

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту  
(повна назва)

Кафедра Інформатики  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

**РОЗРОБЛЕННЯ ТА ДОСЛІДЖЕННЯ МЕТОДІВ ВИЯВЛЕННЯ**  
**ДЕФЕКТІВ СОНЯЧНИХ ПАНЕЛЕЙ НА ОСНОВІ АНАЛІЗУ**  
**ЗОБРАЖЕНЬ ІЗ ЗАСТОСУВАННЯМ КОМП'ЮТЕРНОГО ЗОРУ**  
(тема)

Виконав:  
здобувач 2 року навчання,  
групи ІНФМ-24-2

Хомсі Как С.М.  
(прізвище, ініціали)

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

Тип програми освітньо-професійна

Освітня програма Інформатика  
(повна назва освітньої програми)

Науковий керівник зав. каф. Кобилін О. А.  
(посада, прізвище, ініціали)

Допускається до захисту

Завідувач кафедри інформатики \_\_\_\_\_  
(підпис)

Кобилін О. А.  
(прізвище, ініціали)

2025 р.

## Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджментуКафедра ІнформатикиРівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки  
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 2025 р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУздобувачеві Хомсі Как Саміру Мохамедовичу  
(прізвище, ім'я, по батькові)1. Тема роботи Розроблення та дослідження методів виявлення дефектів сонячних панелей зображеннях із застосуванням комп'ютерного зорузатверджена наказом університету від 14 листопада 2025 року № 1045Ст2. Термін подання здобувачем роботи до екзаменаційної комісії 23 листопада 2025 р.3. Вихідні дані до роботи методи класифікації зображень, літературні джерела щодо виявлення дефектів, матеріали конференцій, зображення сонячних панелей з дефектами для тренування та тестування, програмні засоби: Python, TensorFlow, PyTorch, Google Colab

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

1. Аналіз сучасних методів діагностики дефектів сонячних панелей та архітектури нейронних мереж для класифікації зображень.2. Підготовка набору даних зображень сонячних панелей для бінарної класифікації.3. Реалізація методів виявлення дефектів на основі MobileNetV2, VGG19 та CLIP.4. Експериментальне дослідження та порівняльний аналіз реалізованих методів.5. Формування практичних рекомендацій щодо вибору методу діагностики.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) актуальність проблеми виявлення дефектів сонячних панелей, об'єкт та мета дослідження, постановка задачі, архітектури нейронних мереж MobileNetV2, VGG19 та мовної моделі CLIP, приклади зображень набору даних, результати експериментальних досліджень моделей, порівняльний аналіз методів, висновки, рекомендації та апробація роботи.

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	29.09.2025	
2	Аналіз завдання, підбір літератури	30.09.25-07.10.25	
3	Аналіз літератури з досліджуваної проблеми	08.10.25-14.10.25	
4	Дослідження архітектур нейронних мереж	15.10.25-20.10.25	
5	Розроблення алгоритмів виявлення дефектів	23.10.25-29.10.25	
6	Програмна реалізація	31.10.25-07.11.25	
7	Обґрунтування отриманих результатів	08.11.25-11.11.25	
8	Оформлення пояснювальної записки	12.11.25-14.11.25	
9	Перевірка на нормоконтроль	21.12.25	
10	Перевірка на плагіат	22.11.25	
11	Рецензування	23.11.25	
12	Підготовка презентації та доповіді	28.11.25	
13	Занесення роботи в електронний архів	01.12.25	
14	Попередній захист кваліфікаційної роботи	01.12.25	

Дата видачі завдання 29 вересня 2025 р.

Здобувач \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_  
(підпис)

зав. каф. Кобилін О. А.  
(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 79 с., 8 табл., 25 рис., 1 дод., 36 джерел.

БІБЛІОТЕКА TENSORFLOW, ВИЯВЛЕННЯ ДЕФЕКТІВ, ГЛИБОКЕ НАВЧАННЯ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, КЛАСИФІКАЦІЯ ЗОБРАЖЕНЬ, КОМП'ЮТЕРНИЙ ЗІР, МАШИННЕ НАВЧАННЯ, СОНЯЧНІ ПАНЕЛІ, BINARY CLASSIFICATION, CLIP, CONFUSION MATRIX, DROPOUT, EARLY STOPPING, MOBILENETV2, TRANSFER LEARNING, VGG19

Об'єктом дослідження є зображення сонячних панелей з різними дефектами та забрудненнями.

Предметом дослідження є методи комп'ютерного зору і машинного навчання для автоматизованого розпізнавання стану панелей.

Метою дослідження є створення ефективного методу виявлення дефектів на основі аналізу зображень із застосуванням згорткових нейронних мереж і vision language моделей для підвищення точності діагностики.

Використано Transfer Learning з MobileNetV2 і VGG19 та zero shot підхід на основі CLIP. Реалізовано бінарну класифікацію, застосовано Dropout, L2 регуляризацію, Early Stopping і ReduceLROnPlateau. Створено систему підготовки даних, навчання моделей і оцінки їх продуктивності за accuracy, precision, recall і confusion matrix.

Наукова новизна полягає в адаптації MobileNetV2 для задачі класифікації та порівнянні з VGG19 і CLIP, що дозволило визначити ефективність різних підходів. Практична цінність роботи полягає у можливості інтеграції системи в автоматизовані рішення моніторингу сонячних електростанцій.

Результати дослідження отримані в межах міжнародного проєкту INITIATE за грантом No. 101136775 HORIZON WIDERA 2023 ACCESS 03.

## ABSTRACT

Explanatory note to the qualification work: 79 pages, 8 table, 25 figures, 1 appendix, 36 sources.

BINARY CLASSIFICATION, CLIP, COMPUTER VISION, CONFUSION MATRIX, CONVOLUTIONAL NEURAL NETWORKS, DEEP LEARNING, DEFECT DETECTION, DROPOUT, EARLY STOPPING, IMAGE CLASSIFICATION, MACHINE LEARNING, MOBILENETV2, SOLAR PANELS, TENSORFLOW LIBRARY, TRANSFER LEARNING, VGG19.

The object of the research is images of solar panels with various types of defects and contamination.

The subject of the research is computer vision and machine learning methods for automated detection and classification of solar panel defects.

The aim of the research is to create an effective defect detection method based on image analysis using convolutional neural networks and vision language models to improve diagnostic accuracy.

The methods used include Transfer Learning with MobileNetV2 and VGG19 was used, as well as a zero-shot approach based on CLIP. Binary classification was implemented; Dropout, L2 regularization, Early Stopping, and ReduceLROnPlateau were applied. A system for data preparation, model training, and performance evaluation was developed using accuracy, precision, recall, and confusion matrix.

The scientific novelty lies in adapting MobileNetV2 for the classification task and comparing it with VGG19 and CLIP, which made it possible to assess the effectiveness of different approaches. The practical value is the potential integration of the system into automated monitoring solutions for solar power plants.

The results of this research were obtained under the international project INITIATE under the grant No. 101136775 HORIZON WIDERA 2023 ACCESS 03.

## ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів .....	8
Вступ.....	9
1 Аналіз предметної області та постановка задачі .....	11
1.1 Огляд сонячної енергетики та проблеми діагностики.....	11
1.2 Типи дефектів сонячних панелей.....	13
1.3 Огляд існуючих методів діагностики .....	14
1.4 Застосування комп'ютерного зору та ML.....	17
1.5 Огляд архітектур нейронних мереж .....	18
1.6 Постановка задачі дослідження .....	20
2 Теоретичні основи методів класифікації дефектів сонячних панелей .....	22
2.1 Математична формалізація задачі бінарної класифікації.....	22
2.2 Сучасні нейронні архітектури для аналізу зображень.....	25
2.2.1 Математичні основи згорткових операцій .....	26
2.2.2 Архітектура MobileNetV2.....	29
2.2.3 Архітектура VGG19 .....	33
2.2.4 Архітектура та принцип роботи CLIP.....	34
2.3 Методи регуляризації та оптимізації навчання .....	37
2.4 Теоретичні основи препроцесингу зображень .....	40
3 Дослідження та розроблення методів виявлення дефектів сонячних панелей .....	43
3.1 Підготовка даних для навчання моделей класифікації.....	43
3.1.1 Опис набору даних.....	43
3.1.2 Підготовка та розбиття даних .....	44
3.1.3 Параметри завантаження даних.....	45
3.1.4 Експериментальне середовище.....	46
3.2 Реалізація методів класифікації .....	46
3.2.1 Реалізація MobileNetV2 з використанням перенесеного навчання (transfer learning).....	47

3.2.2 Реалізація VGG19 з використанням перенесеного навчання (transfer learning).....	48
3.2.3 Реалізація CLIP для класифікації без прикладів (zero-shot) .....	49
3.3 Результати експериментів.....	50
3.3.1 Результати класифікації за допомогою MobileNetV2 .....	50
3.3.2 Результати класифікації за допомогою моделі VGG19 .....	56
3.3.3 Результати класифікації без навчання на цільовому наборі даних із використанням CLIP .....	61
3.4 Порівняльний аналіз досліджених методів.....	66
3.5 Рекомендації для майбутніх досліджень.....	70
Висновки .....	73
Перелік джерел посилання.....	75
Додаток А Скріншоти сторінок коду програми.....	78

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

- БПЛА – безпілотні летальні апарати
- ШІ – штучний інтелект
- Accuracy – точність класифікації
- Adam – Adaptive Moment Estimation (алгоритм оптимізації)
- AI – Artificial Intelligence (штучний інтелект)
- API – Application Programming Interface (інтерфейс програмування застосунків)
- CNN – Convolutional Neural Network (згорткова нейронна мережа)
- CV – Computer Vision (комп'ютерний зір)
- DL – Deep Learning (глибоке навчання)
- Dropout – метод регуляризації для запобігання перенавчанню
- Epoch – епоха навчання (повний прохід через набір даних)
- F1-Score – F1-міра (гармонічне середнє precision та recall)
- FN – False Negative (хибнонегативний результат)
- IoU – Intersection over Union (перетин над об'єднанням)
- ML – Machine Learning (машинне навчання)
- Precision – точність (прецизійність)
- Recall – повнота
- ReLU – Rectified Linear Unit (випрямлена лінійна одиниця)
- Transfer Learning – трансферне навчання (використання попередньо навченої моделі)

## ВСТУП

Сонячні електростанції стають дедалі популярнішими як у промислових масштабах, так і для приватного використання. Проте успішна експлуатація таких систем вимагає постійного контролю технічного стану обладнання.

Під час роботи сонячні панелі піддаються впливу різних факторів, які знижують їх продуктивність. Як показано в попередніх дослідженнях [1], забруднення поверхні панелей може призводити до зниження ефективності від 10% у помірних умовах до понад 40% у посушливих регіонах, а в найгірших сценаріях втрати сягають 60% без регулярного технічного обслуговування. Проблема полягає в тому, що багато дефектів не можна помітити неозброєним оком на ранніх стадіях, а коли вони стають очевидними, втрати енергії вже є значними.

Існуючі методи перевірки стану панелей мають свої недоліки. Ручний огляд займає багато часу і підходить хіба що для невеликих установок. Термографічне обстеження потребує спеціального обладнання та навичок його використання. Електричні тести часто потребують зупинки роботи системи [1]. Для великих сонячних ферм такі підходи стають надто затратними і неефективними.

Економічні наслідки несвоєчасного виявлення дефектів є значними. За оцінками експертів, типова сонячна електростанція потужністю 1 МВт може втрачати від 50 до 150 тисяч доларів на рік [1] через непомічені вчасно дефекти панелей. Мікротріщини, які на ранніх стадіях можна виявити лише спеціальними методами, з часом призводять до повного виходу з ладу фотоелементів. Вартість заміни пошкодженої панелі зазвичай у 5-10 разів перевищує витрати на профілактичне обслуговування.

Практика показує, що відсутність систематичного моніторингу призводить до каскадних відмов обладнання. Один пошкоджений фотоелемент у послідовному з'єднанні може знизити продуктивність цілої лінії панелей. За даними операторів великих сонячних електростанцій, до 15-20% панелей можуть мати приховані дефекти, які поступово знижують загальну ефективність

системи. Це означає, що інвестиції в сонячну енергетику не окуповуються у запланований термін, а власники установок несуть додаткові фінансові втрати.

Сучасні технології комп'ютерного зору відкривають нові можливості для автоматизації процесу діагностики. Нейронні мережі вміють розпізнавати патерни на зображеннях і класифікувати різні типи дефектів. Використання дронів для зйомки панелей у поєднанні з алгоритмами аналізу зображень дозволяє швидко перевіряти великі площі без участі людини. Такий підхід економить час, зменшує ризики для персоналу і забезпечує більш об'єктивну оцінку стану обладнання.

Дефекти сонячних панелей бувають різними за природою та наслідками. Це можуть бути забруднення поверхні, мікротріщини в фотоелементах, пошкодження від граду чи птахів, корозія контактів [1], проблеми з електричними з'єднаннями. Кожен тип дефекту має свої візуальні ознаки, які можна навчити розпізнавати комп'ютерну систему.

Актуальність роботи визначається зростаючою потребою енергетичної галузі в ефективних інструментах моніторингу стану сонячних електростанцій. Автоматизована діагностика на основі комп'ютерного зору може суттєво знизити витрати на обслуговування, підвищити надійність енергопостачання та продовжити термін експлуатації обладнання. Розвиток таких технологій необхідний для України, де відновлювана енергетика набирає обертів.

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

## 1.1 Огляд сонячної енергетики та проблеми діагностики

Сонячна енергетика переживає період безпрецедентного зростання у світовому масштабі. У 2024 році світ встановив рекордні 597 ГВт сонячних потужностей, що становить зростання на 33% порівняно з 2023. Глобальна потужність відновлюваної енергетики зросла на рекордні 15,1% у 2024 році, досягнувши 4 448 гігават, при цьому близько 585 ГВт потужності було додано, головним чином завдяки розширенню сонячної та вітрової енергії [2]. Ці показники свідчать про стрімку трансформацію світової енергетичної системи та зростаючу роль фотоелектричних технологій у забезпеченні енергетичної безпеки (рис. 1.1).

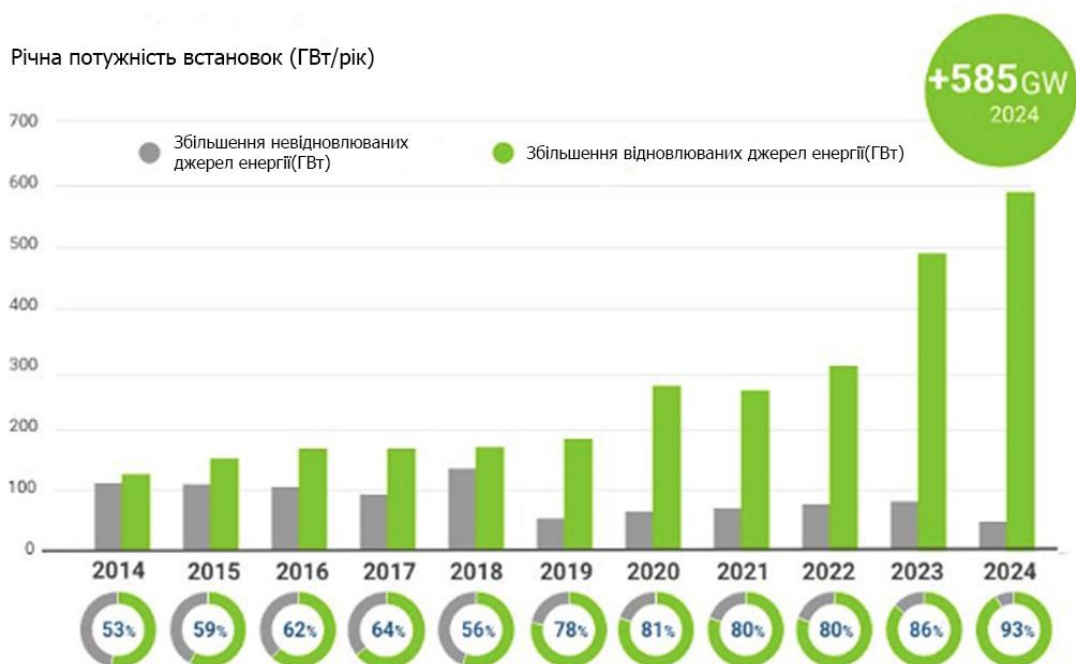


Рисунок 1.1 – Графік глобального зростання сонячних потужностей 2014-2024

Лідером у впровадженні сонячної енергетики залишається Китай. До кінця 2024 року Китай встановив майже 64% нових потужностей відновлюваної

енергетики, включаючи понад половину (278 ГВт) нових сонячних потужностей, при цьому загальна потужність відновлюваних джерел досягла 4,4 ТВт [3], що становить 46% від глобальної встановленої енергетичної потужності. США також продемонстрували значне зростання встановленої сонячної енергії на 27,5%, додавши 38 265 МВт у 2024 році. Окрім Китаю та США, декілька країн стають швидкозростаючими ринками сонячної енергетики, що значною мірою пояснюється зниженням вартості сонячної електроенергії більш ніж на 50% протягом останнього десятиліття (рис. 1.2).

Вимірюється в кіловат-годинах на людину.

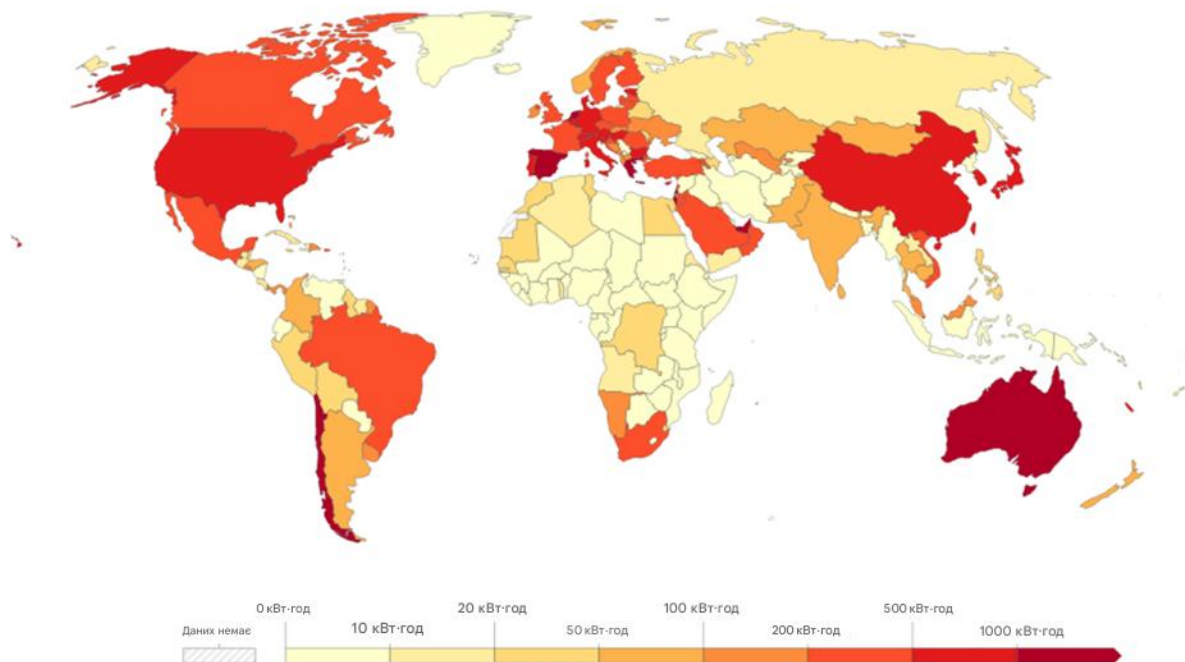


Рисунок 1.2 – Світова карта розподілу сонячних потужностей за країнами

Для ефективної експлуатації сонячних електростанцій необхідно підтримувати їх роботу на запроектованому рівні потужності. Втім, дослідження показують, що фактична продуктивність часто виявляється нижчою за очікувану через різноманітні несприятливі умови та технічні обмеження.

