моделей реалізована з використанням бібліотек Matplotlib та Seaborn, що дозволяє будувати графіки точності, втрат та ROC-кривих.

Таким чином, обране програмне середовище забезпечує повну реалізацію експериментального дослідження, включаючи обробку даних, навчання моделей різних архітектур та аналіз отриманих результатів.

3.2 Програмна підготовка набору даних до навчання моделей

Підготовка набору даних є одним із ключових етапів реалізації системи класифікації медичних зображень, оскільки саме від якості вхідних даних залежить ефективність навчання моделей та достовірність отриманих результатів. У задачах аналізу рентгенівських знімків особливу складність становлять неоднорідність зображень, наявність шумів та різна якість вихідних даних.

У межах даного дослідження програмна підготовка даних включає завантаження та структурування набору зображень, приведення їх до єдиного формату, а також формування цільових міток для багатоміткової класифікації. Важливим аспектом є забезпечення узгодженості розмірності та кольорового представлення зображень, що необхідно для коректної роботи моделей глибокого навчання.

Окрім цього, на етапі підготовки даних застосовуються методи попередньої обробки, спрямовані на покращення якості зображень та підвищення інформативності ознак. Такі процедури дозволяють зменшити вплив артефактів і підвищити стійкість моделей до варіацій у вхідних даних.

Таким чином, програмна підготовка набору даних забезпечує формування якісного та структурованого вхідного середовища для навчання моделей, що є необхідною умовою досягнення високої точності класифікації у задачах медичної діагностики.

3.2.1 Завантаження та структура даних

На початковому етапі виконується завантаження даних із табличного файлу, що містить інформацію про зображення та відповідні їм мітки захворювань. Дані організовані таким чином, що для кожного запису визначено ім'я файлу зображення та набір патологій, які до нього відносяться. Це дозволяє ефективно формувати відповідність між зображенням і його семантичним описом.

Формування структури даних передбачає створення окремих списків шляхів до зображень та їх міток, що спрощує подальше використання даних у процесі навчання моделей. Реалізація цього підходу наведена у лістингу 3.1, де здійснюється зчитування CSV-файлу та побудова повних шляхів до зображень на основі їх розташування у файловій системі.

Лістинг 3.1 – Завантаження та формування структури даних

```
def load_dataset(csv_path, image_dir):  
    df = pd.read_csv(csv_path)  
    image_paths = []  
    labels = []  
    for _, row in df.iterrows():  
        image_path = os.path.join(image_dir, row['Image  
Index'])  
        image_paths.append(image_path)  
        labels.append(row['Finding Labels'])  
    return image_paths, labels
```

3.2.2 Попередня обробка рентгенівських знімків

Попередня обробка зображень є необхідним етапом перед подачею даних до моделей глибокого навчання, оскільки дозволяє забезпечити уніфікований формат вхідних даних та підвищити стабільність процесу навчання. У даному дослідженні всі рентгенівські знімки приводяться до однакового розміру, що відповідає вимогам архітектур CNN, ResNet та Vision Transformer. Крім того, виконується нормалізація значень пікселів, що дозволяє зменшити вплив різних масштабів інтенсивності на процес навчання.

Реалізація попередньої обробки наведена у лістингу 3.2, де здійснюється зчитування зображення, декодування у формат RGB, масштабування до розміру 224×224 пікселів та нормалізація значень до діапазону $[0; 1]$.

Лістинг 3.2 – Попередня обробка рентгенівських знімків

```
def preprocess_image(image_path):  
    image = tf.io.read_file(image_path)  
    image = tf.image.decode_jpeg(image, channels=3)  
    image = tf.image.resize(image, [224, 224])  
    image = tf.cast(image, tf.float32) / 255.0  
    return image
```

3.2.3 Формування цільових міток для багатоміткової класифікації

У задачах медичної діагностики одне зображення може відповідати кільком патологіям одночасно, що зумовлює необхідність використання багатоміткової класифікації. Для коректного навчання моделей текстові мітки перетворюються у числове представлення у вигляді багатовимірного вектора, де кожна позиція відповідає окремому класу захворювання.

Реалізація формування таких міток наведена у лістингу 3.3, де кожен набір текстових міток перетворюється у multi-hot вектор, що містить значення 1 для наявних патологій та 0 для відсутніх. Це дозволяє ефективно використовувати функції втрат для багатоміткової класифікації та забезпечує коректне навчання моделей.

Лістинг 3.3 – Формування міток для багатоміткової класифікації

```
def encode_labels(labels_list, class_names):
    label_map = {label: i for i, label in
enumerate(class_names)}
    encoded_labels = []
    for labels in labels_list:
        multi_hot = np.zeros(len(class_names))
        for label in labels.split('|'):
            if label in label_map:
                multi_hot[label_map[label]] = 1
        encoded_labels.append(multi_hot)
    return np.array(encoded_labels)
```

3.3 Програмна реалізація моделей класифікації

Після підготовки набору даних наступним етапом є програмна реалізація моделей класифікації, які використовуються для розпізнавання патологій на рентгенівських знімках. У межах даного дослідження реалізовано кілька архітектур глибокого навчання, що відрізняються підходами до вилучення ознак та обробки зображень.

Реалізація моделей передбачає визначення їх архітектури, налаштування гіперпараметрів, а також інтеграцію з підготовленими даними. Особлива увага приділяється узгодженню структури моделей із форматом вхідних даних та типом задачі багатоміткової класифікації.

3.3.1 Реалізація моделі CNN

Згорткова нейронна мережа використовується як базова модель для класифікації зображень, оскільки вона ефективно виділяє просторові ознаки та локальні патерни. Реалізація моделі наведена у лістингу 3.4. На вихідному шарі застосовано сигмоїдну функцію активації, що забезпечує можливість багатоміткової класифікації.

Лістинг 3.4 – Програмна реалізація моделі CNN

```
def build_cnn_model(num_classes: int) -> tf.keras.Model:
    model = tf.keras.Sequential([
        tf.keras.layers.Input(shape=(224, 224, 3)),
        tf.keras.layers.Conv2D(32, (3, 3),
activation='relu', padding='same'),
        tf.keras.layers.MaxPooling2D((2, 2)),
        tf.keras.layers.Conv2D(64, (3, 3),
activation='relu', padding='same'),
        tf.keras.layers.MaxPooling2D((2, 2)),
        tf.keras.layers.Flatten(),
        tf.keras.layers.Dense(512, activation='relu'),
        tf.keras.layers.Dropout(0.3),
        tf.keras.layers.Dense(num_classes,
activation='sigmoid')
    ])
    model.compile(
        optimizer=tf.keras.optimizers.Adam(),
        loss='binary_crossentropy',
        metrics=[
            'binary_accuracy',
            tf.keras.metrics.AUC(name='auc',
multi_label=True)
        ])
    return model
```

3.3.2 Реалізація моделі ResNet

Модель ResNet використовується для підвищення ефективності вилучення ознак за рахунок застосування залишкових з'єднань, які дозволяють будувати глибші нейронні мережі без втрати якості навчання. Такий підхід забезпечує стабільне поширення градієнтів під час навчання та дозволяє моделі ефективніше виявляти складні патерни у медичних зображеннях.

У даному дослідженні використано попередньо навчений варіант ResNet50, що дозволяє застосувати переваги трансферного навчання. Базова частина моделі використовується як екстрактор ознак і не навчається на початковому етапі, що зменшує обчислювальні витрати та пришвидшує збіжність моделі. Додаткові шари формують класифікаційну частину, адаптовану до задачі багатоміткової класифікації.

Реалізація моделі наведена у лістингу 3.5, де використано глобальне усереднення ознак, механізм регуляризації Dropout та вихідний шар із сигмоїдною активацією. Для навчання застосовується оптимізатор Adam із зменшеним коефіцієнтом навчання, що забезпечує більш стабільну адаптацію попередньо навченої моделі до нових даних.

Лістинг 3.5 – Програмна реалізація моделі ResNet

```
def build_resnet_model(num_classes: int) -> tf.keras.Model:
    base_model = tf.keras.applications.ResNet50(
        include_top=False,
        weights='imagenet',
        input_shape=(224, 224, 3)
    )
    base_model.trainable = False
    inputs = tf.keras.Input(shape=(224, 224, 3))
    x =
    tf.keras.applications.resnet50.preprocess_input(inputs)
    x = base_model(x, training=False)
```

Продовження лістингу 3.5

```
x = tf.keras.layers.GlobalAveragePooling2D()(x)
x = tf.keras.layers.Dropout(0.3)(x)
outputs = tf.keras.layers.Dense(num_classes,
activation='sigmoid')(x)
model = tf.keras.Model(inputs, outputs)
model.compile(
optimizer=tf.keras.optimizers.Adam(learning_rate=1e-4),
loss='binary_crossentropy',
metrics=[
'binary_accuracy',
tf.keras.metrics.AUC(name='auc',
multi_label=True)
]
)
return model
```

3.3.3 Реалізація моделей ViT-v1 та ViT-v2

Моделі Vision Transformer реалізують підхід до обробки зображень, заснований на механізмі самоуваги, що дозволяє ефективно враховувати глобальні залежності між різними ділянками зображення. На відміну від згорткових мереж, такі моделі не використовують локальні фільтри, а працюють із послідовністю векторів, що формуються з окремих фрагментів зображення.

