Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут комп'ютерних наук та штучного інтелекту
Кафедра математичного моделювання та аналізу даних

До захисту допущено
Кафедрою ММтАД протокол
№___ від_____ завідувач
кафедри _____ Володимир Струков
(підпис) (ім'я, прізвище)
«___»_____ 2026 р.

Кваліфікаційна робота
здобувача першого (бакалаврського) рівня вищої освіти

Оцінка біометричних показників людини з використанням глибокого навчання

(назва роботи)

Спеціальність (спеціалізація) 122 Комп'ютерні науки

Освітня програма Комп'ютерні науки

Виконавець _____ Керімов Фарід
(підпис) (ім'я, прізвище)

Науковий керівник _____ *Миколенко О.П*
(підпис) (ім'я, прізвище)

2026 рік

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський національний університет імені В. Н. Каразіна

Навчально-науковий інститут комп'ютерних наук та штучного інтелекту
Кафедра Математичного моделювання та аналізу даних
Рівень вищої освіти (освітньо-кваліфікаційний рівень) Бакалавр
Галузь знань: F – Інформаційні технології
Спеціальність: F3 – Комп'ютерні науки
Освітня програма: _Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри _ММ та АД_
Струков Володимир Михайлович
«___» _____року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

_Керімов Фарід Анар Оглу _____
(прізвище, ім'я, по батькові студента)

1. Тема роботи: Оцінка біометричних показників людини з використанням глибокого навчання_ керівник роботи Миколенко О.П., к.е.н., доцент_____,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від "_29_"__січня__2026_року №

2. Строк подання студентом роботи01 червня 2026 року

3. Перелік питань, які потрібно розробити:

1. проаналізувати предметну область біометричної ідентифікації та класифікації людини за цифровими зображеннями обличчя;
2. розглянути сучасні підходи комп’ютерного зору та машинного навчання для розпізнавання біометричних характеристик людини;
3. проаналізувати класичні алгоритми аналізу облич (Haar Cascades, HOG, LBP) та сучасні нейромережеві підходи;
4. обґрунтувати вибір набору даних UTKFace для проведення експериментальних досліджень;
5. розробити процедуру підготовки та попередньої обробки цифрових зображень обличчя для навчання нейронних мереж;
6. забезпечити коректне розділення набору даних на навчальну, валідаційну та тестову вибірки;
7. реалізувати етапи попередньої обробки даних, включаючи зміну розміру зображень, нормалізацію та аугментацію;
8. спроєктувати та реалізувати базові архітектури штучних нейронних мереж (ANN та CNN) для задачі класифікації статі людини;
9. побудувати та навчити моделі глибокого навчання на основі технології перенесення знань (ResNet50 та MobileNetV2);
10. провести експериментальне порівняння розроблених моделей, проаналізувати отримані результати, оцінити точність класифікації та сформулювати висновки щодо практичного застосування створеної системи.

План роботи

№ з/п	Назва етапів роботи	Термін виконання етапів роботи
1	Визначення проблематики та аналіз існуючих методів оцінки біометричних показників.	05.05 - 08.05
2	Дослідження архітектур та принципів роботи згорткових нейронних мереж.	09.05 - 13.05
3	Аналіз, парсинг та фільтрація обраного набору даних зображень (UTKFace).	14.05 - 18.05
4	Дослідження та реалізація методів попередньої обробки: зміна розміру, нормалізація, аугментація.	19.05 - 21.05

5	Проєктування базових архітектур нейронних мереж (Fully Connected ANN, Simple CNN).	22.05 - 25.05
6	Навчання базових моделей та первинний аналіз динаміки метрик.	26.05 - 27.05
7	Обґрунтування вибору та інтеграція переднавчених моделей (ResNet50, MobileNetV2).	28.05 - 29.05
8	Проведення перенесення навчання (transfer learning) та тонкого налаштування (fine-tuning).	30.05 - 31.05
№ з/п	Назва етапів роботи	Термін виконання етапів роботи
9	Проведення фінального тестування та порівняльний аналіз метрик усіх моделей.	01.06 - 01.06
10	Детальний аналіз помилок класифікації, підбиття підсумків та визначення напрямків покращення.	02.06 - 03.06
11	Оформлення пояснювальної записки та підготовка презентації на передзахист.	03.06 - 03.06

4. Дата видачі завдання_____

Студент Керімов Фарід_ 

ініціали, прізвище підпис

Керівник роботи Миколенко О.П. 

ініціали, прізвище підпис

АНОТАЦІЯ

Кваліфікаційна робота присвячена дослідженню, програмній реалізації та експериментальному оцінюванню методів машинного навчання для визначення біометричного профілю людини за цифровими зображеннями обличчя. У роботі розглянуто задачу автоматизованої класифікації статі людини на основі аналізу фотографій обличчя із застосуванням сучасних методів комп'ютерного зору та глибокого навчання.

Метою роботи є підвищення точності автоматизованого визначення статі людини за цифровими зображеннями обличчя шляхом розробки та експериментального дослідження моделей штучних нейронних мереж. Об'єктом дослідження є процес автоматизованого аналізу біометричних характеристик людини за цифровими зображеннями. Предметом дослідження є методи попередньої обробки зображень та алгоритми машинного і глибокого навчання для класифікації статі людини.

У роботі використано відкритий набір даних UTKFace, що містить фотографії людей різного віку, статі та етнічного походження. Для проведення експериментів було виконано попередню обробку зображень, яка включала зміну розміру, нормалізацію піксельних значень та аугментацію даних. Набір даних було поділено на навчальну, валідаційну та тестову вибірки для забезпечення об'єктивного оцінювання якості моделей.

У межах роботи реалізовано та досліджено декілька підходів до класифікації статі людини. Як базові моделі використано повнозв'язну штучну нейронну мережу (ANN) та просту згорткову нейронну мережу (CNN). Для підвищення точності класифікації застосовано технологію перенесення навчання на основі архітектур ResNet50 та MobileNetV2.

Проведене експериментальне дослідження показало, що найкращі результати продемонструвала модель ResNet50, яка забезпечила точність класифікації близько 94%. Модель MobileNetV2 показала дещо нижчу

точність, проте продемонструвала значно меншу кількість параметрів та вищу ефективність використання обчислювальних ресурсів. Базові моделі ANN та CNN поступилися архітектурам перенесення навчання за основними показниками якості.

Практичне значення роботи полягає у створенні програмної системи автоматизованого визначення статі людини за цифровими зображеннями обличчя, яка може бути використана в системах біометричної ідентифікації, контролю доступу, відеоспостереження та інтелектуального аналізу даних. Отримані результати підтверджують ефективність застосування сучасних методів глибокого навчання для розв’язання задач аналізу біометричних характеристик людини.

КЛЮЧОВІ СЛОВА: БІОМЕТРИЧНИЙ ПРОФІЛЬ, КЛАСИФІКАЦІЯ СТАТІ, КОМП’ЮТЕРНИЙ ЗІР, МАШИННЕ НАВЧАННЯ, ГЛИБОКЕ НАВЧАННЯ, НЕЙРОННІ МЕРЕЖІ, CNN, ANN, RESNET50, MOBILENETV2, TRANSFER LEARNING, UTKFACE.

ABSTRACT

The qualification thesis is devoted to the research, software implementation, and experimental evaluation of machine learning methods for determining a human biometric profile based on facial image analysis. The study focuses on the task of automatic gender classification using digital face images and modern computer vision techniques.

The purpose of the thesis is to improve the accuracy of automated gender recognition from facial images by developing and experimentally evaluating artificial neural network models. The object of the research is the process of automated analysis of human biometric characteristics using digital images. The subject of the research is image preprocessing methods and machine learning algorithms for gender classification.

The study uses the publicly available UTKFace dataset, which contains facial images of people of different ages, genders, and ethnic backgrounds. During the experimental stage, image preprocessing techniques were applied, including resizing, normalization, and data augmentation. The dataset was divided into training, validation, and testing subsets to ensure an objective evaluation of model performance.

Several approaches to gender classification were implemented and analyzed. Basic models included an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN). To improve classification accuracy, transfer learning techniques based on the ResNet50 and MobileNetV2 architectures were employed.

Experimental results demonstrated that transfer learning models significantly outperformed the baseline approaches. The ResNet50 architecture achieved the highest classification accuracy, while MobileNetV2 provided a favorable balance between recognition performance and computational efficiency. The obtained results confirm the effectiveness of deep learning methods for solving biometric classification tasks based on facial image analysis.

The practical value of the thesis lies in the development of a software solution for automatic gender recognition from facial images. The proposed approach can be applied in biometric identification systems, access control solutions, intelligent surveillance systems, and other computer vision applications. The obtained results demonstrate the potential of modern neural network architectures for the analysis of human biometric characteristics.

KEYWORDS: BIOMETRIC PROFILE, GENDER CLASSIFICATION, COMPUTER VISION, MACHINE LEARNING, DEEP LEARNING, ARTIFICIAL NEURAL NETWORK, CONVOLUTIONAL NEURAL NETWORK, CNN, ANN, RESNET50, MOBILENETV2, TRANSFER LEARNING, UTKFACE.

ЗМІСТ

Вступ.....10
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ВИБІР МЕТОДІВ 11
1.1 Огляд задач оцінки біометричних показників 11
1.1.1 Ретроспективний аналіз класичних систем та еталонних моделей у задачах біометричного аналізу 13
1.2 Аналіз методів глибокого навчання для обробки зображень 16
(згорткові нейромережі) 16
1.3 Еволюція підходів до аналізу зображень: від класичного машинного навчання до згорткових архітектур 17
1.4 Постановка задачі на дослідження 25
Висновки до озділу 1 26
РОЗДІЛ 2. ПІДГОТОВКА ДАНИХ ТА ПРОЄКТУВАННЯ МОДЕЛЕЙ 28
2.1 Анализ и предварительная обработка выбранного набора данных (аугментация, нормализация) 28
2.1.2 Зміна просторового розширення та масштабуваня 30
2.1.3 Математична нормалізація даних 31
2.1.4 Стратифікація та формування вибірок 31
2.1.5 Методи аугментації даних 32
2.2 Проєктування базової архітектури нейронної мережі (Baseline) 33
2.2.1 Повнозв'язна штучна нейронна мережа (Fully Connected ANN) 33
2.2.2 Проста згорткова нейронна мережа (Simple CNN) 36
2.3 Використання технологій перенесення навчання на базі існуючих архітектур 39
2.3.1 Інтеграція архітектури ResNet50 39
2.3.2 Інтеграція архітектури MobileNetV2 41
2.4 Вибір метрик для оцінки ефективності моделе.....42
2.4.1 Цільова функція втрат 42
Висновки до розділу 2 44
РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ 46

3.1 Опис апаратно-програмного середовища та конфігурації гіперпараметрів 46
3.2 Навчання базових архітектур та первинний аналіз динаміки метрик ... 47
3.3 Адаптація та навчання глибоких архітектур перенесення знань 51
3.4 Порівняльний аналіз та технічна оцінка ефективності алгоритмів 53
3.5 Детальний аналіз помилок класифікації алгоритмів та вплив параметрів зображень 56
Висновки до розділу 3 58
ВИСНОВКИ 60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 62

СПИСОК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ

ANN (Artificial Neural Network) - штучна нейронна мережа.

CNN (Convolutional Neural Network) - згорткова нейронна мережа.

LBP (Local Binary Patterns) - локальні бінарні шаблони.

HOG (Histogram of Oriented Gradients) - гістограма орієнтованих градієнтів.

SVM (Support Vector Machine) - метод опорних векторів.

PCA (Principal Component Analysis) - метод головних компонент.

LDA (Linear Discriminant Analysis) - лінійний дискримінантний аналіз.

GPU (Graphics Processing Unit) - графічний процесор.

CPU (Central Processing Unit) - центральний процесор.

API (Application Programming Interface) - прикладний програмний інтерфейс.

RGB (Red, Green, Blue) - колірна модель представлення цифрових зображень.

MLP (Multilayer Perceptron) - багатошаровий перцептрон.

ReLU (Rectified Linear Unit) - функція активації нейронної мережі.

KLT (Kanade–Lucas–Tomasi) - алгоритм відстеження характерних точок на зображенні.

FACS (Facial Action Coding System) - система кодування рухів м’язів обличчя.

UTKFace - відкритий набір зображень обличчя для задач комп’ютерного зору та машинного навчання.

TensorFlow - програмний фреймворк для створення та навчання моделей глибокого навчання.

Keras - високорівневий програмний інтерфейс для розробки нейронних мереж.

OpenCV (Open Source Computer Vision Library) - бібліотека алгоритмів комп’ютерного зору.

ВСТУП

У роботі проведено комплексне дослідження методів автоматичного визначення біометричних характеристик людини (зокрема статі) за цифровими фотографіями обличчя з використанням методів глибокого навчання.

Об'єктом дослідження є процеси аналізу зображень обличчя та визначення біометричних характеристик людини за допомогою методів штучного інтелекту. Предметом дослідження є згорткові нейронні мережі (cnn) методи глибокого навчання та алгоритми перенесення навчання (transer learning) для задачі бінарної класифікації зображень.

У першому розділі проаналізовано предметну область, оглянуто класичні підходи та переваги згорткових архітектур. У другому розділі описано процес підготовки набору даних utkace, включаючи методи аугментації та нормалізації, а також спроєктовано архітектури моделей: від базового персептрона до адаптованих моделей resnet50 та mobilenetv2. Третій розділ присвячено практичній реалізації, де шляхом порівняльного аналізу доведено високу ефективність технології перенесення навчання.

Найкращі результати продемонструвала модель на базі resnet50 із застосуванням тонкого налаштування, досягнувши точності 94,2% на тестовій вибірці. Доведено, що для комерційних рішень на мобільних пристроях найбільш збалансованою є архітектура mobilenett2 (точність 92,1%).

Ключові слова: комп'ютерний зір, згорткові нейронні мережі, глибоке навчання, біометричні показники, класифікація статі, перенесення навчання, resnet50 mobilenetv2 utface.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ВИБІР МЕТОДІВ

1.1 Огляд задач оцінки біометричних показників

Оцінка біометричних показників людини за обличчям є одним із ключових напрямів сучасного комп’ютерного зору та штучного інтелекту. Обличчя містить багато візуальної інформації, яку можна аналізувати за допомогою цифрових камер без фізичного контакту. Тому такі технології широко застосовують у безпеці, відеоспостереженні, контролі доступу, маркетингу та різних інтелектуальних сервісах [Almabdy, Deep Convolutional Neural Network-Based Approaches for Face Recognition, 2019].