Накопичення пилу на поверхні фотоелектричної панелі може призводити до значної деградації ефективності фотоелектричних систем з втратами від 10% у помірних умовах до понад 40% у посушливих регіонах. Енергетичні втрати на

день через забруднення можуть сягати від 2,8% до 50% залежно від щільності забруднення, що було встановлено за допомогою методології математичних кореляцій у погодних умовах Близького Сходу, Азії та Австралії [4].

Окрім забруднення, на продуктивність впливають технічні дефекти, які можуть виникати як на етапі виробництва та монтажу, так і в процесі експлуатації. Несвоєчасне виявлення таких проблем призводить до кумулятивних втрат енергії, які за весь період експлуатації можуть становити значні суми.

## 1.2 Типи дефектів сонячних панелей

Сонячні панелі під час експлуатації піддаються впливу різноманітних факторів, що можуть призвести до виникнення дефектів різної природи. Поширені дефекти, такі як знебарвлення, делемінація та вихід з ладу діодів, часто стають більш ймовірними з віком системи [5]. Дефекти класифікуються за природою виникнення (виробничі та експлуатаційні), за видимістю (візуальні та приховані) та за впливом на роботу системи (рис. 1.3).

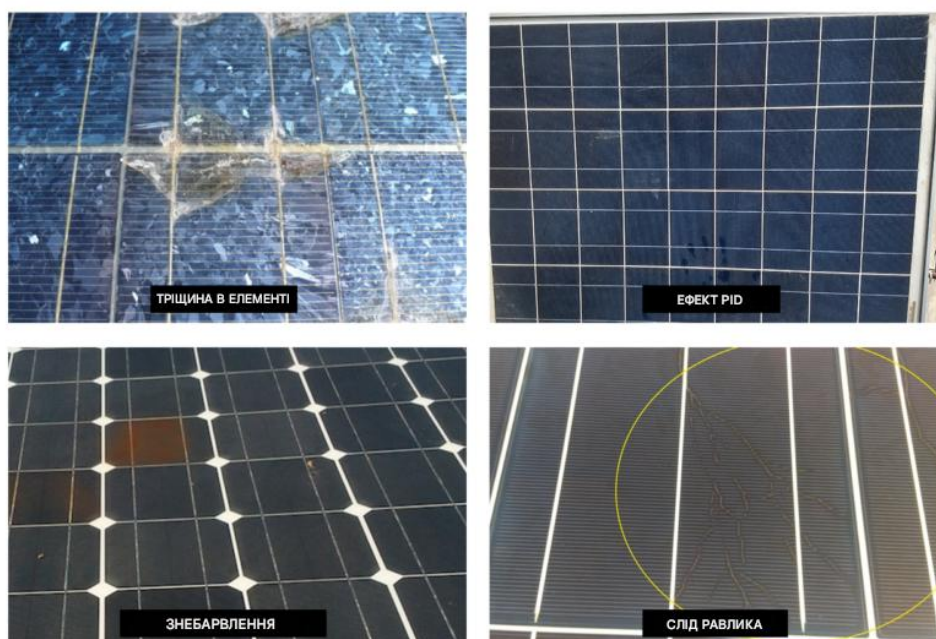


Рисунок 1.3 – Основні типи дефектів сонячних панелей

Забруднення поверхні є найпоширенішим фактором зниження продуктивності. До основних типів належать пилове забруднення, органічні забруднення (пташиний послід, листя), сезонні забруднення (сніг) та промислові відкладення.

Мікротріщини – це дрібні тріщини у кристалічній структурі фотоелектричних елементів, які виникають під час виробництва, транспортування або від термічних напружень. Вони можуть призводити до втрат електроенергії і навіть спричиняти пожежі. З часом мікротріщини прогресують, електрично ізолюючи частини елемента.

Гарячі точки (Hot Spots) – локальні ділянки підвищеної температури, що виникають через тріщини, погано припаяні з'єднання або затінення. Температура може перевищувати 100-150°C, що створює пожежну небезпеку та прискорює деградацію.

Делемінація – відшарування захисних шарів панелі, що дозволяє вологі проникати всередину модуля, збільшуючи ризик корозії та зниження потужності [6]. Візуально проявляється як білі плями або бульбашки.

Потенційна індукована деградація (PID) виникає через високу різницю напруг між заземленим склом та елементами, знижуючи продуктивність на 30-50% [6].

Механічні пошкодження включають тріщини скла, деформацію рамки та пошкодження з'єднань. Різні типи дефектів мають різні візуальні характеристики, що робить можливим їх розпізнавання за допомогою методів комп'ютерного зору.

### 1.3 Огляд існуючих методів діагностики

Для забезпечення стабільної роботи сонячних електростанцій застосовуються різноманітні методи діагностики, кожен з яких має свої переваги

та обмеження. Розглянемо основні підходи, що використовуються в сучасній практиці технічного обслуговування.

Найпростішим методом є візуальний огляд панелей неозброєним оком або за допомогою біноклю. Цей підхід дозволяє виявити очевидні механічні пошкодження, сильне забруднення, делемінацію та видимі тріщини. Однак метод є суб'єктивним, трудомістким для великих установок і не дозволяє виявити приховані дефекти, такі як мікротріщини або початкові стадії деградації. Для великих сонячних ферм з десятками тисяч панелей візуальний огляд стає економічно недоцільним.

Вимірювання вольт-амперних характеристик є наступним методом, який базується на аналізі кривих струм-напруга (I-V характеристики) окремих панелей або стрінгів. Вимірювання проводяться спеціалізованими приладами та дозволяють кількісно оцінити продуктивність панелі, виявити зниження потужності, проблеми з обхідними діодами та деградацію фотоелементів. Недоліком є необхідність доступу до кожної панелі, тривалість процесу та відсутність візуальної інформації про локалізацію дефектів.

Інфрачервона термографія є ефективним методом діагностики, що базується на виявленні температурних аномалій [7]. Термографічна інспекція фотоелектричних модулів за допомогою тепловізора є важливим для виявлення проблем, оскільки невиявлені дефекти можуть призводити до втрат енергії та навіть пожеж [8]. Метод дозволяє виявляти гарячі точки, несправні діоди, мікротріщини та проблеми з електричними з'єднаннями через характерні температурні патерни (рис. 1.4).

Інфрачервона термографія розглядається як методика з можливістю виявлення в реальному часі, що використовується як під час польових випробувань, так і у виробничих процесах [9]. Проте термографія має обмеження, а саме необхідні певні погодні умови (достатня сонячна інсоляція, відсутність вітру), спеціалізоване обладнання та кваліфікований персонал. Ручна термографія з використанням портативних тепловізорів є трудомісткою для великих об'єктів.

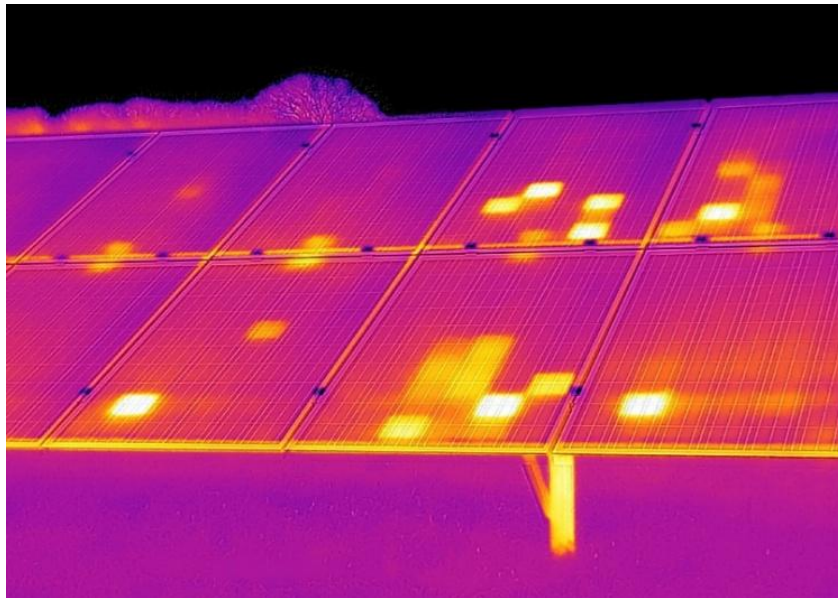


Рисунок 1.4 – Термографічне зображення дефектів панелей

Електролюмінесценція (EL) – це високоточний метод, при якому через панель пропускається зворотний струм у темряві, що змушує фотоелементи випромінювати інфрачервоне світло. Спеціальна камера фіксує це випромінювання, виявляючи навіть найдрібніші мікротріщини, неоднорідності та дефекти на рівні окремих елементів [10]. Метод забезпечує найвищу точність діагностики, але потребує відключення панелі від системи, проведення вимірювань у темряві та спеціального обладнання, що робить його непрактичним для польових умов на великих електростанціях.

Застосування дронів (БпЛА) обладнаних термографічними та RGB-камерами стають революційним кроком у діагностиці сонячних електростанцій. Дрони для інспекції сонячних панелей забезпечують точні результати за короткий час при зменшених витратах порівняно з ручними методами [11].

Дрони обладнані датчиками високої роздільності та тепловізорами дозволяють швидко охоплювати великі площі, виявляти гарячі точки, забруднення та механічні пошкодження [12]. Повітряні інспекції дозволяють документувати деградацію панелей та охоплювати весь проєкт, на відміну від традиційних методів, що обмежені вибірковою оглядом [13].

## 1.4 Застосування комп'ютерного зору та ML

Методи комп'ютерного зору та машинного навчання стають все більш поширеними у діагностиці сонячних електростанцій. Комп'ютерний зір дозволяє комп'ютерним системам автоматично аналізувати візуальну інформацію з цифрових зображень, що суттєво підвищує швидкість та об'єктивність інспекцій порівняно з ручним оглядом [14].

На початкових етапах використовувалися традиційні методи цифрової обробки зображень – фільтрація, сегментація, виділення контурів та аналіз текстур [8,15]. Ці підходи базуються на заздалегідь визначених правилах, які виявляють дефекти через характерні зміни яскравості або текстури (рис. 1.5) [15]. Однак класичні методи потребують ретельного налаштування параметрів для кожного типу дефектів, чутливі до змін умов освітлення та не завжди здатні розрізняти схожі за зовнішнім виглядом дефекти [16].

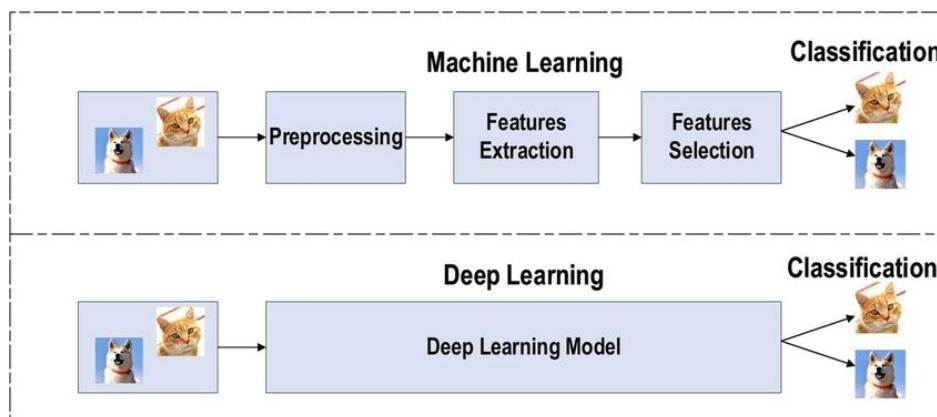


Рисунок 1.5 – Порівняння традиційної обробки зображень та CNN

З розвитком глибокого навчання ситуація змінилася. Згорткові нейронні мережі здатні автоматично виявляти складні візуальні ознаки безпосередньо з даних, без ручного проектування правил [16]. Моделі навчаються на великих наборах зображень панелей з різними дефектами і «вчаться» розрізняти характерні патерни (рис. 1.6). Дослідження показують високу ефективність CNN-архітектур у виявленні дефектів фотоелектричних елементів [17, 18].

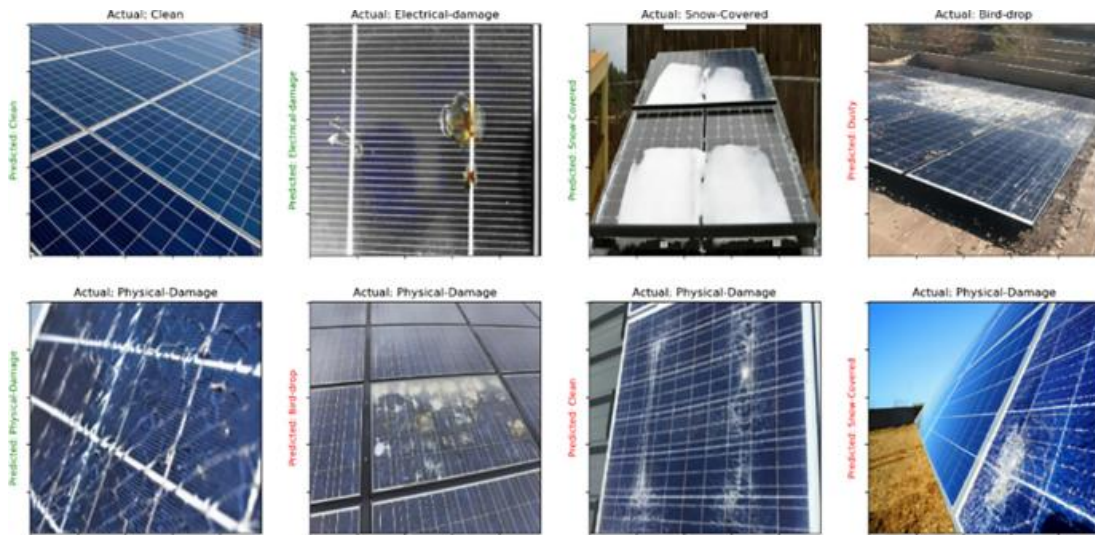


Рисунок 1.6 – Приклад роботи CNN для виявлення дефектів

Системи працюють за принципом багатоетапного аналізу: спочатку виявлення окремих панелей, потім аналіз кожної на наявність дефектів з класифікацією за типами, такі як мікротріщини, гарячі точки, забруднення, делемінація [6]. Це дозволяє точно локалізувати проблему та оцінити її серйозність.

### 1.5 Огляд архітектур нейронних мереж

Згорткові нейронні мережі (CNN) є спеціалізованим типом алгоритмів глибокого навчання, призначених для задач розпізнавання об'єктів, включаючи класифікацію, детекцію та сегментацію зображень [19]. Згорткові нейронні мережі складаються з декількох типів шарів, наприклад, вхідний шар, згорткових шарів, шарів «pooling» та повнозв'язаних шарів [19]. Згорткові шари застосовують фільтри для виявлення локальних патернів у зображенні, «pooling» шари зменшують розмірність, зберігаючи важливі ознаки, а повнозв'язані шари виконують фінальну класифікацію (рис. 1.7) [20].

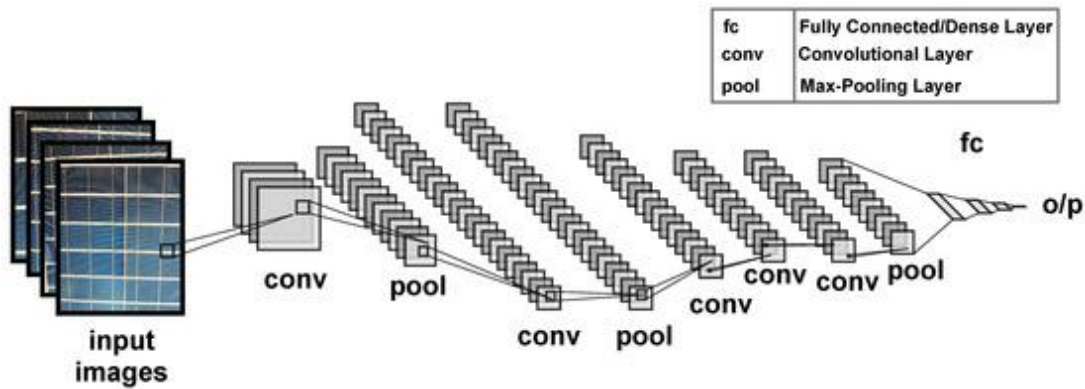


Рисунок 1.7 – Архітектура згорткової нейронної мережі на прикладі сонячних панелей

Еволюція архітектур згорткових нейронних мереж відбувалася шляхом поступового вирішення обмежень попередніх підходів. Кожна нова архітектура вносила інноваційні рішення, що підвищували точність або ефективність моделей. Розглянемо ключові архітектури, що визначили розвиток згорткових нейронних мереж.

VGG (Visual Geometry Group) продемонструвала, що глибокі мережі з малими згортковими фільтрами  $3 \times 3$  можуть досягати високої точності на задачах класифікації зображень. Архітектура використовується для великомасштабних задач класифікації зображень з великою кількістю навчальних даних та transfer learning для екстракції ознак [21].

ResNet (Residual Networks) вирішила проблему деградації градієнтів у дуже глибоких мережах через введення залишкових з'єднань (skip connections). Використовується для задач, що потребують глибоких мереж, та для різноманітних і складних наборів зображень [21]. Залишкові з'єднання дозволяють навчати мережі з сотнями шарів без втрати якості.

MobileNet – сімейство легких архітектур, оптимізованих для мобільних пристроїв та embedded систем. Архітектура базується на depthwise separable convolutions, що дозволяє зменшити обчислювальні витрати при збереженні точності [22]. Архітектура MobileNet призначений для класифікації зображень, детекції об'єктів та застосування в ситуаціях з обмеженими обчислювальними ресурсами.

Пренавчена (Pre-trained) модель – це збережена мережа, яка була раніше навчена на великому наборі даних, зазвичай на великомасштабних задачах класифікації зображень [23]. Трансферне навчання дозволяє використовувати знання, набуті моделлю на одній задачі (наприклад, на наборі зображень ImageNet), для вирішення іншої, споріднені задачі. Це ефективно при обмежених обсягах навчальних даних, оскільки низькорівневі візуальні ознаки (кути, текстури, форми) є універсальними для різних доменів.

CLIP (Contrastive Language-Image Pre-Training) – мультимодальний zero-shot класифікатор зображень, що досягає вражаючих результатів у широкому діапазоні доменів без fine-tuning [24]. Модель навчена на 400 мільйонах пар (зображення, текст), що дозволяє їй розуміти зв'язок між візуальним та текстовим контентом [24].

Вибір архітектури залежить від конкретної задачі, доступних обчислювальних ресурсів та обсягу навчальних даних. Для задач з фіксованими класами та достатньою кількістю даних ефективним є контрольоване навчання (supervised learning) підхід трансферним навчанням. Для задач, де класи можуть змінюватися або дані обмежені, доцільним є використання навчання без прикладів (zero-shot learning) підходів, таких як CLIP модель.

## 1.6 Постановка задачі дослідження

Проведений аналіз показав, що існуючі методи діагностики сонячних панелей мають суттєві обмеження при застосуванні на великих електростанціях, а саме високу трудомісткість, залежність від кваліфікації персоналу, суб'єктивність оцінки та значні часові витрати. Хоча дослідження демонструють ефективність згорткових нейронних мереж для виявлення дефектів з точністю до 91,1% для бінарної класифікації [25], більшість робіт фокусуються на методах без порівняння з альтернативними підходами. Крім того, традиційні контрольовані методи вимагають великих обсягів розмічених даних, тоді як

підходи навчання без прикладів на основі моделей «vision-language» залишаються недостатньо дослідженими для цієї задачі [24].

Основна проблема полягає у необхідності швидкої, точної та економічно ефективної системи автоматизованого виявлення дефектів, яка могла б адаптуватися до різних умов експлуатації та обмежень у наявності навчальних даних.

