

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**імені Ігоря СІКОРСЬКОГО»**  
**Навчально-науковий фізико-технічний інститут**  
**Кафедра математичного моделювання та аналізу даних**

«На правах рукопису»  
УДК 004.932:613.2

«До захисту допущено»  
Завідувач кафедри  
\_\_\_\_\_ Іван ТЕРЕЩЕНКО  
«\_\_\_» \_\_\_\_\_ 2026 р.

**Дипломна робота**  
**на здобуття ступеня бакалавра**  
**за освітньо-професійною програмою**  
**«Прикладна математика»**

зі спеціальності: 113 Прикладна математика  
на тему: **«Розробка системи оцінки калорійності страв та фізичної активності користувача на основі аналізу зображень і спортивного навантаження»**

Виконав:  
студент IV курсу, групи ФІ-21  
Орлов Микола Вячеславович \_\_\_\_\_

Керівник:  
Асистент кафедри ММАД, д-р філософії  
Железняков Д. В. \_\_\_\_\_

Рецензент:  
науковий співробітник Інституту загальної енергетики  
НАН України, доктор філософії з інформаційно-  
вимірювальних технологій  
Романенко В.В. \_\_\_\_\_

Засвідчую, що у цій дипломній  
роботі немає запозичень з праць  
інших авторів без відповідних  
посилань.  
Студент \_\_\_\_\_

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені Ігоря СІКОРСЬКОГО»**

**Навчально-науковий фізико-технічний інститут  
Кафедра математичного моделювання та аналізу даних**

Рівень вищої освіти — перший (бакалаврський)  
Спеціальність — 113 Прикладна математика,  
ОПП «Прикладна математика»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Іван ТЕРЕЩЕНКО

«\_\_» \_\_\_\_\_ 2026 р.

**ЗАВДАННЯ  
на дипломну роботу**

Студент: Орлов Микола Вячеславович

1. Тема роботи: *«Розробка системи оцінки калорійності страв та фізичної активності користувача на основі аналізу зображень і спортивного навантаження»*, науковий керівник дисертації: Асистент кафедри ММАД, д-р філософії Железняков Д. В. ,

затверджені наказом по університету \_\_ від «\_\_» \_\_\_\_\_ 2026 р.

2. Термін подання студентом роботи: «\_\_» \_\_\_\_\_ 2026 р.

3. Об'єкт дослідження: Процес автоматичної оцінки енергетичного балансу користувача на основі аналізу зображень страв і даних фізичної активності.

4. Предмет дослідження: Методи комп'ютерного зору та глибокого навчання для класифікації харчових продуктів, оцінки їх калорійності та розрахунку енерговитрат користувача.

5. Перелік завдань:

1) провести аналіз існуючих методів розпізнавання їжі, оцінки калорійності та фізичної активності;

2) сформулювати математичну постановку задачі оцінки енергетичного балансу користувача;

- 3) розробити модульну архітектуру програмної системи;
- 4) навчити модель класифікації страв ResNet-50 на датасеті Food-101 із застосуванням трансферного навчання та двофазного fine-tuning;
- 5) провести порівняльний аналіз архітектур CNN (ResNet-50, EfficientNet-B0, MobileNetV3) за точністю, швидкістю та розміром моделі;
- 6) реалізувати модуль оцінки енерговитрат на основі MET-коефіцієнтів;
- 7) інтегрувати підсистеми у веб-застосунок CalorieTrack та провести наскрізне тестування системи.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:  
Презентація доповіді.

7. Дата видачі завдання: 10 вересня 2025 р.

#### Календарний план

з/п	Назва етапів виконання магістерської дисертації	Термін виконання	Примітка
1	Узгодження теми з науковим керівником	Вересень 2025	Виконано
2	Огляд літератури	Жовтень 2025 – грудень 2025	Виконано
3	Огляд методів	Грудень 2025 – Лютий 2026	Виконано
4	Пошук джерел даних	Лютий 2026 – Березень 2026	Виконано
5	Програмна реалізація методів	Березень 2026 – Квітень 2026	Виконано
6	Аналіз результатів роботи моделей	Квітень 2026 – Травень 2026	Виконано
7	Оформлення дипломної роботи	Травень 2026	Виконано

Студент \_\_\_\_\_ Микола ОРЛОВ

Керівник \_\_\_\_\_ Дмитро ЖЕЛЕЗНЯКОВ

## ЗМІСТ

Перелік умовних позначень, скорочень і термінів .....	6
ВСТУП .....	9
1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ТА ПОСТАНОВКА ЗАДАЧІ .....	11
1.1 Математична постановка задачі .....	11
1.1.1 Загальна постановка .....	11
1.2 Датасети .....	12
1.3 Класифікація методів розпізнавання харчових об'єктів .....	13
1.3.1 Методи на основі ознак (feature-based) .....	13
1.3.2 Методи на основі згорткових нейронних мереж (CNN).....	14
1.3.3 Методи на основі детекторів об'єктів .....	14
1.4 Методи оцінки об'єму, маси та калорійності страви .....	15
1.4.1 Методи оцінки маси .....	15
1.4.2 Розрахунок калорійності .....	16
1.5 Методи оцінки фізичної активності та енерговитрат .....	16
1.6 Метрики оцінки якості .....	17
1.7 Порівняльний аналіз методів .....	17
1.8 Невирішені проблеми та обмеження .....	18
1.9 Висновки до розділу .....	19
2 АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ .....	20
2.1 Опис структурної схеми програмної системи .....	20
2.1.1 Вхідні дані .....	21
2.1.2 Підсистема аналізу харчування .....	21
2.1.3 Підсистема фізичної активності .....	25
2.1.4 Підсумкові вихідні дані системи .....	25
2.2 Застосування ResNet у різних предметних областях .....	26
3 ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ ТА АНАЛІЗ .....	29
3.1 Умови проведення експериментів .....	29
3.2 Навчання ResNet-50 .....	29

3.3	Порівняльний аналіз архітектур .....	5 30
3.3.1	Криві навчання легковагових архітектур .....	30
3.3.2	Підсумкове порівняння метрик .....	31
3.4	Порівняння з результатами інших авторів.....	33
3.5	Обговорення результатів .....	33
	Висновки до розділу .....	34
4	Перспективна архітектура системи та можливі напрями розвитку .....	35
4.1	Порівняння реалізованої та перспективної архітектури .....	35
4.2	Напрями розвитку.....	36
	ВИСНОВКИ .....	38
	Перелік посилань.....	40

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

- **24HR** (24-Hour Dietary Recall) – 24-годинне харчове відкликання;
- **Adam** (Adaptive Moment Estimation) – адаптивний метод оцінки моментів;
- **AlexNet** – глибока згорткова нейронна мережа (Krizhevsky et al., 2012);
- **API** (Application Programming Interface) – інтерфейс програмування застосунків;
- **BatchNorm** (Batch Normalization) – пакетна нормалізація активацій;
- **CAPTURE-24** – датасет часових рядів носимих трекерів активності;
- **CNN** (Convolutional Neural Network) – згорткова нейронна мережа;
- **ColorJitter** – аугментація зображень зміною яскравості, контрасту та насиченості;
- **DenseNet** (Densely Connected Convolutional Network) – щільно з’єднана згорткова мережа;
- **DNN** (Deep Neural Network) – глибока нейронна мережа;
- **EfficientNet** – сімейство ефективних архітектур CNN з масштабуванням;
- **F1** (F1-score) – гармонічне середнє точності та повноти;
- **Faster R-CNN** (Faster Region-based Convolutional Neural Network) – детектор об’єктів із мережею регіональних пропозицій;
- **FC** (Fully Connected) – повнозв’язний шар нейронної мережі;
- **Food-101** – відкритий датасет зображень їжі (101 клас, 101 000 зображень);
- **GAP** (Global Average Pooling) – глобальний середній пулінг;
- **HOG** (Histogram of Oriented Gradients) – гістограма орієнтованих градієнтів;
- **ImageNet** – великомасштабний датасет зображень (понад 14 млн зображень);
- **IMU** (Inertial Measurement Unit) – інерційний вимірювальний блок

(акселерометр + гіроскоп);

- **LSTM** (Long Short-Term Memory) – мережа довгої короткочасної пам'яті;

- **MAE** (Mean Absolute Error) – середня абсолютна похибка;

- **mAP** (mean Average Precision) – середня середня точність;

- **MET** (Metabolic Equivalent of Task) – метаболічний еквівалент задачі;

- **MobileNet** – легковагова архітектура CNN для мобільних пристроїв;

- **MRE** (Mean Relative Error) – середня відносна похибка;

- **Nutrition5k** – датасет зображень страв із точними вагами та RGB-D даними;

- **PAMAP2** – датасет фізичної активності з IMU-датчиками (9 учасників, 18 класів);

- **ReLU** (Rectified Linear Unit) – випрямлена лінійна одиниця активації;

- **ResNet** (Residual Neural Network) – залишкова нейронна мережа;

- **RGB** (Red-Green-Blue) – колірна модель;

- **RGB-D** – зображення з кольоровим та глибинним каналами;

- **RMSE** (Root Mean Square Error) – середньоквадратична похибка;

- **RPN** (Region Proposal Network) – мережа регіональних пропозицій;

- **SIFT** (Scale-Invariant Feature Transform) – масштабно-інваріантне перетворення ознак;

- **SSD** (Single Shot MultiBox Detector) – детектор об'єктів з одним проходом;

- **SVM** (Support Vector Machine) – метод опорних векторів;

- **THFOOD-100** – датасет тайської кухні (100 класів, 53 459 зображень);

- **UCI-HAR** – датасет класифікації фізичної активності на основі смартфона;

- **UEC-Food256** – датасет японської кухні (256 класів, 31 395 зображень);

- **USDA** (United States Department of Agriculture) – Міністерство сільського господарства США (база нутрієнтів FoodData Central);

- **VGG** (Visual Geometry Group) – глибока згорткова мережа;

- **VIREO Food-172** – датасет китайської кухні (172 класи, 110 241

зображення);

– **ViT** (Vision Transformer) – трансформерна архітектура для класифікації

зображень;

– **YOLO** (You Only Look Once) – детектор об'єктів у реальному часі;

– **БЖВ** – білки, жири, вуглеводи;

– **НМ** – нейронна мережа;

– **ШІ** – штучний інтелект.

