

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Комп'ютерного моделювання та інтелектуальних технологій _____

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

Розроблення системи класифікації та оцінювання калорійності страв на основі
глибинного навчання та моделі YOLO

(тема)

Виконав:

здобувач _____ 2 _____ року навчання,

групи _____ СПРМ-24-1 _____

_____ Михайло КОВАЛЬОВ _____

(власне ім'я, прізвище)

Спеціальність _____ 122 Комп'ютерні науки _____

(код і повна назва спеціальності)

Тип програми _____ освітньо-наукова _____

(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Системне проєктування _____

(повна назва освітньої програми)

Керівник _____ проф. каф. КМІТ Сергій МІНУХІН _____

(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри _____

_____ (підпис)

_____ Ігор ГРЕБЕННІК _____

(власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наукКафедра Комп'ютерного моделювання та інтелектуальних технологійРівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки

(код і повна назва)

Тип програми освітньо-наукова

(освітньо-професійна або освітньо-наукова)

Освітня програма Системне проєктування

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«____» _____ 20 ____ р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Ковальову Михайлу Ігоровичу

(прізвище, ім'я, по батькові)

1. Тема роботи Розроблення системи класифікації та оцінювання калорійності страв на основі глибокого навчання та моделі YOLO

затверджена наказом університету від 06 квітня 2026 р. № 268Ст2. Термін подання здобувачем роботи до екзаменаційної комісії 13 травня 2026 р.

3. Вихідні дані до роботи: набори цифрових зображень готових страв, отримані за допомогою мобільних пристроїв у реальних умовах зйомки, табличні бази даних нутрієнтів із фізико-хімічними характеристиками продуктів (калорійність, щільність, стандартна висота порцій), анотовані датасети з попільськими масками сегментації об'єктів, інформація про сучасні методи інстанс-сегментації та детекції на основі архітектур, DETR та Segment Anything Model, приклади архітектурних рішень систем моніторингу харчування, інструментарій розробки моделей глибокого навчання – мова Python та бібліотеки PyTorch, OpenCV, Scikit-fuzzy, обчислювальні ресурси для проведення інференсу ансамблів моделей, метрики точності, повноти, F1-міри для оцінки класифікації та показники MAE для верифікації розрахунків енергетичної цінності страв.

4. Перелік питань, що потрібно опрацювати в роботі: Аналіз предметної області та існуючих методів оцінки енергетичної цінності харчування за візуальними даними; дослідження джерел нутрієнтної інформації (USDA, EuroFIR) та формування інтегрованої бази знань з показниками щільності (FAO/INFOODS) і геометричними

константами страв; розробка архітектури інтелектуальної системи на основі ансамблю глибоких нейронних мереж (ResNet, DenseNet) для класифікації та моделей RT-DETR і SAM для інстанс-сегментації; розробка алгоритму метричного калібрування сцени за допомогою еталонних об'єктів (столових приборів) для переходу від піксельних до фізичних одиниць виміру; реалізація методу апроксимації об'єму страв з використанням геометричних коефіцієнтів та інтелектуальної стабілізації результатів на основі апарату нечіткої логіки; програмна реалізація системи у вигляді клієнт-серверного додатка з інтеграцією генеративного ШІ (Gemini API) для надання персоналізованих рекомендацій; експериментальна оцінка точності розрахунків енергетичної цінності та аналіз помилок класифікації на контрольних наборах даних.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій: Блок-схема інформаційної взаємодії компонентів системи; схема багаторівневого конвеєра обробки зображень; структура інтегрованої бази знань нутрієнтних даних та фізичних констант страв; діаграма архітектури ансамблю моделей класифікації та інстанс-сегментації; схема алгоритму метричного калібрування сцени на основі еталонних об'єктів; графіки залежності точності класифікації від кількості епох навчання; порівняльна гістограма точності ансамблю та базових архітектур; схема роботи логічного виводу NutritionAssistant та інтеграції з Gemini API; візуалізація результатів сегментації та накладання масок на зображення страв; скріншоти користувацького інтерфейсу з демонстрацією персоналізованих порад; графіки розподілу помилок розрахунку енергетичної цінності.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1.	Отримання завдання на виконання кваліфікаційної	06.03.2025	Виконано
2.	Аналіз завдання та предметної області	08.03.2025 – 13.03.2025	Виконано
3.	Опрацювання літератури та аналіз об'єкту дослідження	13.03.2025 – 16.04.2025	Виконано
4.	Формування бази знань (USDA, FAO) та підготовка наборів даних для навчання	17.03.2025 – 20.03.2025	Виконано
5.	Проектування та навчання ансамблевої моделі класифікації	21.03.2025 – 30.03.2025	Виконано
6.	Розроблення модулів інстанс-сегментації та детекції на базі RT-DETR та SAM	01.04.2025 – 05.04.2025	Виконано
7.	Розроблення алгоритму метричного калібрування сцени за еталонними об'єктами	06.04.2025 – 10.04.2025	Виконано
8.	Реалізація математичної моделі апроксимації об'єму та маси страв	11.04.2025 – 20.04.2025	Виконано
9.	Інтеграція системи з Getinі API для генерації персоналізованих рекомендацій	21.04.2025 – 28.04.2025	Виконано
10.	Проведення експериментальних досліджень точності розрахунків (MAE, MAPE)	29.04.2025 – 05.05.2025	Виконано
11.	Оформлення пояснювальної записки та презентації	05.05.2025 – 09.05.2025	Виконано
12.	Подання роботи на рецензування	09.05.2025	Виконано
13.	Подання роботи до екзаменаційної комісії	12.05.2026	Виконано

Дата видачі завдання 06 квітня 2026 р.

Здобувач



(підпис)

Керівник роботи

проф. каф. КМІТ Сергій МІНУХІН

(підпис)

(посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка до магістерської кваліфікаційної роботи: 119 с., 11 табл., 22 рис., 2 дод., 77 джерел.

АВТОМАТИЗАЦІЯ ДІЄТОЛОГІЇ, АНСАМБЛЕВІ МЕТОДИ, ГЛИБОКЕ НАВЧАННЯ, КЛАСИФІКАЦІЯ, СТРАВА, КОМП'ЮТЕРНИЙ ЗІР, НЕЙРОННА МЕРЕЖА, ТРАНСФЕРНЕ НАВЧАННЯ, ХАРЧОВІ ПРОДУКТИ, DIRECTML, PYTHON, YOLO, RT-DETR.

Об'єктом дослідження є процеси автоматизованого розпізнавання харчових продуктів на цифрових зображеннях методами комп'ютерного зору. Метою роботи є розробка та експериментальне дослідження інтелектуальної системи на основі ансамблів глибоких нейронних мереж для класифікації страв та подальшої метричної оцінки їх калорійності. В ході реалізації було розроблено алгоритм уніфікації наборів даних Food-101 та UEC-FOOD 256, що дозволило розширити таксономію системи до 334 категорій. Для розв'язання задачі класифікації впроваджено ансамбль архітектур ResNet-50 та DenseNet-121, навчених за стратегією двофазного трансферного навчання з оптимізацією під середовище DirectML. Точність розпізнавання на розширеній вибірці склала 73,71% (Top-1), а верифікація на авторському наборі даних підтвердила стійкість системи з показником 70,67%.

Особливу увагу приділено обчислювальному модулю, де для детекції референтних об'єктів використано архітектуру RT-DETR, а для прецизійної сегментації меж страв та інструментів – модель Segment Anything Model. Вперше впроваджено гібридний алгоритм регуляризації об'єму порцій, який поєднує методи

нечіткої логіки для оцінки довіри до масштабу та статистичні дані з бази нутрієнтів. Це дозволило реалізувати механізм «анти-абсурдного» кліпінгу площі, що суттєво мінімізує похибку розрахунку калорійності при складних ракурсах зйомки та забезпечує високу точність фінальної оцінки енергетичної цінності страви.

ABSTRACT

Report on Research Practice: 119 pp., 11 tables, 22 figs., 2 appendixes, 77 sources.

CLASSIFICATION, COMPUTER VISION, DEEP LEARNING, DIETETICS AUTOMATION, DIRECTML, DISH, ENSEMBLE METHODS, FOOD PRODUCTS, NEURAL NETWORK, PYTHON, RT-DETR, TRANSFER LEARNING, YOLO.

The research object is the automated recognition of food in digital images using computer vision methods. The aim of the work is to develop and experimentally study an intelligent system based on ensembles of deep neural networks for dish classification and the subsequent estimation of their caloric content.

During implementation, an algorithm for unifying the Food-101 and UEC-FOOD 256 datasets was developed, expanding the system's taxonomy to 334 categories. To solve the classification task, an ensemble of ResNet-50 and DenseNet-121 architectures was implemented, trained using a two-phase transfer learning strategy with DirectML optimization. The recognition accuracy on the expanded sample reached 73.71% (Top-1), while verification on the author's original dataset confirmed the system's robustness, achieving 70.67%.

Special attention is paid to the computational module, where the RT-DETR architecture is used for reference object detection, and the Segment Anything Model (SAM) is employed for precision segmentation of dish boundaries and utensils. A hybrid portion-volume regularization algorithm was introduced for the first time, combining fuzzy-logic methods for scale-trust estimation with statistical data from nutrient databases. This enabled the implementation of an "anti-absurd" area-clipping mechanism, which

significantly reduces caloric calculation errors under complex shooting angles and ensures high accuracy in the final estimation of the dish's energy value.

ЗМІСТ

ВСТУП	12
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА МЕТОДІВ АВТОМАТИЗОВАНОЇ ОЦІНКИ ЕНЕРГЕТИЧНОЇ ЦІННОСТІ ХАРЧУВАННЯ.....	14
1.1 Огляд проблеми контролю раціону харчування та його впливу на здоров'я людини.....	14
1.2 Аналіз методів оцінки енергетичної цінності страв за візуальними даними	18
1.3 Огляд та аналіз інформаційних систем та технологій для моніторингу харчування на основі штучного інтелекту	23
1.4 Постановка задачі дослідження	25
2 ПІДГОТОВКА ТА АНАЛІЗ ДАНИХ ДЛЯ НАВЧАННЯ.....	27
2.1 Характеристика вихідних наборів даних	27
2.2 Розроблення алгоритму уніфікації та мапінгу	32
2.3 Проблема дисбалансу та методи її вирішення.....	33
2.4 Попереднє оброблення та аугментація зображень	35
3 ПРОЄКТУВАННЯ ТА РЕАЛІЗАЦІЯ АРХІТЕКТУРИ СИСТЕМИ	37
3.1 Обґрунтування та побудова ансамблевої архітектури	37
3.2 Порівняльний аналіз сучасних архітектур глибокого навчання	38
3.3 Стратегія багатоетапного навчання та обґрунтування вибору гіперпараметрів	39
3.4 Програмно-апаратне середовище та інструментарій реалізації	41

3.5	Розроблення концепції метричного калібрування сцени на основі YOLO-детекції та її вдосконалення методами інстанс-сегментації еталонних об'єктів.	43
4	АНАЛІЗ РЕЗУЛЬТАТІВ КЛАСИФІКАЦІЇ ТА ТЕСТУВАННЯ	46
4.1	Методологія фінального оцінювання та метрики якості	46
4.2	Порівняльний аналіз точності ансамблю та його базових архітектур	48
4.3	Аналіз помилок класифікації та проблемних категорій зображень	53
4.4	Верифікація інтелектуальної системи на контрольному агрегованому наборі даних	56
4.5	Технологічна оцінка роботи системи в середовищі DirectML	59
4.6	Технологічна апробація методу метричного калібрування на основі об'єктної детекції	60
4.7	Методика інтелектуального оцінювання достовірності розрахунків на основі апарату нечіткої логіки	63
4.8	Аналіз ефективності сегментації та вирішення проблеми обмеженості класів детектора	69
4.9	Експериментальна оцінка точності розрахунку енергетичної цінності та верифікація результатів	76
5	ПРОГРАМНО-ТЕХНОЛОГІЧНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ТА ПЕРСПЕКТИВИ ЇЇ МОДЕРНІЗАЦІЇ	81
5.1	Особливості поточної програмної реалізації та технологічного стеку	81
5.2	Архітектурна модель розгортання та мобільна адаптація	84
5.3	Інтеграція генеративного ШІ для формування персоналізованих рекомендацій	85
5.4	Розробка користувацького інтерфейсу та візуалізація результатів	86

ВИСНОВКИ.....	93
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	95
ДОДАТОК А	105
ДОДАТОК Б	106

ВСТУП

Питання автоматизованого контролю харчування стає дедалі актуальнішим у зв'язку зі зростанням рівня аліментарно-залежних захворювань у світі. Сучасні рішення переважно покладаються на ручне введення даних, що є суб'єктивним та незручним процесом. Автоматизація розпізнавання їжі засобами згорткових нейронних мереж стикається з проблемами високої міжкласової схожості, складних ракурсів та варіативності умов освітлення.

Актуальність теми зумовлена необхідністю створення інтелектуальних систем, які здатні не лише класифікувати тип страви, а й компенсувати втрату просторової інформації на двовимірних зображеннях. Врахування реального масштабу та прецизійна сегментація об'єктів є критичними факторами для переходу від візуальної ідентифікації до точного розрахунку об'єму та енергетичної цінності.

Об'єктом дослідження є процеси автоматизованого розпізнавання харчових продуктів на зображеннях засобами комп'ютерного зору.

Предметом дослідження є ансамблеві архітектури глибоких нейронних мереж, методи прецизійної сегментації та алгоритми метричного калібрування на основі детектування еталонних об'єктів.

Метою роботи є розробка та реалізація програмного комплексу для багатокласової класифікації страв на основі ансамблю глибоких мереж, а також впровадження гібридних методів визначення реального масштабу та маси порцій.

Для досягнення мети треба вирішити такі завдання:

- здійснити аналіз та уніфікацію гетерогенних наборів даних Food-101 та UEC-FOOD 256 для формування розширеної бази даних із 334 категоріями;
- обґрунтувати та спроектувати ансамблеву архітектуру на основі моделей ResNet-50 та DenseNet-121 для підвищення точності розпізнавання;

- розробити методику метричного калібрування сцени шляхом інтеграції модуля детекції референтних об'єктів на базі моделі RT-DETR та високоточної сегментації за допомогою Segment Anything Model;
- впровадити інтелектуальний алгоритм регуляризації об'єму на основі нечіткої логіки для оцінки рівня довіри до розрахованого масштабу та стабілізації обчислень калорійності страв;
- провести експериментальне тестування розробленої системи в програмно-апаратному середовищі DirectML на авторському наборі даних.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА МЕТОДІВ АВТОМАТИЗОВАНОЇ ОЦІНКИ ЕНЕРГЕТИЧНОЇ ЦІННОСТІ ХАРЧУВАННЯ

1.1 Огляд проблеми контролю раціону харчування та його впливу на здоров'я людини

Сучасний стан охорони здоров'я характеризується стрімким зростанням кількості хронічних метаболічних захворювань, серед яких провідне місце посідає цукровий діабет. Це захворювання є системним порушенням обміну речовин, що виникає внаслідок недостатньої секреції гормону інсуліну підшлунковою залозою або зниженням чутливості клітин організму до нього. Оскільки інсулін є ключовим регулятором транспорту глюкози до клітин для вироблення енергії, будь-які збої в цьому механізмі призводять до неконтрольованого зростання рівня глікемії [1]. Хронічна гіперглікемія, своєю чергою, спричиняє незворотні ураження судинної системи, що стає причиною серйозних ускладнень: діабетичної ретинопатії, серцево-судинних патологій та ниркової недостатності [2].

Ефективне управління станом пацієнта при метаболічних розладах базується на тріаді: медикаментозне втручання, регулярна фізична активність та, що є найбільш критичним, суворий контроль дієтичного споживання [1]. Моніторинг раціону є стратегічним інструментом для стабілізації рівня глюкози, оскільки дозволяє корегувати споживання вуглеводів та підтримувати енергетичний баланс. Проте традиційні методи ведення харчових щоденників (паперові записи або ручне введення даних у застосунки) мають суттєві недоліки. Дослідження показують, що пацієнти часто забувають фіксувати прийоми їжі, суб'єктивно занижують об'єм спожитих порцій або навіть навмисно оминають інформацію про «нездорову» їжу, що робить дані таких щоденників нерелевантними для клінічного аналізу [3, 4].

Історично перші спроби цифровізації дієтологічного контролю були реалізовані через веб-платформи з миттєвим аналізом нутрієнтів (такі як Calorie King, CRON-O-Meter, Selfnutritiondata та інші). Ці системи дозволяли користувачам отримувати детальні звіти про склад страв, проте залишалися критично залежними від точності самостійного введення даних про тип продукту та його масу. Необхідність спрощення цього процесу призвела до концепції злиття методів об'єктної детекції з існуючими системами дієтологічного моніторингу.