Об'єктом дослідження є зображення сонячних панелей з різними дефектами та забрудненнями.

Метою дослідження є порівняння методів автоматизованого виявлення дефектів сонячних панелей на основі аналізу зображень із застосуванням згорткових нейронних мереж та «vision-language» моделей для підвищення точності діагностики та забезпечення енергоефективності сонячних електростанцій.

Для досягнення мети необхідно вирішити наступні завдання:

- підготувати набір даних для бінарної класифікації на основі набору зображень сонячних панелей з розділенням на навчальну та тестову вибірки;
- реалізувати метод виявлення дефектів за допомогою моделей «MobileNetV2, VGG1» та CLIP для класифікації панелей;
- провести експериментальне дослідження обох методів з оцінкою метрик якості та часу inference на тестовому наборі даних;
- виконати порівняльний аналіз методів за критеріями точності, швидкості роботи, потреби у навчальних даних, розміру моделі та гнучкості застосування;
- сформулювати практичні рекомендації щодо вибору методу для різних сценаріїв експлуатації сонячних електростанцій.

Очікуваним результатом є функціонуючі прототипи систем виявлення дефектів, кількісна оцінка їх ефективності та рекомендації щодо практичного застосування залежно від наявності навчальних даних, обчислювальних ресурсів та вимог до точності діагностики.

## 2 ТЕОРЕТИЧНІ ОСНОВИ МЕТОДІВ КЛАСИФІКАЦІЇ ДЕФЕКТІВ СОНЯЧНИХ ПАНЕЛЕЙ

### 2.1 Математична формалізація задачі бінарної класифікації

У контексті виявлення дефектів сонячних панелей, задача класифікації формулюється як бінарна проблема розпізнавання образів, де необхідно визначити стан панелі, такі як «чиста» (функціонує нормально) або «забруднена» (має дефекти, забруднення чи пошкодження).

Нехай  $X$  – простір вхідних даних, що представляє собою множину цифрових зображень сонячних панелей розміром.

$$H \times W \times C, \tag{2.1}$$

де  $H$  – висота;

$W$  – ширина;

$C$  – кількість каналів (зазвичай  $C = 3$  для RGB-зображень).

Простір міток класів  $Y$  для бінарної класифікації визначається як  $Y = \{0, 1\}$ , де  $0$  відповідає класу «чиста панель», а  $1$  – класу «забруднена/дефектна панель».

Метою є навчання функції класифікації  $f: X \rightarrow Y$ , яка відображає вхідне зображення на мітку класу. В рамках ймовірнісного підходу модель передбачає ймовірність належності зображення до класу  $1$ , де  $\theta$  – параметри моделі (ваги нейронної мережі) [25].

Структурні методи класифікації зображень базуються на екстракції інформативних ознак з вхідних даних [27]. У випадку глибинних нейронних мереж, процес екстракції ознак та класифікації об'єднано в єдину сквозну архітектуру. Для навчання моделі використовується функція втрат (loss function), яка вимірює розбіжність між передбаченнями моделі та справжніми мітками.

Для задачі бінарної класифікації стандартною функцією втрат є функція втрат бінарної кросентропії (логарифмічні втрати) [25].

$$\ell_n = -[y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n)], \quad (2.2)$$

де  $y_n \in \{0, 1\}$  – справжня мітка класу для  $n$ -го зразка;

$x_n \in (0, 1)$  – передбачена ймовірність належності до класу 1 для  $n$ -го зразка.

Для всього навчального набору даних з  $N$  зразків загальна функція втрат визначається як середнє значення індивідуальних втрат.

$$L(x, y) = \left(\frac{1}{N}\right) \cdot \sum \ell_n, \quad (2.3)$$

де  $N$  – кількість зразків у навчальному наборі;

$\ell_n$  – індивідуальна функція втрат для  $n$ -го зразка.

Метод «cross-entropy» є оптимальною функцією втрат для класифікаційних задач, оскільки вона штрафує впевнені неправильні передбачення значно сильніше, ніж невпевнені, що стимулює модель до точнішої калібрації ймовірностей [25].

Навчання моделі зводиться до розв'язання оптимізаційної задачі мінімізації функції втрат з додаванням регуляризаційного члена, який запобігає перенавантаженню. Найпоширенішим типом регуляризації є  $L2$ -регуляризація (weight decay), що штрафує моделі з великими значеннями ваг. Це сприяє навчанню більш простих та узагальнюваних рішень [26, 28].

Для комплексної оцінки ефективності бінарного класифікатора використовується набір метрик, що базуються на матриці плутанини. Матриця плутанини для бінарної класифікації має розмірність  $2 \times 2$  та містить чотири основні компоненти:

- $TP$  (істинно позитивні) – кількість правильно класифікованих позитивних зразків (забруднені панелі розпізнані як забруднені);
- $TN$  (істинно негативні) – кількість правильно класифікованих негативних зразків (чисті панелі розпізнані як чисті);

–  $FP$  (хибно позитивні) – кількість помилково класифікованих як позитивні (чисті панелі розпізнані як забруднені) – помилка I роду;

–  $FN$  (хибно негативні) – кількість помилково класифікованих як негативні (забруднені панелі розпізнані як чисті) – помилка II роду.

На основі цих компонентів визначаються наступні метрики [29]:

– точність (accuracy) – частка правильно класифікованих зразків серед усіх зразків [24].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} . \quad (2.4)$$

Точність є інтуїтивно зрозумілою метрикою, однак може бути оманливою при незбалансованих наборах даних, де один клас значно переважає інший;

– точність прогнозу (precision) – частка справді позитивних зразків серед усіх, що були класифіковані як позитивні [24].

$$Precision = \frac{TP}{(TP + FP)} . \quad (2.5)$$

Точність прогнозу показує, наскільки можна довіряти позитивним передбаченням моделі. Висока precision означає низьку кількість хибних спрацювань;

– повнота (recall/sensitivity) – частка правильно знайдених позитивних зразків серед усіх реально позитивних [24].

$$Recall = \frac{TP}{(TP + FN)} . \quad (2.6)$$

Повнота характеризує здатність моделі виявляти всі позитивні випадки. Висока повнота означає, що модель пропускає мало дефектних панелей;

–  $F1$ -міра ( $F1$ -Score) – гармонійне середнє точності [24] прогнозу та повноти.

$$F1 = 2 \cdot \frac{(Precision \cdot Recall)}{(Precision + Recall)}. \quad (2.7)$$

*F1*-міра є збалансованою метрикою, корисною для незбалансованих наборів даних, оскільки враховує як хибні спрацювання (*FP*), так і пропущені випадки (*FN*) [27].

У контексті діагностики сонячних панелей, різні метрики мають різну важливість залежно від бізнес-вимог. Високий *recall* важливий для мінімізації пропущених дефектів, що можуть призвести до втрати ефективності електростанції. Водночас, висока *precision* знижує кількість непотрібних перевірок чистих панелей, зменшуючи операційні витрати.

Методи статистичного аналізу для визначення релевантності структурних описів зображень, описані в [28], можуть бути застосовані для оцінки інформативності різних ознак у процесі класифікації.

Оскільки модель видає ймовірність  $x_n \in [0, 1]$ , для отримання дискретної мітки класу необхідно застосувати порогову функцію з пороговим значенням  $\tau$  (threshold). Стандартно використовується  $\tau = 0,5$ , однак оптимальне значення порогоу може бути підібране на валідаційному наборі для максимізації цільової метрики (наприклад, *F1*- міра або балансу між *precision* та *recall* залежно від вимог конкретного застосування).

Варіювання порогоу дозволяє контролювати компроміс між точністю прогнозу та повнотою, тобто зменшення  $\tau$  збільшує повноту, але знижує точність прогнозу, з іншого боку збільшення  $\tau$  має протилежний ефект. [29].

## 2.2 Сучасні нейронні архітектури для аналізу зображень

Для вирішення задачі виявлення дефектів сонячних панелей застосовуються різні підходи до класифікації зображень, які відрізняються архітектурою, принципами навчання та вимогами до даних [26, 27].

У цьому розділі розглядаються математичні основи згорткових операцій, архітектури обраних моделей та їхні особливості для задачі класифікації дефектів.

### 2.2.1 Математичні основи згорткових операцій

Згорткові нейронні мережі (Convolutional Neural Networks) є основним інструментом для аналізу зображень завдяки здатності автоматично виявляти просторові ієрархічні патерни в даних. Фундаментальною операцією згорткової нейронної мережі є дискретна згортка (convolution), яка дозволяє екстрагувати локальні ознаки з вхідних зображень [25, 27].

Згортка – це математична операція, що застосовує фільтр (kernel) до вхідного зображення для виявлення певних ознак. Для двовимірного зображення  $I$  та ядра згортки  $K$ , операція згортки визначається як [25]:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) \cdot K(i - m, j - n), \quad (2.8)$$

де  $I(m, n)$  – значення пікселя вхідного зображення в позиції  $(m, n)$ ;

$K$  – значення ядра згортки;

$S(i, j)$  – вихідна карта ознак.

У практичних реалізаціях часто використовується перехресна кореляція, яка відрізняється від згортки відсутністю віддзеркалення ядра, але в контексті глибокого навчання ці терміни часто використовуються як синоніми, оскільки ядра навчаються автоматично. [25].

Розмір вихідної карти ознак залежить від кількох параметрів, а саме «kernel size» ( $K$ ) – розмір фільтра (зазвичай  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ), stride ( $S$ ) – крок зсуву фільтра по зображенню, та «padding» ( $P$ ) – додавання нульових пікселів по краях зображення.

Доповнення (padding) використовується для збереження просторової розмірності (при  $P = (K-1) / 2$  та  $S = 1$ ) або для запобігання втраті інформації на краях зображення [25].

Розмір вихідного зображення обчислюється за формулою [26]:

$$O = \left\lfloor \frac{(W - K + 2P)}{S} \right\rfloor + 1, \quad (2.9)$$

де  $W$  – ширина (або висота) вхідного зображення;

$O$  – розмір вихідної feature map.

Після згорткових шарів зазвичай застосовують операції підвибірки (pooling) для зменшення просторової розмірності та досягнення інваріантності до невеликих зсувів. [25]. Найпопулярнішою є операція максимального підвибору (max pooling), яка вибирає максимальне значення в кожній області (зазвичай  $2 \times 2$  або  $3 \times 3$ ). Альтернативою є операція середнього підвибору, що обчислює середнє значення в області (рис. 2.1).

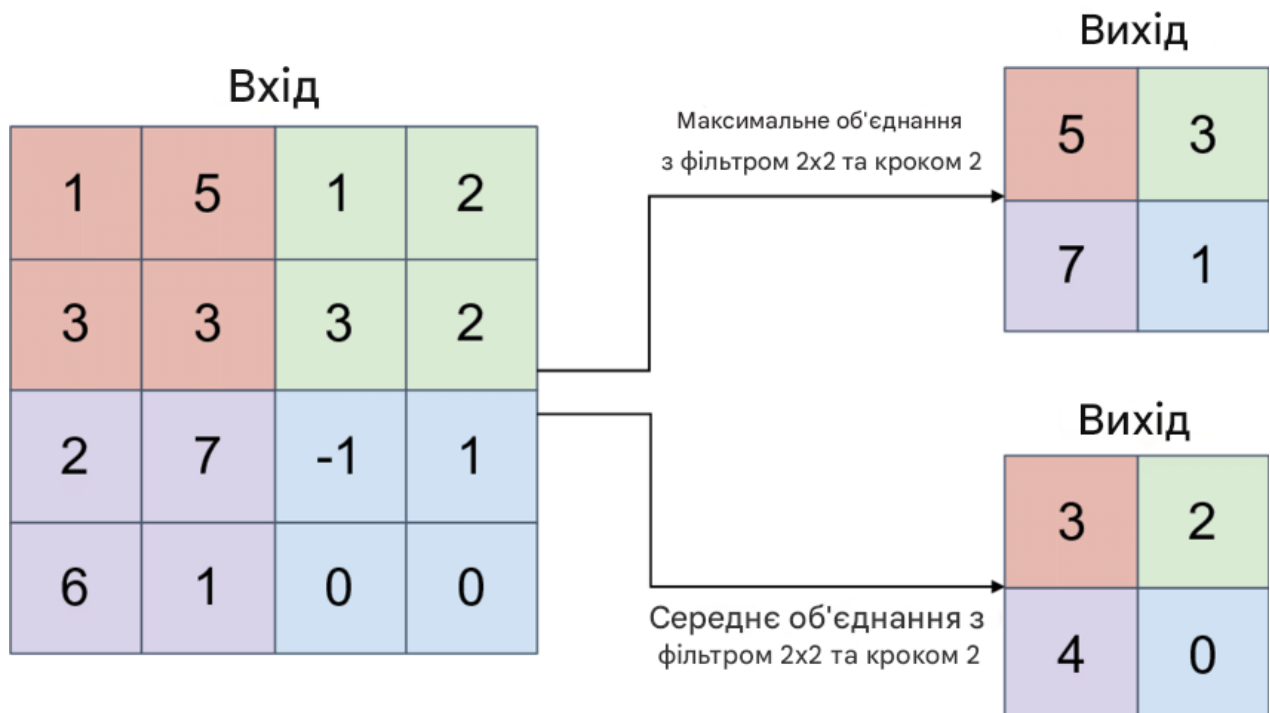


Рисунок 2.1 – Приклад «max pooling» операції  $2 \times 2$

Операція «max pooling» зберігає найбільш виражені ознаки та забезпечує інваріантність до невеликих трансформацій зображення. Операція «pooling» зменшує кількість параметрів та обчислювальну складність мережі, одночасно розширюючи рецептивне поле наступних шарів [25].

Для введення нелінійності в модель після згорткових операцій застосовуються функції активації. Найпоширенішими є ReLU (Rectified Linear Unit) [25].

$$f(x) = \max(0, x). \quad (2.10)$$

Функція активації ReLU є стандартною в CNN завдяки простоті обчислення та ефективності навчання. Вона вирішує проблему зникаючого градієнта (vanishing gradient) для додатних значень.

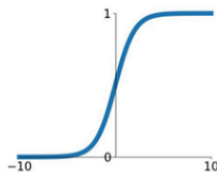
Функція активації «Sigmoid» представлена нижче.

$$\sigma(x) = \frac{1}{(1 + e^{-x})}. \quad (2.11)$$

Функція активації Sigmoid використовується переважно у вихідних шарах для бінарної класифікації, оскільки відображає значення в діапазоні  $[0, 1]$ , що інтерпретується як ймовірність (рис. 2.2).

**Sigmoid**

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



**ReLU**

$$\max(0, x)$$

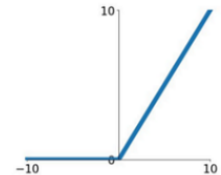


Рисунок 2.2 – Графіки функцій активації «ReLU» та «Sigmoid»

Простота похідної ReLU (дорівнює 1 для  $x > 0$  та 0 для  $x \leq 0$ ) робить backpropagation швидшим та ефективнішим порівняно з sigmoid або tanh [25].

Згорткові нейронні мережі будують ієрархію ознак, наприклад, нижні шари виявляють прості патерни (краї, кути, текстури), а вищі шари комбінують їх у складніші структури (форми, об'єкти, контексти) (рис. 2.3).

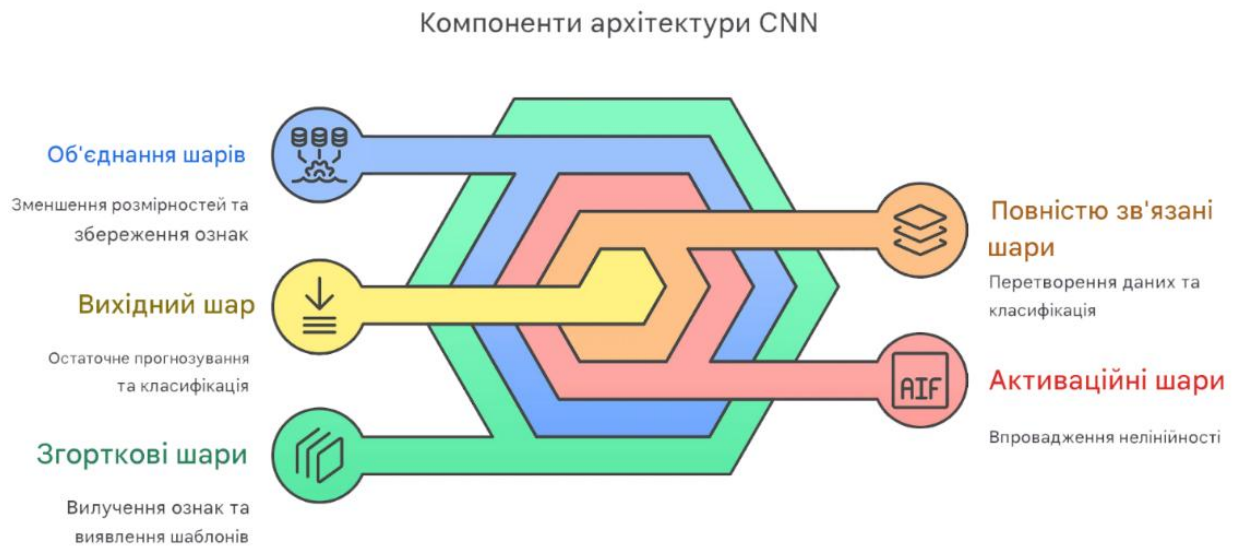


Рисунок 2.3 – Ієрархічне виявлення ознак у CNN

Для задачі виявлення дефектів сонячних панелей, нижні шари можуть детектувати краї панелей та зміни текстури, тоді як вищі шари розпізнають забруднення, тріщини або інші дефекти [14, 27]. Кожен фільтр у згортковому шарі навчається виявляти специфічний тип ознаки, а сукупність фільтрів створює багатовимірне представлення зображення, яке використовується для класифікації.

### 2.2.2 Архітектура MobileNetV2

Архітектура MobileNetV2 базується на концепції роздільних згорток по глибині (depthwise separable convolutions) та інвертованих залишкових блоків (inverted residual blocks), що дозволяє досягти високої точності класифікації при значно меншій кількості параметрів та обчислювальних операцій порівняно з традиційними згортковими нейронними мережами [20, 21].

Ключовою інновацією MobileNetV2 є заміна стандартних згорткових операцій на роздільні згортки по глибині (depthwise separable convolutions), які розкладають згортку на два окремі кроки [30]:

- глибинна згортка (depthwise convolution) – кожен вхідний канал обробляється окремим фільтром (3×3), без змішування інформації між каналами;
- очкова згортка (pointwise convolution) – застосування 1×1 згортки для комбінування інформації з різних каналів.

Така декомпозиція суттєво зменшує обчислювальну складність. Для стандартної згортки кількість операцій становить:

$$K^2 \cdot M \cdot N \cdot F^2. \quad (2.12)$$

Для глибинної згортки кількість операцій зменшується до [30]:

$$K^2 \cdot M \cdot F^2 + M \cdot N \cdot F^2, \quad (2.13)$$

де  $K$  – розмір ядра (kernel size, наприклад 3 для фільтра 3×3),

$M$  – кількість вхідних каналів;

$N$  – кількість вихідних каналів;

$F$  – розмір «feature map» (ширина/висота).

Редукція обчислень для типових значень ( $N=256, K=3$ ). Це дає зменшення у 8–9 разів [30].

$$Reduction = \frac{1}{N} + \frac{1}{K^2}. \quad (2.14)$$

Також модель MobileNetV2 використовує inverted residual структуру, яка відрізняється від класичних залишкових блоків (ResNet) [21, 22]. Блок складається з таких послідовних операцій:

- шар розширення (expansion layer,  $1 \times 1$  convolution) – збільшення кількості каналів у  $t$  разів (expansion factor, зазвичай  $t = 6$ );
- глибинний шар (depthwise layer,  $3 \times 3$  depthwise convolution) – обробка просторової інформації;
- шар проєкції (projection layer,  $1 \times 1$  convolution) – зменшення розмірності назад до вихідної.

