



Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ Комп'ютерних наук \_\_\_\_\_  
(повна назва)  
Кафедра \_\_\_\_\_ Штучного інтелекту \_\_\_\_\_  
(повна назва)  
Рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_  
Спеціальність \_\_\_\_\_ 122 Комп'ютерні науки \_\_\_\_\_  
(код і повна назва)  
Тип програми \_\_\_\_\_ освітньо-професійна \_\_\_\_\_  
(освітньо-професійна або освітньо-наукова)  
Освітня програма \_\_\_\_\_ Науки про дані (Data Science) \_\_\_\_\_  
(повна назва)

ЗАТВЕРДЖУЮ:  
Зав. кафедри \_\_\_\_\_  
(підпис)  
« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві \_\_\_\_\_ Гур'єву Михайлу Ігоровичу \_\_\_\_\_  
(прізвище, ім'я, по батькові)

1. Тема роботи \_\_\_\_\_ Розробка та дослідження методів машинного навчання в задачах аналізу біомедичних зображень для визначення патологій очей \_\_\_\_\_

затверджена наказом університету від 24 листопада 20 25 р. № 1057Ст

2. Термін подання студентом роботи до екзаменаційної комісії 16 грудня 20 25 р.

3. Вихідні дані до роботи \_\_\_\_\_ Науково-технічні публікації, дані інтернет-джерел, документація tensorflow, набори тренувальних даних з фотографіями очного дна \_\_\_\_\_

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1) Аналіз предметної галузі \_\_\_\_\_

2) Збір даних та попередня обробка \_\_\_\_\_

3) Методи машинного навчання \_\_\_\_\_

4) Оцінка та результати експериментів \_\_\_\_\_



## РЕФЕРАТ

Пояснювальна записка: 74 с., 28 рис., 1 дод., 26 джерел.

КЛАСИФІКАЦІЯ, МАШИННЕ НАВЧАННЯ, НЕЙРОННІ МЕРЕЖІ, ПАТОЛОГІЇ ОЧЕЙ, ПОПЕРЕДНЯ ОБРОБКА, РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ, ТРАНСФЕРНЕ НАВЧАННЯ, PYTHON.

Об'єкт дослідження – методи на основі машинного навчання для аналізу та класифікації біомедичних зображень за фотографіями очного дна.

Метою кваліфікаційної роботи є розробка, опис, впровадження та експериментальна оцінка сучасних підходів до машинного навчання і створення глибоких нейронних мереж, для підвищення точності, надійності та їх практичного застосування для автоматизованого виявлення патологій ока на основі зображень очного дна.

Методи дослідження – теоретичний аналіз, розгляд роботи з даними, обробка вхідних зображень, оптимізація тренування моделей, аналіз та дослідження технологій для побудови моделей машинного навчання, для використання у розробці.

Результатами дослідження є проведений аналіз поставленої проблеми, розгляд наявних моделей та їх будови, опис процесу збору та обробки тренувальних даних моделі, використання технік аугментації даних та балансування класів, опис можливих способів покращення якості моделей машинного навчання, тренування моделей з допомогою трансферного навчання та тестування якості отриманих моделей.

## **ABSTRACT**

Master's thesis contains: 74 p., 28 fig., 1 ann., 26 references.

CLASSIFICATION, IMAGE RECOGNITION, MACHINE LEARNING, NEURAL NETWORKS, OCULAR PATHOLOGIES, PREPROCESSING, PYTHON, TRANSFER LEARNING.

The object of research is machine learning-based methods for analyzing and classifying biomedical images based on photographs of the fundus.

The purpose of of this qualification work is to develop, describe, implement, and experimentally evaluate modern approaches to machine learning and deep convolutional neural networks to improve the accuracy, reliability, and practical applicability of automated detection of eye pathologies based on fundus images.

Research methods include theoretical analysis, data review, input image processing, model training optimization, analysis and research of technologies for building machine learning models for use in development.

The results of the research are an analysis of the problem, a review of existing models and their structure, a description of the process of collecting and processing model training data, the use of data augmentation and class balancing techniques, a description of possible ways to improve the quality of machine learning models, training models using transfer learning, and testing the quality of the resulting models.

## ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів .....	8
Вступ.....	9
1 Аналіз предметної галузі .....	11
1.1 Опис предметної галузі.....	11
1.2 Аналіз існуючих аналогічних систем .....	18
1.2.1 Система LumineticsCore компанії Digital Diagnostics .....	19
1.2.2 Вебсервіс компанії RetinaLyze .....	20
1.2.3 EyeArt від компанії Eyenuk.....	21
1.2.4 Airdoc від компанії Beijing Airdoc Technology Co.....	23
1.2.5 ARDA від Google/Verily .....	23
1.3 Постановка задачі.....	24
2 Збір даних та попередня обробка .....	26
2.1 Збір датасету .....	27
2.2 Попередня обробка даних .....	29
2.3 Основні стратегії попередньої обробки для стандартизації та нормалізації зображень.....	32
2.4 Збільшення кількості та різноманітності даних.....	36
2.5 Методи балансування класів.....	40
2.5.1 Оверсемплінг, робота з простором ознак.....	41
2.5.2 Статичні та динамічні зважені функції втрат .....	43
2.5.3 Focal Loss .....	45
3 Методи машинного навчання .....	48
3.1 Бейзлайн модель Random Forest .....	48
3.2 Моделі глибокого навчання. і базові архітектури .....	49
3.2.1 VGG16.....	49
3.2.2 ResNet50.....	51
3.2.3 EfficientNet (B0–B7).....	52
3.2.4 MobileNet .....	53

3.2.5 Vision Transformer .....	55
3.3 Трансферне навчання.....	56
3.4 Інтерпретованість в медицині.....	58
4 Оцінка та результати експериментів.....	60
4.1 Метрики в оцінці медичних систем .....	60
4.2 Тестування розроблених моделей .....	64
Висновки .....	67
Перелік джерел посилання .....	68
Додаток А Програмний код тренування моделей.....	72
Додаток Б Відомість кваліфікаційної роботи.....	74

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ВМД – вікова макулярна дегенерація;

CAD – Computer-Aided Diagnosis – системи, що допомагають лікарям в інтерпретації медичних зображень для підвищення точності діагнозу;

CE – Cross-Entropy – функція втрат, яка вимірює різницю між прогнозованим розподілом ймовірностей та фактичним значенням;

CNN – Convolutional Neural Network – нейромережі, для обробки зображень, використовуючи згорткові шари для виділення ознак;

CSL – Cost-sensitive learning – метод навчання, який враховує різну «вартість» помилок;

DL – Deep Learning – розділ машинного навчання, заснований на використанні багатошарових нейронних мереж для моделювання даних;

ED – Elastic Deformation – техніка аугментації даних, що викривлює зображення для тренування стійкості моделі до змін форми об'єктів;

FN – False Negative – помилковий результат тесту, що вказує на відсутність певної умови, хоча насправді вона присутня;

FP – False Positive – помилковий результат тесту, який вказує на наявність певного стану, коли насправді його немає;

GAN – Generative Adversarial Network – архітектура з двох нейромереж, що змагаються між собою для створення нових даних;

NRR – Neuroretinal Rim – нейроретинальний поясок, тканина головки зорового нерва;

ONH – Optic Nerve Head – диск зорового нерва; ділянка сітківки, де зоровий нерв виходить з ока або сліпа пляма;

ROI – Region of Interest – обрана ділянка зображення, що становить найбільший інтерес для аналізу або обробки;

XAI – eXplainable Artificial Intelligence – методи, що дозволяють зробити рішення та логіку штучного інтелекту зрозумілими для людини.

## ВСТУП

Швидкий розвиток технологій медичної діагностики та глобальне зростання кількості захворювань очей, через специфіку сучасного світу, створили нагальну потребу в автоматизованих, точних і масштабованих діагностичних системах з високим рівнем надійності.

Такі захворювання, як діабетична ретинопатія, глаукома, вікова макулярна дегенерація, катаракта та патологічна міопія, продовжують загрожувати зору мільйонів людей у всьому світі. Раннє виявлення хвороби відіграє вирішальну роль у запобіганні незворотньої втрати зору, проте доступ до високоякісного скринінгу залишається нерівномірним у різних регіонах та клінічних умовах. Ці виклики зробили аналіз зображень на основі машинного навчання надзвичайно актуальним напрямком досліджень, що має значний соціальний та медичний вплив.

В останні роки машинне навчання, зокрема глибоке навчання, продемонструвало значні успіхи у виявленні малопомітних візуальних біомаркерів на біомедичних зображеннях. Фотографування сітківки стало одним із найперспективніших методів для таких методів, воно є неінвазивним, широко використовується та є дуже інформативним. Однак забезпечення надійної, достовірної та узагальненої ефективності в реальних діагностичних робочих процесах все ще залишається складним науковим завданням. Нещодавній комплексний огляд галузі підкреслює, що, хоча підходи глибокого навчання, такі як CNN та віжен трансформери, показали великі перспективи у виявленні різних патологій сітківки, ключовими напрямками досліджень залишаються поліпшення інтерпретованості моделей, обробка класифікації захворювань з декількома хворобами та вирішення проблеми зміщення доменів в наборах даних [1].

Різноманітність пристроїв для фотографування, відмінності в популяціях пацієнтів, дисбаланс класів, висока ціна розмітки даних, та необхідність медично інтерпретованих результатів роблять цю галузь

складною, проте вирішення даної задачі принесе величезну користь суспільству. Вирішення цих питань безпосередньо сприяє поліпшенню якості скринінгу, зменшенню діагностичного навантаження на офтальмологів та сприянню раннього втручання у випадках що могли привести до сліпоты.

З огляду на перелічені фактори, розробка та дослідження методів покращення якості машинного навчання для виявлення патологій очей у біомедичному аналізі зображень є не тільки науково актуальними, але й соціально значущими. Ця тема відповідає глобальним пріоритетам охорони здоров'я, спрямованим на розширення доступу до ранньої діагностики, поліпшення клінічного прийняття рішень та підвищення ефективності стратегій охорони здоров'я, орієнтованих на профілактику порушень зору.

Згідно зі Всесвітнім звітом ВООЗ про зір, щонайменше 1 мільярд випадків порушення зору залишаються невирішеними в усьому світі, багато з яких можна було запобігти за допомогою своєчасної та комплексної діагностики допомоги [2]. Це дослідження знаходиться на перетині комп'ютерного зору, медицини та інновацій в охороні здоров'я на основі даних – галузей, які разом визначають майбутнє інтелектуальних клінічних технологій. Проведена робота пройде по всім важливим етапам розробки і покращення моделей машинного навчання від збору даних і їх обробки до тренування моделей та оцінки їх ефективності, з практичними прикладами.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Опис предметної галузі

Людське око – дивовижний орган, який дозволяє нам сприймати світ навколо. Це єдина частина людського тіла, де можна спостерігати за мікросудинами і зоровим нервом, частиною центральної нервової системи, абсолютно неінвазивним способом, не роблячи жодних розрізів. Очне дно являє собою внутрішню задню поверхню ока, що забезпечує доступ до центральної нервової та судинної систем. Фотографія очного дна фіксує двовимірну проекцію заднього сегмента, візуалізуючи життєво важливі структури, включаючи сітківку, диск зорового нерву (ОНН), макулу, судинну оболонку, склоподібне тіло та судини сітківки. Для автоматизованого аналізу зображень конкретні анатомічні орієнтири повинні бути визначені як ROI, оскільки їх вигляд або локалізовані патологічні зміни визначають клінічний діагноз.

Центральне місце в зорі займає макула – овальна жовтувата ділянка сітківки. На знімках очного дна її можна відрізнити за темнішим кольором. У макулі розташована фовеа – ділянка з найвищою щільністю колбочкових фоторецепторів. Завдяки своїй ролі в центральній гостроті зору макула і фовеа є основними ділянками інтересу для таких патологій, як вікова дегенерація макули (AMD) і патологічна міопія.

Надійна архітектура машинного навчання, розроблена для багатокласової класифікації, повинна включати механізми для визначення пріоритетності цих окремих анатомічних регіонів. Оскільки різні патології є високо локалізованими – глаукома впливає на ОНН, AMD і патологічна міопія впливають на макулу, а DR проявляється вздовж судинної системи – покладання виключно на глобальні характеристики зображення знижує ефективність. Шум з непатологічних областей може заважати побачити тонкі патологічні особливості, необхідні для раннього виявлення. Отже,

обчислювальні методики повинні надавати пріоритет точному виявленню меж (наприклад, межі диска/чашки та центру макули) для локалізації аналізу перед початком вилучення особливостей, з чим допоможе впоратися попередня обробка.

Далі буде систематично описано норму та п'ять патологічних класів та відповідні особливості очного дна, на які в теорії має звертати увагу ML-модель для встановлення клінічної істини [3].

Класифікація «Норма» служить обов'язковим бейслайном про який має знати модель. Вона вимагає відсутності будь-яких патологічних уражень і суворого дотримання встановлених анатомічних співвідношень. Зокрема, нормальне очне дно має рівномірну та цілісну судинну структуру, чітке окреслення нейроретинального краю (NRR) та фізіологічну відповідність правилу ISNT, згідно з яким товщина краю повинна відповідати такому порядку: нижня(I) > верхня(S) > носова(N) > скронева(T), як показано на рисунку 1.1. Встановлення надійного варіанта нормальної бази є надзвичайно важливим для систем ML, оскільки вона визначає оптимальний діапазон характеристик, перш ніж вони будуть класифіковані як аномальні [4].

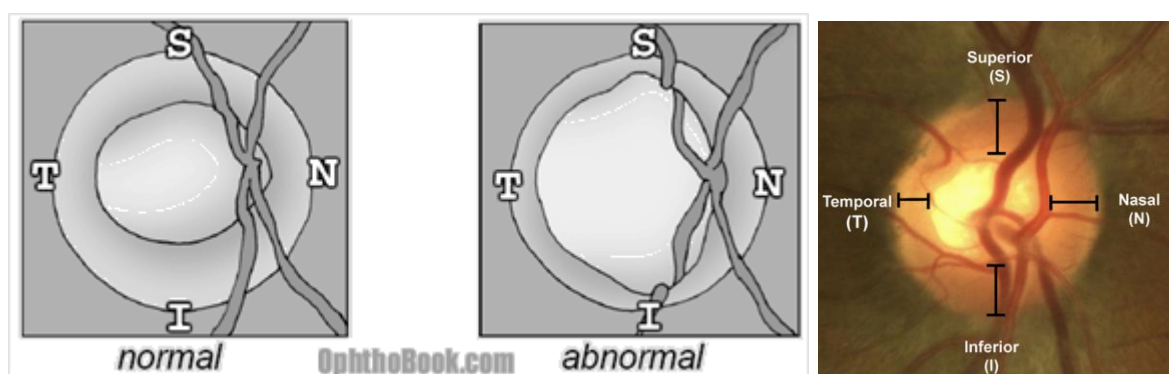


Рисунок 1.1 – Візуалізація правила ISNT

Діабетична ретинопатія виникає внаслідок пошкодження мікросудин сітківки, що призводить до витоку, порушення перфузії та

неоваскуляризації. Характерними ураженнями очного дна є мікроаневризми (часто найперші помітні зміни), крововиливи (точкоподібні та плямоподібні), тверді екsudати, ватні плями та венозні аномалії, приклади яких можна побачити на рисунку 1.2 [5]. Клінічна об'єктивна оцінка тяжкості ДР базується на модифікованій класифікації Airlie House, яка присвоює індекс тяжкості діабетичної ретинопатії (DRSS) на основі ідентифікації та просторового розподілу цих уражень у СFP. Тому автоматизовані системи повинні виконувати кількісний аналіз уражень.

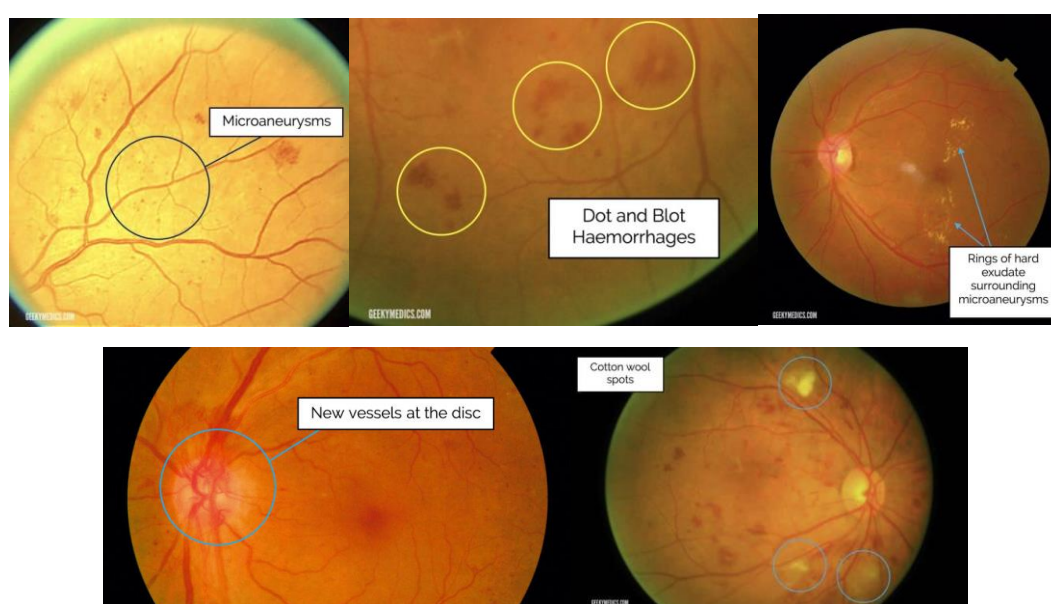


Рисунок 1.2 – Зображення пошкоджень на фото властивих для діабетичної ретинопатії

Складніше завдання полягає не лише в класифікації за наявністю/відсутністю, а й у точній кількісній оцінці уражень. Ця оцінка вимагає точної сегментації на рівні пікселів як дрібних структур (мікроаневризм), так і більших уражень (ексудати, крововиливи) по всьому полю сітківки, щоб правильно присвоїти відповідний ступінь тяжкості діабетичної ретинопатії.

