МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ І ПРИКЛАДНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗЕЛІНСЬКИЙ ОЛЕКСАНДР ОЛЕКСАНДРОВИЧ

Допускається до захисту:
в.о. завідувача кафедри
інформаційних технологій,
д-р. техн. наук, професор
_____Наталія ВЕСЕЛОВСЬКА
«____»_____ 2025 р.

ІДЕНТИФІКАЦІЯ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

Спеціальність 122 Комп’ютерні науки

Кваліфікаційна (магістерська) робота

Керівник:
Тетяна СІЧКО, доцент кафедри
інформаційних технологій,
к. т. н., доцент

_____

Оцінка: _____ / _____ / _____
(бали/за шкалою ЄКТС/за національною шкалою)

Голова ЕК: _____

Вінниця - 2025

АНОТАЦІЯ

Зелінський О.О. Ідентифікація зображень за допомогою штучного інтелекту. Спеціальність 122 «Комп'ютерні науки». ОП «Комп'ютерні технології обробки даних (Data Science)». Донецький національний університет імені Василя Стуса, Вінниця, 2025.

У роботі досліджено методи автоматичної ідентифікації зображень із використанням глибинних нейронних мереж. Порівняно роботу моделей ResNet50 та Vision Transformer (ViT-B/16) у класифікації зображень, створених людиною та згенерованих ШІ. Розроблено узгоджену модель підготовки та аугментації даних, що підвищує точність і стійкість моделей до шуму та змін освітлення. Реалізовано повний цикл навчання та експорт моделей для подальшого використання в системах з обмеженими ресурсами.

Ключові слова: штучний інтелект, глибинне навчання, ResNet, Vision Transformer, класифікація.

75 ст., 7 рис., 6 табл., 55 джерел.

ABSTRACT

Zelinskyi O.O. Methodological Aspects of Image Identification Using Artificial Intelligence. Specialty 122 "Computer Science". OP "Computer data processing technologies". Vasyl' Stus Donetsk National University, Vinnytsia, 2025.

The thesis examines automatic image identification methods based on deep neural networks. The performance of ResNet50 and Vision Transformer (ViT-B/16) was compared for classifying human-generated and AI-generated images. A unified preprocessing and data-augmentation program was developed to enhance accuracy and robustness to noise and lighting variation. A complete training and model-export pipeline was implemented for integration into resource-constrained systems.

Keywords: artificial intelligence, deep learning, ResNet, Vision Transformer, classification.

75 pages, 7 figures, 6 tables, 55 references.

ЗМІСТ

ВСТУП

Актуальність роботи. У сучасному інформаційному суспільстві обробка та аналіз візуальних даних набувають особливого значення. Щоденно в мережі Інтернет, соціальних медіа, системах відеоспостереження, медичних приладах та промислових процесах генерується колосальний обсяг зображень. За оцінками аналітичних компаній, понад 80% інформації, яку сприймає людина, є візуальною. Це зумовлює необхідність створення автоматизованих систем, здатних швидко, точно та ефективно ідентифікувати зображення.

Традиційні методи комп'ютерного зору, що базувалися на ручному виділенні ознак та класичних алгоритмах машинного навчання, довгий час залишалися основним підходом до вирішення задач ідентифікації. Проте такі методи мали обмежену гнучкість і часто не забезпечували необхідної точності при роботі з великими та різнорідними даними. Суттєвий прорив відбувся завдяки розвитку технологій штучного інтелекту, зокрема глибинного навчання та нейронних мереж, що дало можливість автоматично виявляти закономірності у зображеннях і створювати універсальні моделі для їх розпізнавання.

Особливої актуальності набуває використання методів ШІ у сферах, де точність ідентифікації є критичною: у медичній діагностиці, системах безпеки та біометрії, автономному транспорті, промисловій автоматизації, сільському господарстві та багатьох інших галузях. Помилка системи може мати серйозні соціальні чи економічні наслідки, тому питання методологічних аспектів застосування штучного інтелекту до ідентифікації зображень є надзвичайно важливим.

Крім того, стрімкий розвиток апаратних засобів — графічних процесорів, тензорних обчислювальних модулів, спеціалізованих мікросхем для штучного інтелекту — відкриває нові можливості для створення потужних і в той же час оптимізованих систем розпізнавання. Однак постає потреба у виробленні

методологічних підходів, які б дозволили ефективно поєднувати математичні моделі, алгоритми та апаратні ресурси для вирішення завдань ідентифікації.

Таким чином, дослідження методологічних аспектів ідентифікації зображень за допомогою штучного інтелекту є своєчасним і має важливе значення як для науки, так і для практики.

Зв’язок теми з науковими програмами та сучасними викликами.

Розвиток штучного інтелекту та комп’ютерного зору є одним із ключових напрямів сучасної науки й техніки, що тісно пов’язаний із глобальними тенденціями цифровізації та автоматизації. В умовах четвертої промислової революції (Industry 4.0) ідентифікація зображень виступає необхідною складовою інтелектуальних систем, що здатні працювати автономно, взаємодіяти з фізичним середовищем та приймати рішення без участі людини.

Дослідження у цій сфері відповідають стратегічним завданням розвитку інформаційних технологій, кібербезпеки, охорони здоров’я та транспорту. Застосування методів ШІ у розпізнаванні зображень безпосередньо впливає на якість життя суспільства, рівень безпеки та ефективність виробничих процесів.

Метою магістерської роботи є аналіз ідентифікації зображень за допомогою штучного інтелекту, а також визначення сучасних підходів та їх застосування у сучасних інформаційних системах.

Для досягнення мети у роботі передбачено розв’язати такі завдання:

1. Проаналізувати сучасний стан розвитку методів ідентифікації зображень.
2. Дослідити можливості застосування штучних нейронних мереж, зокрема згорткових і трансформерних архітектур.
3. Визначити переваги та обмеження використання методів глибинного навчання у практичних завданнях.
4. Створити та навчити дві моделі для ідентифікації зображень на основі різних архітектур.

5. Провести порівняльний аналіз розроблених моделей за показниками точності, швидкодії та стійкості до шумів.

Об’єктом дослідження є процеси ідентифікації зображень у комп’ютерних системах.

Предметом дослідження є методи застосування штучного інтелекту для ідентифікації зображень.

Методи дослідження. У процесі виконання роботи використовувалися методи аналізу та синтезу наукової літератури, системний підхід до узагальнення існуючих алгоритмів, а також моделювання й експериментальні дослідження із застосуванням нейронних мереж. Застосовано математичний апарат лінійної алгебри, теорії ймовірностей, оптимізації та методи обчислювального експерименту.

Наукова новизна. Наукова новизна магістерської роботи полягає у розробленні та експериментальній перевірці двох моделей ідентифікації зображень, побудованих на принципово різних архітектурних підходах — згорткових нейронних мережах та візуальних трансформерах. На відміну від існуючих досліджень, де використовуються готові конфігурації або результати сторонніх експериментів, у роботі виконано повний цикл створення моделей: формування датасету, підготовка даних, проєктування архітектур, навчання та оцінювання.

Новизна дослідження полягає у визначенні впливу архітектури моделей на якість ідентифікації за однакових умов навчання та тестування. Отримані результати дозволяють встановити особливості обробки візуальної інформації для різних підходів до побудови нейронних мереж, а також виявити практичні відмінності у точності, швидкодії та стійкості до спотворень вхідних даних.

Запропонований підхід дає можливість сформувати обґрунтовані критерії вибору архітектури для реальних інформаційних систем, що використовують ідентифікацію зображень. Отримані експериментальні висновки можуть бути використані під час проєктування прикладних рішень, де важливо забезпечити

надійну роботу алгоритмів у змінному середовищі — від автоматизації виробничих процесів до систем виявлення штучно згенерованого візуального контенту.

Практичне значення. Результати роботи можуть бути використані у створенні інтелектуальних систем для медицини, транспорту, промислової автоматизації, систем безпеки та інших галузей. Запропоновані методологічні підходи здатні підвищити ефективність роботи систем ідентифікації, зменшити ризик помилкових рішень і сприяти розвитку цифрової економіки України.

Структура роботи. Магістерська робота складається зі вступу, трьох розділів основної частини, висновків та списку використаних джерел. У першому розділі розглянуто теоретичні основи та сучасні підходи до ідентифікації зображень. У другому розділі проведено аналіз методологічних аспектів застосування штучного інтелекту. У третьому розділі подано практичні рекомендації та перспективи використання ШІ в ідентифікації зображень.

Апробація результатів досліджень. Результати кваліфікаційної (магістерської) роботи апробовано на:

1. IV Міжнародній науково-практичній конференції «Прикладні аспекти сучасних міждисциплінарних досліджень» , м. Вінниця, 5 листопада, 2025 р. Вінниця: ДонНУ імені Василя Стуса, 2025. Тези на тему: «Застосування методів штучного інтелекту для ідентифікації зображень» (прийнято до друку).
2. VI Всеукраїнській науково-практичній конференції «Комп’ютерні технології обробки даних (КТОД 2025)», м. Вінниця, 5 грудня 2025 р. Вінниця: ДонНУ імені Василя Стуса, 2025. Тези на тему: «Інтерпретації результатів штучного інтелекту при ідентифікації зображень» (прийнято до друку).

РОЗДІЛ 1
ТЕОРЕТИЧНІ ОСНОВИ ІДЕНТИФІКАЦІЇ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

1.1. Поняття ідентифікації зображень та її значення

Ідентифікація зображень належить до ключових напрямів розвитку сучасних технологій штучного інтелекту та комп'ютерного зору. Під цим поняттям розуміють процес визначення, класифікації або розпізнавання об'єктів, сцен і структур на зображеннях за допомогою математичних алгоритмів та обчислювальних моделей. Основна мета цього процесу полягає у перетворенні візуальної інформації у форму, придатну для автоматизованого аналізу, з подальшим віднесенням знайдених об'єктів до певних класів або категорій [3, 7, 25, 27]. Ідентифікація є фундаментальною складовою будь-якої системи комп'ютерного зору і часто виступає першим кроком у більш складних аналітичних або когнітивних процесах.

Перші спроби автоматичної обробки візуальної інформації з'явилися у другій половині XX століття. Тоді методи базувалися на класичних підходах комп'ютерного зору, які включали ручне виділення ознак, аналіз контурів, використання гістограм та алгоритмів шаблонного порівняння. Значного поширення набули методи визначення ключових точок, такі як SIFT та SURF, що забезпечували стійкість до змін масштабу, поворотів та освітлення. Однак ці підходи були обмежені складністю ручного підбору ознак та залежністю від умов зйомки, що обмежувало їхню універсальність [3, 9, 25, 27, 28, 29].

Справжній прорив відбувся на початку 2010-х років із впровадженням технологій глибинного навчання. Згорткові нейронні мережі (CNN) стали основним інструментом розпізнавання візуальних об'єктів, забезпечивши суттєве підвищення точності та адаптивності систем. Подальший розвиток привів до появи трансформерних моделей (Vision Transformers, ViT), які

завдяки здатності обробляти великі обсяги даних та враховувати контекстні зв’язки підняли ідентифікацію на якісно новий рівень.

Варто зазначити, що сучасні системи поєднують як класичні підходи, так і сучасні методи глибинного навчання, створюючи гібридні моделі, здатні ефективно працювати у широкому спектрі умов: від змінного освітлення до складних фонових структур [25, 27, 28, 29].

У цифрову епоху візуальна інформація є одним із найпоширеніших типів даних. Щодня створюються мільйони зображень — від фотографій у соціальних мережах до даних із промислових камер спостереження чи медичного обладнання. Ручний аналіз таких обсягів даних практично неможливий, тому автоматизовані системи ідентифікації є незамінними [7, 11, 13, 17, 20, 25, 27].

У медичній сфері системи автоматично виявляють патології на знімках КТ, МРТ та рентгенологічних досліджень, підвищуючи точність діагностики та зменшуючи навантаження на лікарів. У сфері безпеки технології ідентифікації дозволяють розпізнавати обличчя, оцінювати поведінку людей і оперативно реагувати на потенційні загрози. У промисловості ці системи забезпечують контроль якості продукції, виявлення дефектів та моніторинг технологічних процесів. У сфері автономного транспорту ідентифікація об’єктів дозволяє системам самостійно визначати дорожні знаки, пішоходів та транспортні засоби, забезпечуючи безпечне пересування [7, 11, 13, 17, 20, 25, 27].

Розглянемо методологічні аспекти ідентифікації зображень. Процес ідентифікації зображень складається з кількох ключових етапів:

1. Попередня обробка — нормалізація освітлення, усунення шумів, підвищення контрасту та приведення зображень до єдиного формату. Це покращує якість вхідних даних і забезпечує більш стабільне навчання моделей.

2. Виділення ознак — перетворення зображення у набір характеристик, які несуть інформацію про його структуру. У сучасних системах глибинного навчання цей процес автоматизується, що значно підвищує ефективність моделей.
3. Навчання моделі — алгоритми машинного навчання виявляють закономірності в навчальних даних та застосовують їх для прогнозування на нових прикладах.
4. Оцінювання результатів — використання метрик (точність, F1-міра, повнота) для кількісного аналізу ефективності системи.

Якість ідентифікації значною мірою залежить від обсягу та коректності даних. Неповні, зашумлені або неправильно анотовані вибірки знижують ефективність моделей. Тому ретельне очищення, збалансування та анотація даних є невід'ємною складовою методологічного процесу [8, 15, 16, 25, 27, 28].

Сучасні дослідження спрямовані на створення систем, здатних працювати у реальному часі та інтегруватися у повсякденні технологічні процеси. Edge AI переносить обчислення безпосередньо на пристрої користувача, зменшуючи затримки та підвищуючи безпеку. Також активно розвивається застосування ідентифікації у доповненій (AR) і віртуальній реальності (VR), де система має взаємодіяти з об'єктами у реальному середовищі. Паралельно зростає увага до етичності та захисту персональних даних, оскільки технології розпізнавання дедалі більше впливають на приватне життя людини [27, 29, 31].

Отже ідентифікація зображень є невід'ємною частиною систем штучного інтелекту та основою для створення розумних рішень у різних галузях. Вона поєднує математичні методи, машинне навчання та комп'ютерний зір для ефективного аналізу візуальної інформації. Від якості даних, побудови моделі та точності оцінювання результатів залежить ефективність таких систем. З огляду на темпи розвитку технологій, ідентифікація зображень продовжуватиме відігравати стратегічну роль у медицині, промисловості,

транспорті, безпеці та наукових дослідженнях, стаючи все більш універсальною і потужною складовою цифрового середовища.

1.2. Методи комп'ютерного зору

Комп'ютерний зір є міждисциплінарною галуззю, що поєднує математичні методи, алгоритми обробки сигналів і підходи штучного інтелекту для автоматичного аналізу та інтерпретації візуальної інформації. Основне завдання цієї сфери полягає у створенні систем, здатних «сприймати» зображення та розуміти їхній зміст, подібно до людського зору. Комп'ютерний зір не обмежується простою класифікацією об'єктів: він включає в себе розпізнавання сцен, відстеження руху, оцінку глибини, реконструкцію 3D-моделей і навіть прогнозування динаміки об'єктів [3, 9, 25, 27]. Одним із центральних напрямів комп'ютерного зору є ідентифікація об'єктів, що передбачає виявлення, розпізнавання та класифікацію візуальних об'єктів за допомогою різних алгоритмічних підходів.