Сучасні системи аналізу обличчя можуть виконувати кілька задач одночасно. Найчастіше це визначення статі, оцінка віку та розпізнавання емоцій. Кожна з цих задач має свої особливості та складності при розробці програмного забезпечення.

Визначення статі (Gender Classification) це задача бінарної класифікації. Її мета визначити, чиє обличчя на зображенні: чоловіка чи жінки. Такі системи часто використовують у цифровій рекламі та аналітичних сервісах, де контент змінюється залежно від аудиторії. Хоча людині ця задача здається простою, алгоритми можуть помилятися через низьку якість фото, освітлення, макіяж, окуляри або зміну зачіски. Також на точність впливають положення голови та вираз обличчя

Ще одним важливим напрямком є оцінка віку людини (age estimation). Ця задача складніша, бо процес старіння у всіх різний. Для її вирішення часто використовують регресію або багатокласову класифікацію, де вік поділяють на категорії чи діапазони. На результат впливають генетика, спосіб життя, борода, косметика, окуляри та якість освітлення. Тому навіть сучасні моделі

глибокого навчання не завжди точно визначають реальний вік [He et al., Deep Residual Learning for Image Recognition, 2016].

Ще однією важливою задачею є розпізнавання емоцій (Emotion Recognition). Тут алгоритм аналізує міміку обличчя, щоб визначити емоційний стан людини. Зазвичай системи розпізнають базові емоції: радість, сум, здивування, страх, злість і нейтральний стан. Класифікація спирається на психологічні моделі експресії обличчя [Ekman, An Argument for Basic Emotions, 1992]. Глибокі нейронні мережі аналізують дрібні зміни м’язів і мікровирази, що дає високу точність автоматичного розпізнавання [Li, Deep Facial Expression Recognition: A Survey, 2020]. Такі технології використовують у маркетингових дослідженнях, системах підтримки клієнтів, освітніх платформах та інтерактивних сервісах. Основні характеристики задач аналізу біометричних показників наведені у табл1.1.

Отже задачі визначення віку, статі та емоцій людини є важливими напрямками сучасного комп’ютерного зору. Попри розвиток нейронних мереж і глибокого навчання, точність таких систем все ще залежить від якості зображення, освітлення та індивідуальних особливостей людини.

Таблиця 1.1 - Порівняння основних задач аналізу обличчя

Задача	Тип машинного	Головна складність для алгоритму
Визначення ст	Бінарна класиф	Макіяж освітлення нетипові зачіски
Оцінка віку	Регресія / Багат класифікація	Індивідуальні особливості старіння людини
Розпізнавання емоцій	Багатокласова класифікація	

Схожістьмікровиразівдля різних емоцій

З наведених у табл. 1.1 даних видно, що різні задачі аналізу обличчя мають власні особливості та потребують застосування різних підходів машинного навчання. Визначення статі належить до задач бінарної класифікації та характеризується відносно невеликою кількістю класів, проте на точність розпізнавання можуть впливати освітлення, макіяж і особливості зовнішності людини. Оцінка віку та розпізнавання емоцій є більш складними задачами через значну варіативність зовнішніх ознак і суб'єктивність проявів. З огляду на це задача визначення статі є доцільним об'єктом дослідження для оцінювання ефективності сучасних методів машинного та глибокого навчання.

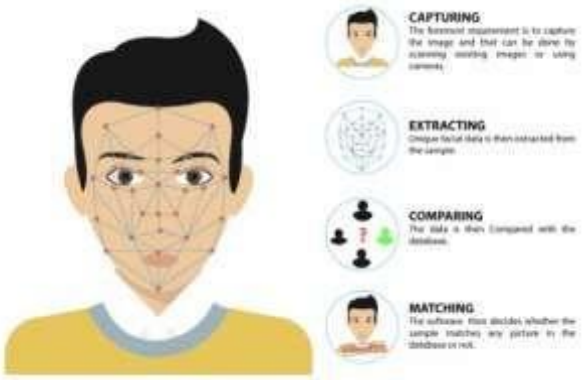


Рисунок 1.1 – Приклад роботи системи оцінки біометричних показників

Раніше ці задачі намагалися вирішувати виключно математичними методами, вимірюючи відстані між очима, носом та губами. Однак цей підхід виявився неефективним у випадках коли людина повертає голову. Сучасні

розробники використовують нейронні мережі, які автоматично виявляють приховані ознаки на рівні пікселів, забезпечуючи значно кращу та швидшу роботу систем

1.1.1 Ретроспективний аналіз класичних систем та еталонних моделей у задачах біометричного аналізу евклідової відстані між вектором проекції нового зображення та векторами відомих еталонів. Ця система довела можливість автоматизації аналізу зображень, проте демонструвала високу чутливість до змін освітлення та масштабних деформацій.

Для подолання дефектів чутливості до світла П. Белхюмером,

Д. Крігманом та А. Пулэ було спроєктовано систему Fisherfaces [Belhumeur et al.,

Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, 1997]. У цій роботі для вирішення задачі класифікації облич умов критично нерівномірного освітлення та варіативності міміки було застосовано модель лінійного дискримінантного аналізу (Linear Discriminant Analysis, LDA). На відміну від PCA, модель LDA максимізувала відношення міжкласової дисперсії до внутрішньокласової, що дозволило системі ігнорувати зовнішні збурення (наприклад, тіні) та фокусуватися виключно на сталих індивідуальних біометричних патернах обличчя.

У задачах бінарної класифікації статі та експрес-оцінки вікових груп на неструктурованих графічних даних («in-the-wild») класичним проривом стала система, розроблена Г. Леві та Т. Хасснером [Levi & Hassner, Age and Gender Classification using Convolutional Neural Networks, 2015]. Автори запропонували компактну модель згорткової нейронної мережі (CNN), що складалася з трьох згорткових шарів (із послідовним застосуванням шарів максимізації Max Pooling та нормалізації Local Response Normalization) та двох повнозв'язних шарів. Для вирішення задачі визначення статі модель використовувала фінальний шар Softmax з двома виходами, а для оцінки віку

- класифікатор на 8 фіксованих вікових інтервалів. Ця праця dovela, що навіть відносно неглибокі нейромережеві архітектури за умови коректної регуляризації (зокрема, використання Dropout шарів) здатні демонструвати високу точність обробки зображень із низькою роздільною здатністю та складним фоном.

Для точнішої, безперервної оцінки біологічного віку людини вагомим внеском стала система dex (deep expectation), яка здобула перемогу на престижному міжнародному конкурсі CHALEARN LAP [Rothe et al., Dex: Deep expectation of apparent age from a single face image, 2015]. Для вирішення задачі визначення точного віку дослідники використали надглибоку архітектуру моделі vgg-16, попередньо навчену на мільйонному наборі даних imagenet, із подальшим точним налаштуванням (fine-tuning) на спеціалізованих біометричних вибірках. Особливість математичної моделі полягала в тому, що задача регресії (обчислення точного числа років) була переформатована в задачу класифікації на 101 дискретний клас (від 0 до 100 років) фінальне значення віку обчислювалося не як жорсткий аргумент максимуму а як математичне очікування за формулою

$$E = \sum_{i=0}^{100} i \cdot p_i$$

де pi- ймовірність, згенерована шаром Softmax для кожного конкретного року i. Такий підхід дозволив системі DEX досягти рекордних показників точності та мінімізувати середню абсолютну похибку (MAE) аналізу.

У галузі розпізнавання емоційного стану та кодування мімічних рухів еталонним інженерним рішенням вважається автоматизована система, розроблена Й. Тіаном, Т. Канаде та Д. Коном [Tian, Kanade, & Cohn, Recognizing Action Units for Facial Expression Analysis, 2001]. Для вирішення задачі розпізнавання дрібних мікровиразів автори розробили модель на основі

штучної нейронної мережі (багатошарового перцептрона, MLP), вхідними даними для якої виступали геометричні вектори зсуву опорних точок обличчя (контурів очей, брів, рота), що відстежувалися в динаміці за допомогою алгоритму Канаде-Лукаса-Томасі (KLT). Модель успішно здійснювала детектування окремих рухових одиниць (Action Units) відповідно до міжнародної системи кодування лицьових рухів (FACS) Пола Екмана, що дозволило комп'ютерним системам аналізувати не лише базові афекти, а й складні комбінації прихованих людських емоцій.

1.2 Аналіз методів глибокого навчання для обробки зображень (згорткові нейромережі)

Коли мова йде про розпізнавання образів та обробку фотографій звичайні повнозв'язні нейронні мережі демонструють дуже низьку ефективність. Справа в тому що будь-яке цифрове зображення складається з великої кількості пікселів. Якщо ми подамо всі ці пікселі на вхід класичної мережі то кількість зв'язків та параметрів для обчислення виросте в геометричній прогресії. Це призводить до того що комп'ютеру просто не вистачає потужності для навчання а сама модель починає перенавчатися тобто просто зазубрює тренувальні картинки і не може працювати з новими кадрами

Для вирішення цієї проблеми були створені згорткові нейронні мережі (Convolutional Neural Networks або CNN) які на сьогодні є головним стандартом у сфері комп'ютерного зору. Основна ідея згорткових мереж полягає в тому що вони враховують двовимірну структуру картинки і шукають локальні ознаки за допомогою спеціальних невеликих матриць - фільтрів

Архітектура стандартної мережі CNN будується з кількох типів шарів які чергуються між собою для поступового виділення ознак. Детальний опис цих шарів та їх функцій наведено у таблиці нижче

Таблиця 1.2 Основні компоненти архітектури згорткової нейромережі

Назва шару мережі	Що саме робить цей шар з картинкою
Згортковий (Convolutional layer)	Сканує зображення маленьким вікном (фільтром) та виділяє специфічні ознаки: лінії межі кути
Агрегаційний (Pooling layer)	Зменшує розмір матриці ознак зазвичай обираючи максимальне значення (Max Pooling). Це економить пам'ять
Повнозв'язний (Fully Connected)	Знаходиться в самому кінці мережі об'єднує всі знайдені ознаки та приймає остаточне рішення щодо класу

З наведених у табл.1.2 даних можна зробити висновок, що сучасні підходи до аналізу біометричних даних суттєво перевершують класичні алгоритми за показниками точності та здатності до узагальнення. Разом із тим підвищення якості розпізнавання супроводжується зростанням обчислювальної складності моделей та вимог до апаратних ресурсів. Отримані результати підтверджують доцільність використання нейронних мереж для задач автоматичного визначення біометричних характеристик людини.

1.3 Еволюція підходів до аналізу зображень: від класичного машинного навчання до згорткових архітектур

Упродовж тривалого періоду розвитку галузі комп’ютерного зору домінували підходи, що базувалися на методології класичного машинного навчання (machine learning). Фундаментом цих методів було послідовне розділення процесу обробки графічної інформації на два автономні етапи:

ручне проектування дескрипторів ознак (hand-crafted features) та їх подальша класифікація за допомогою стандартних математичних класифікаторів.

Одним із найбільш репрезентативних та історично значущих методів у цьому контексті став алгоритм віола джонса що набув масового поширення для ідентифікації облич у режимі реального часу [Viola, Jones, Robust RealTime Face Detection, 2001]. Цей підхід ґрунтується на застосуванні каскадів ознак Хаара (Haar Cascades), які функціонують як прямокутні цифрові маски, здатні обчислювати градієнт яскравості між суміжними піксельними зонами. Для суттєвого зниження обчислювального навантаження в архітектуру інтегровано механізм адаптивного бустингу (AdaBoost), що дозволяє ізолювати лише найінформативніші патерни з величезного масиву можливих ознак, формуючи з них каскадну класифікаційну структуру.

Згенеровані таким чином вектори ознак часто передавалися на вхід машинного опорного вектору (Support Vector Machine, SVM). Математична суть алгоритму SVM полягає в оптимізації розділяючої гіперплощини у багатовимірному просторі ознак таким чином, щоб забезпечити максимальний зазор (margin) між об’єктами різних класів. Окрім каскадів Хаара, у класичних системах аналізу біометричних даних активно використовувалися гістограми орієнтованих градієнтів (Histogram of Oriented Gradients, HOG) та локальні бінарні патерни (Local Binary Patterns, LBP), спрямовані на фіксацію текстурних характеристик і контурів лицьових структур. [Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection, 2005] [Ojala T. et al.
Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, 2002]

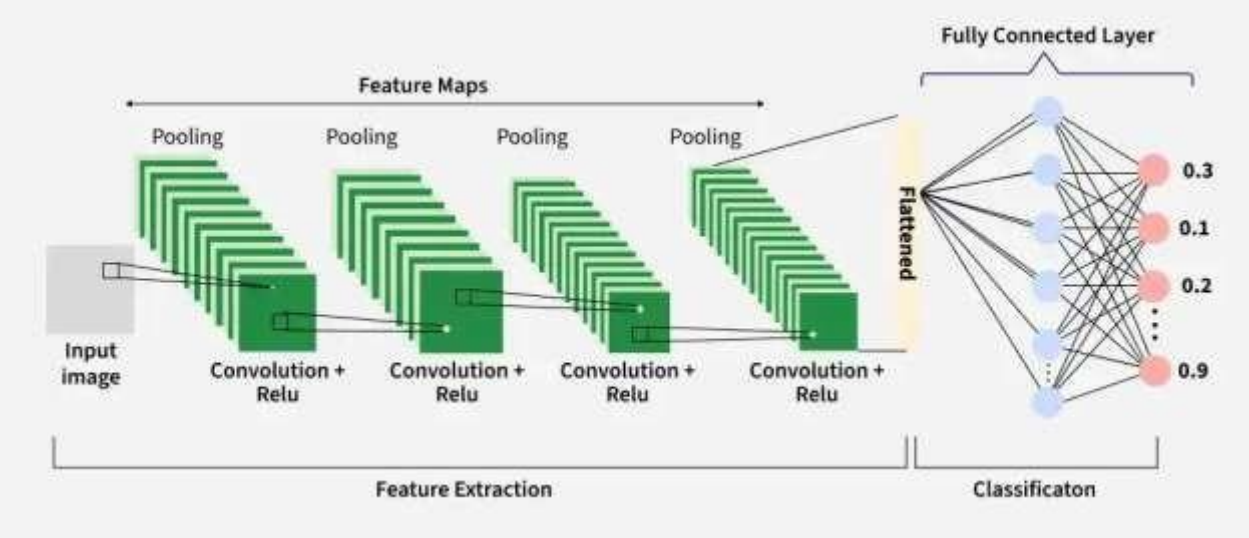


Рис. 1.2 – Структура повнозв'язної нейронної мережі

Проте незважаючи на високу швидкість функціонування, класичні підходи мають критичні недоліки. Головний недолік полягає в тому, що ручні ознаки не мають властивості інваріантності до масштабних просторових трансформацій. За умов мінімальної зміни ракурсу зйомки, появи нерівномірного освітлення, часткового перекриття обличчя (оклюзії) або вираженої мімічної деформації, фіксовані математичні фільтри втрачають здатність коректно ідентифікувати об’єкт. [Szeliski, Computer Vision: Algorithms and Applications, 2010]

Так, системи на основі Haar + SVM виявилися обмеженими і неспроможними до генералізації на неструктурованих реальних даних.