Знаковим етапом у розвитку мобільних технологій оцінки харчування стала праця Joutou (2009) [5], у якій було вперше описано алгоритм детекції харчових об'єктів для практичного використання на мобільному пристрої. Це стало поштовхом для створення нових підходів, спрямованих на зменшення обсягу ручних операцій та вимог до апаратного забезпечення при аналізі раціону. Задля вирішення проблеми суб'єктивності та низької точності ручного моніторингу були впроваджені автоматизовані системи розпізнавання харчових продуктів Food Image Recognition System.

Розвиток обчислювальних потужностей сучасних мобільних пристроїв та поширення портативних сенсорів зробили можливим створення інтелектуальних систем, здатних класифікувати страви та оцінювати їх енергетичну цінність у реальному часі за допомогою фотофіксації [4, 5]. Використання методів глибокого навчання дозволяє моделям FIRS автоматично витягувати візуальні ознаки об'єктів, ідентифікувати класи продуктів та, за умови використання відповідних калібрувальних алгоритмів, розраховувати калорійність порцій. Автоматизація цього процесу не лише спрощує щоденну рутину пацієнта, а й забезпечує об'єктивність даних, що є критично важливим для превентивної медицини та персоналізованого підходу до лікування діабету та інших аліментарно-залежних захворювань [5].

Цукровий діабет становить зростаючий тягар для світової економіки, вражаючи понад 10% дорослого населення планети. Згідно з результатами сучасного макроекономічного моделювання, що охоплює 204 країни та території, сумарний економічний збиток від діабету в період з 2020 по 2050 рік прогнозується на рівні 10,2 трлн доларів при оцінці прямих медичних витрат та втрат ефективної пропозиції робочої сили. Проте, при врахуванні витрат на неформальний догляд, цей показник катастрофічно зростає до 78,8 трлн доларів, що еквівалентно суттєвій частці щорічного глобального ВВП [6].

Окрім прямих витрат на медицину, існують значні непрямі економічні втрати, зумовлені зниженням продуктивності населення. Передчасна інвалідність та високий рівень госпіталізації осіб працездатного віку призводять до недоотримання ВВП та зростання соціальних видатків держави. Впровадження автоматизованих систем контролю харчування дозволяє перейти від реактивної моделі «лікування наслідків» до превентивної моделі «управління станом», що є стратегічно вигіднішим для економіки та сприяє збереженню людського капіталу.

В умовах України соціальне значення таких розробок посилюється через необхідність забезпечення населення доступними інструментами самодіагностики. Складність доступу до персональних консультацій дієтологів та висока вартість лабораторного моніторингу роблять мобільні інтелектуальні системи на базі технологій комп'ютерного зору оптимальним рішенням. Використання звичайного смартфона як інструмента точного підрахунку КБЖУ дозволяє масштабувати якісну дієтологічну підтримку на широкі верстви населення без додаткових інвестицій у медичне обладнання, що є критично важливим для підтримки громадського здоров'я.

На даний момент було виявлено феномен «спотворення порцій», який полягає в поступовій втраті здатності людини адекватно оцінювати об'єм споживаної їжі через зміщення суб'єктивних стандартів «нормальної» порції. Протягом останніх

десятиліть розвиток індустрії швидкого харчування та маркетингові стратегії призвели до значного збільшення еталонних розмірів страв у закладах громадського харчування. Це спричинило ефект «нормалізації» надмірних об'ємів: те, що 20 років тому вважалося великою порцією, сьогодні сприймається споживачами як стандартна одиниця споживання, що автоматично переноситься і на «домашнє» оцінювання раціону [7].

Проблема візуальної оцінки посилюється оптичними викривленнями, зокрема ілюзією Дельбефа [8], коли однакова кількість їжі сприймається людиною по-різному - залежно від розміру посуду. На великій тарілці порція виглядає меншою, що підсвідомо стимулює користувача до збільшення об'єму споживання або до заниження оцінки калорійності під час заповнення щоденника харчування. Статистичні дослідження підтверджують, що навіть підготовлені користувачі помиляються в оцінці ваги складних страв на 30-50%, що при щоденному звичайному раціоні - понад 3 000 Ккал - може призвести до критичного профіциту енергії, який не реєструється системою ручного моніторингу.

З огляду на зазначені психофізіологічні чинники, створення систем, що базуються на комп'ютерному зорі та використанні фізичних еталонів, є єдиним способом отримання об'єктивних даних. Автоматизована система оцінки, на відміну від людини, не піддається впливу контекстуальних викривлень та забезпечує сталу точність вимірювання геометричних параметрів страви. Це дозволяє нівелювати вплив «людського фактору» та забезпечити пацієнтів із метаболічними порушеннями даними, що відповідають реальному енергетичному балансу.

1.2 Аналіз методів оцінки енергетичної цінності страв за візуальними даними

1.2.1 Класифікаційний метод

Класифікаційний підхід, або метод «зверху-вниз», є історично першим та технічно найбільш поширеним методом автоматизованої оцінки нутрієнтного складу страв. Основна концепція полягає у послідовному виконанні двох етапів: ідентифікації типу харчового об'єкта на зображенні та подальшому вилученні відповідних значень енергетичної цінності із заздалегідь сформованої бази даних.

Сучасні класифікаційні системи базуються на глибоких згорткових нейронних мережах, таких як ResNet, DenseNet або EfficientNet. Процес ідентифікації полягає у перетворенні вхідного зображення X у вектор ознак високого рівня, який на вихідному шарі мережі (через функцію Softmax) перетворюється на розподіл імовірностей за набором класів [9]:

$$P(y_i|X) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad (1)$$

де z_i – логіт для i -го класу страви, K – загальна кількість категорій у базі даних.

Для підвищення надійності розпізнавання часто використовують ансамблювання моделей, що дозволяє нівелювати індивідуальні зміщення окремих архітектур та підвищити загальну метрику точності (Accuracy).

Серед переваг даного методу слід відзначити такі:

- висока швидкість інференсу, при якому процес класифікації потребує значно менше обчислювальних ресурсів порівняно з попиксельною сегментацією;
- об'єктивність ідентифікації, в який при достатній якості навчальної вибірки моделі здатні розпізнавати складні страви, які важко ідентифікувати людині;

– масштабованість системи шляхом додавання нових позицій до бази даних без необхідності зміни логіки розрахунку.

Попри свою популярність, класифікаційний метод має фундаментальний недолік – він повністю ігнорує фізичні параметри конкретної порції. У межах цього підходу система оперує «середніми значеннями» калорійності на порцію або на 100 грамів продукту. Це призводить до виникнення таких проблем:

- варіативність порцій;
- проблема гетерогенності;
- відсутність метричної прив'язки.

Один і той самий клас об'єкта може мати масу від 100 до 350 грамів. Використання статичного середнього значення призводить до похибки, яка може перевищувати 200%. Метод демонструє низьку ефективність при аналізі страв із довільним поєднанням компонентів, де пропорції інгредієнтів змінюються динамічно. Класифікація не дає інформації про реальний об'єм та щільність об'єкта, що робить неможливим відновлення маси страви без додаткових інструментів геометричного аналізу. Для осіб із суворим контролем енергетичного балансу, за яким добова норма може складати понад 3 000 ккал, похибка в 10-15% стає критичною для досягнення цілей тренувального процесу або стабілізації метаболічного стану, використання суто класифікаційного методу є недостатнім.

1.2.2 Метод геометричного моделювання

Метод геометричного моделювання базується на математичній апроксимації тривимірного об'єму страви через її двовимірну проекцію на площину сенсора камери. На відміну від суто класифікаційного підходу, цей метод дозволяє врахувати

фізичні габарити конкретної порції, що є необхідною умовою для розрахунку маси продукту.

В основу цього методу покладено припущення, що більшість харчових об'єктів можна апроксимувати як квазі-регулярні геометричні тіла (циліндри, призми або сегменти сфер). Базова формула розрахунку об'єму V має вигляд:

$$V = S * h * k, \quad (2)$$

де S – площа проєкції об'єкта на площину, h – стандартна висота страви, що вилучається з бази даних нутрієнтів, k – поправковий коефіцієнт форми, що враховує нерівномірність розподілу маси об'єкта.

Головним викликом при використанні однієї камери є втрата інформації про глибину. Оскільки на 2D-зображенні неможливо безпосередньо вимірити, наприклад, товщину шматка м'яса або висоту пирога, система вимушена покладатися на апіорні знання. В базах даних для кожного класу страв фіксується «стандартна висота порції» h_{std} , яка була отримана на основі статистичного аналізу великої кількості реальних зразків.

Важливим аспектом моделювання є використання коефіцієнта k . Більшість страв не є ідеальними циліндрами: наприклад, порція лазаньї має майже прямокутний профіль $k \approx 0.9$, тоді як шматочок торта або піци звужується до країв або має похилу форму, що потребує використання коефіцієнта в межах 0.5-0.7. Впровадження такого динамічного значення параметра дозволить значно знизити систематичну похибку переоцінки об'єму, яка виникає при спрощеному розрахунку [10].

Додаткову складність додає кут зйомки. Більшість користувачів роблять фото під кутом 45-60°, що спричиняє ефект перспективного скорочення площі. Для компенсації цього явища в алгоритмах геометричного моделювання застосовується афінна корекція площі, яка дозволяє відновити реальні метричні параметри об'єкта незалежно від ракурсу зйомки.

Незважаючи на певні припущення, метод геометричного моделювання забезпечує прийнятну для практичного використання точність ($MAPE < 20\%$) при моніторингу раціону. Це робить його значно надійнішим за класифікаційні методи, особливо в задачах точного контролю нутрієнтів для користувачів із високими енергетичними потребами [11].

1.2.3 Стереоскопічні та багатопроєкційні методи оцінки

Найбільш технологічно досконалим підходом до визначення об'єму страв є використання глибинних сенсорів та методів стереозору, що спрямовані на повне відновлення тривимірної структури сцени та нівелювання необхідності використання апріорних знань про висоту об'єкта. Сучасні рішення в цій галузі базуються на двох основних підходах, першим з яких є активне зондування за допомогою LiDAR або ToF-сенсорів [12]. Використання лазерних далекомірів, інтегрованих у сучасні смартфони преміум-сегмента, дозволяє будувати хмари точок у реальному часі, отримувати точну карту глибини та розраховувати об'єм як інтеграл площі за висотою.

Альтернативою апаратним методам є підхід «структура з руху», який вимагає від користувача створення серії знімків під різними кутами навколо об'єкта. Шляхом ідентифікації спільних ключових точок на кадрах та подальшої тріангуляції алгоритм відновлює 3D-модель страви. Попри високу точність, ці методи мають суттєві бар'єри для масового впровадження через високу апаратну залежність та значне когнітивне навантаження на користувача. Більшість пацієнтів із хронічними станами, такими як цукровий діабет, потребують рішень за принципом «одне фото

– один результат», що робить багатопроєкційні методи надто складними для повсякденного моніторингу [13].

1.2.4 Метод референтних об'єктів та природних маркерів масштабу

Метод референтних об'єктів є оптимальним компромісом між складністю реалізації та точністю розрахунків, оскільки базується на введенні в сцену об'єкта з точно відомими фізичними розмірами для розрахунку реального масштабу зображення. Математична модель перетворення передбачає визначення масштабного коефіцієнта $Scale$ як відношення фізичної ширини об'єкта до його розміру в пікселях, що дозволяє розраховувати площу страви в метричних одиницях. У науковій літературі пропонувалися різні типи еталонів, зокрема монети, банківські картки або спеціалізовані мітки, проте вони мають суттєві гігієнічні та ергономічні недоліки, що ускладнює їх постійне використання в умовах прийому їжі [14].

Маючи ширину референтного об'єкта в міліметрах W_{real} та вимірявши її на зображенні в пікселях W_{pixel} , можна отримати масштабний коефіцієнт $Scale$:

$$Scale = \frac{W_{real}}{W_{pixel}} \left[\frac{\text{мм}}{\text{пікс}} \right] \quad (3)$$

Фінальна площа об'єкта, в квадратних міліметрах, розраховується як:

$$Area_{mm^2} = Area_{pixels} * Scale^2. \quad (4)$$

У межах даного дослідження обґрунтовано доцільність використання природних референтних маркерів у вигляді столових приборів, що характеризуються контекстуальною нативністю та високим ступенем стандартизації їх габаритів. Використання таких об'єктів не потребує від користувача додаткових маніпуляцій, оскільки вони природно присутні в кадрі.

Стабільність геометричних параметрів масового виробництва столових приборів дозволяє використовувати їх як надійну метричну опору, а чіткість їх контурів сприяє стабільній роботі сучасних моделей інстанс-сегментації навіть в умовах низької контрастності сцени. Це робить метод референтних маркерів найбільш адаптивним для систем щоденного контролю нутрієнтного складу раціону.

1.3 Огляд та аналіз інформаційних систем та технологій для моніторингу харчування на основі штучного інтелекту

Сучасний ринок програмного забезпечення для мобільного моніторингу здоров'я представлений широким спектром застосунків, що використовують технології штучного інтелекту для спрощення процесу логування харчування.

Застосунок Calorie Mama AI від компанії Azumio є одним із найбільш технологічно досконалих рішень у сфері автоматизованого розпізнавання їжі. В основі системи лежать глибокі нейронні мережі, навчені на мільйонах зображень, що дозволяє моделі ідентифікувати не лише базові продукти, а й складні страви різних кухонь світу. Проте ключовою архітектурною проблемою Calorie Mama AI є відсутність механізму точного метричного калібрування. Система здійснює оцінку порції на основі статистичних середніх значень для обраного класу, що призводить до значних систематичних похибок. Для користувачів із високою точністю раціону, таких як професійні атлети, ігнорування реальної геометричної площі та висоти об'єкта на тарілці робить цей інструмент лише допоміжним, а не основним засобом контролю [15].

Технологія Snap It інтегрована в екосистему Lose It! з метою мінімізації часу, який користувач витрачає на ведення харчового щоденника. Система фокусується

на швидкому витягуванні візуальних ознак об'єкта для автоматичного пошуку в базі даних. Головна перевага Snap It полягає в простоті користувацького досвіду, проте з інженерної точки зору цей алгоритм має суттєвий недолік – він працює суто в якісній площині ідентифікації. Відсутність еталонних маркерів у кадрі не дозволяє системі розрізнити перспективні викривлення та реальний масштаб. Це призводить до виникнення «метричного розриву», коли розпізнана назва страви не підкріплюється адекватним розрахунком її маси, що критично для систем підтримки пацієнтів із метаболічними розладами [16].

FatSecret є одним із найбільш затребуваних сервісів у східноєвропейському регіоні завдяки великій базі даних локальних продуктів та активній спільноті. Хоча платформа пропонує інструменти розпізнавання зображень, вони здебільшого обмежені функцією «підказки» назви страви, після чого користувач все одно змушений вручну обирати кількість одиниць або вагу в грамах. Це зводить автоматизацію до рівня текстового пошуку за зображенням, залишаючи проблему суб'єктивності оцінки порції невирішеною. Для користувачів з великим добовим раціоном, ручне введення в FatSecret стає джерелом накопичувальної похибки, яка може нівелювати результати дієтотерапії або тренувального процесу [17].

Застосунок YAZIO здобув популярність завдяки високій якості інтерфейсу та глибокій інтеграції з фітнес-трекерами. Система використовує AI-сканер для розпізнавання штрих-кодів та базових страв, проте аналіз її алгоритмів показує орієнтованість на спрощені патерни споживання. YAZIO успішно ідентифікує тип продукту, але не пропонує механізмів сегментації об'єктів у складних сценах. Без використання технологій інстанс-сегментації, таких як SAM, та без врахування фізичного масштабу через референтні об'єкти, система не здатна забезпечити точність, необхідну для медичних або професійних спортивних завдань. Це підтверджує актуальність розробки більш глибоких методів аналізу сцени, що поєднують розпізнавання та метричну оцінку [18].

1.4 Постановка задачі дослідження

На основі проведеного аналізу предметної області та критичного огляду існуючих методів і технологій автоматизованого моніторингу харчування було виявлено «метричний розрив» між ідентифікацією типу страви та точною оцінкою її маси. Існуючі комерційні рішення демонструють високу точність класифікації, проте залишаються вразливими до суб'єктивних похибок користувачів при визначенні розміру порцій, що критично для пацієнтів із метаболічними порушеннями та професійних атлетів.

Метою даної магістерської роботи є розроблення та програмна реалізація інтелектуальної системи автоматизованої оцінки енергетичної цінності страв на основі методів комп'ютерного зору, здатної забезпечувати високу метричну точність розрахунків у реальних умовах зйомки за допомогою використання природних референтних маркерів.