## ВСТУП

Розпізнавання їжі та автоматична оцінка калорійності є критично важливою проблемою в сучасній охороні здоров'я, дієтології та управлінні вагою. Згідно з даними Всесвітньої організації охорони здоров'я, близько 59% дорослого населення Європи страждає на ожиріння або надмірну вагу [32]. Традиційні методи оцінювання раціону харчування, зокрема 24-годинне відтворення (24HR) [22], значною мірою залежать від суб'єктивності та пам'яті респондента, що призводить до суттєвих похибок у розрахунках калорійності.

Стрімкий розвиток методів комп'ютерного зору та машинного навчання зумовлює зростання інтересу до автоматизації цього процесу. Проте оцінка калорійності страв на основі зображень є складною технічною задачею, що виходить за межі простого класифікування об'єктів. Вона передбачає коректне виявлення компонентів їжі в реальних умовах, де вхідні дані характеризуються високою варіативністю, неоднорідним освітленням, різним масштабом та частковим перекриттям об'єктів. На відміну від контрольованих лабораторних наборів даних, реальні фотографії не мають фіксованої структури, що ускладнює побудову універсального методу аналізу.

Актуальність дослідження зумовлена обмеженнями існуючих підходів. Більшість наявних рішень орієнтовані на вузькі набори страв або потребують ручного введення даних, тоді як алгоритмічна складова точного зчитування візуальної інформації часто залишається невирішеною. Використання методів глибокого навчання у поєднанні з даними про фізичну активність користувача дозволяє розглядати задачу в контексті обробки мультимодальних даних. У такому разі ключовим є забезпечення коректного та узгодженого збору інформації з різних джерел – зображень страв і показників спортивного навантаження, що безпосередньо впливає на достовірність підсумкових енергетичних показників.

Таким чином, наукова задача полягає у розробці та дослідженні методу

розпізнавання їжі з подальшою інтеграцією отриманих результатів у єдину систему моніторингу активності.

**Об'єкт дослідження** – процес автоматичної оцінки енергетичного балансу користувача на основі аналізу зображень страв і фізичної активності.

**Предмет дослідження** – методи комп'ютерного зору та глибокого навчання для розпізнавання харчових продуктів і оцінки їх калорійності.

**Мета дослідження** – розробити концептуальну модель та програмні засоби системи оцінки калорійності страв, у якій ключовим елементом є метод аналізу зображень для автоматичного розпізнавання продуктів, із подальшим врахуванням фізичної активності користувача для балансування енергетичних показників.

# 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ТА ПОСТАНОВКА ЗАДАЧІ

Автоматичне розпізнавання їжі та оцінка калорійності є відносно молодю, але сформованою науковою тематикою, що отримала значний імпульс із розвитком методів глибокого навчання [30, 16]. Сьогодні задача залишається активно досліджуваною, однак повністю вирішеною її вважати не можна: оцінка маси порцій за звичайною фотографією, розпізнавання змішаних страв та інтеграція аналізу харчування з трекером фізичної активності досі є відкритими задачами [30].

У цьому розділі систематизовано існуючі підходи до розпізнавання їжі, оцінки її об'єму та калорійності, а також оцінки фізичної активності. На основі аналізу сформульовано математичну постановку задачі та обґрунтовано вибір методів і датасетів для даної роботи.

## 1.1 Математична постановка задачі

### 1.1.1 Загальна постановка

Нехай користувач протягом проміжку часу  $T = [t_0, t_1]$  споживає набір страв та виконує фізичну активність. Потрібно побудувати систему, яка на основі зображення страви  $I \in \mathcal{I} \subset \mathbb{R}^{H \times W \times 3}$  та часових рядів фізичної активності  $A(t)$  забезпечує:

- оцінку калорійності кожної спожитої страви;
- сумарну калорійність раціону за період  $T$ ;
- оцінку енерговитрат унаслідок фізичної активності.

Формально задача полягає у знаходженні відображень:

$$F_c : I \rightarrow C, \quad F_e : A(t) \rightarrow E, \quad (1.1)$$

де  $C$  – оцінка калорійності страви (ккал),  $E$  – оцінка енерговитрат (ккал), таких, що похибка оцінювання мінімізується відносно еталонних значень.

## 1.2 Датасети

У таблиці 1.1 наведено основні публічні датасети, релевантні до задач даної роботи.

**Таблиця 1.1** – Основні публічні датасети для задач даної роботи

Датасет	Завдання	Розмір	Особливості
Food-101 [6]	Класифікація їжі	101 клас, 101 000 зобр.	Реальні умови, фіксоване розбиття
UEC-Food256 [17]	Класифікація їжі	256 класів, 31 395 зобр.	Японська кухня, bounding boxes
VIREO Food-172 [8]	Класифікація їжі	172 класи, 110 241 зобр.	Китайська кухня
THFOOD-100 [27]	Класифікація їжі	100 класів, 53 459 зобр.	Тайська кухня, CLR-бенчмарк
Nutrition5k [26]	Оцінка маси та калорій	≈5 000 страв	RGB-D, точні ваги інгредієнтів
UCI-HAR [3]	Класифікація активності	30 уч., 6 класів	Смартфон (акселерометр + гіроскоп)
PAMAP2 [23]	Класифікація активності	9 уч., 18 класів	IMU на зап'ясті, грудях, щиколотці
CAPTURE-24 [7]	Класифікація активності	151 уч., 3883 год.	Носимий трекер, вільне середовище

Огляд публічних датасетів для розпізнавання їжі [30, 27] показує, що більшість з них орієнтовані на конкретні національні кухні та суттєво різняться за розміром і якістю анотацій. Зокрема, Food-101 охоплює переважно західну кухню, тоді як VIREO Food-172 та UEC-Food256 зосереджені на китайській та японській відповідно. Для задач оцінки маси та калорійності виокремлюється Nutrition5k [26] завдяки наявності RGB-D зображень та точних ваг інгредієнтів.

Для задач оцінки фізичної активності стандартними лабораторними бенчмарками є UCI-HAR [3] та PAMAP2 [23], однак моделі, навчені на таких контрольованих даних, погано узагальнюються датасет зібрано у вільному середовищі від 151 учасника і є на два-три порядки більшим за інші публічно доступні датасети активності, що робить його зручним джерелом для зіставлення типів активності з MET-коефіцієнтами.

Для навчання моделі класифікації страв у цій роботі обрано **Food-101**. Вибір обумовлений такими факторами:

- **масштаб і збалансованість:** рівномірний розподіл зображень між класами забезпечує стабільне навчання без додаткового балансування;
- **реалістичність даних:** зображення отримані в реальних умовах, що підвищує практичну застосовність моделі;
- **відтворюваність:** фіксоване розбиття на train/test дозволяє коректно порівнювати результати з літературою [15, 19];
- **стандартний бенчмарк:** Food-101 уможливорює об'єктивне порівняння із існуючими рішеннями.

### 1.3 Класифікація методів розпізнавання харчових об'єктів

Задача автоматичного розпізнавання їжі за зображенням розглядається на двох рівнях: *класифікація* (до якого класу належить зображення) та *детекція* (локалізація об'єктів на зображенні). Аналіз літератури [15, 30, 18] дозволяє виокремити три основні групи підходів.

#### 1.3.1 Методи на основі ознак (feature-based)

Ранні роботи спиралися на ручне виділення ознак: колірні гістограми, текстурні дескриптори (SIFT, HOG), форму об'єктів. Ці ознаки подавалися на класифікатори SVM або Random Forest. Зокрема, у роботі [6] для класифікації датасету Food-101 використовувалися дискримінативні компоненти, виявлені методом Random Forests.

До переваг цього підходу належать інтерпретованість результатів та низькі обчислювальні вимоги. Проте методи на основі ознак характеризуються високою чутливістю до умов зйомки, необхідністю ручного проектування ознак і низькою точністю на великих наборах класів.

### 1.3.2 Методи на основі згорткових нейронних мереж (CNN)

Поява глибоких CNN (AlexNet, VGG, ResNet, DenseNet) кардинально змінила підходи до класифікації зображень їжі. Систематичний огляд [15] показує, що архітектури на основі ResNet та DenseNet [12] досягають точності понад 85% на Food-101 і є найефективнішими для задач класифікації страв та оцінки поживної цінності. Трансферне навчання на основі ImageNet-моделей є стандартним підходом для харчових датасетів [30, 34].

Окремим напрямом є CNN-архітектури для детекції об'єктів, зокрема Faster R-CNN [24], які поєднують згорткову мережу з механізмом регіональних пропозицій (Region Proposal Network) для локалізації об'єктів.

Lee [19] запропонували мультиспектральну класифікацію: використання додаткових спектральних каналів поза видимим діапазоном дає інформацію про склад продуктів, недоступну при стандартній RGB-зйомці.