Концептуальним проривом у подоланні зазначених обмежень став перехід до парадигми штучних нейронних мереж (Artificial Neural Networks, ANN). Перші спроби застосування багатошарових перцептронів (MLP) до обробки зображень продемонстрували потенційну здатність моделей самостійно вивчати ієрархію ознак безпосередньо з сирих піксельних матриць за допомогою алгоритму зворотного поширення помилки (backpropagation). Однак використання стандартних повнозв'язних архітектур для аналізу двовимірних графічних об'єктів швидко зіткнулося з фундаментальною

проблемою, відомою як «прокляття розмірності» (curse of dimensionality). [Bellman, Dynamic Programming, 1957].

У повнозв'язній нейромережі кожен нейрон поточного шару з'єднаний абсолютно з усіма нейронами попереднього шару, що є класичним стандартом для архітектур типу багатошарового перцептрона [Haykin, Neural Networks: A Comprehensive Foundation, 2009]. Якщо подати на вхід такої моделі відносно невелике цифрове зображення розмірністю 128x128 пікселів із трьома колірними каналами (RGB), вхідний вектор матиме 49 152 елементи. Відповідно, лише перший прихований шар, що містить помірну кількість нейронів, потребуватиме розрахунку та збереження мільйонів вагових коефіцієнтов.

Такий стрімкий ріст параметрів призводить до двох негативних наслідків: надлишкового споживання обчислювальних ресурсів (пам’яті та часу процесора) і катастрофічного перенавчання моделі (overfitting), коли мережа втрачає здатність до генералізації та просто «зазубрює» тренувальні картини, виявляючись неспроможною працювати з новими кадрами [Goodfellow et al., Deep Learning, 2016]. Окрім цього, перед подачею в повнозв'язний шар матриця зображення обов'язково вирівнюється в одновимірний вектор, що повністю знищує локальну топологію та просторові зв'язки між суміжними пікселями. Графічно структуру такої мережі та критичну щільність її синаптичних зв'язків при спробі обробки графічних даних продемонстровано нижче.

Еволюційним розв'язанням цієї математичної кризи стало проектування згорткових нейронних мереж (Convolutional Neural Networks, CNN). Концептуальна архітектура CNN була натхненна біологічними дослідженнями зорової кори ссавців, проведеними Девідом Х’юбелом та Торстеном Візелом, які відкрили існування простих і складних клітин, що реагують на локалізовані світлові стимули. Впровадження згорткових

архітектур дозволило реалізувати три фундаментальні інженерні принципи: [Hubel, Wiesel, Receptive fields of single neurones in the cat's striate cortex, 1959]

1. Локальні рецептивні поля (local receptive fields) - нейрони шару згортки взаємодіють не з усім зображенням одразу, а лише з обмеженою просторовою областю (вікном згортки).

2. Спільні вагові коефіцієнти (shared weights) - один і той самий набір ваг (ядро згортки або фільтр) переміщується по всій площині зображення, що радикально зменшує загальну кількість параметрів мережі та забезпечує інваріантність до зсуву (shift invariance).

3. Просторова субдискретизація або пулінг (pooling) - операція, спрямована на зменшення геометричної розмірності карт ознак шляхом агрегації значень у локальних вікнах (наприклад, через вибір максимуму - Max Pooling), що знижує чутливість моделі до дрібних просторових спотворень.

Хронологія розвитку згорткових моделей демонструє стійку тенденцію до безперервного ускладнення топологічних зв'язків і оптимізації процесів екстракції ознак [LeCun et al., Deep learning, 2015]. Від перших відносно простих структур до надглибоких сучасних архітектур, кожна нова генерація пропонувала унікальні механізми вирішення проблем згасання градієнта та оптимізації обчислень. Наочну еволюцію ключових архітектур згорткових мереж відображено на відповідному рисунку.

Першою комерційно успішною архітектурою стала мережа LeNet-5, розроблена Янном Лекуном у 1998 році для автоматизованого розпізнавання рукописних символів. Вона продемонструвала класичний шаблон побудови CNN, заснований на послідовному чергуванні згорткових шарів, функцій активації Sigmoid та шарів субдискретизації (subsampling), з фінальним переходом до повнозв’язних шарів.

Нова ера штучного інтелекту розпочалася у 2012 році з появою архітектури flexnet, яка здобула тріумфальну перемогу на престижному конкурсі ImageNet. Ця модель суттєво масштабувала глибину та кількість фільтрів порівняно з LeNet-5. Проте ключовий внесок AlexNet полягав у впровадженні низки інноваційних технологічних рішень: застосуванні нелінійної функції активації ReLU (Rectified Linear Unit), яка дозволила прискорити збіжність градієнтного спуску та частково нівелювати проблему затухання градієнтів; інтеграції шарів Dropout для випадкового відключення нейронів з метою ефективної регуляризації та боротьби з перенавчанням; а також у використанні апаратного прискорення на базі графічних процесорів (GPU).

Наступним кроком у мінімізації помилок класифікації стала розробка архітектури vgg (Visual Geometry Group), представлена у двох варіаціях: vgg-16 та vgg-19. Автори цієї моделі відмовилися від використання великих ядер згортки (таких як 11x11 в alexnet) на користь послідовних блоків із маленькими фільтрами розміром 3x3. Математично доведено, що стек із двох шарів згортки з ядрами 3x3 має таке ж ефективне рецептивне поле, як один шар із ядром 5x5, але при цьому містить значно меншу кількість параметрів та включає більше лінійних нелінійностей, що суттєво підвищує здатність мережі вивчати складні високорівневі абстракції.

Проте нескінченне нарощування глибини мереж призвело до виникнення ефекту деградації (degradation problem), коли при досягненні певної критичної кількості шарів точність навчання починає стрімко падати, що не пов'язано з перенавчанням. Цю фундаменталну проблему успішно розвязала архітектура ResNet (residual network**)** у 2015 році. Інженери запропонували концепцію залишкового навчання (residual learning) шляхом упровадження наскрізних зв'язків (skip pathways або shortcut connections). Ці зв'язки дозволяють пропускати вхідний сигнал безпосередньо через один або

кілька шарів, додаючи вихідне значення x до згенерованого залішкового відображення F(x). Математична формула блоку має вигляд H(x) = F(x) + x. Завдяки такій структурі градієнти під час зворотного поширення можуть безперешкодно протікати через глибокі шари мережі, що дозволило успішно навчати ультраглибокі архітектури, які налічують 50, 101 або навіть 152 шари, демонструючи рекордно низькі показники похибок в задачах комп'ютерного зору.

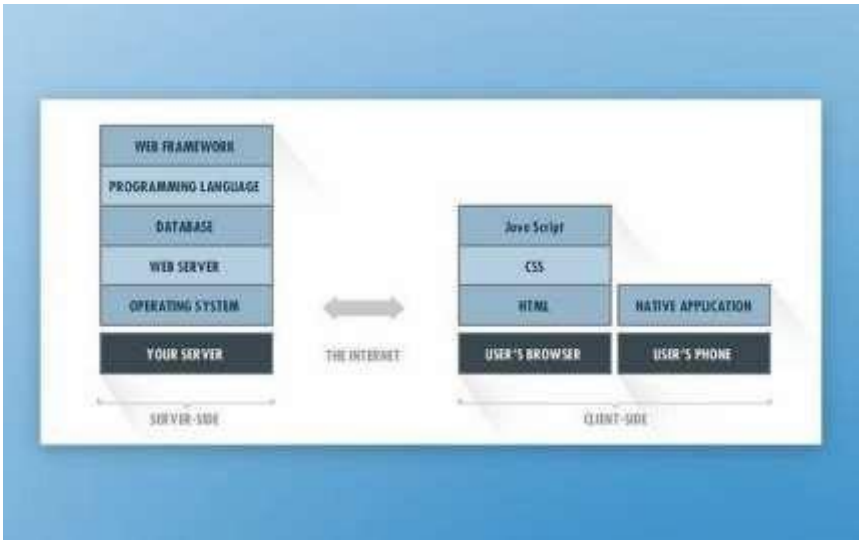


Рис. 1.3. Обґрунтування вибору технологічного стека (Python,TensorFlow/Keras)

Для успішної програмної реалізації спроєктованої системи комп'ютерного зору критично важливим етапом є вибір оптимального інструментарію, який забезпечуватиме необхідну обчислювальну потужність, гнучкість налаштувань та наявність розвиненої екосистеми. Як базову мову програмування було обрано Python. На сьогоднішній день Python де-факто є світовим стандартом у галузі машинного навчання та штучного інтелекту завдяки своєму лаконічному синтаксису та безпрецедентно широкому набору спеціалізованих бібліотек [Raschka, Python Machine Learning, 2019]. Наявність таких інструментів, як NumPy, Pandas та OpenCV дозволяє максимально

оптимізувати процеси попередньої обробки багатовимірних масивів графічних даних тензорі перед їх поданням на вхідний шар нейронної мережі.

Основним обчислювальним ядром для побудови та подальшого навчання згорткових архітектур було обрано відкриту програмну платформу TensorFlow, розроблену дослідницьким підрозділом Google Brain. Згідно з фундаментальними документами архітектури цієї платформи, TensorFlow використовує концепцію графів потоків даних (data flow graphs), що дозволяє надзвичайно ефективно розпаралелювати складні тензорні операції між кількома ядрами центрального процесора (CPU) або переносити їх на графічні прискорювачі (GPU) [Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2016]. Ця властивість є критично важливою для навчання глибоких моделей, де кількість математичних операцій під час алгоритму зворотного поширення помилки сягає мільярдів ітерацій.

Проте безпосереднє програмування низькорівневих графів у TensorFlow є надзвичайно трудомістким процесом, що ускладнює швидке проведення експериментів. Тому для опису топології нейронної мережі було інтегровано високорівневий прикладний програмний інтерфейс (API) Keras. Як зазначає сам творець цієї бібліотеки Франсуа Шолле, головна філософія Keras полягає в орієнтації на когнітивну зручність розробника, що дозволяє інженеру трансформувати теоретичні архітектурні ідеї у робочий програмний код з мінімальними синтаксичними затримками [Chollet, Deep Learning with Python, 2017]. Keras працює безпосередньо поверх базового бекенду TensorFlow і надає зручні абстракції для створення згорткових шарів, об'єднання (пулінгу), функцій активації та оптимізаторів, забезпечуючи ідеальний баланс між швидкістю прототипування та продуктивністю виконання.

Крім того, враховуючи високі апаратні вимоги до процесів тренування глибоких згорткових мереж (зокрема необхідність обробки

десятків тисяч зображень батчами) для практичної реалізації обрано хмарне середовище Google Colaboratory. Воно надає безкоштовний доступ до високопродуктивних графічних процесорів, що є необхідною умовою для уникнення ефекту «вузького місця (botleneck) під час виконання матричних мультиплікацій та суттєвого скорочення загального часу навчання моделі [Bisong, Building Machine Learning and Deep Learning Models on Google Cloud Platform, 2019]

Таблиця 1.3 Основні інструменти технологічного стека

Інструмент	Призначення
Python	Основна мова програмування для реалізації системи
TensorFlow	Фреймворк для створення та навчання нейронних мереж
Keras	Спрощений API для побудови моделей глибокого навчання
OpenCV	Обробка та підготовка зображень
NumPy	Робота з масивами даних та математичними операціями
Matplotlib	Побудова графіків та візуалізація результатів

Використання python, teensorflow та keras е доцільним рішенням для створення системи оцінки біометричних показників людини. Обраний технологічний стек забзпечує зручність розробки, високу продуктивність та можливість ефективного навчання згорткових нейронних мереж для задач комп’ютерного зору.

1.4 Постановка задачі на дослідження

На основі проведеного аналізу предметної області, сучасних методів комп’ютерного зору та підходів глибокого навчання можна сформулювати

основну мету та завдання даної кваліфікаційної роботи. Особлива увага приділяється застосуванню згорткових нейронних мереж для автоматичного аналізу фотографій обличчя та визначення біометричних характеристик людини.

Метою кваліфікаційної роботи є підвищення точності та надійності автоматизованого визначення статі людини в інтелектуальних системах відеоаналітики шляхом застосування оптимізованих архітектур глибокого навчання та технології перенесення знань (transfer learning). Безпосередня розробка програмної системи та проведення експериментальних досліджень є практичними інструментами для досягнення поставленої мети та верифікації ефективності запропонованого підходу на реальних наборах біометричних даних.

Для досягнення поставленої мети необхідно виконати такі теоретичні та практичні задачі.

1. Провести аналіз сучасних методів розпізнавання біометричних характеристик людини та визначити основні проблеми, які виникають під час обробки цифрових зображень.

2. Дослідити принципи роботи згорткових нейронних мереж та обґрунтувати доцільність їх використання у задачах комп’ютерного зору.

3. Виконати аналіз і підготовку набору даних зображень обличчя для проведення експериментального дослідження.

4. Реалізуваи базову модель згорткової нейронної мережі для задачі класифікації статі людини за фотографією.

5. Дослідити можливості технології transfer learning та виконати адаптацію переднавченої моделі mobilenetv2 до обраного набору даних.

6. Провести навчання та тестування створених моделей із використанням сучасних методів оптимізації.

7. Виконати порівняльний аналіз результатів роботи моделей за допомогою метрик accuracy, precision, recall та f1-score, а також побудувати матриці помилок для оцінки якості класифікації.

8. Проаналізувати вплив попередньої обробки зображень та параметрів навчання на точність роботи нейронної мережі.