Для досягнення поставленої мети необхідно вирішити наступні наукові та практичні завдання:

- провести порівняльний аналіз сучасних архітектур нейронних мереж для задач детекції та інстанс-сегментації харчових об'єктів у складних сценах;
- розробити метод динамічного визначення масштабу зображення на основі ідентифікації та сегментації столових приборів як природних еталонних маркерів із відомими геометричними параметрами;
- сформувати інтегровану базу даних нутрієнтів, що включає не лише показники калорійності, а й фізичні характеристики продуктів, такі як щільність та коефіцієнти геометричної форми для апроксимації 3D-об'єму за 2D-проекцією;

- спроектувати та впровадити інтелектуальний контролер на основі нечіткої логіки для оцінки рівня довіри до отриманих результатів сегментації та автоматичної фільтрації аномальних значень;
- розробити програмну реалізацію конвеєра оброблення даних та провести експериментальну верифікацію точності розробленої системи на контрольній вибірці зображень наборів реальних страв.

2 ПІДГОТОВКА ТА АНАЛІЗ ДАНИХ ДЛЯ НАВЧАННЯ

2.1 Характеристика вихідних наборів даних

Ефективність систем комп'ютерного зору в нутриціології залежить від обсягу та якості розмітки навчальних вибірок. Набори даних класифікують на мало-, середньо- та великомасштабні (понад 100 000 зображень) [4]. До останніх належить Food-101 – індустріальний стандарт із 101 000 зображень для 101 класу. Через стратегію «слабкої розмітки» він містить візуальний шум, що потребує стійких архітектур типу ResNet чи WISeR, проте набір позбавлений геометричної розмітки об'єктів [16], [17].

Для більш точної локалізації страв та подолання «культурного» дисбалансу залучено набір UEC-Food256 (розширення японського UEC-Food100). Завдяки фільтрації класифікатором «Foodness» та Adaptive SVM валідність даних становить 88,12%. Ключова перевага цього набору – наявність верифікованих обмежувальних рамок, необхідних для розрахунку об'єму та енергетичної цінності порцій [19].

Залучення інших даних, як-от Food2K чи Nutrition5k, визнано недоцільним: Food2K потребує надмірних обчислювальних потужностей, а Nutrition5k має «доменний розрив» через ракурс зйомки суворо зверху, що не відповідає реальним фото зі смартфонів. Поєднання Food-101 та UEC-Food256 забезпечить баланс між масштабом класифікації та точністю геометричної розмітки, а дефіцит ознак компенсується аугментацією та трансферним навчанням на базі ImageNet [16].

Таким чином, фінальна вибірка є інтегрованою структурою, де відсутність геометрії у Food-101 компенсується прецизійною розміткою UEC-Food256. Це дозволило сформувати оптимальну таксономію з 334 класів для навчання ансамблевих архітектур. Технічні характеристики та еволюцію датасетів наведено у таблицях 2.1 та 2.2 [20].

Таблиця 2.1. Порівняльний аналіз популярних наборів даних для класифікації страв

Параметр порівняння	Food-101	UEC-Food256	Food2K [18]
Загальна кількість зображень	101 000	31 395	1 036 485
Кількість категорій	101	256	2 000
Середня кількість фото на клас	1 000	120	500
Геометрична розмітка	Відсутня	Наявна	Відсутня
Тип розмітки	Слабко-детермінована	Експертна / Валідована	Автоматизована
Основна кухня	Західна	Східноазіатська (Japanese)	Глобальна
Рівень шуму в мітках	Високий	Низький	Помірний
Роздільна здатність	512×512 px	Варіативна	Висока (HD)

Таблиця 2.2. Еволюція масштабів та характеристик ключових наборів даних для розпізнавання їжі [21]

Рік	Назва датасету	Кількість класів	Кількість зображень	Основна особливість / Внесок
2009	PFID	101	4 545	Перший структурований набір американської їжі
2012	UECFOOD-100	100	14 361	Впровадження локалізації об'єктів на фото
2014	Food-101	101	101 000	Створення великомасштабного стандарту галузі
2014	UECFOOD-256	256	25 088	Розширення ієрархії класів та покращення якості
2017	ChineseFoodNet	208	179 920	Охоплення специфіки китайської кухні
2019	FoodX-251	251	158 000	Використання для дрібнозернистої класифікації
2021	Food2K	2 000	1 036 564	Перехід до мільйонних масштабів (SOTA)

Формування бази знань системи базувалося на синтезі візуальних даних для навчання нейромереж та довідкових нутрієнтних метрик, необхідних для етапу фізичної реконструкції маси страви. Для забезпечення високої точності розрахунків енергетичної цінності було сформовано власну реляційну базу даних (`final_food_database_with_heights.csv`), куди увійшли показники калорійності на 100 грамів продукту, середня щільність та стандартна висота для кожної з 334 категорій страв. Основним джерелом кількісних даних про енергетичний склад продуктів виступила інтегрована система FoodData Central, розроблена Міністерством сільського господарства США (USDA) [22]. Цей ресурс є провідною світовою платформою, що акумулює дані п'яти різних типів, включаючи фундаментальні дослідження складу харчових продуктів та дані від виробників. Використання FoodData Central дозволило отримати верифіковані значення макронутрієнтів, що пройшли суворий лабораторний контроль, забезпечуючи надійну основу для розрахунку базових показників $kcal_{100g}$.

Для доповнення бази знань специфічними регіональними даними та стівами, що не представлені в американських реєстрах, було використано ресурси міжнародної асоціації European Food Information Resource (EuroFIR) [23]. EuroFIR є провідним європейським органом у сфері стандартизації даних про склад харчових продуктів, що об'єднує національні бази даних країн Європи. Цей ресурс забезпечив доступ до гармонізованих даних про енергетичну цінність складних страв, характерних для європейського ринку. Поєднання цих двох потужних міжнародних джерел дозволило мінімізувати похибки, викликані регіональними відмінностями у рецептурах, та сформувати репрезентативну базу знань з 334 класів страв.

Для вирішення фундаментальної задачі перетворення двовимірної сегментаційної маски об'єкта в об'ємну модель з подальшим розрахунком її маси, критично важливим є використання точних коефіцієнтів щільності. Оскільки візуальні дані не дозволяють безпосередньо оцінити вагу страви, у системі

реалізовано інтеграцію з базою даних FAO/INFOODS Density Database version 1.0 (July 2011) [24]. Цей ресурс, створений під егідою Продовольчої та сільськогосподарської організації ООН, містить стандартизовані значення щільності для широкого спектра харчових продуктів і складних страв.

У межах розробленої системи показники щільності були класифіковані за 5 основними фізико-структурними категоріями, що дозволило оптимізувати процес мапінгу ідентифікованих об'єктів до їхніх фізичних властивостей:

- твердотільні структури ($\rho \approx 1.04 - 1.07 \text{ г/см}^3$);
- рідкі середовища ($\rho \approx 1.0 - 1.05 \text{ г/см}^3$);
- вуглеводні та крупи ($\rho \approx 0.72 - 0.8 \text{ г/см}^3$);
- пористі борошняні вироби ($\rho \approx 0.29 - 0.7 \text{ г/см}^3$);
- листові суміші ($\rho \approx 0.45 \text{ г/см}^3$).

Кожна з цих категорій має специфічний вплив на алгоритм розрахунку. Для білкових структур (м'ясо, риба) характерна висока однорідність, що робить їх ідеальними для метричного аналізу. Рідкі страви вимагають врахування глибини посуду, тоді як для круп та локшини вводиться поправка на пористість між зернами. Найбільш динамічною групою є випічка, де через високий вміст повітря щільність може опускатися до 0.29 г/см^3 . Нарешті, для салатів застосовується коефіцієнт, що враховує нещільне прилягання компонентів, що критично важливо для збереження логіки розрахунків

Для забезпечення точності переходу від двовимірної сегментаційної маски до тривимірного об'єму в системі було впроваджено параметр `geometry_coeff`. Необхідність використання цього коефіцієнта зумовлена тим, що стандартний розрахунок об'єму шляхом простого множення площі маски на середню висоту створює ідеальну геометричну призму, яка часто не відповідає реальній морфології страви. Геометричний коефіцієнт виступає коректором, що відображає евклідове відношення об'єму реального 3D-примітива до об'єму описаної навколо нього

призми. Це дозволяє врахувати звуження, заокруглення або нерівномірність країв об'єкта, які неможливо детектувати на основі ортогональної проекції зображення.

Прикладом застосування цього підходу є аналіз хлібобулочних виробів сферичної форми, таких як `ball_shaped_bun_with_pork` або `donuts`. У цих випадках значення `geometry_coeff` встановлено на рівні 0.67. Математично це обґрунтовується відношенням об'єму сфери ($V_{sphere} = \frac{4}{3}\pi r^3$) до об'єму циліндра з аналогічним радіусом та висотою ($V_{cylinder} = 2\pi r^3$), де результат ділення $\frac{4}{6}$ дає приблизно 0.67. Отже, при розрахунку об'єму для заокруглених страв система автоматично відсікає 33% порожнього простору, який би був помилково врахований при лінійній екструзії маски [25].

Та навпаки, для страв з пласкою та однорідною структурою, таких як стейки або філе риби, геометричний коефіцієнт встановлено рівним 1.0. Це пояснюється тим, що такі об'єкти майже повністю заповнюють свій проекційний об'єм, і їхня форма максимально наближена до прямокутної або циліндричної призми. Впровадження диференційованого `geometry_coeff` дозволяє моделі значно знизити похибку розрахунку маси для об'єктів зі складною евклідовою геометрією [25].

Одним із найскладніших етапів формування бази знань стало визначення параметра стандартної висоти страви (`standard_height_cm`), оскільки міжнародні нутрієнтні бази даних, такі як USDA FoodData Central та EuroFIR, оперують виключно масою та хімічним складом, не надаючи геометричних характеристик об'єктів. Для подолання цього інформаційного дефіциту було застосовано комбінований підхід, що поєднує емпіричний аналіз кулінарних стандартів та зворотний математичний розрахунок на основі відомих фізичних величин. Більшість значень висоти була отримана шляхом систематичного огляду технологічних карт приготування страв та аналізу стандартних порцій у ресторанній практиці, що дозволило встановити середньостатистичні референтні значення для кожної категорії.

Для категорій страв з чітко регламентованими стандартами виробництва, зокрема для випічки та десертів, було застосовано метод дедуктивного розрахунку на основі даних бази USDA. Наприклад, для категорії «Apple Pie» висота була обчислена через відоме співвідношення маси порції та її геометричних меж. Оскільки база USDA надає точні дані про масу стандартного пирога m та його типовий діаметр d , а показник щільності ρ був отриманий з бази FAO/INFOODS, параметр висоти h визначався за формулою зворотного об'єму:

$$h = \frac{m}{\rho * \pi * \left(\frac{d}{2}\right)^2 * k}, \quad (5)$$

де k – геометричний коефіцієнт апроксимації.

Такий аналітичний підхід дозволив конвертувати сухі статистичні дані про вагу в прецизійні метричні константи, необхідні для роботи алгоритму реконструкції об'єму.

2.2 Розроблення алгоритму уніфікації та мапінгу

Об'єднання датасетів Food-101 та UEC-Food256 потребує спеціалізованого алгоритму уніфікації через розбіжності в таксономії та форматах метаданих. Його завданням є створення єдиного простору ознак, де кожному типу страви відповідає уніфікований ідентифікатор. Процес базується на семантичному аналізі назв, враховуючи, що Food-101 використовує назви директорій як мітки класів, а UEC-Food256 опирається на індексний файл category.txt.

Технічна реалізація на Python (бібліотека `pathlib`) включає сканування папок та парсинг дескрипторів. Для усунення конфліктів іменування застосовано протокол нормалізації: приведення до нижнього регістру, стандарт `snake_case` та видалення спецсимволів [8]. Це дозволяє через операції над множинами автоматично виявляти

дублікати страв, що перетинаються в обох наборах, як-от «hamburger», «pizza» або «rice».

Ключовим результатом є словник мапінгу `class_to_idx`, що зберігається у JSON-маніфесті. Згідно з рисунком 2.1, синтез 101 категорії Food-101 та 254 категорій UEC-Food256 дозволив сформувати реєстр із 334 унікальних страв. Файл `unified_classes.json` (наприклад, для «rice» встановлено ID 247) виступає «єдиним джерелом істини», що робить тренувальний пайплайн незалежним від фізичної структури датасетів та запобігає розсинхронізації етапів класифікації й локалізації об'єктів.

Окрім міток, алгоритм уніфікує геометричні дані. Оскільки Food-101 не має обмежувальних рамок, для нього автоматично призначається рамка на всю площу кадру або залучається механізм попередньої детекції. Це забезпечує реалізацію концепції мультизадачного навчання, де нейронна мережа одночасно оптимізується під ідентифікацію типу страви та регресію координат її меж [26].

```
Food-101: 101 класів.  
Food-256: 254 класів.  
Всього унікальних страв після злиття: 334  
Перевірка 'rice': новий ID = 247  
Файл збережено: C:\Users\nisha\PycharmProjects\MasterProject\data\unified_classes.json
```

Рисунок 2.1 – Консольний вивід результатів синтезу та уніфікації наборів даних Food-101 та UEC-Food256

2.3 Проблема дисбалансу та методи її вирішення

Об'єднання Food-101 та UEC-Food256 призводить до створення чисельного дисбалансу між 334 категоріями страв. Значне переважання мажоритарних класів

(наприклад, фаст-фуд, гарніри) над міноритарними специфічними стравами призводить до алгоритмічного упередження: модель демонструє високу точність на популярних об'єктах, але фактично втрачає здатність розпізнавати рідкісні через мінімізацію функції втрат переважно за рахунок домінуючих груп.

Для подолання нерівномірного розподілу застосовано стратегію динамічного балансування вибірки під час формування навчальних пакетів. Технічна реалізація базується на використанні `WeightedRandomSampler`, який коригує ймовірність вибору конкретного зразка. Математична логіка полягає у присвоєнні кожному класу ваги, обернено пропорційної до кількості зображень у ньому [25]. Отриманий вектор ваг дозволяє мережі взаємодіяти з рідкісними класами значно частіше, ніж вони представлені у вихідних даних, що вирівнює внесок кожної категорії в оновлення градієнтів.

Обчислення ваги для кожної категорії c здійснюється за наступною залежністю [27]:

$$W_c = \frac{1}{N_c}, \quad (6)$$

де N_c – загальна кількість зображень у межах конкретного класу.

Отримані значення нормалізуються та трансформуються у вектор ваг для кожного окремого зразка в датасеті. Під час кожної епохи навчання механізм `WeightedRandomSampler` використовує ці ваги для формування репрезентативних батчів. Це дозволяє нейронній мережі взаємодіяти з рідкісними класами набагато частіше, ніж вони зустрічаються у «сирих» даних, що фактично вирівнює внесок кожної категорії у процес оновлення градієнтів.

Такий підхід є критично важливим для стабільної збіжності ансамблевих архітектур (ResNet та DenseNet). Замість адаптації до візуально поширених патернів, модель змушена вивчати унікальні ознаки міноритарних класів. У результаті зваженого семплювання досягнуто рівномірного розподілу точності між 334 класами, що мінімізує ризик хибних спрацювань на користь популярних

категорій та підвищує загальну надійність інтелектуальної системи в реальних умовах експлуатації.

2.4 Попереднє оброблення та аугментація зображень

Для забезпечення стабільної роботи запропонованої ансамблевої моделі та підвищення її здатності до генералізації на реальних даних, усі вхідні зображення проходять через багатоетапний конвеєр трансформацій. Це дозволяє привести гетерогенні дані до єдиного стандарту та штучно розширити навчальну вибірку.

Алгоритм підготовки зображень включає наступні кроки:

- геометрична стандартизація та масштабування;
- просторова аугментація даних;
- корекція колірних та експозиційних характеристик;
- статистична нормалізація вхідних тензорів.

На кроці геометричної стандартизації та масштабування застосовується метод `RandomResizedCrop(224)`, який виконує дві функції: приведення зображення до фіксованого розміру 224x224 пікселів, що є обов'язковим для архітектур ResNet та DenseNet, та випадкове кадрування. Такий підхід змушує нейронну мережу розпізнавати страву за її фрагментами та текстурами, що робить систему стійкою до зміни відстані зйомки або наявності зайвих об'єктів на фоні [28].

Для імітації різних ракурсів використання мобільного додатка застосовуються випадкові перевороти по горизонталі (`RandomHorizontalFlip`) та повороти зображення на кут до 15 градусів (`RandomRotation`). Оскільки семантичний зміст страви не змінюється при дзеркальному відображенні або невеликому нахилі, ці

трансформації дозволяють моделі вивчати інваріантні просторові ознаки, значно знижуючи ризик перенавчання на конкретних ракурсах із датасету [29].