Класифікація глаукоми полягає в оцінці структурного пошкодження головки зорового нерва, що характеризується стоншенням NRR і розширенням зорової чашки, це називається купінг. Основною кількісною характеристикою для класифікації ML є співвідношення чашки до диска. Хоча співвідношення приблизно 0,6 або більше часто є підозрілим, цей показник необхідно інтерпретувати в контексті. Наприклад, невеликий зоровий нерв може мати патологічне стоншення навіть при співвідношенні C/D 0,4, тоді як великий нерв може бути фізіологічно нормальним при 0,7, рисунок 1.3. Крім того, необхідний аналіз геометрії зорового нерва, оскільки при прогресуванні глаукоми часто спостерігається більш значне збільшення вертикального співвідношення C/D, ніж горизонтального. Оцінка NRR також вимагає перевірки на порушення правила ISNT.

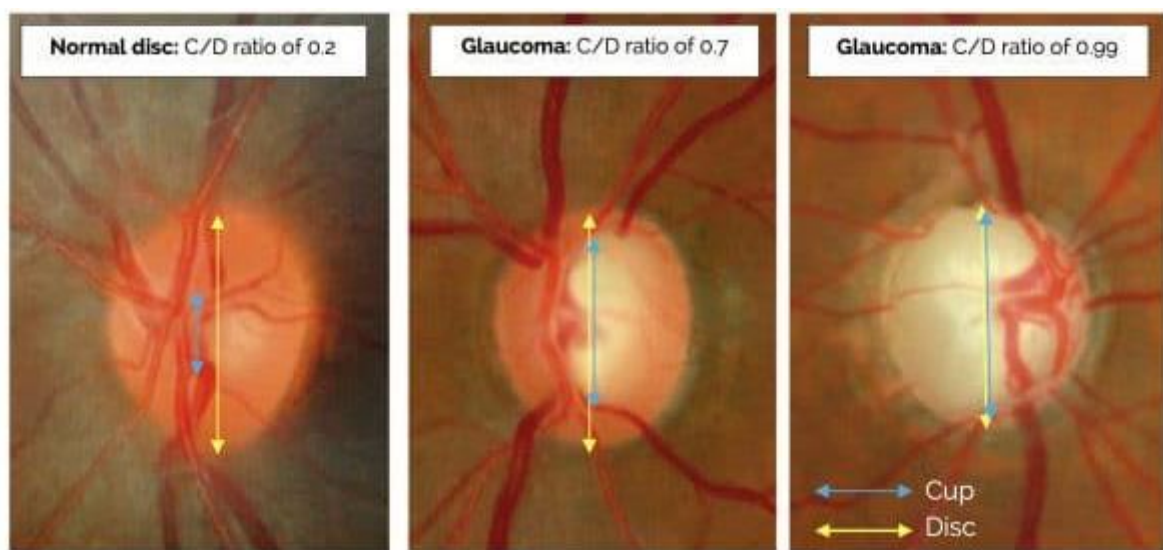


Рисунок 1.3 – Зображення різних співвідношень чашки до диска

Вікова макулярна дегенерація (ВМД) в першу чергу вражає макулу, центральну ділянку сітківки, відповідальну за гостроту зору. На фотографіях очного дна одним з найперших і найхарактерніших ознак ВМД є наявність друз, жовтуватих відкладень, що накопичуються під пігментним епітелієм сітківки і в мембрані Бруха. Ці друзи бувають різних типів:

«тверді» друзи – невеликі з чіткими краями, «м'які» друзи – більші та більш розмиті.

У міру прогресування ВМД можуть розвиватися більш серйозні патологічні зміни. У атрофічній («сухий») формі однією з ключових ознак є географічна атрофія, при якій атрофуються ділянки які часто виглядають як чітко окреслені бліді області; судини нижнього шару можуть стати більш видимими, оскільки пігментована тканина, що їх покриває, зникла.

При неоваскулярній («вологій») формі ВМД ростуть нові аномальні кровоносні судини. Ці судини можуть просочуватися кров'ю або рідиною, викликаючи крововиливи, субретинальну рідину та відшарування пігментного епітелію (ВПЕ), які на фотографіях очного дна виглядають як нерівні, підняті, часто сіро-зелені або червоні ділянки, а також ексудати та рубці (рисунок 1.4).

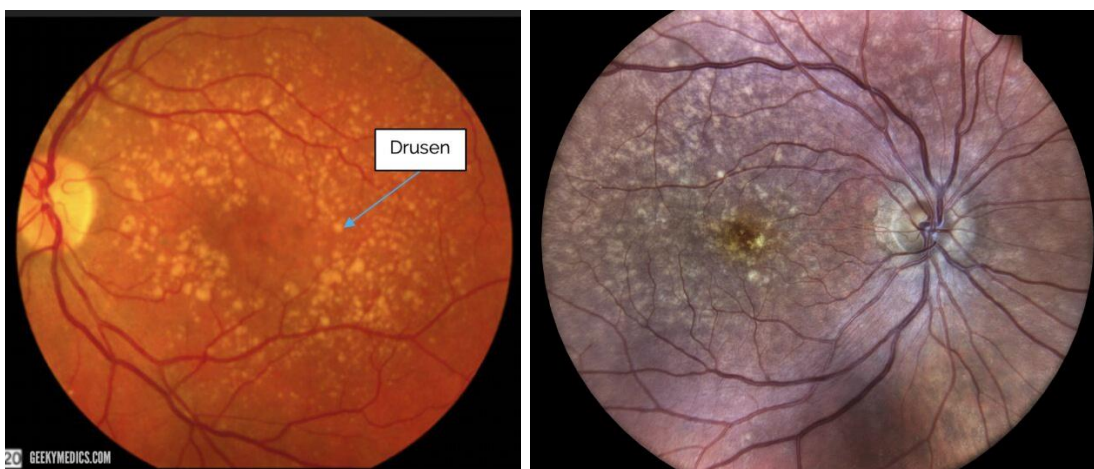


Рисунок 1.4 – Зображення друза на фотографії очного дна

Катаракта – це помутніння природної лінзи ока, а не захворювання сітківки, тому її ознаки на фотографії очного дна проявляються опосередковано. Оскільки лінза стає непрозорою, зображення очного дна втрачає чіткість, демонструючи знижений контраст, підвищену розмитість або зовсім розмитий вигляд. Судини, макула та диск зорового нерва

виглядають менш чіткими, а через розсіювання світла всередині помутнілого кришталіка можуть з'являтися яскраві відблиски або туманні артефакти. При катаракті середньої та важкої ступені камера для фотографування очного дна може взагалі не в змозі зробити якісне зображення, що призводить до отримання тьмяних фотографій низької якості, які обмежують діагностичну деталізацію, як показано на рисунку 1.5.

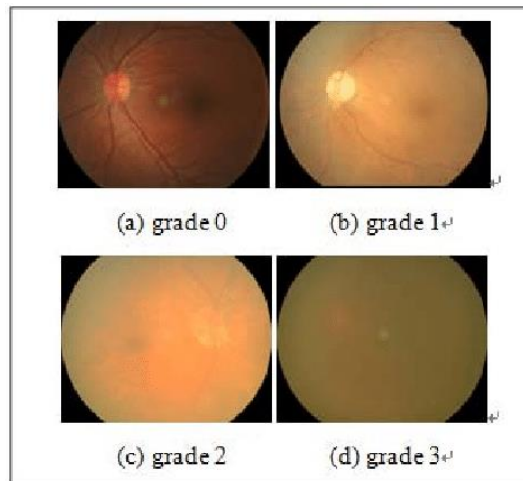


Рисунок 1.5 – Приклади фотографій різного рівня катаракти

Це створює унікальний «парадокс» для навчання ШІ. У багатьох наборах даних про DR з однією хворобою розмиті зображення позначаються як «непіддаються оцінці» або «низькоякісні» і виключаються. У шестикласній системі з декількома захворюваннями модель повинна навчитися класифікувати «розмитість» як позитивний діагностичний результат (катаракта), а не як порушення контролю якості.

Міопія (короткозорість) змінює форму і структуру ока, що призводить до появи характерних особливостей очного дна. При легкій та помірній міопії сітківка виглядає більш видовженою, а диск зорового нерва може здаватися нахиленим, овальним, а не круглим. У перипапільярній ділянці часто спостерігається перипапільярна атрофія – бліде, нерівне кільце навколо диска зорового нерва, спричинене розтягуванням тканин сітківки. Фон

сітківки може здаватися тоншим, демонструючи світліше або більш «металеве» відбиття, відоме як мозаїчний очне дно, через видимість судин хоріоїдеї, що знаходяться під ним.

При сильній міопії зміни стають більш вираженими. Подовжене око може спричинити стоншення сітківки, лакові тріщини (дрібні розриви в глибших шарах сітківки) або навіть ранню дегенерацію поблизу макули. Макула може злегка зміщуватися, а загальне забарвлення очного дна часто змінюється в бік більш жовтуватого або блідого вигляду. Ці ознаки відображають структурне навантаження на сітківку від осьового подовження і є критичними маркерами для відрізнення міопії від інших захворювань сітківки, що впливають на форму, пігментацію або судинні візерунки (рисунок 1.6).

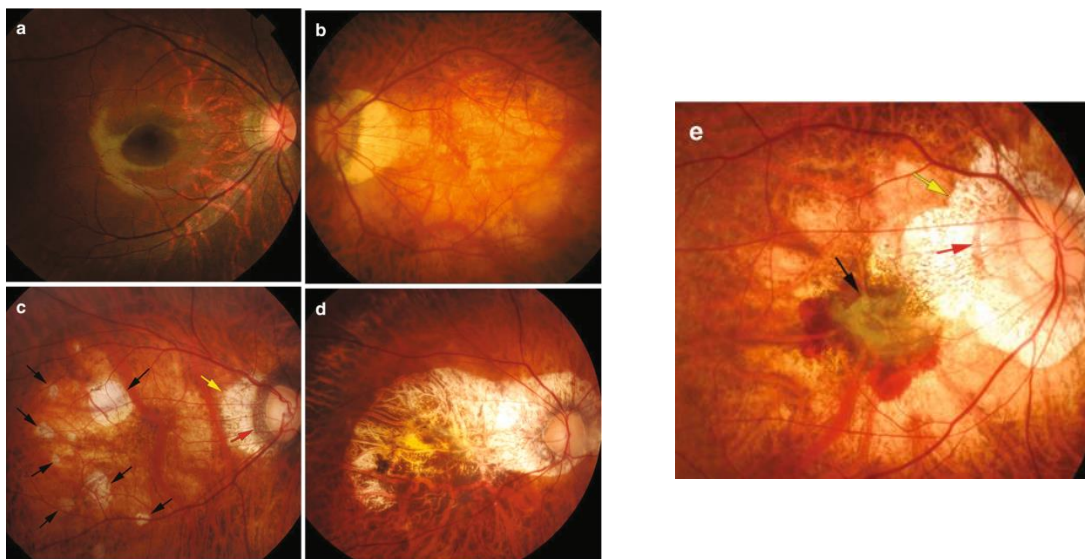


Рисунок 1.6 – Вигляд різних випадків міопії на фотографіях очного дна

Моделі машинного навчання зазвичай добре працюють з ПМ, оскільки фонові текстури «тигроїд» є глобальною, повторюваною особливістю, яку CNN легко засвоюють. Однак структурне спотворення диска зорового нерва при ПМ може серйозно заплутати алгоритми виявлення глаукоми, які покладаються на стандартну геометрію диска.

Останні досягнення в галузі машинного навчання та комп'ютерного зору дозволяють програмному забезпеченню на основі штучного інтелекту навчатися на великих об'ємах фотографій очного дна, що дає змогу виявляти захворювання з високою точністю, яка часто відповідає і навіть перевищує людські показники.

Клінічні дослідження демонструють безпечність та ефективність ШІ у виявленні наявності чи відсутності станів, що загрожують зору, таких як діабетична ретинопатія, вікова макулодистрофія глаукома, катаракта, патологічна короткозорість та при оцінці ризику системних захворювань, таких як гіпертонія, інсульт та серцево-судинні захворювання за допомогою даних отриманих з фотографій сітківки.

У майбутньому скринінг очей за участю методів машинного навчання для аналізу зображень зробить діагностування більш доступним, уможливить широкомасштабний скринінг населення і знизить вартість перевірки стану здоров'я, досягнувши цільової точності порівняної з лікарями близької до 100%.

## 1.2 Аналіз існуючих аналогічних систем

Інтеграція штучного інтелекту (ШІ) в офтальмологічне обстеження є одним з найважливіших досягнень сучасної медичної діагностики. Хоча в перше десятиліття досліджень переважали бінарні класифікаційні системи, орієнтовані виключно на діабетичну ретинопатію (ДР), зараз у цій галузі відбувається парадигмальний зсув у бік комплексних систем обстеження на наявність різних захворювань. Цей перехід обумовлений клінічною реальністю, згідно з якою пацієнти, особливо люди похилого віку, рідко страждають на ізольовані патології.

Аналіз показує деяку неузгодженість в цій галузі: комерційне впровадження в США залишається в основному обмеженим показаннями для однієї хвороби через консервативність регуляторних органів, академічні

дослідження та міжнародні комерційні організації успішно продемонстрували життєздатність багатокласових систем глибокого навчання.

На сьогоднішній день на світовому ринку компаній, які займаються діагностикою хвороб очей за допомогою штучного інтелекту, вже виділилися лідери галузі, які за допомогою інвестицій різної величини створюють сервіси, програмні застосунки та прилади, здатні допомогти виявленню хвороб, які впливають на зір людини, деякі з них займаються мультикласовою діагностикою, деякі сфокусовані на одній чи парі хвороб. Розглянемо показники ефективності та точності низки таких рішень.

### 1.2.1 Система LumineticsCore компанії Digital Diagnostics

Система LumineticsCore, розроблена компанією Digital Diagnostics, є першою в історії автономною системою штучного інтелекту, схваленою FDA для виявлення діабетичної ретинопатії. Система призначена для використання в офтальмологічних умовах, таких як клініки первинної медичної допомоги, з використанням деяких сумісних камер для огляду очного дна.

LumineticsCore призначений для використання медичними працівниками для автоматичного виявлення діабетичної ретинопатії більш ніж легкого ступеня у дорослих (22 роки та старше) з діагнозом діабет, у яких раніше не було діагностовано діабетичну ретинопатію.

У своєму ключовому клінічному дослідженні, що підтверджує дозвіл FDA, алгоритм LumineticsCore продемонстрував чутливість 87,2% і специфічність 90,7% у порівнянні з суворим стандартом оцінки людиною [6]. Висока специфічність є ключовою особливістю, що демонструє ефективність системи в сортуванні пацієнтів і дозволяє уникнути 91% непотрібних візитів до фахівців для пацієнтів, які отримали негативний результат тесту на mtmDR (рисунок 1.7).

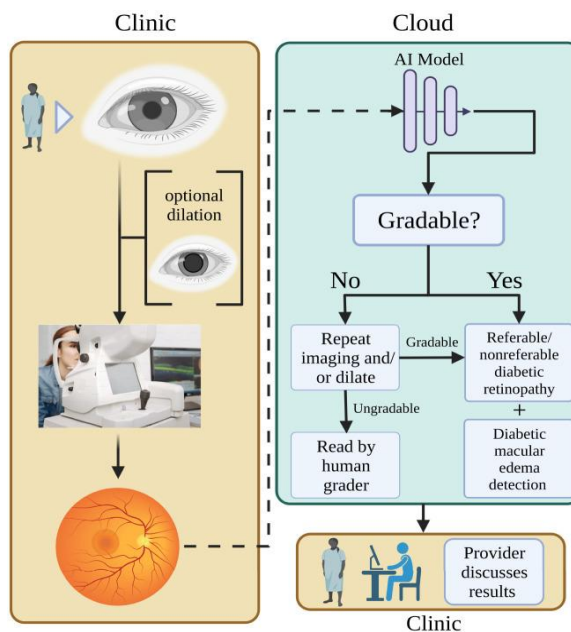


Рисунок 1.7 – Приклад роботи клініки з використанням ШІ моделі для скринінгу

Недоліками системи є:

- недоступність для вільного використання за межами медичних установ;
- дороговизна розробки та встановлення системи потребує великої кількості пацієнтів на діагностику для досягнення рівня грошової ефективності в порівнянні зі стандартними методами;
- LumineticsCore не призначений для виявлення супутніх захворювань, офтальмологічних або інших системних захворювань, крім діабетичної ретинопатії;
- погана сумісність з деякими видами камер відмінних від Topcon TRC-NW400.

### 1.2.2 Вебсервіс компанії RetinaLyze

Retinalyze – це система, розроблена датською компанією RetinaLyze System. Вона функціонує як система підтримки клінічних рішень для

фахівців з офтальмології, аналізуючи зображення очного дна та скани ОСТ. Вона має CE Mark від Європейського Союзу. Retinalyze зосереджується на трьох найпоширеніших захворюваннях, що спричиняють сліпоту: діабетичній ретинопатії, глаукомі та ВМД.