Об’єктом дослідження є процеси аналізу зображень обличчя та визначення біометричних характеристик людини за допомогою методів штучного інтелекту.

Предметом дослідження є згорткові нейронні мережі, методи глибокого навчання та алгоритми перенесення навчання для задачі бінарної класифікації зображень облич.

Практичне значення роботи полягає у можливості створення програмної системи, яка може бути використана у задачах відеоспостереження, цифрової аналітики, систем контролю доступу та інших інтелектуальних сервісх де необхідно автоматично визначати характеристики людини за фотографією

Висновки до озділу 1

У першому розділі кваліфікаційної роботи було проведено аналіз предметної області визначення біометричних характеристик людини за фотографіями обличчя. Розглянуто основні задачі сучасного комп’ютерного зору, серед яких визначення статі, оцінка віку та розпізнавання емоційного стану людини. Також було визначено основні фактори, що впливають на точність роботи систем аналізу зображень, зокрема освітлення, якість фотографій, положення обличчя та індивідуальні особливості людини.

У процесі дослідження було встановлено, що найбільш ефективним підходом для аналізу цифрових зображень є використання згорткових нейронних мереж. CNN дозволяють автоматично виділяти важливі ознаки

обличчя та забезпечують високу точність класифікації навіть у складних умовах. Крім цього, було розглянуто основні компоненти архітектури згорткових мереж та принципи їх роботи.

Також у розділі обґрунтовано вибір технологічного стека для реалізації практичної частини роботи. Для створення програмної системи було обрано мову програмування python та фреймворк tensorflow/keras, які забезпечують зручний процес розробки та підтримують сучасні методи глибокого навчання. Для проведення експериментів планується використання хмарного середовища Google Colab, яке дозволяє використовувати графічні процесори для прискорення навчання моделей.

На основі проведеного аналізу було сформульовано основну мету дослідження та визначено задачі, які необхідно реалізувати у наступних розділах кваліфікаційної роботи.

РОЗДІЛ 2. ПІДГОТОВКА ДАНИХ ТА ПРОЄКТУВАННЯ МОДЕЛЕЙ

2.1 Анализ и предварительная обработка выбранного набора данных (аугментация, нормализация)

В якості основної емпіричної бази для цього дослідження та вирішення завдань автоматичного розпізнавання біометричних характеристик людини за цифровими зображеннями обличчя я обрав масштабний відкритий набір даних utkface. [Zhang et al., Age Progression/Regression via Deep Backbone, 2017].

Вибір саме цього датасету зовсім не випадковий. У сучасному комп'ютерному зорі якість та різноманітність навчальної вибірки безпосередньо впливають на підсумкову ефективність будь-якої нейромережі. Цей набір налічує понад двадцять тисяч цифрових фотографій облич у форматі rgb. Такий обсяг інформації є цілком достатнім для тренування глибоких моделей без ризику миттєвого перенавчання.

Головна цінність колекції utkface полягає в її надзвичайно високій варіативності за ключовими демографічними та біометричними показниками. Вік людей на світлинах охоплює весь життєвий цикл від немовлят до ста шістнадцяти років. Крім того, набір добре збалансований за етнічним та гендерним складом. Це вкрай важливо, якщо ми хочемо створити об'єктивну систему штучного інтелекту без внутрішніх упереджень. Зображення в датасеті відображають цілком реальні умови зйомки. В індустрії подібні дані часто називають неструктурованими або ж такими, що зібрані в «диких» умовах. Для них характерні суттєві перепади в освітленні, різні ракурси зйомки, повороти й нахили голови, різноманітна міміка, а також наявність аксесуарів, що перекривають обличчя, наприклад, сонцезахисні окуляри, капелюхи чи об'ємні зачіски.

Робота з такими сирими матеріалами є доволі складною обчислювальною задачею. Не можна просто так згодувати ці фотографії алгоритму машинного навчання і сподіватися на якісний результат. Перед подачею інформації на вхід нейронної мережі потрібна ретельна підготовка. Процес обробки даних для навчання глибоких нейромереж у цьому дослідженні складався з кількох послідовних етапів, і кожен з них має чітке математичне обґрунтування.

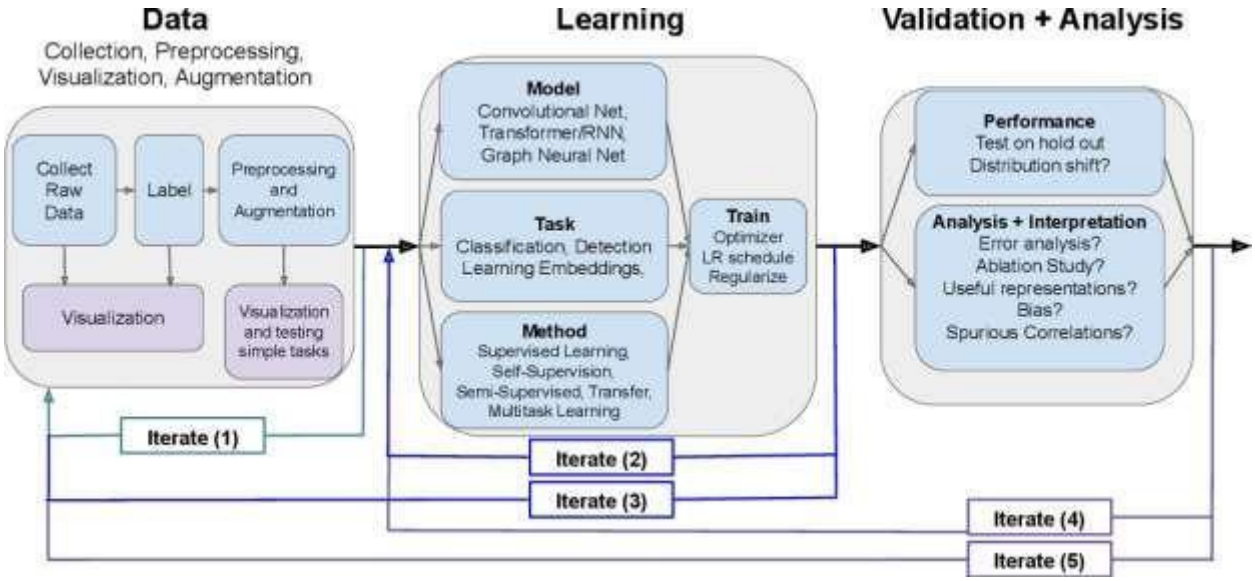


Рис.2.1-Схема Последовательных этапов подготовки и предобработки входных данны

2.1.1 Парсинг метаданих та фільтрація зображень

Першим етапом алгоритму підготовки став парсинг та сувора фільтрація даних. Імена файлів у вихідному наборі UTKFace містять дуже корисну інформацію. Вони мають структуровані мітки у спеціальному текстовому форматі, де послідовно через нижнє підкреслення вказані вік, стать, раса та часова мітка створення файлу. На етапі початкового завантаження файлів в оперативну пам'ять системи я реалізував спеціальний алгоритм для автоматичного вилучення цільової ознаки гендеру. В

архітектурі розроблюваної системи бінарної класифікації мітка нуль алгоритмічно відповідає чоловічій статі, а мітка один вказує на жіночу стать.

У процесі читання та обробки тисяч файлів алгоритм постійно перевіряв структурну цілісність метаданих. Фотографії з пошкодженою структурою імені файлу, пропущеними значеннями або некоректними даними автоматично видалялися із загальної вибірки. Ця дія є критично важливим кроком, щоб запобігти появі шуму в навчальному сигналі. Якщо нейромережа отримає фотографію чоловіка з помилковою міткою жінки, алгоритм градієнтного спуску буде намагатися адаптувати вагові коефіцієнти під хибну інформацію. Це неминуче знизить загальну точність класифікатора на фінальних етапах тестування.

2.1.2 Зміна просторового розширення та масштабуваня

Другим важливим етапом стала зміна розмірів зображень, що в англійській термінології називається resizing. Оригінальні фотографії в досліджуваному наборі мають різні розміри. Для стандартних повнозв'язних та згорткових мереж подача картинок різного розміру є абсолютно неприпустимою, оскільки архітектура нейромережі вимагає суворо фіксованого розміру вхідного тензора на першому шарі.

Щоб забезпечити константну розмірність та оптимізувати споживання обчислювальних ресурсів у хмарному середовищі Google Colab, усі вхідні зображення були програмно масштабовані до квадратного розміру шістдесят чотири на шістдесят чотири пікселі. При цьому було збережено 3 базові колірні канали RGB, що відображають червону, зелену та синю складові кожного пікселя. [Goodfellow et al., Deep Learning, 2016]

Вибір розмірності 64x64пікселі є науково обґрунтованим компромісом. З одного боку, матриця такого розміру зберігає ключові локальні біометричні ознаки обличчя. На такому зображенні алгоритми комп'ютерного зору все ще здатні розрізнити контури очей, форму носа, лінію

губ та особливості росту волосся. З іншого боку, таке розширення суттєво знижує вимоги до обсягу відеопам'яті графічного процесора в процесі навчання моделей. Зображення розміром двісті двадцять чотири на двісті двадцять чотири пікселі вимагали б у десятки разів більше пам'яті для зберігання градієнтів на етапі зворотного поширення помилки, що зробило б неможливим навчання в рамках доступних безкоштовних хмарних потужностей [LeCun et al., Deep Learning, 2015]

2.1.3 Математична нормалізація даних

Третій етап являє собою математичну нормалізацію даних. Вихідні значення пікселів у цифрових фотографіях кодуються цілими числами в суворому діапазоні від нуля до двохсот п'ятдесяти п'яти, де нуль означає повну відсутність кольору, а двісті п'ятдесят п'ять означає максимальну яскравість каналу. [Goodfellow et al., Deep Learning, 2016] Подача таких великих чисел на вхід штучної нейронної мережі є серйозною архітектурною помилкою. Під час множення великих значень пікселів на вагові коефіцієнти нейронів суми активацій починають неконтрольовано зростати. Це призводить до нестабільності градієнтів та ефекту насичення функцій активації, коли алгоритм просто перестає навчатися.

Для підвищення стабільності роботи алгоритмів градієнтного спуску та значного прискорення збіжності цільової функції втрат було виконано процедуру масштабування. Усі цілочисельні значення пікселів перетворили у формат чисел із плаваючою крапкою одинарної точності float32. Потім їх масштабували до безперервного діапазону від нуля до одиниці шляхом математичного ділення матриці інтенсивностей на константу 255 [Chollet, Deep Learning with Python, 2017]

Математично це перетворення виглядає наступним чином:

$$X_{norm} = \frac{X}{255.0}$$

2.1.4 Стратифікація та формування вибірок

Четвертий етап включає правильне формування незалежних вибірок для навчання та тестування. Щоб мінімізувати ризики перенавчання мережі та забезпечити об'єктивну оцінку узагальнюючої здатності алгоритмів, усю доступну очищену вибірку було розділено у суворому співвідношенні. Вісімдесят відсотків даних було виділено виключно для тренування моделей, а двадцять відсотків залишено для валідації та фінального тестування.

Вкрай важливо зазначити, що розділення проводилося не випадковим чином, а методом стратифікації [Kuhn et al., Applied Predictive Modeling, 2013]. Цей статистичний метод гарантує збереження ідентичного балансу цільових класів чоловіків та жінок як у навчальному, так і в тестовому підмножинах. Якби застосовувалося просте випадкове розділення, існував би ризик того, що алгоритм створить дисбаланс. Наприклад, у тестову вибірку могло б потрапити непропорційно багато зображень однієї статі. У такому випадку система оцінки показала б спотворені підсумкові метрики якості, які не відображали б реальну здатність мережі вирішувати поставлену задачу бінарної класифікації.

2.1.5 Методи аугментації даних

П'ятим і завершальним етапом підготовки стала просторова аугментація даних. Теоретичний аналіз сучасних підходів глибокого навчання вказує на критичну важливість застосування методів штучного розширення набору даних. В умовах обмеженої кількості унікальних фотографій нейронні

мережі схильні просто запам'ятовувати вихідні кадри. [Goodfellow et al., Deep Learning, 2016] Для боротьби з цим ефектом застосовуються трансформації.

У програмному конвеєрі було застосовано випадкові горизонтальні відображення та помірні повороти зображень на випадковий кут. Ці матричні операції дозволяють штучно розширити візуальну варіативність навчальної множини без необхідності фізичного збору нових фотографій. Згорткові фільтри завдяки аугментації вчаться інваріантно реагувати на просторові зрушення об'єктів. Модель починає розуміти, що відображене по горизонталі або злегка повернуте обличчя чоловіка все одно залишається обличчям чоловіка, а не стає новим невідомим об'єктом. [Chollet, Deep Learning with Python, 2017].



Рис. 2.2 - Приклади зображень з набору даних UTKFace після попередньої обробки

2.2 Проєктування базової архітектури нейронної мережі (Baseline)

Для оцінки ефективності складних моделей глибокого навчання та методів перенесення навчання у межах цього дослідження необхідно сформувати базову лінію (Baseline) контрольний пункт для порівняння. Як базові архітектури було обрано та спроєктовано з нуля (from scratch) дві моделі різного ступеня складності: багатошаровий повнозв'язний персептрон (ANN) та просту згорткову нейронну мережу (Simple CNN).

2.2.1 Повнозв'язна штучна нейронна мережа (Fully Connected ANN)

Перша базовая архітектура реалізована у вигляді класичного багатошарового персептрона (MLP), де кожен нейрон поточного шару жорстко пов'язаний з усіма нейронами попереднього шару. Головна мета інтеграції цієї моделі в експеримент демонстрація обмеженості стандартних лінійних шарів під час обробки просторових даних високої розмірності (цифрових зображень).