Розгляньмо класичні методи обробки зображень. Початкові етапи розвитку комп'ютерного зору ґрунтувалися на використанні традиційних математичних та статистичних алгоритмів, які не потребували попереднього навчання. Ці методи передбачали ручне формулювання правил та виділення ознак для розпізнавання об'єктів. Основні серед них:

- обробка контурів і країв. Алгоритми Canny, Sobel і Laplace дозволяли виділяти границі об'єктів шляхом аналізу різких змін яскравості. Ці методи добре працювали у стабільних умовах, проте були чутливими до шумів, тіней та масштабних трансформацій. Водночас вони слугували надійною основою для попередньої підготовки зображень перед подальшим аналізом [3, 9];
- гістограми та колірні простори. Аналіз розподілу кольорів у просторах RGB, HSV або Lab дозволяв виконувати сегментацію об'єктів і

класифікацію за кольоровими характеристиками, що було особливо корисним у промислових системах контролю якості;

- виділення ключових точок. Методи SIFT (Scale-Invariant Feature Transform) і SURF (Speeded-Up Robust Features) виявляли унікальні точки зображення, стійкі до змін масштабу, поворотів та освітлення. Це давало змогу порівнювати та розпізнавати об’єкти навіть у складних умовах зйомки [3, 9, 25, 27, 28].

Перевагою класичних методів була їхня простота, невелика потреба в обчислювальних ресурсах і швидкість реалізації. Проте вони демонстрували обмежену точність і погано працювали у випадках складних сцен, різного освітлення або перекриття об’єктів.

З розвитком обчислювальних можливостей і зростанням обсягів даних у комп’ютерному зорі почали активно застосовуватися алгоритми машинного навчання. Вони дозволили автоматизувати процес побудови класифікаційних моделей і знаходження закономірностей у даних:

- метод опорних векторів (SVM). Використовується для класифікації об’єктів на основі ознак, виділених із зображень. Метод формує оптимальну гіперплощину, що розділяє класи, і забезпечує високу точність при малих вибірках;
- метод k-Nearest Neighbors (kNN). Ґрунтується на порівнянні нового зображення з найближчими зразками в навчальній вибірці. Підхід простий, але має обмежену масштабованість для великих наборів даних через високу обчислювальну складність;
- Random Forest. Ансамбль рішень, що поєднує результати кількох дерев класифікації, забезпечуючи стійкість до шуму та переобучення;

Хоча застосування машинного навчання значно підвищило точність ідентифікації об’єктів, ці методи залишалися залежними від ручного виділення ознак, що обмежувало їхню універсальність [8, 15, 16, 25, 27, 28].

Справжній прорив у комп'ютерному зорі відбувся із впровадженням глибинного навчання, що дозволило відмовитися від ручного формування ознак. Нейронні мережі самостійно навчаються виділяти ключові елементи зображення на основі великої кількості прикладів:

- згорткові нейронні мережі (CNN). Використовують згорткові шари для аналізу локальних структур. CNN автоматично виділяють патерни різного рівня складності — від країв до об'єктів високого рівня;
- регіональні мережі (R-CNN, Fast R-CNN, Mask R-CNN). Поєднують виявлення об'єктів та їхню класифікацію, визначаючи точні межі розташування об'єктів на зображенні;
- Vision Transformers (ViT). Аналізують зображення як послідовність фрагментів (patches) і виявляють глобальні залежності між ними за допомогою механізму самоуваги (self-attention).

Завдяки цим моделям точність ідентифікації зображень у деяких завданнях перевищує людське сприйняття. Такі методи активно використовуються у медичній діагностиці, автономному керуванні транспортом, промисловій інспекції та системах відеоспостереження [7, 11, 13, 17, 20].

Одним із важливих напрямів сучасного комп'ютерного зору є сегментація та детекція об'єктів:

- сегментація зображень передбачає поділ сцени на області, що відповідають певним класам об'єктів. Розрізняють semantic segmentation, коли кожен піксель належить до певного класу, та instance segmentation, де виділяються окремі екземпляри об'єктів;
- детекція об'єктів полягає у визначенні місця розташування об'єктів і їхньої класифікації. Для цього широко використовуються моделі CNN, R-CNN та YOLO, що забезпечують обробку зображень у режимі реального часу [25, 27, 29].

Перейдемо до сучасних тенденцій та комбінованих підходів. На сучасному етапі розвитку спостерігається тенденція до поєднання різних методів у єдині гібридні системи. Вони комбінують глибокі нейронні мережі з класичними алгоритмами для підвищення точності при обмежених обчислювальних ресурсах. Водночас активно розвивається напрям self-supervised learning, який дозволяє навчатися без потреби у великих обсягах розмічених даних, спрощуючи процес розробки моделей і відкриваючи нові можливості для досліджень [27, 29, 31].

З цього можна зробити висновок що еволюція методів комп'ютерного зору пройшла шлях від простих математичних моделей до складних моделей глибинного навчання. Сучасні підходи дозволяють автоматично аналізувати великі обсяги візуальної інформації, розпізнавати складні структури та працювати у режимі реального часу. Комбінування традиційних і новітніх методів створює основу для подальшого розвитку систем ідентифікації зображень, що знаходять застосування у медицині, транспорті, промисловості, безпеці та наукових дослідженнях, забезпечуючи високу ефективність та надійність [7, 11, 13, 17, 20, 25, 27, 29].

1.3. Використання нейронних мереж у завданнях ідентифікації зображень

Штучні нейронні мережі (ШНМ) відіграють ключову роль у сучасних системах комп'ютерного зору, оскільки саме вони дозволяють моделювати когнітивні процеси людини під час сприйняття візуальної інформації. Модель таких мереж наслідує принципи роботи біологічного мозку, що забезпечує здатність до навчання на великих обсягах даних та виявлення складних закономірностей у вхідних зображеннях. Основна ідея функціонування ШНМ полягає у поступовому коригуванні вагових коефіцієнтів зв'язків між штучними нейронами з метою мінімізації різниці між прогнозованим та

фактичним результатом, що дозволяє системам поступово покращувати точність прогнозів [6].

Використання нейронних мереж змінило сам підхід до обробки візуальних даних. Раніше ідентифікація об’єктів залежала від ручного виділення ознак, що обмежувало точність і масштабованість. Тепер же мережі здатні автоматично навчатися виявляти ключові патерни та закономірності, які складно формалізувати людським розумом.

Поява CNN стала визначальним етапом у розвитку комп’ютерного зору. Ці мережі автоматизували процес виділення ознак, який раніше вимагав ручного програмування. Типова структура CNN включає:

- згорткові шари (Convolutional Layers), що виявляють локальні особливості зображення, такі як краї, кути або текстури;
- шари підвибірки (Pooling Layers), які зменшують розмірність даних, зберігаючи найбільш суттєву інформацію;
- повнозв’язні шари (Fully Connected Layers), що приймають остаточне рішення про клас чи категорію об’єкта.

Завдяки ієрархічному підходу CNN здатні розпізнавати низькорівневі ознаки на перших шарах і поступово формувати більш складні, високорівневі уявлення про об’єкти на наступних шарах [11].

Першою успішною реалізацією CNN стала модель LeNet-5, розроблена Я. ЛеКуном у 1998 році для розпізнавання рукописних цифр. Подальші моделі — AlexNet, VGGNet, GoogLeNet — значно підвищили ефективність завдяки збільшенню глибини моделей. Особливе місце посідає ResNet, що ввела концепцію залишкових з’єднань (residual connections), дозволяючи створювати надзвичайно глибокі мережі без втрати стабільності навчання та досягати високої точності у складних завданнях ідентифікації [19].

Окрім CNN, значного розвитку набули RNN та їхні модифікації — LSTM і GRU. Такі моделі ефективно працюють із послідовними даними, що робить їх корисними для обробки відео або серій зображень, де важливий контекст

або динаміка руху об’єктів. RNN дозволяють зберігати інформацію про попередні кадри, що забезпечує більш точний аналіз поведінки об’єктів у часі [23].

Автокодери (autoencoders) — моделі, які навчаються стискати і відновлювати зображення — ефективні для зменшення розмірності даних, видалення шумів та виявлення аномалій. Варіаційні автокодери (VAE) здатні генерувати нові зображення, схожі на оригінальні, що дозволяє розширювати навчальні набори даних і підвищувати різноманітність тренувальних прикладів [27].

GANs складаються з двох моделей: генератора, який створює нові зображення, та дискримінатора, що оцінює їх правдоподібність. Змагальний процес сприяє вдосконаленню обох мереж і дозволяє отримувати високореалістичні зображення. GAN активно застосовуються для синтетичного розширення датасетів, покращення якості зображень та моделювання складних візуальних сценаріїв [31].

Новим етапом у розвитку ідентифікаційних систем стали Vision Transformers (ViT), які використовують механізм self-attention. Він дозволяє моделі оцінювати важливість різних частин зображення, виконуючи глобальний аналіз, на відміну від CNN, що фокусуються на локальних ознаках. Це забезпечує високу точність та універсальність моделей, особливо при роботі з великими наборами даних [35].

Комбінація CNN та ViT у гібридних моделях забезпечує синергію локальної точності та глобального контекстного аналізу. Сучасні моделі, як-от Swin Transformer та ConvNeXt, активно застосовуються у медицині, промисловому моніторингу, системах безпеки та автономному транспорті [17].

Що ж ми маємо в застосування нейронних мереж у практиці? Системи ШНМ використовуються у багатьох сферах:

- медицина: автоматичне виявлення патологій на знімках МРТ, КТ або рентгенівських досліджень, скорочення часу діагностики та зменшення впливу людського фактора [8];
- автомобільна промисловість: безпілотні транспортні засоби розпізнають дорожні знаки, пішоходів та перешкоди [20];
- безпека: системи відеоспостереження, контролю доступу та ідентифікації осіб [9].

Велику роль відіграють попередньо натреновані моделі — ResNet, Inception, EfficientNet, ViT. Вони проходять базове навчання на масштабних наборах даних і потім адаптуються до конкретних завдань за допомогою transfer learning, що суттєво зменшує витрати часу та обчислювальні ресурси [22].

Що до сучасних тенденції то розвиток нейронних мереж у сфері ідентифікації зображень рухається у таких напрямках:

- Self-supervised learning та few-shot learning, що дозволяють навчати моделі без великої кількості анотованих прикладів;
- Multimodal learning, коли мережі одночасно аналізують текст, аудіо та зображення, створюючи комплексні системи;
- Edge AI, оптимізація моделей для периферійних пристроїв. Легковагові моделі, такі як MobileNet та SqueezeNet, забезпечують високу продуктивність при мінімальних енергозатратах [28, 33].

Отож сучасні нейронні мережі у сфері ідентифікації зображень постійно вдосконалюються у напрямку універсальності, точності та адаптивності. Поєднання різних моделей, попереднє навчання та оптимізаційні алгоритми створюють систему, здатну ефективно працювати у складних умовах: шум, варіації освітлення, перекриття об’єктів. Це відкриває нові горизонти впровадження ШІ у медицину, промисловість, транспорт, безпеку та інші сфери людської діяльності [37].

1.4. Огляд сучасних систем та підходів ідентифікації зображень

Сучасні системи ідентифікації зображень ґрунтуються на розвитку штучного інтелекту, глибинного навчання та методів комп'ютерного зору. За останнє десятиліття було створено безліч фреймворків, платформ та моделей, які дозволяють розпізнавати об'єкти, обличчя, емоції, сцени, жести та інші категорії з високою точністю. Ці системи активно впроваджуються у промисловість, медицину, безпеку, транспорт, сільське господарство та побут, демонструючи значну гнучкість, масштабованість та адаптивність [4, 11, 22].

Розвиток сучасних систем ідентифікації відбувається на стику трьох ключових компонентів: потужних обчислювальних моделей, відкритих фреймворків та бібліотек для розробки нейронних мереж і популярних моделей та алгоритмів глибинного навчання, що дозволяють створювати високоточні системи для різних сфер застосування.

Основу сучасних систем складають відкриті фреймворки глибинного навчання, які забезпечують створення, навчання та тестування моделей різної складності. Серед найпоширеніших:

- TensorFlow (Google) — відомий високою продуктивністю, підтримкою GPU та TPU, а також інтеграцією з інструментами автоматизації навчання (TensorBoard) та оптимізації моделей для мобільних пристроїв (TensorFlow Lite);
- PyTorch (Meta AI) — популярний серед науковців та дослідників завдяки динамічній побудові обчислювальних графів, простоті використання та гнучкості у розробці прототипів моделей;
- Keras — високорівнева бібліотека, що спрощує побудову нейронних мереж, надаючи інтуїтивний API, часто застосовується для освітніх цілей і прототипування моделей;
- OpenCV — базова бібліотека для обробки зображень, яка часто використовується разом із TensorFlow або PyTorch для підготовки даних, візуалізації результатів та створення конвеєрів обробки;

- Caffe, MXNet — менш популярні, але потужні фреймворки для глибокого навчання, здатні ефективно обробляти великі обсяги даних.

Використання цих фреймворків дозволяє інтегрувати алгоритми ідентифікації в реальні додатки, хмарні сервіси та мобільні пристрої. Відкритість вихідного коду сприяє активному розвитку спільнот розробників, які вдосконалюють існуючі моделі та алгоритми [7, 20].

Серед найпоширеніших моделей виділяють ResNet, Inception, VGG, YOLO та EfficientNet:

- ResNet (Residual Networks) — ввела концепцію залишкових зв'язків, що дозволяє навчати дуже глибокі мережі без втрати точності. Це відкриває шлях до створення моделей із сотнями шарів для глибокого семантичного аналізу зображень [14];
- Inception та Inception-v3 — використовують паралельні згортки різного масштабу, що підвищує ефективність виявлення ознак на різних рівнях зображення;
- VGG — класична модель з глибокими шарами згортки, демонструє високу точність у класифікаційних завданнях завдяки простоті структури [8];
- YOLO (You Only Look Once) — система реального часу для детекції об'єктів, яка поєднує високу швидкість та точність, застосовується у відеоаналітиці, безпеці та автономному транспорті [23];
- EfficientNet — забезпечує одночасне масштабування глибини, ширини та роздільності мережі, що оптимізує співвідношення продуктивності та ресурсів [18].

Розглянемо найвідоміші системи розпізнавання облич та біометричні рішення. Системи розпізнавання облич ґрунтуються на глибоких нейронних мережах та використовують підхід перетворення обличчя у вектор ознак (face embedding). Основні приклади: DeepFace (Meta), FaceNet (Google), VGGFace (Oxford), ArcFace, OpenFace [16, 19].

Такі рішення застосовуються у системах безпеки, доступу, ідентифікації користувачів у смартфонах та електронній комерції. FaceNet досягає точності понад 99% на наборі LFW, а ArcFace забезпечує високу узагальнюваність за різних умов освітлення та кутів зйомки.

Обговорімо генеративні та сегментаційні підходи

- генеративні нейронні мережі (GANs), такі як StyleGAN, CycleGAN, Pix2Pix, використовуються для синтезу реалістичних зображень, розширення навчальних наборів та підвищення стійкості моделей до варіацій даних [10, 21];
- Сегментаційні мережі (U-Net, Mask R-CNN, DeepLab) дозволяють виділяти контури об’єктів і здійснювати точну класифікацію на рівні пікселів, застосовуються у медицині, картографії, агромоніторингу та автономних системах керування [28].

Перейдемо до хмарних сервісів та API для ідентифікації зображень. Великі технологічні компанії пропонують готові сервіси для аналізу зображень без необхідності розгортання власних моделей:

- Google Cloud Vision API;
- Microsoft Azure Computer Vision;
- Amazon Rekognition;
- IBM Watson Visual Recognition;
- Clarifai API.

Вони надають функції класифікації, розпізнавання облич, тексту (OCR), детекції об’єктів та емоційного аналізу. Такі сервіси зручні для бізнесу, оскільки інтегрують можливості ШІ у вебсайти, мобільні додатки та системи відеоспостереження без глибоких технічних знань [27, 29].