Ефективність залежить від хвороби та стратегії застосування. У дослідженні, проведеному в рамках Датської національної програми скринінгу (N=1001 око), Retinalyze досяг чутливості 84,9% та специфічності 89,9% для виявлення ДР. Для виявлення глаукоми в дослідженні, в якому брали участь понад 1800 випадків, було зафіксовано чутливість 86,5% і специфічність 95,5%. Однак у порівняльному дослідженні з LumineticsCore було зафіксовано нижчу специфічність 71,82% для однієї стратегії застосування [7].

До недоліків Retinalyze можна віднести:

- система доступна лише клінікам, оптиками та іншим бізнесам;
- брак ліцензії FDA;
- низька ефективність в тестах (71.82%) показує, що модель може бути чутлива до умов використання.

### 1.2.3 EyeArt від компанії Eyeuk

Eyeuk – міжнародна компанія, що займається медичними технологіями та послугами на основі штучного інтелекту, яка спеціалізується на ШІ-скринінгу очей для автономного виявлення захворювань та ШІ-прогнозуванні біомаркерів для оцінки ризиків та спостереження за захворюваннями. Компанія пропонує продукти, які автономно аналізують зображення сітківки ока пацієнта, виявляють ознаки захворювання і надають звіт для лікарів.

До позитивних сторін даного сервісу можна віднести те що пройдена сертифікація FDA, CE, Health Canada License. Щодо точності то у дослідження, що охоплює 17 реальних досліджень та більше ста тисяч

обстежень, EyeArt продемонстрував високу діагностичну точність у виявленні DR, що потребує направлення до фахівця, досягнувши сукупної чутливості 95% та сукупної специфічності 81%. Крім того, в ході ключового клінічного випробування система досягла чутливості 97% і специфічності 90% для DR, що загрожує зору (рисунок 1.8). Конструкція системи, оптимізована для досягнення максимальної чутливості, що забезпечує виявлення пацієнтів з безсимптомно прогресуючою DR, навіть якщо це призводить до дещо вищого рівня хибнопозитивних результатів порівняно з іншими платформами з високою специфічністю.

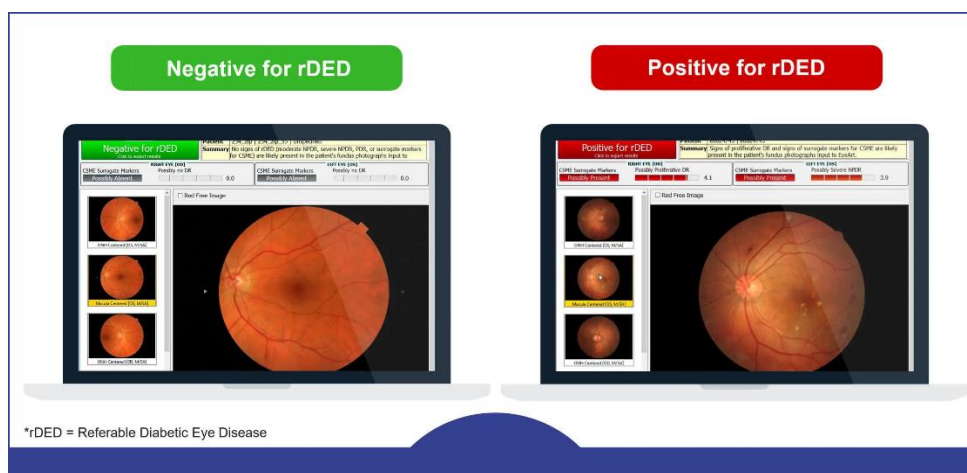


Рисунок 1.8 – Можливі результати діагностики EyeArt

До недоліків належать наступні пункти:

- EyeArt показує знижену ефективність при наявності катаракти;
- рішення не призначене для виявлення супутніх захворювань, інших офтальмологічних захворювань або інших системних захворювань;
- EyeArt призначено лише для використання із зображеннями, отриманими за допомогою Canon CR-2 AF та Topcon NW400;
- система доступна лише клінікам.

#### 1.2.4 Airdoc від компанії Beijing Airdoc Technology Co

Airdoc – це велика китайська компанія, що спеціалізується на технологіях у сфері охорони здоров'я та використовує свою систему Airdoc Retinal AI System для широкої діагностики хронічних захворювань. Система отримала сертифікат CE та сертифікат NMPA класу III для медичних виробів у Китаї. Airdoc має найширший спектр застосування серед розглянутих систем. Окрім ДР, ARAS застосовується для діагностики підозри на глаукому та захворювань макули. Компанія стверджує, що її оцінка стану здоров'я на основі штучного інтелекту дозволяє оцінити 55 ризиків, пов'язаних із хронічними захворюваннями, за допомогою одного зображення сітківки ока, що вказує на основну стратегічну спрямованість на висококорисний скринінг системних ризиків для здоров'я, а не виключно на офтальмологічні застосування [8].

У багатоцентровому дослідженні, що підтверджує ефективність системи скринінгу на основі штучного інтелекту, Airdoc продемонстрував високу ефективність у лікуванні багатьох захворювань. Для виявлення ДР загальна чутливість становила 94.8%, а специфічність 95.4%. Ефективність була особливо високою для виявлення підозри на глаукому, досягнувши специфічності 0,993 і чутливості 0,891. Для захворювань макули (AMD), чутливість становила 0,901, а специфічність – 0,955.

#### 1.2.5 ARDA від Google/Verily

ARDA – це система глибокого навчання, розроблена в результаті співпраці між Google, Verily Life Sciences та такими організаціями, як Aravind Eye Hospitals. Вона призначена в першу чергу для проведення масштабних медичних оглядів, зокрема серед хворих на діабет в Індії та Таїланді. У 2019 році система отримала сертифікат CE, що демонструє визнання її ефективності.

ARDA продемонструвала надзвичайно високу ефективність у реальних умовах клінічного моніторингу після впровадження. У перехресному дослідженні 4537 випадково відібраних пацієнтів ефективність системи у виявленні важкої ДР досягала чутливості 97,0% та специфічністю 96,4% [9]. Важливо, що система має майже нульовий рівень пропущених важких випадків. Ця валідація вважається дуже значущою, оскільки дана робота є одним із перших масштабних звітів про ефективність офтальмологічного пристрою на основі штучного інтелекту що використовувалась для настільки масштабного дослідження.

Незважаючи на високу ефективність, ARDA не має схвалення FDA США, що обмежує її комерційне поширення. Крім того, клінічний моніторинг, необхідний для підтримки такого високого рівня ефективності після впровадження, вважається дорогим і необхідним процесом. Залежність від підтримання стабільності роботи в мінливих умовах використання камери, калібрування обладнання та різноманітності пацієнтів вимагає постійного контролю.

### 1.3 Постановка задачі

Метою роботи є розробка, опис та оцінка сучасних підходів до машинного навчання, для підвищення точності, надійності та практичної застосовності автоматизованого виявлення патологій ока на основі зображень очного дна, для досягнення точності порівняної з реальним лікарем [10].

Розроблені в процесі виконання роботи моделі мають демонструвати частину з описаних методик покращення та тренування моделей, тестування моделей проводитиметься стандартизованою функцією, яка буде показана та описана в одному з розділів.

Успішне втілення поставленої задачі передбачає виконання наступних завдань:

- аналіз предметної галузі;
- проведення аналізу існуючих аналогів;
- збір навчальних даних;
- варіанти попередньої обробки фотографій очного дна;
- опис процесу аугментації датасету;
- методи балансування класів;
- огляд моделей машинного навчання;
- трансферне навчання;
- регуляризація та запуск тренування моделей;
- опис специфіки оцінки медичних моделей;
- оцінка якості медичних моделей.

Таким чином задачею поставленою при виконанні роботи є вироблення рекомендацій для побудови та покращення моделей машинного навчання для біомедичних цілей, що працюють з візуальними даними, на прикладі моделі класифікації патологій очей, для досягнення точності порівняною з живими лікарями.

## 2 ЗБІР ДАНИХ ТА ПОПЕРЕДНЯ ОБРОБКА

На ранніх етапах побудови класифікаційної моделі для медичної візуалізації процес збору чистого, репрезентативного та достатньо великого набору даних відіграє вирішальну роль у визначенні продуктивності кінцевої моделі. Набір даних служить не тільки «паливом» для навчання, але й вираженням того, що модель може і не може зрозуміти. Здатність моделі розрізняти тонкі ознаки захворювань, особливо в таких сферах, як зображення очного дна, повністю залежить від різноманітності та правильності даних, з якими вона стикається під час навчання.

Занадто малий набір даних призводить до перенавчання, коли модель запам'ятовує приклади, а не вивчає узагальнені патерни. Це призводить до високої точності на навчальних зразках, але низької продуктивності на нових зображеннях. І навпаки, великий і різноманітний набір даних зменшує цей ризик, даючи моделі можливість вивчати специфічні для захворювання особливості в різних умовах таких як освітлення, різні налаштування камер, роздільної здатності зображень, положень очей і даних пацієнтів. Для медичних застосувань це є надзвичайно важливим для забезпечення надійності та об'єктивної роботи в реальних умовах.

Іншим критичним аспектом є баланс класів. Медичні набори даних часто страждають від нерівномірного розподілу між категоріями – наприклад, нормальні очі можуть бути набагато поширенішими, ніж рідкісні захворювання. Без збалансування модель стане упередженою до класів більшості і не зможе виявити менш поширені, але важливі стани, які не можна впускати.

Якість зібраних зображень також безпосередньо впливає на процес навчання. Низька роздільна здатність, неправильні мітки або непослідовна попередня обробка можуть спричинити шум, який погіршує модель. Добре підібраний набір даних з точними анотаціями гарантує, що модель вивчає значущі особливості, а не артефакти. У медичній візуалізації навіть невеликі

помилки в маркуванні можуть призвести до значних діагностичних помилок, що робить ретельну підготовку набору даних важливим етапом у всьому дослідницькому процесі.

Розмір, різноманітність, збалансованість і якість даних є основою всього процесу збору даних, які будуть основою всіх подальших процесів побудови моделей. Будь-які вдосконалення архітектури або стратегії навчання будуть обмежені якістю самих даних; тому збір набору даних є одним з найважливіших етапів у розробці надійної моделі класифікації.

## 2.1 Збір датасету

Основною задачею, що я ставив при зборі датасету мати від 800 до 1000 картинок різних ступенів, видів камер та освітлень на кожен з шести класів: норма, катаракта, міопія, глаукома, діабетична ретинопатія та вікова макулярна дегенерація. За основу датасету було обрано набір даних з Kaggle Ocular Disease Recognition. Дані були зібрані в реальних лікарнях і клініках в Китаї, з реальними пацієнтами яких було близько 5000, за допомогою реальних камер для огляду очного дна. Проте цей датасет є вкрай незбалансованим та іноді фотографія має декілька хвороб у стовбці діагнозу. Після підрахунку в переводі на фотографії було отримано такий розподіл даних, який потім сильно скоротився та розбавився даними з інших датасетів, рисунок 2.1.

```
Sum of normal photos: 5353
Sum of diabetes: 3217
Sum of glaucomas: 599
Sum of cataract: 586
Sum of AMD: 532
Sum of hypertension: 378
Sum of myopia: 453
```

Рисунок 2.1 – Фінальна композиція датасету

Ще одним використаним датасетом з якого взято велику кількість фотографій є датасет `eye_diseases_classification` в якому є по 1000 якісних фотографій на 4 класи, він в свою чергу теж є збіркою інших даних з старіших датасетів. З цього набору взято велику частку доданих даних з фотографіями глаукоми та катаракти, всі фотографії були переглянуті та викинуті завеликі або неякісні, також додано даних норми і діабетичної ретинопатії для балансування та зберігання різноманітності.

Для отримання більшої кількості зображень з класу міопії було знайдено набір даних під назвою PALM: Open Fundus Photograph Dataset with Pathologic Myopia Recognition and Anatomical Structure Annotation. Цей набір даних містить 1200 кольорових зображень очного дна, для аналізу патологічної міопії. Окрім кольорових зображень очного дна та міток для категорії патологічної міопії, набір також має дані про процедуру отримання зображень та обладнання.

Доповнення різних класів проводилося фотографіями проводилося з Kaggle датасетів AMDNet23: Fundus Image Dataset for Age-Related Macular Degeneration Disease, DR\_200, RFMiD, HRF, ARMD Curated Dataset 2023, 1kIM. Фінальне розподілення класів можна побачити на рисунку 2.2.

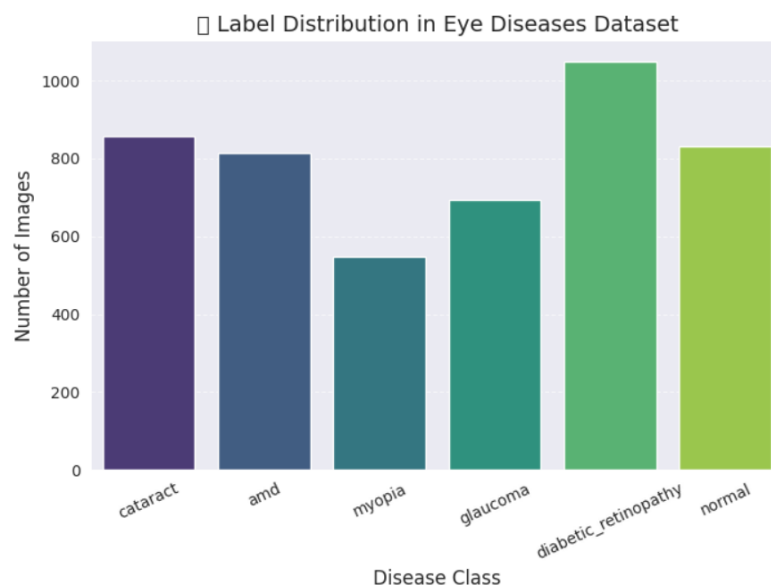


Рисунок 2.2 – Фінальна композиція датасету

Фінальний набір даних має такий набір даних: cataract: 856 фото, amd: 814 фото, myopia: 547 фото, glaucoma: 694 фото, diabetic\_retinopathy: 1050 фото, normal: 830 фото, кожне з яких приведено до розміру 256x256 пікселів. Тож фінальний розмір датасету 6 класів та 4791 фото. Навмисно створено датасет з нерівною кількістю зображень для створення умов для використання функцій регуляризації та баласування класів при створенні та тренуванні моделей, рисунок 2.3 показує приклади фотографій класів.

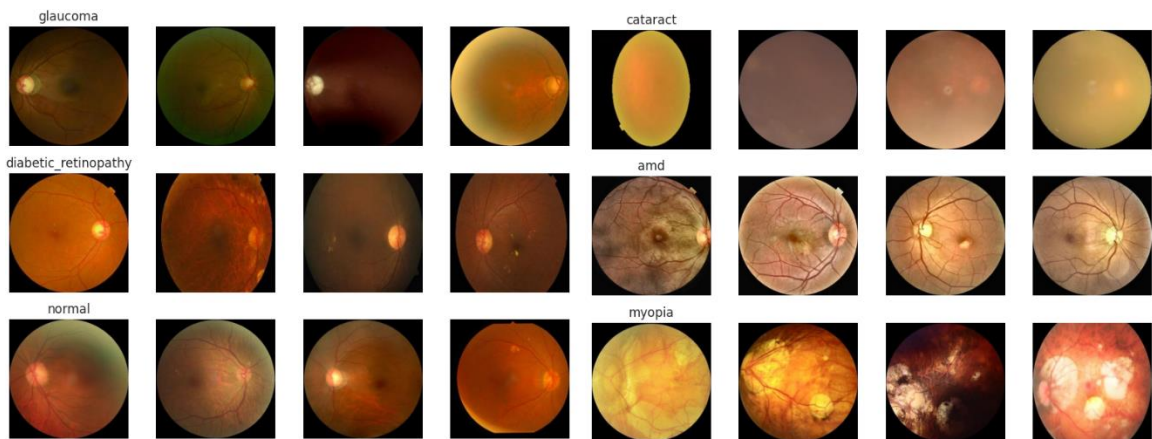


Рисунок 2.3 – Візуалізація прикладів фото кожного класу

## 2.2 Попередня обробка даних

Успішне застосування алгоритмів глибокого навчання (DL), зокрема конволюційних нейронних мереж, для завдань діагностики в офтальмології сильно залежить від якості та кількості навчальних даних. У цьому розділі встановлюється необхідність ретельної обробки медичних зображень, що включає попередню обробку та стратегічне розмноження даних, шляхом вивчення внутрішніх обмежень у медичній сфері, накладених на збір клінічних даних та доступність якісних датасетів з професіоналами розмічених даних.

Алгоритми DL демонструють високу ефективність у різних завданнях комп'ютерного зору; однак для досягнення високої продуктивності та

дженералізації вони потребують значного обсягу даних. У медичній галузі, а саме у фотографуванні очного дна, набори даних часто обмежуються малою кількістю маркованих даних різних етапів та специфічних видів хвороби що часто призводить до оверфітінгу мережі на навчальному наборі та поганій продуктивності на незнайомих рідкісних прикладах. Дефіцит даних в медичних доменах посилюється вартістю та часом, необхідними для анотації даних експертами, особливо для складних або рідкісних випадків патологій.