Процес проходження сигналу (Forward Pass) у спроєктованій ANN складається з таких етапів:

Вхідне згладжування (Flattening): Оскільки повнозв'язний шар здатний приймати на вхід лише одновимірні вектори, першим шаром моделі виступає Flatten. Він перетворює вхідний тривимірний тензор нормованого зображення обличчя розмірністю $(64, 64, 3)$ в одновимірний вектор ознак. Розмірність отриманого вектора обчислюється як добуток просторових координат та кількості каналів [Haykin, Neural Networks and Learning Machines, 2009]

$$Nfeatures = 64 \text{ x } 64 \text{ x } 3 = 12288$$

1. Приховані шари обробки ознак: Модель містить два послідовні приховані повнозв'язні шари (Dense) 256 та 64 нейрони відповідно. Для кожного нейрона на цих шарах виконується лінійне перетворення вхідного вектора x з ваговими коефіцієнтами w та зсувом б:

$$z = \sum_{i=1}^{m} w_i x_i + b$$

Після лінійного перетворення сигнал проходить крізь нелінійну функцію активації ReLU (Rectified Linear Unit), яка математично описується як [Goodfellow et al., Deep Learning, 2016]:

$$f(z) = \max(0, z)$$

Використання ReLU замість класичних функцій (сигмоїди або гіперболічного тангенса) дозволяє уникнути ефекту затухання градієнтів на етапі зворотного поширення помилки (Backpropagation) та забезпечує розрідженість (sparsity) активації нейронів.

Вихідна класифікація: Фінальний шар містить 2 нейрони (що відповідає кількості цільових класів: чоловік / жінка). На цьому шарі використовується функція активації Softmax, яка нормує вихідні логіти мережі у розподіл ймовірностей, сума яких дорівнює одиниці:

$$\sigma(z)_i = \frac{e_{zi}}{\left(\sum_{j=1}^{K} e^{z_j}\right)}, \quad i = 1, K\ (K = 2)$$

Деталізовану структуру та візуалізацію графів обчислень моделі ANN представлено у табл. 2.1.

Таблиця 2.1 Специфікація та структура шарів моделі Fully Connected ANN

Назва шару (Тип)	Розмірніст вихідного	Кількість параметрі	Функція активації
Input_Layer (Вхід)	(None, 64,	0	Немає
Flatten (Згладжування)	(None, 122	0	Немає
Dense_Hidden_1 (Полнозв'язний)	(None, 256	12288x256 3,146,240$	ReLU
Dense_Hidden_2 (Повнозв'язний)	(None, 64)	256 x 64 + 16,448	ReLU
Dense_Output (Вихідний)	(None, 2)	64 x 2 + 2	Softmax
Усього параметрів:	3,162,818 млн)		

Аналітичний недолік архітектури*:* Як видно з табл. 2.1, повнозв'язна модель вимагає понад 3,16 мільйона параметрів для навчання. При цьому переважна більшість ваг припадає на перший прихований шар. Модель повністю ігнорує той факт, що сусідні пікселі обличчя пов'язані просторово локальні ознаки: контур очей, форма носа що робить її вкрай чутливою до найменших зсувів або поворотів голови на фотографії.

2.2.2 Проста згорткова нейронна мережа (Simple CNN)

Для усунення критичних недоліків повнозв'язної мережі була спроєктована архітектура, заснована на принципах локальних рецептивних полів та розділення ваг згорткова нейронна мережа. Спроєктована модель Simple CNN послідовно витягує ієрархічні ознаки з двовимірного зображення, зберігаючи просторову топологію.

Модель побудована за блочним принципом і включає такі компоненти:

1. Перший блок вилучення ознак (Low-level features):

Згортковий шар (Conv2D): містить 32 фільтри ядра згортки розмірністю 3x3. Крок згортки (stride) дорівнює 1, стратегія заповнення країв padding='same' (просторова розмірність карти ознак на виході залишається (64x64). Цей шар фіксує базові низькорівневі геометрічні примітиви: межі, лінії, контрастні переходи світлотіні [LeCun et al., Deep Learning, 2015].

Шар субдискретизації (MaxPooling2D): виконує операцію максимізації в локальному вікні розміром 2 x 2 з кроком 2. Це дозволяє вдвічі зменшити просторову роздільну здатність карти ознак (з 64x64x32), виділяючи найбільш інваріантні ознаки та знижуючи обчислювальне навантаження. Другий блок вилучення ознак (Mid-level features) [Goodfellow et al., Deep Learning, 2016]

Згортковий шар (Conv2D): містить 64 фільтри розмірністю 3x3 з активацією ReLU та padding='same'. Збільшення кількості фільтрів з 32 до 64 на глибших шарах є класичним підходом, оскільки мережа починає комбінувати прості лінії у складніші текстурні патерни та форми елементів обличчя. [Hubel, Wiesel, Receptive Fields..., 1959].

Шар субдискретизації (MaxPooling2D): повторно зменшує просторову розмірність карт ознак з 32x32 до 16x16 Блок класифікації (Голова мережі).

Вихідні двовимірні карти ознак другого блоку перетворюються шаром Flatten в один вектор, який передається на компактний прихований повнозв'язний шар із 64 нейронами (ReLU) для фінальної інтеграції високорівневих ознак. Вихідний шар складається з 2 нейронів з активацією Softmax для бінарного вибору гендера.

Деталізовану структуру та просторову зміну тензорів для моделі Simple CNN наведено у табл. 2.2.

Таблиця 2.2 Специфікація та структура шарів моделі Simple CNN

Назва шару (Тип)	Розмірність вихідного тензора	Кількість параметрів (ваг)	Параметри шару	Функция активации
Input_Layer (Вхід)	(None, 64, 64, 3)	0	-	Немає
Conv2D_1 (Згортка)	(None, 64, 64, 32)	(3x3x3)x32+32=920	32 фільтри 3x3	ReLU
MaxPooling2D_1	(None, 32,32, 32)	0	Окно 2x2 stride=2	Немає

Conv2D_2 (Згортка)	(None, 32, 32, 64)	(3x3x32) x 64 + 64 = 18,496$	64фільтри, 3x3	ReLU
MaxPooling2D_2	(None, 16, 16, 64)	0	Окно 2x2 stride=2	Немає
Flatten (Згладжування)	(None, 16384)	0	16x16x64=16384	Немає
Dense_Classifier	(None, 64)	1638x64+64=1,048,640	64 нейрони	ReLU
Dense_Output	(None, 2)	64x2+2=130	2 нейрони	Softmax
Усього параметрів:	1,068,186 (1.06 млн)			

Аналіз структури моделі, наведеної в табл. 2.2, свідчить про ефективне використання згорткових шарів для виділення просторових ознак зображення. На відміну від повнозв’язної архітектури, згорткова мережа використовує значно меншу кількість параметрів при збереженні здатності виявляти характерні риси обличчя. Це дозволяє підвищити швидкість навчання моделі та знизити ризик перенавчання на обмеженому наборі даних.

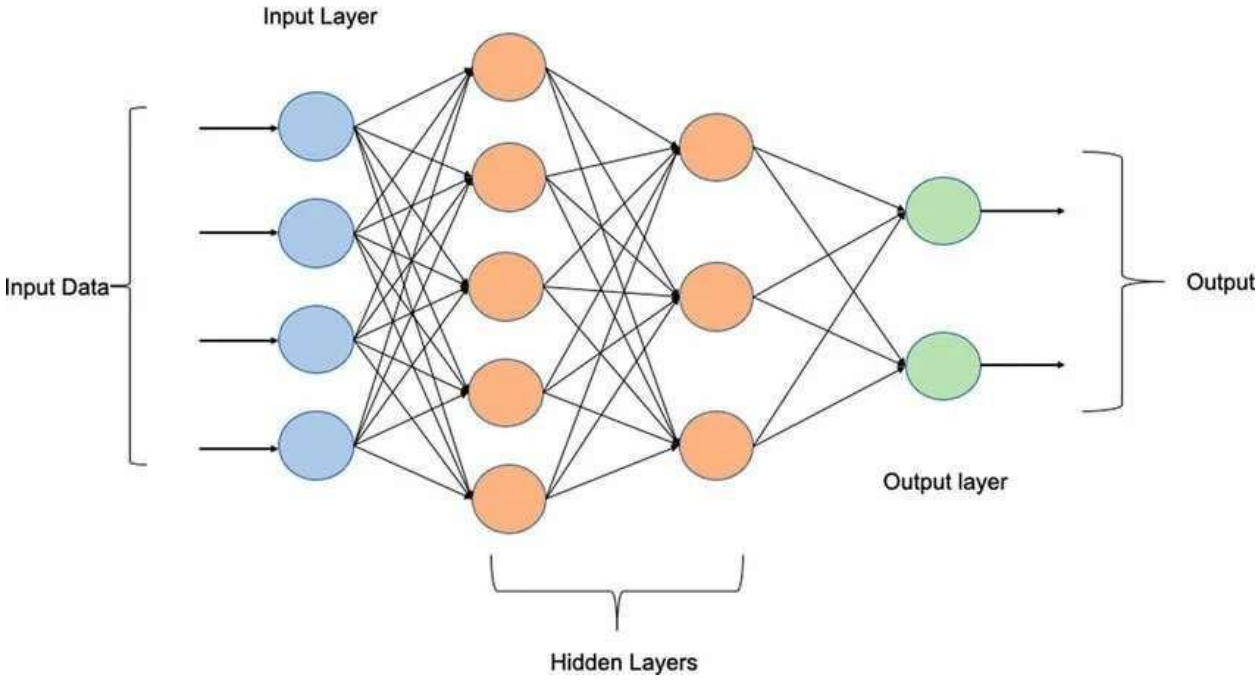


Рис. 2.2 - Граф вычислений и топологія базовой моделі FC ANN

2.3 Використання технологій перенесення навчання на базі існуючих архітектур

Для досягнення дійсно максимальних показників точності в умовах обмежених обчислювальних ресурсів та відносно невеликого обсягу зібраної вибірки у дипломній роботі був застосований передовий метод перенесення навчання. (Transfer Learning) [Pan et al., A Survey on Transfer Learning, 2010] Концепція цього методу полягає у використанні накопичених знань, які були отримані величезною нейронною мережею під час вирішення однієї глобальної задачі, для ефективного вирішення іншої локальної задачі. У рамках нашого дослідження ми орієнтувалися на архітектури ResNet50 та MobileNetV2, щоб порівняти ефективність великовагових та легковагових моделей. [Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge, 2015].

2.3.1 Інтеграція архітектури ResNet50

Як перший і основний екстрактор ознак була обрана надглибока залишкова нейронна мережа архітектури ResNet50. Ця математична модель була попередньо навчена в лабораторіях на колосальному наборі даних ImageNet, який налічує понад мільйон чотириста тисяч зображень загального вигляду, ретельно розбитих на тисячу категорій.

Використання архітектури ResNet50 зумовлене її справді унікальною структурною особливістю, а саме наявністю наскрізних залишкових зв'язків між шарами. У звичайних глибоких мережах сигнал під час проходження через десятки нелінійних шарів поступово спотворюється та загасає. Залишкові зв'язки дозволяють передавати частину вхідного сигналу безпосередньо через один або кілька обчислювальних блоків, повністю оминаючи важкі нелінійні перетворення.

Математична логіка такого залишкового обчислювального блоку чітко описується формулою [He et al., Deep Residual Learning for Image Recognition, 2016]

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

де змінна x виступає вихідним вхідним сигналом блоку, а функція нелінійним перетворенням всередині самого згорткового блоку.

це витончене технологічне рішення успішно нівелює фундаментальну проблему деградації глибоких мереж і запобігає загасанню градієнтів на глибоких рівнях оптимізації. [He et al., Deep Residual Learning for Image Recognition, 2016] Воно дозволяє ефективно навчати мережі, реальна глибина яких досягає десятків і сотень шарів. Завдяки тривалому навчанню на масиві imagennet згорткові шари мережі ResNet50 вже містять високооптимізовані фільтри. Різні шари алгоритму здатні розпізнавати універсальні візуальні примітиви, починаючи від простих прямих ліній на

перших рівнях і закінчуючи складними біометричними контурами облич на глибоких рівнях абстракції.

У межах нашого експериментального дослідження була реалізована класична стратегія вилучення ознак. Величезну модель ResNet50 завантажили в оперативну пам'ять з оригінальними ваговими коефіцієнтами ImageNet, при цьому оригінальну повнозв'язну класифікаційну частину на тисячу класів було програмно відсічено. Вхідний розмір прийманого тензора був жорстко зафіксований на значеннях шістдесят чотири на шістдесят чотири на три канали. Абсолютно всі вагові коефіцієнти згорткового базису були заморожені. Заморожування ваг запобігає катастрофічному руйнуванню вже чудово навчених детекторів ознак потужними та хаотичними градієнтними кроками на ранніх етапах адаптації класифікатора під новий набір даних облич. [Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge, 2015].

Проєктування нової класифікаційної голови виконано з використанням функціонального програмного інтерфейсу Keras. До виходу замороженого згорткового базису був послідовно застосований шар глобального просторового усереднення [Lin et al., Network In Network, 2013]. Цей алгоритмічний шар математично агрегує просторову інформацію згенерованих карт ознак шляхом обчислення середнього арифметичного значення для кожного окремого каналу. Така операція елегантно перетворює багатовимірний вихідний тензор у компактний плоский вектор ознак фіксованої довжини, що містить дві тисячі сорок вісім елементів.

Шар глобального усереднення кардинально знижує ризик перенавчання порівняно зі звичайною операцією Flatten, оскільки він взагалі не має параметрів, що навчаються. Одразу за ним розташовується повнозв'язний адаптаційний шар на шістдесят чотири нейрони з активацією ReLU, а потім фінальний шар на два нейрони з активацією Softmax для

обчислення апостеріорних ймовірностей належності обличчя до чоловічої або жіночої статі.

2.3.2 Інтеграція архітектури MobileNetV2

Крім важкої архітектури resnet50, для дотримання вимог початкового технічного завдання була спроєктована адаптація переднавченої моделі mobilenetv2. Головна проблема класичних згорткових мереж полягає в їхній величезній обчислювальній складності, що робить неможливим їх запуск на слабких мобільних пристроях або вбудованих камерах відеоспостереження. Архітектура mobilenetv2 вирішує цю проблему за рахунок застосування концепції глибинно-роздільних згорток [Sandler et al., MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2018].

Замість того, щоб виконувати повноцінну дорогу тривимірну математичну згортку по всіх просторових пікселях і колірних каналах одночасно, алгоритм mobilenetv2 розбиває цей процес на два незалежні етапи. [Sandler et al., MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2018] Перший етап виконує просторову фільтрацію кожного каналу окремо. Другий етап застосовує точкову згортку розміром один на один піксель для комбінування інформації між каналами. Такий підхід скорочує кількість математичних операцій множення майже в десять разів порівняно зі стандартними мережами при мінімальному падінні передбачувальної точності. [Sandler et al., MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2018]

Процес підготовки моделі mobilenetv2 до перенесення навчання був ідентичним до процедури для resnet50. Базові шари екстрактора ознак були заморожені, оригінальний класифікатор видалений, а замість нього інтегровані шари глобального усереднення та повнозв'язний класифікатор на два гендерні класи. Порівняння продуктивності важкої залишкової мережі та

легкої мобільної архітектури дозволить зробити цінні висновки про доцільність використання тих чи інших алгоритмів у комерційній розробці.