Формування таких фрагментів, або патчів, є початковим етапом роботи моделі та реалізується шляхом розбиття зображення на неперекривні області фіксованого розміру.

Отримані патчі перетворюються у послідовність векторів, яка подається на вхід трансформерної частини моделі. Відповідна реалізація цього етапу наведена у лістингу 3.6.

Лістинг 3.6 – Програмний код, що реалізує формування патчів у Vision Transformer

```
class PatchExtractor(tf.keras.layers.Layer):
    def __init__(self, patch_size):
        super().__init__()
        self.patch_size = patch_size
    def call(self, images):
        patches = tf.image.extract_patches(
            images=images,
            sizes=[1, self.patch_size, self.patch_size, 1],
            strides=[1, self.patch_size, self.patch_size, 1],
            rates=[1, 1, 1, 1],
            padding='VALID')
        batch_size = tf.shape(images)[0]
        patch_dim = patches.shape[-1]
        patches = tf.reshape(patches, [batch_size, -1,
patch_dim])
        return patches
```

Безпосередньо архітектура моделей ViT-v1 та ViT-v2 реалізована як послідовність блоків самоуваги та багат шарового перцептронну, що дозволяє моделі навчатися складним просторовим залежностям. Реалізація наведена у лістингу 3.7, де також передбачено можливість використання різних оптимізаторів, що дозволяє варіювати процес навчання моделей. Завершальний етап моделі включає агрегацію ознак та класифікаційний шар із сигмоїдною функцією активації для задачі багатоміткової класифікації. Використання глобального усереднення дозволяє зменшити розмірність ознак перед класифікацією та підвищити стійкість моделі до варіацій у даних. Застосування кількох блоків самоуваги сприяє більш глибокому аналізу взаємозв'язків між різними ділянками зображення. Це, у свою чергу, дозволяє моделі ефективніше виявляти складні патологічні структури.

Лістинг 3.7 – Програмний код, що реалізує моделі ViT-v1 та ViT-v2

```

def build_vit_model(num_classes, patch_size=32,
optimizer_type='adamw'):
    inputs = tf.keras.Input(shape=(224, 224, 3))
    patches = PatchExtractor(patch_size)(inputs)
    x = tf.keras.layers.Dense(128)(patches)
    for _ in range(4):
        x1 = tf.keras.layers.LayerNormalization()(x)
        attention = tf.keras.layers.MultiHeadAttention(
            num_heads=4, key_dim=128)(x1, x1)
        x2 = x + attention
        x3 = tf.keras.layers.LayerNormalization()(x2)
        mlp = tf.keras.Sequential([
            tf.keras.layers.Dense(256, activation='gelu'),
            tf.keras.layers.Dense(128)]) (x3)
        x = x2 + mlp
    x = tf.keras.layers.GlobalAveragePooling1D()(x)
    outputs = tf.keras.layers.Dense(num_classes,
activation='sigmoid')(x)
    if optimizer_type == 'adamw':
        optimizer = tf.keras.optimizers.AdamW()
    elif optimizer_type == 'sgd':
        optimizer = tf.keras.optimizers.SGD()
    else:
        optimizer = tf.keras.optimizers.Adam()
    model = tf.keras.Model(inputs, outputs)
    model.compile(
        optimizer=optimizer,
        loss='binary_crossentropy',
        metrics=[
            'accuracy',
            tf.keras.metrics.AUC(name='auc',
multi_label=True)
        ])
    return model

```

3.3.4 Реалізація гібридної моделі ViT-ResNet

Гібридна модель ViT-ResNet поєднує переваги згорткових нейронних мереж та трансформерних архітектур, що дозволяє ефективно використовувати як локальні, так і глобальні ознаки зображення. Реалізація моделі наведена у лістингу 3.8, де поєднуються згортковий блок ResNet та шар Multi-Head Attention.

Лістинг 3.8 – Програмна реалізація гібридної моделі ViT-ResNet

```
def build_vit_resnet_model(num_classes):
    base_model = tf.keras.applications.ResNet50(
        include_top=False,
        weights='imagenet',
        input_shape=(224, 224, 3)
    )
    base_model.trainable = False
    inputs = tf.keras.Input(shape=(224, 224, 3))
    x = base_model(inputs)
    x = tf.keras.layers.Reshape((-1, x.shape[-1]))(x)
    x = tf.keras.layers.MultiHeadAttention(
        num_heads=4,
        key_dim=128
    )(x, x)
    x = tf.keras.layers.GlobalAveragePooling1D()(x)
    outputs = tf.keras.layers.Dense(num_classes,
activation='sigmoid')(x)
    model = tf.keras.Model(inputs, outputs)
    model.compile(
        optimizer='adam',
        loss='binary_crossentropy',
        metrics=['accuracy']
    )
    return model
```

У запропонованій реалізації базова модель ResNet50 використовується як попередньо навчений екстрактор ознак, що дозволяє зменшити обчислювальні витрати та підвищити якість початкового представлення даних. Отримані ознаки перетворюються у послідовність векторів, до яких застосовується механізм багатоголової самоуваги. Це дозволяє моделі враховувати глобальні залежності, які не завжди ефективно виявляються лише згортковими шарами.

Завершальний етап включає агрегацію ознак та класифікаційний шар із сигмоїдною функцією активації. Такий підхід дозволяє підвищити якість класифікації за рахунок комбінування різних механізмів вилучення ознак та забезпечує більш гнучке представлення складних структур медичних зображень.

3.4 Реалізація процесу навчання моделей

Процес навчання моделей передбачає налаштування функції втрат, вибір метрик оцінювання та конфігурацію оптимізатора. Реалізація цих компонентів наведена у лістингу 3.9, де використовується функція втрат, а також метрики точності та AUC для комплексної оцінки якості моделей.

Лістинг 3.9 – Програмний код, що реалізує функції втрат та метрик

```
def compile_model(model, optimizer):
    model.compile(
        optimizer=optimizer,
        loss=tf.keras.losses.BinaryCrossentropy(),
        metrics=[
            tf.keras.metrics.BinaryAccuracy(name='accuracy'),
            tf.keras.metrics.AUC(name='auc', multi_label=True)
        ]
    )
    return model
```

У таблиці 3.1 наведено основні гіперпараметри трансформерних моделей, зокрема типи оптимізаторів та функцій втрат, що використовуються у процесі навчання. Вибір оптимізатора безпосередньо впливає на швидкість збіжності моделі та стабільність процесу навчання.

Для моделі ViT-v1 використовується оптимізатор AdamW, який поєднує адаптивне налаштування швидкості навчання з регуляризацією ваг, що дозволяє покращити узагальнюючу здатність моделі. У випадку ViT-v2 застосовується оптимізатор SGD, який характеризується меншою обчислювальною складністю та може забезпечити швидше навчання за певних умов.

Гібридна модель ViT-ResNet використовує оптимізатор Adam разом із функцією втрат Binary Crossentropy with Logits, що дозволяє більш стабільно працювати з неактивованими вихідними значеннями моделі. Такий вибір гіперпараметрів забезпечує баланс між точністю класифікації та ефективністю навчання моделей різних архітектур.

Таблиця 3.1 – Гіперпараметри ViT моделей

Модель	Оптимізатор	Функція втрат
ViT-v1/32	AdamW	Binary Cross Entropy
ViT-v2/32	SGD	Binary Cross Entropy
ViT-ResNet/16	Adam	BCE with Logits

Безпосередній процес навчання моделей реалізовано у лістингу 3.10, де використовується механізм ітеративного оновлення ваг на основі навчальної вибірки з контролем якості на валідаційних даних. Додатково застосовується стратегія адаптивного зменшення швидкості навчання, що дозволяє уникнути застрягання у локальних мінімумах. Використання callback-механізмів дає змогу автоматично регулювати параметри навчання залежно від динаміки функції втрат, що підвищує ефективність оптимізації.

Лістинг 3.10 – Програмний код, що реалізує навчання моделей

```
def train_model(model, train_dataset, val_dataset,
epochs=10):
    reduce_lr = tf.keras.callbacks.ReduceLROnPlateau(
        monitor='val_loss',
        factor=0.3,
        patience=2,
        min_lr=1e-6)
    history = model.fit(
        train_dataset,
        validation_data=val_dataset,
        epochs=epochs,
        callbacks=[reduce_lr])
    return history
```

Ініціалізація оптимізаторів реалізована у лістингу 3.11, де передбачено вибір різних алгоритмів оптимізації залежно від типу моделі. Можливість динамічного вибору оптимізатора спрощує проведення експериментів та порівняння моделей між собою. Це також дозволяє досліджувати вплив різних стратегій оптимізації на якість класифікації.