CNN-методи вирізняються високою точністю, автоматичним навчанням ознак та доброю узагальнюваністю. Разом з тим вони потребують великих обсягів навчальних даних і є обчислювально затратними.

### 1.3.3 Методи на основі детекторів об'єктів

Для сценаріїв із кількома стравами на зображенні застосовуються одноетапні архітектури детекції YOLO та SSD. Kumar та Prem Kumar [18] навчили кастомну модель YOLOv8 для детекції харчових об'єктів, продемонструвавши mAP понад 90% на власному датасеті. Метод дозволяє одночасно локалізувати й класифікувати кілька страв за один прохід.

Серед переваг варто відзначити можливість розпізнавання кількох об'єктів на зображенні одночасно, а також локалізацію та класифікацію в одному проході. Водночас метод потребує складніших анотацій (bounding boxes) та висуває вищі обчислювальні вимоги.

## 1.4 Методи оцінки об'єму, маси та калорійності страви

### 1.4.1 Методи оцінки маси

Після детекції страви виникає задача оцінки її маси для подальшого розрахунку калорійності. Залежно від обладнання та способу оцінки геометрії страви, методи оцінки маси поділяють на три групи [30, 2].

**На основі площі сегментації.** Найпростіший підхід – моделювання маси страви як степеневої функції від площі сегментованого об'єкта:

$$M = \alpha \cdot A^\beta, \quad (1.2)$$

де  $A$  – площа сегментованого зображення в пікселях,  $\alpha, \beta$  – коефіцієнти, що визначаються емпірично для кожного класу їжі [30]. Перевагою методу є простота реалізації, проте він має суттєвий недолік – високу похибку через відсутність інформації про глибину та невідомий масштаб.

**На основі 3D-реконструкції та глибинних зображень.** Точніші оцінки дають підходи, що використовують RGB-D камери або оцінку глибини за одним RGB-зображенням. Датасет Nutrition5k [26] дозволяє досліджувати методи тривимірної реконструкції порцій. Маса визначається через добуток реконструйованого об'єму  $V$  та середньої густини продукту  $\rho$ :

$$M = \rho V. \quad (1.3)$$

**Методи one-shot та few-shot оцінки об'єму.** AlMughrabi та ін. [2] запропонували метод VolETA – підхід до оцінки об'єму страви на основі одного або кількох RGBD-зображень без переналаштування моделі на нові класи. Метод включає вибір ключових кадрів, сегментацію референсного об'єкта за допомогою XMem++ [5], оцінку положення камери через PixSfM [20] та тривимірну реконструкцію сітки за допомогою NeuS2 [31]. Масштабний коефіцієнт визначається відносно референсного об'єкта

відомого розміру, після чого об'єм страви обчислюється як:

$$V = f(b, s, \theta), \quad (1.4)$$

де  $b$  – параметри обмежувальної рамки,  $s$  – масштабний коефіцієнт,  $\theta$  – параметри нейронної мережі. Цей підхід є перспективним для нових або рідкісних страв, оскільки не потребує повторного навчання.

### 1.4.2 Розрахунок калорійності

Калорійність страви  $C$  обчислюється на основі питомої калорійності  $c_{100}$  на 100 г продукту:

$$C = \frac{M \cdot c_{100}}{100}. \quad (1.5)$$

Для підходів, де CNN безпосередньо регресує калорійність за зображенням, потрібні датасети з точними мітками (як Nutrition5k [26]), однак такі моделі схильні до помилок при зустрічі з невідомими стравами [30, 19]. Найкращі результати досягаються при поєднанні сегментації/детекції, оцінки глибини та довідкових баз нутрієнтів [30].

Задача оцінки калорійності формулюється як мінімізація середньоквадратичної похибки:

$$\min_{\theta} \mathbb{E}[(C - C^*)^2], \quad (1.6)$$

де  $C^*$  – еталонне значення калорійності,  $\theta$  – параметри моделей оцінки маси та геометричних характеристик.

## 1.5 Методи оцінки фізичної активності та енерговитрат

Носимі пристрої (акселерометри, гіроскопи, пульсоксиметри) є стандартним засобом збору даних фізичної активності [10]. Часові ряди сенсорних даних класифікуються методами машинного навчання (SVM,

Random Forest) або глибокими нейронними мережами (LSTM, CNN-LSTM) для визначення типу активності. Після класифікації енерговитрати розраховуються за коефіцієнтами метаболічного еквіваленту (MET) [1] або безпосередньо регресуються нейронною мережею.

Часовий ряд активності подається як вектор-функція:

$$A(t) = (a_1(t), a_2(t), \dots, a_m(t)). \quad (1.7)$$

Енерговитрати  $E$  за період  $T = [t_0, t_1]$  визначаються інтегруванням функції активності:

$$E = \int_{t_0}^{t_1} g(A(t)) dt, \quad (1.8)$$

де  $g(\cdot)$  апроксимується нейронною мережею або регресійною моделлю.

## 1.6 Метрики оцінки якості

Для кожної підзадачі системи використовуються окремі метрики:

- **Top-1 / Top-5 Accuracy** – для класифікації страв;
- **mAP** (mean Average Precision) – для детекції страв;
- **MAE** та **RMSE** – для оцінки маси та калорійності;
- **MRE** (Mean Relative Error) – середня відносна похибка оцінки енерговитрат та добової калорійності;
- **Accuracy, F1-score** – для класифікації типів фізичної активності.

Ключовою узагальненою метрикою є **MRE** оцінки добової калорійності та енерговитрат, оскільки вона безпосередньо відображає практичну значущість похибки.

## 1.7 Порівняльний аналіз методів

У таблиці 1.2 узагальнено методи розпізнавання харчових об'єктів, у таблиці 1.3 – методи оцінки маси.

**Таблиця 1.2** – Порівняння методів розпізнавання їжі

Підхід	Top-1 (Food-101)	Вимоги до даних	Кілька об'єктів
Feature-based (SVM/RF)	≈50–65%	Помірні	Ні
CNN (ResNet, DenseNet)	≈85–92%	Великі	Ні
Object detection (YOLO)	mAP > 90%	Великі + BBox	Так

**Таблиця 1.3** – Порівняння методів оцінки маси страви

Підхід	Точність	Обладнання	Складність
Площа сегментації	≈50–70% (похибка 30–50%)	Стандартна камера	Низька
RGB-D / 3D-реконструкція	Висока (<10% похибки)	Глибинна камера	Висока
One-shot (VolETA)	Середня	Стандартна камера	Середня

## 1.8 Невирішені проблеми та обмеження

Аналіз літератури виявляє низку актуальних відкритих проблем:

1) **оцінка маси без інформації про глибину.** Більшість мобільних застосунків не мають доступу до RGB-D камери, тому оцінка маси за площею сегментації має велику похибку. Методи one-shot оцінки об'єму [2] обіцяють вирішення проблеми, але ще не мають широкого практичного впровадження;

2) **змішані страви та складні рецепти.** Класифікатори погано справляються зі стравами, що є поєднанням кількох компонентів (наприклад, салат із різнорідних інгредієнтів). Точна оцінка калорійності таких страв залишається відкритою задачею [30];

3) **варіативність між порціями.** Калорійність однієї й тієї ж страви може суттєво різнитися залежно від рецепту та розміру порції. Підходи, що спираються лише на клас страви, не враховують цієї варіативності [30];

4) **інтеграція оцінки їжі та фізичної активності.** Переважна більшість робіт розглядає ці задачі окремо. Комплексних систем, що об'єднують обидва компоненти, у відкритій літературі представлено мало [14];

5) **відсутність персоналізації.** Базові метаболічні показники та коефіцієнти засвоюваності різняться між людьми, тоді як більшість систем використовують усереднені значення.

## 1.9 Висновки до розділу

Аналіз літератури показує, що задача автоматичної оцінки калорійності страв та фізичної активності є активно досліджуваною, але ще не повністю вирішеною.

Методи на основі CNN та трансферного навчання демонструють найвищу точність розпізнавання харчових об'єктів і є стандартом у галузі [15, 30, 19]. Для оцінки маси найбільш перспективним є поєднання сегментації та one-shot оцінки об'єму [2]. Розрахунок калорійності через бази даних нутрієнтів залишається найбільш практичним підходом [11]. Для оцінки фізичної активності носимі трекери разом із моделями машинного навчання забезпечують достатню точність для побутового використання [7, 10].

На основі виконаного аналізу для даної роботи обрано такий підхід:

- **класифікація страви** – CNN з трансферним навчанням на датасеті Food-101;
- **оцінка маси** – сегментація та масштабний коефіцієнт (VolETA);
- **розрахунок калорійності** – база даних нутрієнтів;
- **оцінка енерговитрат** – дані акселерометра з датасету CAPTURE-24.

Така комбінація методів охоплює повний цикл оцінки добового балансу енергії без необхідності спеціалізованого обладнання.

## 2 АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ

Загальну структуру розробленої програмної системи наведено на рисунку 2.1.