Крім того, клінічні набори даних очного дна часто мають проблему з дисбалансом класів, коли прикладів однієї хвороби набагато більше. Цей дисбаланс перешкоджає здатності моделі навчатися надійним ознакам для недостатньо представлених категорій. Збільшення обсягу даних є ключовою стратегією для штучного збільшення розміру та різноманітності набору даних, що зменшує перенавчання та знижує залежність від дорогих анотованих наборів даних.

Зображення очного дна за своєю природою схильні до погіршення якості через недосконалості систем зйомки фотографій та фізіологічні варіації ока кожної людини. Основними перешкодами в аналізі зображень сітківки є значні коливання контрасту в зображеннях та нерівномірність освітлення. Нерівномірність освітлення зумовлена кількома факторами: обмеженнями приладу для зйомки, внутрішньоочним розсіюванням, коливаннями відбиття світла та положенням ока відносно оптичної осі камери.

Яскравим проявом цих проблем із зйомкою є віньєтування, при якому яскравість і контрастність значно зменшуються до країв зображення. Цей негативний ефект призводить до того, що диск зорового нерва, який зазвичай є найяскравішою анатомічною структурою, іноді виглядає темнішим за інші ділянки сітківки на периферії, як це можна побачити на рисунку 2.4.

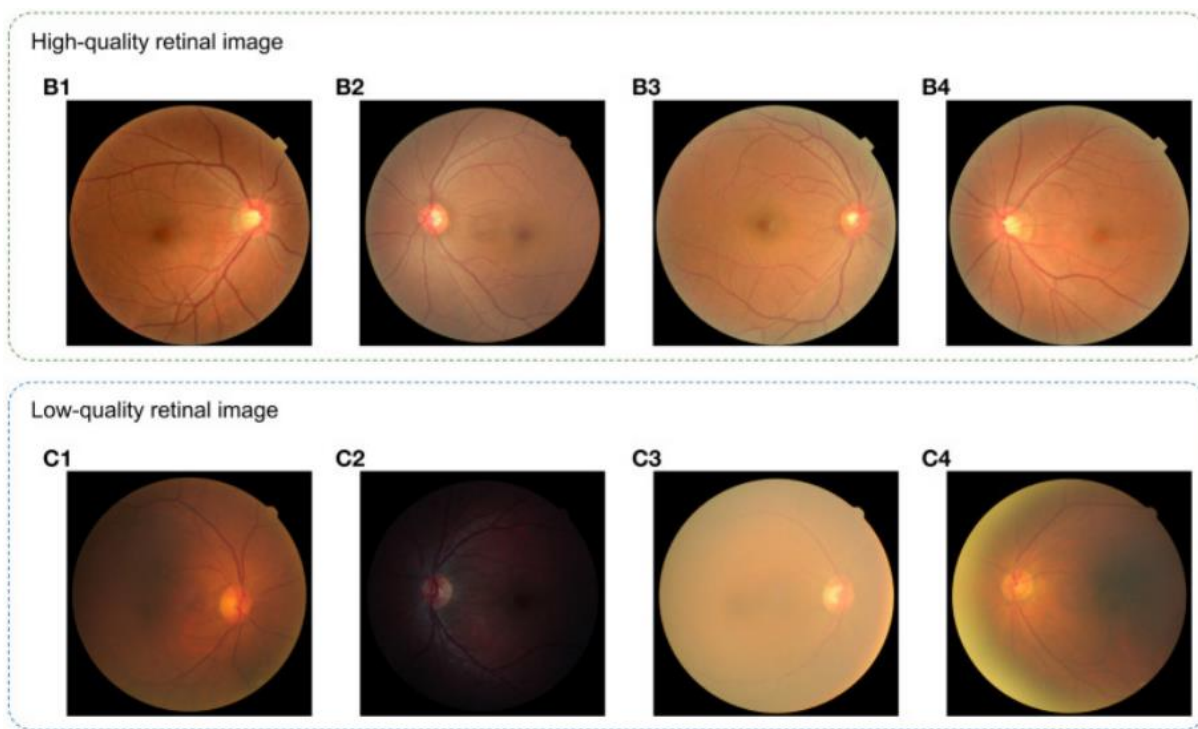


Рисунок 2.4 – Порівняння якісних і неякісних зображень очного диску

Інші проблеми якості включають низький контраст, високий рівень шуму та розмитість, спричинену раптовими рухами пацієнта. Наявність цих просторово неоднорідних характеристик освітлення та контрасту вимагає надійних методів попередньої обробки для стандартизації зображень, що є необхідним перед вилученням ознак за допомогою алгоритмів машинного навчання. При зборі та переборі даних були спроби відсіювання неякісних даних зображень.

Загальна стратегія обробки даних для класифікації зображень очного дна базується на фундаментальній послідовності: стандартизація, а потім введення різноманітності. Попередній перегляд відіграє вирішальну роль у стандартизації, виконуючи такі кроки, як шумозаглушення, нормалізація та підвищення контрасту. Ця стандартизація стабілізує вхідний сигнал, гарантуючи, що модель глибокого навчання фокусується на справжніх патологічних або анатомічних ознаках, а не на шумі або артефактах збору даних.

Після стандартизації вхідних зображень та досягнення високої точності, розширення даних штучно збільшує варіативність набору даних шляхом введення контрольованих, очікуваних варіацій (наприклад, різних кутів, освітлення або незначних деформацій). Навчання на цьому різноманітному, стандартизованому наборі даних збагачує модель варіаціям, що зустрічаються в реальному світі, покращуючи генералізацію, її надійність та здатність до узагальнення незнайомих клінічних даних.

У багатьох методах візуалізації, особливо у фотографії очного дна, інтерес представляє лише кругле поле зору сітківки, тоді як навколишня чорна або темна область не містить корисної інформації. Тому для стандартизації вхідних даних і видалення фону зазвичай застосовується комбінований підхід, що поєднує обрізку та кругле маскування. Обрізка зазвичай починається з виявлення неінформативних областей і вилучення мінімального обмежувального прямокутника, що містить сітківку. Наприклад методу ідентифікує всі пікселі вище порогу сірого кольору і вилучає обмежувальний прямокутник навколо них.

Після обрізки часто застосовується кругова маска – зазвичай шляхом визначення центру та радіуса диска сітківки або шляхом зміни розміру до квадрата, а потім маскування колом радіусом в половину найдовшої сторони та встановлення пікселів зовні до нуля.

Це гарантує, що вхідні дані для моделей мають однаковий масштаб і форму, усуває сторонні артефакти на кордонах і покращує вилучення ознак на нижчих рівнях, зосереджуючи обчислення на відповідній тканині.

### 2.3 Основні стратегії попередньої обробки для стандартизації та нормалізації зображень

Попередня обробка є важливим першим етапом інженерії даних, метою якого є перетворення необроблених зображень очного дна, що

містять артефакти, у стандартизований формат, придатний для надійного навчання ознакам.

Неоднорідний профіль інтенсивності та ефект віньєтування необхідно мінімізувати, щоб модель не інтерпретувала артефакти затінення як клінічно значущі ознаки. Одним із стандартних та ефективних методів корекції затінення є оцінка низькочастотної компоненти фонового освітлення. Це часто робиться шляхом згладжування зображення, зазвичай зеленого каналу, за допомогою медіанного фільтра з великим ядром (наприклад, розмір ядра 56x56). Отримане фонове зображення потім віднімається від зеленого каналу оригінального зображення, що ефективно мінімізує невеликі шмегулятини інтенсивності та нормалізує вигляд зображення.

Інші підходи передбачають згладжування оригінального зображення за допомогою фільтра нижніх частот або середнього чи медіанного фільтра, розмір якого перевищує найбільшу візуально видялючуся характеристику сітківки.

Перетворення колірного простору використовується для вирішення проблеми широкого діапазону контрастності в нативному колірному просторі RGB. Мета цих дій полягає в тому, щоб відокремити компонент інтенсивності від інформації про колір, що дозволяє цілеспрямовано поліпшити яскравість, не порушуючи колірний вміст. Модель CIE Lab, наприклад, вважається дуже підходящою для покращення, оскільки вона чітко відокремлює інтенсивність від двох інших колірних компонентів зображення.

Локальні адаптивні методи покращення контрасту, такі як CLAHE, можуть бути застосовані виключно до цього каналу інтенсивності. Ця процедура нормалізує яскравість і покращує локальний контраст, а подальше перетворення назад в RGB зберігає сприймані відносні колірні значення як на рисунку 2.5.

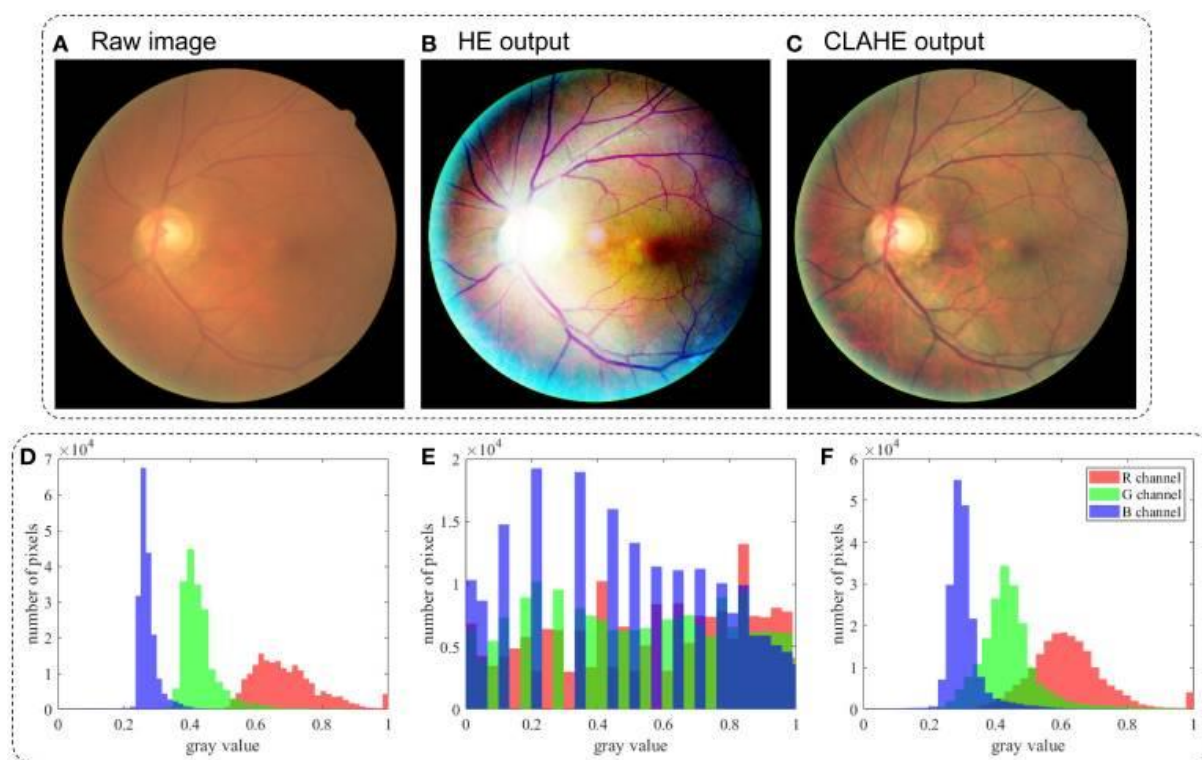


Рисунок 2.5 – Порівняння картинки без обробки, HE та CLAHE

Огляд відновлення зображення сітківки ока. (А) необроблене зображення. (В) зображення, покращене за допомогою HE. (С) зображення, покращене за допомогою CLAHE та гістограми значень пікселів для зображень (А), (В) та (С) відповідно.

CLAHE широко використовується в обробці біомедичних зображень завдяки своїй ефективності в підвищенні доступності найважливіших особливостей. Алгоритм спочатку розділяє зображення на роз'єднані області. У кожній області застосовується локальне вирівнювання гістограми. Після цього, для усунення меж між контекстними областями використовується білінійна інтерполяція, що дає в результаті плавне зображення з локальним поліпшенням.

Нормалізація кольору має на меті зробити зображення незмінним при змінах освітлення, що дозволяє алгоритму розрізняти особливості області інтересу (ROI) на основі внутрішніх характеристик, а не розбіжностей кольорів, спричинених різними пристроями збору даних. Стандартні методи

включають нормалізацію сірого світла, вирівнювання гистограми кольорів та специфікацію гистограми кольорів до стандартного еталонного зображення.

Фільтрування є основною операцією в обробці зображень сітківки, яка використовується для придушення шуму та поліпшення загальної якості зображення. Однак основною технічною проблемою є збереження тонких судин сітківки, які мають невелику площу та слабкий контраст порівняно з фоном. методи, такі як гаусівське фільтрування, застосовуються для загального придушення шуму. Більш ефективні, вдосконалені методи часто використовують адаптивні просторові ядра. Прикладом є оптимальне просторове ядро, яке використовується в двосторонньому фільтруванні, яке представлене лінійною функцією розсіювання з орієнтацією та масштабом, що адаптивно регулюються до локальної структури судин. Цей адаптивний підхід забезпечує збереження деталей зображення під часпогашення шумів на фото.

Високоєфективною стратегією для максимізації видимості особливостей є послідовність шумозаглушення та підвищення контрасту. Фільтр Вінера – це адаптивний фільтр, який виконує шумозаглушення, намагаючись зберегти краї зображення. Коли фільтр Вінера застосовується до зеленого каналу перед адаптивним вирівнюванням гистограми з обмеженим контрастом (CLANE), він виявляється ефективним для максимізації деталей кровоносних судин і загального зображення зеленого каналу (рисунок 2.6) [11].

Крім того, ця конкретна послідовність усуває небажані штучні межі та посилення шуму, які зазвичай асоціюються з алгоритмом CLANE при його окремому використанні, що призводить до значно покращеного зображення зі зменшеним шумом.

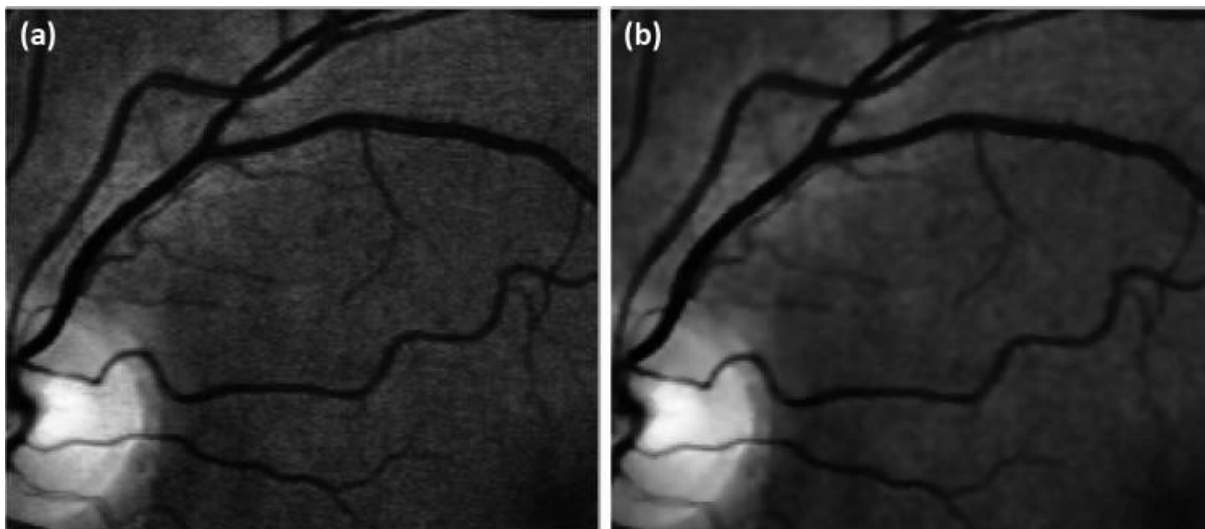


Рисунок 2.6 – Близьке порівняння якості збережених деталей з алгоритмом Вінера та без

Таким чином відбувається процес покращеної підготовки зображень до тренувань з метою покращення якості моделей шляхом обробки вхідних фотографій сітківки.

#### 2.4 Збільшення кількості та різноманітності даних

Аугментація даних розширює доступний набір даних шляхом генерації реалістичних варіацій існуючих зображень, що допомагає моделям краще узагальнювати та зменшувати вірогідність оверфітінгу. У медичних зображеннях як фотографії очного дна аугментація повинна залишатися анатомічно правдоподібною, одночасно вводячи варіації в просторовій орієнтації, яскравості та масштабі. Однією з поширених технік є обертання, при якому зображення обертаються на контрольований кут, щоб відобразити природні відмінності в положенні пацієнта. Наприклад, обертання зображень очного дна на  $\pm 15^\circ$  допомагає моделі вивчити незалежні від орієнтації візерунки в судинах і ураженнях.

Геометричні перетворення діють у просторовій області для імітації змін, що відбуваються під час зйомки зображення, таких як зміни кута нахилу камери, положення пацієнта або відстані. Вони вважаються основоположними для побудови надійності моделі проти природних просторових варіацій.

Перевертання – горизонтальне або вертикальне – є іншим широко використовуваним методом, хоча в деяких медичних галузях для збереження анатомічної правильності доречним є лише горизонтальне перевертання. Варіатор кольору змінює яскравість, контраст, насиченість або відтінок для імітації варіацій зображення, спричинених різними камерами або умовами освітлення. Практичним прикладом є застосування невеликого шуму яскравості для імітації варіативності освітлення сітківки. Збільшення масштабу вносить контрольовані зміни в масштаб, дозволяючи моделям стати стійкими до відмінностей у відстані камери або обрізці. Наприклад, випадкові перетворення збільшення можуть допомогти класифікатору залишатися стабільним, коли ураження з'являються в різних розмірах або положеннях у кадрі, як на рисунку 2.7.