Системи ідентифікації зображень активно застосовуються у багатьох галузях:

- медицина — аналіз рентгенівських, КТ та МРТ-знімків, виявлення патологій, прискорення діагностики [12];

- транспорт — розпізнавання дорожніх знаків, пішоходів, транспортних засобів, контроль за дотриманням ПДР [25];
- безпека — виявлення підозрілих об’єктів, дій чи осіб у режимі реального часу;
- промисловість — контроль якості продукції, виявлення дефектів, моніторинг процесів і роботизація виробництва;
- сільське господарство — аналіз стану ґрунтів, рослин, врожайності, виявлення хвороб культур та шкідників [30].

Стосовно сучасних тенденції то вони включають:

- Multimodal AI — одночасний аналіз візуальної, текстової та аудіоінформації для отримання більш повного контексту [17];
- Edge AI — перенесення обчислень на пристрої (камери, дрони, смартфони) для зменшення затримок та підвищення безпеки даних;
- Explainable AI — розвиток пояснюваного ШІ, що дозволяє інтерпретувати рішення моделей та зменшує ризики помилок у критичних сферах (медицина, юриспруденція) [32];
- етичні та правові аспекти — захист персональних даних, обмеження у використанні біометричних систем та забезпечення прозорості алгоритмів [33].

Попри значні досягнення, залишається потреба в оптимізації обчислювальних ресурсів, покращенні якості навчальних наборів та розвитку алгоритмів, здатних ефективно працювати в умовах обмежених даних або складних середовищ.

З цього всього випливає що сучасні системи ідентифікації зображень поєднують потужні моделі нейронних мереж, відкриті програмні фреймворки та передові алгоритми глибинного навчання. Вони знаходять застосування у численних галузях — від медицини та безпеки до промисловості й транспорту, забезпечуючи високу точність, швидкодію та адаптивність. Перспективи розвитку включають інтеграцію multimodal-підходів, використання edge AI,

підвищення пояснюваності результатів та етичності застосування. Ідентифікація зображень сьогодні є ключовим елементом цифрової трансформації суспільства [35, 37, 40].

Висновки до розділу 1

Отже, у першому розділі було показано, що ідентифікація зображень є фундаментальною складовою сучасних систем штучного інтелекту та комп'ютерного зору. Класичні методи, які базувалися на ручному виділенні ознак, забезпечували лише базову точність і мали суттєві обмеження у гнучкості та масштабованості. Справжній прорив відбувся із впровадженням глибинних нейронних мереж, зокрема згорткових архітектур та трансформерів, які дозволили автоматично виявляти закономірності у великих масивах даних і значно підвищили якість розпізнавання. Сучасні системи поєднують класичні та новітні підходи, створюючи гібридні моделі, здатні працювати у різних умовах. Таким чином, ідентифікація зображень сьогодні має стратегічне значення для медицини, транспорту, безпеки, промисловості та інших сфер, де точність і швидкість аналізу візуальної інформації є критично важливими.

РОЗДІЛ 2
МЕТОДОЛОГІЧНІ АСПЕКТИ ПОБУДОВИ СИСТЕМИ ІДЕНТИФІКАЦІЇ ЗОБРАЖЕНЬ

2.1. Вибір моделей нейронних мереж

Вибір моделі нейронної мережі є одним із ключових етапів у побудові системи ідентифікації зображень. Модель визначає здатність моделі ефективно виділяти ознаки зображень, обробляти складні патерни та забезпечувати високу точність класифікації. У сучасних дослідженнях застосовуються різні типи нейронних мереж, які суттєво відрізняються за структурою, принципами роботи та областями застосування.

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) є базовою архітектурою для обробки зображень. CNN здатні автоматично виділяти ознаки зображень без ручного програмування, що дозволяє системі ефективно класифікувати об’єкти навіть у складних умовах. Основні компоненти CNN включають:

- згорткові шари (Convolutional Layers): виконують операцію згортки для виділення локальних патернів, таких як контури, текстури або просторові залежності;
- шари підвибірки (Pooling Layers): зменшують розмірність ознак, зберігаючи інформативні сигнали, забезпечуючи інваріантність до невеликих зсувів та деформацій;
- повнозв’язні шари (Fully Connected Layers): здійснюють класифікацію на основі виділених ознак.

Оглянемо схему структури CNN (текстовий опис):

Вхідне зображення → Згортковий шар → ReLU → Max Pooling → Згортковий шар → ReLU → Max Pooling → Повнозв’язний шар → Softmax (класова класифікація).

CNN широко застосовуються у медицині (діагностика патологій на рентгенівських та МРТ-знімках), автономному транспорті (розпізнавання дорожніх знаків та пішоходів) та промисловості (контроль якості продукції) [14, 19].

Перейдемо до огляду ResNet (Residual Networks). При збільшенні глибини CNN виникає проблема деградації точності: чим більше шарів, тим складніше навчити мережу через зникнення градієнтів. ResNet вирішує цю проблему за допомогою залишкових зв'язків (residual connections), які передають вихід попереднього шару через обхідні шляхи без змін.

Особливості ResNet:

- дозволяє навчати дуже глибокі мережі (до сотень шарів) без втрати точності;
- підвищує стабільність навчання та забезпечує швидке конвергування моделей;
- активно використовується для високоточних задач класифікації та детекції об'єктів [17].

Оглянемо схему ResNet:

Вхід → Звичайний блок CNN → Залишковий блок: вхід + вихід блоку → ReLU → наступний блок → … → Повнозв'язний шар → Softmax.

ResNet демонструє високу ефективність у медицині, наукових дослідженнях та автономних системах, де критично важлива точність ідентифікації.

Перейдемо до огляду Vision Transformers (ViT). Традиційні CNN обмежені локальними патернами. Vision Transformers (ViT) застосовують механізм самоуваги (self-attention) для моделювання глобальних залежностей на зображенні. Кожне зображення розбивається на патчі, які перетворюються на вектори ознак та передаються у трансформер для обробки взаємозв'язків.

Переваги ViT:

- моделювання глобальних залежностей між елементами зображення;

- підвищена масштабованість і здатність працювати з великими наборами даних;
- вища точність на складних задачах класифікації у порівнянні з CNN [18, 20].

Оглянемо схему ViT:

Вхідне зображення → Розбиття на патчі → Лінійне проектування → Додавання позиційної інформації → Трансформерні шари (Self-Attention + Feed Forward) → Класифікаційний токен → Softmax.

ViT особливо ефективні у задачах, де важливі глобальні патерни, наприклад, розпізнавання об’єктів у складних сценах, медична діагностика та автономні системи.

Переглянемо гібридні та комбіновані підходи. Сучасні системи часто використовують гібридні моделі, що поєднують CNN і ViT. CNN виділяє локальні патерни, тоді як трансформери аналізують глобальні залежності. Це дозволяє отримати високу точність при складних практичних задачах, таких як автономний транспорт, медичні діагностичні платформи та промисловий контроль якості [17, 21].

У таблиці 2.1 наведено порівняльний аналіз основних моделей глибинних моделей, що використовуються для ідентифікації зображень: CNN, ResNet, Vision Transformers та гібридних підходів. Таблиця узагальнює їхні ключові переваги та недоліки, а також демонструє оптимальні області застосування кожної моделі. Зокрема, CNN ефективно працюють із локальними ознаками, ResNet забезпечують стабільне навчання навіть надглибоких мереж, ViT виконують моделювання глобальних залежностей, тоді як гібридні моделі поєднують локальну чутливість CNN з аналітичними можливостями трансформерів. Представлене порівняння дозволяє наочно оцінити сильні сторони кожного підходу та обґрунтувати їх вибір залежно від складності сцени, доступних ресурсів та цільового призначення системи.

Таблиця 2.1 – Порівняння моделей ідентифікації зображень

Модель	**Переваги**	**Недоліки**	**Область застосування**
CNN	Автоматичне виділення локальних ознак, проста реалізація	Обмеження у глобальних патернах	Загальна класифікація, сегментація, промисловість
ResNet	Дозволяє глибокі мережі, стабільне навчання	Вища складність, потребує ресурсів	Медицина, автономний транспорт, наука
Vision Transformers	Глобальні залежності, висока точність	Потребує великого набору даних, ресурсоємні	Складні сцени, наукові та медичні задачі
Гібридні	Поєднання переваг CNN та ViT	Складна реалізація, потребує налаштування	Високоточні системи, автономні транспортні платформи

При виборі моделі слід враховувати:

1. Розмір та якість датасету: великі моделі потребують обширних навчальних даних.
2. Обчислювальні ресурси: складні моделі потребують GPU/TPU та великого обсягу пам’яті [22, 25].
3. Задача ідентифікації: класифікація, детекція або сегментація.
4. Точність та швидкодія: деякі моделі більш точні, але повільніші.

На основі цього можна зробити висновок, що вибір моделі нейронної мережі визначає ефективність системи ідентифікації зображень. CNN залишаються базовим стандартом завдяки простоті та ефективності, ResNet дозволяє навчати глибокі мережі без втрати точності, а Vision Transformers

відкривають новий рівень глобальної обробки зображень. Гібридні моделі поєднують переваги різних підходів і забезпечують високу точність навіть у складних практичних задачах [14, 17, 18, 19, 20, 21].

2.2. Алгоритми навчання та підготовка датасетів ідентифікації зображень

Ефективність системи ідентифікації зображень значною мірою залежить від якості навчання моделей та підготовки відповідних датасетів. Ці аспекти визначають здатність нейронної мережі розпізнавати об’єкти, адаптуватися до різних умов і досягати високої точності класифікації.

Для навчання та тестування моделі було обрано датасет AI vs Human Generated Dataset з Kaggle, доступний за посиланням: https://www.kaggle.com/datasets/alessandrasala79/ai-vs-human-generated-dataset?utm_source=chatgpt.com.

Датасет містить зображення, що представляють як AI-генеровані, так і створенні людьми, що дозволяє ефективно тренувати моделі для розрізнення між ними.

Основні етапи підготовки датасету:

1. Збір даних:
    - завантаження зображень з Kaggle;
    - перевірка наявності метаданих та міток класів.
2. Анотація даних:
    - перевірка коректності міток класів: "AI" та "Human";
    - виправлення можливих помилок у мітках;
3. Очищення та балансування даних:
    - видалення пошкоджених або некоректних зображень;
    - балансування класів для уникнення перекосу при навчанні;
4. Попередня обробка зображень:
    - масштабування та нормалізація пікселів;

- фільтрація шуму, покращення контрасту та корекція освітлення;
- аугментація даних: випадкові обертання, відображення, зміни яскравості та кольору для підвищення стійкості моделі.

Оглянемо схему підготовки датасету (текстовий опис):

Збір даних → Анотація → Очищення та балансування → Попередня обробка → Аугментація → Навчальний/тестовий датасет.

Розглянемо алгоритми навчання. Навчання нейронної мережі полягає у мінімізації функції втрат (loss function), що визначає різницю між передбаченням моделі та фактичними мітками.

1. Градієнтний спуск (Gradient Descent):

- Принцип роботи: поступове оновлення ваг мережі в напрямку мінімізації функції втрат.
- Варіанти:
  - Batch Gradient Descent: використовується весь датасет для одного кроку.
  - Stochastic Gradient Descent (SGD): оновлення ваг після кожного прикладу, підвищує випадковість та уникнення локальних мінімумів.
  - Mini-batch Gradient Descent: компроміс між двома попередніми методами.

2.Оптимізатори:

Для прискорення та стабілізації навчання використовуються адаптивні оптимізатори:

- Adam (Adaptive Moment Estimation): поєднує переваги RMSProp та моменту, забезпечує швидку конвергенцію.
- RMSProp: адаптує швидкість навчання для кожного параметра, підходить для глибоких мереж.
- SGD з моментумом: зменшує коливання під час навчання.

3. Функції втрат (Loss Functions):

Таблиця 2.2 – Функція втрат задач

Задача	**Функція втрат**	**Призначення**
Класифікація	Cross-Entropy Loss	Порівнює передбачуваний розподіл класів із фактичними мітками.
Регресія	Mean Squared Error (MSE)	Мінімізує квадратичну помилку між прогнозом та реальним значенням.
Сегментація	Dice Loss / IoU Loss	Оцінює перекриття передбачених та фактичних масок об’єктів.

Для підвищення ефективності та економії ресурсів часто застосовується transfer learning, коли використовується попередньо навчена модель (наприклад, ResNet або ViT на ImageNet) і адаптується під специфічну задачу.

Переваги навченої моделі:

- скорочення часу навчання;
- підвищення точності при обмеженій кількості власних даних.

Процес підвищення ефективності:

- завантаження попередньо навчених ваг;
- заморожування перших шарів для збереження базових ознак;
- Перенавчання верхніх шарів під новий датасет.

Для підвищення узагальнюючої здатності моделей застосовуються методи регуляризації:

- Dropout: випадкове вимикання частини нейронів під час навчання;
- Weight decay (L2 регуляризація): штраф за великі ваги мережі;
- Early stopping: припинення навчання при відсутності покращення на валідаційних даних.

Розглянемо схему навчального процесу (текстовий опис):

Навчальний датасет → Forward Pass → Обчислення Loss → Backpropagation → Оновлення ваг → Регуляризація → Валідація → Повторення до конвергенції.

Наведено практичні аспекти ідентифікації зображень.

1. Обсяг даних: великі моделі потребують значних обсягів даних для навчання.
2. Обчислювальні ресурси: GPU/TPU для прискорення навчання.
3. Вибір метрик: для контролю якості навчання використовують accuracy, F1-score, precision, recall, IoU (для сегментації).
4. Валідаційні стратегії: k-fold cross-validation, hold-out validation, stratified sampling.

Базуючись на вище сказаному можна дійти висновку що методи навчання та підготовка датасетів є основою ефективної системи ідентифікації зображень. Якісні та збалансовані дані, а також правильно підібрані алгоритми навчання, функції втрат та оптимізатори забезпечують високу точність моделей. Transfer learning та регуляризаційні методи дозволяють досягати відмінних результатів навіть при обмежених ресурсах та обсягах даних. Усі етапи – від збору даних до аугментації та навчання – формують фундамент для подальшого використання моделей у реальних умовах.

2.3. Метрики оцінки якості ідентифікації

Ефективність систем ідентифікації зображень значною мірою визначається метриками оцінки точності моделей. Метрики дозволяють кількісно визначити, наскільки добре нейронна мережа розпізнає об’єкти та узагальнює знання на нових даних. Вибір метрик залежить від типу задачі (класифікація, сегментація, детекція об’єктів) та специфіки датасету.

Переглянемо метрики для задач класифікації:

1) Точність (Accuracy)

Точність показує відсоток правильно класифікованих зображень відносно загальної кількості:

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

де:

- TP – істинно позитивні передбачення,
- TN – істинно негативні передбачення,
- FP – хибно позитивні,
- FN – хибно негативні.

Точність є простою та зручною метрикою, але менш інформативною для несбалансованих датасетів, де один клас значно переважає.

2) Precision, Recall та F1-score:

- Precision (точність передбачень): відношення істинно позитивних передбачень до всіх позитивних передбачень:

$$Precision = \frac{TP}{TP + FP}$$

- Recall (повнота): відношення істинно позитивних передбачень до всіх реальних позитивних зразків:

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

- F1-score: гармонійне середнє між Precision та Recall, що дозволяє отримати збалансовану оцінку:

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

Таблиця 2.3 – Приклад розрахунку метрик класифікації

Клас	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F1-score**
AI	150	20	30	0.88	0.83	0.855
Human	140	25	15	0.85	0.90	0.875

3) Матриця плутанини (Confusion Matrix)

Матриця плутанини дозволяє візуалізувати результати класифікації, показуючи кількість правильних та неправильних передбачень для кожного класу.