Лістинг 3.11 – Програмний код, що реалізує ініціалізація оптимізаторів

```
def get_optimizer(optimizer_type='adam'):
    if optimizer_type == 'adam':
        return tf.keras.optimizers.Adam()
    elif optimizer_type == 'adamw':
        return tf.keras.optimizers.AdamW()
    elif optimizer_type == 'sgd':
        return tf.keras.optimizers.SGD()
    else:
        return tf.keras.optimizers.Adam()
```

3.5 Реалізація засобів оцінювання результатів

Оцінювання результатів навчання моделей є завершальним етапом програмної реалізації, що дозволяє визначити ефективність побудованих архітектур у задачі класифікації медичних зображень. Для цього використовуються як числові метрики якості, так і графічні методи аналізу, що забезпечують комплексну оцінку поведінки моделей. У межах дослідження реалізовано засоби для обчислення основних показників якості, побудови ROC-кривих та візуалізації внутрішніх механізмів трансформерних моделей.

3.5.1 Обчислення показників Accuracy та AUC

Для кількісного оцінювання якості моделей використовуються метрики Accuracy та AUC, які дозволяють визначити точність класифікації та здатність моделі відрізняти позитивні та негативні класи. Accuracy відображає частку правильно класифікованих прикладів, тоді як AUC є більш інформативною метрикою для багатоміткових задач.

Реалізація обчислення метрик наведена у лістингу 3.12, де використовується вбудований механізм оцінювання моделі на тестовій вибірці. Отримані значення метрик формуються у вигляді структурованого набору даних, що дозволяє зручно аналізувати результати різних моделей.

Лістинг 3.12 – Програмний код, що реалізує обчислення метрик якості моделі

```
def evaluate_model(model, test_dataset):  
    results = model.evaluate(test_dataset)  
    metrics = dict(zip(model.metrics_names, results))  
    return metrics
```

3.5.2 Побудова ROC-кривих і кривих навчання

Графічний аналіз результатів навчання дозволяє більш детально оцінити поведінку моделей у процесі класифікації. ROC-крива відображає залежність між часткою істинно позитивних та хибнопозитивних результатів, що дозволяє оцінити якість моделі незалежно від порогового значення.

Реалізація побудови ROC-кривої наведена у лістингу 3.13, де здійснюється накопичення прогнозів моделі та відповідних істинних значень, після чого обчислюються показники FPR та TPR. Отримана крива дозволяє візуально оцінити ефективність моделі та порівняти різні архітектури між собою.

Лістинг 3.13 – Програмний код, що реалізує побудову ROC-кривої

```
def plot_roc_curve(model, dataset):  
    y_true = []  
    y_pred = []  
    for images, labels in dataset:  
        preds = model.predict(images)  
        y_true.append(labels.numpy())  
        y_pred.append(preds)  
    y_true = np.concatenate(y_true)  
    y_pred = np.concatenate(y_pred)  
  
    fpr, tpr, _ = roc_curve(y_true.ravel(), y_pred.ravel())  
    roc_auc = auc(fpr, tpr)  
  
    plt.plot(fpr, tpr, label=f"AUC = {roc_auc:.2f}")  
    plt.xlabel("False Positive Rate")  
    plt.ylabel("True Positive Rate")  
    plt.legend()  
    plt.show()
```

3.5.3 Формування attention maps для трансформерних моделей

Для трансформерних моделей важливим аспектом є інтерпретованість результатів, яка досягається за рахунок аналізу механізму самоуваги. Attention maps дозволяють візуалізувати, на які ділянки зображення модель звертає найбільшу увагу під час прийняття рішення.

Реалізація формування attention maps наведена у лістингу 3.14, де здійснюється отримання ваг самоуваги з відповідного шару моделі та їх подальша обробка. Візуалізація цих даних у вигляді теплової карти дозволяє проаналізувати поведінку моделі та оцінити її здатність фокусуватися на релевантних областях зображення.

Лістинг 3.14 – Програмний код, що реалізує формування attention map для Vision Transformer

```
def visualize_attention_map(model, image):
    image = tf.expand_dims(image, axis=0)
    predictions = model(image)
    attention_layer = None
    for layer in model.layers:
        if isinstance(layer, tf.keras.layers.MultiHeadAttention):
            attention_layer = layer
            break
    if attention_layer is None:
        return
    attention_output = attention_layer(image, image)
    attention_map = tf.reduce_mean(attention_output, axis=-1)
    attention_map = tf.image.resize(
        tf.expand_dims(attention_map, -1),
        (224, 224))
    plt.imshow(tf.squeeze(attention_map), cmap='jet')
    plt.colorbar()
    plt.show()
```

4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

4.1 План експериментальних досліджень

План експериментальних досліджень побудовано з урахуванням необхідності об'єктивного порівняння різних архітектур глибокого навчання в задачі багатоміткової класифікації рентгенівських знімків грудної клітки.

У межах дослідження передбачено реалізацію та навчання п'яти моделей: CNN, ResNet, ViT-v1, ViT-v2 та гібридної моделі ViT-ResNet, що відповідає підходу, використаному в аналізованій статті. Кожна модель навчається за однакових умов підготовки даних і використовує однаковий формат вхідних зображень, що забезпечує коректність подальшого порівняння.

Експеримент включає етапи підготовки даних, навчання моделей, оцінювання якості та аналізу результатів. Особлива увага приділяється впливу архітектури моделі та використанню попереднього навчання на кінцеву точність класифікації.

Отримані результати використовуються для побудови графіків навчання, ROC-кривих та проведення порівняльного аналізу ефективності моделей з метою визначення найбільш доцільного підходу для задач підтримки прийняття рішень у медичній діагностиці.

4.1.1 Мета та гіпотеза експерименту

Метою експерименту є оцінювання ефективності різних архітектур глибокого навчання у задачі багатоміткової класифікації рентгенівських знімків грудної клітки та визначення впливу сучасних підходів, зокрема трансформерних моделей і попереднього навчання, на якість класифікації.

Особливу увагу приділено дослідженню можливостей Vision Transformer у порівнянні з класичними згортковими нейронними мережами.

Гіпотеза експерименту полягає в тому, що трансформерні моделі, зокрема Vision Transformer, здатні забезпечити вищу якість класифікації медичних зображень порівняно з традиційними CNN за рахунок використання механізму самоуваги [16]. Додатково передбачається, що застосування попереднього навчання на великих наборах даних дозволяє суттєво покращити результати моделей, особливо у випадку гібридних архітектур, таких як ViT-ResNet.

4.1.2 Критерії порівняння моделей

Порівняння моделей у межах експерименту здійснюється на основі кількісних показників, що характеризують якість багатоміткової класифікації.

Основними критеріями є точність класифікації на тренувальній, валідаційній та тестовій вибірках, що дозволяє оцінити здатність моделі до узагальнення та виявити можливі ознаки перенавчання. Додатково використовується показник площі під ROC-кривою (AUC), який дає змогу оцінити здатність моделі розрізняти класи незалежно від обраного порогу класифікації.

Окрім числових метрик, для порівняння моделей використовуються графічні інструменти аналізу, зокрема ROC-криві та криві навчання. Вони дозволяють візуально оцінити динаміку навчання моделей і порівняти їх поведінку на різних етапах тренування. Для трансформерних моделей додатково враховується інтерпретованість результатів за допомогою attention maps, що дає змогу оцінити якість виділення інформативних областей зображення та підвищує довіру до отриманих результатів у контексті медичної діагностики.

4.2 Характеристика набору даних NIH Chest X-ray

Набір даних NIH Chest X-ray є одним із найбільш масштабних відкритих ресурсів для досліджень у галузі автоматизованого аналізу медичних зображень, зокрема рентгенографії грудної клітки. Він був сформований у межах дослідницької ініціативи Національного інституту здоров'я США та містить понад 100 тисяч рентгенівських знімків, отриманих від більш ніж 30 тисяч пацієнтів. Зображення представлені у фронтальній проєкції, що є стандартом для первинної діагностики захворювань органів грудної клітки, і мають відповідні текстові анотації, що описують наявні патології.

Кожне зображення у наборі даних супроводжується однією або кількома мітками, які відповідають певним захворюванням або патологічним станам. Усього розглядається до 14 основних класів, серед яких, зокрема, пневмонія, ателектаз, кардіомегалія, емфізема, пневмоторакс та інші. Важливою особливістю набору є багатомітковий характер анотацій, що означає можливість одночасної присутності кількох патологій на одному зображенні. Це значно ускладнює задачу класифікації та вимагає застосування спеціалізованих моделей і підходів до навчання.

Формування міток у наборі даних здійснювалося автоматично на основі обробки текстових медичних звітів за допомогою методів обробки природної мови. Такий підхід дозволив отримати великий обсяг розмічених даних, однак водночас призвів до появи певного рівня шуму та неточностей у мітках. У результаті частина зображень може містити некоректні або неповні анотації, що створює додаткові труднощі для навчання моделей і впливає на достовірність отриманих результатів.

Ще однією суттєвою характеристикою набору є значний дисбаланс між класами. Деякі патології представлені великою кількістю прикладів, тоді як інші зустрічаються значно рідше. Така нерівномірність розподілу даних ускладнює процес навчання, оскільки моделі можуть бути схильні до

переважного розпізнавання більш поширених класів. У зв'язку з цим виникає необхідність використання відповідних методів обробки даних, зокрема аугментації та налаштування функцій втрат, що дозволяє частково компенсувати вплив дисбалансу.