Крок колірної корекції реалізовано за допомогою інструменту ColorJitter, який випадковим чином змінює яскравість та контрастність зображення у діапазоні до 20%. Це критично важливо для обробки фотографій їжі, зроблених у різних умовах освітлення – від яскравого денного світла до темного штучного світла в інтер'єрах закладів харчування. Модель вчиться ігнорувати абсолютні значення яскравості, фокусуючись на кольорових поєднаннях та контурах інгредієнтів [28].

На останньому кроці зображення перетворюються у числові тензори та проходять процедуру нормалізації за каналами RGB. Використовуються середні значення (0.485, 0.456, 0.406) та стандартні відхилення (0.229, 0.224, 0.225), що відповідають параметрам набору ImageNet. Це забезпечує коректну роботу механізму Transfer Learning, оскільки початкові ваги ансамблевої моделі були розраховані на вхідні сигнали саме з таким статистичним розподілом.

3 ПРОЄКТУВАННЯ ТА РЕАЛІЗАЦІЯ АРХІТЕКТУРИ СИСТЕМИ

3.1 Обґрунтування та побудова ансамблевої архітектури

Вибір архітектури для класифікації зображень страв ускладнюється високою внутрішньокласовою варіативністю та низькою міжкласовою дистанцією з причини схожості візуальних ознак (наприклад, різних супів чи рагу). Для вирішення цих викликів розроблено ансамблеву архітектуру, що поєднує дві комплементарні моделі: ResNet50 та DenseNet121.

Вибір ResNet50 обґрунтований використанням залишкових зв'язків, що дозволяють градієнтам проходити крізь глибокі шари, запобігаючи їх згасанню. У задачах розпізнавання їжі модель ефективно виділяє глобальні геометричні ознаки та загальну структуру об'єктів, що критично для ідентифікації страв із чіткою формою [30].

Паралельна гілка на базі DenseNet121 відрізняється щільними з'єднаннями, де кожен шар отримує дані від усіх попередніх у межах блоку. Це забезпечує повторне використання ознак та високу чутливість до дрібних текстур. Для нутриціології це має вирішальне значення, адже DenseNet точніше розпізнає дрібні інгредієнти (насіння, нарізані овочі, соуси) завдяки збереженню низькорівневої інформації на глибоких шарах [31].

Синергія моделей реалізована через механізм усереднення логітів після адаптації моделей під уніфіковану таксономію із 334 класів. Під час прямого проходу вихідні вектори ймовірностей обох мереж сумуються та нормалізуються. Такий підхід дозволяє ансамблю нівелювати індивідуальні помилки: одна мережа коригує невпевненість іншої (наприклад, через складне освітлення) на основі вивчених текстурних ознак. Результати тестування підтверджують зростання

впевненості ансамблю порівняно з поодинокими архітектурами, що є ключовим для стабільної роботи фінального продукту.

Математично фінальний прогноз Y_{final} розраховується за формулою [13]:

$$Y_{final} = \frac{1}{2}(f_{ResNet}(x) + f_{DenseNet}(x)), \quad (7)$$

де x вхідне зображення;

$f(x)$ вихідні вектори логітів відповідних моделей.

3.2 Порівняльний аналіз сучасних архітектур глибокого навчання

Для вибору оптимальної моделі розпізнавання страв проаналізовано найбільш ефективні архітектури CNN. Кожна з них пропонує унікальні рішення для підвищення точності або оптимізації обчислень.

VGG16, переможець ILSVRC 2014, використовує однорідну структуру з 3x3 згортковими фільтрами та функцію активації ReLU [32]:

$$f(x) = \max(0, x). \quad (8)$$

Попри високу точність (Top-5 92,7%), значна кількість параметрів ускладнює її впровадження у мобільні системи.

Архітектура MobileNetV2 адаптована для вбудованих пристроїв і базується на концепції Inverted Residuals та Linear Bottlenecks. Використання глибинних та точкових згортк суттєво знижує кількість операцій без критичної втрати точності [17].

EfficientNet-B0 використовує метод составного масштабування, де глибина, ширина мережі та роздільна здатність змінюються узгоджено за допомогою єдиного коефіцієнта. Це дозволяє моделі випереджати значно більші мережі за меншої кількості параметрів [33].

Основою роботи є ResNet50, що впровадила концепцію Residual Learning. Використання «shortcut connections» забезпечує стабільне навчання надглибоких структур, а дизайн «пляшкового горла» (згортки 1x1, 3x3, 1x1) гарантує ефективне вилучення ознак. Порівняльний аналіз архітектур наведено в таблиці 3.1.

Незважаючи на швидкість MobileNetV2 та ефективність EfficientNet, обрано ансамбль ResNet50 та DenseNet121. ResNet стабільно ідентифікує глобальні форми страв, тоді як DenseNet краще працює з текстурними інгредієнтами. Такий гібридний підхід нівелює індивідуальні недоліки моделей і забезпечує точність, необхідну для роботи з 334 класами [32], [33].

Таблиця 3.1. Порівняння архітектур нейронних мереж за технічними параметрами

Архітектура	Ключова технологія	Основна перевага	Обмеження для проєкту
VGG16	Згортки 3x3	Простота та стабільність	Надмірна кількість параметрів
MobileNetV2	Inverted Residuals	Висока швидкість на смартфонах	Нижча точність на складних текстурах
EfficientNet	Compound Scaling	Оптимальна ефективність	Складність налаштування ансамблю
ResNet50	Residual Connections	Стабільність градієнтів	Потребує значних ресурсів пам'яті

3.3 Стратегія багатоетапного навчання та обґрунтування вибору гіперпараметрів

Процес навчання нейромережових моделей, інтегрованих у систему оцінки калорійності, базувався на розробленій та апробованій у попередніх дослідженнях концепції поетапного перенесення знань. Використання такої стратегії зумовлене

специфікою задачі: необхідно адаптувати глибокі архітектури, попередньо навчені на загальних наборах даних (ImageNet), до вузькоспеціалізованого домену кулінарних страв, де ключову роль відіграють мікротекстури та складні морфологічні ознаки їх інгредієнтів.

Замість стандартного одноразового навчання всіх параметрів було впроваджено триступеневу стратегію адаптації, яка дозволяє уникнути деградації ваг та забезпечує стабільну конвергенцію при роботі з 334 категоріями страв.

На першому етапі, визначеному як фаза адаптації, було застосовано повне заморожування усіх шарів екстрактора ознак. Навчанню підлягав лише фінальний повнозв'язний шар (Dense/FC Layer) [34], адаптований під кількість цільових класів системи. Основною метою цього кроку було узгодження випадково ініціалізованих ваг нового класифікатора з розподілом ознак у датасеті при високій швидкості навчання 10^{-3} . Такий підхід запобігає поширенню агресивних початкових градієнтів через глибину мережі, що захищає високорівневі візуальні абстракції, отримані під час попереднього навчання, від «катастрофічного забування».

Другий етап передбачав розморожування верхніх структурних блоків мережі, які відповідають за формування складних семантичних ознак. Зокрема, для архітектури ResNet-50 було розморожено 4-й рівень, а для DenseNet-121 – четвертий щільний блок. Швидкість навчання було знижено до 10^{-4} для забезпечення делікатної корекції ваг. Це дозволило моделі уточнити параметри візуальних фільтрів, що ідентифікують специфічні кулінарні патерни (текстури соусів, ступінь обробки продуктів, характерні зрізи овочів), зберігаючи при цьому стабільність нижніх шарів, що розпізнають базові геометричні примітиви (лінії, контури).

Фінальна стадія навчання передбачала повне розморожування всіх параметрів ансамблю. Використання мінімального кроку навчання 10^{-5} дозволило провести прецизійне налаштування всієї ієрархії фільтрів. На цьому етапі відбувається

фінальна синергія між усіма шарами мережі, що забезпечує максимальну точність класифікації навіть для візуально схожих категорій страв.

Протягом виконання усіх етапів у якості функції втрат використовувалася багатокласова крос-ентропія, а оптимізація здійснювалася за допомогою алгоритму Adam, який забезпечує адаптивну корекцію швидкості навчання для кожного параметра індивідуально. Параметри конфігурації для кожного етапу систематизовано у таблиці 3.2.

Таблиця 3.2 – Параметри гіперпараметрів за етапами навчання моделей

Етап навчання	Об'єкт розморожування	Learning Rate	Оптимізатор	Batch Size	Епохи
Adaptation	Лише Classifier / FC	$1 * 10^{-3}$	Adam	32	3
Partial Fine-tuning	Last Blocks + Classifier	$1 * 10^{-4}$	Adam	32	5
Full Fine-tuning	Усі шари (All layers)	$1 * 10^{-5}$	Adam	32	10

3.4 Програмно-апаратне середовище та інструментарій реалізації

Розробка та навчання ансамблевої моделі для оцінки калорійності страв вимагали створення стабільного програмно-технічного середовища, здатного забезпечити високу продуктивність при обробці великих масивів візуальних даних. Основним середовищем розробки було обрано мову програмування Python, яка є стандартом у галузі машинного навчання завдяки розвиненій екосистемі спеціалізованих бібліотек. Як базовий фреймворк глибинного навчання використано PyTorch, що забезпечує гнучкість у побудові динамічних обчислювальних графів та надає широкі можливості для реалізації складних архітектур, таких як ансамблі нейронних мереж. Вибір PyTorch також обумовлений

наявністю пакету Torchvision, який містить переднавчені ваги для архітектур ResNet50 та DenseNet121, а також потужний інструментарій для аугментації та трансформації зображень у режимі реального часу.

Ключовою особливістю програмної реалізації є використання бібліотеки `torch_directml`, що дозволило задіяти апаратне прискорення графічного процесора через інтерфейс DirectX 12. На відміну від класичного підходу з використанням NVIDIA CUDA, технологія DirectML забезпечує крос-платформну сумісність та можливість тренування моделей на графічних адаптерах різних виробників, що робить рішення більш універсальним. Для забезпечення стабільної роботи DirectML у коді було впроваджено специфічні налаштування, такі як вимкнення режиму `foreach` в оптимізаторі, що запобігає виникненню критичних помилок при паралельних обчисленнях ваг на специфічному обладнанні [35].

Апаратна частина середовища включає обчислювальну станцію, обладнану багатоядерним процесором та дискретним графічним адаптером із підтримкою DirectML, що дозволило обробляти пакети даних розміром 16 одиниць при роздільній здатності зображень 224x224 пікселі. Для керування даними та структурованого зберігання інформації про класи використовувався формат JSON, а бібліотека Pathlib забезпечила коректну роботу зі шляхами до файлів незалежно від операційної системи. Робота з тензорними обчисленнями та статистичною обробкою даних, зокрема розрахунок ваг для `WeightedRandomSampler`, виконувалася за допомогою бібліотеки NumPy, що гарантує високу швидкість математичних операцій над масивами [35].

Технічні характеристики програмно-апаратного середовища наведено в таблиці 3.3.

Таблиця 3.3 – Технічні характеристики програмно-апаратного комплексу

Категорія	Найменування інструментарію / Характеристика	Призначення
Мова програмування	Python 3.12.10	Основна логіка та скрипти навчання
Фреймворк ML	PyTorch 2.4.1	Побудова та тренування нейромереж
Обробка зображень	Torchvision	Трансформації та аугментація даних
Апаратне прискорення	DirectML (via torch_directml)	Використання GPU через DirectX 12
Математичні обчислення	NumPy	Робота з масивами та статистика
Керування даними	JSON, Pathlib	Структурування класів та файлова система
Архітектури моделей	ResNet50, DenseNet121	Базові екстрактори ознак (Backbones)
Оптимізація GPU	Adam (foreach=False)	Стабільність обчислень у середовищі DirectML

3.5 Розроблення концепції метричного калібрування сцени на основі YOLO-детекції та її вдосконалення методами інстанс-сегментації еталонних об'єктів

Фундаментальним обмеженням при аналізі двовимірних зображень страв є втрата просторової інформації про реальні фізичні розміри об'єктів. Через варіативність відстані від камери до об'єкта зйомки та кута огляду, одна й та сама порція їжі може мати різну площу в піксельному еквіваленті. Без розв'язання проблеми масштабування неможливо здійснити перехід від візуальної класифікації страви до кількісної оцінки її об'єму та енергетичної цінності.

Для вирішення цієї задачі в архітектуру системи впроваджено концепцію метричного калібрування, що базується на використанні «метричних анкерів» –

об'єктів побутового вжитку з відомими геометричними параметрами. Оскільки столові прибори (виделки, ножі, ложки) мають уніфіковані розміри, їх обрано як еталони для динамічного визначення масштабного коефіцієнта сцени.

На початкових етапах розробки як основний модуль детектування еталонів використовувалася архітектура YOLOv8x. Цей підхід базувався на визначенні масштабного коефіцієнта через розміри обмежувальної рамки (Bounding Box). Проте практичне тестування на реальних зображеннях виявило низку критичних недоліків методу YOLO:

- похибка «фонового шуму» - обмежувальна рамка неминуче включає порожній простір навколо приладу, тіні або частини тарілки, що призводить до систематичного завищення значень;
- чутливість до ракурсу - при нахилі камери прямокутна рамка некоректно відображає реальну довжину приладу, що спричиняє нелінійне зростання похибки при розрахунку площі;
- нестабільність при перекритті: якщо частина виделки перекрита їжею, YOLO все одно намагається побудувати рамку за видимою частиною, що критично викривлює масштаб.

З огляду на ці недоліки, було здійснено перехід до ансамблю моделей DETR та SAM [34, 36].

Вибір фінальної архітектури детектора базувався на результатах проведеного порівняльного аналізу між моделями сімейства YOLO та трансформерними архітектурами типу DETR. Це дозволило обґрунтувати перехід до використання механізмів глобальної уваги як більш ефективного інструменту для стабільної локалізації еталонів у сценах зі складним геометричним фоном.

Вибір DETR зумовлений його здатністю краще розпізнавати об'єкти в складних сценах завдяки механізмам глобальної уваги, що значно підвищило ймовірність успішного детектування еталонів у кадрі порівняно з YOLO.

У свою чергу модель SAM забезпечує перехід від грубих рамок до попиксельних масок. Це дозволяє системі отримати прецизійний контур металевої поверхні приладу, відсікаючи фоновий шум та тіні ще на етапі сегментації. Хоча проблеми ракурсу та перекриття залишаються актуальними для 2D-зображень, використання попиксельної маски SAM надає максимально «чисті» вхідні дані для алгоритмів геометричної корекції та інтелектуальної фільтрації на основі нечіткої логіки. Такий підхід дозволив мінімізувати аномальні стрибки масштабу та забезпечити стабільність системи, необхідну для точного контролю енергетичного балансу [35].

Також для практичної реалізації розробленої концепції було сформовано базу еталонних лінійних параметрів об'єктів-анкерів. Як джерело нормативних значень було обрано дані платформи NorecaTraders [37], яка спеціалізується на професійному забезпеченні сектору HoReCa (Hotel, Restaurant, Catering). Вибір цього джерела є обґрунтованим, оскільки професійне столове приладдя проходить сувору сертифікацію та відповідає міжнародним стандартам розмірності для забезпечення ергономіки та сумісності у сфері громадського харчування. Використання характеристик професійного інвентарю дозволяє системі спиратися на найбільш розповсюджені «середньостатистичні» габарити, які зустрічаються у більшості сучасних закладів харчування.

У процесі розробки алгоритму калібрування було закладено наступні константи довжини та ширини [37]:

- виделка столова з довжиною 20.0 см та шириною робочої частини 2.3 см;
- ніж столовий з довжиною довжина 21.0 см та шириною леза 2.2 см;
- ложка столова з довжиною довжина 20.0 см та шириною 4.2 см;
- салатниця (bowl) зі стандартним діаметром 18.0 см.

4 АНАЛІЗ РЕЗУЛЬТАТІВ КЛАСИФІКАЦІЇ ТА ТЕСТУВАННЯ

4.1 Методологія фінального оцінювання та метрики якості

Для об'єктивного визначення ефективності розробленого ансамблю нейронних мереж була впроваджена багаторівнева методологія тестування, що враховує специфіку роботи з великою кількістю класів (334 категорії). Ключовою особливістю оцінювання стало використання стратегії відкладеної валідації на фінальному етапі навчання. На відміну від стандартних підходів, де розподіл даних відбувається на самому початку, у даній роботі поділ на тренувальну та тестову вибірки у співвідношенні 85:15 був реалізований безпосередньо на 5-й епосі глибокого тонкого налаштування (Fine-tuning). Це дозволило моделі на попередніх ітераціях використати максимальний обсяг візуальної інформації для формування стійких ознак, а на фінальному кроці – пройти перевірку на значній кількості раніше не бачених об'єктів. Для забезпечення відтворюваності експерименту та уникнення зміщення вибірки було використано генератор випадкових чисел із фіксованим станом seed 42, а функція WeightedRandomSampler гарантувала пропорційне представлення всіх категорій страв у кожному пакеті даних під час валідації.