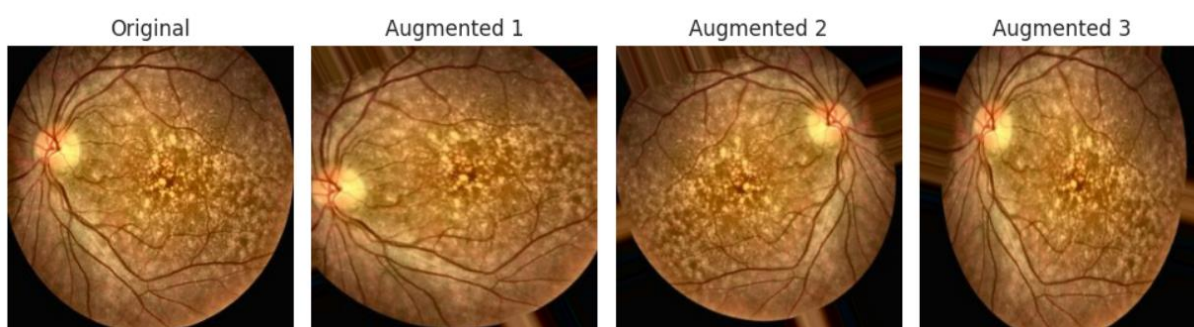


Рисунок 2.7– Приклад роботи аугментацій одного зображення

Розширення на основі ядра передбачає використання фільтрів згортки для зміни текстури зображення. Такі техніки, як підвищення різкості або розмиття зображення, імітують зміни фокусу або якості зйомки. Введення контрольованого шуму на основі ядра діє як стратегія регуляризації,

змушуючи модель надійно витягувати особливості незалежно від тонких відмінностей текстури, тим самим покращуючи її можливості узагальнення.

Хоча стандартні геометричні та фотометричні методи мають широке застосування, медичні зображення вимагають методів, які спеціально імітують реалістичну фізіологічну мінливість і враховують специфічні для даної області артефакти.

Еластична деформація (ED) – це спеціалізована техніка просторового спотворення, яка локально викривляє зображення, створюючи нежорсткі перетворення. Цей метод є особливо актуальним у медичній візуалізації, оскільки живі людські об'єкти природно піддаються незначним анатомічним змінам, рухам пацієнта та фізіологічним змінам. Завдяки введенню цих складних нежорстких модифікацій, ED дозволяє моделі краще узагальнювати та точно ідентифікувати анатомічні структури, навіть якщо вони виглядають дещо по-різному на різних сканах.

Для аналізу медичних зображень доповнені дані повинні суворо дотримуватися обмежень клінічного реалізму. Біоінспіровані техніки, такі як ED, є потужними, але їх потужність повинна контролюватися, щоб запобігти генерації даних з низькою точністю, які можуть внести шкідливий шум у процес навчання, рисунок 2.8. Дослідження, що вивчають оптимальне застосування ED до зображень сітківки чітко визначили ці обмеження на основі оцінки клінічних експертів в межах від 1 до 9.

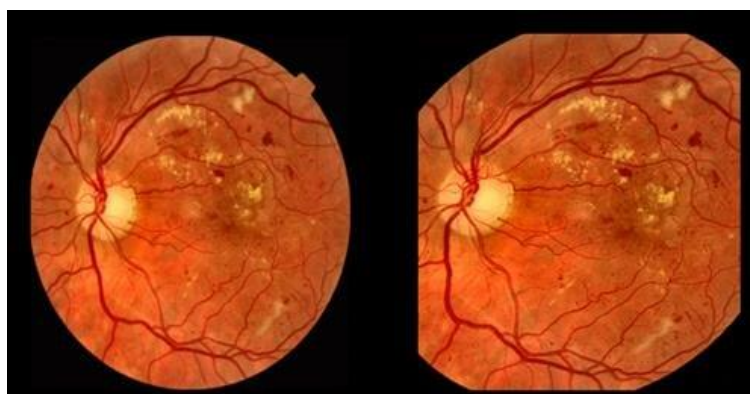


Рисунок 2.8 – Порівняння зображень еластичної деформації

Генеративні конкуруючі мережі (GAN) є просунутою версією вирішення проблеми недостачі даних, пропонуючи шлях до пом'якшення не тільки дефіциту кількості даних, але й проблем різноманіття та упередженості наборів даних. Умовні генеративні конкуруючі мережі cGAN генерують високоякісні синтетичні зображення шляхом навчання генераторної мережі проти дискримінаційної мережі.

Цей процес дозволяє створювати нові зразки, що мають характеристики, отримані з базового розподілу даних, обходячи залежність від дорогих анотованих наборів даних. Pix2Pix HD – це архітектура cGAN з високою роздільною здатністю, навчена на парних зображеннях очного дна та відповідних картах судин сітківки. Модель генерує синтетичні результати з високою роздільною здатністю залежно від судинної структури, рисунок 2.9 [12].

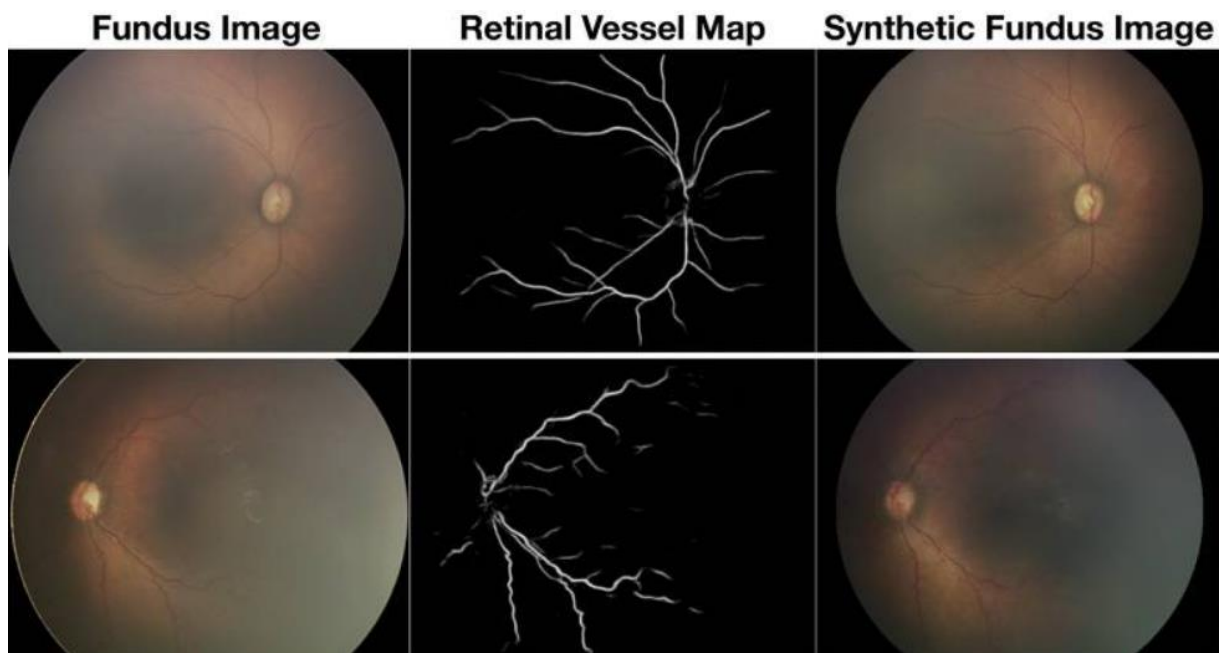


Рисунок 2.9 – Близьке порівняння якості збережених деталей з алгоритмом Вінера та без нього

Ця здатність підкреслює використання GAN для збільшення розміру та різноманітності наборів даних, зокрема шляхом синтезу зображень, що

поєднують існуючі патологічні особливості в нових конфігураціях, тим самим покращуючи генералізацію алгоритму для клінічного застосування.

Ефективність та якість моделей глибокого навчання для класифікації зображень очного дна в основному визначаються застосуванням попередньої обробки та збільшенням обсягу даних. Попередня обробка виконує важливу функцію стандартизації даних – виправлення системних недоліків, таких як віньєтування, шум і нерівномірне освітлення, а також вибіркоче посилення патологічних сигналів за допомогою таких технік, як SLANE. Збільшення обсягу даних базується на цій основі шляхом введення контрольованої різноманітності, тим самим покращуючи надійність моделі та запобігаючи оверфітінгу.

## 2.5 Методи балансування класів

Застосування алгоритмів глибокого навчання до завдань класифікації реальних даних часто стикається з проблемою незбалансованого розподілу класів, що є дуже властивим для таких галузей, як медична діагностика та спеціалізований аналіз зображень. Високий дисбаланс класів виникає, коли певні класи значно недопредставлені порівняно з домінуючими класами більшості, що призводить до упередження на користь більшості. Цей перекошений розподіл вносить байєс частоти в традиційні моделі машинного навчання, змушуючи їх надавати пріоритет загальній точності класифікації за рахунок показників продуктивності, критичних для класу меншості.

У сферах, де безпека має критичне значення, таких як комп'ютерна діагностика (CAD) очних захворювань, наприклад діабетичної ретинопатії та міопії та глаукоми, меншість класів часто представляє патологію, яку необхідно виявити на ранній стадії, щоб запобігти серйозним ускладненням. У цьому контексті невдача у точній класифікації випадків меншості безпосередньо призводить до потенційної помилкової діагностики пацієнтів

із серйозними захворюваннями, які можна було попередити. Отже, загальна точність є недостатнім показником корисності моделі. Оцінка ефективності повинна базуватися на надійних показниках, таких як чутливість (відтворюваність для класу меншості), специфічність та AUC.

Успішне впровадження спеціалізованих методів для протидії цій упередженості підтверджується клінічно значущими поліпшеннями, такими як повідомлене збільшення гармонійного середнього значення чутливості та специфічності для виявлення глаукоми з 89,09% до 92,59% на єгипетському наборі даних [13].

Стратегії, розроблені для усунення дисбалансу класів, можна умовно розділити на дві категорії: стратегії на рівні даних, які модифікують розподіл навчального набору, та стратегії на рівні алгоритмів, які коригують процес навчання або функцію cost. Методи на рівні даних включають повторну вибірку (надмірну вибірку, недостатню вибірку), збільшення обсягу даних та генерацію синтетичних даних. Методи на рівні алгоритмів охоплюють навчання з урахуванням вартості та модифікації функції втрат. Необхідність застосування цих методів збалансування є надзвичайно важливою для забезпечення переходу моделей глибокого навчання від академічної до практичної та клінічної застосовності.

### 2.5.1 Оверсемплінг, робота з простором ознак

Стратегії на рівні даних мають на меті маніпулювання розподілом навчальних даних для надання класифікатору збалансованого уявлення про всі класи. Хоча традиційні методи надмірного вибірки є простими та фундаментальними, вони стикаються із значними викликами при застосуванні до сучасних високорозмірних даних, таких як медичні зображення.

Техніка синтетичної надмірної вибірки меншин (SMOTE) є популярним фундаментальним методом надмірної вибірки. Вона працює

шляхом генерації синтетичних спостережень для класу меншин. Для вибраного екземпляра класу Синтетичні спостереження генеруються у випадковому діапазоні між цим спостереженням і його  $K$ -найближчими сусідами класу, зазвичай п'ять.

Ефективність традиційних геометричних методів, включаючи SMOTE, суттєво обмежується при роботі з високорозмірними даними, такими як зображення. Ці методи значною мірою покладаються на обчислення евклідових відстаней у просторі вхідних даних. При застосуванні до необроблених даних пікселів «прокляття розмірності» призводить до втрати дискримінаційного значення метрик відстані. Як наслідок, це може призвести до створення синтетичних зразків, які перекриваються з класом більшості або вводять шум. Таке погіршення оптимальних меж прийняття рішень знижує здатність моделі розрізняти класи. Емпіричні дослідження підтверджують, що для високорозмірних даних застосування SMOTE часто не дозволяє зменшити упередженість щодо класу більшості і може бути навіть менш ефективним, ніж просте виконання випадкового недобірки [14]. Крім того, процес генерації синтетичних зразків вносить сильну позитивну кореляцію між новоствореними зразками та оригінальними зразками меншості, що підвищує ризик перенавчання до локалізованих артефактів обмеженого розподілу меншості.

Щоб пом'якшити властиві недоліки геометричного оверсемплінгу в сирих картинках, сучасні методики працюють або в латентних просторах ознак, або за допомогою складного генеративного моделювання.

Оверсемплінг простору ознак використовує ознаки для стабілізації процесу передискретизації. Він передбачає зменшення нестабільності високих вимірів шляхом відокремлення вилучення ознак від стратегії передискретизації. Спочатку попередньо навчається кодер зображень – часто з використанням контрастного навчання – для генерації надійних і високодискримінаційних представлень ознак. Контрастне навчання

покращує представлення ознак шляхом вивчення подібностей і відмінностей між даними, забезпечуючи універсальність ознак для кожного класу, незалежно від дисбалансу. Потім оверсемплінг виконується в цьому низькорозмірному, значущому просторі ознак.<sup>7</sup>

Ефективність таких методів, як CL-SR, демонструє значну зміну в стратегічній меті сучасного надлишкового вибіркування для складних даних. Акцент більше не робиться виключно на коригуванні кількості класів, а в першу чергу на оптимізації якості множини ознак. Працюючи в надійному просторі ознак, залежність від простих, ненадійних метрик евклідової відстані в сирих даних мінімізується.<sup>7</sup>

Генеративні фреймворки, що використовують дифузійні моделі, представляють найсучасніший підхід до створення високоякісних, реалістичних синтетичних даних, тим самим вирішуючи проблеми як кількості, так і якості даних. Використання генеративних дифузійних моделей, продемонструвало успіх у вирішенні проблеми дисбалансу класів у складних медичних наборах даних, таких як виявлення глаукоми у раніше вказаному єгипетському дослідженні виявлення глаукоми.

### 2.5.2 Статичні та динамічні зважені функції втрат

Стратегії на рівні алгоритмів вирішують проблему дисбалансу класів шляхом модифікації цілі оптимізації, що відома як навчання з урахуванням вартості (CSL). CSL зміщує процес навчання в бік менш репрезентованих класів шляхом призначення штрафів, тим самим гарантуючи, що помилкові класифікації критичних, рідкісних класів спричиняють значно вищі штрафи моделі. Це безпосередньо впливає на величину та напрямок градієнтів під час навчання.

Впровадження CSL зазвичай передбачає модифікацію стандартної функції втрат кросс-ентропії з метою створення ваг, специфічних для кожного класу.

Найпростіший статичний підхід CSL визначає ваги класів статично на основі частоти класу. Це часто передбачає встановлення ваг, пропорційних оберненій частоті класу. Існують різні схеми перетворення вектора штрафу, включаючи лінійний, логарифмічний та квадратний корінь. Однак просте зваження за оберненою частотою класу може призвести до поганої продуктивності для частих класів, оскільки воно має тенденцію надмірно штрафувати більшість класів без необхідності [15]. Щоб запобігти цьому, більш просунуті статичні підходи CSL пропонують зваження на основі оберненої ефективної кількості зразків, метрики, яка враховує структуру даних і надмірність, а не тільки сирий підрахунок, пропонуючи більш репрезентативне та правильне вимірювання рідкості класу.

Динамічні зважені функції втрат розширюють CSL, дозволяючи вагам штрафів адаптуватися не тільки на основі частоти класу, але й на основі результату прогнозування моделі. Наприклад, динамічно зважена збалансована втрата призначає класам ваги на основі як частоти класу, так і прогнозованої ймовірності складності класифікації випадку.

Здатність динамічної схеми зважування самостійно адаптувати свої ваги відповідно до оцінок прогнозу дозволяє моделі коригувати випадки з різним рівнем складності. Ця структура гарантує, що оновлення градієнта обумовлюються саме зразками важких меншинних класів – тими, які модель не може правильно класифікувати, – що призводить до надійної генералізації в дуже незбалансованих наборах даних, включаючи медичну візуалізацію.

Фундаментальна різниця між цими двома підходами CSL полягає в методології: статичний CSL оптимізує рідкість на рівні класу, застосовуючи однакове високе покарання до всіх помилкових класифікацій меншин. Динамічний CSL, навпаки, оптимізує складність на рівні зразків. Включаючи оцінку прогнозу, динамічні схеми можуть відфільтрувати прості приклади в класі меншості, забезпечуючи, що потужність мережі спрямована на вирішення найскладніших граничних випадків. Ця

пріоритетність складних випадків над просто рідкісними дозволяє охопити дуже важливі краї вибірки.

### 2.5.3 Focal Loss

Focal Loss (FL) – це прогрес у динамічному оверсемплінгу вибірки, розроблений спеціально для вирішення проблеми сильного дисбалансу класів, що часто зустрічається в завданнях з високою складністю прогнозування.

FL був спочатку запроваджений для вирішення проблеми нестабільності навчання в одноступеневих детекторах об'єктів, таких як RetinaNet, які оцінюють вибірку можливих місць розташування об'єктів [16]. Ця вибірка призводить до значного дисбалансу класів, коли прості приклади фону значно переважають цілі на передньому плані. У стандартній функції лоссу кросс-ентропії, навіть якщо кожен простий негативний приклад сприяє незначній індивідуальній втраті, їхній великий обсяг створює кумулятивну втрату, яка домінує в загальному градієнті. Цей кумулятивний ефект заважає ефективному навчанню на кількох критичних, складних прикладах, що погіршує продуктивність детектора.