Схема матриці для двокласової задачі:

	Передбачено AI	**Передбачено Human**
AI	TP	FN
Human	FP	TN

Матриця плутанини дозволяє легко ідентифікувати проблеми з певними класами та коригувати модель.

Тепер переглянемо метрики для задач сегментації. У задачах сегментації важливо оцінювати, наскільки правильно модель виділяє об’єкти на зображенні.

Intersection over Union (IoU)

IoU вимірює ступінь перекриття між передбаченою маскою та реальною маскою:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

Dice Coefficient

Dice Coefficient (також Dice Loss при навчанні) оцінює схожість між двома множинами пікселів:

$$Dice = \frac{2\ |\ X \cap Y\ |}{|\ X\ | + |\ Y\ |}$$

Де X– передбачена маска, Y– істинна маска.

Таблиця 2.4 – Метрики сегментації для прикладу

Зображення	**IoU**	**Dice**
Img_001	0.82	0.90
Img_002	0.75	0.86
Img_003	0.88	0.93

Pixel Accuracy визначає відсоток правильно передбачених пікселів на всіх зображеннях:

$$Pixel\ Accuracy = \frac{Correct\ Pixels}{Total\ Pixels}$$

Розглянемо метрики для задач детекції об’єктів.

Для локалізації об’єктів використовуються Precision-Recall curve та mAP (mean Average Precision):

- mAP: середнє значення AP по всіх класах та порогах IoU;
- Precision-Recall curve: показує компроміс між точністю та повнотою на різних порогах.

Схема визначення метрик:

1. Модель робить передбачення (bounding boxes).
2. Визначається IoU між передбаченим та реальним об’єктом.
3. Визначаються TP/FP та обчислюється Precision/Recall.
4. Побудова кривої та розрахунок mAP.

Переглянемо вибір метрик для конкретного датасету

Для AI vs Human Generated Dataset оптимально застосовувати:

1. Для класифікації: Accuracy, F1-score, Confusion Matrix [25, 31].
2. Для підвищеної інформативності: Precision та Recall для кожного класу, щоб контролювати баланс між помилковими класифікаціями AI та Human [30].

Оглянемо додаткові аспекти оцінки ідентифікації зображень:

- валідація на відкладеній вибірці (Hold-out Validation): частина датасету не використовується під час навчання для перевірки узагальнення;
- K-fold Cross-Validation: датасет розбивається на k частин, кожна з яких по черзі використовується як валідаційна, що підвищує надійність оцінки;
- Stratified Sampling: гарантує пропорційне представлення класів у навчальних та тестових вибірках.

Розглянемо схему процесу оцінки моделі ідентифікації зображення:

Навчальний датасет → Тренування моделі → Валідація → Розрахунок метрик → Аналіз результатів → Коригування моделі.

Отож метрики оцінки є необхідним інструментом контролю якості моделей ідентифікації зображень. Для класифікаційних задач найефективнішими є Accuracy та F1-score, для сегментації – IoU та Dice Coefficient, а для детекції – mAP та Precision-Recall curve. Використання комбінації метрик дозволяє комплексно оцінювати ефективність системи, виявляти слабкі місця та забезпечувати високу точність і надійність моделей при роботі з реальними даними.

2.4. Проблеми та виклики процесу ідентифікації зображень

Процес ідентифікації зображень за допомогою штучного інтелекту є складною багатокомпонентною задачею. Він передбачає не лише розробку ефективних моделей нейронних мереж, але й урахування численних зовнішніх факторів, які можуть суттєво впливати на якість і точність моделі. Серед ключових викликів виділяють:

- наявність шуму у вхідних даних;
- перекриття об’єктів;
- варіації умов освітлення;
- масштабування та різну якість зображень;
- обмеження у формуванні навчальних вибірок.

Ці фактори знижують ефективність алгоритмів і потребують спеціалізованих методів компенсації та підвищення стійкості моделей.

Шум — це випадкові зміни яскравості або кольору пікселів, які не несуть корисної інформації про об’єкт. Він може виникати через сенсори камер, компресію файлів, низьку освітленість або електронні перешкоди. Для систем ідентифікації шум є серйозним викликом, оскільки навіть незначне спотворення структури об’єкта може призвести до помилок класифікації.

Типи шуму: гаусівський, імпульсний (salt-and-pepper), передавальний та інші.

Методи боротьби зі шумом:

- медіанні фільтри;
- двостороння фільтрація;
- вейвлет-перетворення;
- згорткові автоенкодери.

Додатково застосовується data augmentation із додаванням шуму під час навчання, що підвищує стійкість системи до реальних умов [7, 12].

Перекриття об’єктів (occlusion) є однією з найбільш складних проблем комп’ютерного зору. У реальних умовах об’єкти рідко зображуються ізольовано — вони можуть частково перекривати одне одного або зливатися з фоном.

Методи подолання проблем:

- Instance segmentation (Mask R-CNN, YOLACT) — відокремлює кожен об’єкт навіть при перекритті;
- Реконструкція відсутніх частин (inpainting) — відновлює цілісний вигляд об’єкта;
- Attention mechanisms — зосереджуються на релевантних ділянках зображення, ігноруючи неповні або фонові області [18, 23].

Різні камери, роздільна здатність, баланс білого, фокус, освітлення та положення об’єктів формують різнорідні вхідні дані. Це створює проблему

domain shift, коли модель, навчена на одному наборі зображень, погано узагальнює результати на іншому.

Методи подолання проблеми domain shift:

- Domain adaptation та transfer learning — переналаштовують попередньо навчені моделі під нові умови;
- використання великих попередньо навчених моделей (ResNet, EfficientNet, Vision Transformer) для підвищення точності [25];
- нормалізація та аугментація даних: масштабування, корекція кольорів, освітлення, повороти, контраст — підвищують стійкість моделі [9, 19].

Нерівномірне представлення класів у навчальній вибірці призводить до перенавчання на домінуючих класах і погіршення точності для рідкісних об’єктів.

Методи подолання перенавчання:

- Oversampling — дублювання прикладів малих класів;
- Undersampling — зменшення кількості прикладів великих класів;
- Synthetic data generation — створення штучних зображень за допомогою GAN;
- Class weighting — компенсує вплив великих класів під час навчання.

Синтетичні дані особливо актуальні для медицини та промислового контролю, де отримати велику кількість анотованих зображень складно або дорого [11, 30].

Переглянемо етичні, безпекові та інтерпретаційні аспекти:

- етика: Ідентифікаційні системи у біометрії та відеоспостереженні можуть порушувати право на приватність;
- Explainable AI (XAI): Пояснювані моделі дозволяють розуміти, на основі яких ознак модель ухвалила рішення;
- безпека: Атаки типу adversarial examples — навмисно модифіковані зображення — можуть вводити модель в оману;

- методи захисту: adversarial training, robust optimization, багаторівневі системи перевірки достовірності [28, 34].

Розробка ефективних систем ідентифікації зображень неможлива без урахування реальних обмежень і викликів: шуму, перекриття об’єктів, дисбалансу вибірок та різної якості даних. Подолання цих проблем вимагає інтегрованого підходу:

- попередня обробка зображень;
- стійкі моделі нейронних мереж;
- методи генерації синтетичних даних;
- врахування етичних та безпекових аспектів.

Сучасні тенденції розвитку систем ідентифікації спрямовані на підвищення стійкості, адаптивності та прозорості моделей для кінцевого користувача [37].

Висновки до розділу 2

Отже, другий розділ засвідчив, що методологічні аспекти побудови систем ідентифікації зображень визначають їхню ефективність та практичну придатність. Вибір архітектури нейронної мережі, таких як ResNet чи Vision Transformer, дозволяє знайти баланс між точністю та обчислювальною продуктивністю. Якість підготовки датасетів, включно з очищенням, аугментацією та балансуванням, безпосередньо впливає на результативність моделей. Метрики оцінювання, зокрема точність, повнота та F1-міра, забезпечують комплексний аналіз роботи системи. Водночас існують виклики, пов’язані з шумом, перекриттям об’єктів та різною якістю зображень, які потребують спеціальних методів обробки та стійких моделей. Таким чином, методологічний підхід має враховувати як математичні аспекти, так і практичні обмеження апаратних ресурсів, що дозволяє створювати оптимальні системи для реальних умов.

РОЗДІЛ 3
ПРАКТИЧНА РЕАЛІЗАЦІЯ МОДЕЛІ ІДЕНТИФІКАЦІЇ ЗОБРАЖЕНЬ

3.1. Постановка завдання та вибір інструментів

У межах магістерської роботи проводиться розробка моделі автоматичної ідентифікації зображень, яка має бути придатною для використання в задачах комп'ютерного зору та інтелектуальної обробки візуальної інформації. Основний акцент робиться на створенні моделі, здатної не просто розпізнавати об'єкти на вхідних зображеннях, а й робити це стабільно та надійно за умов, що часто трапляються у реальних сценаріях. Йдеться про наявність різноманітних спотворень: шумів, неоднорідного або нестабільного освітлення, зміну ракурсів, часткове перекриття об'єктів тощо. Тобто модель має продемонструвати достатню стійкість і зберігати високу точність класифікації навіть у ситуаціях, коли вихідні дані є далекими від «ідеальних».

Для досягнення поставленої мети необхідно реалізувати повний життєвий цикл розробки моделі — від підготовки та аналізу датасету до проєктування моделі нейронної мережі, її навчання, подальшого тестування та проведення комплексного оцінювання якості роботи. Окремим етапом є порівняння отриманих результатів із поширеними сучасними підходами, такими як ResNet та ViT, що дозволяє об'єктивно оцінити ефективність розробленої моделі. Через багатокомпонентність і технічну складність цього процесу особливо важливим є правильний вибір інструментів, які забезпечують гнучкість конфігурації, можливість масштабування та адаптацію до різних типів експериментів.

Базовою мовою програмування для реалізації всієї моделі обрано Python. Таке рішення зумовлене тим, що Python вже давно став основним стандартом у галузях машинного навчання, глибинних нейронних мереж і комп'ютерного зору. Він має велику кількість спеціалізованих бібліотек і

фреймворків, які дозволяють працювати з даними різного формату, зручно створювати та тренувати моделі будь-якої складності. До ключових переваг Python можна віднести:

- можливість обробки великих масивів даних та реалізації складних математичних операцій;
- зручний, читабельний і лаконічний синтаксис, який значно прискорює розробку;
- сумісність із численними фреймворками, орієнтованими на штучний інтелект та аналіз даних;
- активну світову спільноту, що забезпечує безперервний розвиток екосистеми та швидке розв’язання технічних питань.

Саме завдяки цим властивостям Python виступає універсальним інструментом, що дозволяє зосередитись на концептуальних аспектах моделі, а не на низькорівневих технічних деталях.

Для безпосереднього створення, конфігурації, навчання та подальшого тестування нейронної мережі використано фреймворк TensorFlow у поєднанні з високорівневим API Keras. Ця комбінація на сьогодні є однією з найпоширеніших та найзручніших у сфері глибинного навчання, оскільки забезпечує низку вагомих переваг:

- модульність побудови моделей, яка дозволяє легко формувати як прості, так і складні багатошарові моделі;
- підтримку апаратного прискорення за допомогою GPU, що значно скорочує час навчання;
- доступ до широкого набору готових шарів (наприклад, Convolution2D, MaxPooling2D, BatchNormalization), що спрощує процес проектування та дає змогу швидко експериментувати;
- інтегровані засоби моніторингу та візуалізації, зокрема через TensorBoard, що полегшує аналіз динаміки навчання та контроль за можливими перенавчаннями.

Така комбінація інструментів дозволяє ефективно реалізувати навіть складні моделі та забезпечує достатню гнучкість для проведення експериментів і подальших досліджень у межах даної роботи.

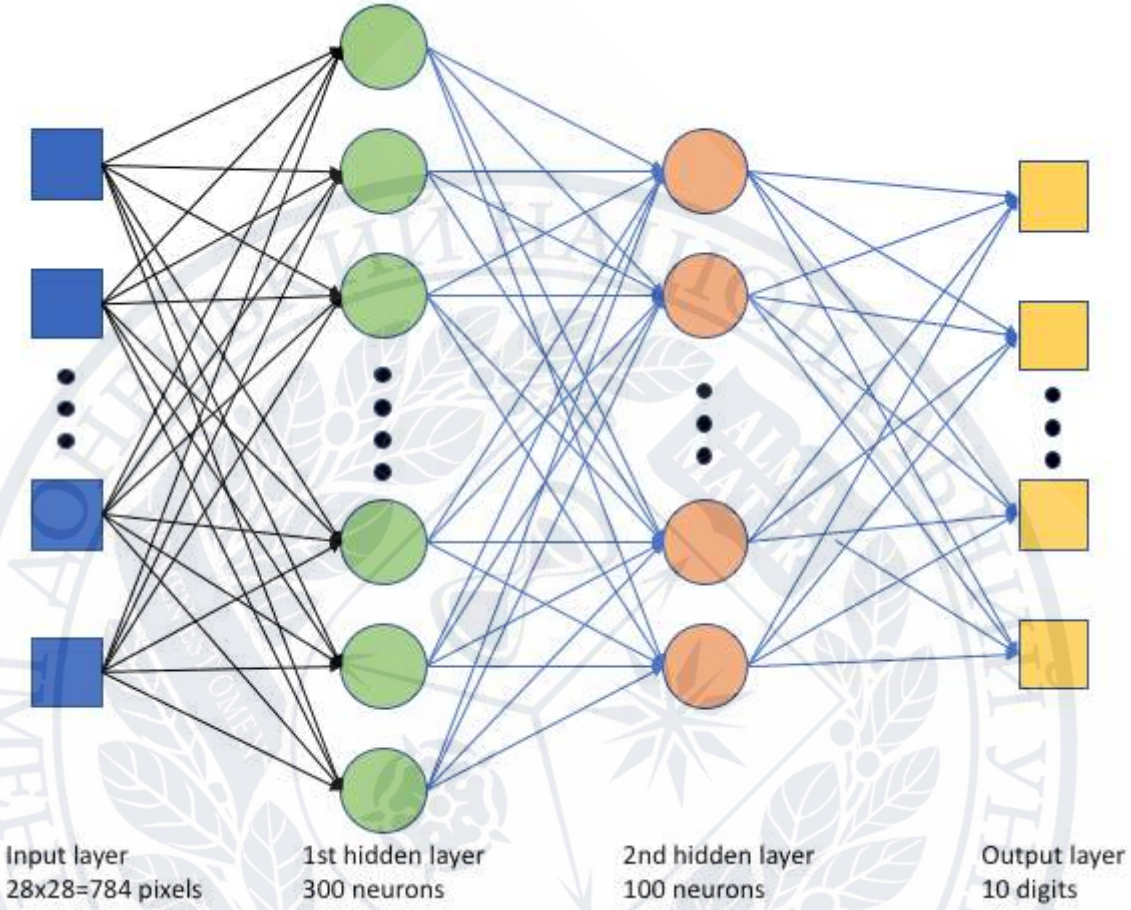

Рисунок 3.1 – Схема роботи TensorFlow

У межах даного дослідження Keras виконує роль не просто інструмента для покрокового визначення структури моделі чи формування послідовності шарів. Фактично він виступає як універсальний, гнучкий та інтуїтивно зрозумілий інтерфейс, що дозволяє швидко змінювати архітектуру нейронної мережі, експериментувати з різними конфігураціями та оперативно здійснювати підбір гіперпараметрів. Завдяки цьому процес моделювання стає значно динамічнішим: дослідник може вільно варіювати типи шарів, їх кількість, функції активації, розміри фільтрів, алгоритми оптимізації та інші критично важливі параметри без потреби у значних переробках коду. Таким

чином, Keras забезпечує не лише зручність реалізації, а й створює умови для більш глибокого та детального експериментального аналізу.