Зображення у наборі даних мають різну роздільну здатність і якість, що обумовлено особливостями процесу їх отримання в клінічних умовах. Для використання у задачах глибокого навчання вони попередньо приводяться до єдиного формату, зокрема масштабу 224×224 пікселів, що відповідає вимогам більшості сучасних архітектур нейронних мереж. Додатково застосовується нормалізація значень пікселів і, за необхідності, методи аугментації, такі як випадкові обертання та віддзеркалення, що сприяє підвищенню узагальнюючої здатності моделей.

Таким чином, набір даних NIH Chest X-ray є репрезентативним джерелом медичних зображень для задач автоматизованої діагностики, що дозволяє проводити комплексні дослідження ефективності різних підходів глибокого навчання. Водночас його особливості, зокрема багатомітковість, шум у мітках та дисбаланс класів, створюють додаткові виклики, які необхідно враховувати при побудові та оцінюванні моделей класифікації.

4.3 Результати дослідження базових моделей

У межах експериментального дослідження проведено навчання та оцінювання базових моделей класифікації рентгенівських знімків грудної клітки, зокрема згорткової нейронної мережі CNN та моделі ResNet. Дані архітектури розглядаються як базові підходи у задачах комп'ютерного зору та використовуються для подальшого порівняння з більш сучасними трансформерними моделями. Обидві моделі були адаптовані до задачі багатоміткової класифікації, що передбачає можливість одночасного виявлення кількох патологій на одному зображенні.

У процесі навчання моделей було отримано значення точності на тренувальній, валідаційній та тестовій вибірках, що дозволяє оцінити їх здатність до узагальнення. Аналіз кривих навчання показує, що модель CNN демонструє стабільне зростання точності на тренувальній вибірці, однак поступається більш складним архітектурам за якістю узагальнення. Водночас модель ResNet, завдяки використанню residual-зв'язків, забезпечує більш ефективне навчання глибокої мережі та демонструє вищі показники точності, що свідчить про її кращу здатність до виділення інформативних ознак зображення.

Додатково було проведено оцінювання моделей за допомогою ROC-кривих та показника площі під кривою (AUC), що дозволяє більш повно охарактеризувати якість класифікації. Отримані результати свідчать про те, що модель ResNet перевищує CNN за всіма основними метриками, демонструючи кращу здатність до розрізнення класів навіть у складних умовах багатоміткової класифікації. Це підтверджує доцільність використання більш глибоких архітектур для задач медичної діагностики.

Таким чином, базові моделі забезпечують задовільний рівень якості класифікації рентгенівських знімків, однак мають певні обмеження, пов'язані з недостатньою здатністю враховувати складні глобальні залежності між ознаками. Отримані результати використовуються як еталон для подальшого порівняння з трансформерними моделями, що дозволяє оцінити доцільність їх застосування у задачах підтримки прийняття рішень у медичній діагностиці.

4.3.1 Результати моделі CNN

У результаті навчання базової моделі CNN отримано стабільну динаміку зменшення значення функції втрат на тренувальній вибірці, що свідчить про поступове покращення здатності моделі апроксимувати залежності у даних. Початкове значення втрат становить близько 0.214,

після чого спостерігається його монотонне зниження до рівня приблизно 0.201 наприкінці навчання. Аналогічна тенденція характерна і для валідаційної вибірки, хоча крива має незначні коливання (рисунок 4.1).

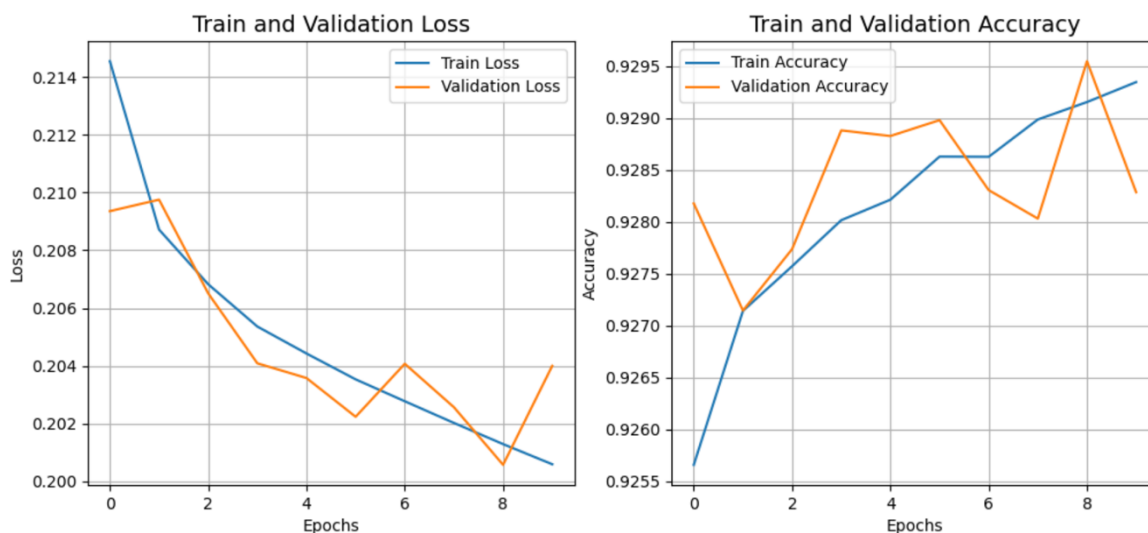


Рисунок 4.1 – Криві втрат та точності моделі CNN на тренувальній і валідаційній вибірках

Аналіз кривих точності показує, що модель CNN демонструє поступове зростання точності на тренувальній вибірці від приблизно 0.926 до 0.929, що підтверджує ефективність процесу навчання. Валідаційна точність також зростає, однак характеризується більшими флуктуаціями, що може свідчити про вплив обмеженої складності моделі та чутливість до варіацій у даних. Незважаючи на досягнення задовільних показників точності, спостерігається певний розрив між тренувальними та валідаційними метриками, що вказує на обмежену здатність моделі CNN до узагальнення у задачі багатоміткової класифікації рентгенівських знімків. Додатково слід зазначити, що після декількох епох приріст точності суттєво сповільнюється, що може свідчити про досягнення моделі свого граничного рівня ефективності.

4.3.2 Результати моделі ResNet

У результаті навчання моделі ResNet спостерігається більш виражене та стабільне зниження значення функції втрат порівняно з базовою моделлю CNN. Початкове значення втрат на тренувальній вибірці становить близько 0.211 і поступово зменшується до приблизно 0.185 наприкінці навчання, що свідчить про ефективне засвоєння ознак моделлю. Валідаційна крива втрат демонструє різке зниження на початкових етапах, після чого стабілізується з незначними коливаннями, що вказує на достатньо хорошу здатність моделі до узагальнення (рисунок 4.2).

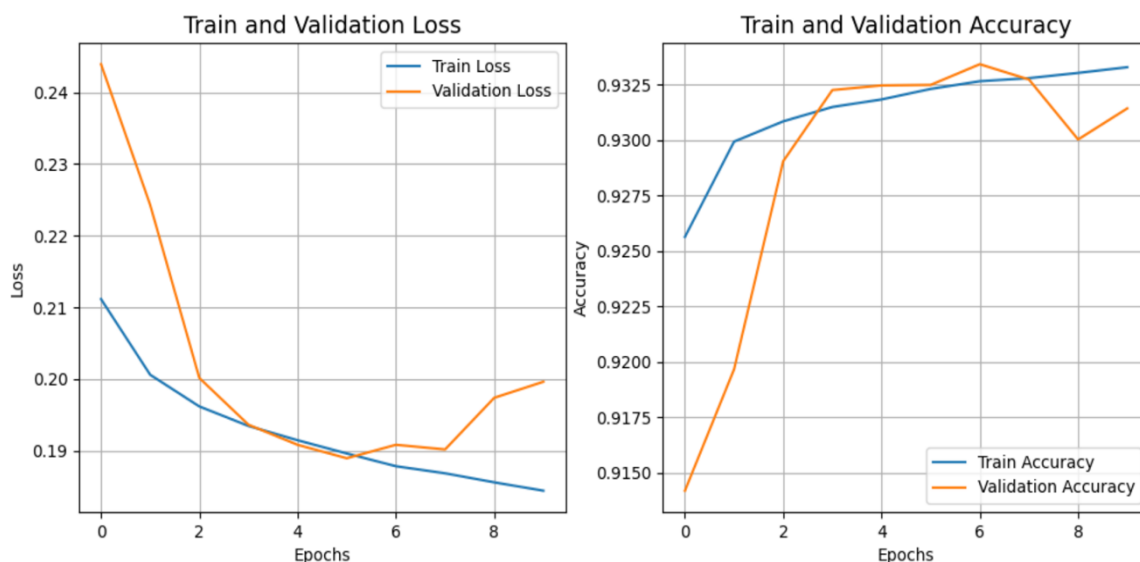


Рисунок 4.2 – Криві втрат та точності моделі ResNet на тренувальній і валідаційній вибірках

Аналіз кривих точності показує, що модель ResNet досягає вищих значень порівняно з CNN, причому тренувальна точність зростає від приблизно 0.926 до 0.933. Валідаційна точність також швидко збільшується на перших епохах і досягає рівня понад 0.932, після чого коливається в невеликому діапазоні.