2.4 Вибір метрик для оцінки ефективності моделей

Об'єктивна оцінка якості роботи спроєктованих алгоритмічних моделей, їх валідація та коректний порівняльний аналіз вимагають використання комплексної системи взаємодоповнюваних математичних метрик. Спирання виключно на один показник точності може легко призвести до хибних висновків про здатність алгоритму надійно працювати в реальних умовах експлуатації. У межах цього дослідження були обрані наступні ключові показники програмної оцінки.

2.4.1 Цільова функція втрат

Як головну оптимізаційну цільову функцію для всіх навчених моделей було обрано розріджену категоріальну крос-ентропію. Вибір саме цієї функції зумовлений тим, що вона математично ідеально підходить для задач суворої бінарної або багатокласової класифікації, коли мітки цільових класів у навчальному наборі даних представлені у вигляді простих цілочисельних індексів нуль або один.

Математично функція крос-ентропії виражається наступним чином:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M} y_{ic} \backslash ln(p_{ic})$$

У представленій формулі змінна N позначає загальний обсяг пакета даних, тобто розмір батча вибірки. Параметр pі,уі вказує на передбачену моделлю ймовірність істинного класу для кожного конкретного об'єкта вибірки. Алгоритм зворотного поширення помилки спрямований на мінімізацію цього значення в процесі навчання. Чим ближче значення

ймовірності правильної відповіді до одиниці, тим ближче значення логарифма до нуля, і тим менший загальний штраф отримує мережа.

2.4.2 Метрики якості класифікації

Найбільш зрозумілою та базовою метрикою ефективності роботи нейронної мережі є загальна точність класифікації, широко відома під терміном Accuracy. Вона визначає частку абсолютно правильно класифікованих зображень облич серед загального обсягу наданої системі валідаційної або тестової вибірки.

Формула розрахунку загальної точності має вигляд:

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

У цій класичній статистичній формулі змінна TP позначає кількість істинно позитивних результатів, змінна TN позначає кількість істинно негативних результатів, змінна FP позначає кількість хибнопозитивних помилок першого роду, а змінна FN вказує на кількість хибнонегативних помилок другого роду алгоритму. Використання метрики загальної точності є абсолютно математично коректним підходом у нашому конкретному випадку, оскільки сформований робочий зріз набору даних UTKFace жорстко збалансований за кількістю зображень чоловіків та жінок. Якби у вибірці було дев'яносто відсотків чоловіків і десять відсотків жінок, дана метрика втратила б свій об'єктивний сенс.

Однак для проведення по-справжньому детального аналізу помилок нейромережевої класифікації у складних граничних випадках потрібне залучення апарату матриці помилок. Матриця помилок являє собою таблицю,

яка дозволяє візуально оцінити, які саме категорії зображень мережа плутає найчастіше. На основі комірок цієї матриці обчислюються дві додаткові метрики: точність передбачення позитивного класу Precision та повнота виявлення цільового класу Recall.

Метрика Precision наочно показує, яка частка фотографій, алгоритмічно класифікованих мережею як певний гендер, дійсно належить до цього гендеру в реальності. Метрика Recall, у свою чергу, демонструє, яку частку зображень конкретного гендеру мережа змогла успішно знайти в загальному масиві даних. Оскільки в реальних прикладних задачах між цими двома метриками постійно виникає якісний дисбаланс, для їх об'єктивного об'єднання застосовується гармонічне середнє значення, відоме в статистиці як F1-міра.

Формула розрахунку гармонічного середнього:

$$F_1 = 2 \cdot \frac{(Precision \cdot Recall)}{(Precision + Recall)}$$

Використання F1-міри дозволяє отримати максимально сувору оцінку реальної продуктивності моделі навіть в екстремальних ситуаціях, коли нейронна мережа через помилки навчання починає схилятися до вибору тільки одного мажоритарного класу, ігноруючи інші.

Висновки до розділу 2

У другому розділі ми детально розглянули процес підготовки даних та створення архітектур нейронних мереж для розпізнавання статі людини. Як основну базу для навчання було обрано набір фотографій utkface, який чудово підходить для нашої задачі завдяки великій кількості різноманітних облич.

Ми зясували, що нейромережа не може працювати з необробленими картинками. Тому дані пройшли кілька етапів підготовки. Спочатку ми дістали правильні мітки статі з назв файлів та відсіяли пошкоджені фото. Потім усі зображення зменшили до єдиного розміру шістдесят чотири на шістдесят чотири пікселі, щоб системі вистачило пам'яті для швидких розрахунків. Після цього значення кольорів математично нормалізували. Це значно прискорює та стабілізує процес навчання. Для захисту від перенавчання ми поділили набір на тренувальну та тестову частини. Також ми застосували аугментацію, тобто штучно розширили набір за рахунок випадкових поворотів та віддзеркалень фотографій.

Для практичного експерименту ми спроєктували кілька різних моделей. Базовими стали звичайна повнозв'язна мережа багатошаровий персептрон та проста згорткова мережа. Аналіз показав, що повнозв'язні моделі дуже погано працюють із зображеннями, оскільки втрачають просторові зв'язки між пікселями. Натомість згорткова архітектура набагато краще виділяє ознаки обличчя за допомогою матричних фільтрів.

Щоб отримати дійсно високу точність, ми вирішили використати передову технологію перенесення навчання. Для цього були адаптовані дві вже готові моделі: важка resnet50 та легка mobilenetv2. Ми заморозили їхні основні шари, щоб зберегти базові знання алгоритмів комп'ютерного зору, і додали нові шари спеціально для нашої задачі класифікації статі.

Наприкінці розділу ми визначилися з метриками для оцінки результатів. Крім базового показника загальної точності accuracy ми будемо використовувати цілу систему оцінки, куди входить функція втрат та матриця помилок. Додатково рахуватимуться метрики pprecision recall і f1 міра. Це дозволить нам максимально об'єктивно порівняти всі моделі та побачити, в яких саме ситуаціях вони помиляються найчастіше.

РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

3.1 Опис апаратно-програмного середовища та конфігурації гіперпараметрів

Успішне навчання сучасних глибоких згорткових нейронних мереж вимагає колосальних обчислювальних потужностей. Це пов'язано насамперед з необхідністю виконання мільярдів операцій матричного множення тензорів на кожному кроці оптимізації ваг. З огляду на це всі експериментальні програмні розрахунки для цієї дипломної роботи проводилися у спеціалізованому хмарному обчислювальному середовищі Google Colab. Ця платформа надає безкоштовний доступ до високопродуктивних графічних прискорювачів. Для забезпечення необхідної швидкості обробки батчів тензорів був задіяний хмарний графічний процесор, який надає достатній обсяг відеопам'яті для паралельної обробки графічної інформації. Відсутність графічного прискорювача збільшила б час компіляції та навчання моделей з кількох годин до кількох тижнів.

Програмна реалізація алгоритмів була повністю виконана мовою програмування Python, яка де-факто визнана світовим стандартом в індустрії штучного інтелекту. Для безпосереднього конструювання архітектур мереж використовувався високорівневий прикладний інтерфейс бібліотеки Keras, що працює поверх потужного тензорного обчислювального ядра TensorFlow. Використання цього стека технологій гарантує високу надійність розрахунків. Для забезпечення суворої наукової відтворюваності одержуваних результатів генератори псевдовипадкових чисел для бібліотек обчислень були зафіксовані на єдиному базовому значенні перед запуском кожного експерименту.

Процес навчання суворо контролювався заданим набором гіперпараметрів. Вибір цих значень завжди зумовлений пошуком тонкого

балансу між швидкістю збіжності математичного алгоритму та загальною стабільністю одержуваних на валідації результатів. Розмір пакета даних, тобто батча, був встановлений константно на значенні тридцять два зображення. Такий розмір забезпечує оптимальне співвідношення між точністю статистичної оцінки градієнта функції втрат та обсягом споживаної оперативної пам'яті обчислювального пристрою.

Як базовий алгоритм оптимізації градієнтного спуску абсолютно у всіх експериментах застосовувався адаптивний метод Adam. Цей алгоритм динамічно обчислює індивідуальні швидкості навчання для кожного вагового параметра нейронної мережі на основі збору оцінок першого та другого статистичних моментів градієнта. Базова стартова швидкість навчання для простих моделей персептрона та легкої згорткової мережі була встановлена на рівні однієї тисячної. Для глибоких моделей на базі технологій перенесення навчання стартова швидкість навчання була примусово знижена до однієї десятитисячної. Таке зниження необхідне для того, щоб уникнути занадто різких коригувань ваг нового адаптаційного шару класифікатора, які могли б дестабілізувати заморожений базис.

Навчання кожної досліджуваної нейронної мережі теоретично планувалося протягом п'ятдесяти повних проходів по набору тренувальних даних, які називаються епохами. Однак для ефективного запобігання фатальному перенавчанню був впроваджений програмний механізм ранньої зупинки процесу. Цей алгоритм безперервно відстежував значення функції втрат на окремій незалежній валідаційній вибірці. Якщо цей показник не демонстрував жодних покращень протягом п'яти послідовних епох, процес математичного навчання автоматично переривався системою, а в пам'ять відновлювалися ті вагові коефіцієнти шарів мережі, які показали найкращий результат валідації в минулому. Додатково в конвеєр застосовувався

програмний механізм динамічного зниження швидкості навчання при виході функції втрат на довге плато.

Навчання базових архітектур та первинний аналіз динаміки метрик Першим практичним етапом експериментального дослідження стало безпосереднє навчання класичної повнозв'язної нейронної мережі з нуля. Як і передбачалося на попередньому етапі теоретичного проектування архітектури, багатошаровий персептрон зіткнувся з фундаментальними обчислювальними труднощами під час обробки щільної графічної інформації.

Протягом перших п'яти епох модель демонструвала надзвичайно стрімке зниження значення функції втрат на тренувальному наборі даних. Точність розпізнавання класів на навчальній вибірці швидко перевищила показник у вісімдесят п'ять відсотків. Однак графік метрик на валідаційній вибірці, яку модель не бачила в процесі оптимізації, показав зовсім іншу картину. Починаючи з сьомої епохи навчання валідаційна точність повністю перестала зростати, жорстко зафіксувавшись на рівні близько сімдесяти двох відсотків, тоді як тренувальна точність продовжувала прагнути до абсолютних ста відсотків впевненості.

Таке колосальне розходження кривих графіків є в машинному навчанні класичним симптомом важкого перенавчання мережі. Модель не навчилася шукати загальні концепції чоловічого або жіночого обличчя. Вона просто почала механічно запам'ятовувати конкретні значення пікселів тренувальних зображень, повністю втративши алгоритмічну здатність до математичного узагальнення візуальних ознак. Вбудований механізм ранньої зупинки штатно перервав марне навчання персептрона на дванадцятій епосі. Висока архітектурна чутливість повнозв'язної мережі до найменших просторових зрушень обличчя на фотографії зробила цю модель абсолютно непридатною для точноїкласифікації біометричних показників.

Другим етапом стало навчання спроектованої нами простої згорткової нейронної мережі. Впровадження локальних згорткових фільтрів та шарів просторової субдискретизації кардинально змінило динаміку збіжності алгоритму градієнтного спуску на краще.

Легка згорткова мережа навчалася значно стабільніше. Криві функції втрат для тренувального та незалежного валідаційного наборів даних знижувалися плавно і синхронно протягом перших п'ятнадцяти епох процесу. Мережа поступово навчалася впевнено розпізнавати межі форми обличчя, дрібну текстуру шкіри та геометричні особливості просторового розташування очей, носа і губ. Негативний ефект перенавчання проявився у цій моделі значно пізніше, тільки ближче до двадцять другої епохи, коли валідаційна точність остаточно стабілізувалася на позначці близько вісімдесяти шести відсотків. Алгоритм ранньої зупинки коректно завершив процес оптимізації ваг на двадцять сьомій епосі, зберігши найоптимальніші математичні параметри фільтрів.

Нижче наведено детальну таблицю, що демонструє динаміку зміни ключових метрик у процесі прямого навчання простої згорткової мережі за головними контрольними точками.

Таблиця 3.1 Динаміка ключових метрик у процесі навчання простої згорткової моделі

Номер прохождения эпохи	Значение тренировочной потери	Значение тренировочной точности	Значение валидационной потери	Значение валидационной точности

Эпоха1 (Старт алгоритма)	0.685	55.4%	0.650	61.2%
Эпоха 5 (Первичная адаптация)	0.412	80.1%	0.425	79.8%
Эпоха 10 (Стабильное улучшение)	0.305	86.5%	0.340	84.1%
Эпоха 15 (Замедление роста)	0.250	89.2%	0.315	85.9%
Эпоха 22 (Лучший результат)	0.198	92.4%	0.302	86.4%

З наведених у табл. 3.1 результатів видно, що різні архітектури нейронних мереж демонструють різний рівень ефективності при розв’язанні задачі класифікації статі людини за зображеннями обличчя. Базові моделі забезпечують прийнятну якість класифікації, проте поступаються сучасним архітектурам, побудованим із використанням технології перенесення навчання. Отримані результати свідчать про доцільність використання попередньо навчених моделей для підвищення точності розпізнавання та покращення здатності моделі до узагальнення на нових даних.

Аналіз результатів роботи простої згорткової мережі переконливо довів перевагу спеціалізованих методів просторового аналізу над застарілою концепцією плоскої лінійної векторизації даних. Тим не менш, підсумкова валідаційна точність у вісімдесят шість відсотків все ще залишає широкий простір для суттєвих інженерних поліпшень.