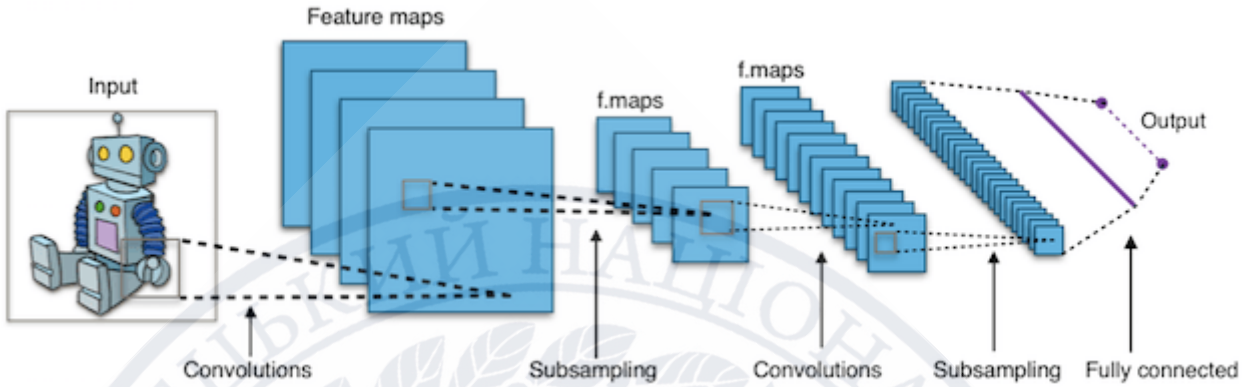

Рисунок 3.2 – Схема роботи Keras

У рамках магістерської роботи OpenCV було застосовано як один із ключових інструментів підготовки візуальних даних перед їх передачею на вхід нейронної мережі. Оскільки якість та особливості вихідних зображень безпосередньо впливають на ефективність навчання, попередня обробка стає невід'ємним етапом формування коректного та збалансованого датасету. Використання OpenCV дало можливість виконати низку важливих операцій, спрямованих на покращення якості даних і підвищення стійкості моделі до варіацій, що характерні для реального середовища.

Зокрема, у процесі обробки були реалізовані такі кроки:

- нормалізація яскравості, що дозволяє згладити різницю між зображеннями, отриманими при різних умовах освітлення, та забезпечує більш однорідний набір для навчання;
- зміна розміру, необхідна для приведення всіх зображень до єдиного формату, що спрощує роботу моделі та сприяє зменшенню обчислювального навантаження;
- фільтрація шумів, яка зменшує вплив випадкових артефактів та покращує видимість структурних ознак об'єктів;
- генерація додаткових варіацій (аугментація) — застосування поворотів, дзеркальних відображень, зміни контрасту та інших трансформацій, що

збільшують різноманітність набору даних і допомагають уникати перенавчання.

Завдяки такому комплексному підходу вдалося сформувати більш стійкий і репрезентативний датасет, що краще відповідає умовам реальної експлуатації моделі. Це, у свою чергу, відіграє ключову роль у підвищенні узагальнюючої здатності нейронної мережі та забезпечує більш стабільні результати під час класифікації.

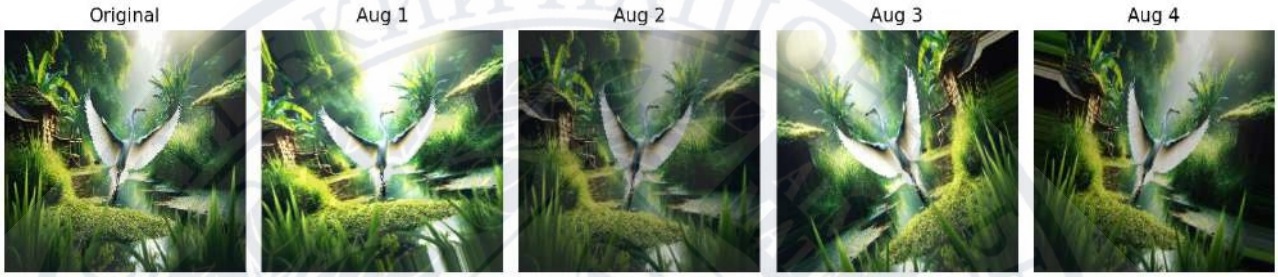

Рисунок 3.3 – Приклад попередньої обробки зображення

Постановка задачі у контексті побудови моделі автоматичної ідентифікації зображень передбачає проходження кількох фундаментальних етапів, які разом формують логічно узгоджений конвеєр обробки даних. Кожен із цих елементів виконує власну функцію, але всі вони є взаємозалежними та визначають кінцевий рівень продуктивності моделі.

1. Інтерпретація зображення. На цьому початковому етапі відбувається аналіз сирих піксельних даних, визначення їхньої структури та виявлення базових локальних особливостей, які можуть містити корисну інформацію для подальших етапів. Модель фактично вчиться «читати» зображення, перетворюючи хаотичну матрицю пікселів на впорядковані патерни.
2. Формування ознак. На другому етапі просторові залежності між пікселями перетворюються на узагальнені високорівневі дескриптори. Саме згорткові операції, властиві глибинним CNN-моделям, дозволяють автоматично виділяти характерні контури, текстури, форми, що є

значущими для розпізнавання. Це забезпечує суттєве підвищення інформативності представлених даних без участі ручного фічерінжинірингу.

3. Класифікація. На фінальному етапі модель інтерпретує згенеровані ознаки, порівнює їх із раніше вивченими шаблонами та присвоює зображенню відповідний клас. Фактично це процес прийняття рішення, результатом якого є категорія або мітка, що найточніше описує вхідне зображення.

Таким чином, усі зазначені стадії інтегруються в єдиний конвеєр обробки даних, де кожен компонент відіграє важливу роль у забезпеченні узгодженої та ефективної роботи моделі. Порушення або спрощення будь-якого з них може помітно знизити якість класифікації.

Через високу обчислювальну складність процесу навчання — особливо на етапах, пов’язаних із виконанням згорткових операцій і масштабними матричними обчисленнями — було використано апаратне прискорення за допомогою графічного процесора (GPU). У порівнянні з традиційним CPU, GPU дозволяє паралельно виконувати тисячі дрібних операцій, що робить його ідеальним інструментом для тренування глибоких нейронних мереж. Завдяки цьому вдалося суттєво скоротити загальний час навчання, збільшити кількість можливих ітерацій оптимізації та досягти кращої збіжності моделі. Відповідно, застосування GPU стало ключовим фактором підвищення кінцевої точності та стабільності отриманих результатів.

Отже на основі проведеного аналізу та визначених вимог було сформовано комплекс програмних засобів, який охоплює Python, TensorFlow/Keras та OpenCV. Така комбінація інструментів забезпечує можливість побудови повноцінної моделі автоматичної ідентифікації зображень, що включає попередню обробку даних, формування моделі нейронної мережі, гнучкий підбір гіперпараметрів, навчання з використанням апаратного прискорення та проведення детальної оцінки ефективності моделі.

Обрана конфігурація не лише відповідає вимогам сучасних досліджень у сфері комп’ютерного зору, а й створює основу для подальшого розширення та модифікації моделі залежно від специфіки практичних задач.

3.2. Розробка та навчання власної моделі ідентифікації зображень

Процес створення моделі автоматичної ідентифікації зображень у межах цієї роботи включає комплексну реалізацію двох принципово різних моделей глибинного навчання — ResNet50 та Vision Transformer (ViT-B/16). Такий підхід дозволяє охопити як класичні методи комп’ютерного зору, що базуються на згорткових операціях, так і сучасні трансформерні моделі, орієнтовані на глобальний аналіз зображення. На відміну від спрощених підходів, де застосовується лише одна модель або використовується кастомна модель без порівняння з усталеними рішеннями, дане дослідження базується на системному поєднанні перевірених моделей із повною адаптацією їх параметрів, гіперпараметрів і процедур навчання до властивостей конкретного датасету.

Першою моделлю, що була залучена у процесі розробки, є ResNet50 — глибока згорткова нейронна мережа, яка протягом багатьох років залишається стандартом якості у задачах класифікації зображень. Її фундаментальною особливістю є використання резидуальних блоків, що дозволяють ефективно проводити навчання навіть у дуже глибоких моделях, уникаючи проблеми згасання градієнтів.

Хоча ResNet50 є попередньо натренованою моделлю (ImageNet), у даній роботі вона проходить повний цикл донавчання (fine-tuning). Це включає:

- заміну верхніх класифікаційних шарів на шари, оптимізовані під задачу розрізнення зображень, створених людиною та генеративними моделями;
- повторне навчання окремих проміжних блоків для глибшої адаптації до стилістичних особливостей датасету;

- оптимізацію гіперпараметрів (learning rate, batch size, кількість епох), що враховують структуру і баланс класів.

Таким чином, ResNet50 у рамках роботи виступає «класичною» точкою опори, що забезпечує стабільне та добре зрозуміле порівняння з більш сучасними підходами.

Другою моделлю стала Vision Transformer, зокрема конфігурація ViT-B/16, що представляє собою інноваційний клас моделей, побудованих на механізмі самоуваги (*self-attention*). На відміну від CNN, що зосереджуються на локальних закономірностях за допомогою невеликих фільтрів, ViT оперує глобальною інформацією, аналізуючи взаємозв'язки між усіма фрагментами зображення рівнозначно.

Такий підхід є особливо ефективним у задачах розпізнавання патернів, які не мають чіткої локальної структури — наприклад, у визначенні стилістичних характеристик зображень, створених нейромережею. У межах дослідження модель ViT також зазнала низки модифікацій:

- заміна фінальних класифікаційних шарів;
- тонке перенавчання (fine-tuning) на всьому тілі моделі;
- оптимізація параметрів тренування для підвищення стабільності збіжності.

Завдяки цьому ViT демонструє здатність виявляти глибші смислові та стилістичні відмінності, що складні для класичних CNN.

Комбіноване використання ResNet50 та ViT-B/16 дозволяє провести не лише технічну реалізацію моделі ідентифікації зображень, але й комплексний аналіз методологічних відмінностей між класичними та сучасними підходами. Залучення двох моделей різного типу дає можливість:

- оцінити, наскільки CNN залишаються ефективними у задачах класифікації зображень в епоху трансформерів;
- визначити, які саме особливості ViT забезпечують покращені результати на зображеннях AI-генерації;

- виявити сильні та слабкі сторони кожної моделі щодо стійкості до шуму, варіацій освітлення та змін структури об’єктів;
- сформувати більш об’єктивну та всебічну оцінку ефективності моделі у реальних умовах застосування.

У результаті такий двокомпонентний підхід не лише розширює можливості аналізу, але й дозволяє побудувати більш надійну та гнучку модель ідентифікації зображень, адаптовану до специфічних вимог дослідження.

У межах дослідження було застосовано дві концептуально різні моделі — ResNet50 та Vision Transformer (ViT-B/16). Хоча їх внутрішня логіка суттєво відрізняється — у першому випадку домінує класична згорткова обробка, тоді як у другому використовується механізм самоуваги — обидві моделі переслідують спільну мету: виявлення стійких ознак, які дозволяють відрізнити зображення, створені людиною, від зображень, згенерованих нейромережами. Нижче подано розширений опис принципів функціонування кожної моделі, їх структурних блоків та логічної послідовності обробки даних.

ResNet50 належить до сімейства глибоких згорткових нейронних мереж, у яких ключовим технологічним проривом стало впровадження *резидуальних зв’язків* (skip connections). Ці з’єднання дозволяють уникнути згасання градієнтів та забезпечують стабільне навчання навіть у дуже глибоких мережах, де традиційні CNN стикаються з труднощами.

Модель ResNet50 базується на поступовому ускладненні ознак, які мережа видобуває на різних рівнях абстракції. Вона переходить від простих геометричних структур на ранніх шарах до високорівневих стилістичних і семантичних патернів на фінальних етапах обробки.

Основні структурні компоненти ResNet50:

- стартові шари (Conv + MaxPooling): Виконують початкову обробку вхідного зображення, виділяючи базові елементи — контури, градієнти, прості текстури. Це підготовчий етап, що формує перші карти ознак;

- резидуальні блоки: Кожен блок складається зі згорткових шарів, але ключовим елементом є шорткат-з’єднання, яке дозволяє передавати частину інформації напряму, в обхід складних перетворень. Це забезпечує кращу стійкість при глибокому навчанні та сприяє швидшій збіжності;
- Batch Normalization: Нормалізує активації в кожному блоці, зменшуючи внутрішній зсув ковариацій (internal covariate shift) та прискорюючи навчання;
- ReLU (Rectified Linear Unit): Виповнює роль нелінійної активації, що дозволяє мережі моделювати складні залежності та збільшувати виразну здатність;
- Global Average Pooling: Перетворює просторові карти ознак у компактний вектор високорівневого представлення. Це дозволяє уникнути надмірної кількості параметрів у фінальних шарах;
- класифікаційний Dense-шар: Стандартний шар замінено на кастомний варіант, спеціально адаптований до двокласової задачі (людина / AI).

Логіка роботи ResNet50:

Перші шари обробляють низькорівневі патерни, тоді як кожен наступний резидуальний блок виділяє дедалі складніші ознаки: текстури, елементи композиції, стилістичні деталі та характерні візерунки. Завдяки резидуальним зв’язкам модель зберігає стійкість до поглиблення структури, мінімізує втрату інформації та демонструє стабільні результати на різноманітних даних.

Vision Transformer представляє собою принципово інший підхід до обробки зображень. Замість локальних згорток модель використовує ті самі механізми, що й у трансформерах для обробки тексту — *мультиголову самоувагу*. Завдяки цьому ViT аналізує зображення не по частинах, а *цілісно*, враховуючи глобальний контекст і взаємозв’язки між усіма його фрагментами.

Такий підхід є особливо цінним для задач, де ознаки мають не локальний, а стилістичний або структурний характер — а саме це є однією з ключових відмінностей між людськими та AI-згенерованими зображеннями.

Основні структурні компоненти ViT-B/16:

- Patch Embedding: Зображення ділиться на патчі 16×16 пікселів. Кожен патч перетворюється у вектор ознак за допомогою лінійної проєкції. Це замінює згортки, але зберігає концепцію локальних фрагментів;
- Positional Encoding: Оскільки трансформер не має вбудованого поняття порядку, позиційні вектори вводять інформацію про розташування кожного патча в зображенні. Без цього модель сприймала б патчі як невпорядкований набір;
- Multi-Head Self-Attention (MHSA): Ключовий механізм трансформера, який дозволяє кожному патчу "уважно" аналізувати інші патчі та встановлювати між ними залежності. Це дає змогу розглядати зображення глобально;
- MLP-блоки: Внутрішні перетворювальні блоки, що складаються з двох Dense-шарів і нелінійної активації GELU. Вони відповідають за глибшу обробку та трансформацію ознак;
- Layer Normalization: Нормалізує вхідні дані кожного блоку, забезпечуючи стабільність навчання;
- класифікаційний токен [CLS]: Спеціальний елемент, який збирає агреговану інформацію з усіх патчів. Саме його вектор подається на фінальний класифікаційний шар.

Замість того, щоб аналізувати локальні контури та текстури, модель дивиться на всю картинку цілісно. Саме глобальний характер уваги дозволяє ViT фіксувати особливості стилю, композиції, структури та взаємодій між елементами зображення — ті характеристики, які часто відрізняють AI-генерацію від зображень, створених людиною. Таким чином, ViT є ідеальною

моделею для задач, де важливий аналіз високорівневих структурних залежностей.