4.4 Результати дослідження трансформерних моделей

У межах експериментального дослідження проведено навчання та оцінювання трансформерних моделей, зокрема ViT-v1, ViT-v2 та гібридної моделі ViT-ResNet, які базуються на використанні механізму самоуваги для аналізу зображень. На відміну від класичних згорткових нейронних мереж, трансформерні архітектури дозволяють враховувати глобальні залежності між різними ділянками зображення, що є важливим для задач медичної діагностики, де патологічні ознаки можуть бути розподілені по всій області знімка.

Результати навчання показали, що моделі Vision Transformer демонструють більш стабільну динаміку зростання точності та кращу узагальнюючу здатність порівняно з базовими моделями. Зокрема, ViT-v1 і ViT-v2 досягають вищих значень точності на валідаційній вибірці, що свідчить про їх ефективність у задачі багатоміткової класифікації. При цьому відмінності між ViT-v1 та ViT-v2 обумовлені використанням різних оптимізаторів, що впливає на швидкість збіжності та стабільність навчального процесу.

Особливу увагу привертає гібридна модель ViT-ResNet, яка поєднує переваги згорткових мереж та трансформерів. Отримані результати свідчать про те, що дана модель демонструє найвищі показники точності та AUC серед усіх досліджених підходів, що підтверджує ефективність використання попереднього навчання та комбінування різних архітектур. Такий підхід дозволяє більш точно виділяти інформативні ознаки зображення та підвищує якість класифікації.

Таким чином, трансформерні моделі показали кращі результати порівняно з базовими CNN-архітектурами, що підтверджує доцільність їх використання у задачах автоматизованого аналізу медичних зображень.

4.4.1 Результати моделі ViT-v1

У результаті навчання моделі ViT-v1 спостерігається стабільне зниження значення функції втрат як на тренувальній, так і на валідаційній вибірках. Початкове значення втрат на тренувальній вибірці становить близько 0.24 і поступово зменшується до приблизно 0.209 наприкінці навчання, що свідчить про ефективне засвоєння ознак моделлю. Валідаційна крива втрат також демонструє стійке зниження без значних коливань, що вказує на хорошу здатність моделі до узагальнення (рисунок 4.3).

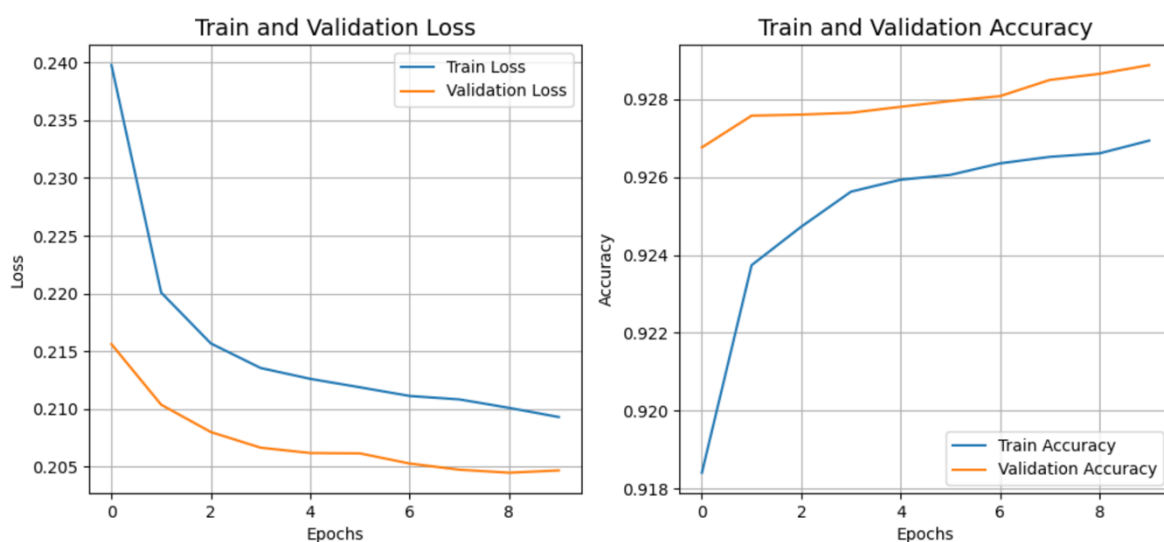


Рисунок 4.3 – Криві втрат та точності моделі ViT-v1 на тренувальній і валідаційній вибірках

Аналіз кривих точності показує, що модель ViT-v1 демонструє впевнене зростання точності на тренувальній вибірці від приблизно 0.918 до 0.927. Валідаційна точність перевищує тренувальну на більшості епох і досягає значень близько 0.929, що свідчить про ефективність використання механізму самоуваги. Відносно невеликий розрив між тренувальною та валідаційною точністю підтверджує здатність моделі узагальнювати знання.

4.4.2 Результати моделі ViT-v2

У процесі навчання моделі ViT-v2 спостерігається різке зниження значення функції втрат на початкових етапах, після чого крива стабілізується. Початкове значення втрат є значно вищим порівняно з іншими моделями, однак вже після кількох епох воно стрімко зменшується та досягає значень близько 2–3, що свідчить про швидку збіжність моделі. Валідаційна крива втрат має подібну динаміку, що вказує на узгодженість процесу навчання (рисунок 4.4).

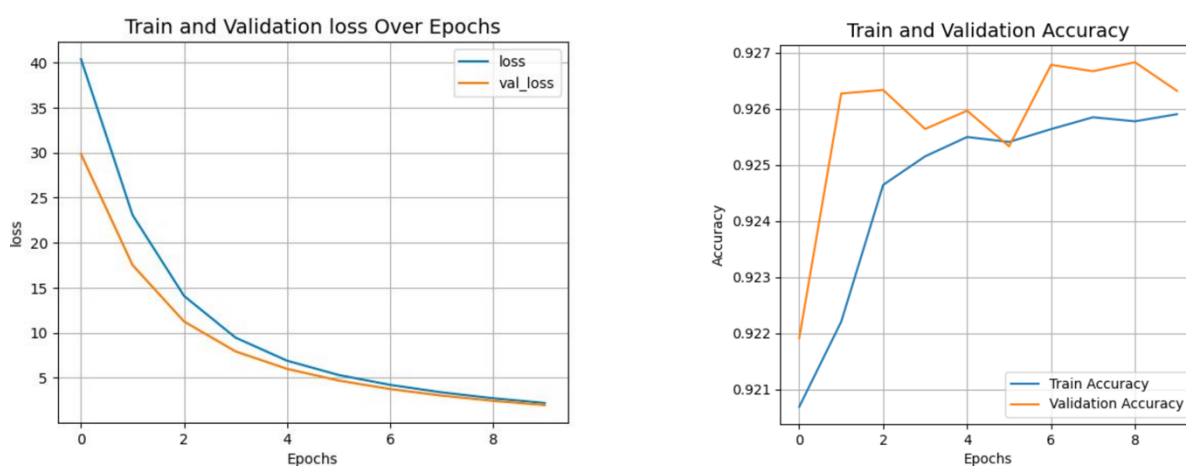


Рисунок 4.4 – Криві втрат та точності моделі ViT-v2 на тренувальній і валідаційній вибірках

Аналіз точності показує, що модель ViT-v2 демонструє стабільне зростання як на тренувальній, так і на валідаційній вибірках. Тренувальна точність підвищується від приблизно 0.921 до 0.926, тоді як валідаційна точність досягає значень близько 0.927 і залишається стабільною протягом останніх епох. Порівняно з ViT-v1, дана модель демонструє дещо вищі показники точності, що може бути зумовлено використанням іншого оптимізатора.

4.4.3 Результати моделі ViT-ResNet

Гібридна модель ViT-ResNet демонструє найвищі показники точності серед усіх досліджених архітектур, що підтверджує ефективність поєднання згорткових мереж і трансформерного підходу. Аналіз динаміки навчання показує, що модель вже на початкових етапах досягає високих значень точності на тренувальній вибірці, які зростають від приблизно 0.931 до понад 0.940 наприкінці навчання (рисунк 4.5). При цьому крива точності має плавний характер без різких коливань, що свідчить про стабільність процесу навчання.