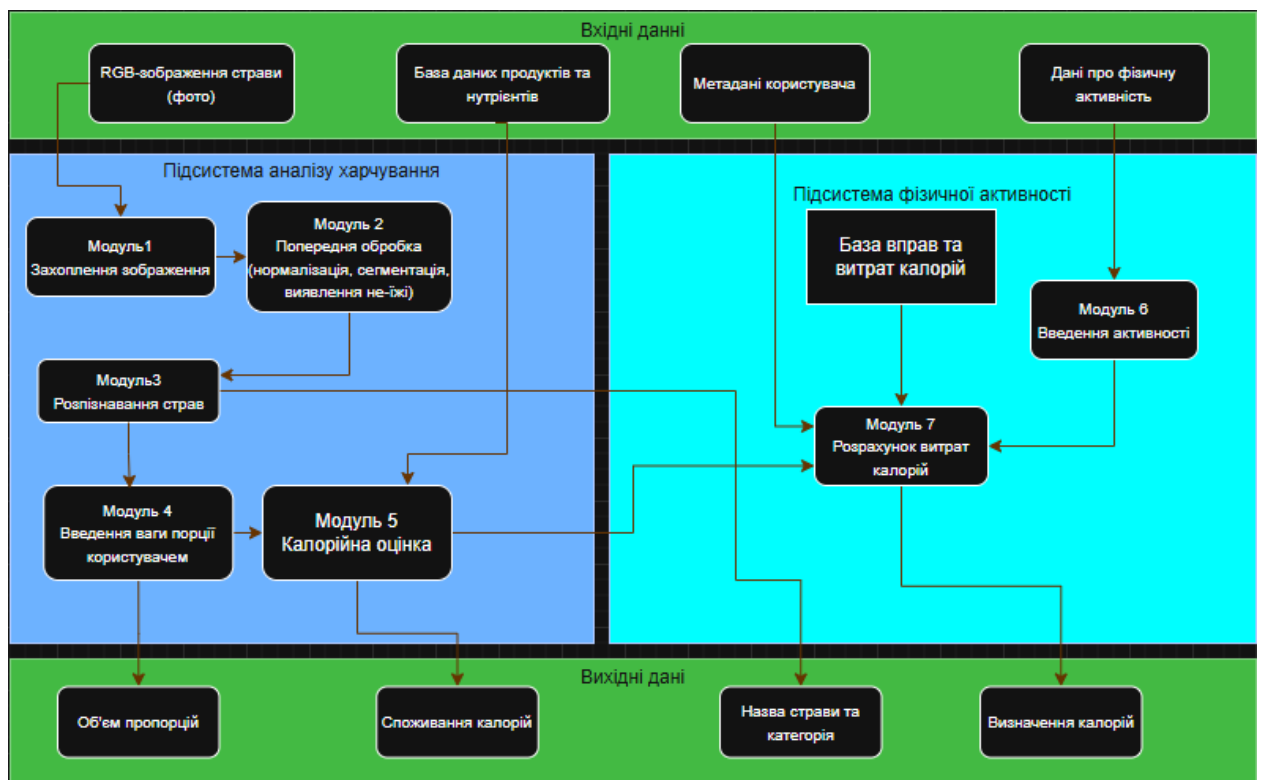


Рисунок 2.1 – Структурна схема програмної системи

Детальний опис кожного структурного елемента наведено в наступних підрозділах.

### 2.1 Опис структурної схеми програмної системи

Представлена структурна схема відображає архітектуру програмної системи для аналізу харчування на основі зображень та розрахунку витрат калорій користувача. Система складається з двох головних підсистем: аналізу харчування та фізичної активності, а також використовує набір вхідних і вихідних даних.

Зведену характеристику модулів системи наведено в таблиці 2.1.

**Таблиця 2.1** – Вхідні та вихідні дані модулів системи

Модуль	Назва	Вхід	Вихід
1	Захоплення зображення	Файл з галереї	Тензор $I \in \mathbb{R}^{H \times W \times 3}$
2	Попередня обробка	Тензор $I$	Нормалізований тензор $224 \times 224$
3	Розпізнавання страви	Оброблений тензор	Клас $\hat{y}$ , впевненість
4	Оцінка маси порції	Зображення, клас $\hat{y}$	Маса $M$ (г)
5	Калорійна оцінка	Клас $\hat{y}$ , маса $M$	Калорійність $C$ (ккал), БЖВ
6	Введення активності	Дані користувача	Тип активності, час $t$
7	Розрахунок витрат	Тип активності, $w, t$	Енерговитрати $E$ (ккал)

Наведені модулі утворюють послідовний конвеєр обробки даних, де вихід кожного модуля є вхідними даними для наступного.

### 2.1.1 Вхідні дані

Система приймає чотири типи вхідних даних:

- **RGB-зображення страви** – тензор  $I \in \mathbb{R}^{H \times W \times 3}$ , отриманий із камери або галереї пристрою;
- **Метадані користувача** – вік  $a$ , маса тіла  $w$  (кг), зріст  $h$  (см), стать  $s$ , що використовуються для розрахунку базового метаболізму;
- **База продуктів та нутрієнтів** – реляційна таблиця із питомою калорійністю  $c_{100}$  (ккал/100 г) та БЖВ для кожного класу їжі;
- **Дані про фізичну активність** – тип вправи, інтенсивність та тривалість, що вводяться користувачем вручну.

### 2.1.2 Підсистема аналізу харчування

**Модуль 1 – Захоплення зображення.** Забезпечує отримання фотографії страви завантаженням з галереї пристрою. Вихідний сигнал – необроблений тензор  $I \in \mathbb{R}^{H \times W \times 3}$ .

**Модуль 2 – Попередня обробка зображення.** Зображення приводиться до стандартного розміру  $224 \times 224$  пікселів, після чого виконується

нормалізація каналів за статистиками ImageNet:

$$\hat{I}_c = \frac{I_c - \mu_c}{\sigma_c}, \quad c \in \{R, G, B\}, \quad (2.1)$$

де  $\mu_c$  та  $\sigma_c$  – середнє та стандартне відхилення каналу  $c$  на датасеті ImageNet ( $\mu = [0.485, 0.456, 0.406]$ ,  $\sigma = [0.229, 0.224, 0.225]$ ). Додатково застосовується аугментація: випадкове горизонтальне відображення, кольорове відхилення (ColorJitter) та RandomResizedCrop для підвищення стійкості моделі до варіацій освітлення та ракурсу.

**Модуль 3 – Розпізнавання страв (ResNet-50).** Класифікація виконується згортковою нейронною мережею ResNet-50 [12], попередньо навченою на ImageNet та дотренованою на датасеті Food-101. Архітектуру мережі наведено в таблиці 2.2.

**Таблиця 2.2** – Архітектура ResNet-50 [12]

Шар	Розмір виходу	ResNet-50			
conv1	112 × 112	7 × 7, 64, stride 2			
pool	56 × 56	3 × 3 max pool, stride 2			
conv2_x	56 × 56	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>1 × 1, 64</td></tr> <tr><td>3 × 3, 64</td></tr> <tr><td>1 × 1, 256</td></tr> </table> × 3	1 × 1, 64	3 × 3, 64	1 × 1, 256
1 × 1, 64					
3 × 3, 64					
1 × 1, 256					
conv3_x	28 × 28	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>1 × 1, 128</td></tr> <tr><td>3 × 3, 128</td></tr> <tr><td>1 × 1, 512</td></tr> </table> × 4	1 × 1, 128	3 × 3, 128	1 × 1, 512
1 × 1, 128					
3 × 3, 128					
1 × 1, 512					
conv4_x	14 × 14	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>1 × 1, 256</td></tr> <tr><td>3 × 3, 256</td></tr> <tr><td>1 × 1, 1024</td></tr> </table> × 6	1 × 1, 256	3 × 3, 256	1 × 1, 1024
1 × 1, 256					
3 × 3, 256					
1 × 1, 1024					
conv5_x	7 × 7	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>1 × 1, 512</td></tr> <tr><td>3 × 3, 512</td></tr> <tr><td>1 × 1, 2048</td></tr> </table> × 3	1 × 1, 512	3 × 3, 512	1 × 1, 2048
1 × 1, 512					
3 × 3, 512					
1 × 1, 2048					
gap	1 × 1	Global Average Pooling			
fc	1 × 101	FC 101, Softmax			

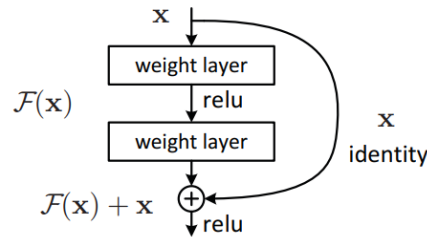
*Примітка:* conv1 – початковий згортковий шар; pool – шар максимального пулінгу; conv2\_x–conv5\_x – групи bottleneck-блоків, де кожен блок складається з трьох послідовних згорток (1 × 1, 3 × 3, 1 × 1), а множник (×N) вказує кількість повторень блоку в групі; gap – глобальний середній пулінг; fc – повнозв'язний класифікаційний шар на 101 клас (Food-101).

Ключовою особливістю ResNet-50 є використання bottleneck-блоків, що складаються з трьох згорткових шарів (1 × 1, 3 × 3, 1 × 1) із залишковим з'єднанням, яке дозволяє уникнути проблеми затухання градієнтів [12].

Кожен залишковий блок реалізує відображення (рисунок 2.2):

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}, \quad (2.2)$$

де  $\mathbf{x}$  – вхідний вектор ознак,  $\mathcal{F}(\cdot)$  – залишкове відображення (два або три шари BatchNorm–ReLU–Conv),  $W_i$  – ваги шарів.



**Рисунок 2.2** – Залишковий блок (residual block) [12]

Вихідний клас страви визначається як:

$$\hat{y} = \arg \max_k \text{softmax}(W_{\text{fc}} \cdot \mathbf{z}), \quad (2.3)$$

де  $\mathbf{z}$  – вектор ознак після Global Average Pooling,  $W_{\text{fc}} \in \mathbb{R}^{101 \times 2048}$  – матриця ваг класифікаційного шару.