Focal Loss вирішує цю проблему шляхом переформування стандартної кросс-ентропії додаючи модулюючий коефіцієнт, щоб зменшити вагу добре класифікованих прикладів.

Параметр гамма контролює швидкість зменшення лоссу для добре класифікованих прикладів, тим самим ефективно фокусуючи навчання на розрідженому наборі складних, неправильно класифікованих прикладів. Зі збільшенням гамма ефект зниження ваги стає більш вираженим. Наприклад, якщо коефіцієнт встановлено на 2, приклад, який модель класифікує з високою впевненістю, матиме збиток, зменшений у 100 разів порівняно зі стандартним збитком кросс ентропії. Це агресивне придушення гарантує, що прості приклади вносять незначний внесок в оновлення ваг мережі, тим

самим запобігаючи тому, щоб велика кількість простих негативних прикладів перевантажувала процес оптимізації.

У медичній галузі Focal Loss виявився дуже ефективним у сегментації медичних зображень, де цільові ураження або пухлини часто займають значно менший об'єм порівняно з фоном, створюючи середовище з надзвичайним дисбалансом щільності переднього плану та фону [17].

Кількісні порівняння підкреслюють його перевагу над бінарною кросс-ентропією у завданнях сегментації. Наприклад, для сегментації пухлин головного мозку використання Focal Loss призвело до суттєвого поліпшення діагностичних показників: показник F1 покращився з 74,52% для binary CE до 81% для Focal Loss, а точність зростає з 99,03% до 99,44%.

Однак ефективність Focal Loss не є абсолютною для всіх завдань комп'ютерного зору. Дані аналізу зображень очного дна, зокрема сегментації судин, вказують на обмеження в обробці тонких, дрібних деталей. В одному дослідженні було виявлено, що категоріальна кросс-ентропія є кращим вибором для чіткої сегментації тонких кровоносних судин порівняно з категоріальним Focal Loss (рисунки 2.10) [18].

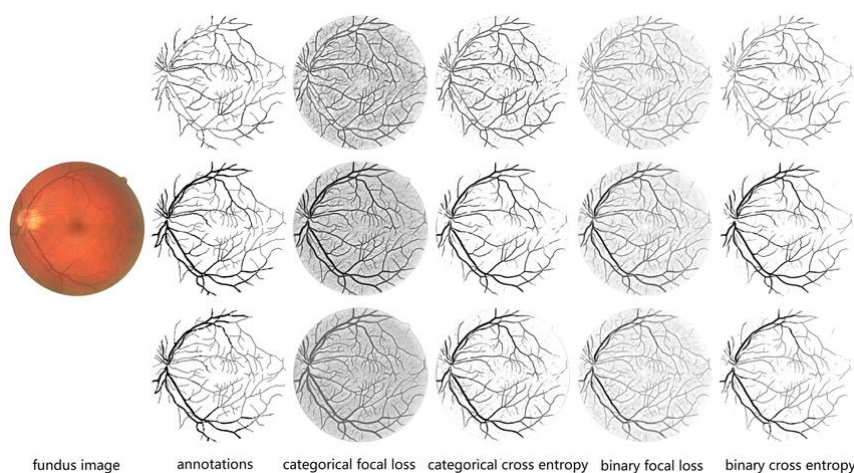


Рисунок 2.10 – Порівняння функцій втрат для судин

Ця суперечність вказує на важливу відмінність: Focal Loss є високооптимізованим для надійного виявлення об'єктів в умовах

надзвичайної нерівноваги шляхом агресивного придушення фону. Стандартна перехресна ентропія, підтримуючи більш збалансований градієнтний внесок по всіх пікселях, може бути більш придатною для високоточного уточнення меж складних, тонких структур. Для широких, рідкісних уражень ідеально підходить агресивний градієнтний фокус FL, і навпаки, для надзвичайно тонких анатомічних структур, таких як тонкі судини, високий параметр гамма FL може ненавмисно придушувати внесок пікселів меж, що були трохи неправильно класифіковані, що призводить до розмиття або нестабільності, з якими стандартна категоріальна SE справляється більш плавно. Тому вибір функції втрат, а саме параметрів повинен базуватися на масштабі та складності цільових ознак.

Рішення щодо оптимальної стратегії балансування класів вимагає зваженого підходу, який враховує складність реалізації та вартість комп'ютеру у порівнянні з конкретними перевагами методів на рівні алгоритмів у плані контролю моделі. У складних завданнях мультикласової класифікації оптимальна продуктивність моделі часто досягається шляхом поєднання кількісних методів для даних та алгоритмічних методів.

## 3 МЕТОДИ МАШИННОГО НАВЧАННЯ

### 3.1 Бейзлайн модель Random Forest

Бейзлайн модель класифікації зображень – це перша, найпростіша система, яка створюється, щоб мати можливість порівнювати з нею всі наступні моделі. Мета полягає не в тому, щоб досягти найвищої точності, а в тому, щоб надати еталонний рівень продуктивності моделі на даних та допомогти зрозуміти, чи дійсно більш складні підходи додають якості моделі.

Класичні методи машинного навчання давно використовуються для аналізу медичних зображень, особливо часто вони використовувалися до появи глибоких згорткових мереж. Ці моделі базуються на рукотворних ознаках – величинах, отриманих із інформації на зображеннях. У контексті візуалізації сітківки ока, такими ознаками можуть бути, кольори на гістограмах зображень, розташування судин або контрасти інтенсивності на у важливих регіонах навколо диска зорового нерва або макули.

Random Forest – це ансамблевий метод, що складається з декількох дерев рішень, кожне з яких навчається на підмножині даних, та випадковій підмножині ознак у кожному вузлі [19]. Під час інференсу голоси окремих дерев агрегуються за допомогою голосування більшістю для отримання остаточного прогнозу класу. Таке поєднання методу забезпечує різноманітність дерев і зменшує перенавчання.

Однак існують також певні обмеження, Random Forest виходить з незалежності між зразками; він не вивчає просторові ієрархії в зображенні, що означає, що він може не вловити тонкі закономірності, які можуть використовувати глибокі мережі. Об'явлення моделі можна побачити на рисунку 3.1.

```
rf = RandomForestClassifier(n_estimators=200, random_state=42, n_jobs=-1)
rf.fit(X_train, y_train)
```

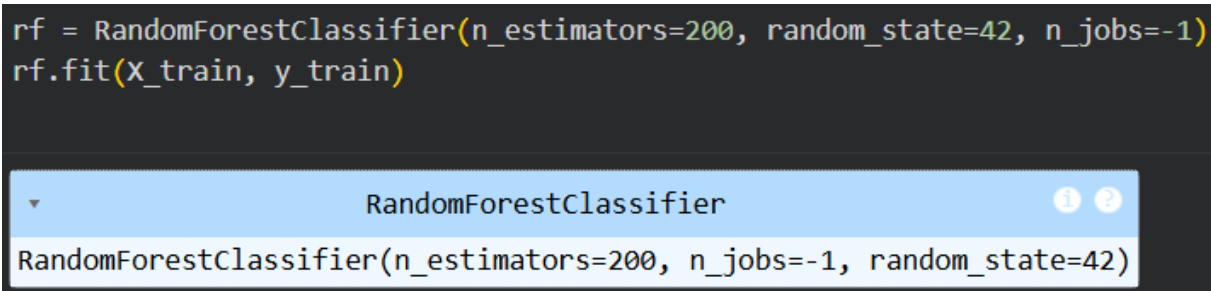


Рисунок 3.1 – Програмний код побудови моделі RandomForest

Застосовуючи стандартний `RandomForestClassifier` до ознак, витягнутих із зображень очного дна розміром  $256 \times 256$  після попередньої обробки та аугментації, створюється стандартна база з невеликими зусиллями та коротким часом тренуванням, яка точно продемонструє, яку цінність додають моделі глибокого навчання та інші складніші техніки в подальших експериментах з моделлю.

### 3.2 Моделі глибокого навчання. і базові архітектури

На відміну від базової моделі, побудованої на штучно створених ознаках, архітектури глибокого навчання, зокрема конволюційні нейронні мережі (CNN), здатні навчатися багатим ієрархічним представленням безпосередньо з необроблених піксельних даних. Ці моделі, попередньо навчені на великих наборах візуальних даних і точно налаштовані та натреновані для специфічного завдання, дозволяють виявляти тонкі ознаки на зображеннях сітківки, які можуть бути важко виявити за допомогою ознак з RF.

#### 3.2.1 VGG16

VGG16 – це класична глибока конволюційна мережа, розроблена групою з Оксфорду. Вона має 16 шарів: 13 конволюційних шарів, за якими

йдуть 3 повністю з'єднані шари. , Конволюційна частина використовує виключно ядра  $3 \times 3$  з кроком 1, а мережа застосовує max-pooling після груп згорток, щоб систематично зменшувати просторову роздільну здатність. В кінці два великі повністю з'єднані шари (кожен з 4096 нейронами) подаються в кінцевий шар класифікації. Ця конструкція полегшує ієрархічне вивчення ознак: низькорівневі такі як краї та текстури у ранніх шарах і більш складні форми як структури судин, області зорового диска у вищих шарах (рисунок 3.2) [20].

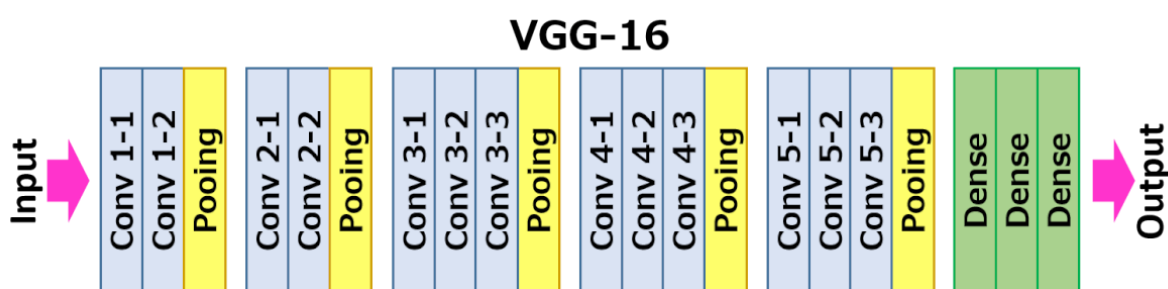


Рисунок 3.2 – Архітектура мережі VGG-16

VGG16 часто використовується як базова модель для трансферного навчання в медичній візуалізації завдяки своїй простоті та добре зрозумілій поведінці. Попередньо навчені ваги широко доступні, а архітектура є прозорою: відносно легко перевірити активації та зрозуміти, що вивчає мережа.

Однак VGG16 має багато параметрів, а його повністю з'єднані шари роблять його обчислювально важким. Ці фактори можуть зробити точне налаштування на невеликих медичних наборах даних більш схильним до перенавчання, а робота може бути повільнішою у порівнянні з більш оптимізованими, сучасними архітектурами.

### 3.2.2 ResNet50

У 2015 році команда Microsoft Reserch представила архітектуру, яка стала проривом у комп'ютерному зорі, вирішивши проблему зі зникаючими градієнтами у тренуванні дуже глибоких моделей розміром більше 50 шарів. Пропуск з'єднань є одним з ключових аспектів, що роблять ResNet особливо потужним, покращуючи генералізацію та покращуючи градієнт. У традиційній моделі нейронної мережі інформація проходить через усі шари мережі, без можливості «пропустити» проміжні шари. Однак у ResNet пропуск з'єднань дозволяє частині інформації оминати один або декілька конволюційних шарів і бути доданою безпосередньо до кінцевого результату більш глибокого шару.

Зокрема, ResNet50 складається з блоків-ботлнеків: кожен блок має три конволюційні шари (1x1, потім 3x3, потім 1x1), де перший зменшує розмірність, другий виконує основну конволюцію, а третій відновлює розмірність, це допомагає дозволяє тримати під контролем кількість параметрів, що важливо для глибоких моделей (рисунок 3.3).

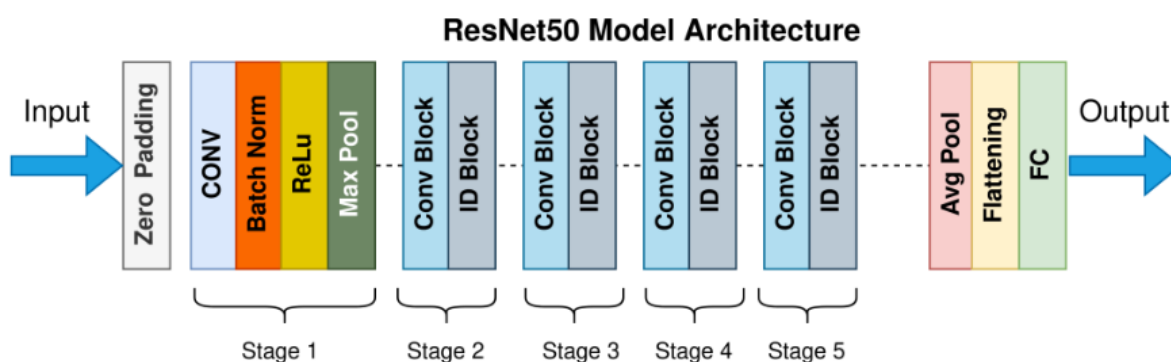


Рисунок 3.3 – Архітектура мережі ResNet50

ResNet50 часто використовується в сценаріях, де потрібне точне представлення, але ризик перенавчання високий, наприклад, у діагностиці медичних сканів. Його пропускні з'єднання допомагають мережі надійно

вивчати як низькорівневі краї та невеликі ураження, так і високорівневі анатомічні структури.

Більше того, у завданнях візуалізації сітківки варіанти ResNet показали високу ефективність до 97% у поєднанні з етапами попередньої обробки, такими як CLANE, про які говорилося у попередньому розділі [21].

### 3.2.3 EfficientNet (B0–B7)

У 2019 році дослідники Google AI представили EfficientNet, сучасну модель комп'ютерного зору, створену для розпізнавання об'єктів і закономірностей на зображеннях. Вона була розроблена в першу чергу для класифікації зображень, що передбачає віднесення зображення до однієї з декількох заздалегідь визначених категорій. Однак сьогодні EfficientNet також служить основою для більш складних завдань, таких як виявлення об'єктів, сегментація та трансферне навчання.

EfficientNet – це сімейство моделей, розроблених із використанням комбінованого масштабування, яке рівномірно масштабує три виміри мережі: кількість шарів, кількість каналів та роздільну здатність зображень. Замість того, щоб масштабувати ці фактори незалежно один від одного, EfficientNet визначає оптимальний баланс між ними. Починаючи з базової мережі EfficientNet-B0, кожна наступна модель B1, B2, ..., B7 масштабує ці виміри відповідно до комбінованого коефіцієнта, що дозволяє обрати компромісну за потужністю модель в залежності від можливих обчислювальних витрат. Загальна кількість шарів в EfficientNet-B0 становить 237, а в EfficientNet-B7 – 813. Всі моделі складаються з однакових блоків та підблоків у різній кількості (рисунок 3.4).

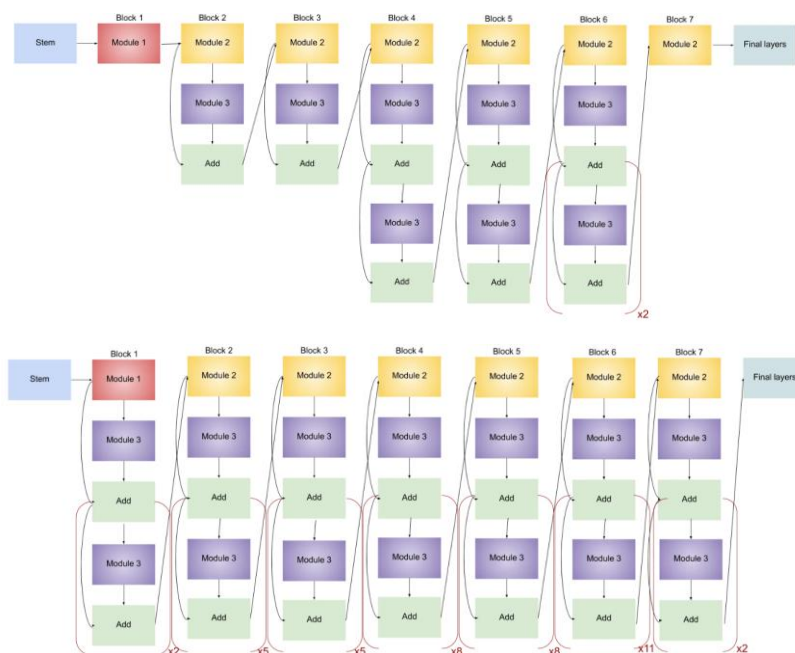


Рисунок 3.4 – Порівняння архітектури мереж EfficientNet-B0 та EfficientNet-B7

У таких завданнях, як класифікація раку шкіри, EfficientNet-B7 продемонстрував високу продуктивність, залишаючись компактним. На наборах даних рентгенівських знімків грудної клітки EfficientNets продемонстрував хорошу продуктивність з поліпшенням AUC і низькою вартістю інференції. Його ефективність робить його привабливим для використання в клінічних умовах, де обчислювальні ресурси можуть бути обмеженими.

EfficientNet є свідченням винахідливості сучасних архітектур глибокого навчання. Його масштабована конструкція в поєднанні з ефективними методами навчання робить його універсальним інструментом для безлічі завдань комп'ютерного зору.