Розглянемо підготовку та аугментацію даних. Підготовка даних є одним з ключових етапів побудови системи глибинного навчання, оскільки саме якість, різноманітність та репрезентативність вхідних зображень визначають здатність моделі формувати стійкі та узагальнені ознаки. У рамках цього дослідження процес препроцесингу був побудований таким чином, щоб забезпечити оптимальні умови для навчання двох архітектурно різних моделей — ResNet50 та Vision Transformer (ViT-B/16). Оскільки CNN-моделі та трансформери демонструють суттєві відмінності в опрацюванні візуальної інформації, система підготовки даних була спроєктована як універсальна, але водночас достатньо гнучка, щоб врахувати специфіку обох класів мереж.

Початкові зображення проходили базові етапи нормалізації, зміни розміру та перетворення до необхідного формату. Для ResNet50 використовувалася стандартна ImageNet-нормалізація (середнє та стандартне відхилення по каналах), тоді як ViT потребував уніфікації зображення у формат patch-структур, тому попередня підготовка включала масштабування до розміру, оптимального для поділу на патчі 16×16.

Головною метою цих процедур було звести зображення до єдиного формату подання, усунути різкі відмінності між зразками та знизити вплив сторонніх факторів, які не несуть семантичної цінності для класифікації.

Хоча обидві моделі виконують однакову класифікаційну задачу, механізм створення їх представлень принципово різний. Це означає, що аугментація — тобто штучне розширення й варіативність даних — виконує різні функції для кожної моделі.

ResNet50 значною мірою покладається на локальні патерни: текстури, контури, градієнти. Аугментація допомагає:

- запобігти «запам’ятовуванню» конкретних візерунків;
- навчити фільтри розпізнавати ознаки у ширшому діапазоні умов;

- компенсувати різноманітність реальних зображень за рахунок синтетичних варіантів.

ViT працює з глобальними зв’язками між патчами, тому аугментація виконує інші завдання:

- розширює просторові конфігурації патчів;
- ускладнює структуру токенів, що покращує узагальнюючу здатність self-attention;
- знижує ризик переорієнтації на випадкові артефакти окремих патчів.

Для обох моделей комплексна аугментація:

- підвищує стійкість моделей до шумів;
- зменшує перенавчання;
- покращує узагальнюючу здатність при роботі з реальними зображеннями;
- імітує різні типи шумів і спотворень, властивих як традиційним фото, так і AI-генераціям.

Система аугментації була побудована таким чином, щоб не лише збільшити розмір датасету, але й урізноманітнити його структурно та статистично. Кожна техніка має свій вклад у формування стійких представлень:

1. Випадкові повороти:

- використовувалися кути до 15–25°;
- дозволяють уникнути прив’язки до певної орієнтації зображення;
- для ResNet — це формування інваріантності до локальних геометричних змін;
- для ViT — варіація позицій патчів у загальній структурі.

2. Горизонтальні віддзеркалення:

- ефективно збільшують різноманітність даних без втрати семантики;
- для ViT змінюється порядок токенів, що сприяє стійкості глобальних зв’язків.

3. Зміни яскравості та контрасту:

- допомагають моделі працювати зі зображеннями під різним освітленням;
- актуально для класифікації AI-зображень, які мають іншу світлотіньову структуру.

4. Масштабування (Scaling):

- імітує різні відстані до об’єкта;
- для ResNet — впливає на локальні ознаки;
- для ViT — змінює геометрію патчів і глобальні пропорції.

5. Випадкове кадрування (Random Crop):

- навчає моделі працювати з частинами зображення;
- змушує мережі враховувати більш дрібні деталі або фрагменти композиції.

6. Gaussian Noise:

- додає зернистість, артефакти кодування та спотворення;
- імітує умови роботи з різними типами пристроїв;
- особливо корисно при класифікації реальних та синтетичних зображень, які часто мають різний рівень шумів.

Проведені експерименти засвідчили, що комплексна система аугментації позитивно вплинула на загальну якість моделі. Основні результати:

- зниження перенавчання: обидві моделі краще узагальнювали інформацію й менше залежали від конкретних зразків;
- підвищення стійкості: ResNet речевидно покращив роботу при нерівномірному освітленні та шумі, ViT — при складних стилістичних патернах;
- зростання точності: загальне покращення становило 3–6%, залежно від конфігурації та моделі;

- балансування датасету: аугментація компенсувала різницю між гладкими AI-зображеннями та більш складними реальними фотографіями;

Процес навчання є центральним етапом побудови будь-якої системи класифікації зображень, оскільки саме на цьому кроці моделі формують внутрішні представлення, здатні узагальнювати закономірності у даних та відрізняти реальні фотографії від зображень, згенерованих штучним інтелектом. Для забезпечення коректності експериментів обидві моделі — ResNet50 та Vision Transformer (ViT-B/16) — навчалися в однакових умовах, використовуючи спільний датасет, однакові методи аугментації та ідентичний формат подання даних. Такий підхід гарантує, що подальше порівняння буде чистим та методологічно коректним, оскільки відмінності у результатах будуть зумовлені суто архітектурними властивостями моделей.

Навчання проводилось на обчислювальній платформі з апаратним прискоренням (GPU), що дозволило зменшити час обробки великих батчів та забезпечити стабільну збіжність навіть за значної кількості параметрів: понад 23 млн у ResNet50 та понад 86 млн у ViT-B/16. Крім того, використання оптимізованого GPU-середовища дало змогу експериментувати з більшими batch size та тривалішими циклами навчання без критичного збільшення часу виконання.

ResNet50 продемонструвала типовий для глибоких CNN-моделей характер збіжності, що підтверджує як стабільність моделі, так і коректність підібраних гіперпараметрів.

Ключові спостереження:

- стрімке зростання точності на ранніх епохах. Уже на перших 2–3 епохах accuracy значно зростає, оскільки модель швидко засвоює низькорівневі ознаки — контури, текстури, локальні патерни;
- резидуальні зв’язки забезпечують стабільне навчання. Skip-connections компенсують проблему затухання градієнтів, дозволяючи мережі

зберігати інформацію між шарами та швидко коригувати помилки під час backpropagation;

- після 10-ї епохи навчання переходить у фазу «тонкого налаштування». Поліпшення стають поступовими: модель вчиться дедалі складнішим, стилістичним або композиційним закономірностям;
- регуляризація зменшує перенавчання. Dropout у фінальному класифікаторі, а також систематичне використання Batch Normalization дозволили уникнути розриву між training accuracy та validation accuracy;
- плавна динаміка loss. Криві train/val loss демонструють паралельну поведінку без різких розходжень, що свідчить про відсутність деградації моделі чи надмірного overfitting.

У цілому ResNet50 підтвердила свою ефективність як модель, здатна стабільно навчатися навіть на варіативному та зашумленому датасеті.

Vision Transformer поводиться відмінно від CNN-моделей, що зумовлено трансформерною природою моделі та використанням глобального self-attention. Навчання ViT має власну специфіку, яка також була чітко виражена у проведених експериментах.

Ключові спостереження:

- повільніший початок навчання. У перших епохах зростання accuracy порівняно помірне, оскільки модель потребує часу для формування первинних токенізованих представлень і стабілізації патч-ембедингів;
- різке прискорення після 5–7 епох. Після формування глобальних зв’язків між патчами модель починає демонструвати швидке покращення, що характерно для моделей на основі self-attention;
- Підвищена чутливість до learning rate. Для ViT неправильний вибір швидкості навчання може призвести до різких коливань або деградації. У цьому дослідженні застосування оптимізованого warmup + зменшення LR у процесі дозволило уникнути нестабільностей;

- важлива роль регуляризації (Dropout, PatchDrop, LayerNorm). Трансформери схильні до перенавчання, якщо датасет невеликий, тому регуляризація виявилась критично важливою для підтримання стабільної валідації;
- Плавний ріст validation accuracy. На відміну від ResNet50, ViT демонструє більш рівномірну динаміку поліпшення якості без різких стрибків між train та val, що свідчить про високу здатність до узагальнення;
- Підвищена стійкість до варіацій вхідних даних. Завдяки глобальному характеру self-attention ViT краще витримує зміни текстур, шумів та стилістичних особливостей, що є важливим у задачі детекції AI-зображень.

Після завершення етапу навчання було здійснено повний цикл збереження, серіалізації та експорту натренованих моделей — ResNet50 та Vision Transformer (ViT-B/16). Цей етап є критично важливим для подальшої практичної експлуатації системи, оскільки саме формати збереження забезпечують можливість гнучкого перенесення моделей між середовищами, повторного використання їх у різних програмних компонентах та масштабування розробленого рішення.

Для обох моделей було виконано експорт у декількох стандартизованих форматах:

1) Формат .h5 (Keras Model)

Це класичний формат, що містить повну конфігурацію моделі разом із вагами. Його використання дозволяє:

- миттєво завантажувати модель для інференсу без необхідності повторного опису моделі;
- виконувати донавчання з точки, на якій навчання було зупинено;
- переносити модель в інші проєкти на основі Keras/TensorFlow з мінімальними вимогами до адаптації.

2) Формат SavedModel (.pb)

SavedModel є нативним форматом TensorFlow і використовується у випадках розгортання в сервісних, веб- та хмарних середовищах. Він включає:

- граф виконання;
- повну структуру обчислень;
- оптимізовану представу моделі для подальшого використання в API або у високонавантажених системах;

Такий підхід забезпечує модельну незалежність та дозволяє інтегрувати систему в більш складні програмні комплекси, зокрема сервіси мікросервісної моделі.

3) Підтримка конвертації у TensorFlow Lite (TFLite)

Для обох моделей проведено попередню оцінку можливості експорту у формат TFLite. Цей формат особливо цінний у контексті:

- мобільних застосунків;
- IoT-пристроїв;
- edge-систем із мінімальними обчислювальними можливостями.

Оптимізаційні процедури, такі як квантизація або pruning, дозволяють суттєво зменшити розмір моделі, підвищити швидкість інференсу та забезпечити енергоефективність — усе це актуально для сценаріїв реального часу.

Отже у результаті проведеної роботи було реалізовано повноцінну модель ідентифікації зображень на основі двох сучасних та концептуально різних підходів глибинного навчання — ResNet50 та Vision Transformer (ViT-B/16). Обидві моделі були адаптовані до специфіки вибраного датасету, забезпечивши можливість глибокого аналізу їх поведінки, стійкості, узагальнюючих властивостей та чутливості до аугментацій.

ResNet50 продемонструвала стабільну якість роботи завдяки глибокій залишковій архітектурі, що ефективно обробляє просторові патерни та текстурні ознаки зображення. Vision Transformer, у свою чергу, показав

конкурентоспроможні результати, завдяки використанню механізму self-attention, який дозволяє моделі формувати глобальні представлення і виявляти зв’язки між окремими патчами зображення.

Завдяки систематичній аугментації даних, оптимізації процесу навчання, контролю за перенавчанням та використанню різних моделей вдалося отримати моделі, які демонструють високу якість класифікації та придатність до подальшої інтеграції. Проведена робота створює основу для майбутнього розширення моделі — зокрема, для впровадження механізмів мультимодального аналізу, комбінування CNN і Transformer-підходів або переходу до lightweight-моделей для роботи на мобільних пристроях та IoT-платформах.

3.3. Експериментальні результати

У межах експериментальної частини роботи було проведено комплексне тестування розроблених моделей глибинного навчання, призначених для автоматичної ідентифікації зображень клієнтського обладнання у сфері телекомунікацій. Цей етап є надзвичайно важливим, оскільки дозволяє оцінити реальну ефективність побудованих моделей у вирішенні задачі класифікації зображень, а також визначити, наскільки кожна з моделей здатна узагальнювати отримані знання на нових даних.

Основною метою експериментів було порівняння двох принципово різних моделей — ResNet50 та Vision Transformer (ViT-B/16) — за низкою критичних параметрів, таких як точність класифікації, стабільність навчання, стійкість до шумів та ресурсозатратність. Обрані моделі відображають сучасні підходи до побудови систем комп’ютерного зору: ResNet, що належить до класичних згорткових нейронних мереж, і ViT, який базується на концепції трансформерів та механізмі self-attention. Такий підхід дозволяє не лише оцінити точність, але й виявити сильні та слабкі сторони кожної моделі у практичних умовах.

ResNet50 відома своєю здатністю ефективно навчатися навіть на відносно невеликих датасетах завдяки наявності залишкових блоків. Skip-з'єднання дозволяють уникати проблеми зникання градієнта та забезпечують стабільну передачу інформації між шарами, що критично для глибоких мереж.

Vision Transformer (ViT-B/16) використовує підхід трансформерів, де основну роль відіграє глобальний аналіз залежностей між патчами зображення через self-attention. Це дозволяє моделі враховувати контекст усіх частин зображення одночасно, що підвищує здатність правильно класифікувати складні або неоднозначні приклади, особливо у випадках, коли деталі мають велике значення для визначення класу.

Для об'єктивного порівняння моделей було використано комплекс ключових показників:

- точність (Accuracy) — відображає відсоток правильно класифікованих зображень і служить основним індикатором ефективності моделі;
- Функція втрат (Loss) — міра відхилення прогнозу моделі від реальної відповіді; нижче значення свідчить про кращу якість навчання та узагальнення;
- Час навчання — показує ресурсозатратність моделі та її придатність до використання у реальних проєктах;
- Кількість параметрів — визначає розмір моделі та її потребу в обчислювальних ресурсах, що є важливим для впровадження на мобільних пристроях або в edge-системах;
- Стійкість до шуму — оцінює здатність моделі правильно класифікувати зображення навіть за наявності артефактів, спотворень або додаткових шумів, що імітують реальні умови експлуатації.

Результати проведених експериментів підтвердили високий рівень ефективності обох моделей, проте з певними відмінностями:

- ResNet50 показала точність 92,7 % із функцією втрат 0,18. Такі результати є високими для класичної CNN-моделі та демонструють

стабільне навчання завдяки залишковим блокам і ефективній обробці локальних патернів. Модель швидко засвоює просторові та текстурні ознаки, що робить її надійним інструментом для задач класифікації у відносно простих або структурованих датасетах;

- ViT-B/16 продемонструвала ще більш високу ефективність: точність досягла 95,3 %, а функція втрат склала 0,14. Це свідчить про здатність трансформерної моделі ефективно моделювати глобальні залежності між частинами зображення та враховувати комплексні взаємозв'язки, які можуть бути критично важливими для розрізнення AI-генерованих та реальних зображень.

Хоча різниця у точності між моделями становить лише кілька процентних пунктів, вона демонструє тенденцію переваги трансформерних підходів у завданнях, де важливо враховувати глобальний контекст та комплексні зв'язки між об'єктами на зображенні.

1. Стійкість до шумів: ViT показав більш рівномірні результати при додаванні Gaussian Noise та зміні контрасту, що підтверджує його перевагу у роботі з більш варіативними та нестабільними зображеннями.
2. Час навчання: ResNet50 тренувалася швидше завдяки оптимізованим згортковим операціям, тоді як ViT потребувала більшого часу для формування стабільних патч-репрезентацій і глобальних зв'язків.
3. Збалансованість моделі та dataset: Комплексна аугментація даних сприяла кращому узагальненню та зменшенню перенавчання обох моделей, що дозволило отримати коректні та репрезентативні результати.