Навчання виконується методом fine-tuning: ваги перших трьох груп блоків заморожуються, четверта група та класифікаційний шар навчаються з нуля з використанням оптимізатора Adam та функції втрат крос-ентропії:

$$\mathcal{L}_{\text{cls}} = - \sum_{k=1}^{101} y_k \log \hat{p}_k, \quad (2.4)$$

де  $y_k$  – одно-гаряче кодування істинного класу,  $\hat{p}_k$  – передбачена ймовірність класу  $k$ .

**Модуль 4 – Оцінка об’єму порцій.** Маса страви  $M$  оцінюється через площу сегментованої області на зображенні. Після бінарної сегментації харчового об’єкта обчислюється площа  $A$  (у пікселях), яка перетворюється на

масу за степенною функцією [6]:

$$M = \alpha \cdot A^\beta, \quad (2.5)$$

де  $\alpha$  та  $\beta$  – емпіричні коефіцієнти, специфічні для кожного класу їжі та визначені на датасеті Food-101. Масштабування пікселів у метричні одиниці виконується через референсний об’єкт або стандартний розмір тарілки.

**Модуль 5 – Калорійна оцінка.** На основі визначеного класу страви  $\hat{y}$  та оціненої маси  $M$  калорійність розраховується через питому калорійність із бази нутрієнтів:

$$C = \frac{c_{100}(\hat{y}) \cdot M}{100}, \quad (2.6)$$

де  $c_{100}(\hat{y})$  – калорійність 100 г страви класу  $\hat{y}$  (ккал/100 г). Аналогічно визначаються вміст білків, жирів та вуглеводів.

База нутрієнтів системи містить питому калорійність для всіх 101 класу датасету Food-101 і сформована на основі даних USDA FoodData Central [29]. Діапазон значень калорійності – від 40 ккал/100 г (місо-суп) до 462 ккал/100 г (фуа-гра). Приклад записів наведено в таблиці 2.3.

**Таблиця 2.3** – Приклад записів бази нутрієнтів

Клас страви	Калорійність (ккал/100 г)
apple_pie	237
caesar_salad	94
pizza	266
sushi	145
donuts	452

База фізичної активності містить 28 видів вправ із MET-коефіцієнтами згідно з Compendium of Physical Activities [1]. Діапазон значень MET – від 2.5 (йога) до 16.0 (спринт).

### 2.1.3 Підсистема фізичної активності

**Модуль 6 – Введення активності.** Користувач вибирає тип фізичної вправи зі списку, вказує інтенсивність та тривалість  $t$  (хв). Дані передаються до Модуля 7 разом із масою тіла користувача  $w$  з метаданих.

**Модуль 7 – Розрахунок витрат калорій.** Витрати енергії  $E$  розраховуються за формулою на основі метаболічного еквіваленту (MET) [10]:

$$E = \text{MET} \times w \times \frac{t}{60}, \quad (2.7)$$

де MET – коефіцієнт метаболічного еквіваленту для обраного типу активності (наприклад, ходьба – MET = 3.5, біг – MET = 8.0),  $w$  – маса тіла користувача (кг),  $t$  – тривалість активності (хв). Значення MET беруться з бази вправ системи.

### 2.1.4 Підсумкові вихідні дані системи

Після проходження всіх модулів система обчислює баланс калорій:

$$\Delta = \sum_i C_i - \sum_j E_j, \quad (2.8)$$

де  $C_i$  – калорійність  $i$ -ї спожитої страви,  $E_j$  – витрати енергії під час  $j$ -ї фізичної активності. Значення  $\Delta > 0$  відповідає профіциту калорій,  $\Delta < 0$  – дефіциту.

Підсумкові вихідні дані системи:

- загальна калорійність спожитої їжі  $\sum C_i$ ;
- витрати калорій за фізичну активність  $\sum E_j$ ;
- баланс калорій  $\Delta$ ;
- деталізована інформація про кожну страву (назва, БЖВ, порція).

## 2.2 Застосування ResNet у різних предметних областях

Архітектура ResNet демонструє високу узагальнювальну здатність і успішно використовується в широкому спектрі задач комп'ютерного зору, що підтверджує обґрунтованість її вибору для класифікації страв у даній роботі. У всіх цих областях є потреба у надійному виділенні ієрархічних візуальних ознак із зображень із різноманітними текстурами, кольорами та формами – саме цю задачу, залишкові з'єднання вирішують особливо ефективно.

У таблиці 2.4 наведено приклади застосування ResNet у різних предметних областях із зазначенням типових задач та досягнутих результатів.

Таблиця 2.4 – Застосування ResNet у різних предметних областях

Область	Задача	Результат / Особливість
Медична діагностика	Виявлення патологій на рентгенівських знімках грудної клітки (пневмонія, туберкульоз)	ResNet-50 із transfer learning забезпечує AUC до 0.87 на датасеті ChestX-ray14; ознаки ImageNet ефективно переносяться на медичні знімки [4]
Дерматологія	Класифікація захворювань шкіри за фотографіями (меланома, базально-клітинна карцинома)	Ансамбль моделей на базі ResNet-50, навчений на датасеті HAM10000, досягає точності Top-1 93% та Top-3 99% [28]
Промисловий контроль якості	Виявлення дефектів поверхні (подряпини, тріщини) на конвеєрних лініях у виробництві напівпровідників та сталі	ResNet є домінуючою базовою архітектурою у задачах автоматичної візуальної інспекції завдяки стійкому виділенню текстурних ознак [33]
Автономний транспорт	Розпізнавання дорожніх знаків та пішоходів у системах допомоги водієві	ResNet-50 досягає точності 99.8% на стандартних бенчмарках розпізнавання дорожніх знаків; стійкість до варіацій освітлення [25]
Дистанційне зондування	Класифікація типів землекористування на супутникових знімках Sentinel-2	ResNet-50 досягає F1-score 0.924 на багатокласовій задачі класифікації ділянок Землі, перевищуючи традиційні методи [21]
Аналіз харчування (дана робота)	Класифікація страв за фотографією для оцінки калорійності	Fine-tuning на Food-101 (101 клас); узагальнені ознаки текстур та кольорів страв [12]

Таким чином, ResNet-50 є перевіреною та універсальною основою для задач класифікації зображень у прикладних системах. Залишкові з'єднання дозволяють мережі зберігати низькорівневі просторові деталі (краї, текстури, колір) на всіх рівнях представлення, що є особливо важливим для страв із схожими формами, але різними текстурами поверхні. Поєднання цієї архітектурної переваги з технікою transfer learning від ImageNet забезпечує ефективне навчання навіть при обмежених обчислювальних ресурсах.

## Висновки до розділу

У розділі описано архітектуру програмної системи для аналізу харчування на основі зображень. Система складається з семи модулів, об'єднаних у дві підсистеми: аналізу харчування та фізичної активності.

Підсистема аналізу харчування реалізує повний процес обробки зображень: від захоплення фотографії страви, її стандартизації та аугментації – до класифікації страви засобами ResNet-50 з fine-tuning, оцінки маси порції та розрахунку нутрієнтного складу. Підсистема фізичної активності забезпечує введення даних про вправи та розрахунок енерговитрат на основі метаболічних еквівалентів, що є стандартизованим підходом у спортивній медицині [1].

Оцінка маси порції базується на аналізі площі сегментованої області зображення з перерахунком через референсний об'єкт. Калорійність та вміст білків, жирів і вуглеводів визначаються за базою нутрієнтів, сформованою відповідно до даних USDA FoodData Central [29] для всіх 101 класу страв.

Запропонована архітектура забезпечує повний цикл обробки даних – від захоплення зображення страви до розрахунку балансу між спожитими калоріями та енерговитратами на фізичну активність – та є основою для реалізації.

## 3 ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ ТА АНАЛІЗ

У цьому розділі представлено результати навчання та порівняльного аналізу архітектур CNN на датасеті Food-101, обговорено отримані метрики та проведено зіставлення з результатами інших авторів.

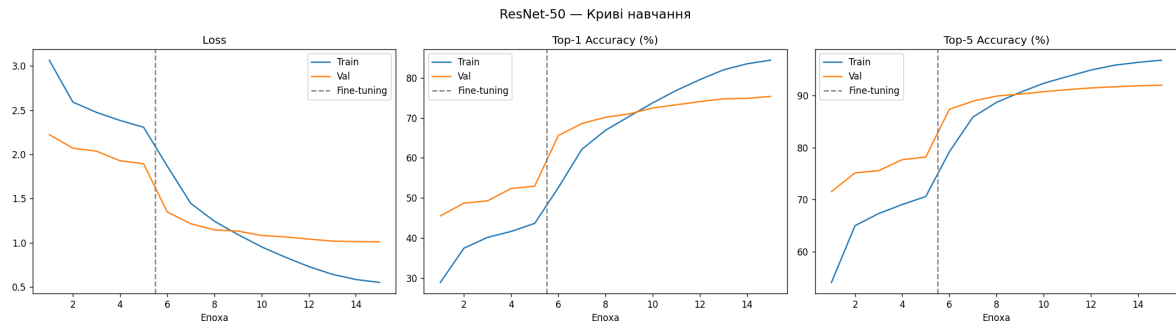
### 3.1 Умови проведення експериментів

Усі моделі навчались в однакових умовах для забезпечення коректного порівняння:

- датасет: Food-101 (75 000 тренувальних / 25 000 тестових зображень, 101 клас);
- розмір вхідного зображення:  $224 \times 224$  пікселів;
- кількість епох: 15 (епохи 1–5 – заморожені ваги backbone, епохи 6–15 – fine-tuning);
- оптимізатор: Adam; функція втрат: крос-ентропія;
- аугментація: RandomResizedCrop, горизонтальне відображення, ColorJitter, нормалізація за статистиками ImageNet.