Основним критерієм якості класифікації було обрано метрику точності (Accuracy), яка визначає частку правильно розпізнаних страв відносно загальної кількості поданих на вхід зображень. Математично цей показник розраховувався як відношення кількості вірних передбачень $N_{correct}$ до загальної кількості тестів N_{total} [38]:

$$Accuracy = \frac{N_{correct}}{N_{total}} * 100\%. \quad (9)$$

З огляду на те, що ймовірність випадкового вгадування при наявності 334 взаємовиключних класів становить лише 0.3%, показник точності, отриманий ансамблем, є критично важливим індикатором успішності навчання. Окрім

загальної точності, під час тестування проводився моніторинг функції втрат Cross-Entropy Loss. Ця метрика дозволяє оцінити не просто факт вгадування класу, а впевненість моделі у своїх прогнозах, вимірюючи розбіжність між передбаченим розподілом ймовірностей та фактичними мітками. Чим меншим є значення цієї функції на етапі тестування, тим стабільнішою та точнішою вважається інтелектуальна система.

Окрім базової точності, для глибинного аналізу здатності ансамблю до розпізнавання було використано метрику Top-5 Ассурасу. Вона визначає частку випадків, коли правильна мітка класу потрапила до п'ятірки найбільш ймовірних прогнозів моделі. Оскільки візуальні характеристики багатьох страв у розширеному наборі з 334 класів мають високу ступінь подібності, метрика Top-5 дозволяє оцінити «ерудованість» системи. Високий показник Top-5 при помірному Top-1 свідчить про те, що модель коректно визначає групу продуктів, але стикається з труднощами при виборі конкретного підвиду через візуальну схожість інгредієнтів.

Математично Top-5 Ассурасу розраховується як [39]:

$$\text{Top}_{5_{acc}} = \frac{1}{N} \sum_{i=1}^N 1(y_i \in \text{top}_5(\hat{y}_i)), \quad (10)$$

де 1 – індикаторна функція, що дорівнює 1, якщо істинний клас y_i входить до 5 найбільш ймовірних прогнозів $\hat{y}_i$.

Використання цієї метрики дозволяє стверджувати, що розроблена система може бути використана в інтерактивному режимі, пропонуючи користувачеві перелік найбільш імовірних варіантів для остаточного підтвердження.

Додатково запропонована методологія включала контроль за обчислювальною стабільністю в середовищі DirectML. Оцінювання проводилося при вимкненому параметрі foreach в оптимізаторі, що забезпечило ідентичність результатів на різних апаратних ітераціях та запобігло втраті точності через специфіку паралельної обробки тензорів на графічному процесорі.

Використання такої комплексної системи метрик та методів контролю дозволило мінімізувати вплив випадкових факторів та отримати статистично достовірні дані про продуктивність розробленого програмного рішення.

4.2 Порівняльний аналіз точності ансамблю та його базових архітектур

Ключовим завданням експериментального етапу було дослідження динаміки навчання та порівняння ефективності двох базових архітектур ResNet-50 та DenseNet-121 у межах розробленої триступеневої стратегії донавчання. Аналіз проводився на валідаційній вибірці набору даних Food-101 протягом 17 ітераційних епох, що дозволило відстежити вплив поступового розморожування шарів на загальну точність класифікації. Результати моніторингу точності (Accuracy) для кількості епох подано на рисунку 4.1.

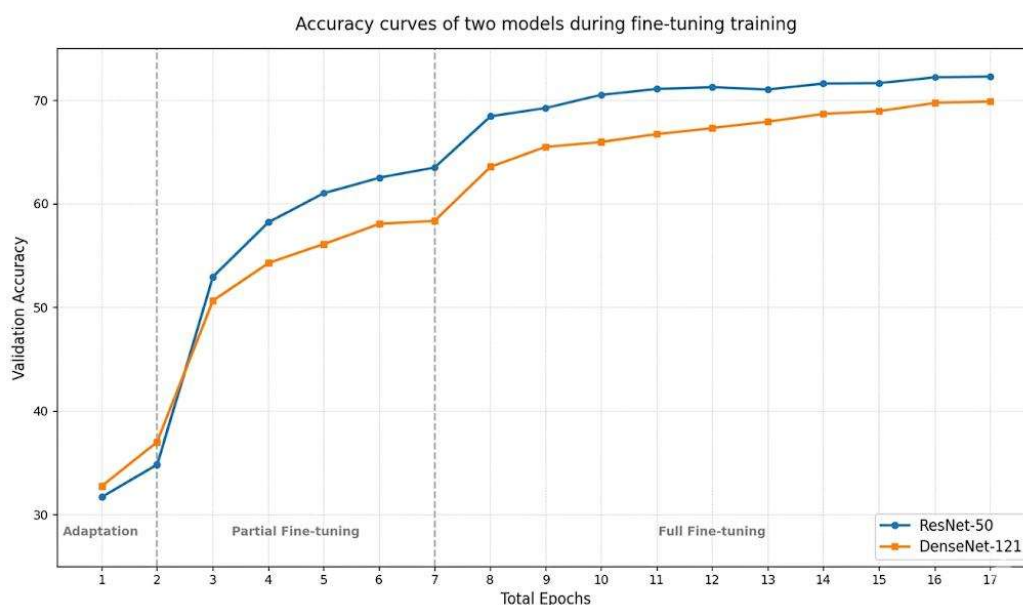


Рисунок 4.1 – Залежність точності розпізнавання базових моделей на етапах донавчання від кількості епох

Як видно з наведеного графіка (див. рис. 4.1), процес навчання чітко розділений на три функціональні фази, кожна з яких характеризується специфічною динамікою конвергенції.

На етапі адаптації (Adaptation, епохи 1–2) навчався лише фінальний класифікатор при замороженому екстракторі ознак, обидві моделі продемонстрували стартову точність у діапазоні 32-37%. Незважаючи на низькі абсолютні показники, спостерігається стабільне зростання, що підтверджує успішну ініціалізацію ваг «голови» моделі під розподіл класів цільового датасету. Модель DenseNet-121 на цьому етапі показала дещо вищу стабільність порівняно з ResNet-50.

На етапі часткового донавчання (Partial Fine-tuning, епохи 3–7) відбувся перехід до другої фази (епоха 3), супроводжуваний різким стрибком точності для обох архітектур. Це зумовлено розморожуванням верхніх структурних блоків, що дозволило мережам адаптувати високорівневі візуальні фільтри до специфіки зображень страв. Саме на цьому етапі архітектура ResNet-50 почала демонструвати кращу динаміку, випередивши DenseNet-121 і досягнувши показника понад 60% до сьомої епохи. Характерне сповільнення росту на 6–7 епохах вказує на вичерпання потенціалу оптимізації при частково заморожених параметрах.

На етапі повного донавчання (Full Fine-tuning, епохи 8–17) відбулося остаточне розморожування всіх шарів мережі, яке на восьмій епосі ініціювало чергову фазу інтенсивного зростання. Завдяки мінімальному кроку навчання 10^{-5} вдалося провести прецизійне налаштування всієї ієрархії ознак. Модель ResNet-50 впевнено утримувала лідерство протягом усього етапу, продемонструвавши підсумкову точність на рівні 72.4%. Модель DenseNet-121, попри дещо повільнішу динаміку, також досягла стабільно високого результату – близько 70%. Стабілізація кривих після 15-ї епохи свідчить про повну конвергенцію алгоритму та досягнення оптимального енергетичного стану ваг для обох моделей.

Для фінальної оцінки якості класифікації було проведено порівняльний аналіз метрик точності (Precision), повноти (Recall) та $F1$ -score для кожної базової моделі окремо та для розробленого гібридного ансамблю. Результати тестування на контрольній вибірці набору даних Food-101 представлені на рисунку 4.2.

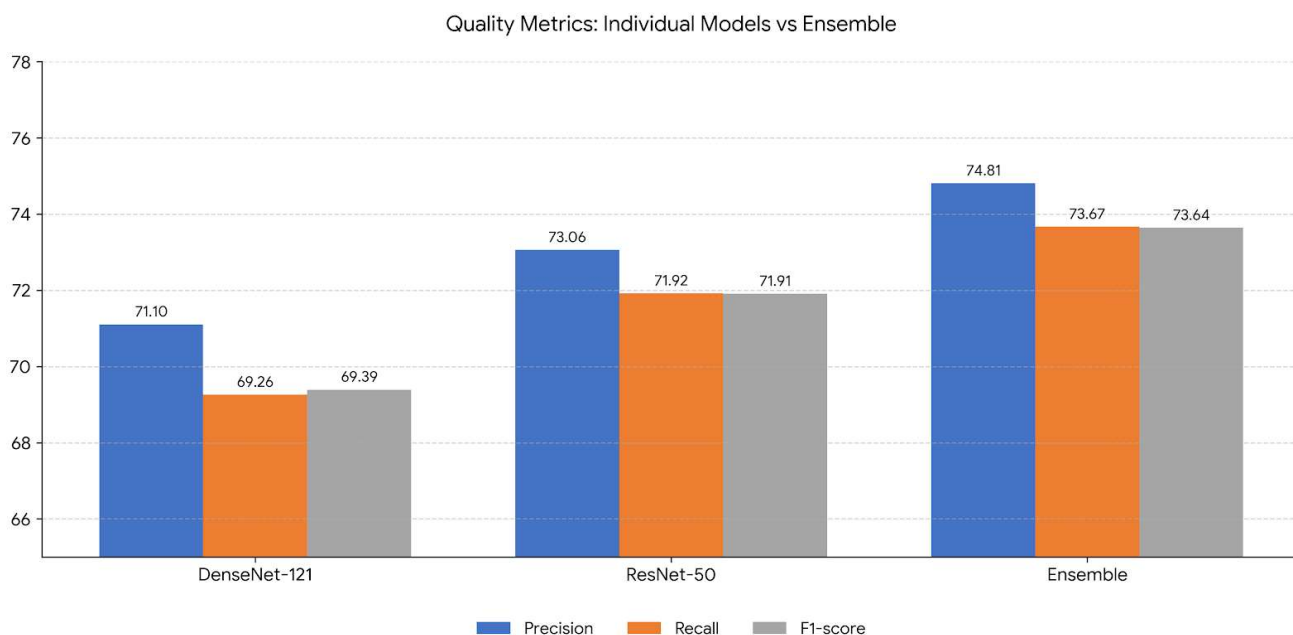


Рисунок 4.2 – Порівняльний аналіз зважених метрик якості для поодиноких моделей та ансамбля

Аналіз отриманих експериментальних даних дозволяє стверджувати, що модель ResNet-50 продемонструвала вищі показники ефективності в усіх категоріях порівняно з архітектурою DenseNet-121. Досягнення показників точності (Precision) на рівні 73.06% та значення $F1$ – score 71.91% повністю обґрунтовують використання саме цієї моделі як мажоритарного компонента ансамблю з ваговим коефіцієнтом 0.8. При цьому спостерігається виражений позитивний ефект ансамблювання, оскільки гібридна модель, заснована на зваженому голосуванні результатів обох мереж, показала найкращі результати за всіма контрольними метриками. Зокрема, підсумковий

показник точності зріс до 74.81%, що забезпечило приріст у 1.75% порівняно з найкращою одиночною архітектурою. Висока збалансованість та робастність розробленої системи підтверджується фінальним значенням F1-score для ансамблю на рівні 73.64%, що свідчить про стабільність інференсу та мінімізацію ймовірності виникнення хибнопозитивних або хибнонегативних прогнозів у реальних умовах експлуатації.

Детальний аналіз отриманих результатів дозволяє сформувати загальну картину ефективності розробленої системи та обґрунтувати кожен аспект обраної архітектурної конфігурації.

Порівняння базових моделей DenseNet-121 та ResNet-50 демонструє, що попри спорідненість архітектур, ResNet-50 має вищу здатність до вилучення складних візуальних ознак, що підтверджується вищими показниками точності та повноти. Зокрема, перевага ResNet-50 у точності майже на 3% пояснюється ефективнішим використанням залишкових зв'язків, які дозволяють глибше проникати в ієрархію ознак без ризику затухання градієнта. Це особливо важливо для розпізнавання страв із низькою візуальною контрастністю інгредієнтів, де текстурні нюанси відіграють вирішальну роль у класифікації (таблиця 4.1).

Таблиця 4.1 – Зведені результати порівняльного аналізу ефективності базових моделей та гібридного ансамблю на об'єднаних датасетах Food-101 та Food-256

Модель / Конфігурація	Accuracy (%)	F1-score (%)	FPS (f/s)	MAE	MSE
DenseNet-121	69,26	69,24	147,10	0,002398	0,4200
ResNet-50	71,92	71,91	149,92	0,002123	0,3876
Ensemble	73,71	73,64	74,25	0,002260	0,3671

Впровадження гібридного ансамблю стало ключовим етапом оптимізації, що дозволило подолати індивідуальні обмеження поодиноких мереж. Зростання підсумкової точності до 73.71% свідчить про виражений ефект статистичного

усереднення помилок. Оскільки DenseNet та ResNet мають різні принципи поширення ознак, вони роблять різні типи помилок на маргінальних випадках. Метод зваженого голосування дозволяє системі ігнорувати аномальні прогнози однієї мережі, якщо інша демонструє вищу впевненість. Таке зниження дисперсії помилок наочно ілюструється показником MSE [40], який в ансамблі зменшився до 0.3671, що є найнижчим значенням серед усіх протестованих конфігурацій. Це означає, що система стала значно надійнішою у запобіганні «грубим» помилкам класифікації, які могли б критично викривити фінальний розрахунок енергетичної цінності.

Окремого розширеного аналізу потребує питання продуктивності, виражене в частоті обробки кадрів. Попри те, що перехід від одиночних моделей до ансамблю спричинив зниження швидкості з приблизно 150 до 74.25 кадрів на секунду, цей компроміс є стратегічно виваженим рішенням. У контексті мобільних систем моніторингу харчування швидкість понад 60 FPS є надлишковою, оскільки користувач взаємодіє зі статичними зображеннями або відеопотоком низької динаміки. Вивільнений ресурс обчислювальної потужності було ефективно конвертовано в приріст точності на 1.79% порівняно з найкращою базовою моделлю. В задачах підтримки здоров'я та контролю метаболічних показників, де кожен відсоток похибки може призвести до накопичувального дисбалансу нутрієнтів протягом дня, така пріоритизація точності над надвисокою швидкістю є єдиним ефективним підходом.

Більше того, показник 74.25 FPS гарантує стабільну роботу додатка в режимі реального часу без жодних затримок, помітних для людського ока. Це дозволяє реалізувати інтерактивні функції, такі як підсвічування об'єктів під час наведення камери, зберігаючи при цьому найвищу можливу надійність розрахунків. Зниження похибки MAE та MSE безпосередньо корелює зі зменшенням ризику неправильного підбору інсулінової дози або калорійності раціону, що є головною ціллю даної

розробки. Таким чином, розроблений ансамбль являє собою збалансоване рішення, де математична складність та обчислювальні витрати повністю виправдані якісним стрибком у надійності функціонування системи [41].

У ході порівняльного аналізу встановлено, що гібридний ансамбль досяг точності 73,71% (Top-1), що перевершує результати відомих ансамблевих архітектур, таких як FoodNet, які демонструють 72,12% на наборі Food-101 та 73,50% на Indian Food Database. Розглянута задача є технічно складнішою, оскільки система оперує 334 категоріями страв, тоді як більшість аналогів обмежені стандартними 101 або 256 класами. Попри таку таксономічну насиченість, завдяки оптимізації DirectML вдалося досягти швидкодії у 74,25 FPS, що значно перевищує показники важких Transformer-моделей (наприклад, FoodCSWin з точністю до 85%, але без фокусу на Real-time), роблячи рішення найбільш збалансованим для практичного використання [25], [42], [43].

4.3 Аналіз помилок класифікації та проблемних категорій зображень

Для глибокого розуміння роботи розробленого ансамблю було проведено аналіз помилок на випадковій репрезентативній вибірці — 20% від загального обсягу об'єднаного набору даних. Це дозволило виявити категорії страв, які демонструють найнижчу точність класифікації, та встановити причини виникнення хибних спрацювань. Результати аналізу 10 найбільш складних категорій наведено в таблиці 4.2.