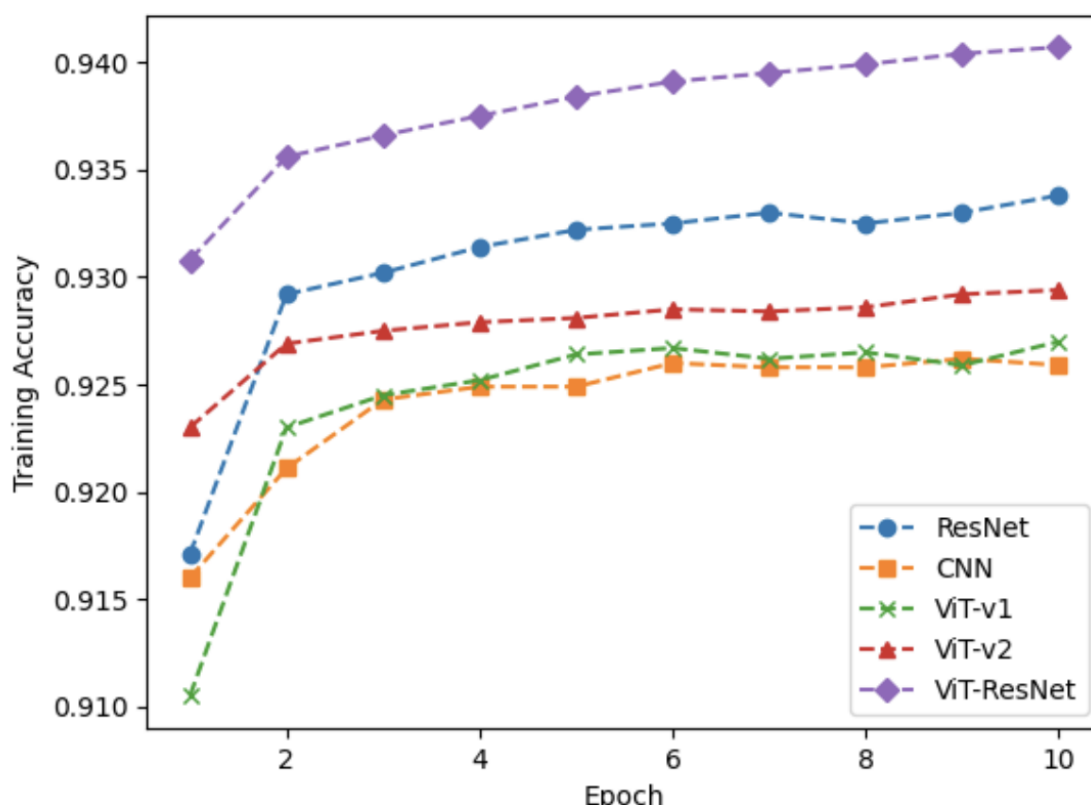


Рисунок 4.5 – Порівняння точності моделей на тренувальній вибірці

На валідаційній вибірці модель також демонструє високі результати, досягаючи значень точності близько 0.932, що є одним із найкращих показників серед усіх моделей (рисунк 4.6). Важливо відзначити, що

розрив між тренувальною та валідаційною точністю залишається незначним, що свідчить про добру узагальнюючу здатність моделі та відсутність вираженого перенавчання. Незважаючи на незначні коливання на окремих епохах, загальна тенденція залишається стабільною, що підтверджує надійність отриманих результатів.

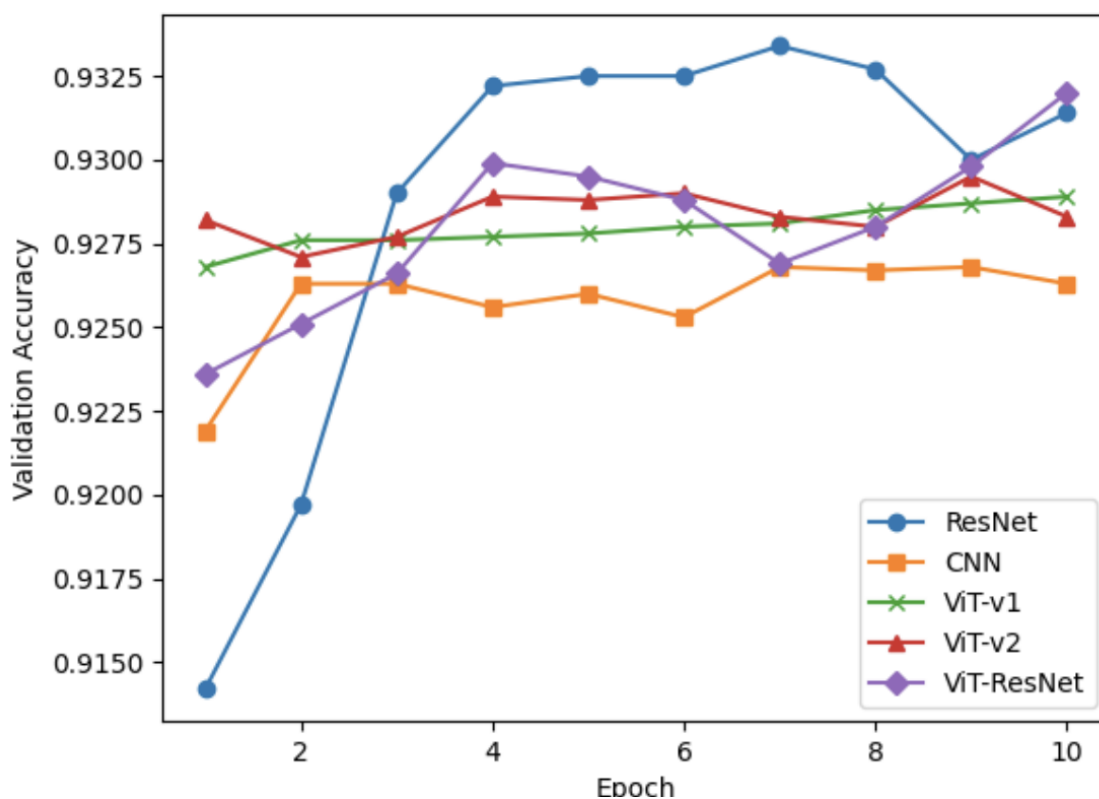


Рисунок 4.6 – Порівняння точності моделей на валідаційній вибірці

Отримані результати свідчать про те, що використання попереднього навчання та гібридної архітектури дозволяє суттєво покращити якість класифікації рентгенівських знімків. Поєднання здатності згорткових мереж ефективно виділяти локальні ознаки з можливостями трансформерів враховувати глобальні залежності забезпечує перевагу моделі ViT-ResNet над як базовими CNN-архітектурами, так і чистими трансформерними моделями.

4.5 Аналіз впливу попереднього навчання на результати класифікації

У межах проведеного дослідження особливу увагу приділено аналізу впливу попереднього навчання на ефективність моделей класифікації, зокрема гібридної архітектури ViT-ResNet. Попереднє навчання на великомасштабних наборах даних дозволяє моделі сформувати базове уявлення про структуру зображень і характерні ознаки, що суттєво спрощує подальше донавчання на спеціалізованих медичних даних. Це особливо важливо для задач медичної діагностики, де доступні набори даних часто мають обмежений обсяг і високий рівень шуму.

Отримані результати показують, що використання попереднього навчання забезпечує підвищення точності класифікації та стабільності навчального процесу. Зокрема, модель ViT-ResNet, яка використовує попередньо навчену згорткову частину, демонструє кращі показники як на тренувальній, так і на валідаційній вибірках порівняно з моделями, що навчалися з нуля. Це свідчить про ефективність перенесення знань з великих загальних датасетів на медичну предметну область.

Додатковим підтвердженням ефективності попереднього навчання є результати візуалізації attention map, які дозволяють проаналізувати, на які області зображення звертає увагу модель під час класифікації.

Модель фокусується (рисунок 4.7) на анатомічно значущих ділянках грудної клітки, що свідчить про коректне виділення інформативних ознак навіть у випадку відсутності патологій. Така поведінка моделі підвищує довіру до результатів її роботи та підтверджує доцільність використання попереднього навчання для покращення інтерпретованості моделей.

Таким чином, попереднє навчання є важливим фактором підвищення ефективності сучасних моделей глибокого навчання у задачах медичної класифікації зображень.

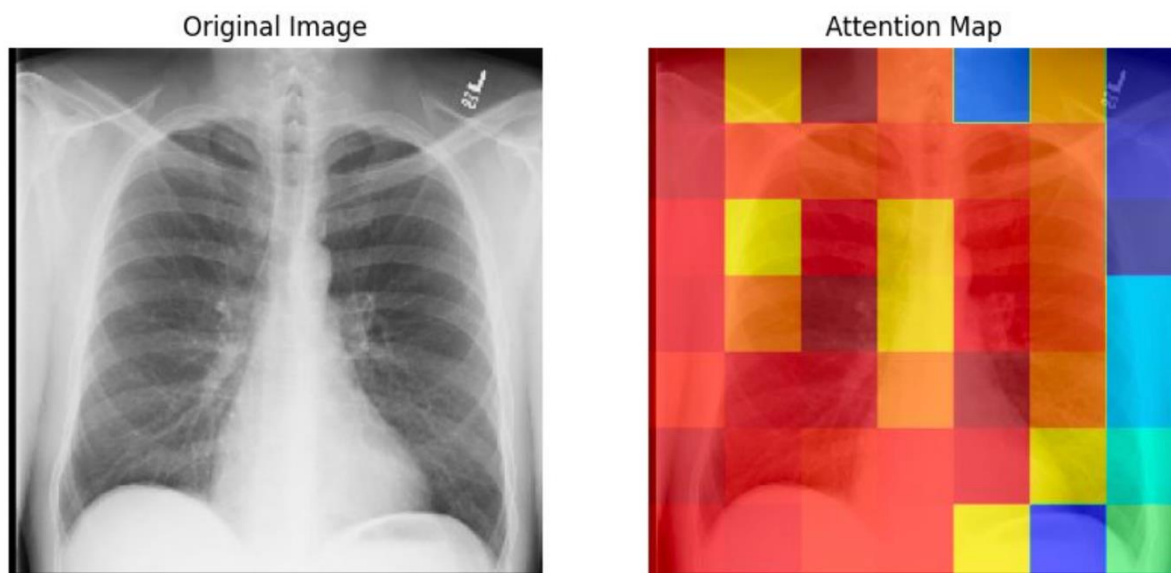


Рисунок 4.7 – Візуалізація attention map для рентгенівського знімка без патологій

4.6 Порівняльний аналіз отриманих результатів

Порівняльний аналіз отриманих результатів дозволяє оцінити ефективність різних архітектур глибокого навчання у задачі багатоміткової класифікації рентгенівських знімків грудної клітки. Відповідно до отриманих значень точності та AUC, наведених у таблиці 4.1, усі моделі демонструють близькі результати, однак між ними спостерігаються суттєві відмінності у здатності до узагальнення. Зокрема, модель CNN показує найнижчі результати серед усіх розглянутих підходів, що пояснюється її відносно простою архітектурою та обмеженими можливостями виділення складних ознак.