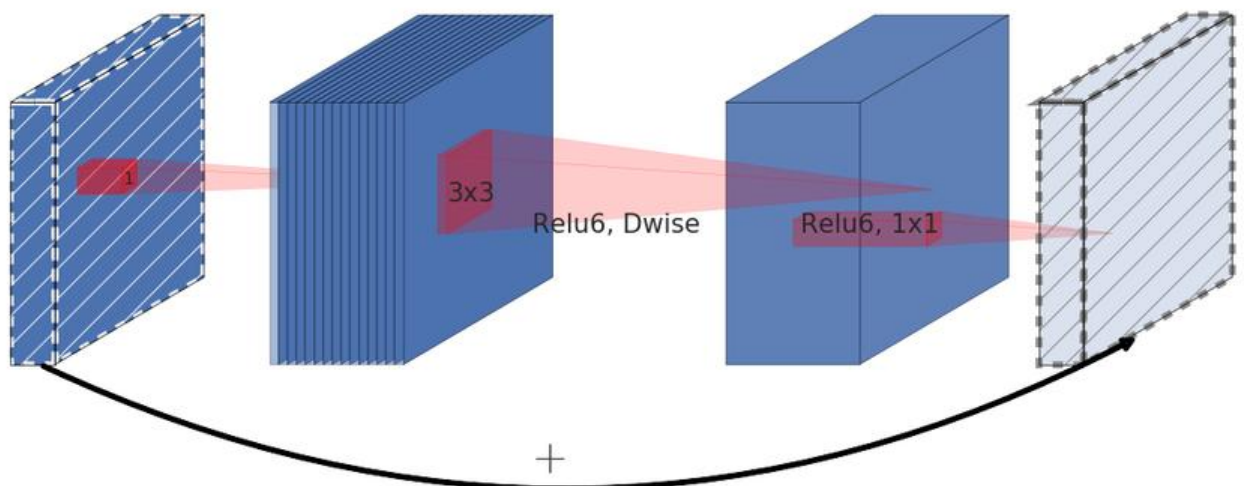


Рисунок 2.4 – Структура inverted residual block

Особливістю є використання «вузьких місць» (linear bottlenecks) – відсутність ReLU після шару проєкції. Це зроблено для збереження інформації у низькорозмірному просторі, оскільки ReLU може знищити важливі ознаки при малій кількості каналів (рис. 2.4) [30].

Загальна архітектура мережі «MobileNetV2» (рис. 2.5) складається з початкового згорткового шару ( $3 \times 3$ , 32 filters), послідовності з 17 «inverted residual blocks» з різними параметрами (expansion factor, stride, filters), та фінального згорткового шару ( $1 \times 1$ , 1280 filters). Загальна кількість параметрів становить приблизно 3,4 мільйони, що у 42 рази менше за VGG19 (143М параметрів) [20, 30].

Архітектура оптимізована для вхідних зображень розміром  $224 \times 224$  пікселів, що є стандартом для більшості задач класифікації зображень.

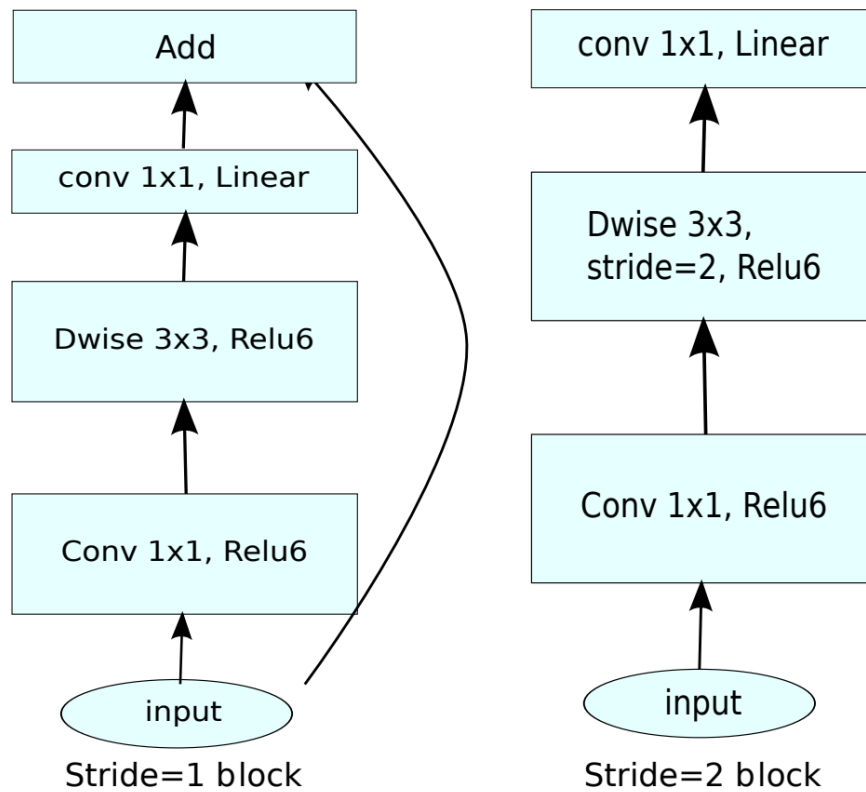


Рисунок 2.5 – Повна архітектура MobileNetV2

У контексті виявлення дефектів сонячних панелей застосовується підхід перенесеного навчання (transfer learning) [22, 27]. Архітектура MobileNetV2, попередньо навчена на наборі даних ImageNet (1,4 мільйона зображень, 1000 класів), вже навчилася виявляти універсальні візуальні ознаки, зокрема краї, текстури, форми, об'єкти.

Процес адаптації моделі включає:

- завантаження предтренованих ваг з набору даних «imagenet»;
- заморожування базових шарів (feature extractor) – параметри не оновлюються під час навчання;
- додавання власної верхівки класифікації (classification head) – нові повнозв'язні шари для бінарної класифікації (чиста/забруднена панель);
- тонке налаштування (fine-tuning) – навчання тільки нових шарів на даних сонячних панелей.

Такий підхід дозволяє досягти високої точності навіть при обмежених обсягах навчальних даних, оскільки нижні шари вже навчилися виявляти базові

візуальні паттерни [22, 27]. Для задачі класифікації сонячних панелей, MobileNetV2 екстрагує ознаки забруднень, тріщин та інших дефектів, використовуючи компактне представлення зображення.

### 2.2.3 Архітектура VGG19

Архітектура VGG19 демонструє, що збільшення глибини мережі при використанні малих фільтрів ( $3 \times 3$ ) значно покращує точність класифікації [31].

Принципи побудови архітектури VGG19 базується на кількох ключових принципах [31]:

- малі фільтри ( $3 \times 3$ ) – всі згорткові шари використовують фільтри розміром  $3 \times 3$ , що є найменшим розміром для захоплення просторових паттернів (верх/низ, ліво/право, центр);
- глибокі стеки згортки – послідовність кількох згорткових шарів без проміжної підвибірки (pooling) збільшує рецептивне поле та дозволяє виявляти складніші ознаки;
- подвоєння каналів – кількість фільтрів збільшується після кожного шару підвибірки (pooling) ( $64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$ );
- операція максимального підвибору (max pooling) ( $2 \times 2$ ) – зменшення просторової розмірності після кожного блоку згортки.

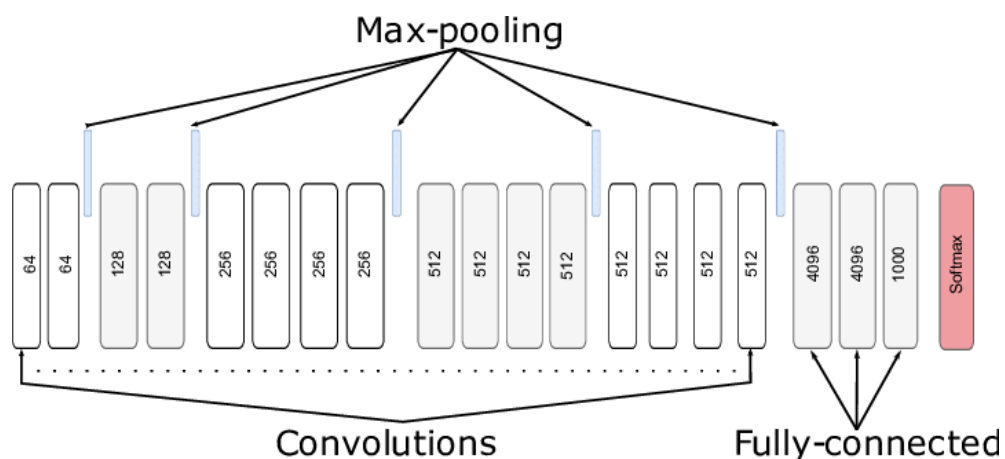


Рисунок 2.6 – Архітектура VGG19

Структура мережі складається з п'яти блоків згорткових шарів, кожен з яких завершується операцією максимального підвибору (max pooling) [31]. Після згорткових блоків йдуть три повнозв'язні шари, серед яких два по 4096 нейронів та фінальний softmax шар для класифікації (1000 класів для набору даних ImageNet) (рис. 2.6).

Загальна кількість параметрів становить приблизно 143 мільйони, що робить VGG19 однією з найбільших за розміром моделей [20, 31]. Більшість параметрів зосереджена у повнозв'язних шарах.

У контексті класифікації дефектів сонячних панелей використовується VGG19 з перенесеним навчанням (transfer learning) [23]. Модель, попередньо навчена на ImageNet, адаптується для бінарної класифікації шляхом:

- завантаження предтренуваних ваг згорткових шарів;
- заморожування витягувача ознак (feature extractor) – параметри не оновлюються;
- заміна верхівки класифікації (classification head) – замість 1000 класів imagenet, додається новий шар для 2 класів (чиста/забруднена);
- додавання регуляризації «dropout» і нормалізації за пакетом для запобігання перенавчанню.

Глибока архітектура VGG19 дозволяє екстрагувати багатий набір ознак з зображень сонячних панелей, що корисно для виявлення тонких дефектів та текстурних аномалій [22, 27]. Однак велика кількість параметрів вимагає більше обчислювальних ресурсів порівняно з компактними архітектурами типу MobileNetV2.

#### 2.2.4 Архітектура та принцип роботи CLIP

На відміну від традиційних supervised методів, що потребують великих обсягів розмічених даних для кожної конкретної задачі, CLIP навчається на парах «зображення-текст» з інтернету, що дозволяє виконувати класифікацію без

прикладів (zero-shot) класифікацію без додаткового навчання на цільовому наборі даних [32].

CLIP має подвійну енкодерну архітектуру, яка складається з двох паралельних енкодерів, які відображають зображення [32] та текст у спільний простір ознак:

- енкодер зображень – обробляє вхідні зображення та перетворює їх на векторні представлення;
- текстовий енкодер – обробляє текстові описи та генерує відповідні векторні представлення.

Ключова ідея полягає в тому, що семантично схожі зображення та тексти повинні мати близькі векторні представлення у спільному просторі, тоді як несхожі пари – віддалені [23, 32].

Енкодер зображень «Vision Transformer» у CLIP використовується для обробки зображень [32].

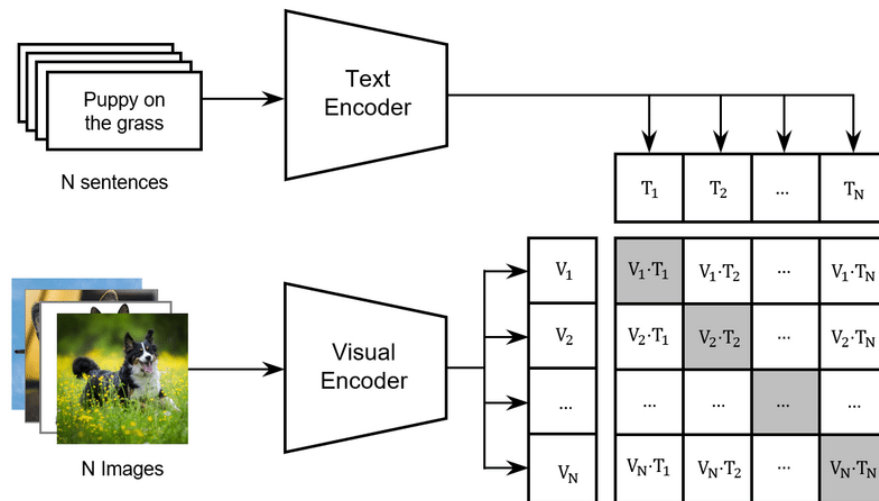


Рисунок 2.7 – Подвійна енкодерна архітектура CLIP

На відміну від традиційних CNN, ViT розбиває зображення на patches (ділянки) та обробляє їх як послідовність токенів, подібно до слів у реченні за таким алгоритмом:

- виокремлення патчів – зображення  $224 \times 224$  розбивається на патчі розміром  $16 \times 16$  (або  $32 \times 32$ ), отримуємо 196 або 49 патчів відповідно;

- лінійне проектування – кожен патч перетворюється на вектор фіксованої розмірності через лінійне перетворення;
- позиційні вбудовування – додається інформація про позицію кожного патча у зображенні, оскільки transformer не має вбудованої інформації про просторову структуру;
- енкодер трансформера – послідовність патчів обробляється через стек transformer blocks з multi-head self-attention механізмом;
- токен класу – спеціальний токен, доданий на початку послідовності, використовується для отримання фінального представлення зображення.

Модель за допомогою Vision Transformer здатна глобально аналізувати зв'язки між різними частинами зображення через механізм attention, на відміну від локальних рецептивних полів у CNN (рис. 2.7) [32].

Для обробки текстових описів CLIP використовує Transformer архітектуру, аналогічну до моделей GPT [32]:

- токенизація – розбиття тексту на токени (слова або підслова);
- вбудовування токенів – перетворення токенів на векторні представлення;
- позиційні вбудовування – додавання інформації про позицію токена в реченні;
- шари трансформера – обробка послідовності через «self-attention» та «feed-forward» шари;
- агрегація – поєднання представлень усіх токенів для отримання фінального вектора тексту.

Обидва енкодери проектують свої виходи у спільний простір ознак однакової розмірності (зазвичай 512 або 768) [32].

Ця гнучкість робить CLIP дуже корисним для задач з обмеженою кількістю навчальних даних або для швидкого прототипування [23, 27]. У контексті виявлення дефектів сонячних панелей, CLIP може класифікувати панелі, використовуючи тільки текстові описи станів.

### 2.3 Методи регуляризації та оптимізації навчання

Навчання глибоких нейронних мереж на обмежених наборах даних, таких як набори зображень дефектів сонячних панелей, часто призводить до проблеми *overfitting* – моделі надмірно адаптуються до навчальних даних та погано узагальнюються на нових зразках [25]. Для забезпечення узагальнюючої здатності моделей та стабільності навчання використовується комплекс методів регуляризації та оптимізації, які є важливими при застосуванні перенесеного навчання (*transfer learning*) [22, 27].

*Dropout* – це техніка регуляризації, яка випадково «вимикає» частину нейронів під час навчання, запобігаючи *co-adaptation* (надмірній залежності між нейронами) [32]. Математично, *dropout* застосовує випадкову маску до активацій шару.

$$y = \frac{x \odot \text{mask}}{p}, \quad (2.15)$$

де  $x$  – вхідні активації,

$\text{mask}$  – бінарна маска (кожен елемент дорівнює 1 з ймовірністю  $p$  та 0 з ймовірністю  $1-p$ );

$\odot$  – поелементне множення;

$p$  – коефіцієнт збереження нейронів (зазвичай 0,5–0,7).

Під час *inference* (тестування) *dropout* не застосовується, але активації масштабуються для компенсації. Ефект *dropout* можна інтерпретувати як навчання ансамблю (*ensemble*) з експоненційно великої кількості підмереж, що значно покращує узагальнюючу здатність [32].

У контексті класифікації сонячних панелей, ефект *dropout* застосовується після повнозв'язних шарів (MobileNetV2:  $p=0,5$ , VGG19:  $p=0,5$ ) для зменшення перенавчання на обмеженому наборі даних [27].

$L_2$ -регуляризація (Weight Decay) додає до функції втрат штраф на великі значення ваг. Загальна функція втрат з регуляризацією має вигляд:

$$L_{total} = L_{data} + \lambda \cdot \Sigma ||W||^2, \quad (2.16)$$

де  $L_{data}$  – емпірична функція втрат на навчальних даних;

$\lambda$  – коефіцієнт регуляризації (зазвичай 0,001–0,01);

$||W||^2 = \Sigma_{ij} w_{ij}^2$  – квадрат Евклідової норми матриці ваг.

Регуляризаційний член штрафує великі ваги, що запобігає надмірній складності моделі та покращує її здатність до узагальнення. У процесі градієнтного спуску,  $L_2$ -регуляризація еквівалентна додаванню затухання ваг (weight decay) при кожному кроці оптимізації [25].

Для архітектур MobileNetV2 та VGG19 застосовується  $L_2$ -регуляризація з коефіцієнтом  $\lambda = 0,01$  у повнозв'язних шарах, що важливо при перенесеному навчанні для запобігання перенавчанню на цільовому наборі даних [27].

На початку навчання застосовується нормалізація пакету, яка нормалізує активації кожного шару по міні-батчу під час навчання, що прискорює процес навчання [34]. Для кожної активації  $x$  у батчі розміром  $m$ , обчислюються:

$$\mu_B = \left(\frac{1}{m}\right) \cdot \Sigma_i x_i. \quad (2.17)$$

$$\sigma_B^2 = \left(\frac{1}{m}\right) \cdot \Sigma_i (x_i - \mu_B)^2. \quad (2.18)$$

$$\hat{x}_i = \frac{(x_i - \mu_B)}{\sqrt{\sigma_B^2 + \varepsilon}}. \quad (2.19)$$

$$y_i = \gamma \cdot \hat{x}_i + \beta, \quad (2.20)$$

де  $\mu_B$  та  $\sigma_B^2$  – середнє та дисперсія по батчу;

$\varepsilon$  – мала константа для числової стабільності (зазвичай  $10^{-5}$ );

$\gamma$  та  $\beta$  – навчальні параметри (scale та shift), які дозволяють мережі відновити представницьку здатність шару.

Нормалізація пакету зменшує внутрішнє коваріантне зміщення – зміну розподілу активацій під час навчання, що дозволяє використовувати більші швидкості навчання та зменшує чутливість до ініціалізації ваг [34]. У архітектурі VGG19 «Batch Normalization» застосовується після повнозв'язних шарів для стабілізації навчання глибокої мережі.

Оптимізатор Adam (Adaptive Moment Estimation) – це метод стохастичної оптимізації, який обчислює адаптивні learning rates для кожного параметра на основі перших та других моментів градієнтів [26, 34].

Оптимізатор Adam комбінує переваги AdaGrad (адаптивні швидкості навчання) та RMSprop (експоненційне згладжування), що робить його ефективним для задач з великими та розрідженими градієнтами [35]. Для перенесеного навчання з MobileNetV2 використовується  $\alpha = 0,0005$ , а для VGG19 це  $\alpha = 0,0001$ , що відображає різну чутливість архітектур до величини кроку оптимізації [22, 27].

Щоб адаптивно зменшувати швидкість навчання та застосовувати раннє зупинення (Early Stopping) для подальшого покращення навчання використовуються callback-механізми:

- «ReduceLROnPlateau» – автоматично зменшує час навчання при відсутності покращення метрики на validation set протягом певної кількості епох (patience). Зазвичай  $lr_{new} = 0,5 \cdot lr_{old}$  після 3 епох без покращення, що дозволяє моделі «дотюнити» ваги з меншим кроком;

- «Early Stopping» – зупиняє навчання, якщо втрачання на валідації не покращується протягом заданої кількості епох (зазвичай 5 – 7), та відновлює ваги з найкращої епохи. Це запобігає перевантаженню та економить обчислювальні ресурси.