### 3.2.4 MobileNet

MobileNet – це сімейство CNN, оптимізоване для ефективного використання на мобільних або вбудованих пристроях. Ключовим

елементом цих моделей є глибинна роздільна згортка – один фільтр на вхідний канал, за якою слідує точкова 1x1 згортка для об'єднання каналів. Це значно зменшує кількість параметрів і обчислювальні витрати порівняно зі стандартними згортками. MobileNet також пропонує гіперпараметри, такі як коефіцієнт ширини, що контролює кількість каналів та коефіцієнт роздільної здатності вхідного зображення, що дозволяє досягти кращого балансу між швидкістю та точністю (рисунок 3.5).

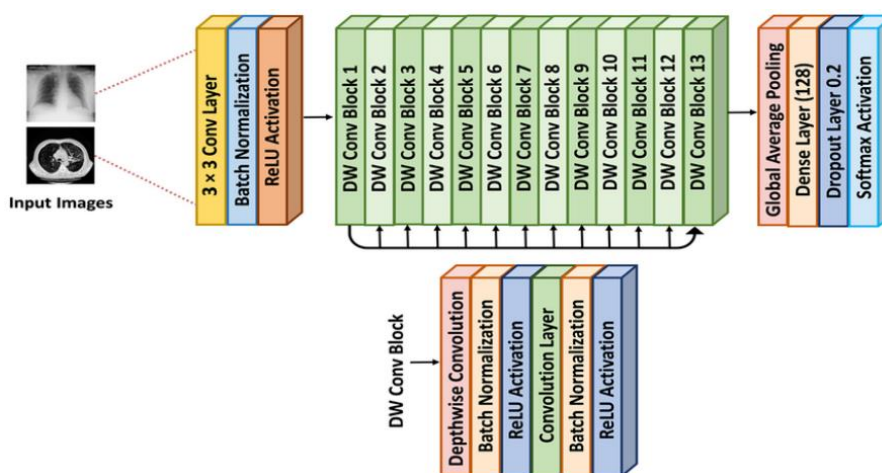


Рисунок 3.5 – Архітектура мережі MobileNet

MobileNet є дуже актуальним для реального застосування в клінічних умовах, де може знадобитися проведення тестування на мобільних пристроях таких як інструменти скринінгу, системи пунктів надання медичної допомоги.

Завдяки своїм невеликим розмірам і низькій затримці, він може використовуватися в телемедицині або портативних діагностичних інструментах, наприклад, для аналізу очного дна за допомогою камери смартфона.

Більш того, в обмежених умовах MobileNet може бути більш придатним для точного налаштування, ніж дуже велика CNN.

На зображенні 3.6 можна побачити порівняння великої кількості архітектур моделей на датасеті ImageNet, у порівнянні з попередньою найкращою моделлю CNN, EfficientNet-B7 була в 6,1 раза швидшою та в 8,4 раза меншою за розміром [22].

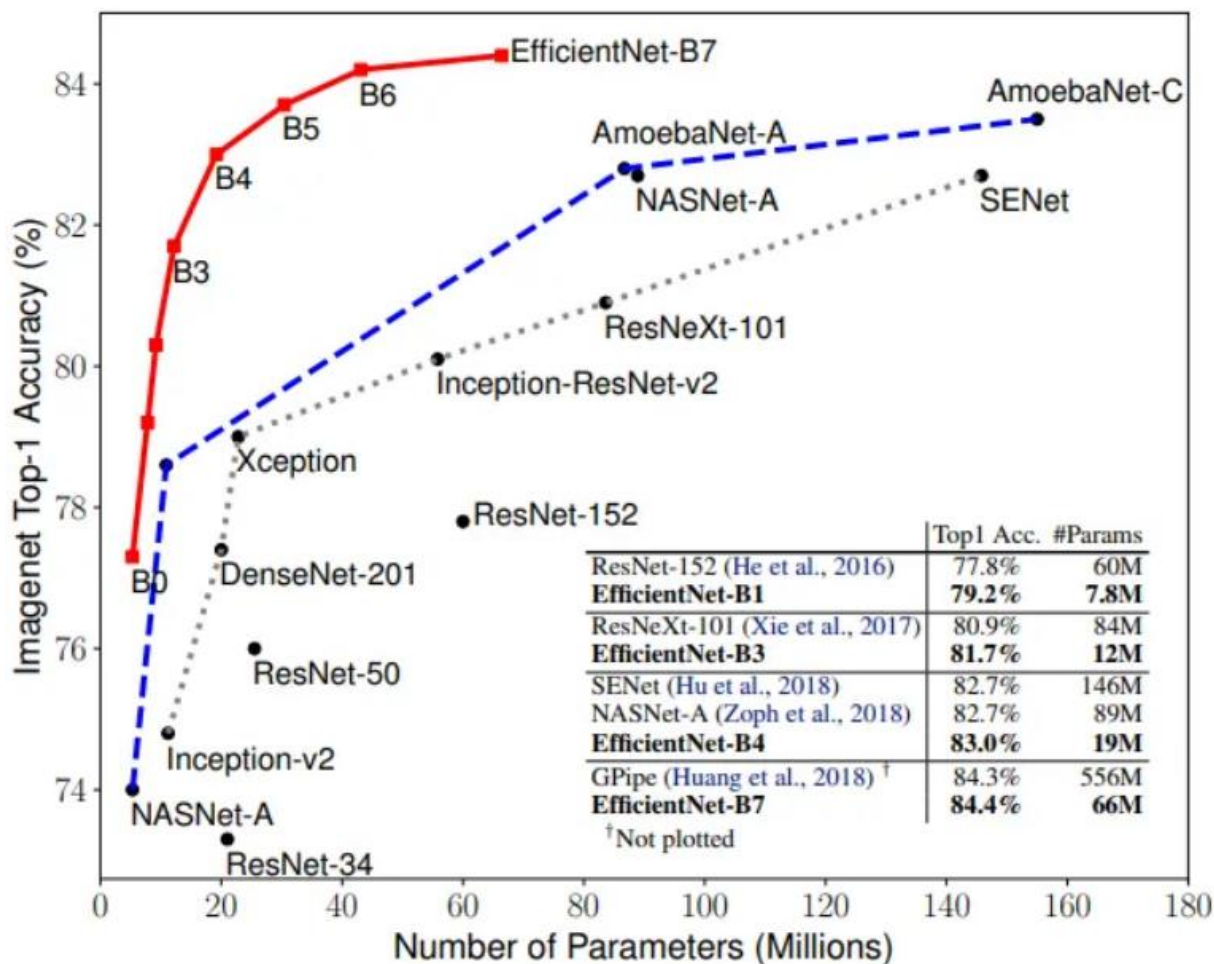


Рисунок 3.6 – Порівняння розміру та точності моделей на датасеті ImageNet

### 3.2.5 Vision Transformer

Vision Transformer (ViT) представляє собою зміну парадигми в комп'ютерному зорі, переносячи архітектури на основі трансформаторів, спочатку розроблені для обробки природної мови, в область зображень. Замість обробки піксель за пікселем або за допомогою ковзних згорткових

фільтрів, ViT розділяє вхідне зображення на фрагменти фіксованого розміру, вирівнює кожен фрагмент і вбудовує їх у простір нижчої розмірності. Ці ембедінги доповнюються позиційними кодуваннями для збереження просторового розташування, а потім обробляються як послідовність, яка проходить через стандартний кодер трансформера.

Однією з найсильніших переваг ViT є його здатність пов'язувати кожен патч з кожним іншим патчем за допомогою самоуваги, а не покладатися на локальні рецептивні поля, як у CNN. У контексті візуалізації сітківки це є дуже корисним: патологічні ознаки, такі як мікроаневризми, крововиливи або друзи, можуть бути просторово розподілені, і їх значення може залежати від їхнього зв'язку з іншими ділянками (макула, диск зорового нерва, судини). Глобальний механізм уваги допомагає моделі розуміти ці взаємозв'язки між патчами ефективніше, ніж суто локальні конволюційні фільтри.

ViT можуть отримати значну користь від попереднього навчання на великих наборах візуальних даних, а потім бути точно налаштовані на медичному наборі даних. У медичній візуалізації, де марковані дані часто є дефіцитними, попередньо навчені ViT можуть забезпечити міцну основу, зменшуючи оверфітінг та покращуючи дженералізацію. Проте таке навчання потребує велику кількість зображень та комп'юту, що ускладнює тренування на даних з великою роздільною здатністю.

Що стосується візуалізації сітківки, то в останніх дослідженнях почали вивчати самоконтрольоване попереднє навчання ViT для кращого використання немаркованих даних очного дна.

### 3.3 Трансферне навчання

Анотовані медичні набори даних часто є невеликими через те що медичні дані є дорогими в розмітці і потребують анотації експертами, а навчання глибоких мереж з нуля вимагає великих обсягів даних і значних

обчислювальних ресурсів. Трансферне навчання зменшує ці вимоги, використовуючи раніше набуті знання, що дозволяє ефективно навчати моделі навіть за обмежених медичних даних. Більше того, ознаки, вивчені на загальних зображеннях, часто можуть добре переноситися на медичні зображення – особливо низькорівневі ознаки.

Трансферне навчання означає повторне використання моделі навченої на одній задачі, для іншої, але пов'язаної з нею задачі, замість навчання всієї моделі з нуля. У комп'ютерному зорі це зазвичай означає використання мережі, попередньо навченої на великому загальному наборі даних як ImageNet, а потім її фінетюнінг до завдання медичної візуалізації, такого як класифікація захворювань сітківки. Для такого тренування моделі зазвичай заморожується якась кількість шарів вихідної моделі та замінюється фінальний класифікуючий шар, рисунок 3.7. Таким чином тренувальний час та потрібні ресурси значно скорочуються, а якість моделі значно краще від тренованої з рандомної ініціалізації моделі [23].

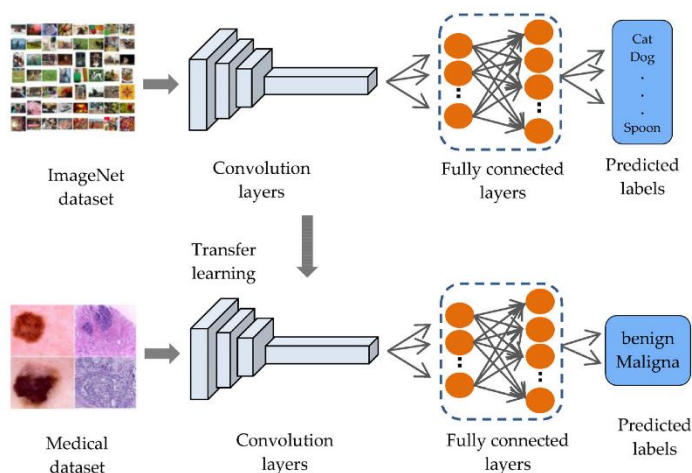


Рисунок 3.7 – Візуалізація схеми трансферного навчання

Підсумовуючи, трансферне навчання пропонує прагматичний і потужний шлях для використання глибокого навчання в завданнях медичної візуалізації, таких як класифікація захворювань сітківки. Воно зменшує

вимоги до даних і обчислювальних потужностей, прискорює конвергенцію і часто дає сильну генералізацію, що є критично важливим, коли анованих медичних даних не вистачає. Проте, воно вимагає ретельного налаштування і продуманого врахування відмінностей між доменами, потенційних упереджень і інтерпретованості.

### 3.4 Інтерпретованість в медицині

У медичній візуалізації, включаючи класифікацію захворювань сітківки, інтерпретованість є не просто бонусом, а необхідною умовою. Лікарям важливо розуміти, чому модель ставить певний діагноз, перш ніж довіряти її результатам або діяти відповідно до них. Без прозорості інструменти ШІ ризикує бути сприйнятим як «чорна скринька, що обмежує його застосування в реальних клінічних умовах. ХАІ розкриває, які області зображення або особливості впливають на прогнози моделі, тим самим даючи лікарям можливість перевіряти рішення, прийняті на основі ШІ, для їх прийняття або відкидання.

Більше того, пояснюваність сприяє безпеці моделі та виявленню упередженості: візуалізуючи, на чому зосереджується модель, можна виявити, чи звертає вона увагу на нерелевантні або помилкові області, що може вказувати на небажану поведінку.

Одним з найбільш широко використовуваних інструментів пояснюваності в глибокому навчанні для обробки зображень є Grad-CAM (Gradient-weighted Class Activation Mapping). У контексті візуалізації сітківки Grad-CAM використовувався для накладання карт уваги на фотографії очного дна, щоб показати, які анатомічні ділянки елементів очного дна модель вважала найважливішими при діагностуванні захворювання (рисунок 3.8) [24].

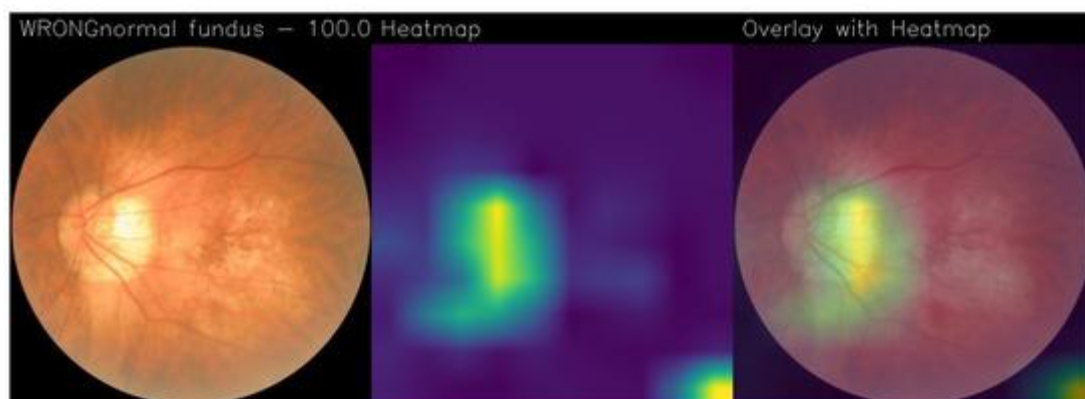


Рисунок 3.8 – Фото очного дна з міопією та теплова карта уражених регіонів за Grad-CAM

До негативних характеристик таких видів моделей можна віднести їх чутливість до даних, архітектури та відсутність універсальної методики тестування та стандартизації в галузі.

У завданнях з високими ризиками, таких як діагностика захворювань сітківки, наявність інтерпретованих пояснень є критично важливою для дотримання нормативних вимог, довіри лікарів до інструменту та прийняття пацієнтами.

## 4 ОЦІНКА ТА РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ

### 4.1 Метрики в оцінці медичних систем

Розроблена модель потребує обов'язкової оцінки якості для розуміння подальших дій та розбору слабких та сильних сторін системи [25]. Точна валідація медичних моделей машинного навчання починається з чіткого розуміння чотирьох основних результатів, отриманих з матриці помилок. Цей базовий інструмент підсумовує ефективність класифікаційної моделі шляхом підрахунку справжніх позитивних, справжніх негативних, помилкових позитивних і помилкових негативних прогнозів у порівнянні з реальними класами картинок. Кожен наступний показник, що використовується для оцінки, розраховується на основі цих підрахунків.

Матриця помилок визначає, де модель була успішною, а де зазнала невдачі, надаючи важливі кількісні дані: True Positive випадок, коли модель правильно визначає наявність медичного стану, True Negative випадок, коли модель правильно визначає відсутність хвороби, класифікуючи здорового пацієнта як не хворого. False Positive (FP) результат випадок, коли модель неправильно прогнозує наявність захворювання, яке насправді відсутнє. Це характеризується як «хибна тривога», False Negative (FN) випадок, коли модель неправильно прогнозує відсутність захворювання, яке насправді присутнє.

У медичних ситуаціях, де на кону стоять високі ризики, наслідки неправильної класифікації є асиметричними, тобто вартість помилки FP рідко дорівнює вартості помилки FN. Це означає, що модель потрібно будувати з урахуванням такого перекосу.

Помилки типу FN мають серйозні наслідки, оскільки призводять до затримки діагностики та лікування, що може призвести до затримати надання невідкладної допомоги. Пропуск справжнього позитивного випадку, чи то через помилку моделі, чи низьку якість фотографії, є

провалом критичного виявлення. І навпаки, FP спричиняють непотрібні терапевтичні втручання, включаючи додаткові процедури, дорогі повторні тести та призначення невірної лікування, а також викликають занепокоєння у пацієнтів. У умовах масштабного скринінгу високий рівень FP споживає цінні ресурси, такі як час фахівців та ресурси лікарень. Мінімізація FP є особливо важливою, коли підтверджувальні повторні процедури самі по собі є інвазивними або несуть значний ризик.

Співвідношення прийнятних помилок типу False Negative до помилок типу False Positive визначається тяжкістю захворювання та ризиками подальших втручань. У випадку небезпечної, швидко прогресуючої хвороби де ціна помилки дуже висока модель повинна бути спрямована на мінімізацію помилок такого типу, навіть якщо це призведе до більшої кількості помилок, що призводять до глибшої перевірки.

Чутливість і специфічність є основними показниками в клінічній епідеміології, що оцінюють внутрішню здатність моделі правильно класифікувати пацієнтів відповідно до їхнього справжнього клінічного стану. Чутливість вимірює ефективність моделі у правильному виявленні позитивних випадків, коли захворювання дійсно присутнє. Принцип розрахунку полягає у порівнянні кількості справжніх правильно виявлених хворих осіб із загальною кількістю осіб, які дійсно хворіють на це захворювання тобто сума справжніх позитивних результатів та хибних негативних результатів. З клінічної точки зору, тест з високою чутливістю гарантує, що буде пропущено мало позитивних випадків.