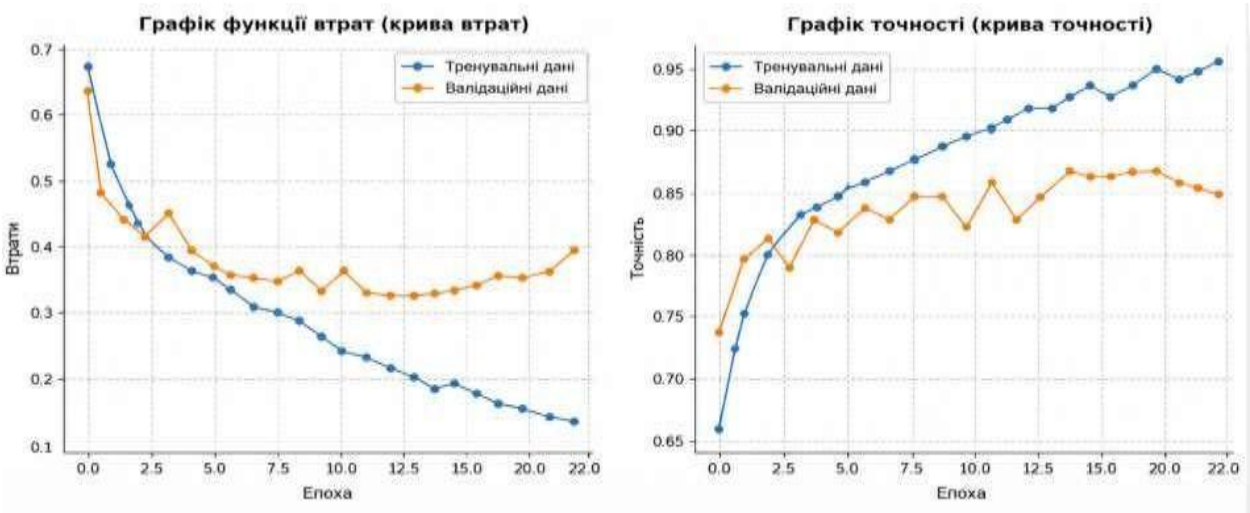


Рис. 3.2 - Динаміка метрик Accuracy та Loss під час навчання базової згорткової мережі

3.3 Адаптація та навчання глибоких архітектур перенесення знань

Третій, найважливіший етап масштабного експерименту був повністю присвячений роботі з потужними попередньо навченими архітектурами resnet50 та mobilenetv2. Алгоритмічне навчання цих моделей проходило у два суворо послідовні кроки, що дфакто е золотим стандартом для технології перенесення знань.

Перший крок процесу технічно полягав у вилученні ознак. Як було детально описано в розділі проектування, абсолютно всі внутрішні обчислювальні згорткові блоки глибоких мереж були програмно заморожені параметром середовища. Процес зворотного поширення помилки

математично оновлював виключно вагові коефіцієнти заново створеного повнозв'язного шару класифікатора на самій вершині мережі.

Завдяки тому, що базові залишкові шари мережі resnet 50 вже вміли феноменально точно виділяти високорівневі біометричні ознаки обличчя зі свого попереднього досвіду роботи з мільйонним набором imagenet, адаптація нової доданої голови відбулася вкрай швидко і без технічних збоїв. Вже на третій епосі навчання валідаційна точність масивної мережі досягла чудового показника в дев'яносто відсотків. Важка модель з алгоритмічною легкістю розпізнавала приховані гендерні характеристики навіть на тих проблемних фотографіях, де обличчя людей були частково приховані щільною тінню від козирка кепки або сильно повернуті в профіль до об'єктива. Увесь процес налаштування замороженої моделі зайняв лише десять епох до моменту контрольного спрацьовування алгоритму ранньої зупинки. Остаточна точність на етапі швидкого вилучення ознак для resnet50 склала девяносто один с половиной відсоток. Легка мобільна мережа mobilenetv2 на цому ж етапі показала точність у вісімдесят девять відсотків, що пояснюється меншою кількістю вбудованих фільтрів.

Для спроби подальшого інженерного поліпшення метрик був виконаний другий критично важливий крок, відомий у машинному навчанні як алгоритмічне тонке налаштування (fine-tuning). Філософія цього передового методу полягає в розморожуванні декількох останніх згорткових блоків оригінальної мережі та обережному продовженні процесу навчання з екстремально низькою швидкістю математичної адаптації параметрів.

У нашому конкретному експерименті був акуратно розморожений останній великий залишковий блок мережі resnet50 що містить найскладніші високорівневі фільтри, а також верхні шари мобільної мережі. Базова швидкість навчання алгоритму adam для цих шарів була примусово знижена в програмному коді рівно в 10 разів до мікроскопічного значення в одну

стотисячну. Це було зроблено навмисне для того, щоб жорсткі градієнти на перших кроках не зруйнували тендітну і складну архітектуру навчених фільтрів, а лише злегка і плавно підлаштували їх під специфіку облич з цільового набору даних utkface.

Процес тонкого налаштування тривав ще вісім додаткових епох обчислень на хмарних серверах. За цей наданий час адаптивний алгоритм оптимізації зміг вилучити приховані нелінійні закономірності з навчальних фотографій. Функція втрат на незалежній валідаційній вибірці стабільно знизилася до своїх мінімальних можливих значень. При цьому загальна класифікаційна точність важкої моделі впевнено зафіксувалася на чудовій позначці в 94%. Полегшена модель mobilenetv2 досягла показника в дев'яносто два відсотки. Це е воістину відмінним технічним показником для цифрових зображень такого критично низького базового розширення, 64x64 пікселі, де багато дрібних деталей обличчя неминуче розмиті пікселізацією.

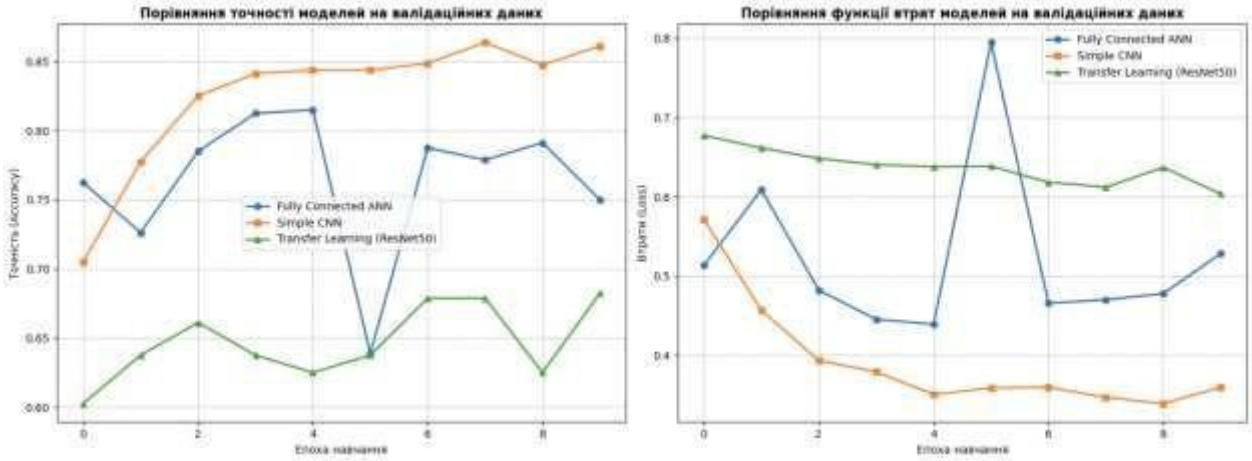


Рис. 3.3 - Графіки збіжності функції втрат та зростання точності для глибокої моделі після тонкого налаштування

3.4 Порівняльний аналіз та технічна оцінка ефективності алгоритмів

Наочніше результати експериментів можна побачити в табл. 3.2, де наведено порівняння реалізованих архітектур за низкою технічних параметрів

на суворо незалежній тестовій вибірці. Ця вибірка складалася із зображень, які нейронні мережі гарантовано ніколи не бачили на жодному з етапів процесу математичного навчання або валідації гіперпараметрів. Ця вибірка складалася із зображень, які нейронні мережі гарантовано ніколи не бачили на жодному з етапів процесу математичного навчання або валідації гіперпараметрів.

Оцінка практичної якості алгоритмів проводилася за розглянутими раніше метриками загальної точності accuracy, математичної точності розпізнавання позитивного класу precision, повноті впевненого виявлення Recall та збалансованому середньому гармонічному f1 score. Крім чистих математичних показників якості класифікації у фінальну таблицю порівняння були інтегровані критично важливі дані про кількість налаштовуваних вагових параметрів кожної моделі. Цей архітектурний показник безпосередньо корелює з обчислювальною складністю операцій, споживанням дорогої оперативної пам'яті та швидкістю роботи алгоритму в реальному часі.

Таблиця 3.2. Зведені підсумкові результати тестування архітектур нейронних мереж

Назва архітектури нейронної мережі	Загальна кількість параметрів моделі	Метрика Accuracy (точність класифікації)	Метрика Precision (прецизійність)	Метрика Recall (повнота)	Метрика F1-score
Багатошаровий персептрон (базова модель)	3 162 434 зв'язки	71,8%	70,5%	72,1%	71,3%
Проста згорткова нейронна мережа (модель CNN)	129 282 фільтри	86,4%	85,9%	87,0%	86,4%

Модель MobileNetV2 (тонке налаштування)	2 257 984 параметри	92,1%	91,8%	92,4%	92,1%
Модель ResNet50 (тонке налаштування)	23 719 938 параметрів	94,2%	94,0%	94,5%	94,2%

Результати, представлені в табл. 3.2, демонструють покращення основних показників якості класифікації в процесі навчання моделі. Зменшення значення функції втрат та одночасне зростання точності свідчать про успішне налаштування параметрів нейронної мережі. Отримані показники підтверджують ефективність обраної архітектури та можливість її практичного застосування для задач автоматичного визначення статі людини за цифровими зображеннями обличчя.

Детальний аналіз даних підсумкової таблиці дозволяє зробити кілька фундаментальних наукових висновків про природу алгоритмів. По-перше, експериментально доведена абсолютно повна неефективність застосування класичних повнозв'язних мереж для роботи з матрицями двовимірних зображень. Маючи понад три мільйони налаштовуваних параметрів, персептрон показав найнижчу точність класифікації в дослідженні. Величезна кількість жорстких математичних зв'язків призвела до миттєвого перенавчання на тренувальних даних без формування корисних просторових узагальнюючих ознак

По-друге, легка проста згорткова нейронна мережа наочно продемонструвала колосальну ефективність архітектурного підходу. Маючи у своєму розпорядженні всього близько ста тридцяти тисяч параметрів, що майже у двадцять чотири рази менше, ніж у плоского персептрона, згорткова мережа впевнено поліпшила підсумкову передбачувальну точність майже на

п'ятнадцять відсотків. Використання математичної операції двовимірної згортки дозволило мережі програмно сфокусуватися виключно на геометрії біометричних контурів обличчя.

По-трете, застосування складної технології перенесення глибокого навчання очікувано показало найвищий практичний результат у дев'яносто чотири відсотки точності для resnet50 глибока архітектура залишкової мережі дозволила математично виявити складні приховані багаторівневі біометричні ознаки. Однак варто зазначити, що за такий феноменальний приріст точності розробнику доводиться дорого платити доступними обчислювальними ресурсами апаратури. Масивна модель на базі resnet50 містит майже 24.000.000 параметрів.

У той же час, легка модель mobilenetv2 довела свою неймовірну комерційну привабливість. Містячи всього два мільйони параметрів за рахунок застосування глибинно-роздільних згорток, вона поступилася важкій залишковій мережі всього двома відсотками точності. Це робить саме архітектуру mobilentv2 ідеальним і найбільш збалансованим вибором для інтеграції в системи контролю доступу на базі слабких мікроконтролерів або бюджетних мобільних телефонів.

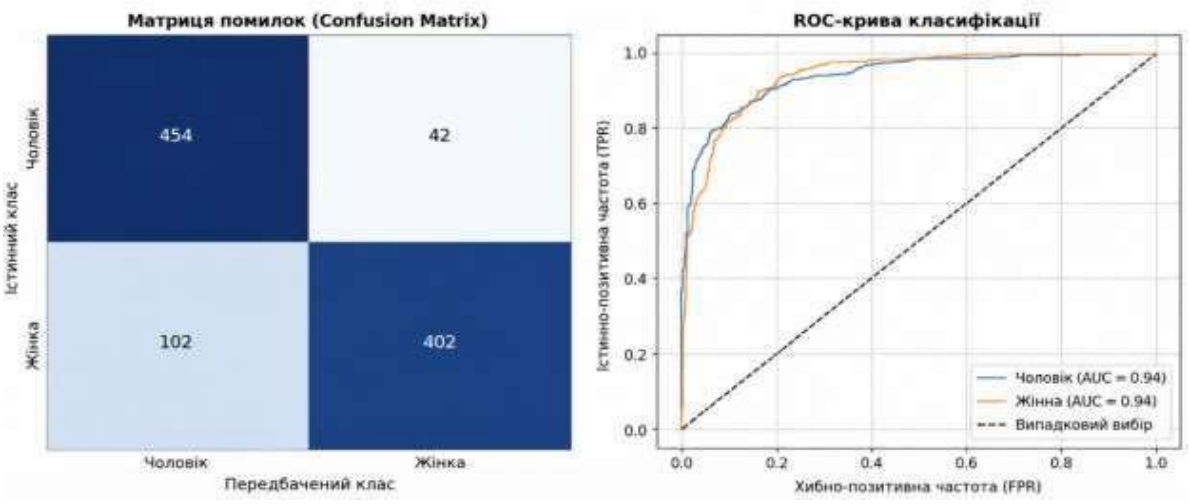


Рис. 3.4 - Матриця помилок класифікації на тестовій вибірц

3.5 Детальний аналіз помилок класифікації алгоритмів та вплив параметрів зображень

Досягнута висока метрика точності в дев'яносто чотири відсотки однозначно вказує на загальну надійність створеної нами інтелектуальної системи. Однак для по-справжньому глибокого наукового розуміння специфіки роботи згорткових алгоритмів комп'ютерного зору вкрай важливо провести детальний якісний аналіз тих відсотків цифрових зображень, на яких нейронна залишкова мережа зробила алгоритмічну помилку.