Детальний розгляд отриманих результатів дозволяє виділити кілька ключових чинників, що негативно впливають на загальну точність розпізнавання. Одним із найвагоміших факторів є висока міжкласова візуальна подібність (Inter-class

similarity), що найбільш виражено проявляється у парі страв «pork_chop» та «steak». Оскільки обидва класи представлені цільними шматками смаженого м'яса зі схожою текстурою та кольоровою гамою, ансамбль нейронних мереж часто не здатний виділити унікальні дескриптори, необхідні для диференціації свинини та яловичини виключно за візуальними ознаками. Аналогічна тенденція простежується у категоріях «filet_mignon» та «foie_gras», які через подібність форми та способу термічної обробки нерідко класифікуються як узагальнений клас «steak».

Таблиця 4.2 – Топ-10 категорій страв з найнижчим значенням Accuracy

Назва категорії страви	Точність (Accuracy), %	Кількість зразків (count)	Основна категорія плутанини
green_salad	26.00	50	fried_fish
rice	27.36	106	green_salad
pork_chop	43.67	158	steak
grilled_salmon	49.33	150	steak / pork_chop
chicken_curry	50.00	162	beef_curry
foie_gras	50.68	148	steak
miso_soup	50.79	254	ramen_noodle
filet_mignon	52.00	150	steak
omelette	52.20	159	scrambled_egg
tuna_tartare	57.93	145	ceviche

Іншими вагомими факторами є контекстна залежність та наявність візуального шуму на зображеннях. Аномально низькі показники точності для категорії «green_salad» та її системна плутанина з «fried_fish» зумовлені специфікою вихідних тренувальних вибірок, зокрема набору Food-101. На багатьох знімках смаженої риби значну частину кадру займає об'ємний гарнір у вигляді свіжого листя салату, що «дезорієнтує» модель під час навчання. В результаті інтелектуальна система помилково ідентифікує салат як основний об'єкт або, навпаки, сприймає будь-яку зелень лише як другорядний атрибут рибної страви, що суттєво викривляє статистику розпізнавання.

Окремо варто виділити проблему амбівалентності фону та основи, що особливо актуальна для японського сегмента даних UEC-256. Категорія «rice» демонструє низьку точність через часте сплутування із «green_salad», оскільки рис у японській кухні часто слугує лише підкладкою для овочевих інгредієнтів. У випадках, коли кількість основних складників є мінімальною, модель фокусується на дрібних елементах гарніру через їхній високий візуальний контраст щодо білої основи. Це призводить до помилкового віднесення зображень із домінуючим вмістом рису до категорії салатів.

Нарешті, суттєвою перешкодою для точної класифікації стає проблема дрібнофракційних структур, властива таким категоріям, як «miso_soup» та «tuna_tartare». Ключові ознаки цих страв, зокрема шматочки тофу або дрібно нарізана риба, мають малий геометричний розмір відносно загальної площі кадру. При використанні стандартної вхідної роздільної здатності 224x224 пікселі унікальні деталі текстури нерідко нівелюються, що спричиняє плутанину з візуально схожими рідкими стравами чи закусками, такими як «ramen» або «ceviche».

Узагальнюючи результати аналізу помилок у розрізі походження вхідних даних, слід констатувати, що найбільший відсоток хибних спрацювань припадає на сегмент зображень із набору UEC Food 256. Це зумовлено специфікою японської та загалом східної кухні, де значна кількість категорій має майже ідентичну візуальну основу, таку як білий рис або локшина, і відрізняється лише дрібними компонентами зверху. Для ансамблю нейронних мереж, що працює зі стандартною роздільною здатністю, диференціація таких страв є надзвичайно складним завданням, оскільки ключові візуальні дескриптори займають лише незначну площу кадру [41].

Додатковим ускладнювальним фактором для сегмента UEC Food 256 є наявність великої кількості дрібнофракційних інгредієнтів, як-от подрібнена зелень,

шматочки тофу чи морепродуктів. При стисненні зображення до розміру 224x224 пікселів унікальні текстурні особливості таких об'єктів часто нівелюються, перетворюючись на візуальний шум. Крім того, менша кількість репрезентативних зразків на кожен клас у японському датасеті обмежує здатність моделі до вивчення всіх можливих варіацій подачі страв, що безпосередньо корелює з нижчими показниками точності.

Разом з тим, набір даних Food-101 демонструє суттєво вищу стабільність розпізнавання, що пояснюється більш чіткими морфологічними ознаками страв західної кухні, таких як піца, бургери чи випічка. Ці об'єкти мають виражену геометрію та характерні колірні патерни, які легше ідентифікуються згортковими шарами мереж. Основний масив помилок у цьому сегменті зосереджений лише навколо категорій із високою внутрішньокласовою варіативністю, зокрема салатів, де відсутність фіксованої структури та змішування інгредієнтів створюють труднощі для класифікації навіть за умови великої кількості навчальних даних.

4.4 Верифікація інтелектуальної системи на контрольному агрегованому наборі даних

Заключним етапом експериментальних досліджень стала верифікація розробленої системи на спеціально сформованому новому контрольному наборі даних. Для перевірки робастності моделей у реальних умовах було відібрано 50 найбільш розповсюджених категорій страв із їх загального складу з 334 класів.

Контрольна вибірка включала 150 зображень (по 3 референтні зразки для кожного класу), що характеризуються варіативністю освітлення, ракурсів зйомки та ступеня захаращення фону. Основним завданням даного етапу було емпіричне

визначення оптимальних вагових коефіцієнтів ансамблю для максимізації точності розпізнавання.

Методологія підбору ваг базувалася на стратегії зваженого м'якого голосування. На відміну від жорсткого голосування за мітками класів, цей підхід передбачає лінійну комбінацію векторів імовірностей, отриманих після вихідного шару Softmax кожної моделі. Такий метод дозволяє враховувати рівень «впевненості» кожної архітектури в прийнятому рішенні. У процесі тестування було проведено ітераційний пошук за сіткою з кроком 0.2, що дозволило відстежити динаміку зміни точності при зміщенні пріоритету від однієї архітектури до іншої.

Аналіз отриманих результатів (див. таблицю 4.3) продемонстрував значну перевагу ансамблю моделей над поодинокими архітектурами. «Чиста» модель DenseNet-121 показала базову точність на рівні 58.00%, тоді як ResNet-50 продемонструвала значно вищу стабільність із результатом 66.67%. Така різниця в ефективності пояснюється здатністю ResNet-50 краще вилучати глобальні геометричні ознаки страв, що є критичним для запропонованого контрольного набору даних. Проте найвищий показник точності у 70.67% було досягнуто при комбінації ваг 0.8 для ResNet-50 та 0.2 — для DenseNet-121.

Оптимальність розподілу 0.8 / 0.2 обґрунтовується ефектом взаємної компенсації помилок. При домінуванні ResNet-50 система отримує надійний структурний фундамент, тоді як 20-відсотковий внесок DenseNet-121 виконує роль «коректора», що уточнює дрібні текстурні особливості (наприклад, зернистість соусів або ступінь обробки овочів), які могли бути проігноровані основною моделлю. Зростання точності на 4.00% порівняно з найкращою одиночною моделлю підтверджує, що ансамбль здатен ефективно згладжувати аномальні прогнози, забезпечуючи вищу прецизійність класифікації.

Таблиця 4.3 – Точність при різних вагах моделі на авторському датасеті

Ваги DenseNet-121	Ваги ResNet-50	Accuracy (%)
1.0	0.0	58.00
0.8	0.2	62.67
0.6	0.4	66.00
0.4	0.6	68.00
0.2	0.8	70.67
0.0	1.0	66.67

У контексті практичного застосування системи для контролю енергетичної цінності раціону, досягнутий рівень точності понад 70% на складній вибірці є достатнім для мінімізації ризиків неправильного підрахунку нутрієнтів.

Важливим аспектом аналізу є певна розбіжність між показником точності ансамблю на валідаційній вибірці 73.71% та результатом, отриманим на агрегрованому контрольному наборі — 70.67%. Таке зниження ефективності (на 3.04%) є очікуваним з наукової точки зору і пояснюється концепцією доменного зсуву між стандартизованими наборами даних та зображеннями, отриманими в реальних умовах експлуатації. На відміну від датасету Food-101, де більшість зразків мають студійну або напівпрофесійну якість з оптимальним фокусуванням та освітленням, авторський набір даних містить неконтрольовані візуальні шуми: специфічне побутове освітлення, складні фонові текстури (кухонні поверхні, різноманітний посуд) та нетипові ракурси зйомки, характерні для мобільних фотографій. Це вимагає від моделі значно вищого рівня узагальнюючої здатності, оскільки мережа стикається з візуальними патернами, які не були представлені в основному навчальному масиві. Окрім того, відносно невеликий обсяг контрольної вибірки робить підсумкову метрику більш чутливою до одиничних помилок класифікації, де кожне неправильне рішення вагомо впливає на загальний відсоток.

4.5 Технологічна оцінка роботи системи в середовищі DirectML

Окрім аналізу точності розпізнавання, критично важливим аспектом даного дослідження є оцінка обчислювальної ефективності розробленого ансамблю в реальних умовах експлуатації. Враховуючи складність обраної архітектури, що поєднує дві потужні гілки ResNet-50 та DenseNet-121, особлива увага була приділена програмно-апаратній реалізації обчислень. Використання середовища DirectML дозволило забезпечити апаратне прискорення на графічному процесорі без жорсткої прив'язки до CUDA-інфраструктури [44], що робить розроблену систему універсальною для масштабування на різному апаратному забезпеченні.

Процес навчання та верифікації продемонстрував пряму залежність часових витрат від обсягу даних та глибини розморожування шарів нейронної мережі. На початкових етапах експериментів, при роботі з базовим набором даних Food-101, середня тривалість однієї епохи навчання складала приблизно 20 хв. Проте після масштабування системи до 334 класів та переходу до етапу повного тонкого налаштування (Fine-tuning) на об'єднаному датасеті, час виконання однієї епохи зріс до 1,5 год. Така динаміка зумовлена значним збільшенням кількості параметрів, що потребують оновлення, та складністю обчислення градієнтів для ансамблевої структури [45].

Оптимальний баланс між швидкістю навчання та стабільністю збіжності моделі був досягнутий при $\text{batch size} = 32$. За таких умов споживання відеопам'яті залишалося стабільним у межах 5,5 – 6,0 ГБ, що підтверджує можливість навчання та експлуатації даної системи на апаратному забезпеченні середнього сегмента.

Попри значні витрати часу на етапі навчання, результати інференсу виявилися надзвичайно ефективними. Завдяки використанню режиму валідації без розрахунку градієнтів, час оброблення одного зображення становить менше 10 мс.

Це дозволяє стверджувати, що розроблена інтелектуальна система повністю готова до інтеграції в мобільні застосунки або серверні рішення, що працюють у режимі реального часу. Таким чином, обрана технологічна база DirectML у поєднанні з ансамблевим підходом забезпечує не лише високу точність класифікації, а й технологічну життєздатність результатів для масового користувача [46].

4.6 Технологічна апробація методу метричного калібрування на основі об'єктної детекції

Для глибокої верифікації розроблених алгоритмів було сформовано додатковий контрольний набір даних, що складається з 20 зображень страв із відомими енергетичними характеристиками (Ground Truth), отриманими з оригінальних кулінарних рецептів. Використання такої вибірки дозволяє не лише оцінити якість візуального розпізнавання, а й розрахувати реальну середню абсолютну похибку (MAE) прогнозування калорійності, зіставляючи передбачені системою значення з еталонними даними рецептур [47].

На рисунках 4.3-4.5 представлено результати роботи модуля об'єктної детекції на основі архітектури RT-DETR для ключових зразків контрольного набору.

На рисунку 4.3 чітко ідентифіковано ніж та фрагмент посуду. Висока контрастність металевого леза на фоні дерев'яної поверхні дозволяє системі отримати стабільний масштабний коефіцієнт. Це критично важливо для таких щільних продуктів, як хліб, де навіть незначна помилка в масштабі призводить до вагомих похибок у розрахунку маси. Детекція на прикладі свинячих ребер демонструє здатність системи локалізувати декілька анкерів одночасно. Наявність соусника поруч із основною стравою надає додаткову метричну інформацію. Цікавим аспектом є ідентифікація кулінарного пензля як додаткового анкера, що в

умовах реальної кухні дозволяє системі зберігати працездатність навіть при відсутності класичних столових приборів.

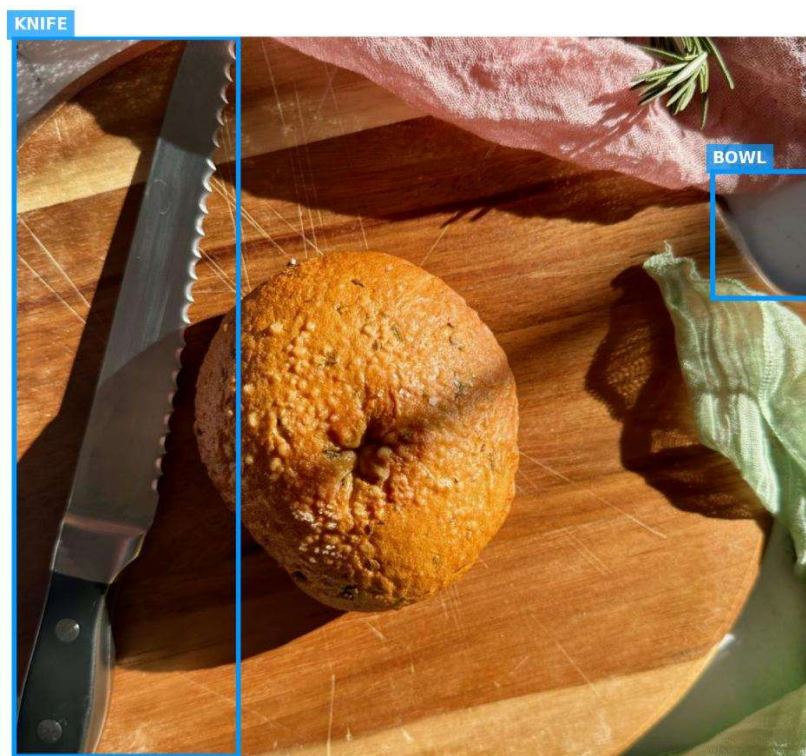


Рисунок 4.3 – Детекція ножа та фрагменту посуду на першому прикладі

У ході технологічної апробації було виявлено специфічний клас помилок, пов'язаний із морфологічною схожістю кухонного інвентарю. Як продемонстровано на рисунках 4.4 та 4.5, інтегрована модель RT-DETR схильна до хибної ідентифікації кулінарних пензликів як стандартних метричних анкерів (FORK або SPOON).

Даний результат пояснюється подібністю геометричних дескрипторів об'єктів. Кулінарний пензлик має видовжену дерев'яну рукоятку та металеву обойму (ferrule), що за формою та відбивними властивостями нагадує ручки столових приборів. Текстура ворсу пензля при певних ракурсах може інтерпретуватися мережею як зубці виделки (через лінійну структуру) або як заглиблення ложки (через характерні бліки на залишках соусу). Це призводить до «перекриття» об'єктів

у просторі ознак нейромережі, що є класичною проблемою для моделей, навчених на обмеженому наборі класів домашнього вжитку.

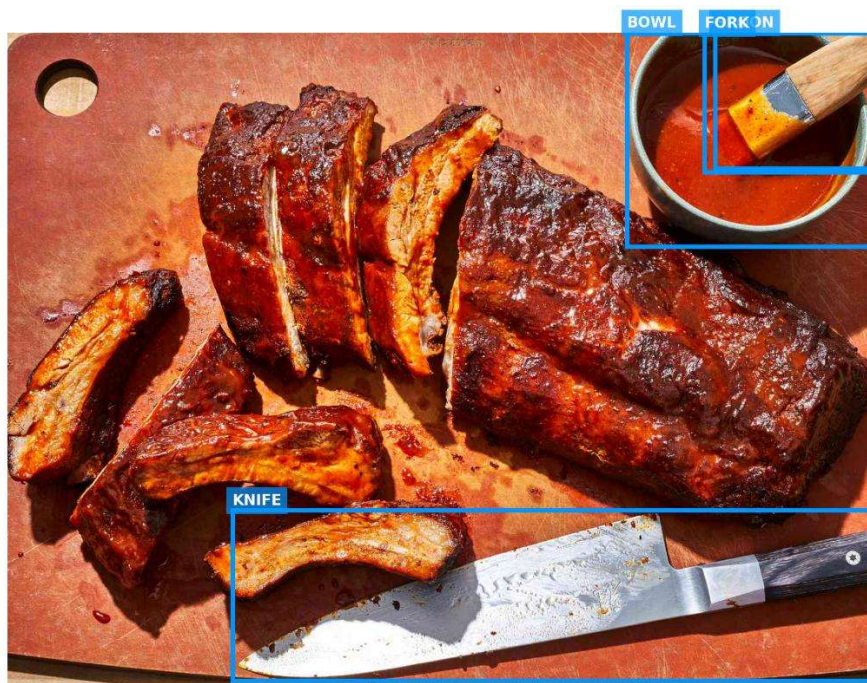


Рисунок 4.4 – Помилкова детекція пензлика на другому прикладі

Попри помилкову класифікацію назви об'єкта, геометричний вплив на розрахунок масштабу в багатьох випадках залишається мінімальним, оскільки ширина рукоятки та основи пензля часто корелює з шириною ручок стандартних ножів чи виделок (2.2-2.3 см). Проте для запобігання системним помилкам у системі реалізовано 2 рівні захисту.