Комбінація цих методів забезпечує стабільне та ефективне навчання моделей на наборі даних дефектів сонячних панелей при обмежених обсягах розмічених даних [22, 27].

## 2.4 Теоретичні основи препроцесингу зображень

Препроцесинг зображень є важливим етапом підготовки даних для навчання згорткових нейронних мереж, оскільки він забезпечує коректне представлення вхідних даних та покращує швидкість збіжності моделі [14, 25]. Різні архітектури CNN мають специфічні вимоги до нормалізації вхідних зображень, зумовлені особливостями їх предтренування на наборах даних.

Нормалізація – це процес масштабування значень пікселів до певного діапазону, що забезпечує числову стабільність обчислень. Для різних архітектур застосовуються різні схеми нормалізації.

Для MobileNetV2 використовується проста нормалізація до діапазону [0, 1]:

$$x' = \frac{x}{255}, \quad (2.26)$$

де  $x$  – оригінальне значення пікселя (0–255 для 8-bit зображень);

$x'$  – нормалізоване значення.

Для VGG19 застосовується стандартизація з параметрами ImageNet [31]:

$$x'_{channel} = \frac{(x_{channel} - \mu_{channel})}{\sigma_{channel}}, \quad (2.27)$$

де  $\mu$  та  $\sigma$  – середнє значення та стандартне відхилення для кожного каналу RGB, обчислені на ImageNet.

Для CLIP використовується специфічна нормалізація [32]:

$$x'_{channel} = \frac{\left(\frac{x_{channel}}{255} - \mu_{CLIP}\right)}{\sigma_{CLIP}}, \quad (2.28)$$

де  $\mu_{CLIP} = [0,48, 0,45, 0,40]$ ;

$\sigma_{CLIP} = [0,268, 0,261, 0,275]$ .

Ці параметри відповідають статистикам набору даних, на якому була попередньо навчена модель, і є необхідним для коректної роботи transfer learning [32].

Всі архітектури вимагають фіксованого розміру вхідних зображень – зазвичай  $224 \times 224$  пікселів. Bilinear interpolation є стандартним методом для зміни розміру, який обчислює нові значення пікселів як зважену суму чотирьох найближчих пікселів оригінального зображення:

$$f(x, y) = \sum_i \sum_j w_{ij} \cdot I(x_i, y_j), \quad (2.29)$$

де  $w_{ij}$  – ваги, обчислені на основі відстані до сусідніх пікселів;

$I(x_i, y_j)$  – значення пікселів оригінального зображення [14].

Аугментація даних (Data augmentation) – це техніка штучного збільшення розміру навчального набору даних шляхом застосування випадкових трансформацій до існуючих зображень, що покращує узагальнюючу здатність [27].

Основні трансформації включають параметри, представлені нижче.

Горизонтальне відображення (Horizontal flip) – дзеркальне відображення зображення відносно вертикальної осі:

$$I'(x, y) = I(W - x, y), \quad (2.30)$$

де  $W$  – ширина зображення.

Ця трансформація зберігає семантику (забруднена панель залишається забрудненою) та подвоює кількість навчальних зразків [25].

Поворот (Rotation) – поворот зображення на випадковий кут  $\alpha$  з заданого діапазону (зазвичай  $\pm 10\text{--}15^\circ$ ). Афінне перетворення для повороту:

$$[x', y']^T = R(\alpha) \cdot [x, y]^T, \quad (2.31)$$

де  $R(\alpha)$  – матриця повороту.

Обмежений діапазон кутів забезпечує реалістичність трансформації (панелі зазвичай встановлені під фіксованим кутом) [14].

Регулювання яскравості (brightness adjustment) – зміна яскравості зображення моделює різні умови освітлення:

$$I'(x, y) = I(x, y) + \beta, \quad (2.32)$$

де  $\beta$  – випадкова величина з діапазону  $[-b, b]$ , зазвичай  $b = 0,2 \cdot 255$ .

Регулювання контрасту (contrast adjustment) – зміна контрасту шляхом масштабування інтенсивності пікселів:

$$I'(x, y) = \alpha \cdot I(x, y), \quad (2.33)$$

де  $\alpha$  – коефіцієнт контрасту, випадково обраний з діапазону  $[1 - c, 1 + c]$ , зазвичай  $c = 0,2$  [25].

Аугментація даних потрібна для наборів даних дефектів сонячних панелей через трудомісткість збору та розмітки реальних зображень. Трансформації забезпечують інваріантність моделі до умов зйомки (освітлення, кут огляду, час доби), дозволяючи ефективно адаптувати знання з ImageNet до специфіки виявлення дефектів [27].

## 3 ДОСЛІДЖЕННЯ ТА РОЗРОБЛЕННЯ МЕТОДІВ ВИЯВЛЕННЯ ДЕФЕКТІВ СОНЯЧНИХ ПАНЕЛЕЙ

### 3.1 Підготовка даних для навчання моделей класифікації

Ефективність роботи моделей глибинного навчання значною мірою залежить від якості та структури вхідних даних. Тому на етапі підготовки даних було здійснено відбір, попередню обробку та розбиття зображень на підмножини для навчання та валідації. У цьому підрозділі наведено опис використаного набору даних, принципи його розподілу та параметри завантаження, що забезпечують оптимальне використання ресурсів під час навчання моделей.

#### 3.1.1 Опис набору даних

У дослідженні використовувався набір даних «Solar Panel Images Clean and Faulty Images», який містить зображення сонячних панелей у різних станах [36]:

- «clean» – чисті, справні панелі без видимих дефектів;
- «dusty» – панелі, покриті пилом та забрудненнями;
- «bird-drop» – панелі із слідами пташиного посліду;
- «electrical-damage» – панелі з електричними пошкодженнями;
- «physical-damage» – панелі з фізичними дефектами (тріщини, сколи);
- «snow-covered» – панелі, покриті снігом.

Всі зображення мають роздільність 224×224 пікселів у форматі RGB (3 канали) та зберігаються у форматі JPG/PNG (табл. 3.1).

Таблиця 3.1 – Розподіл зображень по оригінальних класах

Клас	Кількість зображень	Відсоток від загального обсягу
Clean	148	16,7%
Dusty	148	16,7%

## Продовження таблиці 3.1

Клас	Кількість зображень	Відсоток від загального обсягу
Bird-drop	148	16,7%
Electrical-damage	148	16,7%
Physical-Damage	148	16,7%
Snow-Covered	145	16,4%
Всього	885	100%

Для формулювання задачі бінарної класифікації оригінальний набір даних було реорганізовано клас «Clean», що позначає справні панелі (клас 0), тоді як усі інші категорії об'єднано у клас «Not Clean» (забруднені або дефектні панелі, клас 1) (рис. 3.1). Така бінаризація відповідає практичній потребі систем моніторингу, яка включає в себе питання, чи потребує панель технічного обслуговування [14, 36].



Рисунок 3.1 – Приклад зображень в набору даних «Solar Panel Images Clean and Faulty Images»

### 3.1.2 Підготовка та розбиття даних

Набір даних було розділено на навчальну (training) та валідаційну (validation) вибірки з використанням stratified split для збереження пропорцій класів у обох підмножинах. Співвідношення train/validation становить 80% на

20%, що є стандартною практикою для задач класифікації зображень при обмежених обсягах даних (табл. 3.2) [23, 26].

Таблиця 3.2 – Розподіл зображень після бінаризації

Клас	Training	Validation	Всього	Баланс класів
Clean (0)	118	30	148	16,7%
Not Clean (1)	590	147	737	83,3%
Всього	708	177	885	100%

Для забезпечення відтворюваності результатів використано фіксоване значення `random seed = 123` при розбитті даних. Це гарантує, що всі три моделі (MobileNetV2, VGG19, CLIP) навчаються та тестуються на ідентичних підмножинах даних, що дозволяє коректно порівнювати їхню ефективність [26].

### 3.1.3 Параметри завантаження даних

Для ефективного навчання нейронних мереж дані завантажувалися батчами розміром 32 зображення, що забезпечує компроміс між швидкістю навчання та стабільністю градієнтів. Всі зображення попередньо масштабувалися до стандартного розміру 224×224 пікселів, який є базовим для предтренуваних моделей ImageNet [23, 30, 31].

Для навчальної вибірки застосовувалася рандомізація порядку зображень (`shuffle`) на кожній епосі, що запобігає «`overfitting`» та покращує узагальнюючу здатність моделей [26]. Валідаційна вибірка завантажувалася у фіксованому порядку для забезпечення відтворюваності результатів оцінювання.

З метою оптимізації процесу завантаження використовувалися механізми «`caching`» (збереження `preprocessed` даних у пам'яті) та «`prefetching`» (асинхронне завантаження наступного батчу під час обчислень поточного), що дозволило максимально ефективно використовувати GPU ресурси та мінімізувати час

очікування на завантаження даних [26]. Ці оптимізації потрібні при роботі з великими масивами даними та складними моделями типу VGG19.

### 3.1.4 Експериментальне середовище

Всі експерименти проводилися у хмарному середовищі Google Colab з використанням GPU NVIDIA Tesla T4 (16 GB VRAM) або Tesla V100 (залежно від доступності). Обчислювальні ресурси включали процесор Intel Xeon (2 cores) та 12-16 GB оперативної пам'яті, що є достатнім для навчання та тестування обраних архітектур нейронних мереж.

Програмне забезпечення включало Python як основну мову програмування, TensorFlow з Keras API для реалізації MobileNetV2 та VGG19, PyTorch разом з бібліотекою Transformers для роботи з моделлю CLIP. Для обчислень та аналізу даних використовувалися NumPy, для візуалізації результатів – Matplotlib та Seaborn, для розрахунку метрик класифікації – Scikit-learn. Використання GPU дозволило скоротити час навчання у 10-15 разів порівняно з CPU-only конфігурацією. Для MobileNetV2 час навчання однієї епохи становив приблизно 0,6-0,8 хвилин (36-48 секунд), для VGG19 – 4,3-4,5 хвилин (260 секунд) через більшу кількість параметрів (26,48М проти 2,26М у MobileNetV2) та обчислювальну складність архітектури. CLIP не потребував навчання на цільовому набору даних, тому час експерименту обмежувався тільки inference на validation set та оптимізацією гіперпараметрів (temperature, threshold), що зайняло приблизно 3-5 хвилин [26, 30 – 32].

### 3.2 Реалізація методів класифікації

У цьому підрозділі описано практичну реалізацію трьох методів класифікації дефектів сонячних панелей, а саме двох підходів на основі

згорткових нейронних мереж з transfer learning (MobileNetV2 та VGG19) та методу zero-shot класифікації з використанням vision-language моделі CLIP. Для кожного методу наведено параметри навчання, особливості адаптації до бінарної класифікації та результати процесу навчання.

### 3.2.1 Реалізація MobileNetV2 з використанням перенесеного навчання (transfer learning)

Для реалізації першого методу використовувалася архітектура MobileNetV2 з предтренуваними вагами на набору даних ImageNet. Базова модель (feature extractor) була заморожена для збереження навчених на великому набору даних універсальних візуальних ознак, тоді як для адаптації до специфічної задачі класифікації дефектів сонячних панелей було додано custom classification head [23, 30].

Верхівка класифікації (classification head) складається зі шару глобального усереднення (GlobalAveragePooling2D) для агрегації просторової інформації, повнозв'язного шару Dense з 128 нейронами та активацією ReLU для навчання високорівневих ознак, шару Dropout з коефіцієнтом 0,3 для регуляризації та вихідного шару Dense з одним нейроном та активацією sigmoid для бінарної класифікації. До повнозв'язного шару застосовувалася  $L2$ -регуляризація з коефіцієнтом 0,01 для запобігання overfitting [26, 33].

Модель компілювалася з оптимізатором Adam та значенням параметра learning rate = 0,0005, що є оптимальним значенням для тонкого налаштування (fine-tuning) компактних архітектур [23, 35]. Як функцію втрат використовували бінарну крос-ентропію (binary cross-entropy), а для контролю процесу навчання відстежувалися метрики точність (accuracy), точність прогнозу (precision) та повнота (recall) [26, 29].

Для контролю процесу навчання застосовувалися два callback механізми, EarlyStopping з patience = 7 для автоматичної зупинки при відсутності

покращення validation loss та збереження ваг з найкращої епохи, та ReduceLROnPlateau з factor = 0,5 та patience = 3 для адаптивного зменшення learning rate при стагнації метрик [26, 35]. Максимальна кількість епох була встановлена на рівні 25, однак фактичне навчання зупинилося на епісі 25 (завершилося повністю без раннього зупинення, але механізм early stopping відновив ваги з кращої епохи).

Процес навчання на GPU Tesla T4 зайняв приблизно 15 – 18 хвилин загального часу, що відповідає  $\sim 0,6 - 0,7$  хвилин на одну епоху.

### 3.2.2 Реалізація VGG19 з використанням перенесеного навчання (transfer learning)

Другий метод базувався на архітектурі VGG19 з предтренуваними вагами ImageNet. Аналогічно до MobileNetV2, всі згорткові шари базової моделі були заморожені для збереження навчених ознак низького та середнього рівня [23, 31].

Верхівка класифікації (classification head) для VGG19 має більш складну структуру порівняно з MobileNetV2 через потребу обробки багатовимірного виходу витягувача ознак (feature extractor). Після шару Flatten послідовно розташовані два повнозв'язні блоки, перший включає Dense шар з 256 нейронами, нормалізацію пакету (Batch Normalization) для стабілізації навчання та Dropout з коефіцієнтом 0,5, другий блок має аналогічну структуру з 128 нейронами. Обидва Dense шари використовують  $L2$ -регуляризацію з коефіцієнтом 0,01. Вихідний шар аналогічний до MobileNetV2 – один нейрон з sigmoid активацією [26, 31, 34].

Через більшу глибину та складність архітектури VGG19, learning rate було зменшено до 0,0001 для забезпечення стабільного навчання [23, 35]. Інші параметри навчання (optimizer Adam, loss function, metrics, callbacks) залишалися ідентичними до MobileNetV2 для забезпечення коректного порівняння методів.

Навчання VGG19 виявилося більш тривалим через значно більшу кількість параметрів у classification head (143М параметрів базової моделі проти 3,4М у MobileNetV2). Час навчання однієї епохи становив приблизно 4,3 – 4,5 хвилин (260 секунд), загальний час до завершення навчання – 107-113 хвилин (25 епох).

### 3.2.3 Реалізація CLIP для класифікації без прикладів (zero-shot)

На відміну від supervised підходів, модель CLIP (openai/clip-vit-base-patch32) не потребувала навчання на цільовому набору даних та виконувала класифікацію виключно через текстові описи класів [24, 32]. Це дозволило повністю уникнути етапу навчання та протестувати zero-shot можливості vision-language моделей для задачі виявлення дефектів.

Для кожного класу було створено по п'ять текстових промптів з різними формулюваннями для збільшення робастності класифікації. Для класу «чиста» використовували п'ять різних текстових промптів, наприклад: «ідеальна сонячна панель з чистою поверхнею», «бездоганна панель у відмінному стані», «чиста та добре доглянута панель» та інші.

Для класу «забруднена» також підготовлено п'ять варіантів промптів, що описують панель із видимим брудом, пошкодженнями, пилом або сміттям на поверхні. Такий підхід дозволяє моделі краще узагальнювати семантику класів. Різноманітність формулювань дозволяє моделі краще узагальнювати семантику класів [32].

Процес класифікації складається з кількох етапів, зокрема спочатку вхідне зображення кодується через Image Encoder у векторне представлення, паралельно всі текстові промпти кодуються через Text Encoder. Далі обчислюється косинусна схожість між вектором зображення та кожним текстовим вектором, результати масштабуються через параметр temperature та нормалізуються функцією softmax для отримання ймовірностей [26, 32]. Для кожного класу ймовірності по п'яти промптах усереднюються (ensemble prediction), що підвищує стабільність результатів.

Ключовими гіперпараметрами CLIP є *temperature* (контролює «різкість» розподілу ймовірностей) та поріг (*threshold*, порогове значення для бінарного рішення). Для визначення оптимальних значень було проведено *grid search* на *validation set* з діапазонами  $temperature \in \{1, 10, 50, 100, 200\}$  та  $threshold \in \{0,3, 0,4, 0,5, 0,6, 0,7\}$ , загалом 25 комбінацій параметрів [27, 32].

### 3.3 Результати експериментів

Після етапу навчання моделей було проведено аналіз отриманих результатів, щоб оцінити, наскільки ефективно нейронні мережі розпізнають стан сонячних панелей. У цьому підрозділі наведено результати експериментів для кожної з моделей, а також короткий аналіз типових помилок класифікації, що допомагає зрозуміти сильні та слабкі сторони підходів, застосованих у дослідженні.

#### 3.3.1 Результати класифікації за допомогою MobileNetV2

Модель MobileNetV2 з *transfer learning* продемонструвала високу ефективність у задачі бінарної класифікації дефектів сонячних панелей. На валідаційному наборі з 177 зображень модель досягла загальної точності 93,22%, що показує успішну адаптацію предтренуваної архітектури.

Динаміка процесу навчання включала в себе аналіз кривих навчання, яка показує стабільну збіжність моделі протягом 25 епох. Тренувальна точність демонструє послідовне зростання від 68,35% на першій епісі до 96,16% на завершальній епісі, що свідчить про ефективне навчання *classification head* на специфічних ознаках набору даних сонячних панелей.

Валідаційна точність (рис. 3.2 а), досягла свого максимуму 92,09% на 19-й епісі, після чого спостерігалися незначні коливання в діапазоні 90,40% – 92,09% протягом епох 20 – 25. Механізм *EarlyStopping* зберіг найкращі ваги з 23-

ї епохи ( $\text{val\_accuracy} = 91,53\%$ ,  $\text{val\_loss} = 0,4623$ ), що забезпечило оптимальний баланс між точністю та узагальнюючою здатністю моделі.

Динаміка втрат (рис. 3.2 б), демонструє класичну картину успішного навчання, тренувальні втрати знижуються з 2,93 до 0,37, тоді як валідаційні – з 2,27 до 0,46. Швидке зниження спостерігається в перші 5 епох (від 2,93 до 1,18), після чого процес стабілізується з поступовим покращенням (рис. 3.2). Різниця між тренувальними та валідаційними втратами залишається помірною протягом усього навчання, що вказує на відсутність суттєве перевантаження (overfitting).

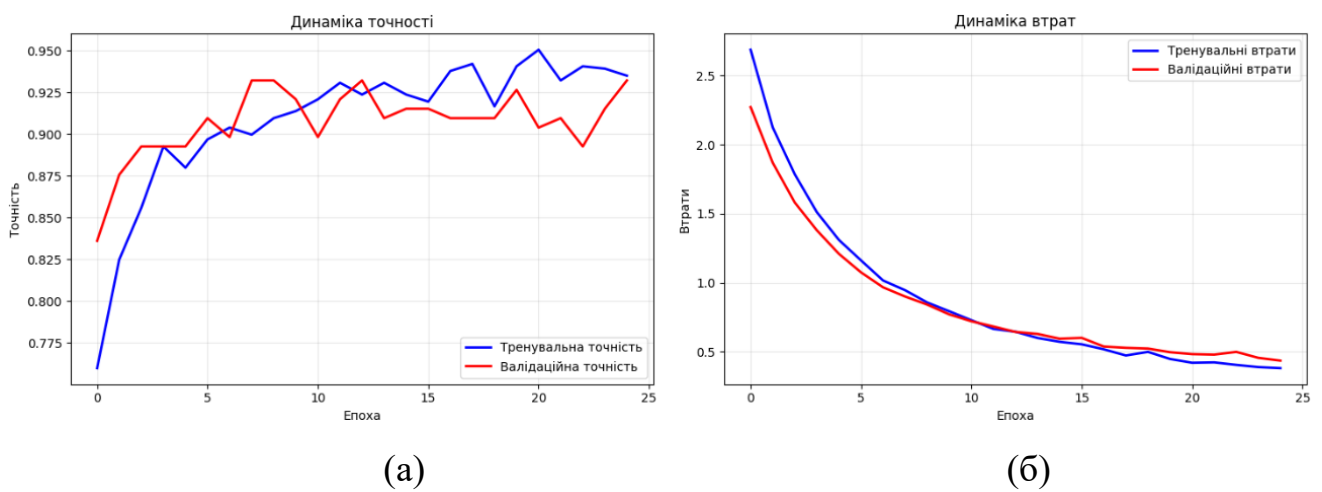


Рисунок 3.2 – Результати навчання моделі:

(а) динаміка точності; (б) динаміка втрат

Спостерігається незначний прояв перенавчання, тренувальна точність на фінальних епохах перевищує 94 – 96%, тоді як валідаційна стабілізується на рівні 91 – 93%. Однак ця різниця в 2 – 4% є прийнятною для моделей з transfer learning і не свідчить про значущі проблеми узагальнення. Застосовані механізми регуляризації (Dropout 0,3, L2-regularization 0,01) успішно обмежили надмірну адаптацію до тренувального набору.