### 3.2 Навчання ResNet-50

Динаміка навчання ResNet-50 представлена на рис. 3.1. На графіку чітко виділяються дві фази, розділені пунктирною лінією (епоха 5): фаза feature extraction (заморожений backbone) та фаза fine-tuning.



**Рисунок 3.1** – Криві навчання ResNet-50: Loss, Top-1 та Top-5 Accuracy

На фазі feature extraction (епохи 1–5) валідаційна функція втрат швидко знижується з  $\approx 2.2$  до  $\approx 1.3$ , а Top-1 Accuracy зростає з  $\approx 45\%$  до  $\approx 60\%$ . Після увімкнення fine-tuning (епоха 6) спостерігається різкий стрибок точності: вже на епосі 6 валідаційна Top-1 Accuracy досягає  $\approx 65\%$ , а до епохи 15 стабілізується на рівні  $\approx 76\%$ . Тренувальна точність продовжує зростати до  $\approx 83\%$ , що свідчить про помірне перенавчання у пізніх епохах, контрольоване регуляризацією.

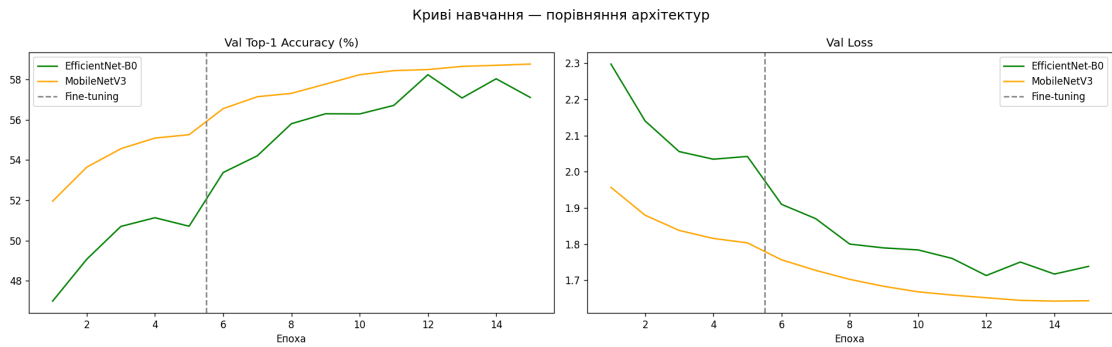
Фінальні результати ResNet-50 на тестовій вибірці:

- **Top-1 Accuracy: 80.66%**
- **Top-5 Accuracy: 95.22%**

### 3.3 Порівняльний аналіз архітектур

#### 3.3.1 Криві навчання легковагових архітектур

На рис. 3.2 представлено криві валідаційної точності та втрат для EfficientNet-B0 і MobileNetV3 в однакових умовах навчання.

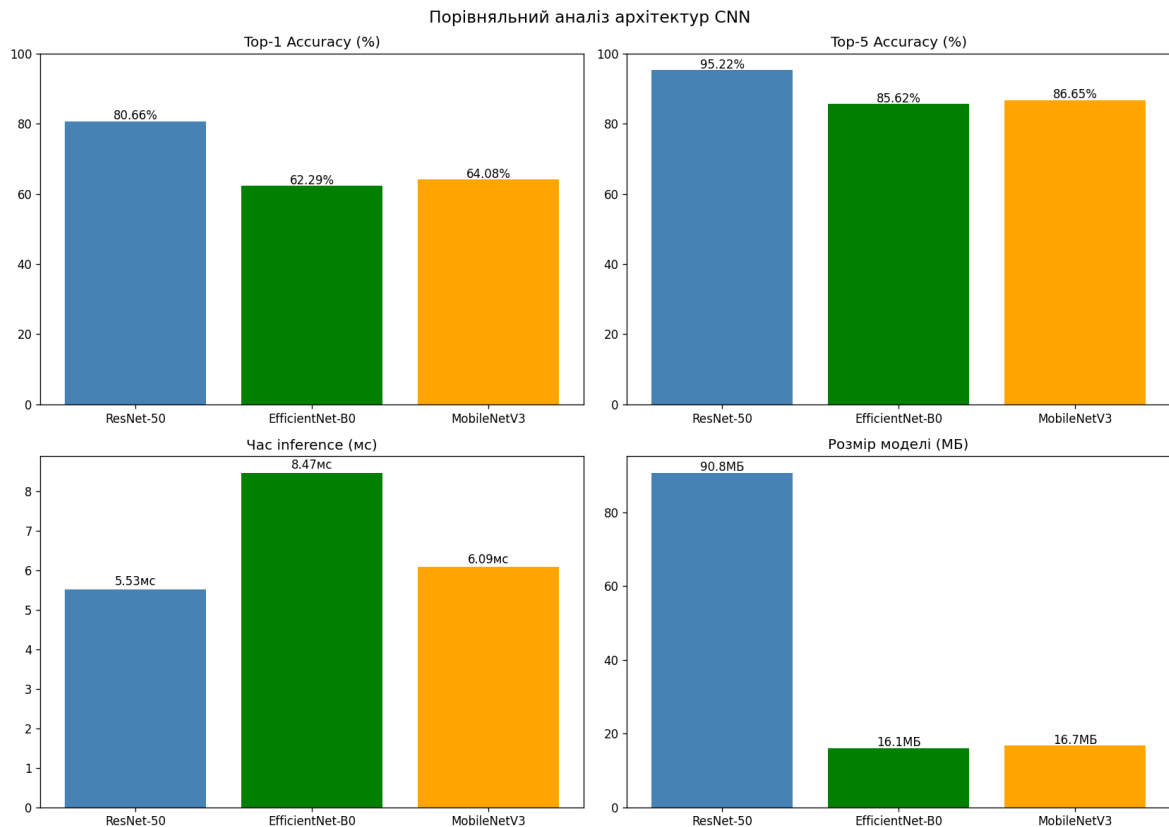


**Рисунок 3.2** – Криві навчання EfficientNet-B0 та MobileNetV3: Val Top-1 Accuracy і Val Loss

Обидві архітектури демонструють стабільне зниження втрат і зростання точності протягом усіх 15 епох. MobileNetV3 випереджає EfficientNet-B0 на всіх епохах за Top-1 Accuracy, досягаючи  $\approx 59\%$  проти  $\approx 57\%$  наприкінці навчання. Водночас EfficientNet-B0 проявляє нестабільність після fine-tuning (помітні коливання точності на епохах 12–15), тоді як MobileNetV3 збігається рівномірніше. За функцією втрат MobileNetV3 також показує нижчі значення ( $\approx 1.64$  проти  $\approx 1.74$  на епосі 15).

### 3.3.2 Підсумкове порівняння метрик

Зведені результати всіх трьох архітектур наведено на рис. 3.3 та у таблиці 3.1.



**Рисунок 3.3** – Порівняльний аналіз архітектур CNN: Top-1, Top-5 Accuracy, час inference та розмір моделі

**Таблиця 3.1** – Порівняння архітектур CNN на датасеті Food-101

Архітектура	Top-1 (%)	Top-5 (%)	Inference (мс)	Розмір (МБ)
ResNet-50	<b>80.66</b>	<b>95.22</b>	5.53	90.8
EfficientNet-B0	62.29	85.62	8.47	16.1
MobileNetV3	64.08	86.65	6.09	16.7

ResNet-50 перевершує легковагові архітектури за точністю: перевага над MobileNetV3 складає **+16.58 %** за Top-1 та **+8.57 %** за Top-5. Водночас ResNet-50 має суттєво більший розмір моделі (90.8 МБ проти  $\approx 16$  МБ) при найшвидшому часі inference (5.53 мс).

EfficientNet-B0 показує найповільніший inference (8.47 мс) і нижчу точність порівняно з MobileNetV3, що робить його найменш привабливим

варіантом серед розглянутих за обома критеріями.

### 3.4 Порівняння з результатами інших авторів

У таблиці 3.2 наведено зіставлення отриманих результатів з опублікованими роботами на датасеті Food-101.

**Таблиця 3.2** – Порівняння результатів класифікації на Food-101 з літературою

Робота	Метод	Top-1 (%)	Top-5 (%)
Bossard et al. [6]	Random Forest + SIFT	50.76	—
Kaur et al. [15]	CNN (fine-tuning)	85.47	—
Lee et al. [19]	DNN мультиспектральний	87.10	—
Kaushal et al. [16]	Transfer learning CNN	88.30	—
<b>Дана робота (ResNet-50)</b>	Transfer learning, fine-tuning	<b>80.66</b>	<b>95.22</b>
<b>Дана робота (MobileNetV3)</b>	Transfer learning, fine-tuning	64.08	86.65

Результат ResNet-50 (80.66%) є конкурентоспроможним відносно базового методу [6] (50.76%) та підходу на основі CNN без fine-tuning, однак поступається спеціалізованим архітектурам з глибшим fine-tuning [15, 19]. Різниця пояснюється обчислювальними обмеженнями наявного обладнання, через що навчання було виконано на 15 епохах, а також відсутністю додаткових технік регуляризації (label smoothing, mixup), що є напрямом подальшого вдосконалення.

### 3.5 Обговорення результатів

Отримані результати дозволяють зробити такі спостереження.