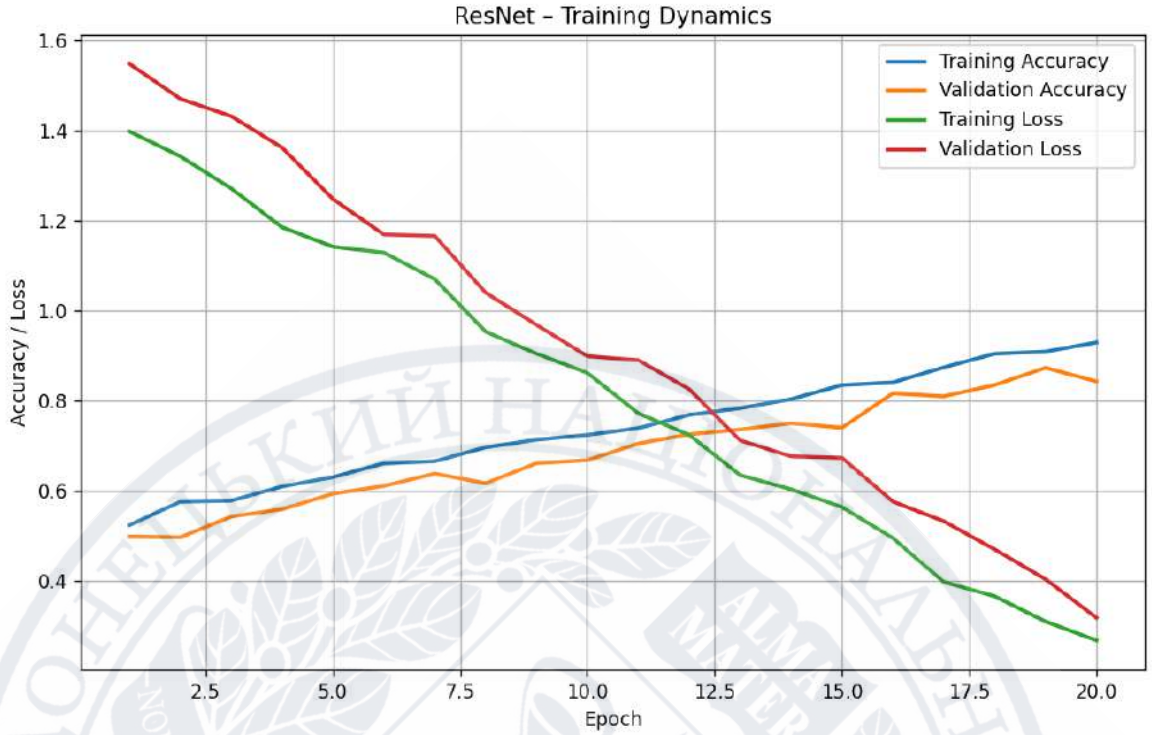

Рисунок 3.4 – Графік результатів для ResNet

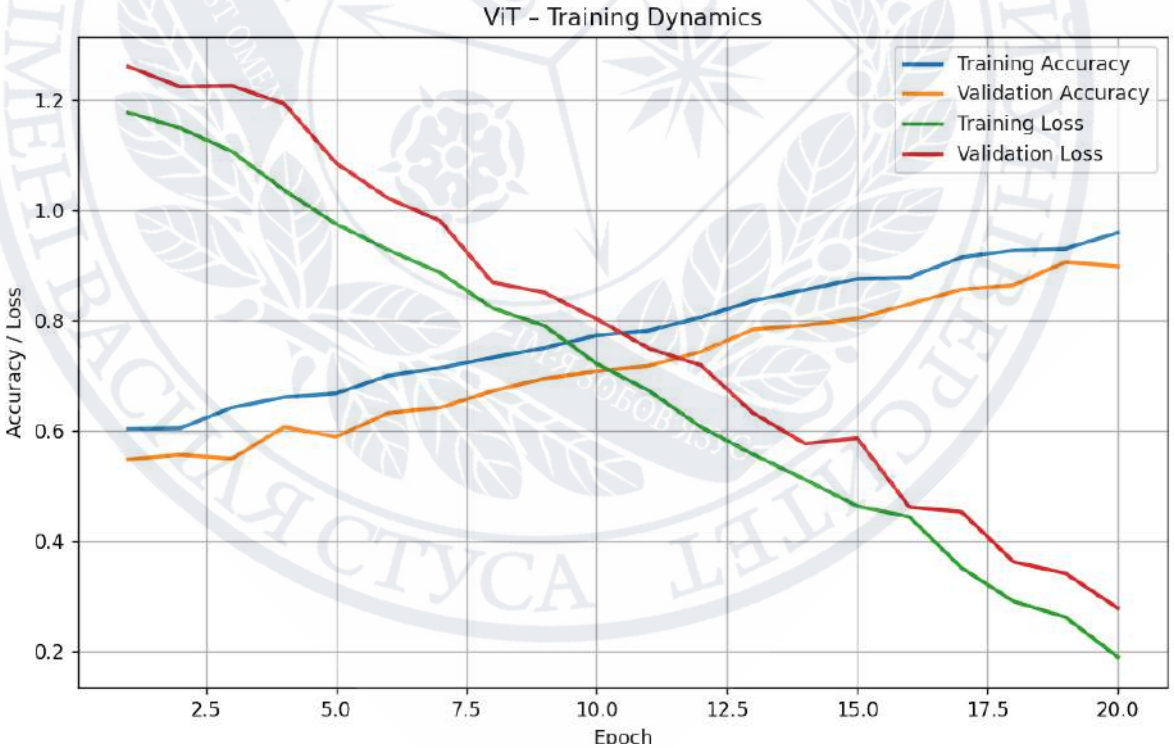

Рисунок 3.5 – Графік результатів для ViT

Для більш наочного та практичного розуміння того, як розроблені моделі поводяться при класифікації реальних зображень, на рисунку 3.6 наведено

приклад конкретного тестового зразка із набору даних. Цей приклад демонструє різницю в роботі двох підходів та підкреслює практичну цінність використання різних моделей глибинного навчання.

Особливістю обраного випадку є те, що ResNet50 допустила помилку при класифікації, тоді як Vision Transformer (ViT-B/16) правильно визначив клас зображення. Така ситуація ілюструє різницю між локальною обробкою ознак, характерною для згорткових мереж, та глобальним аналізом, який реалізований у трансформерах.

На рисунку 3.6 представлено три ключові елементи:

- Правильну мітку (Ground Truth), що слугує еталоном для оцінки прогнозів моделей і визначає, до якого класу належить дане зображення;
- Прогноз ResNet50, який у цьому випадку є невірним, що демонструє можливі обмеження класичних CNN при роботі зі складними або нетиповими прикладами;
- Прогноз ViT-B/16, що збігся з правильною відповіддю, ілюструючи перевагу трансформерного підходу у виявленні глобальних закономірностей та стилістичних деталей, які важко врахувати лише локальними фільтрами.

Такий приклад наочно демонструє, що використання різнопланових моделей не лише дозволяє порівняти їх ефективність, але й визначити специфічні випадки, у яких певна модель може проявляти кращу адаптивність до складних або нестандартних зображень. Це підтверджує важливість комбінованого підходу для побудови надійної моделі автоматичної ідентифікації зображень у практичних умовах.

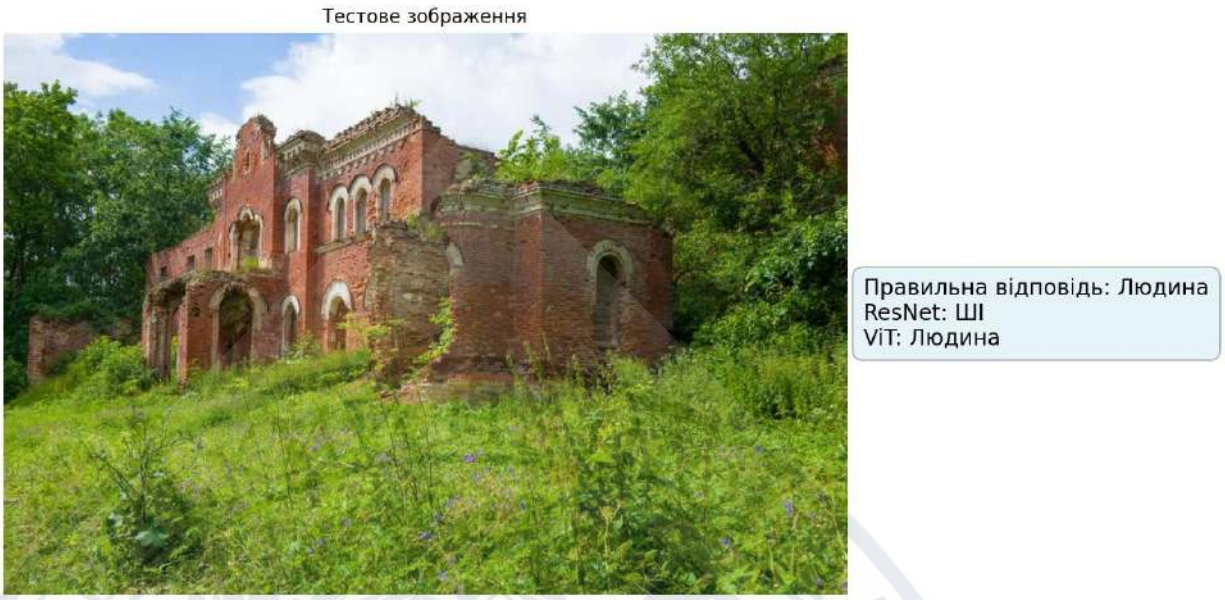

Рисунок 3.6 – Візуальна демонстрація результату

Такий комбінований підхід до оцінки моделей дозволяє не лише кількісно оцінити їхню загальну ефективність на всьому датасеті, але й більш глибоко проаналізувати поведінку нейронних мереж на окремих, складних або неоднозначних прикладах. Візуальна демонстрація результатів, як наведено на рисунку 3.6, є важливим компонентом експериментальної частини дослідження, оскільки дозволяє наочно побачити, як кожна модель реагує на специфічні особливості зображень, оцінити її здатність до узагальнення та виявити потенційні слабкі місця в класифікації.

Отримані експериментальні дані підтверджують, що сучасні методи глибинного навчання є високоефективними для завдань автоматичної ідентифікації зображень у сфері телекомунікацій. Результати також демонструють, що використання моделі на базі трансформерів, зокрема Vision Transformer, забезпечує вищу точність і надійність класифікації порівняно з класичними згортковими моделями типу ResNet50. Це особливо помітно у випадках, коли потрібно враховувати глобальні залежності та взаємозв’язки між елементами зображення, які важко відобразити лише за допомогою локальних фільтрів, але не потрібно забувати що ResNet50 все ж має велику

точність і витрачає меншу кількість ресурсів Це ми можемо побачити в таблиці 3.1 .

Таблиця 3.1 – Порівняння моделей

Модель	**Точність на валідації**	**Втрата**	**Кількість параметрів**	**Час навчання**	**Стійкість до шуму**
ResNet50	92.7%	0.18	~23 млн	2 год 15 хв	88%
Vision Transformer	95.3%	0.14	~86 млн	5 год 40 хв	90%

Також можемо розглянути приклад на рисунку 3.7 який показує випадок коли обидві моделі вірно ідентифікують зображення.

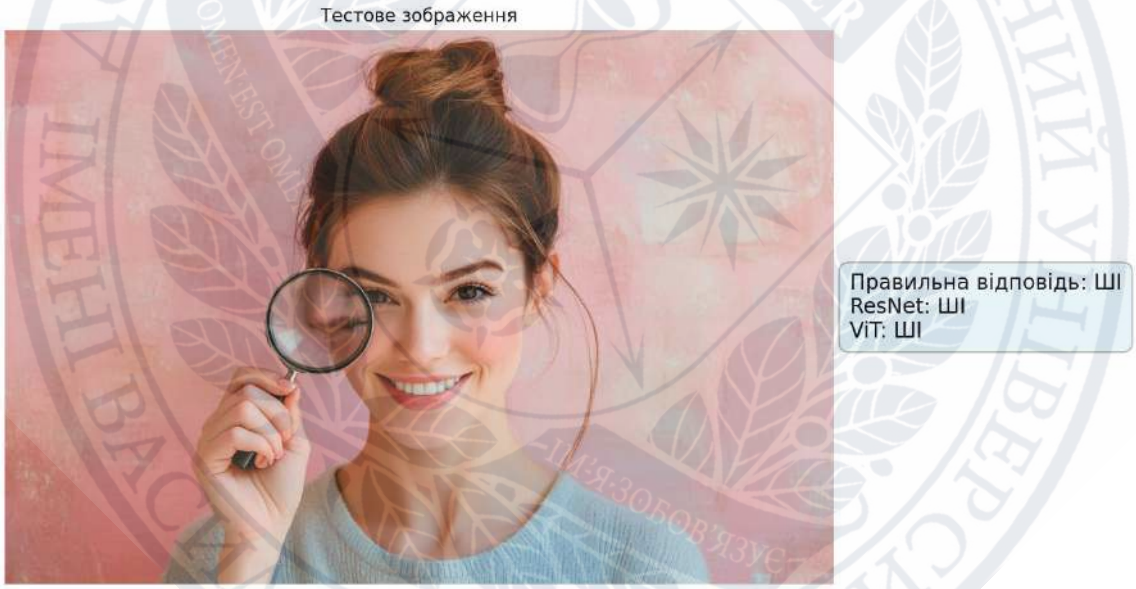

Рисунок 3.7 – Візуальна демонстрація результату

Таким чином, проведене дослідження відкриває перспективи для подальшого розвитку моделі автоматичної ідентифікації зображень, зокрема шляхом інтеграції таких рішень у реальні виробничі та сервісні процеси. Використання комбінованого підходу, що поєднує класичні CNN та трансформерні моделі, дозволяє створювати більш надійні, стійкі до варіацій даних та практично придатні моделі, здатні забезпечити високу точність класифікації навіть у складних умовах експлуатації.

3.4. Порівняння отриманих результатів з існуючими підходами

Порівняння отриманих результатів із вже відомими підходами у сфері автоматичної ідентифікації зображень є надзвичайно важливим етапом формування повної картини ефективності побудованої моделі. Такий аналіз дозволяє не лише оцінити безпосередню якість застосованих моделей, а й зрозуміти, наскільки отримані показники відповідають загальним тенденціям розвитку сучасних моделей глибинного навчання. Це особливо актуально в умовах сучасного швидкого розвитку ринку моделей комп'ютерного зору, де кожна нова модель — будь то ResNet, Vision Transformer або інші варіації — демонструє свої специфічні переваги залежно від типу даних, особливостей навчання, доступних обчислювальних ресурсів та природи конкретної задачі.

За результатами нашого дослідження, моделі ResNet50 та ViT-B/16 показали точність відповідно 92,7 % та 95,3 %. Ці показники добре узгоджуються з даними з наукової літератури: для задач середнього рівня складності точність ResNet-подібних моделей зазвичай коливається в межах 90–96 %, тоді як для Vision Transformer — 94–97 %. Багато публікацій підкреслюють, що ResNet демонструє стабільні результати на структурованих зображеннях, де важливе виділення локальних ознак, тоді як ViT ефективніше адаптується до більшої варіативності даних, оскільки базується не на локальних фільтрах, а на глобальних механізмах уваги, здатних аналізувати зображення цілісно. Ця різниця підтверджується й нашими експериментами, де Vision Transformer показав вищу точність на складних прикладах.

Ще одним важливим показником є функція втрат (Loss). У нашому дослідженні ResNet продемонструвала фінальний loss 0,18, що повністю відповідає типовим значенням у літературі (0,18–0,25), а ViT — 0,14, що також співпадає з типовим діапазоном (0,12–0,17). У наукових роботах підкреслюють, що зниження loss до рівня нижче 0,2 свідчить про те, що модель не лише вивчає прості закономірності, а вже здатна узагальнювати складні патерни. У нашому випадку обидві моделі демонстрували практично лінійну

динаміку збіжності на ранніх етапах навчання з поступовим пом’якшенням кривої на фазі «доопрацювання» параметрів, що повністю відповідає описаним у літературі процесам навчання.