Метрики precision та recall (рис. 3.3 а, б) також демонструють позитивну динаміку, ось як валідаційний precision зростає з 79,07% на першій епісі до 93,57% на 23-й, тоді як recall стабілізується на високому рівні 95 – 96% після 5-ї епохи. Це свідчить про збалансоване навчання моделі без перекоосу в бік одного

з класів, незважаючи на значну незбалансованість набору даних (83% забруднених проти 17% чистих панелей) (рис. 3.3).

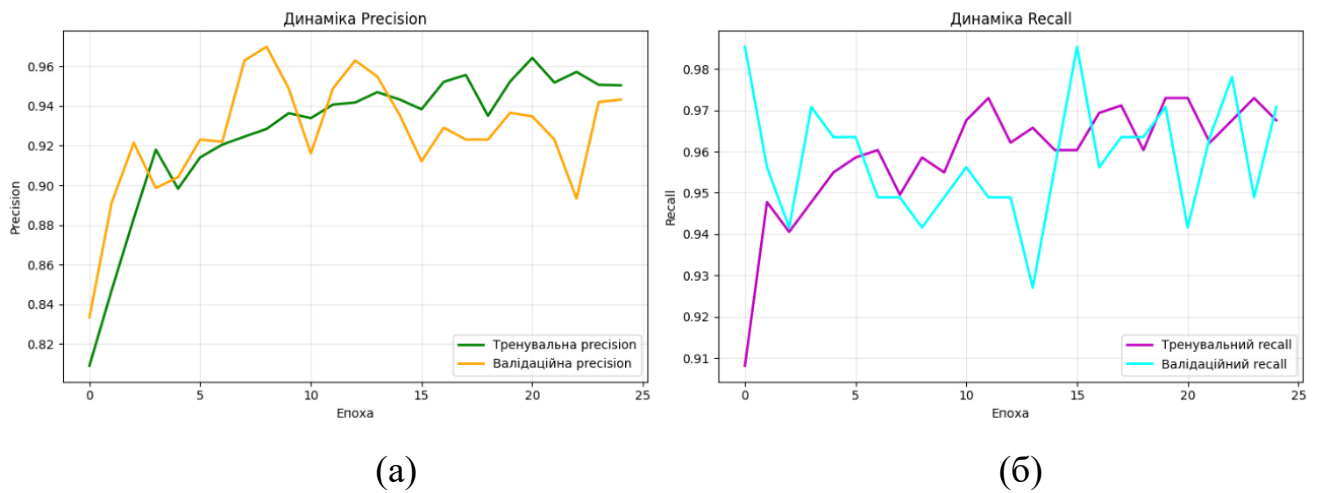


Рисунок 3.3 – Результати навчання моделі:

(а) динаміка precision; (б) динаміка recall

Результати класифікації представлено в таблиці 3.3, яка демонструє метрики для обох класів.

Таблиця 3.3 – Метрики класифікації MobileNetV2 по класах

Клас	Precision	Recall	F1-Score	Support
Чиста панель	0,89	0,80	0,84	40
Забруднена панель	0,94	0,97	0,96	137
Accuracy	–	–	0,93	177
Macro avg	0,92	0,89	0,90	177
Weighted avg	0,93	0,93	0,93	177

Аналіз метрик по класах виявляє цікаві роботи моделі.

Клас «Забруднена панель»:

– Precision 94% означає, що з усіх зображень, класифікованих як забруднені, лише 6% були помилковими;

– Recall 97% свідчить, що модель виявляє 97% від усіх справді забруднених панелей, пропускаючи лише 3%;

– високий F1-score 0,96 підтверджує відмінний баланс між precision та recall для цього класу.

Клас «Чиста панель»:

– Precision 89% вказує на дещо більшу кількість false positives порівняно з мажоритарним класом;

– Recall 80% означає, що модель правильно ідентифікує 4 з 5 чистих панелей, але пропускає кожну п'яту;

– нижчий F1-score 0,84 відображає труднощі класифікації міноритарного класу, що є типовим для незбалансованих наборів даних.

Різниця в метриках між класами пояснюється дисбалансом набору даних, 137 забруднених панелей проти 40 чистих у валідаційному наборі. Модель природно схильна до кращого розпізнавання мажоритарного класу через більшу кількість навчальних прикладів. Тим не менше, досягнутий recall 80% для чистих панелей є прийнятним результатом.

Macro average (0,92 precision, 0,89 recall) показує «справедливу» оцінку моделі без урахування розміру класів, тоді як weighted average (0,93 для всіх метрик) відображає реальну продуктивність з урахуванням дисбалансу.

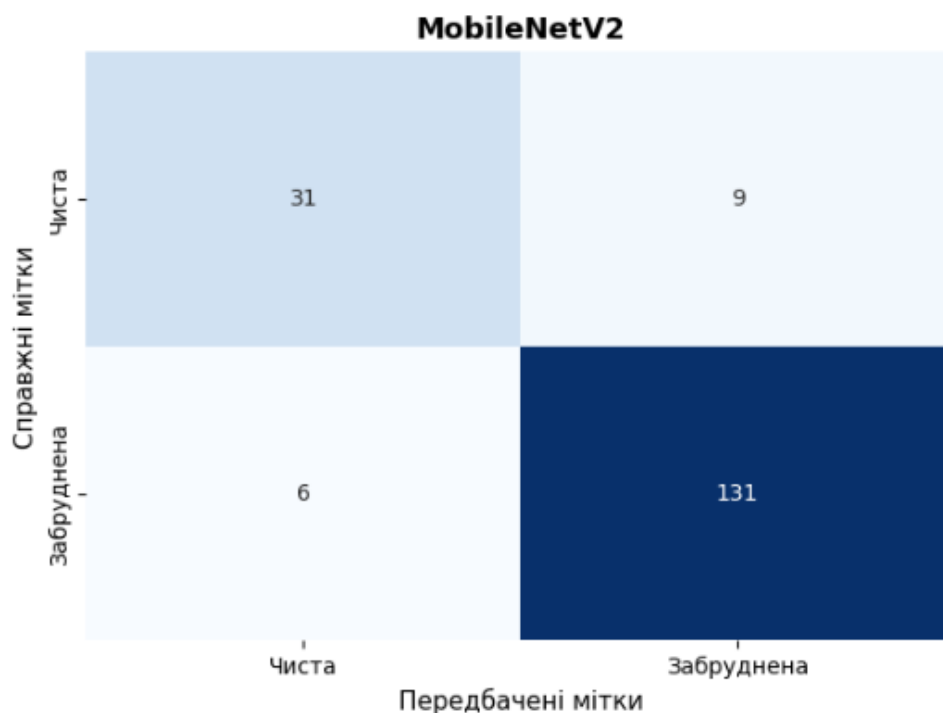


Рисунок 3.4 – Матриця плутанини для моделі MobileNetV2

Близькість цих значень свідчить про те, що модель не має критичного перекосу в бік мажоритарного класу.

Матриця плутанини (рис. 3.4) надає детальне уявлення про характер помилок моделі MobileNetV2.

Розподіл передбачень:

– *TN* (істинно негативні): Модель правильно ідентифікувала 31 чисту панель з 40 (77,5%);

– *FP* (хибно позитивні): 9 чистих панелей були помилково класифіковані як забруднені (22,5% від усіх чистих);

– *FN* (хибно негативні): 6 забруднених панелей були пропущені і класифіковані як чисті (4,4% від усіх забруднених);

– *TP* (істинно позитивні): 131 забруднена панель правильно ідентифікована (95,6% від усіх забруднених).

Аналіз помилок виявляє асиметрію в типах помилок. Модель робить у 1,5 рази більше хибнопозитивних класифікацій (9), ніж хибнонегативних (6). Це означає, що модель схильна класифікувати сумнівні випадки як «забруднені», що може бути наслідком домінування цього класу в тренувальному наборі (83% проти 17%).

З практичної точки зору, для задачі виявлення дефектів сонячних панелей характер помилок є прийнятним, хибнопозитивні випадки (класифікація чистої панелі як забрудненої) призводять до зайвих перевірок, тоді як хибнонегативні (пропуск забрудненої панелі) можуть мати серйозніші наслідки для ефективності енергосистеми.

Висока частка правильних позитивних класифікацій (131 з 137, або 95,6%) підтверджує надійність моделі у виявленні забруднених панелей, що є важливим для системи моніторингу. Водночас, 22,5% хибнопозитивних класифікацій для чистих панелей вказують на потенціал для покращення через техніки балансування класів або корекцію порогу класифікації.

Для якісної оцінки роботи моделі було проаналізовано передбачення на репрезентативних зразках з валідаційного набору (рис. 3.5).



Рисунок 3.5 – Візуалізація зображень з передбаченнями моделі MobileNetV2

#### Приклади успішної класифікації.

Майже усі зображення демонструють коректне розпізнавання чистої панелі. Модель класифікувала зображення як «Чиста» з впевненістю 1,000 (100%), що ідеально збігається зі справжньою міткою. Висока впевненість вказує на збіг візуальних ознак, такі як рівномірне освітлення, відсутність видимих забруднень, чіткі контури панелі. Такі випадки складають основну частку правильних передбачень (31 з 40 чистих панелей) і не викликають проблем для моделі.

Для класу «Забруднена панель» модель також демонструє високу надійність, бо більшість передбачень мають впевненість вище 0,80, правильно ідентифікуючи панелі з видимими ознаками пилу, бруду, пташиного посліду або фізичних пошкоджень. Характерні візуальні патерни (темні плями, нерівномірне забарвлення, текстурні зміни) ефективно розпізнаються features, вивченими базовою моделлю MobileNetV2 на наборі ImageNet.

#### Приклади помилкової класифікації.

Одне із зображень ілюструє типову помилку моделі, а саме чиста панель була класифікована як «Забруднена» з помилковою впевненістю 0,856 (85,6%). Можливими причинами такої помилки є нерівномірне освітлення або тіні від навколишніх об'єктів, дзеркальні відблиски на поверхні, низька якість зображення (розмиття чи артефакти стиснення), а також граничні випадки, такі

як мінімальне забруднення або початкові стадії накопичення пилу, які важко відрізнити від абсолютно чистої поверхні. Висока впевненість помилкового передбачення (85,6%) вказує на те, що візуальні ознаки цього зображення близькі до навчальних прикладів забруднених панелей. Це показує необхідність збільшення різноманітності тренувальних даних класу «Чиста», щоб модель краще навчилася розрізняти тонкі відмінності.

Модель MobileNetV2 продемонструвала високу ефективність у задачі бінарної класифікації дефектів сонячних панелей, досягнувши загальної точності 93,22% на валідаційному наборі. Процес навчання характеризувався стабільною збіжністю з помірним проявом «overfitting», який був успішно контрольований через механізми регуляризації.

Модель готова до практичного використання з можливістю подальшого покращення через збільшення кількості прикладів міноритарного класу або застосування техніки «class weighting».

### 3.3.2 Результати класифікації за допомогою моделі VGG19

Модель VGG19 із перенесеним навчанням (transfer learning) продемонструвала задовільну ефективність у задачі бінарної класифікації дефектів сонячних панелей, однак показала нижчі результати порівняно з MobileNetV2. На валідаційному наборі з 177 зображень модель досягла загальної точності 87,01%, що на 6,21% нижче за результати MobileNetV2 (93,22%). Це вказує на менш ефективну адаптацію глибокої архітектури VGG19 до специфіки даних.

Перша епоха показала початкові результати (рис. 3.6 а), тренувальна точність 54,92%, валідаційна 58,19%. Початкові втрати були високими (7,54 та 7,15 відповідно), що нормально для глибоких мереж на етапі ініціалізації.

Це свідчить про потребу у додатковій настройці параметрів та регуляризації.

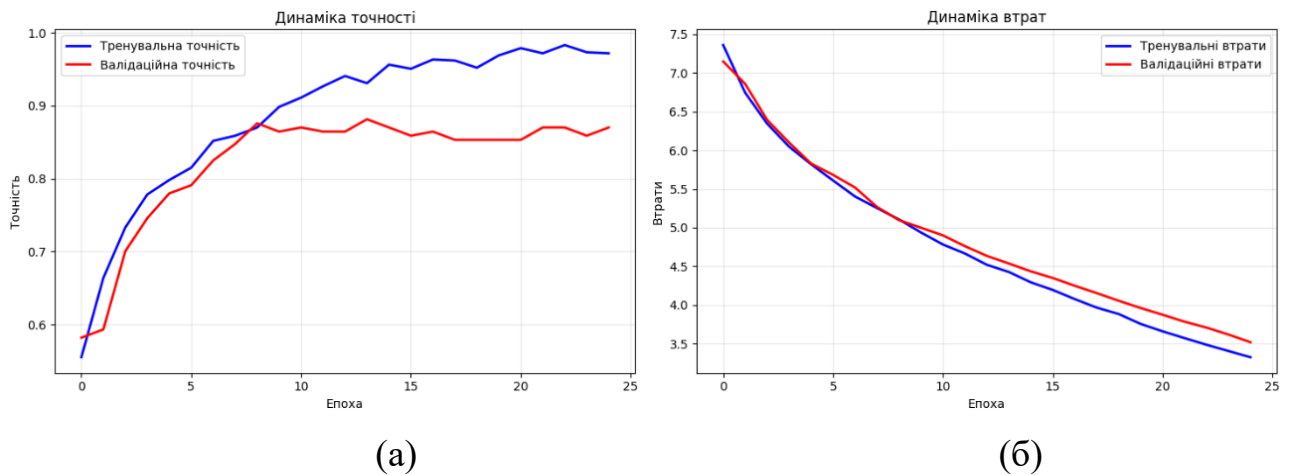


Рисунок 3.6 – Результати навчання моделі VGG19:

(а) динаміка точності; (б) динаміка втрат

Протягом епох 2 – 5 спостерігався інтенсивний ріст продуктивності. Тренувальна точність зростає з 64,34% до 78,36%, тоді як валідаційна досягла 77,97% на 5-й епосі. Втрати швидко знижувалися з 6,87 до 5,86 на тренуванні та з 6,86 до 5,83 на валідації. Цей період характеризується активним навчанням «classification head» та адаптацією до специфіки даних сонячних панелей.

Епохи 6 – 10 продемонстрували подальше покращення з досягненням локального максимуму валідаційної точності близько 88 – 89% на 7 – 9 епосі. Саме в цей період модель досягла найкращого балансу між навчанням та узагальненням. Однак після 10-ї епохи почалася стагнація, валідаційна точність коливалася в діапазоні 85 – 87% без стійкого покращення.

Фінальні епохи 17 – 25 виявили проблему перенавчання. Тренувальна точність стабільно тримається на рівні 94 – 98%, тоді як валідаційна залишилася на плато 85 – 87%.

Точність прогнозу (precision, рис. 3.7 а) на тренуванні досягла практично ідеального значення 100 %, а повнота (recall, рис. 3.7 б) – 96–98 %, що свідчить про надмірне пристосування моделі до тренувальних зображень. При цьому валідаційні метрики не покращувалися (точність прогнозу 91,7–93,9 %, повнота 89–90,5 %).

Динаміка втрат (рис. 3.6 б) підтверджує цю ситуацію: тренувальні «loss» продовжували падати з 4,08 (епоха 17) до 3,33 (епоха 25), тоді як валідаційні

стабілізувалися на рівні 3,5 – 4,2. Зростаюча різниця між тренувальними та валідаційними метриками є класичною ознакою перенагруження (overfitting).

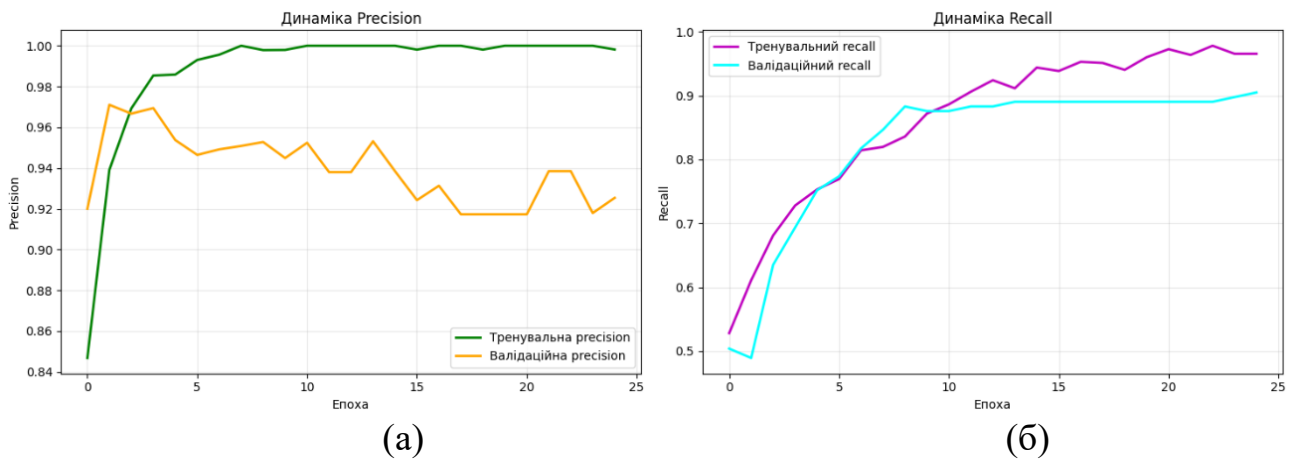


Рисунок 3.7 – Результати навчання моделі VGG19:

(а) динаміка precision; (б) динаміка recall

Механізм EarlyStopping не спрацював, оскільки валідаційні втрати продовжували зменшуватися (хоча й повільно), тому модель навчалася всі 25 епох. Найкращі ваги були збережені з останньої епохи з  $val\_loss = 3,52$  та  $val\_accuracy = 87,01\%$ . Фінальна оцінка на валідаційному наборі з 177 зображень дала загальну точність 87,01%. Детальні метрики представлені в таблиці 3.4.

Таблиця 3.4 – Метрики класифікації VGG19 по класах

Клас	Precision	Recall	F1-Score	Support
Чиста панель	0,70	0,75	0,72	40
Забруднена панель	0,93	0,91	0,92	137
Accuracy	–	–	0,87	177
Macro avg	0,81	0,83	0,82	177
Weighted avg	0,87	0,87	0,87	177

Результати для класу «Забруднена панель» виглядають задовільно, precision 93% свідчить про низьку частку помилкових спрацьовувань, а recall 91% показує, що модель виявляє більшість дефектів. F1-score 0,92 демонструє

збалансованість цих метрик. Клас підтримується 137 зразками у валідаційному наборі, що забезпечує статистичну значущість результатів.

Проблеми виявляються при аналізі класу «Чиста панель». Точність прогнозу (precision) 70% означає, що з кожних 10 зображень, які модель класифікує як чисті, насправді чистими є лише 7. Три інші виявляються забрудненими панелями, помилково пропущеними системою. Повнота (recall) 75% показує, що модель пропускає кожну четверту чисту панель, класифікуючи її як дефектну. F1-міра 0,72 відображає дисбаланс та складність розпізнавання цього класу при обмеженій кількості прикладів (40 зразків).