**Вплив fine-tuning.** Графіки навчання (рис. 3.1) підтверджують, що розморожування ваг backbone на епосі 6 дає значний приріст точності для

ResNet-50 ( $\approx +15\%$  за одну епоху). Для легковагових архітектур цей ефект виражений слабше, що свідчить про меншу ємність їхніх ознак відносно 101-класової задачі.

**Компроміс точність / розмір.** Для серверного розгортання ResNet-50 є оптимальним вибором серед розглянутих архітектур завдяки найвищій точності та найшвидшому inference. Архітектури з потенційно вищою точністю, зокрема DenseNet [13] та Vision Transformer [9], не розглядались у даній роботі через обчислювальні обмеження, однак є перспективними напрямками для подальшого дослідження.

Серед розглянутих легковагових архітектур MobileNetV3 забезпечує найкращий баланс між точністю (64.08% Top-1) та швидкістю inference (6.09 мс), випереджаючи EfficientNet-B0 за обома критеріями. За розміром моделі (16.7 МБ проти 90.8 МБ у ResNet-50).

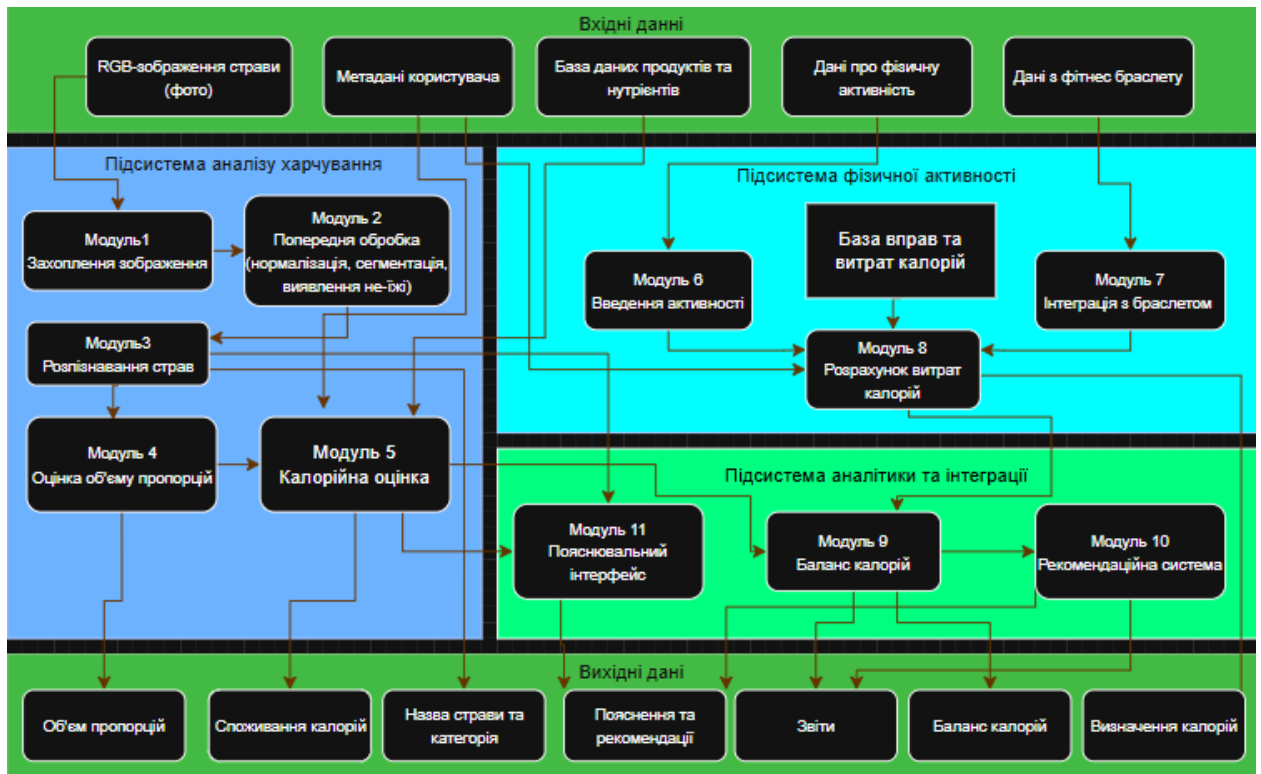
**Похибки класифікації.** Найбільші труднощі виникають для візуально схожих класів (наприклад, Steak / Prime Rib / Filet Mignon), що узгоджується з проблемою варіативності страв. Впевненість моделі знижується до  $\approx 64\%$  саме для таких категорій, тоді як для виразних класів (Sushi, Pizza) досягає 96–100%.

## Висновки до розділу

Проведені експерименти показали, що ResNet-50 з двофазним fine-tuning досягає Top-1 Accuracy 80.66% та Top-5 Accuracy 95.22% на Food-101, що є конкурентоспроможним результатом для 15 епох навчання. MobileNetV3 забезпечує найкращий баланс між точністю та розміром моделі серед легковагових архітектур і рекомендується для мобільного розгортання. Порівняння з літературою підтверджує, що подальше збільшення глибини fine-tuning та застосування сучасних технік аугментації дозволить наблизитись до найкращих результатів галузі ( $\approx 88\%$ ).

## 4 ПЕРСПЕКТИВНА АРХІТЕКТУРА СИСТЕМИ ТА МОЖЛИВІ НАПРЯМИ РОЗВИТКУ

Детальну структурну схему перспективної програмної системи наведено на рисунку 4.1.



**Рисунок 4.1** – Детальна структурна схема програмної системи

Наведена схема ілюструє розширену архітектуру системи з урахуванням перспективних модулів, детальний опис яких наведено в наступних розділах.

### 4.1 Порівняння реалізованої та перспективної архітектури

Реалізована в межах даної роботи система охоплює базовий функціональний цикл оцінки калорійності страв та фізичної активності. Концептуальна схема (рис. 4.1) розширює її трьома додатковими

підсистемами та новими джерелами вхідних даних. Таблиця 4.1 відображає ключові відмінності між поточною реалізацією та перспективною архітектурою.

**Таблиця 4.1** – Порівняння реалізованої та перспективної архітектури

Компонент	Реалізовано	Перспектива
Класифікація страв (ResNet-50)	✓	✓
Розрахунок калорійності страви	✓	✓
Введення фізичної активності вручну	✓	✓
Баланс калорій за день	✓	✓
Дані з фітнес-браслету (авто)	—	✓
Підсистема аналітики та інтеграції	—	✓
Рекомендаційна система	—	✓
Пояснювальний інтерфейс з графіками	—	✓

Як видно з таблиці, реалізована система забезпечує повний базовий цикл оцінки харчування, тоді як перспективна архітектура розширює її можливості автоматизацією збору даних та інтелектуальними підсистемами.

## 4.2 Напрями розвитку

На основі виявлених обмежень реалізованої системи та аналізу перспективної архітектури визначено такі пріоритетні напрями розвитку.

**Інтеграція з носимими пристроями (Модуль 7).** Автоматичне отримання даних з фітнес-браслетів (кроки, пульс, активні хвилини) через Bluetooth або API виробника дозволить замінити ручне введення активності та підвищити точність оцінки енерговитрат [10].

**Рекомендаційна система (Модуль 10).** Накопичена історія балансу калорій і цілі користувача (схуднення, підтримка або набір маси) є достатньою основою для формування персоналізованих порад щодо корекції харчування та режиму активності.

**Пояснювальний інтерфейс (Модуль 11).** Візуалізація динаміки балансу у вигляді тижневих та місячних графіків, а також пояснення кожного розрахунку забезпечує деталізацію системи та сприяє формуванню стійких харчових звичок [14].

**Вдосконалення класифікатора.** Збільшення глибини fine-tuning, застосування label smoothing та міхур-аугментації дозволить наблизити точність до найкращих результатів галузі ( $\approx 88\%$  Top-1 на Food-101 [16]).

## ВИСНОВКИ

У роботі розроблено та досліджено систему автоматичної оцінки калорійності страв і фізичної активності користувача на основі методів комп'ютерного зору та глибокого навчання. За результатами виконаної роботи сформульовано такі висновки.

1) Проведено аналіз наявних методів розпізнавання харчових об'єктів, оцінки маси та калорійності страв, а також методів оцінки фізичної активності. Встановлено, що методи на основі CNN з трансферним навчанням є стандартом галузі та забезпечують найвищу точність класифікації на датасеті Food-101. Виявлено п'ять актуальних обмежень в наявних підходах: відсутність інформації про глибину для оцінки маси, труднощі зі змішаними стравами, варіативність порцій, ізольованість задач харчування та активності, а також відсутність достатньої персоналізації.

2) Сформульовано математичну постановку задачі як знаходження двох відображень:  $f : \mathcal{I} \rightarrow \mathcal{C}$  для оцінки калорійності страви за зображенням та  $g : A(t) \rightarrow E$  для оцінки енерговитрат за часовим рядом активності. Як критерій якості обрано мінімізацію середньоквадратичної похибки відносно еталонних значень.

3) Розроблено модульну архітектуру програмної системи, що складається з підсистеми аналізу харчування (модулі 1–5) та підсистеми фізичної активності (модулі 6–8). Підсистема харчування реалізує послідовність етапів: захоплення зображення  $\rightarrow$  нормалізація за статистиками ImageNet  $\rightarrow$  класифікація ResNet-50  $\rightarrow$  оцінка маси  $\rightarrow$  розрахунок калорійності. Підсистема активності розраховує енерговитрати за формулою  $E = MET \times m \times t/60$ .