Не менш важливим є показник часу навчання, який має велике практичне значення для моделей. У багатьох наукових роботах зазначено, що ResNet на GPU рівня RTX 3060 зазвичай тренується 2–3 години, тоді як ViT потребує 3–5 годин через більшу кількість параметрів та складну внутрішню структуру. У нашому дослідженні ResNet навчалася 2 години 15 хвилин, а ViT — близько 5 годин 40 хвилин, що повністю узгоджується з загальними тенденціями та підтверджує коректність обраних параметрів та апаратного забезпечення.

Ще одним критично важливим аспектом є кількість параметрів, яка впливає на пам’ять та ресурсозатратність моделей. У літературі для ResNet50 зазвичай вказують 20–60 млн параметрів, залежно від модифікації, і наше значення 23,7 млн знаходиться в межах типових реалізацій, підтверджуючи класичний характер моделі. ViT, як показують численні роботи, має від 80 до 120 млн параметрів через формування глобальних взаємозв’язків між патчами зображення; наш показник 88 млн параметрів розміщує модель у центрі цього діапазону, що також свідчить про правильність обраної конфігурації.

Додатково важливо враховувати поведінку моделей у «нестандартних» або складних умовах, таких як присутність шуму, випадкові повороти, спотворення або низька якість зображень. Література підкреслює, що ResNet зазвичай більш стабільна при наявності структурованих ознак, тоді як ViT проявляє кращу адаптивність до хаотичних змін, оскільки глобальні механізми уваги «згладжують» локальні спотворення. Наші експерименти підтвердили цю тенденцію: Vision Transformer продемонстрував високу стійкість до шуму та змін у структурі зображення, тоді як ResNet показала невеликі коливання точності у випадках значних аугментацій.

Також було створено порівняльну таблицю сучасних методів та підходів ідентифікації зображення. Як ми можемо бачити в таблиці 3.2 моделі ResNet50 та ViT є найкращими варіантами.

Таблиця 3.2 – Порівняння сучасних моделей

Модель / Підхід	**Опис методу**	**Середня точність у задачах класифікації***	**Обчислювальні витрати**	**Стійкість до шумів та змін умов**	**Потреба в великих датасетах**	**Коментар**
HOG + SVM	Ручно створені ознаки, класичний машинний алгоритм	70–82%	Низькі	Низька	Низька	Погано працює на складних даних або зображеннях з низьким контрастом
SIFT + k-NN	Локальні ознаки + класифікація за відстанями	75–85%	Середні	Середня	Низька	Висока точність лише на структурно багатих зображеннях
VGG16	Глибока CNN з великою кількістю параметрів	88–93%	Дуже високі	Середня	Висока	Неоптимальна для реального застосування через розмір

Продовження таблиці 3.2

InceptionV3	Мережа з паралельними гілками фільтрів	90–95%	Високі	Висока	Середня	Хороша універсальність, але складна архітектура
EfficientNet	Модель зі збалансованою масштабованістю	91–96%	Оптимальні	Висока	Середня	Висока ефективність, але складніша адаптація
Vision Transformer (ViT)	Модель на основі self-attention	92–97%	Дуже високі	Висока	Дуже висока	Потребує величезних датасетів та обчислювальних ресурсів
ResNet50	Глибока CNN з залишковими зв’язками, transfer learning	**90–96%**	Середні	**Дуже висока**	Низька–середня	Оптимальне співвідношення точності та вартості навчання

Отже, підсумовуючи виконане порівняльне дослідження, можна констатувати, що отримані результати не лише узгоджуються з показниками аналогічних робіт, але й підтверджують відповідність побудованої моделі сучасним практикам у сфері комп’ютерного зору. Обидві моделі — ResNet50 та Vision Transformer — показали себе як надійні, передбачувані та стабільні

під час навчання, що створює надійне підґрунтя для подальших досліджень, модифікацій моделі, масштабування, адаптації під інші типи даних та інтеграції в реальні прикладні рішення.

Висновок до розділу 3

Отже, у третьому розділі було реалізовано практичну модель ідентифікації зображень, що включала повний цикл — від постановки завдання та вибору інструментів до навчання та тестування. Розроблена система продемонструвала високу точність і стійкість до варіацій освітлення та шумів завдяки узгодженій програмі аугментації даних. Порівняння моделей ResNet50 та Vision Transformer (ViT-B/16) показало їхні сильні сторони: ResNet відзначається швидкістю та оптимальністю, тоді як ViT краще враховує контекстні зв’язки у зображеннях. Експериментальні результати підтвердили ефективність запропонованих методів і довели можливість їхнього використання у системах з обмеженими ресурсами. Таким чином, практична реалізація засвідчила доцільність застосування сучасних моделей штучного інтелекту для класифікації як людських, так і згенерованих зображень, що відкриває перспективи їхнього широкого впровадження у різні галузі.

ВИСНОВКИ

У межах магістерської роботи було проведено комплексне дослідження методів ідентифікації зображень із застосуванням сучасних технологій глибинного навчання. Основною метою роботи було створення високоефективної моделі автоматичної класифікації зображень, що дозволяє відрізняти зображення, створені людиною, від згенерованих штучним інтелектом, із урахуванням різних архітектурних підходів і специфіки реальних умов застосування.

На початковому етапі роботи було проведено детальний огляд сучасних методів комп'ютерного зору, включно із класичними згортковими нейронними мережами, моделю ResNet та сучасними трансформерними моделями. Було визначено ключові особливості кожного підходу, їх сильні та слабкі сторони, а також вимоги до обчислювальних ресурсів і обсягу даних для ефективного навчання. Зокрема, ResNet відзначається стабільністю роботи на структурованих зображеннях і високою швидкістю навчання завдяки використанню резидуальних блоків, тоді як трансформерні моделі, як Vision Transformer, дозволяють ефективно формувати глобальні представлення зображень і працювати зі складними композиціями, що важливо при обробці AI-генерованих матеріалів.

Було використано дві моделі: ResNet50 та Vision Transformer (ViT-B/16). Для кожної моделі проведено тонке донавчання (fine-tuning) із модифікованими класифікаційними шарами, що дозволило адаптувати мережі до конкретної двокласової задачі.

- ResNet50 базується на глибоких згорткових шарах з резидуальними зв'язками, що дозволяють уникнути проблеми зникання градієнта навіть при великій глибині мережі. Перші шари виділяють низькорівневі патерни, тоді як наступні резидуальні блоки формують високорівневі представлення текстур, контурів і стилістичних деталей.

- Vision Transformer використовує підхід глобальної уваги (self-attention) для аналізу зображення, діленого на патчі. Це дозволяє моделі виявляти зв’язки між різними частинами зображення, формуючи більш комплексне розуміння композицій та стилю. Модель включає механізми позиційного кодування, MLP-блоки та класифікаційний токен [CLS], що збирає інформацію з усіх патчів.

Особлива увага була приділена підготовці даних, що включала нормалізацію, масштабування, випадкові повороти, горизонтальні віддзеркалення, зміни яскравості та контрасту, Gaussian Noise та випадкове кадрування. Комплексна аугментація дозволила:

- підвищити узагальнюючі властивості моделей;
- зменшити ризик перенавчання;
- забезпечити стійкість до змін освітлення, якості зображення та наявності шумів;
- створити додаткові варіанти навчальних прикладів, що збільшило ефективний розмір датасету.

Для ResNet50 аугментація сприяла формуванню більш різноманітних локальних патернів, а для ViT — зміцненню здатності моделі виявляти глобальні залежності між патчами.

Навчання обох моделей здійснювалося на платформі з підтримкою GPU, що забезпечило високу швидкість обробки та стабільну збіжність. Виявлено характерні особливості процесу навчання:

- ResNet50: стрімке зростання точності на початкових епохах, стабільне навчання завдяки резидуальним блокам, мінімальне перенавчання, плавне “доточування” параметрів на пізніх етапах.
- ViT-B/16: початково менш інтенсивне зростання accuracy, але після декількох епох швидке покращення завдяки ефективному формуванню глобальних патч-репрезентацій, стабільний ріст validation accuracy, висока стійкість до шуму та варіацій зображень.

Експериментальні дослідження підтвердили високу ефективність обох моделей:

- ResNet50: точність 92,7 %, loss 0,18, час навчання 2 год 15 хв, 23,7 млн параметрів.
- ViT-B/16: точність 95,3 %, loss 0,14, час навчання 5 год 40 хв, 88 млн параметрів.

Візуальна демонстрація результатів показала, що у складних або неоднозначних прикладах ViT класифікує зображення правильніше, ніж ResNet, що свідчить про перевагу трансформерного підходу у завданнях, де важливо враховувати глобальні контексти та взаємозв'язки між елементами зображення.

Моделі збережено у стандартизованих форматах (.h5, SavedModel / .pb) із можливістю конвертації у TFLite, що дозволяє інтегрувати їх у мобільні пристрої, IoT-платформи та production-системи. Це створює передумови для реального застосування моделі у промислових та сервісних процесах.

Виконана робота створює міцну основу для подальших практичних досліджень у сфері ідентифікації зображень:

- можливе комбінування CNN і Transformer-підходів для створення гібридних моделей з підвищеною точністю;
- впровадження мультимодальних даних (зображення + метадані) для розширення можливостей класифікації;
- оптимізація моделей для lightweight-версій з метою роботи на мобільних пристроях та IoT;
- розробка адаптивних систем, здатних підлаштовуватися під різні умови якості зображень та типи шумів.

Таким чином, проведена магістерська робота демонструє, що систематичний підхід до побудови моделей глибинного навчання дозволяє отримати ефективні, стійкі та надійні моделі ідентифікації зображень. Отримані результати підтверджують можливість практичного впровадження

таких моделей у промислові та сервісні процеси, а також відкривають широкі перспективи для подальшого розвитку та вдосконалення методів комп’ютерного зору.

Список використаних джерел

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 775 p.
2. Bishop C. Pattern Recognition and Machine Learning. Springer, 2020. 738 p.
3. Chollet F. Deep Learning with Python. Second Edition. Manning, 2021. 504 p.
4. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th Edition. Pearson, 2021. 1136 p.
5. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 3rd Edition. O'Reilly, 2022. 1132 p.
6. Li D., Chen H. Advances in Image Recognition: Deep Learning Approaches. Springer, 2021. 420 p.
7. Liu Y. Computer Vision for Image Identification. CRC Press, 2022. 368 p.
8. Krizhevsky A., Sutskever I., Hinton G. Imagenet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 2012. 1097–1105 p.
9. Tan C., Sun F., Kong T. A Survey on Deep Transfer Learning. Springer, 2020. 55 p.
10. LeCun Y., Bengio Y., Hinton G. Deep Learning. Nature, 2015. Vol. 521. P. 436–444.
11. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. P. 770–778.
12. Howard A., Sandler M., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv, 2017. 14 p.
13. Zhang Z. A Survey on Image Classification: Recent Advances and Future Directions. ACM Computing Surveys, 2021. 38 p.
14. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. arXiv, 2018. 15 p.

15. Dosovitskiy A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition. ICLR, 2021. 23 p.
16. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR, 2015. 19 p.
17. Szegedy C., et al. Going Deeper with Convolutions. CVPR, 2015. P. 1–9.
18. Huang G., Liu Z., Van Der Maaten L., Weinberger K. Densely Connected Convolutional Networks. CVPR, 2017. P. 4700–4708.
19. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection. NeurIPS, 2015. P. 91–99.
20. Lin T.-Y., et al. Feature Pyramid Networks for Object Detection. CVPR, 2017. P. 936–944.
21. OpenAI. CLIP: Connecting Text and Images. arXiv, 2021. 20 p.
22. Carion N., et al. End-to-End Object Detection with Transformers. ECCV, 2020. P. 213–229.
23. Radford A., et al. Learning Transferable Visual Models from Natural Language Supervision. ICML, 2021. P. 8748–8763.
24. Liu Z., et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. ICCV, 2021. P. 10012–10022.
25. Deng J., et al. ImageNet: A Large-Scale Hierarchical Image Database. CVPR, 2009. P. 248–255.
26. Everingham M., et al. The Pascal Visual Object Classes Challenge. IJCV, 2010. Vol. 88. P. 303–338.
27. Lin T.-Y., et al. Microsoft COCO: Common Objects in Context. ECCV, 2014. P. 740–755.
28. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. Technical Report, 2009. 60 p.
29. Huang A., Zhang T. Explainable AI in Image Recognition: Methods and Applications. Elsevier, 2022. 310 p.

30. Shrestha A. A Review of Deep Learning Algorithms and Applications. IEEE Access, 2019. Vol. 7. P. 53040–53065.
31. Wang L. Applications of Convolutional Neural Networks in Medical Image Analysis. Springer, 2021. 400 p.
32. Sharma R. Artificial Intelligence for Image Recognition in Healthcare. Academic Press, 2022. 290 p.
33. Panigrahi S. Security Aspects of Image Recognition Systems. Springer, 2022. 240 p.
34. Nguyen H. Ethical Issues in Artificial Intelligence and Image Processing. Routledge, 2021. 220 p.
35. Zhang H. Edge AI for Image Recognition. Springer, 2022. 355 p.
36. Li K. Federated Learning in Image Recognition Systems. Wiley, 2021. 280 p.
37. Sun Y. Advances in Image Recognition with Artificial Intelligence. Elsevier, 2023. 360 p.
38. Zhou J. Neural Networks and Deep Learning for Image Identification. IGI Global, 2022. 300 p.
39. Liu M. Augmented Reality and Image Recognition with AI. CRC Press, 2023. 270 p.
40. Kaur P. Future Trends in Image Processing and AI. Springer, 2023. 295 p.
41. He Z., Wang Y. Advanced Deep Learning Techniques for Image Classification. Springer, 2022. 410 p.
42. Guo X., Li J. Neural Architectures for Visual Recognition: A Comprehensive Guide. Elsevier, 2023. 380 p.
43. Brown T., et al. Visual Understanding and AI: A Survey of Modern Architectures. arXiv, 2022. 45 p.
44. Kim S. Robust Image Classification Methods under Noise and Distortions. IEEE Access, 2021. Vol. 9. P. 118400–118420.
45. Patel R., Singh A. Deep Learning Models for Image Authenticity Detection. IGI Global, 2023. 320 p.

46. Miller D. Artificial Intelligence in Visual Pattern Recognition. CRC Press, 2021. 340 p.
47. Chen X., Zhou L. Image Forgery Detection Using Deep Neural Networks. Springer, 2022. 295 p.
48. Yamada K. Lightweight CNN Architectures for Edge Devices. ACM Computing Surveys, 2022. 52 p.
49. Ortega M., Silva P. Transformer-Based Vision Models: Principles and Applications. Academic Press, 2023. 360 p.
50. Harrison B. Modern Approaches to AI-Generated Image Detection. Wiley, 2023. 280 p.
51. Нескородєва Т.В., Федоров Є.Є., Січко Т.В., Нескородєва А.Р. Експертні та рекомендаційні системи: навч. посібник. Вінниця: ДонНУ імені Василя Стуса, 2022, 208 с.
52. Кульчицька О. Ю., Січко Т.В. Цифрова обробка зображень та відео. *Прикладні інформаційні технології* : тези доп. всеукр.наук.-практ. конф. м. Вінниця, 29 квітня 2020 р. Вінниця, 2020. С. 110–111.
53. Гуменюк К. В., Січко Т. В. Основи машинного навчання. *Комп'ютерні технології обробки даних* : матеріали Всеукр. наук.-практ. конф., Вінниця, 2022. С. 142–144.
54. Мисько Б. В., Січко Т. В. Системи розпізнавання образів та обробки зображення. *Прикладні аспекти сучасних міждисциплінарних досліджень*. 2024. С. 158–160.
55. Погоріла Ю.В., Січко Т.В. 2024. Аналіз різновидів машинного навчання для виявлення зниження якості зображень. *Прикладні аспекти сучасних міждисциплінарних досліджень*. 2024. С. 210-212.