Специфічність вимірює ефективність моделі у правильній ідентифікації негативних випадків, коли захворювання дійсно відсутнє. Принцип розрахунку полягає у порівнянні кількості правильно відхилених здорових осіб із загальною кількістю осіб, які дійсно здорові – сума помилкових позитивних та справжніх негативних результатів. Висока специфічність тесту гарантує, що лише невелика кількість здорових осіб буде помилково визнана позитивною [26].

Хоча чутливість і специфічність забезпечують внутрішні характеристики тесту, для прийняття клінічних рішень необхідні прогностичні показники, що враховують ймовірність захворювання пацієнта після отримання результатів тесту. На відміну від чутливості та специфічності, ці показники значною мірою залежать від поширеності захворювання серед вибірки в датасеті та балансу класів.

Точність вимірює загальну частку всіх прогнозів які є правильними відносно загальної кількості зроблених прогнозів. Хоча вона дає швидкий знімок ефективності, у медичній сфері вона має фатальний недолік. У наборах даних, де захворювання є рідкісним, модель може досягти оманливо високої точності, просто завжди прогнозуючи «здоровий» стан для кожного пацієнта. Наприклад, захворювання, яке вражає 1 з 1000 пацієнтів, дозволяє нефункціональній моделі досягти 99,9% точності при нульовому рівні істинних позитивних результатів, що означає, що вона пропускає кожного пацієнта, який дійсно потребує допомоги.

Пресіжен стосується надійності позитивного результату тесту. Принцип розрахунку полягає в порівнянні кількості правильних визначень із загальною кількістю випадків, коли модель зробила позитивний прогноз. Це відображає ймовірність того, що пацієнт дійсно хворіє на захворювання, з огляду на його позитивний результат тесту. Це корисно, коли вартість помилкових позитивних результатів є високою, що дає впевненість перед переходом до подальших терапевтичних заходів. Негативна прогностична цінність є оберненою версією пресіжена для негативних випадків, що стосується надійності негативного результату тесту.

Для незбалансованих даних є F1. F1-score – це важливий комплексний показник, призначений для надання збалансованого підсумку ефективності моделі щодо позитивного класу шляхом гармонізації точності та відтворюваності. Принцип розрахунку полягає у визначенні гармонійного середнього значення точності та відтворюваності. Метрика F1 є надзвичайно важливою для таких застосувань, як виявлення рідкісних

захворювань, де невдача у виявленні випадків або велика кількість помилкових сигналів мають серйозні клінічні наслідки для діагностики захворювань.

Багато медичних моделей видають безперервний показник ймовірності, а не жорстку класифікацію. Для оцінки цих моделей потрібні інструменти, які оцінюють дискримінаційну здатність за всіма можливими пороговими значеннями прийняття рішень.

Крива ROC – це потужний інструмент візуалізації, який використовується для ілюстрації ефективності бінарної класифікаційної моделі при різних порогових значеннях. Принцип візуалізації полягає в побудові графіка справжньої позитивної частоти на осі Y проти помилкової позитивної частоти на осі X. Ця візуалізація дозволяє медичним дослідникам бачити результат компромісу між чутливістю та специфічністю на кожній потенційній межі прийняття рішення. Крива, яка сильно вигинається у верхній лівий кут, вказує на високу точність тесту з відмінною дискримінацією. Аналіз ROC є вирішальним для вибору потенційно оптимальних моделей до того, як буде остаточно визначено фіксований клінічний поріг.

Площа під кривою ROC узагальнює дискримінаційну здатність моделі в єдину статистику, незалежну від порогу. Імовірно, представляє ймовірність того, що класифікатор присвоїть випадково вибраному позитивному екземпляру вищий бал, ніж випадково вибраному негативному екземпляру. Це означає, що AUC близький до 1,0 відображає відмінну внутрішню здатність ранжувати пацієнтів за ризиком. Однак AUC може бути оманливо високим у сильно незбалансованих медичних наборах даних. Приклад вигляду кривої ROC з виділеною AUC можна побачити на рисунку 4.1.

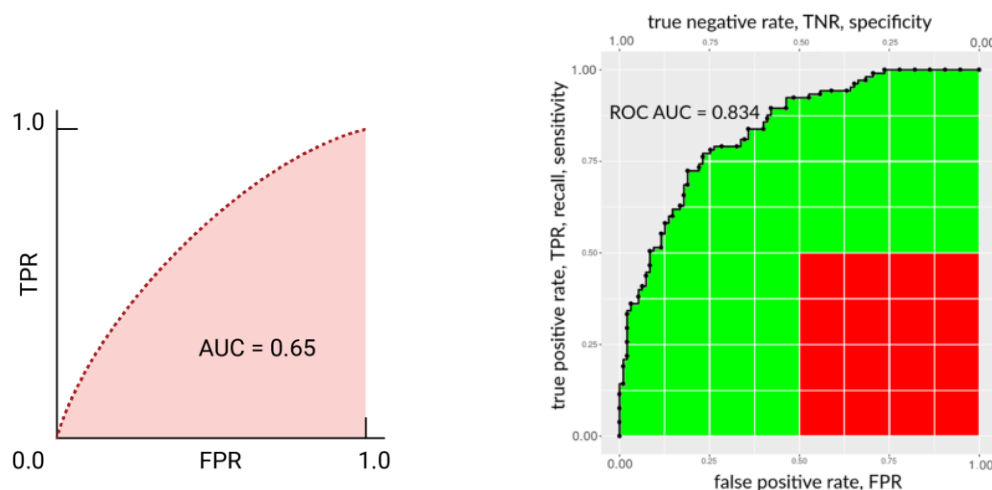


Рисунок 4.1 – ROC і AUC двох гіпотетичних моделей

Знаючи основні критерії оцінок моделей можемо розробити код для пайплайну тестування моделей натренованих з урахуванням усіх попередніх порад побудови моделей.

#### 4.2 Тестування розроблених моделей

Опираючись на вищеописані методи тестування медичних моделей машинного навчання розробимо власну функцію для тестування. Функція оцінки працює таким чином: спочатку вона бере справжні мітки з тестової вибірки і порівнює їх з прогнозами, згенерованими навченою нейронною мережею. Модель видає розподіл ймовірностей по всіх класах, і ці ймовірності перетворюються в дискретні прогнозовані мітки за допомогою операції `argmax`. Маючи в розпорядженні `y_true` та `y_pred`, функція обчислює основні показники ефективності, такі як точність, і створює звіт про класифікацію, що включає точність, відтворюваність та F1-оцінку для кожної категорії захворювань. Це дає загальне уявлення про те, наскільки добре модель розрізняє AMD, катаракту, діабетичну ретинопатію, глаукому, міопію та нормальні зображення очного дна.

Основна діагностична інформація походить з матриці помилок. Використовуючи записи цієї матриці, функція обчислює чутливість та специфічність для кожної категорії. Чутливість вимірює, наскільки ефективно модель ідентифікує випадки конкретного захворювання (відсоток справжніх позитивних результатів), а специфічність вимірює, наскільки добре вона уникає помилкової класифікації інших захворювань як цього конкретного захворювання (відсоток справжніх негативних результатів). Загалом, функція забезпечує комплексний і клінічно значущий процес оцінки, адаптований для багатокласової класифікації медичних зображень, її роботу можна побачити на рисунку 4.2.



Рисунок 4.2 – Вивід функції при різних моделях з високою та низькою точністю моделей

Розбираючи отримані візуалізації та метрики можна побачити, що модель з першого набору скрінів має дуже високу продуктивність для виявлення діабетичної ретинопатії та має найбільші труднощі з глаукомою. Клас глаукоми моделлю часто інтерпретується як катаракта або норма, ще означає, що модель може потребувати фін-тюнінгу на цьому класі з завантаженням більшої кількості зображень або переробкою попередньої обробки, через можливий занадто темний вигляд певних важливих ділянок на фотографіях, оскільки за показником чутливості модель вірно розпізнає глаукому лише в 50% випадків.

Високі показники чутливості, специфічності та інших оціночних показників зазвичай досягаються завдяки збалансованому та добре підготовленому набору даних у поєднанні з моделлю, яка була навчена з правильними налаштуваннями. Коли дані для навчання та валідації точно відображають візуальну мінливість, що спостерігається на реальних зображеннях очного дна, модель ефективніше вивчає специфічні для захворювання закономірності.

Погані результати часто виникають через такі проблеми, як дисбаланс класів, неправильні мітки класів або обмежену кількість прикладів навчання для певних захворювань. Якщо захворювання, зустрічається в наборі даних набагато рідше, модель може розвинути баєс щодо прогнозування більшості класів, що призведе до низької чутливості до рідкісних класів. Неправильно анотовані зображення, можуть і знизити як чутливість, так і специфічність. Крім того, нестабільна якість зображень, різниця в яскравості або ненормалізовані колірні канали можуть спричинити небажані відхилення, які мережа інтерпретує як значущі, що призводить до поганої продуктивності. Разом ці виправлення допомагають підвищити продуктивність за всіма показниками і створити модель, яка поводить себе більш послідовно в оцінках клінічного типу

## ВИСНОВКИ

В кваліфікаційні роботі було показано, як архітектури глибокого навчання можуть забезпечити надійний і клінічно значущий аналіз зображень очного дна за умови попередньої обробки, добре підібраних наборів даних та новітніх методів навчання з використанням та комбінацією різних інструментів і методичної оцінки. Основні результати включають розробку рекомендацій для повного навчання та валідації моделі, порівняльний аналіз декількох найсучасніших моделей для трансферного навчання при роботі з обмеженими наборами медичних зображень. У сукупності ці результати показують, що автоматизований аналіз зображень сітківки може досягти високої точності, надійності та стабільності при різних захворюваннях, таких як міопія, глаукома, катаракта, діабетична ретинопатія та вікова дегенерація макули.

У практичній офтальмології рекомендації, розроблені в цьому проєкті, можуть бути фундаментом для побудови допоміжних діагностичних інструментів, які допомагають медикам у ранньому виявленні та моніторингу аномалій сітківки. Забезпечуючи послідовну та швидку інтерпретацію зображень, такі системи можуть допомогти скоротити затримки в діагностиці, особливо в умовах, де бракує фахівців. Описані методи можуть бути основою для впровадження інструментів автоматизованого аналізу очного дна в лікарнях, та при масштабних ініціативах зі скринінгу в малорозвинених країнах.

В роботі було описано та проаналізовано предметну область фотографій очного дна, розказано про компанії які займаються розробкою та використанням подібних моделей та посавлено задачі для виконання. Описано процес збору та обробки даних з різних джерел з метою створення зваженого стандартизованого та оформленого датасету. Описано можливі інструменти для створення якісних моделей для роботи з візуальними даними та трансферним навчанням та процес тестування їх якості.

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ**

1. Muchuchuti S., Viriri S. Retinal disease detection using deep learning techniques: a comprehensive review. *Journal of imaging*. 2023. Vol. 9, no. 4. P. 84. URL: <https://doi.org/10.3390/jimaging9040084> (date of access: 22.11.2025).
2. World report on vision. *World Health Organization (WHO)*. URL: <https://www.who.int/publications/world-report-on-vision> (date of access: 14.11.2025).
3. Filatov V., Semenets V., Zolotukhin O. Synthesis of semantic model of subject area at integration of relational databases. *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*, Sozopol, Bulgaria, 6–8 September 2019. URL: <https://doi.org/10.1109/caol46282.2019.9019532> (date of access: 14.11.2025).
4. The ISNT rule: how often does it apply to disc photographs and retinal nerve fiber layer measurements in the normal population? / L. Y.-C. Poon et al. *American journal of ophthalmology*. 2017. Vol. 184. P. 19–27. URL: <https://doi.org/10.1016/j.ajo.2017.09.018> (date of access: 17.11.2025).
5. Fundoscopic appearances of retinal pathologies. <https://geekymedics.com>. URL: <https://geekymedics.com/fundoscopic-appearances-of-retinal-pathologies/> (date of access: 15.11.2025).
6. Artificial intelligence and diabetic retinopathy: AI framework, prospective studies, head-to-head validation, and cost-effectiveness / A. E. Rajesh et al. *Diabetes care*. 2023. Vol. 46, no. 10. P. 1728–1739. URL: <https://doi.org/10.2337/dci23-0032> (date of access: 16.11.2025).
7. Grzybowski A., Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: idx-dr and retinalyze. *Journal of clinical medicine*. 2021. Vol. 10, no. 11. P. 2352. URL: <https://doi.org/10.3390/jcm10112352> (date of access: 20.11.2025).

8. A multicenter clinical study of the automated fundus screening algorithm / F. Li et al. *Translational vision science & technology*. 2022. Vol. 11, no. 7. P. 22. URL: <https://doi.org/10.1167/tvst.11.7.22> (date of access: 20.11.2025).

9. Performance of a deep learning diabetic retinopathy algorithm in India / A. Brant et al. *JAMA network open*. 2025. Vol. 8, no. 3. P. e250984. URL: <https://doi.org/10.1001/jamanetworkopen.2025.0984> (date of access: 20.11.2025).

10. Intelligent automation systems / Avrunin O. et al. *Novabook*. 2021. (date of access: 24.11.2025).

11. Alwazzan M. J., Ismael M. A., Ahmed A. N. A hybrid algorithm to enhance colour retinal fundus images using a wiener filter and CLAHE. *Journal of digital imaging*. 2021. URL: <https://doi.org/10.1007/s10278-021-00447-0> (date of access: 24.11.2025).

12. Deepfakes in ophthalmology / J. S. Chen et al. *Ophthalmology science*. 2021. Vol. 1, no. 4. P. 100079. URL: <https://doi.org/10.1016/j.xops.2021.100079> (date of access: 25.11.2025).

13. DiffuPT: class imbalance mitigation for glaucoma detection via diffusion based generation and model pretraining / Y. Nawar et al. 2025 *IEEE/CVF winter conference on applications of computer vision (WACV)*, Tucson, AZ, USA, 26 February – 6 March 2025. 2025. P. 4098–4107. URL: <https://doi.org/10.1109/wacv61041.2025.00403> (date of access: 25.11.2025).

14. Blagus R., Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC bioinformatics*. 2013. Vol. 14, no. 1. URL: <https://doi.org/10.1186/1471-2105-14-106> (date of access: 26.11.2025).

15. Kang H., Vu T., Yoo C. D. Learning imbalanced datasets with maximum margin loss. *2021 IEEE international conference on image processing (ICIP)*, Anchorage, AK, USA, 19–22 September 2021. 2021.

URL: <https://doi.org/10.1109/icip42928.2021.9506389> (date of access: 26.11.2025).

16. Focal loss for dense object detection / T.-Y. Lin et al. *IEEE transactions on pattern analysis and machine intelligence*. 2020. Vol. 42, no. 2. P. 318–327. URL: <https://doi.org/10.1109/tpami.2018.2858826> (date of access: 27.11.2025).

17. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation M. Yeung et al. *Computerized medical imaging and graphics*. 2022. Vol. 95. P. 102026. URL: <https://doi.org/10.1016/j.compmedimag.2021.102026> (date of access: 27.11.2025).

18. Alwazzan M. J., Ismael M. A., Ahmed A. N. A hybrid algorithm to enhance colour retinal fundus images using a wiener filter and CLAHE. *Journal of digital imaging*. 2021. URL: <https://doi.org/10.1007/s10278-021-00447-0> (date of access: 28.11.2025).

19. What is random forest?. *Iguazio*. URL: <https://www.iguazio.com/glossary/what-is-random-forest/> (date of access: 28.11.2025).

20. Karamehić S., Jukić S. Brain tumor detection and classification using VGG16 deep learning algorithm and python imaging library. *Bioengineering studies*. 2023. Vol. 4, no. 2. P. 1–13. URL: <https://doi.org/10.37868/bes.v4i2.id252> (date of access: 28.11.2025).

21. Performance evaluation of densenet121 and resnet50 with clahe-based preprocessing for diabetic retinopathy detection | Spectrum of Engineering Sciences. *Spectrum of Engineering Sciences*. URL: <https://thesesjournal.com/index.php/1/article/view/980> (date of access: 28.11.2025).

22. Kolhe P. B., Shelke Ramesh D., Agarwal N. EnhanceNet: rethinking model scaling for convolutional neural network. *Lecture notes in networks and*

*systems*. Singapore, 2025. P. 1–15. URL: [https://doi.org/10.1007/978-981-96-1747-0\\_1](https://doi.org/10.1007/978-981-96-1747-0_1) (date of access: 28.11.2025).

23. Mukhlif A. A., Al-Khateeb B., Mohammed M. A. Incorporating a novel dual transfer learning approach for medical images. *Sensors*. 2023. Vol. 23, no. 2. P. 570. URL: <https://doi.org/10.3390/s23020570> (date of access: 29.11.2025).

24. A method for ocular disease diagnosis through visual prediction explainability / A. Santone et al. *Electronics*. 2024. Vol. 13, no. 14. P. 2706. URL: <https://doi.org/10.3390/electronics13142706> (date of access: 29.11.2025).

25. Kobzev V., Semenets V., Filatov V. Components of the information system for monitoring the quality of education in Kharkov National University of Radio Electronics *7th Int. scientific and technical conf. Information systems and technologies (IST-2018)*. Kharkiv-Koblevo. 2018. P. 51–54 (date of access: 29.11.2025).

26. Pedigo M. Sensitivity and specificity: a complete guide. URL: <https://www.datacamp.com/tutorial/sensitivity-specificity-complete-guide> (date of access: 29.11.2025).