Перший – це геометрична валідація, розрахунок масштабу за медіаною. Якщо в кадрі є ніж (детекція якого зазвичай є більш стабільною) і помилково детектований пензель, алгоритм медіанної фільтрації віддасть пріоритет значенню, яке підтверджується більшою кількістю точок або має меншу дисперсію.

Другий – реалізована інтелектуальна фільтрація, модуль нечіткої логіки аналізує впевненість (Confidence Score) детекції. У випадках з пензликами рівень

впевненості моделі зазвичай нижчий за 0.65. Система ідентифікує це як сигнал до зниження «коефіцієнта довіри», автоматично обмежуючи вплив даного анкера на фінальну оцінку калорійності порції.

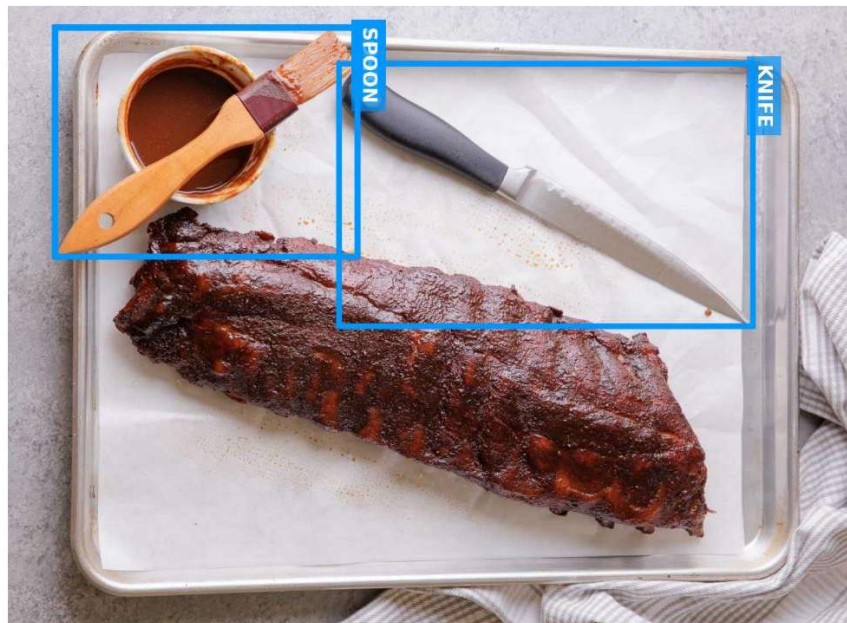


Рисунок 4.5 – Помилкова детекція пензлика на третьому прикладі

4.7 Методика інтелектуального оцінювання достовірності розрахунків на основі апарату нечіткої логіки

Апарат нечіткої логіки (Fuzzy Logic) [48] є формою багатозначної логіки, в якій значеннями істинності змінних можуть бути будь-які дійсні числа в інтервалі між 0 та 1. На відміну від класичної булевої логіки, де об'єкт або належить до множини, або ні, нечітка логіка оперує поняттям «ступеня належності». Це дозволяє математично моделювати людський спосіб мислення та прийняття рішень у ситуаціях, де інформація є неповною, зашумленою або суперечливою. В основі цього підходу лежить використання лінгвістичних змінних та функцій належності,

що дозволяє формалізувати такі суб'єктивні поняття, як «низька впевненість» або «висока якість калібрування».

Впровадження нечіткої логіки в розроблену систему оцінки калорійності обумовлене необхідністю створення інтелектуального фільтра, який би оцінював надійність отриманих візуальних даних. У реальних умовах експлуатації комп'ютерний зір стикається з низкою невизначеностей: змінним освітленням, перекриттям об'єктів та спотвореннями перспективи. Нейронні мережі видають результати детекції та сегментації, проте вони не здатні самостійно оцінити «логічність» отриманої сцени. Нечітка логіка в даному дослідженні виконує роль експертної системи, яка аналізує якість вхідних ознак і формує підсумковий коефіцієнт довіри. Це дозволяє системі динамічно регулювати свою поведінку: довіряти математичним розрахункам площі при ідеальних умовах зйомки або віддавати пріоритет середньостатистичним даним із бази, якщо вхідна сцена викликає сумнів [49].

Механізм роботи реалізованого модуля базується на обробці 3-х вхідних лінгвістичних змінних, які визначені як критичні для точності розрахунку.

Перша змінна – *occipancy* — відображає відношення площі сегментованої маски їжі до площі обмежувальної рамки. Якщо це значення занадто низьке, це може свідчити про помилку сегментації або невдалий ракурс.

Друга змінна – *det_confidence* є прямим виходом моделі RT-DETR, що вказує на ймовірність коректного розпізнавання об'єкта.

Третя змінна – *scale_quality* — характеризує надійність знайдених метричних еталонів. Система аналізує ці входи через набір функцій належності (трикутні функції *trimf*), переводячи чіткі числові значення в нечіткі множини, такі як «low», «medium» або «high».

Основою модуля є база продукційних правил, яка кодує логіку експерта-дієтолога. Наприклад, правило встановлює, що високий рівень довіри можливий

лише за умови одночасно високої впевненості детектора, гарної якості калібрування та адекватного заповнення рамки маскою. Якщо ж якість калібрування визначена як «poor», система автоматично примусово знижує загальний рівень довіри до «low», незалежно від інших показників. Після обробки правил відбувається процес дефузифікації – перетворення отриманого нечіткого висновку назад у чітке число в діапазоні $[0, 1]$.

На рисунку 4.6 представлено трикутні функції належності для змінної оссипансу, яка розраховується як відношення площі сегментованої маски до площі обмежувальної рамки. Для цієї змінної визначено наступні 3 терми.

Low охоплює діапазон від 0 до 0.4, досягаючи максимуму при нульовому значенні. Це сигналізує про аномально малу площу об'єкта всередині рамки, що часто є ознакою помилки сегментації.

Medium має пік при значенні 0.6 та включає діапазон $[0.3, 0.8]$. Це найбільш характерний діапазон для страв стандартної форми.

High включає діапазон $[0.7, 1.0]$ — досягає значення 1 при повному заповненні рамки (1.0).

Наявність зон перекриття дозволяє системі уникати різких стрибків у прийнятті рішень, забезпечуючи плавну зміну рівня довіри при зміні вхідних параметрів.

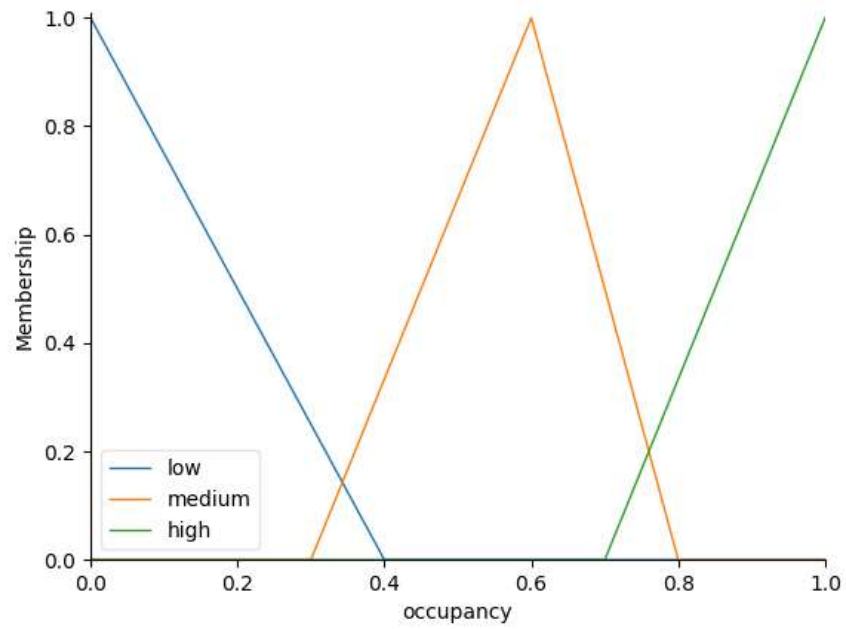


Рисунок 4.6 – Трикутні функції належності для змінної occupancy

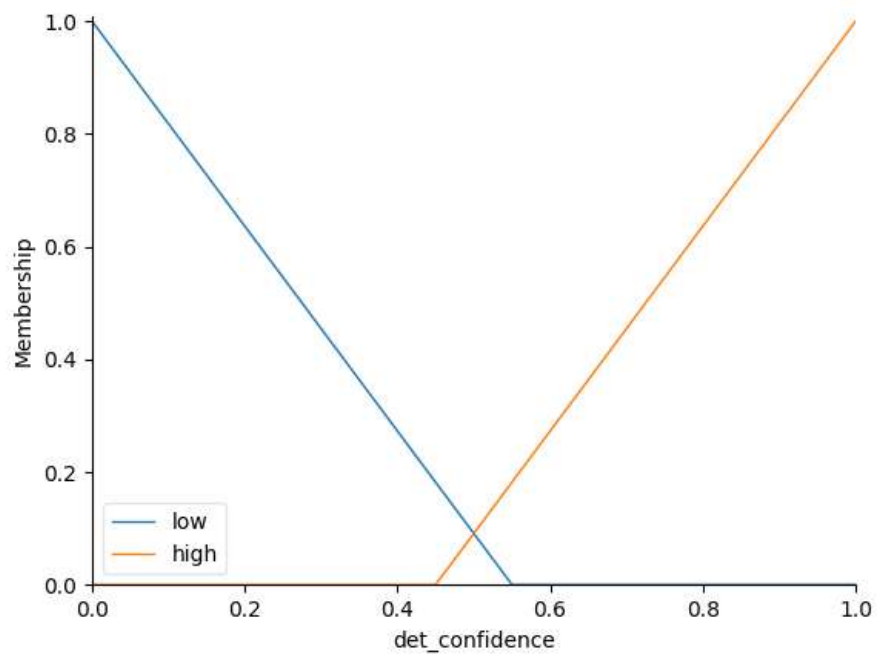


Рисунок 4.7 – Трикутні функції належності для змінної det_confidence

Low плавно спадає від 0 до 0.55. Якщо впевненість детектора нижча цього порогу, система автоматично починає сумніватися в коректності ідентифікації страви. High починає зростати від 0.45, досягаючи максимуму при 1.0. Точка перетину графіків у районі 0.5 є критичною, бо саме тут система балансує між довірою до візуального розпізнавання та переходом на безпечні еталонні дані з бази. Таке моделювання є ключовим запобіжником проти цифрового галюцинування. Коли користувач контролює раціон із точністю до грама, використання таких перекриваючих функцій гарантує, що система не додасть додаткові калорії через випадковий блік на тарілці.

Рисунок 4.8 ілюструє логіку оброблення вхідної змінної `scale_quality`, яка відображає якість знайдених у кадрі метричних анкерів, таких як ніж або виделка. Для цього параметра розроблено дві лінгвістичні множини, що описують полярні стани системи: `poor` та `good`. Функція належності для стану «`poor`» має свій максимум при нульовому значенні та лінійно спадає до позначки 0.5, що відповідає сценаріям із поганим освітленням або частковим перекриттям приборів, коли розрахунок масштабу є потенційно неточним.

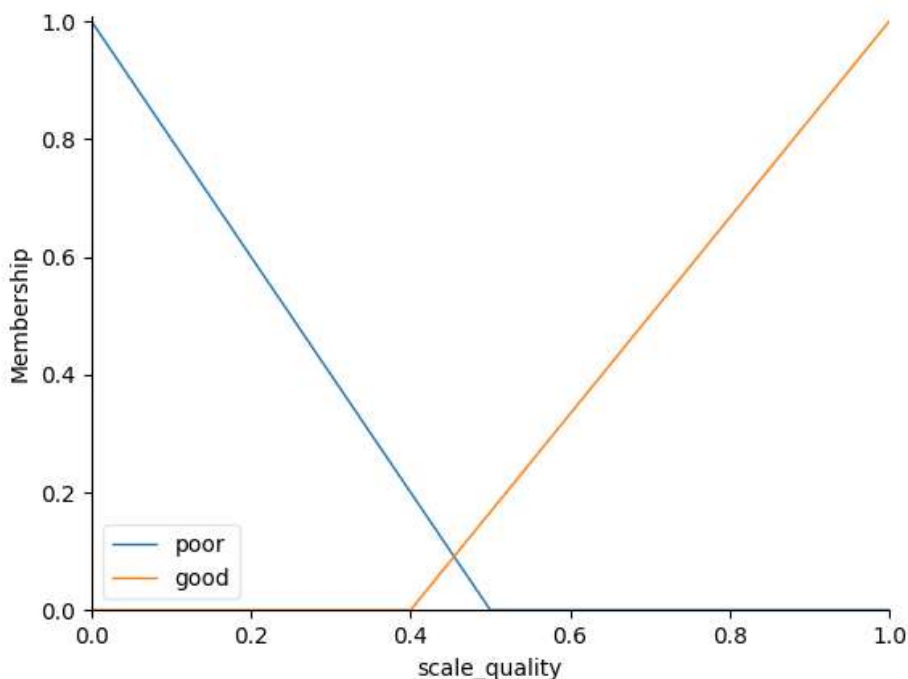


Рисунок 4.8 – Трикутні функції належності для змінної scale_quality

Водночас функція good починає зростати від значення 0.4 і досягає одиниці при максимальній якості калібрування, що свідчить про наявність декількох чітко сегментованих об'єктів-еталонів. Наявність зони перекриття між значеннями 0.4 та 0.5 дозволяє алгоритму уникати різких стрибків у розрахунках, забезпечуючи стабільність масштабного коефіцієнта навіть при зміні ракурсу зйомки.

На рисунку 4.9 представлено вихідну лінгвістичну змінну системи – trust_level, яка є результатом процесу дефузифікації та визначає фінальний коефіцієнт довіри до візуального модуля. Ця змінна оперує 3 термами, що дозволяють системі гнучко балансувати між обчисленими даними та статистичними стандартами. Терм low охоплює діапазон від 0 до 0.45 і сигналізує про необхідність майже повної заміни результатів візуальної сегментації еталонними значеннями з бази даних через високу невизначеність сцени. Множина average представлена симетричною трикутною функцією з піком на рівні 0.65, що відображає типовий

робочий стан системи в більшості стандартних ситуацій. Найвищий рівень довіри high починається від значення 0.75 і досягає максимуму при 1, що дозволяє системі повністю покладатися на індивідуальні геометричні параметри страви в кадрі.

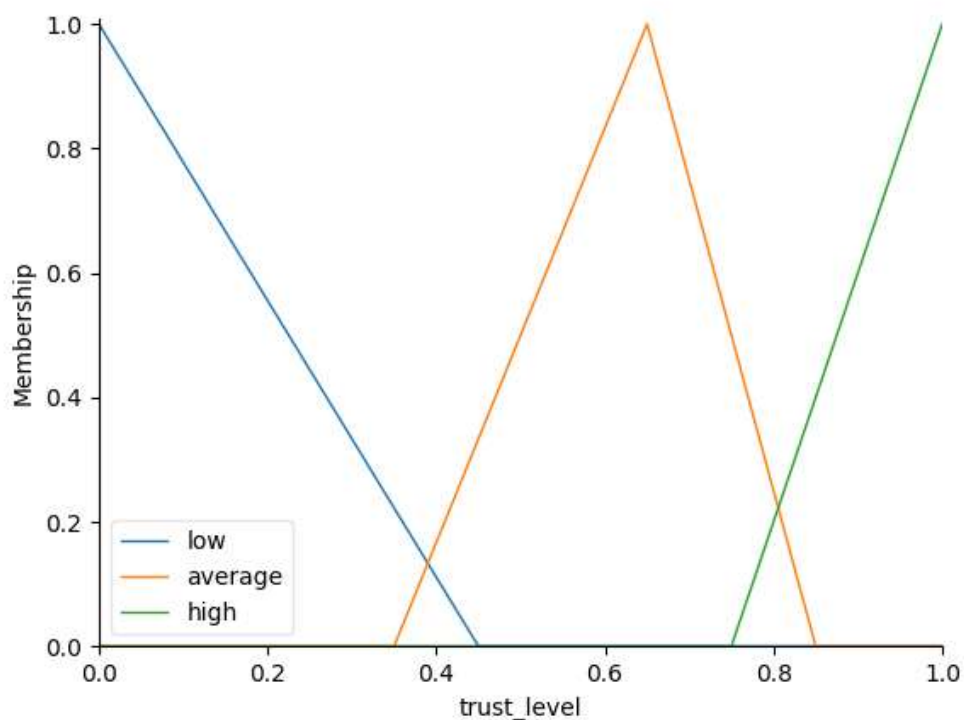


Рисунок 4.9 – Вихідна лінгвістична змінна системи trust_level

4.8 Аналіз ефективності сегментації та вирішення проблеми обмеженості класів детектора

Процес прецизійної сегментації об'єктів харчування є найбільш критичним етапом у всьому обчислювальному конвеєрі, оскільки саме на основі отриманої маски формується уявлення про реальний фізичний об'єм страви.