Модель ResNet демонструє покращення результатів порівняно з CNN, досягаючи вищих значень точності як на тренувальній, так і на тестовій вибірках. Це підтверджує ефективність використання residual-зв'язків для побудови глибших нейронних мереж.

Водночас трансформерні моделі ViT-v1 та ViT-v2 забезпечують результати, співставні з ResNet, що свідчить про їх здатність ефективно працювати з глобальними залежностями у зображеннях.

Найвищу точність на тестовій вибірці демонструє гібридна модель ViT-ResNet, яка поєднує переваги згорткових та трансформерних підходів. Це підтверджує доцільність комбінування різних архітектур для підвищення якості класифікації медичних зображень.

Таблиця 4.1 – Результати оцінювання якості моделей

Модель	Точність на тестовій вибірці	Точність на тренувальній вибірці	Точність на валідаційній вибірці	AUC
CNN	91	92.62	92.68	0.82
ResNet	93	93.38	93.34	0.86
ViT-v1	92.63	92.7	92.89	0.86
ViT-v2	92.83	92.94	92.95	0.84
ViT-ResNet	93.9	93.02	92.07	0.85

У межах порівняльного аналізу отриманих результатів доцільно окремо розглянути ROC-криві для моделей Vision Transformer та CNN, оскільки вони дають змогу оцінити не лише загальну точність класифікації, а й здатність моделі розрізняти окремі класи захворювань за різних значень порога прийняття рішення.

Для моделі Vision Transformer (рисунок 4.8) характерна відносно стабільна форма ROC-кривих для більшості класів, що свідчить про узгоджену роботу моделі при класифікації різних патологій. Значення площі під кривою для окремих класів знаходяться поблизу 0.5, що вказує на складність задачі та наявність шуму у даних, однак загальна тенденція демонструє здатність моделі ефективно обробляти багатоміткові залежності.

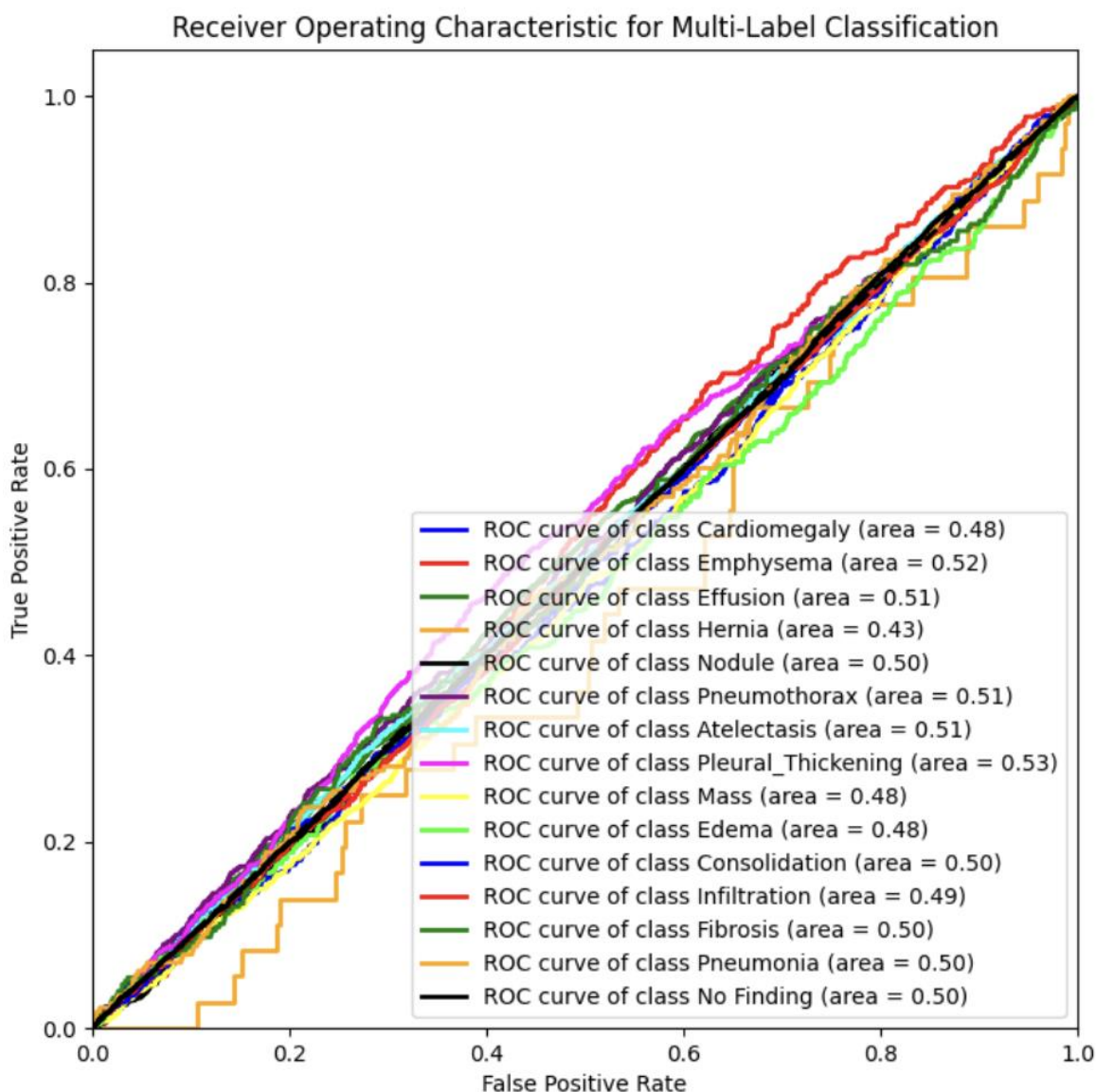


Рисунок 4.8 – ROC-криві багатоміткової класифікації для моделі Vision Transformer на повному наборі даних

Нижче представлено ROC-криві для моделі CNN (рисунок 4.9), отримані на обмеженому наборі даних. У даному випадку спостерігається більша варіативність кривих для різних класів, що свідчить про менш стабільну роботу моделі. Для окремих патологій значення AUC суттєво відрізняються, що вказує на залежність якості класифікації від представленості класів у навчальних даних.