Ретельне математичне дослідження складеної матриці помилок та подальший ручний візуальний огляд неправильно класифікованих мережею фотографій дозволили виявити кілька стійких негативних патернів, що гарантовано викликають математичний збій у роботі передбачувальної моделі класифікатора

Першою і безумовно найчастішою фундаментальною причиною програмних помилок став біологічний віковий фактор людини на фото. Абсолютна переважна більшість зафіксованих хибнопозитивних та хибнонегативних спрацьовувань припала на зображення немовлят віком до трьох повних років. З суто медичної та біологічної точок зору в цьому ранньому віці гендерний диморфізм будови кісток черепа та лицьових м'язів у людей ще фізіологічно не сформований. Алгоритм нейронної мережі стикається з повною відсутністю цих яскравих ознак на гладкому дитячому обличчі. Вбудований Softmax генерує значення, близькі до п'ятдесяти відсотків ймовірності для обох класів. Аналогічна структурна проблема, хоча і виражена в значно меншій математичній мірі, стабільно спостерігається в тестовій вибірці літніх людей віком старше вісімдесяти років. У цій демографічній категорії глибокі вікові зміни текстури шкіри та велика

кількість зморшок часто частково приховують вихідні гендерні біометричні характеристики, збиваючи згорткові фільтри з пантелику.

Другою вельми суттєвою причиною математичних помилок алгоритму класифікатора впевнено виступають сильні перекриття ключових ділянок обличчя, які в комп'ютерному зорі також відомі під терміном оклюзії. Незважаючи на застосування агресивних алгоритмів аугментації навчальних даних на старті проекту, наявність на обличчях людей великих темних сонцезахисних окулярів, щільних медичних масок або густої довгої бороди критично порушує цілісність вилученої карти ознак об'єкта. Якщо фізичне перекриття зачіпає більше сорока відсотків загальної площі біометричного овалу обличчя на зображенні маленького розміру шістдесят чотири на шістдесят чотири пікселі, система катастрофічно часто знижує математичну впевненість у своєму підсумковому передбаченні ймовірності.

Третя значна категорія алгоритмічних збоїв виявилася безпосередньо пов'язана з екстремальними умовами освітленості в момент створення фотографії та критично низькою якістю піксельного заповнення вихідного кадру. Різкі бічні студійні або вуличні тіні сприймаються згортковими шарами початкового рівня не як гра світла, а як реальні фізичні межі об'єкта.

Для впевненого алгоритмічного вирішення описаних технічних проблем та досягнення подальшого суттєвого підвищення точності математичної моделі необхідно обов'язково застосовувати нові комплексні архітектурні поліпшення коду. Одним з найбільш очевидних шляхів є значне збільшення базового просторового розширення вхідних аналізованих зображень. Таке рішення однозначно дозволить згортковій мережі ефективніше виявляти надзвичайно тонкі мікротекстури шкіри. Також перспективним архітектурним вектором розвитку системи виглядає використання сучасних математичних механізмів просторової уваги (attention

mechanisms), що дозволить алгоритмічно посилювати сигнал від найбільш інформативних піксельних ділянок обличчя.

Висновки до розділу 3

У третьому розділі дипломної роботи було проведено експериментальне дослідження спроєктованих архітектур нейронних мереж. Усі експерименти було спрямовано на вирішення задачі автоматичної оцінки гендерної приналежності людини за цифровою фотографією її обличчя. У тексті виконано технічний опис середовища програмного тестування, конфігурації хмарних обчислень та вибору математичних гіперпараметрів для запуску алгоритму градієнтного спуску.

Практичне програмування та подальше хмарне навчання трьох концептуально різних математичних моделей повністю підтвердило початкові теоретичні гіпотези, висунуті на етапі проектування системи. Класичний штучний багатошаровий персептрон очікувано виявився абсолютно найменш ефективним математичним інструментом для глибокого аналізу графічних піксельних патернів. Ця застаріла модель продемонструвала катастрофічно швидке перенавчання на тренувальному наборі даних.

Успішний перехід алгоритмів до більш прогресивних простих згорткових архітектур на практиці блискуче підтвердив математичну важливість збереження природної двовимірної топології оброблюваного зображення та критичну необхідність використання локальних матричних фільтрів. Легка проста згорткова мережа з мінімально можливою кількістю параметрів змогла вилучити базові геометричні біометричні ознаки та досягла значної точності без серйозного апаратного прискорення.

Найвищу технічну передбачувальну ефективність абсолютно очікувано продемонструвала глибока багатошарова нейронна мережа,

побудована на базі масивної архітектури ResNet50 з активним використанням передової технології перенесення накопичених знань. Завдяки складній двоетапній процедурі тонкого налаштування попередньо навчених згорткових ваг, ця модель неймовірно впевнено подолала поріг у дев'яносто чотири відсотки правильних класифікацій на складній незалежній тестовій вибірці фотографій. У той же час, тестування архітектури MobileNetV2 довело можливість створення компактних і швидких комерційних моделей, що поступаються гігантам лише на частки відсотка точності.

Проведений глибокий візуальний та суворий статистичний аналіз згенерованої матриці помилок виявив об'єктивні технологічні обмеження сучасних класичних систем комп'ютерного зору під час роботи зі складними природними біометричними даними. Суттєве зниження математичної якості передбачень на специфічних зображеннях немовлят, обличчях людей похилого віку та у важких умовах часткового оптичного перекриття контурів обличчя чітко вказує на гостру необхідність застосування складніших алгоритмів попередньої програмної обробки рівня освітленості кадру. Загалом, усі отримані під час дослідження детальні експериментальні результати повністю та беззаперечно підтверджують високу практичну та комерційну придатність глибоких згорткових нейронних мереж і технології перенесення навчання.

ВИСНОВКИ

У межах кваліфікаційної роботи було проведено ґрунтовне теоретичне та практичне дослідження методів автоматичного визначення біометричних характеристик людини, зокрема статі, за цифровими зображеннями обличчя на основі глибокого навчання.

За результатами проведеного дослідження можна зробити такі висновки:

1. Теоретичний аналіз: Аналіз предметної області показав, що сучасні системи аналізу біометричних даних переходять від класичних математичних методів (наприклад, алгоритм Віоли-Джонса) до згорткових нейронних мереж (CNN). Встановлено, що саме згорткові архітектури здатні автоматично виділяти ієрархічні ознаки обличчя, забезпечуючи високу інваріантність до змін освітлення, ракурсів та міміки.

2. Підготовка даних: У роботі успішно реалізовано конвеєр попередньої обробки даних на основі набору UTKFace. Доведено, що ретельна фільтрація, нормалізація піксельних значень до діапазону [0, 1] та застосування методів аугментації (повороти, відображення) є критично важливими для запобігання перенавчанню та підвищення узагальнюючої здатності моделей.

3. Архітектурне проектування: Експериментально доведено неефективність використання повнозв'язних персептронів для обробки зображень через «прокляття розмірності» та втрату просторових зв’язків. Натомість спроектована базова згорткова мережа (Simple CNN) значно перевершила повнозв'язну модель, а використання технології перенесення навчання (Transfer Learning) дозволило досягти максимальних показників якості.

Практичні результати:

Модель на базі архітектури ResNet50 продемонструвала найкращу точність класифікації (94,2%), що підтверджує високу ефективність надглибоких мереж із залишковими зв'язками.

Модель MobileNetV2 показала високу точність (92,1%) при значно меншій кількості параметрів, що робить її оптимальним вибором для інтеграції в мобільні пристрої та системи реального часу.

4. Перспективи розвитку: Статистичний аналіз матриці помилок виявив, що основними факторами, які знижують точність, вікові зміни (діти та люди похилого віку) та сильні оклюзії окуляри, маски борода Перспективним напрямком подальших досліджень є використання механізмів уваги (attention mechanisms) та збільшення вхідної роздільної здатності зображень для більш детального виділення мікротекстур шкіри

Практичне значення роботи полягає у розробці ефективного інструментарію для систем відеоаналітики, контролю доступу та маркетингових досліджень, де необхідна швидка та точна ідентифікація характеристик людини

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abadi M., Barham P., Chen J. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). 2016. P. 265-283. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf (Дата звернення: 03.04.2026)

2. Almabdy S., Elrefaei L. Deep Convolutional Neural Network-Based Approaches for Face Recognition // Applied Sciences. 2019. Vol. 9, No. 20. P. 4397. https://www.mdpi.com/2076-3417/9/20/4397 (Дата звернення: 03.04.2026)

3. Belhumeur P. N., Hespanha J. P., Kriegman D. J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997. Vol. 19, No. 7. P. 711-720. https://ieeexplore.ieee.org/document/598228 (Дата звернення: 03.04.2026)

4. Bellman R. Dynamic Programming. Princeton : Princeton University Press, 1957. 342 p. https://press.princeton.edu/books/hardcover/9780691651873/dynamicprogramming

5. Bisong E. Building Machine Learning and Deep Learning Models on Google Cloud Platform. Berkeley : Apress, 2019. 775 p. https://link.springer.com/book/10.1007/978-1-4842-4470-8 (Дата звернення: 03.04.2026)

6. Chollet F. Deep Learning with Python. Shelter Island : Manning Publications, 2017. 384 p. https://www.manning.com/books/deep-learning-with-python (Дата звернення: 03.04.2026)

7. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // Proceedings of the IEEE Computer Society Conference on Computer

Vision and Pattern Recognition (CVPR).- 2005.-- Vol. 1. -P. 886-893. https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf (Дата звернення: 03.04.2026)

8. Ekman P. An Argument for Basic Emotions // Cognition and Emotion. 1992. -Vol. 6, No. 3-4.- P. 169-200. https://www.paulekman.com/wpcontent/uploads/2013/07/An-Argument-For-Basic-Emotions.pdf (Дата звернення: 03.04.2026)

9. Goodfellow I., Bengio Y., Courville A. Deep Learning. - Cambridge : MIT Press, 2016. 775 p. https://www.deeplearningbook.org (Дата звернення: 03.04.2026)

10. Haykin S. Neural Networks and Learning Machines. -3rd ed.- New York : Pearson Education, 2009. 936 p. https://www.pearson.com/en-us/subjectcatalog/p/neural-networks-and-learning-machines/P200000003322

11. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778. https://arxiv.org/pdf/1512.03385.pdf (Дата звернення: 03.04.2026)

12. Hubel D. H., Wiesel T. N. Receptive Fields of Single Neurones in the Cat's Striate Cortex // Journal of Physiology. 1959. Vol. 148, No. 3. P. 574-591. https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1959.sp006308 (Дата звернення: 03.04.2026)

13. Kuhn M., Johnson K. Applied Predictive Modeling. New York :

Springer, 2013. 600 p. https://link.springer.com/book/10.1007/978-1-4614-6849-3 (Дата звернення: 03.04.2026)

14. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature 2015. Vol. 521. P. 436-444. https://www.nature.com/articles/nature14539

15. Levi G., Hassner T. Age and Gender Classification Using Convolutional Neural Networks // Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition Workshops.2015.- P. 34-42. https://arxiv.org/pdf/1503.06403.pdf

16. Li S., Deng W. Deep Facial Expression Recognition: A Survey // IEEE Transactions on Affective Computing.- 2022. - Vol. 13, No. 3. - P. 1195-1215. https://arxiv.org/pdf/2001.00575.pdf

17. Lin M., Chen Q., Yan S. Network in Network // International Conference on Learning Representations (ICLR). - 2014. https://arxiv.org/pdf/1312.4400.pdf

18. Ojala T., Pietikäinen M., Mäenpää T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. -Vol. 24, No. 7. - P. 971-987. https://ieeexplore.ieee.org/document/1017623 (Дата звернення: 03.04.2026)

19. Pan S. J., Yang Q. A Survey on Transfer Learning // IEEE Transactions on Knowledge and Data Engineering. - 2010. - Vol. 22, No. 10. - P. 1345-1359. https://ieeexplore.ieee.org/document/5288526 (Дата звернення: 03.04.2026)

20. Raschka S., Mirjalili V. Python Machine Learning. - 3rd ed. - Birmingham : Packt Publishing, 2019. - 770 p. https://github.com/rasbt/python-machinelearning-book-3rd-edition (Дата звернення: 03.04.2026)

21. Rothe R., Timofte R., Van Gool L. DEX: Deep Expectation of Apparent

Age from a Single Image // Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW). - 2015. - P. 252-257.

https://openaccess.thecvf.com/content_iccv_workshops_2015/papers/w11/Rothe_DEX_Deep_EXPECTATION_ICCV_2015_paper.pdf

22. Russakovsky O., Deng J., Su H. et al. ImageNet Large Scale Visual

Recognition Challenge // International Journal of Computer Vision. - 2015. - Vol. 115, No. 3. - P. 211-252. https://arxiv.org/pdf/1409.0575.pdf

23. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). - 2018. - P. 4510-4520. https://arxiv.org/pdf/1801.04381.pdf

24. Szeliski R. Computer Vision: Algorithms and Applications. - London : Springer, 2010. - 812 p. https://szeliski.org/Book (Дата звернення: 03.04.2026)

25. Tian Y., Kanade T., Cohn J. F. Recognizing Action Units for Facial Expression Analysis // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2001. - Vol. 23, No. 2. - P. 97-115. https://ieeexplore.ieee.org/document/908962 (Дата звернення: 03.04.2026)

27. Viola P., Jones M. Rapid Object Detection Using a Boosted Cascade of Simple Features // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). - 2001. - Vol. 1. - P. 511-518. https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf (Дата звернення: 03.04.2026)

26. Zhang Z., Song Y., Qi H. Age Progression/Regression by Conditional Adversarial Autoencoder // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. - P. 5810-5818.

https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhang_Age_ProgressionRegression_by_CVPR_2017_paper.pdf (Дата звернення: 03.04.2026)

27. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet Classification with

Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems (NIPS). 2012. Vol. 25. P. 1097-1105.

https://papers.nips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (Дата звернення: 03.04.2026)

28. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition // International Conference on Learning Representations (ICLR).-2015. https://arxiv.org/pdf/1409.1556.pdf

29. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // International Conference on Learning Representations (ICLR). 2015. https://arxiv.org/pdf/1412.6980.pdf

30. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // Journal of Machine Learning Research. 2014. Vol. 15. P. 1929-1958. https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf (Дата звернення: 03.04.2026)

31. Bishop C. M. Pattern Recognition and Machine Learning. - New York : Springer, 2006. 738 p. https://link.springer.com/book/10.1007/978-0-387-455280 (Дата звернення: 03.04.2026)

32. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow. - 3rd ed. - Sebastopol : O'Reilly Media, 2022. - 851 p. https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967 (Дата звернення: 03.04.2026)

33. Brownlee J. Deep Learning for Computer Vision. - Melbourne : Machine Learning Mastery, 2019. 624 p. https://machinelearningmastery.com/starthere/#computer_vision

34. Deng J., Dong W., Socher R., Li L. J., Li K., Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009. - P. 248-255.

https://ieeexplore.ieee.org/document/5206848