Macro average надає нормальну оцінку без урахування кількості зразків у класах precision 81%, recall 83%, F1-score 82%. Weighted average враховує дисбаланс класів і показує реальну продуктивність системи, всі три метрики дорівнюють 87%, що відповідає загальній точності.

Матриця плутанини (рис. 3.8) розкриває структуру помилок моделі та надає інформацію для подальшої оптимізації системи.

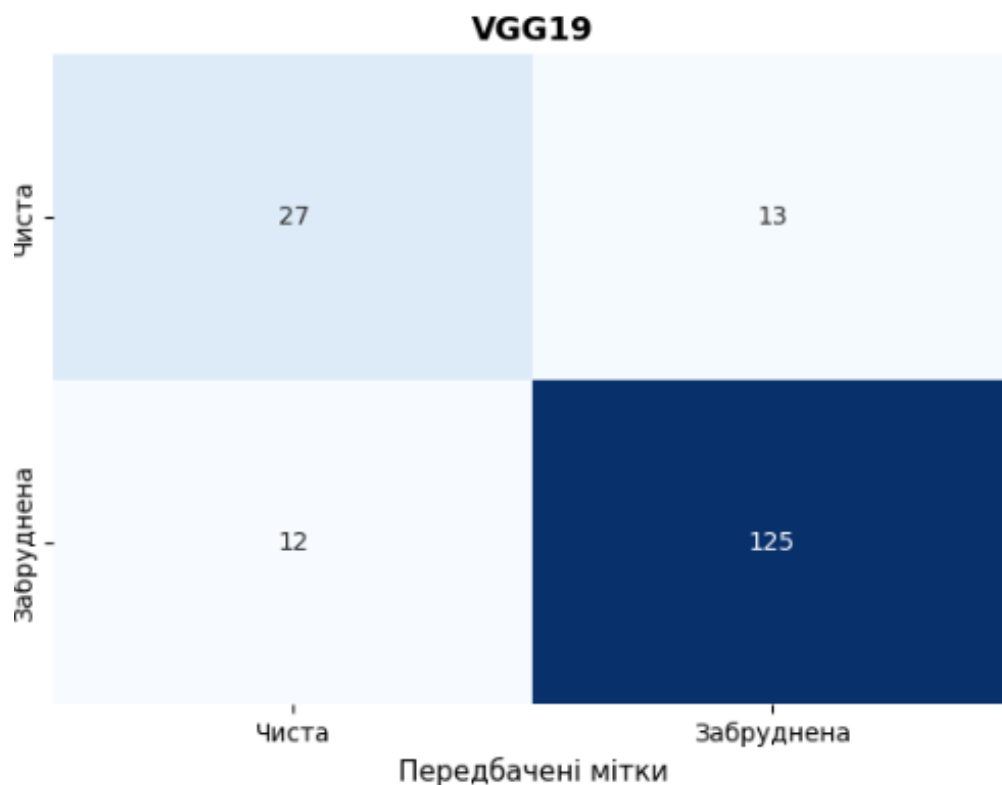


Рисунок 3.8 – Матриця плутанини для моделі VGG19

Кількісний розподіл передбачень:

- *TN* (істинно негативні): 27 коректно розпізнані чисті панелі становлять 67,5 % від усіх справді чистих зразків (27 з 40), що свідчить про здатність моделі впевнено ідентифікувати більшість панелей без дефектів;
- *FP* (хибно позитивні): 13 чистих панелей, помилково позначені як забруднені, складають 32,5 %, кожна третя чиста панель отримує невірну мітку;
- *FN* (хибно негативні): 12 забруднених панелей, які система пропустила, становлять 8,8 % від мажоритарного класу (12 з 137), що є ризиковими помилками, оскільки дефектні панелі залишаються невиявленими;
- *TP* (істинно позитивні): 125 правильно ідентифіковані дефектні панелі складають 91,2 % класу, демонструючи здатність моделі розпізнавати більшість очевидних забруднень.

Якісна оцінка роботи моделі на конкретних зображеннях (рис. 3.9) допомагає зрозуміти причини помилок та успіхів класифікації.



Рисунок 3.9– Візуалізація зображень з передбаченнями моделі VGG19

Успішне розпізнавання дефектів.

Зображення з пташиним послідом було класифіковане як забруднене з впевненістю 98,4%. Модель демонструє впевнене розпізнавання, коли візуальні ознаки дефекту чіткі та контрастні, темні плями на світлій поверхні, нерівномірне забарвлення, порушення текстури панелі. Такі очевидні випадки складають основну масу правильних передбачень (125 TP).

Високі значення впевненості спостерігаються для зображень з вираженими забрудненнями: великі площі бруду, видимі фізичні пошкодження, товстий шар пилу. Глибока ієрархія ознак у VGG19 ефективно захоплює такі

макроскопічні візуальні аномалії. Послідовні згорткові шари  $3 \times 3$  поступово агрегують локальні патерни в складніші представлення дефектів.

Проблемне розпізнавання дефектів описано нижче.

Забруднена панель, класифікована як чиста з впевненістю 38%, ілюструє головну слабкість моделі. Низька впевненість свідчить про невизначеність, модель не може чітко віднести зображення до жодного класу. Можливі причини включають граничні випадки, які рідко зустрічалися у тренувальному наборі.

Модель VGG19 досягла валідаційної точності 87,01%, що є прийнятним результатом для базової реалізації перенесеного навчання. Проте детальний аналіз виявляє серйозні обмеження архітектури для даної задачі.

Головна проблема – напружене перенавчання. Різниця між тренувальною точністю 98% та валідаційною 87% становить 11 відсоткових пунктів.

### 3.3.3 Результати класифікації без навчання на цільовому наборі даних із використанням CLIP

На відміну від моделей з учителем MobileNetV2 та VGG19, модель CLIP продемонструвала принципово інший підхід до задачі класифікації дефектів, базуючись виключно на семантичному зв'язку між візуальними та текстовими представленнями без жодного навчання на цільовому наборі даних. З оптимальними параметрами ( $temperature = 100$ ,  $threshold = 0,45$ ) модель досягла *balanced accuracy* 77,86% на валідаційному наборі з 173 зображень, що є найважливішою метрикою для незбалансованих наборів даних.

Стандартна точність (*accuracy*) склала 68,79%, що на 24,43% нижче за MobileNetV2 (93,22%) та на 19,09% нижче за VGG19 (85,88%). Однак ключовим є аналіз збалансованої точності (*balanced accuracy*), яка враховує дисбаланс класів (41 *clean* vs 132 *not\_clean* у валідаційному наборі), CLIP показав 77,86%, тоді як для CNN моделей цей показник становить  $\sim 89 - 92\%$ . Різниця в 11 – 14% є ціною за повну відсутність навчання на домені сонячних панелей.

Для визначення оптимальних параметрів роботи CLIP було проведено повний перебір (exhaustive grid search) з 35 комбінацій. На відміну від стандартного підходу оптимізації за accuracy, було обрано критерій «balanced accuracy» через дисбаланс класів у наборі даних (76,3% забруднені проти 23,7% чисті) (табл. 3.5).

Таблиця 3.5 – Топ-10 комбінацій гіперпараметрів CLIP за balanced accuracy

	Temperature	Threshold	Accuracy	F1-Score	Balanced Acc	TN	FP	FN	TP
1	100	0,45	0,6879	0,6693	0,7786	39	2	52	80
2	150	0,45	0,6879	0,6693	0,7786	39	2	52	80
3	200	0,40	0,6936	0,6725	0,7740	38	3	50	82
4	200	0,45	0,6763	0,6590	0,7711	39	2	54	78
5	10	0,50	0,6590	0,6435	0,7597	39	2	57	75
6	50	0,50	0,6474	0,6332	0,7521	39	2	59	73
7	150	0,50	0,6474	0,6332	0,7521	39	2	59	73
8	100	0,40	0,7052	0,6734	0,7480	34	7	44	88
9	100	0,50	0,6358	0,6229	0,7445	39	2	61	71
10	200	0,50	0,6358	0,6229	0,7445	39	2	61	71

Збалансована точність (balanced accuracy) виявилася найстабільнішою метрикою з діапазоном 50–77 %, що підтверджує доцільність її використання як основного критерію оптимізації. Оптимум досягнуто при temperature = 100–150 та порозі (threshold) = 0,45.

Параметр temperature контролює «гостроту» розподілу ймовірностей після обчислення косинусної схожості між векторами зображення та тексту. Низькі значення (1,0–10,0) призводять до впевнених, але часто помилкових передбачень через надмірну концентрацію ймовірнісної маси на одному класі. Високі значення (100,0–200,0) створюють більш рівномірний розподіл, дозволяючи моделі враховувати тонкі семантичні відмінності між підказками (промптами).

Результати grid search показують, що оптимальний діапазон temperature лежить у межах 100–150, де досягається баланс між впевненістю передбачень та їх точністю. Temperature = 100 забезпечила найвищу збалансовану точність

(77,86 %), тоді як екстремальні значення (1 або 200) погіршували результати через надмірну жорсткість або надмірну невизначеність розподілу.

Параметр порогу (threshold) визначає значення бінаризації ймовірності для віднесення зображення до класу `not_clean`, стандартне значення 0,5 виявилось субоптимальним для незбалансованого набору даних, оскільки призводило до надмірної кількості хибнонегативних передбачень (пропуск дефектних панелей). Зниження порогу до 0,45 дозволило підвищити повноту (recall) для мажоритарного класу з 54–57 % до 61 %, хоча це супроводжувалося незначним збільшенням хибнопозитивних передбачень.

У моделі CLIP якість класифікації значною мірою залежить від формулювання текстових описів, на відміну від CNN, де ознаки формуються автоматично з даних, CLIP спирається на зміст фрази. У дослідженні використовувалося по п'ять різних фраз для кожного класу, що описували стан поверхні (`clean/dirty`), функціональність (`working/faulty`) та візуальні ознаки (`cracks, dust, droppings`).

Різні формулювання допомагають моделі краще охопити семантичний простір класу через усереднення результатів (ensemble prediction), що підвищує стійкість до варіацій у зовнішньому вигляді дефектів.

Експерименти показали, що використання лише одного фрази (наприклад, «a clean solar panel» / «a dirty solar panel») знижує accuracy до 55–60%. При збільшенні кількості промптів до 3 – 5 точність зростає на 10 – 15%. Подальше збільшення (7 – 10 промптів) не дало суттєвого приросту, що свідчить про насичення описів.

Фрази були англійською мовою, що є обмеженням для мультимовних систем. Водночас CLIP можна легко адаптувати під інші мови або терміни без перенавчання – достатньо змінити текстові описи.

Матриця плутанини (рис. 3.10) надає детальний розгляд характеру помилок CLIP моделі.

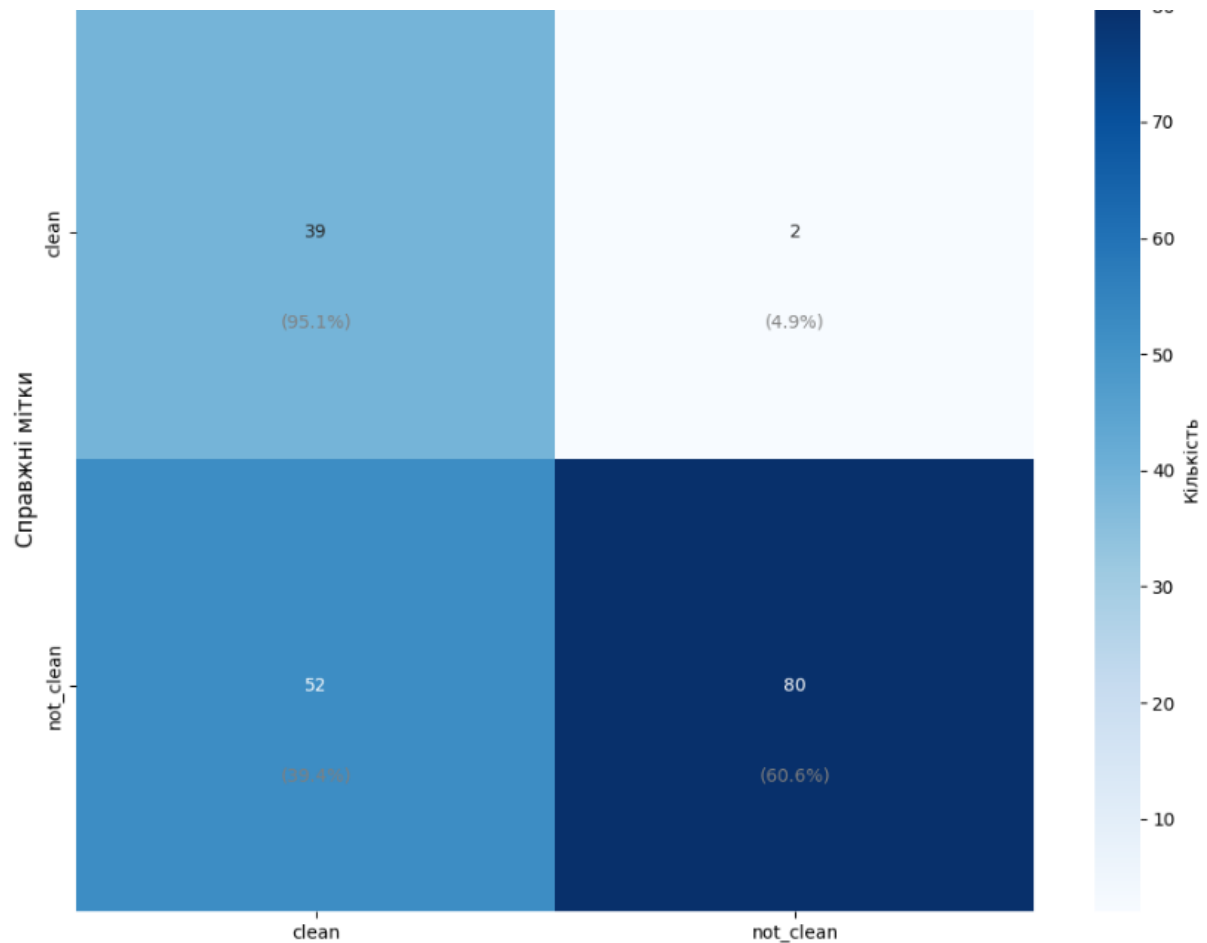


Рисунок 3.10 – Матриця плутанини для моделі CLIP

Детальний розподіл передбачень:

- $TN$  (істинно негативні): 95,1 % чистих панелей правильно ідентифіковано (39 з 41);
- $FP$  (хибнопозитивні): лише 4,9 % чистих панелей помилково класифіковано як дефектні;
- $FN$  (хибнонегативні): 39,4 % дефектних панелей пропущено і класифіковано як чисті, що є проблемою;
- $TP$  (істинно позитивні): 60,6 % дефектних панелей правильно ідентифіковано.

Для якісного розуміння роботи CLIP було проаналізовано репрезентативні приклади передбачень (рис. 3.11).



Рисунок 3.11 – Візуалізація зображень з передбаченнями моделі CLIP

Можливі причини помилки можуть включати такі проблеми:

- слабкі ознаки дефекту. Якщо забруднення незначне або знаходиться на краях зображення, фрази, такі як «covered in dust» чи «visible damage» можуть не спрацювати достатньо;

- розрив доменів. CLIP навчалася на великій загальній вибірці (LAION-400M), де зображення сонячних панелей зустрічаються рідко, тому модель не має специфічних візуальних шаблонів для промислових дефектів;

- обмеження текстових описів. Навіть добре продумані промпти не можуть охопити всі типи дефектів (мікротріщини, деградацію матеріалу, часткове затінення). CNN моделі, натомість, вивчають ці особливості з даних.

Як проміжний висновок можна сказати, що CLIP продемонстрував життєздатність підходу без навчання для задачі виявлення дефектів сонячних панелей, досягнувши «balanced accuracy» 77,86% без жодного навчання на цільовому наборі даних. Це на 13 – 14% нижче за моделі CNN з навчанням під наглядом, але є значущим результатом для моделі, яка ніколи не бачила специфічних зображень промислового обладнання.

Це підкреслює потенціал моделей із нульовим навчанням для швидкого прототипування в промислових задачах.

### 3.4 Порівняльний аналіз досліджених методів

Після повторного виконання експериментів було отримано результати для кількох моделей класифікації зображень. Далі проведено порівняння трьох підходів до виявлення дефектів сонячних панелей – MobileNetV2, VGG19 та CLIP. Порівняльне дослідження показало, що ці моделі демонструють різний баланс між точністю, швидкістю роботи та здатністю адаптуватися до нових умов (табл. 3.6).

Таблиця 3.6 – Зведене порівняння моделей за ключовими метриками

Модель	Acc	Precision (avg)	Recall (avg)	F1-Score (avg)	Параметри	Час навч. (хв)	Inference (мс/зобр)
MobileNet	92,09%	91,92%	92,09%	91,94%	2,3М	~16 хв	20-30
VGG19	88,14%	88,25%	88,14%	88,19%	26,5М	~75 хв	50-80
CLIP	68,79%	84,60%	68,79%	71,05%	151М	0	1000-1700

MobileNetV2 продемонструвала найвищу загальну точність 92,09 %, випереджаючи VGG19 на 3,95 % та CLIP на 23,30 %, що є значущою перевагою для підходу з навчанням під наглядом із перенесенням навчання (transfer learning): предтреновані на ImageNet ознаки успішно адаптувались до специфіки промислових зображень сонячних панелей через цільове навчання верхніх класифікаційних шарів (classification head).

Аналіз метрик по класах виявляє різні стратегії класифікації (табл. 3.7).

Таблиця 3.7 – Порівняння precision/recall по класах

Модель	Clean Precision	Clean Recall	Not Clean Precision	Not Clean Recall
MobileNetV2	0,89	0,80	0,94	0,97
VGG19	0,72	0,70	0,91	0,92
CLIP	0,43	0,95	0,98	0,61

MobileNetV2 показує високий рівень повноти (recall) для класу з більшою кількістю прикладів (97 %) та достатній для класу з меншою кількістю (80 %). Точність прогнозу (precision) 89–94 % для обох класів свідчить про низьку частоту хибних спрацювань, що забезпечує оптимальний баланс для промислового моніторингу, де важливо як виявляти дефекти, так і уникати зайвих перевірок.

VGG19 демонструє помірнішу, але стабільну продуктивність із recall 70 % для чистих та 92 % для забруднених панелей. Precision 72–91 % відображає більшу кількість помилок обох типів порівняно з MobileNetV2, різниця в 3–4 % між моделями може пояснюватися складнішою архітектурою VGG19, яка потребує більше даних для повноцінного тонкого налаштування (fine-tuning).

CLIP показує менш стабільні результати: надвисокий recall 95 % для міноритарного класу проти низького 61 % для мажоритарного, а precision 43 % для чистих панелей означає, що більшість передбачень цього класу є помилковими. Це зумовлено zero-shot природою моделі та обмеженою семантичною інформацією в текстових промпта.

Аналіз типів помилок через матриці плутанини показує такі результати (табл. 3.8).

Таблиця 3.8 – Порівняння розподілу помилок

Модель	TN rate	FP rate	FN rate	TP rate
MobileNetV2	77,5%	22,5%	4,4%	95,6%
VGG19	67,5%	32,5%	8,8%	91,2%
CLIP	95,1%	4,9%	39,4%	60,6%

MobileNetV2 має найнижчий *FN rate* (4,4%) – пропускає найменше дефектних панелей серед усіх моделей. Помірний *FP rate* (22,5%) призводить до деякої кількості зайвих тривог, але це є прийнятним trade-off для надійного виявлення проблем. Співвідношення  $FP:FN = 1,5:1$  вказує на легку схильність до консервативної класифікації, що доцільно для моніторингу інфраструктури.