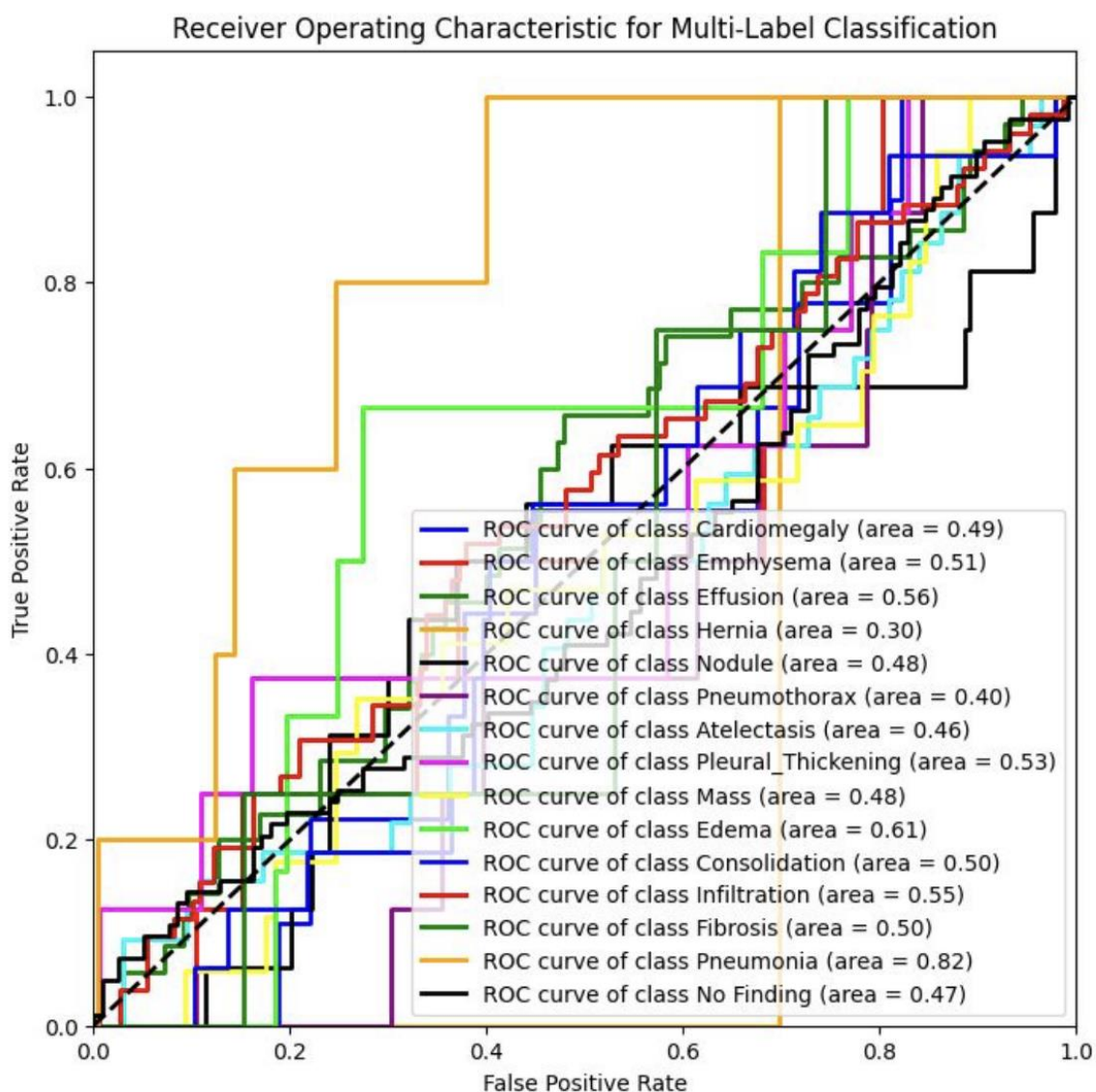


Рисунок 4.9 – ROC-криві багатоміткової класифікації для моделі CNN на обмеженому наборі даних

Таким чином, аналіз ROC-кривих підтверджує висновки, отримані за табличними показниками точності та AUC. Використання моделі Vision Transformer є більш доцільним у задачах автоматизованої багатоміткової класифікації рентгенівських знімків грудної клітки, особливо за наявності достатнього обсягу навчальних даних. Водночас результати моделі CNN свідчать про те, що класичні згорткові архітектури можуть використовуватися як базові моделі для порівняння.

ВИСНОВКИ

У результаті виконання кваліфікаційної роботи було досягнуто поставленої мети – проведено комплексне дослідження методів використання систем штучного інтелекту для підтримки прийняття рішень у задачах медичної діагностики на прикладі класифікації рентгенографічних зображень грудної клітки. У ході роботи було здійснено аналіз сучасного стану предметної галузі, досліджено особливості застосування глибоких нейронних мереж у медичних задачах, а також обґрунтовано доцільність використання різних архітектур, зокрема CNN, ResNet та Vision Transformer. Теоретичні положення було підтверджено експериментальними дослідженнями на реальному наборі медичних даних, що відповідає вимогам до кваліфікаційних робіт магістерського рівня. Робота була апробована на конференції [20], що підтверджує її актуальність, практичну значущість та відповідність сучасному рівню розвитку наукових досліджень у галузі штучного інтелекту.

У практичній частині роботи було реалізовано та досліджено декілька моделей глибокого навчання для задачі багатоміткової класифікації рентгенівських знімків. Проведене експериментальне дослідження дозволило отримати кількісні оцінки якості моделей за такими метриками, як точність на тренувальній, валідаційній та тестовій вибірках, а також значення площі під ROC-кривою. Отримані результати показали, що базова модель CNN демонструє найнижчі показники точності, зокрема близько 91% на тестовій вибірці, що підтверджує її обмежену здатність до узагальнення у складних задачах медичної діагностики. Модель ResNet продемонструвала покращені результати, досягнувши приблизно 93% точності, що пояснюється використанням залишкових зв'язків і глибшої архітектури.

Моделі Vision Transformer показали конкурентоспроможні результати, особливо у випадках, коли використовувались оптимізовані

параметри або попереднє навчання. Зокрема, було встановлено, що трансформерні підходи здатні ефективніше враховувати глобальні залежності в зображеннях, що є критично важливим для медичних задач. Найкращі результати були отримані для гібридної моделі ViT-ResNet, яка поєднує переваги згорткових мереж і механізму самоуваги. Дана модель продемонструвала найвищу тестову точність на рівні приблизно 93.9%, що свідчить про доцільність комбінування різних архітектурних підходів.

Порівняння отриманих результатів із існуючими аналогами показало, що розглянуті моделі відповідають сучасному рівню розвитку технологій у галузі комп'ютерного зору для медичної діагностики. Класичні згорткові мережі, такі як CNN, можуть використовуватись як базові моделі, однак поступаються більш сучасним архітектурам за точністю та стабільністю. ResNet демонструє високий рівень ефективності та залишається одним із найбільш надійних підходів у практичних застосуваннях. У свою чергу, Vision Transformer є перспективним напрямком розвитку, який забезпечує кращу масштабованість і здатність до обробки складних структур даних. Отримані результати узгоджуються з сучасними дослідженнями, у яких зазначається перевага трансформерних моделей при роботі з великими наборами медичних зображень.

Разом із цим у роботі виявлено низку обмежень, що впливають на якість результатів. Зокрема, використання підмножини набору даних може обмежувати здатність моделей до узагальнення. Крім того, значний вплив має дисбаланс класів, що ускладнює навчання моделей для рідкісних патологій. Також слід відзначити високу обчислювальну складність трансформерних моделей, що може бути критичним фактором при їх впровадженні у реальні медичні системи.

Таким чином, у кваліфікаційній роботі повністю виконано поставлені завдання, досягнуто мети дослідження та отримано як теоретично обґрунтовані, так і практично значущі результати.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Topol E. Deep medicine: How artificial intelligence can make healthcare human again. *Nature Medicine*. 2019. Vol. 25, no. 1. P. 44–56.
2. Litjens G. et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017. Vol. 42. P. 60–88.
3. Esteva A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017. Vol. 542. P. 115–118.
4. Rajpurkar P. et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv.org*. URL: <https://arxiv.org/abs/1711.05225> (date of access: 04.02.2025).
5. Lundervold A. S., Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*. 2019. Vol. 29, no. 2. P. 102–127.
6. Shen D., Wu G., Suk H. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*. 2017. Vol. 19. P. 221–248.
7. Artificial intelligence in healthcare. *World Health Organization*. URL: <https://www.who.int/publications/i/item/9789240029200> (date of access: 10.02.2025).
8. Artificial intelligence in medicine: Current trends and future possibilities. *National Institutes of Health*. URL: <https://www.nih.gov/artificial-intelligence> (date of access: 15.02.2025).
9. TensorFlow documentation. Machine learning for healthcare applications. URL: <https://www.tensorflow.org> (date of access: 18.02.2025).
10. PyTorch documentation. Deep learning framework for medical image analysis. URL: <https://pytorch.org/docs/stable/index.html> (date of access: 20.02.2025).
11. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 775 p.

12. Bishop C. M. Pattern Recognition and Machine Learning. New York : Springer, 2006. 738 p.
13. Chollet F. Deep Learning with Python. New York : Manning Publications, 2018. 384 p.
14. Artificial intelligence in medical imaging. *Radiological Society of North America*. URL: <https://www.rsna.org/artificial-intelligence> (date of access: 01.03.2025).
15. Multimodal machine learning: A survey and taxonomy / T. Baltrušaitis et al. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019. Vol. 41, no. 2. P. 423–443.
16. Artificial intelligence and machine learning in healthcare. *IBM*. URL: <https://www.ibm.com/artificial-intelligence/healthcare> (date of access: 05.03.2025).
17. Artificial intelligence in medicine overview. *Stanford Medicine*. URL: <https://med.stanford.edu/artificial-intelligence> (date of access: 09.03.2025).
18. Artificial intelligence in radiology. *Radiology Society of North America*. URL: <https://pubs.rsna.org/artificial-intelligence> (date of access: 12.03.2025).
19. Multimodal artificial intelligence in healthcare. *Nature Reviews Bioengineering*. URL: <https://www.nature.com/articles/multimodal-ai-healthcare> (date of access: 15.03.2025).
20. Жук А.В., наук.керівник Кудрявцева М.С. Методи використання систем штучного інтелекту для підтримки прийняття рішень в задачах медичної діагностики. 28-ий Міжнародний молодіжний форум «Радіoeлектроніка та молодь у ХХІ столітті». Харків: ХНУРЕ, 16–18 квітня 2024. С. 10–12.
21. Філатов В., Єрохін А., Золотухін О., Кудрявцева М. Методи інтелектуального аналізу процесів у медичних інформаційних системах. *Information Extraction and Processing*. 2020. № 48 (124). С. 92–98. URL: <https://doi.org/10.15407/vidbir2020.48.092> (дата звернення: 10.03.2025).