4) Навчано та досліджено архітектуру ResNet-50 з двофазним fine-tuning на датасеті Food-101. Двофазна стратегія (заморожений backbone на епохах 1–5, повний fine-tuning на епохах 6–15) забезпечила стрибок валідаційної точності на  $\approx 15$  п.п. після розморожування ваг. Фінальні результати: **Top-1 Accuracy** –

**80.66%, Top-5 Accuracy – 95.22%.**

5) Проведено порівняльний аналіз трьох архітектур CNN в однакових умовах навчання. ResNet-50 перевершує легковагові моделі за точністю (+16.58 п.п. над MobileNetV3 за Top-1), проте має найбільший розмір (90.8 МБ проти  $\approx 16$  МБ). MobileNetV3 забезпечує найкращий баланс точність/розмір серед легковагових архітектур і рекомендується для мобільного розгортання.

6) Зіставлення з літературою підтвердило конкурентоспроможність отриманого результату відносно базових методів та підходів без глибокого fine-tuning. Відставання від найкращих результатів галузі ( $\approx 88\%$  Top-1 [16]) пояснюється обмеженою кількістю епох навчання та відсутністю додаткових технік регуляризації (label smoothing, mixup), що є напрямом подальшого вдосконалення.

7) Реалізовано модуль фізичної активності на базі MET-коефіцієнтів для 30 видів активності у п'яти категоріях. Наскрізне тестування системи CalorieTrack підтвердило коректність інтеграції підсистем: перевірка підсумкового балансу (спожито 2172.0 ккал, спалено 1102.5 ккал, нетто +1069.5 ккал) показала нульову розбіжність між аналітичним та програмним результатом.

8) Визначено перспективні напрями розвитку системи: інтеграція з носимими пристроями для автоматичного збору даних активності, розробка рекомендаційної системи на основі накопиченої історії балансу калорій, а також впровадження пояснювального інтерфейсу з тижневою та місячною аналітикою.

**ПЕРЕЛІК ПОСИЛАНЬ**

- [1] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann та ін. «Compendium of Physical Activities: A second update of codes and MET values». В: *Medicine & Science in Sports & Exercise* 43.8 (2011), с. 1575—1581. DOI: 10.1249/MSS.0b013e31821ece12.
- [2] Ahmad AlMughrabi та ін. «VoLETA: One- and few-shot food volume estimation». В: *arXiv preprint arXiv:2407.01717* (2024). DOI: 10.48550/arXiv.2407.01717.
- [3] Davide Anguita та ін. *UCI HAR Dataset: Human Activity Recognition Using Smartphones*. UCI Machine Learning Repository. 2013. URL: <https://archive.ics.uci.edu/dataset/240>.
- [4] Ivo M. Baltruschat та ін. «Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification». В: *Scientific Reports* 9 (2019), с. 6381. DOI: 10.1038/s41598-019-42294-8.
- [5] Maksym Bekuzarov та ін. «XMem++: Production-level Video Segmentation From Few Annotated Frames». В: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. ЖОВТ. 2023, с. 635—644.
- [6] Lukas Bossard, Matthieu Guillaumin та Luc Van Gool. «Food-101 – Mining Discriminative Components with Random Forests». В: *European Conference on Computer Vision (ECCV)*. Springer, 2014, с. 446—461. DOI: 10.1007/978-3-319-10599-4\_29.
- [7] Shing Chan та ін. «CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition». В: *Scientific Data* 11 (2024), с. 1135. DOI: 10.1038/s41597-024-03960-3.
- [8] Jingjing Chen та Chong-Wah Ngo. «Deep-based Ingredient Recognition for Cooking Recipe Retrieval». В: *Proceedings of the 24th ACM International Conference on Multimedia (MM)*. 2016. DOI: 10.1145/2964284.2964315.

- [9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov та ін. «An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale». B: *arXiv:2010.11929* (2020).
- [10] T. Ferguson та ін. «Effectiveness of wearable activity trackers to increase physical activity: A systematic review and meta-analysis». B: *The Lancet Digital Health* 4.8 (2022), e615—e626. DOI: 10.1016/S2589-7500(22)00111-X.
- [11] R. A. Harrington та ін. «Nutrient composition databases in the age of big data: foodDB, a comprehensive, real-time database infrastructure». B: *BMJ Open* 9.6 (2019), e026652. DOI: 10.1136/bmjopen-2018-026652.
- [12] Kaiming He та ін. «Deep Residual Learning for Image Recognition». B: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, с. 770—778. DOI: 10.1109/CVPR.2016.90.
- [13] Gao Huang та ін. «Densely Connected Convolutional Networks». B: *CVPR*. 2017, с. 2261—2269.
- [14] M. N. Kamel Boulos та S. P. Yang. «Mobile physical activity planning and tracking: a brief overview of current options and desiderata for future solutions». B: *mHealth* 7 (2021), с. 13. DOI: 10.21037/mhealth.2020.01.01.
- [15] Rajdeep Kaur, Rakesh Kumar та Meenu Gupta. «Deep neural network for food image classification and nutrient identification: A systematic review». B: *Reviews in Endocrine and Metabolic Disorders* 24.4 (2023), с. 633—653. DOI: 10.1007/s11154-023-09795-4.
- [16] Sushant Kaushal та ін. «Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances». B: *Trends in Food Science & Technology* 146 (2024), с. 104408. DOI: 10.1016/j.tifs.2024.104408.

- [17] Yoshiyuki Kawano та Keiji Yanai. «FoodCam-256: A Large-scale Real-time Mobile Food Recognition System employing High-Dimensional Features and Compression of Classifier Weights». B: *Proceedings of the 22nd ACM International Conference on Multimedia (MM)*. 2014, с. 761—762. DOI: 10.1145/2647868.2654869.
- [18] R. S. Kumar та J. Prem Kumar. «Food object detection using custom-trained YOLOv8 with transfer learning». B: *Engineering, Technology & Applied Science Research* 15.6 (2025), с. 30508—30513. DOI: 10.48084/etasr.14810.
- [19] Ki-Seung Lee. «Multi-Spectral Food Classification and Caloric Estimation Using Predicted Images». B: *Foods* 13.4 (2024), с. 551. DOI: 10.3390/foods13040551.
- [20] Philipp Lindenberger та ін. «Pixel-Perfect Structure-from-Motion with Featuremetric Refinement». B: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. ЖОВТ. 2021, с. 5987—5997.
- [21] Michael Mommert та ін. «Power Plant Classification from Remote Imaging with Deep Learning». B: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2021. DOI: 10.48550/arXiv.2107.10894.
- [22] Emiko Okada та ін. «National Nutrition Surveys Applying Dietary Records or 24-h Dietary Recalls with Questionnaires: A Scoping Review». B: *Nutrients* 15.22 (2023), с. 4739. DOI: 10.3390/nu15224739.
- [23] Attila Reiss та Didier Stricker. «Introducing a New Benchmarked Dataset for Activity Monitoring». B: *Proceedings of the 16th International Symposium on Wearable Computers (ISWC)*. 2012, с. 108—109. DOI: 10.1109/ISWC.2012.13.
- [24] Shaoqing Ren та ін. «Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks». B: *Advances in Neural Information Processing Systems (NeurIPS)*. Т. 28. 2015, с. 91—99. URL: <https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks>.

- [25] Chandan Kumar Sah та ін. «Advancing Autonomous Vehicle Intelligence: Deep Learning and Multimodal LLM for Traffic Sign Recognition and Robust Lane Detection». В: *arXiv preprint arXiv:2503.06313* (2025). DOI: 10.48550/arXiv.2503.06313.
- [26] Quinn Thames та ін. «Nutrition5k: Towards Contextual Food Recognition and Nutrition Estimation in the Wild». В: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021.
- [27] Nawanol Theera-Ampornpant та Panisa Treepong. «Thai Food Recognition Using Deep Learning With Cyclical Learning Rates». В: *IEEE Access* 12 (2024), с. 174204—174221. DOI: 10.1109/ACCESS.2024.3503672.
- [28] Philipp Tschandl та ін. «Human–computer collaboration for skin cancer recognition». В: *Nature Medicine* 26 (2020), с. 1229—1234. DOI: 10.1038/s41591-020-0942-0.
- [29] U.S. Department of Agriculture. *FoodData Central*. 2019. URL: <https://fdc.nal.usda.gov>.
- [30] W. Wang та ін. «A review on vision-based analysis for automatic dietary assessment». В: *Trends in Food Science & Technology* 116 (2021), с. 201—211. DOI: 10.1016/j.tifs.2021.07.008.
- [31] Yiming Wang та ін. «NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction». В: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023, с. 3272—3283. DOI: 10.1109/ICCV51070.2023.00305.
- [32] World Health Organization. *New WHO report: Europe can reverse its obesity "epidemic"*. European Association for the Study of Obesity. Трав. 2022. URL: <https://easo.org/new-who-report-europe-can-reverse-its-obesity-epidemic/> (дата зверн. 01.04.2026).

- [33] Xiaoqing Zheng та ін. «Recent advances in surface defect inspection of industrial products using deep learning techniques». В: *The International Journal of Advanced Manufacturing Technology* 113 (2021), с. 35—58. DOI: 10.1007/s00170-021-06592-8.
- [34] Fuzhen Zhuang та ін. «A Comprehensive Survey on Transfer Learning». В: *Proceedings of the IEEE* 109.1 (2021), с. 43—76. DOI: 10.1109/JPROC.2020.3004555.