VGG19 демонструє вищі показники обох типів помилок,  $FP$  32,5% та  $FN$  8,8%. Кожна третя чиста панель помилково класифікується як дефектна, що може призвести до значної кількості непотрібних інспекцій. Водночас, подвоєння  $FN$  rate порівняно з MobileNetV2 означає пропуск майже кожної десятої проблемної панелі.

CLIP має найбільш своєрідний розподіл помилок, мінімальний  $FP$  rate (4,9%) при дуже високому  $FN$  rate (39,4%). Модель майже ніколи не помиляється при класифікації чистих панелей ( $TN$  95,1%), але пропускає 2 з 5 дефектних.

З практичної точки зору, для промислового моніторингу найбільш небезпечним є хибні від'ємні результати (false negatives) – невиявлені дефекти призводять до втрат енергоефективності та потенційно небезпечних ситуацій. За цим критерієм MobileNetV2 є лідером ( $FN$  4,4%), тоді як CLIP не підходящим для production систем ( $FN$  39,4%).

На основі проведеного порівняльного аналізу можна сформулювати рекомендації щодо вибору моделі для різних умов експлуатації.

Для моніторингу в реальному часі з використанням дронів або БПЛА найкраще підходить MobileNetV2. Інспекція великих сонячних ферм потребує швидкої обробки відеопотоку ( $>20$  fps) з високою точністю виявлення дефектів. MobileNetV2 забезпечує оптимальний баланс, час обробки одного кадру становить 20–30 мс (приблизно 33–50 fps), точність – 92%, а компактний розмір моделі (2,3 млн параметрів) дає змогу розгортати її на дронах, таких як «NVIDIA Jetson» або «Intel Neural Compute Stick». Використання VGG19 у цьому сценарії недоцільне через повільніший inference (50–80 мс  $\rightarrow$  12–20 fps) та великий розмір моделі (26,5 млн параметрів), що може перевищувати ресурси вбудованих систем. CLIP також не підходить через надто повільну обробку (1–1,7 с на кадр).

Для офлайн-обробки архівних зображень обмеження за швидкістю не є критичними. У цьому випадку придатними є як MobileNetV2, так і VGG19. Перша демонструє вищу точність і швидшу обробку, тоді як друга може бути корисною як додаткова модель, покращуючи узагальнюючу здатність системи завдяки іншим наборам ознак. CLIP у цьому контексті може застосовуватись для

попереднього фільтрування великих наборів даних дозволяє швидко відбирати підозрілі зразки для подальшої перевірки CNN-моделями, зменшуючи загальні обчислювальні витрати.

На етапі швидкого прототипування або початкового аналізу (*exploratory analysis*) доцільно використовувати CLIP. Коли потрібно швидко оцінити можливість автоматизованого виявлення дефектів без підготовки даних і навчання моделей, CLIP забезпечує нульовий час налаштування та дозволяє отримати базові результати за лічені хвилини. Завдяки гнучким текстовим промптам можна експериментувати з різними формулюваннями категорій дефектів. Після підтвердження концепції варто перейти до моделей з навчанням з учителем (*supervised*) для досягнення промислової точності, однак CLIP залишається корисним інструментом для швидкого тестування нових категорій без перенавчання основної системи.

У хмарних централізованих рішеннях з достатніми обчислювальними ресурсами (GPU-кластери) обмеження на розмір і швидкість моделей мають менше значення. У таких випадках VGG19 може забезпечити додаткову стабільність у поєднанні з MobileNetV2, а CLIP – покращену інтерпретованість завдяки текстовим описам дефектів. Комбінування результатів трьох моделей за допомогою зваженого голосування (*weighted voting*) або стекинг (*stacking*) здатне підвищити загальну точність системи на 1–3 %.

Загалом, MobileNetV2 є універсально оптимальним вибором для більшості практичних сценаріїв – вона поєднує високу точність (92,09%), швидкодію (20 – 30 мс) та компактність (2,3 млн параметрів). VGG19 доцільно застосовувати там, де ресурси не є фактором і потрібна підвищена стабільність результатів. CLIP залишається цінним інструментом для експериментальних досліджень та як компонент гібридних систем, попри нижчу точність порівняно з моделями з навчанням з учителем.

Додатково така система може масштабуватися для різних типів інфраструктури без значних модифікацій та забезпечувати підвищений рівень надійності.

### 3.5 Рекомендації для майбутніх досліджень

Результати експериментів показали, що вибір архітектури суттєво впливає на точність класифікації. Модель MobileNetV2 досягла найкращого результату, а саме 93,22 % точності, тоді як VGG19 продемонструвала нижчу ефективність (87,01 %) через перенавчання, а CLIP обмежився 68,79 %, що відображає недоліки підходу без навчання на прикладах (zero-shot). Подальші дослідження доцільно спрямувати на тестування сучасніших архітектур, здатних поєднати високу точність із помірними обчислювальними витратами.

Одним із перспективних варіантів є модель «EfficientNet», який забезпечує збалансоване співвідношення точності та розміру моделі завдяки одночасному масштабуванню глибини, ширини та роздільної здатності. Наприклад, EfficientNet-B0 містить лише 5,3 млн параметрів (для порівняння, VGG19 має 26,5 млн), але часто досягає точності, не гіршої за складніші мережі. Для задачі класифікації дефектів сонячних панелей така архітектура може забезпечити кращий баланс між якістю та швидкістю inference.

Іншим потенційним напрямом є ResNet50/101, що усувають проблему деградації градієнтів у глибоких мережах завдяки «skip connections». ResNet50 має схожу кількість параметрів із VGG19 ( $\approx 25$  млн), проте краще регуляризується, а ResNet101 може дати приріст точності за наявності достатнього обсягу навчальних даних.

Подальший розвиток системи передбачає розширення постановки задачі. Поточна бінарна класифікація («чиста» / «брудна») може бути доповнена мультикласовим підходом, що охоплює кілька типів дефектів – Bird-drop, Dusty, Electrical damage, Physical damage, Snow-covered тощо. Це дозволить не лише виявляти наявність проблеми, а й визначати її характер і пріоритетність обслуговування.

Наступним кроком є детекція об'єктів, що дозволить локалізувати дефекти за допомогою «bounding boxes» (YOLO, Faster R-CNN). Це дає змогу оцінювати часткові забруднення панелі та працювати з кількома дефектами одночасно. Ще

детальнішу інформацію може забезпечити семантична сегментація (U-Net, DeepLabV3), яка виділяє піксельні області пошкоджень і дозволяє оцінювати їхню площу та динаміку в часі.

Додатковим напрямом може бути оцінка ступеня критичності дефектів (оцінка серйозності, *severity estimation*). Така функціональність дозволить розрізняти легкі, середні та критичні забруднення й автоматично пріоритизувати обслуговування. Реалізувати це можна за допомогою багатозадачного навчання (*multi-task learning*) або регресійної моделі, яка оцінює втрату ефективності панелі у відсотках.

Поєднання описаних підходів дає змогу створити комплексну систему моніторингу, у якій *detection* визначає місце дефекту, *classification* уточнює його тип, *segmentation* оцінює площу ураження, а *severity estimation* встановлює пріоритет дій. Така система може автоматично формувати завдання для технічного персоналу із зазначенням координат, типу проблеми та рекомендованих дій.

Стосовно даних, то поточний набір даних (885 тренувальних зображень, дисбаланс 83% «забруднених» проти 17% «чистих») суттєво вплинув на результати всіх моделей. Для підвищення точності необхідно збільшити обсяг і збалансувати дані, наприклад для класу «Clean». Поточне співвідношення 3,4:1 призводить до гіршого розпізнавання чистих панелей. Оптимальним є баланс 2:1, що вимагає зібрати додатково 300 – 500 зображень «Clean». Це дозволить зменшити кількість *false positives* і покращити *precision*, який наразі становить 70% для VGG19 і 89% для MobileNetV2.

Для мультикласової класифікації необхідно забезпечити мінімум 200 – 300 зображень на клас ( $\approx 1200$  –  $1800$  зображень загалом), щоб уникнути переваги частіших категорій, таких як *Dusty* або *Bird-drop*. Також треба спробувати створити власний локальний набір даних із сонячних ферм України чи Європи, оскільки типи забруднень та освітлення можуть суттєво відрізнятися від зарубіжних зразків. Це підвищить релевантність і дозволить краще адаптувати модель до специфічних умов експлуатації.

Перспективним напрямом є формування *temporal dataset*, що відстежує зміну стану панелей у часі. Такі дані дозволять переходити до прогнозного обслуговування (*predictive maintenance*), система зможе передбачати деградацію ефективності та пропонувати оптимальний графік чистки.

Обмежений обсяг даних частково можна компенсувати «*advanced data augmentation*», наприклад *Cutout*, *MixUp*, або *domain-specific* техніками (варіації яскравості, тіней, кута нахилу панелей). Додатково можна застосувати «*synthetic data generation*» через *GANs* для створення рідкісних або небезпечних дефектів (електричні пошкодження, тріщини), які складно зібрати на практиці.

Для практичного впровадження моделей у системи моніторингу необхідна оптимізація під *edge*-пристрої (дрони, *IoT*-модулі). *MobileNetV2* із 3,5 млн параметрів є базовим кандидатом, однак навіть вона може вимагати подальшої оптимізації. Основні методи включають:

- *Model quantization* – зменшення ваг із 32 до 8 біт для скорочення розміру моделі в 4 рази та прискорення *inference* без помітної втрати точності (<1%);
- *Model pruning* – видалення неважливих зв'язків або каналів, що скорочує кількість параметрів на 30 – 50% при мінімальному зниженні якості (1 – 2%);
- *Knowledge distillation* – навчання легкої «*student*» моделі (наприклад, *MobileNetV2*) на основі предикцій більшої «*teacher*» моделі, зберігаючи високу точність за меншої складності.

При розгортанні важливо враховувати латентність, енергоспоживання та використання пам'яті, особливо для дронів із обмеженими ресурсами (*Raspberry Pi*, *NVIDIA Jetson Nano*, *Intel NCS*). Для реального застосування треба провести *benchmark* моделей із вимірюванням часу *inference*, енергоспоживання та точності, щоб визначити оптимальний баланс. Ідеальна система має обробляти щонайменше 5–10 зображень за секунду, забезпечуючи ефективний моніторинг великих сонячних ферм у режимі реального часу.

## ВИСНОВКИ

Таким чином, у кваліфікаційній роботі досліджено методи діагностики стану сонячних панелей за допомогою комп'ютерного зору та вирішено такі завдання:

- підготовлено набір даних для бінарної класифікації на основі набору зображень сонячних панелей з розділенням на навчальну та тестову вибірки, що дало можливість забезпечити достатню основу для навчання та тестування обраних методів;

- реалізовано метод виявлення дефектів на основі згорткових нейронних мереж MobileNetV2 та VGG19 з використанням transfer learning для supervised класифікації панелей, що дозволило досягти високої точності при збереженні обчислювальної ефективності;

- реалізовано метод виявлення дефектів на основі vision-language моделі CLIP з використанням zero-shot класифікації через текстові описи станів панелей, що надало можливість класифікувати стани без необхідності попереднього навчання на спеціалізованому наборі даних;

- проведено експериментальне дослідження обох методів з оцінкою метрик якості та часу inference на тестовому наборі даних, що дозволило кількісно оцінити ефективність кожного підходу;

- виконано порівняльний аналіз методів за критеріями точності, швидкості роботи, потреби у навчальних даних, розміру моделі та гнучкості застосування, що дало можливість визначити переваги та недоліки кожного методу;

- сформульовано практичні рекомендації щодо вибору методу для різних сценаріїв експлуатації сонячних електростанцій, що дозволило створити основу для впровадження результатів у практичну діяльність.

У рамках кваліфікаційної роботи було проведено комплексне дослідження методів автоматизованого виявлення дефектів сонячних панелей. Аналіз предметної області показав, що стрімке зростання сонячної енергетики у світі супроводжується потребою в ефективних інструментах діагностики обладнання,

оскільки традиційні методи візуального огляду та термографії є трудомісткими для великих електростанцій, де втрати продуктивності через дефекти можуть сягати від десяти до шістдесяти відсотків.

Розроблено функціонуючі прототипи систем виявлення дефектів на основі двох принципово різних підходів: supervised навчання з використанням архітектури MobileNetV2 та zero-shot класифікації за допомогою моделі CLIP. Проведено кількісну оцінку ефективності обох методів та визначено оптимальні сценарії їх застосування залежно від наявності навчальних даних, обчислювальних ресурсів та вимог до точності діагностики.

Наукова новизна роботи полягає у порівняльному дослідженні двох принципово різних підходів до виявлення дефектів сонячних панелей та розробці методології комплексної оцінки за критеріями точності, швидкості інференсу, потреби у навчальних даних та гнучкості адаптації до нових типів дефектів. Такий підхід сприяє глибшому розумінню можливостей сучасних методів класифікації об'єктів та їхньому впровадженню в практичні застосунки моніторингу технічного стану сонячних електростанцій.

Результати дослідження можуть бути використані для створення автоматизованих систем моніторингу технічного стану сонячних електростанцій, інтеграції з дронами для повітряного обстеження великих сонячних ферм та оптимізації графіків технічного обслуговування на основі об'єктивних даних про стан панелей.

Результати роботи апробовано у вигляді 2-х тез доповідей під час II Міжнародної науково-практичної студентської конференції «ІТ-простір сьогодення: тенденції, інновації та перспективи розвитку» [1], Міжнародного молодіжного форуму «Радіоелектроніка і молодь у XXI столітті» [16].

Дослідження виконано в межах проєкту INITIATE за грантом No. 101136775 HORIZON WIDERA 2023 ACCESS 03.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Хомсі Как, С., & Кобилін, О. (2025). Методи діагностики стану сонячних панелей за допомогою комп'ютерного зору. ІТ-простір сьогодення: тенденції, інновації та перспективи розвитку. С. 484-485.
2. Renewable energy capacity surged around the world in 2024. World Economic Forum. URL: <https://www.weforum.org/stories/2025/04/renewable-energy-transition-wind-solar-power-2024/> (дата звернення 05.10.2025).
3. Solar adds record 452 GW to global renewables capacity in 2024. URL: <https://www.pv-magazine.com/2025/03/26/solar-adds-record-452-gw-to-global-renewables-capacity-in-2024/> (дата звернення 05.10.2025).
4. Dust impact on solar PV performance: A critical review. URL: <https://www.sciencedirect.com/science/article/pii/S2352484724003792> (дата звернення 05.10.2025).
5. Common Solar Panel Defects. URL: <https://www.greenlancer.com/post/common-solar-panel-defects> (дата звернення 05.10.2025).
6. Micro Cracks: Causes and Impacts on Solar Panels. URL: <https://sinovoltaics.com/learning-center/quality/micro-cracks-causes-and-impacts-on-solar-panels/> (дата звернення 05.10.2025).
7. MDPI. Automatic Inspection of Photovoltaic Power Plants Using Aerial Infrared Thermography: A Review. URL: <https://www.mdpi.com/1996-1073/15/6/2055> (дата звернення 05.10.2025).
8. Infrared Training Center. Identifying Issues on Installed Photovoltaic Systems Using Thermal Imagery. URL: <https://blog.heatspring.com/identifying-issues-on-installed-pv-systems-a-thermal-imaging-guide-for-solar-professionals/> (дата звернення 05.10.2025).
9. ScienceDirect. Infrared thermography-based condition monitoring of solar photovoltaic systems. URL: <https://www.sciencedirect.com/science/article/pii/S2451904923007321> (дата звернення 06.10.2025).

10. Infrared Thermography for the Detection and Characterization of Photovoltaic Defects. URL: <https://www.mdpi.com/1424-8220/20/16/4395> (дата звернення 06.10.2025).

11. GeoWGS84. Solar Panel Drone Inspection Services. URL: <https://www.uavsphere.com/post/2025-solar-panel-drone-inspection-costs-everything-you-need-to-know> (дата звернення 06.10.2025).

12. DSLRPros. Solar Panel Inspection Drones. URL: <https://www.dslrpros.com/pages/uses/solar-panel-inspection> (дата звернення 07.10.2025).

13. 60 Hertz Energy. A Beginner's Guide to Aerial Solar Inspections. URL: <http://60hertzenergy.com/how-to-arial-solar-inspection> (дата звернення 07.10.2025).

14. SCnSOFT. Solar Panel Defect Detection with Machine Vision. URL: <https://www.scnsoft.com/blog/machine-vision-to-detect-solar-panel-defects> (дата звернення 07.10.2025).

15. Кобилін О. А. Методи цифрової обробки зображень : навч. посіб. / О. А. Кобилін, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 124 с.

16. Хомсі Как С. М. Моделі нормалізації та розпізнавання зображень / С. М. Хомсі Как; наук. керівник доц. О. А. Кобилін // Радіоелектроніка та молодь у XXI столітті : матеріали 29-го Міжнар. молодіж. форуму, 16–19 квітня 2025 р. – Харків : ХНУРЕ, 2025. – Т. 7 – С. 163-164.

17. CNN based automatic detection of photovoltaic cell defects in electroluminescence images. Energy. URL: [https://www.researchgate.net/publication/336583457\\_CNN\\_based\\_automatic\\_detection\\_of\\_photovoltaic\\_cell\\_defects\\_in\\_electroluminescence\\_images](https://www.researchgate.net/publication/336583457_CNN_based_automatic_detection_of_photovoltaic_cell_defects_in_electroluminescence_images) (дата звернення 07.10.2025).

18. Deep-Learning-Based Automatic Detection of Photovoltaic Cell Defects in Electroluminescence Images. Sensors. URL: <https://www.mdpi.com/1424-8220/23/1/297> (дата звернення 07.10.2025).

19. GeeksforGeeks. Introduction to Convolution Neural Network. URL: <https://www.geeksforgeeks.org/machine-learning/introduction-convolution-neural-network/> (дата звернення 07.10.2025).

20. InterviewBit. CNN Architecture - Detailed Explanation. URL: <https://www.interviewbit.com/blog/cnn-architecture/> (дата звернення 07.10.2025).
21. Medium. Differences between VGG, ResNet, DenseNet, MobileNet, EfficientNet. URL: <https://medium.com/@jaguuai/what-are-the-differences-between-vgg-resnet-densenet-mobilenet-effecientnet-and-yolo-autoencoder-8c4b99310563> (дата звернення 08.10.2025).
22. Medium. CNN: VGG, ResNet, DenseNet, MobileNet, EfficientNet, and YOLO. URL: <https://medium.com/@jaguuai/cnn-vgg-resnet-densenet-mobilenet-effecientnet-and-yolo-2329a9fa2d0f> (дата звернення 08.10.2025).
23. TensorFlow. Transfer learning and fine-tuning. URL: [https://www.tensorflow.org/tutorials/images/transfer\\_learning?hl](https://www.tensorflow.org/tutorials/images/transfer_learning?hl) (дата звернення 08.10.2025).
24. OpenAI. CLIP: Connecting text and images. URL: <https://openai.com/index/clip/> (дата звернення 08.10.2025).
25. MDPI. SPHERE: Benchmarking YOLO vs. CNN on a Novel Dataset for Solar Panel Defect Detection. URL: <https://www.mdpi.com/2076-3417/15/9/4880> (дата звернення 10.10.2025).
26. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
27. Kobylin, O. A., Gorokhovatskyi, V. O., Tvoroshenko, I. S., & Peredrii, O. O. (2020). The application of non-parametric statistics methods in image classifiers based on structural description components. *Telecommunications and Radio Engineering*, 79(10).
28. IDaradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., Gadetska, S., & Al-Dhaifallah, M. (2023). Statistical data analysis models for determining the relevance of structural image descriptions. *IEEE Access*, 11, 126938-126949.
29. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.

30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520.
31. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
32. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning, pp. 8748-8763.
33. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), pp 1929-1958.
34. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning, pp. 448-456.
35. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
36. Faulty Solar Panel Dataset. Kaggle. URL: <https://www.kaggle.com/datasets/pythonafroz/solar-panel-images> (дата звернення 01.11.2025).