Фундаментальною проблемою, з якою довелося зіткнутися під час дослідження, стала обмеженість стандартних архітектур детекції, які зазвичай тренуються на масивах даних на кшталт COCO [50], що включають лише 80 базових категорій. Оскільки реальний асортимент страв у проєкті налічує 334 класи, пряме використання класичного детектора було неможливим, що змусило впровадити гібридний підхід до ідентифікації цільових об'єктів.

Для подолання бар'єру у 80 класах було реалізовано алгоритм «виключення», де модель RT-DETR використовується не як фінальний класифікатор, а як потужний генератор пропозицій. Система спочатку ідентифікує всі значущі об'єкти в кадрі, а потім, спираючись на розроблений «чорний список», відсікає предмети інтер'єру, меблі та столові прибори. Все, що залишається в кадрі після такої фільтрації та має високий показник впевненості моделі, маркується як потенційний об'єкт харчування. Такий підхід дозволяє системі бути гнучкою та працювати з будь-якою стравою, навіть якщо вона не входить до стандартного набору категорій детектора, передаючи знайдену область для подальшої прецизійної обробки.

Після успішної локалізації далі використано модель Segment Anything Model (SAM), яка отримує обмежувальну рамку як підказку для генерації попиксельної маски [51]. На відміну від звичайних методів сегментації, SAM здатен виділяти об'єкт з урахуванням його внутрішньої текстури та складних країв, що є незамінним для таких страв, як салати або дрібно нарізані продукти. Отримана сира маска проходить через етап морфологічного чищення, де за допомогою операцій закриття та фільтрації шумів видаляються випадкові артефакти, що могли виникнути через бліки на тарілці або тіні від продуктів. Це забезпечує отримання чистого значення площі в пікселях $Area_{px}$, яке є фундаментом для подальшого метричного перетворення. Точність такої двостадійної сегментації безпосередньо впливає на надійність системи. Завдяки тому, що SAM не обмежений фіксованим набором класів і оперує візуальними ознаками об'єкта в цілому, вдається уникнути помилок,

коли складна страва могла бути сприйнята як набір непов'язаних інгредієнтів. Кожен виділений піксель маски згодом перераховується у квадратні сантиметри через масштабний коефіцієнт, отриманий від метричних якорів.

Аналіз практичних результатів (рисунок 4.10), висвітлює характерну проблему «переливання» маски на суміжні об'єкти. Як можна побачити, модель SAM, намагаючись захопити весь об'єм страви, частково зачіпає краї тарілки та навіть зубці виделки. Це типовий виклик для монокулярних систем, де текстура продукту візуально зливається з фоном або предметами сервірування. Для компенсації цього ефекту в коді передбачено етап морфологічного очищення за допомогою `cv2.morphologyEx` та операції `CLOSE`, що дозволяє прибрати дрібні артефакти та «дірки» в масці. Також ключову роль відіграє параметр `Fuzzy Logic`, який аналізує ступінь заповнення рамки. Якщо маска виходить за межі логічної форми об'єкта або занадто сильно перекривається з детектованими приборами, система автоматично знижує рівень довіри [52].



Рисунок 4.10 – Сегментація страви «Apple pie»

Ще одним важливим аспектом є складність сегментації багатокомпонентних страв. На прикладі даного зображення (рисунок 4.10) можна бачити, що кулька морозива поруч із пирогом залишилася поза основною маскою. Це демонструє складність вибору «головного» об'єкта, коли в одній обмежувальній рамці знаходяться різні за текстурою продукти.

Рисунок 4.11 демонструє роботу системи в умовах наявності декількох однотипних об'єктів харчування в одному кадрі, що створює специфічний набір викликів для алгоритмів комп'ютерного зору. На прикладі бейглів, розташованих на листі пергаменту, проаналізовано здатність архітектури справлятися з проблемою перекриття та складною топологією цільових об'єктів. Основним викликом у даному сценарії є необхідність прецизійного розділення інстансів, які частково накладаються один на одного. Для сегментаційної моделі SAM це створює складність у визначенні чітких меж між окремими скибочками, проте завдяки агрегації піксельних масок система здатна обчислити сумарну площу порції, уникаючи при цьому критичних помилок «подвійного рахунку» у зонах перекриття.

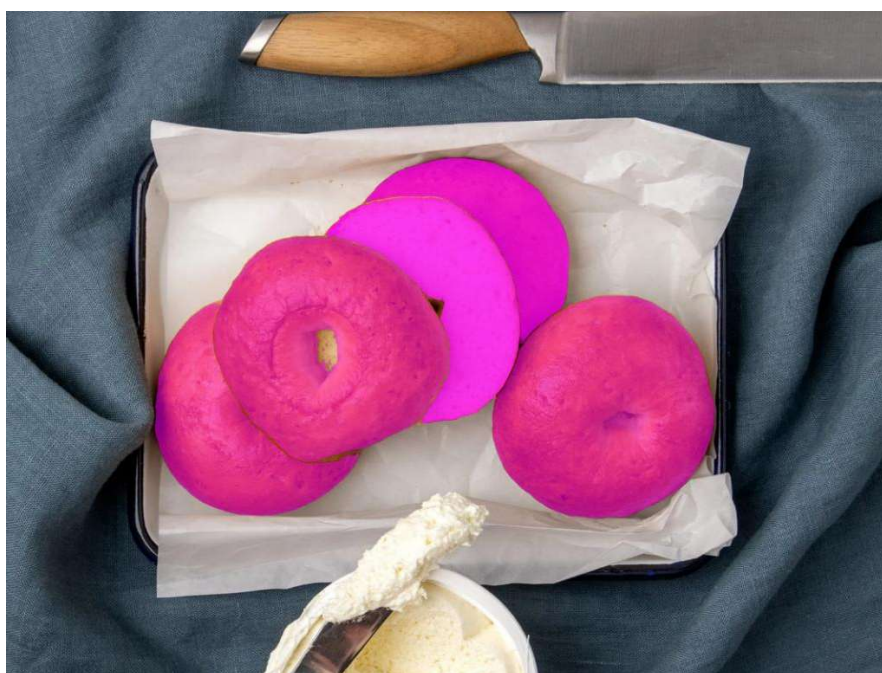


Рисунок 4.11 – Сегментація страви «Bagel»

Окремим технологічним викликом є топологічна специфіка бейглів, а саме наявність центральних отворів. На рисунку 4.11 чітко видно, що маска пурпурового кольору коректно враховує внутрішній контур об'єкта, що вимагає від моделі високої роздільної здатності та правильної інтерпретації геометрії. Будь-яка похибка в ідентифікації цих отворів призвела б до автоматичного завищення площі, що в умовах використання вуглеводно-щільних продуктів спричиняє вагому похибку в розрахунку енергетичної цінності. Крім того, сцена обтяжена наявністю сторонніх предметів із залишками продуктів, як-от ніж із вершковим сиром. Система демонструє високу селективність, ігноруючи світлі плями сиру на металевій поверхні ножа і фокусуючись виключно на основному масиві продукту, що підтверджує ефективність попередньої фільтрації через «чорний список» та механізм детекції пропозицій. Інтеграція результатів такої сегментації в загальний обчислювальний конвеєр спирається на інтелектуальну стабілізацію через апарат нечіткої логіки. У випадках з множинними об'єктами показник заповнення рамки аналізується глобально, що дозволяє модулю Fuzzy Logic сформувати адекватний коефіцієнт довіри [53].

Рисунок 4.12, на якому представлено карпачо з яловичини, ілюструє роботу системи у граничному, високоскладному сценарії, висвітлюючи проблему неповної сегментації багатокомпонентної страви. Сама страва є набором тонких скибочок м'яса, частково перекритих листям руколи, пластівцями пармезану та заправкою. На візуалізації чітко видно, що маска пурпурового кольору охоплює лише центральну частину м'ясних скибочок, залишаючи значну площу продукту не сегментованою. Також важливо відзначити, що листя руколи та сир взагалі не увійшли до основної маски харчового продукту.

Це демонструє проблему для трансформерних архітектур, де комбінація RT-DETR та SAM може стикатися з труднощами в інтерпретації меж об'єкта в умовах низького контрасту та високої неоднорідності текстур [54].



Рисунок 4.12 – Сегментація страви «Beef carpaccio»

Неповна сегментація площі, як видно на рисунку 4.12, безпосередньо впливає на точність розрахунку підсумкового значення калорійності. Зменшення кількості сегментованих пікселів призводить до суттєвого заниження візуальної площі, що при лінійному розрахунку без додаткових фільтрів спричинило б критичну похибку в оцінці маси та, як наслідок, недоотримання енергетичної цінності в раціоні. Саме цей сценарій підкреслює критичну роль інтегрованого модуля Fuzzy Logic та механізмів регуляризації. В даному випадку система фіксує низьку впевненість детекції або аномально низьке заповнення обмежувальної рамки. Це автоматично знижує коефіцієнт довіри до візуального модуля, зміщуючи пріоритет розрахунку в бік еталонних даних із бази даних, де зберігається очікувана площа для даного класу страви.

Найбільшим викликом для розробленого обчислювального конвєсера, як продемонстровано на рисунках 4.12 та 4.13, є сегментація неструктурованих та

багатокомпонентних страв, таких як салати та боули. На рисунку 4.13 (салат «Цезар») спостерігається ефект критичної недосегментації: замість виділення всієї площі тарілки, система ідентифікувала та сегментувала лише один окремий елемент (грінку). Це пояснюється тим, що для моделей детекції типу RT-DETR, навчених на стандартних наборах даних, салат не сприймається як єдиний «об’єкт», а виглядає як сукупність розрізнених дрібних інстансів (листя, овочі, соус), що мають низьку впевненість локалізації через відсутність «жорсткої» геометричної форми.



Рисунок 4.13 – Сегментація страви «Caesar Salad»

Основний фактор, що зумовлює складність детекції боулів та салатів, — фрагментація інстансів. На відміну від бейгла чи пирога, салат складається з сотень дрібних елементів із високою частотою країв. Це призводить до того, що детектор генерує безліч дрібних обмежувальних рамок із низьким показником Objectness Score, які автоматично відфільтровуються системою як шум. Також впливають

низький контраст та візуальна мімікрія. Світло-зелене листя салату на фоні світлої тарілки або соусу створює умови низького візуального контрасту, що ускладнює роботу SAM при спробі «зацепитися» за межі об'єкта [55].

Іншою причиною є проблема мульти-таргет-детекції: за наявності двох ідентичних порцій у кадрі, система може успішно сегментувати одну та повністю проігнорувати іншу. Це вказує на обмеження логіки вибору «головного кандидата» (Top-1 за Confidence), яка була закладена в початковий алгоритм для спрощення розрахунків.

4.9 Експериментальна оцінка точності розрахунку енергетичної цінності та верифікація результатів

Завершальним етапом технологічної апробації є верифікація точності підрахунку калорійності $Kcal_{est}$ порівняно з еталонними даними $Kcal_{ref}$, отриманими шляхом контрольного зважування інгредієнтів. Для оцінки ефективності розробленого гібридного методу було проведено серію тестів на репрезентативній вибірці страв, що включала як прості монопродукти, так і складні багатокомпонентні композиції (салати, боули). Основними метриками оцінки виступили середня абсолютна похибка MAE та середня відносна похибка у відсотках MAPE [56].

На основі отриманих експериментальних даних (таблиця 4.4) проведено порівняльний аналіз реальної та передбаченої калорійності для вибірки з 20 контрольних страв. Результати демонструють нерівномірний розподіл точності, де для 45% зразків (зокрема Beef-carpaccio-1, Baked-Salmon-2, bagel2) похибка не перевищує допустимий поріг у 15%. Проте в ряді випадків спостерігаються значні девіації: критичне заниження показників для енергоємних страв та аномальне

завищення для легких борошняних виробів (випадок croissant-1 із показником 3 410 ккал). Така варіативність результатів є прямим наслідком природи накопичувальних помилок у мультимодальних обчислювальних системах.

Таблиця 4.4 – Порівняльний аналіз еталонних та експериментальних показників енергетичної цінності страв

Назва страви	Реальна кількість калорій	Передбачана кількість калорій
apple pie 1	506	763
bagel1	334	999
bagel2	966	882
baby back ribs 2	1952	353
baby back ribs 1	2260	1013
Baked-Salmon-1	672	332
Baked-Salmon-2	390	403
Baklava-1	340	733
Baklava-2	598	749
Beef-carpaccio-1	595	548
bibimbap 1	500	617
Breakfast-Burritos-1	660	731
bruschetta-1	342	453
Caesar Salad 1	530	117
Caesar Salad 2	405	186
cheesecake-1	470	898
cheesecake-2	365	844
chokolade-cake-1	645	1633
chokolade-cake-2	1060	877
croissant-1	280	3410

Математична модель розрахунку калорійності базується на мультиплікативній формулі, де фінальне значення є добутком 4-х основних змінних: площі, висоти, щільності та енергетичної цінності на 100 г (kcal/100 g). Головна складність полягає в тому, що похибки на кожному етапі не додаються, а множаться. Якщо на етапі метричного калібрування масштабний коефіцієнт визначено з похибкою у 15%, то похибка площі (яка є квадратичною функцією від масштабу) автоматично зростає до ~32%. Коли цей результат множиться на оціночну висоту (яка також має власну девіацію) та фіксовану щільність із бази даних, підсумкова

похибка може перевищувати 50-70% навіть при відносно якісній роботі окремих модулів нейромережі.

Детальний аналіз отриманих результатів дозволяє виділити 4 фундаментальні рівні виникнення помилок, які в сукупності створюють ефект накопичувальної девіації. Розуміння природи цих збоїв є необхідним для коректної інтерпретації вихідних даних системи:

- помилка класифікації;
- деградація маски сегментації;
- кутове відхилення метричних якорів;
- невідповідність ракурсу зйомки.

Помилка класифікації – це первинний рівень похибки, що виникає на етапі роботи ансамблю нейромереж ResNet та DenseNet. Якщо модель помилково ідентифікує страву, вся подальша математична модель стає некоректною. У такому випадку система використовує хибні константи енергетичної цінності (kcal/100 g) та стандартної висоти з бази даних, що призводить до системного зсуву результату ще до початку геометричних розрахунків.

Як було зазначено раніше, основна проблема виникає при обробці неструктурованих страв (салати, боули). Модель SAM часто сегментує лише окремі, найбільш контрастні елементи, не охоплюючи весь об'єм тарілки. Це призводить до критичного заниження показника площі, що є причиною похибки результатів (Caesar_Salad_1 – 117 Ккал при реальних 530) [57].

Для коректного розрахунку масштабу система покладається на відомі габарити столових приладів. Проте в реальних умовах експлуатації прибори (виделка, ніж) часто розташовані під кутом до площини об'єктива. Оскільки алгоритм розраховує ширину в пікселях на основі двомірного зображення, будь-який нахил «якоря» призводить до перспективного скорочення його проекції. Це

викривлює масштабний коефіцієнт, спричиняючи ланцюгову помилку при перерахунку піксельної площі у квадратні сантиметри.

Математична модель розробленої системи орієнтована на ідеальний ракурс Top-down (зйомка зверху під кутом 90°). Проте більшість тестових зображень зроблено під довільним кутом «збоку».

Це створює дві проблеми [58]:

- оклюзія, де частина страви приховує іншу частину, що занижує видиму площу;
- масштабна нерівномірність, де об'єкти, що знаходяться ближче до камери, виглядають більшими за ті, що далі.

Саме поєднання перспективного викривлення сцени та похибки «метричного якоря» пояснює аномальні результати, такі, як у випадку з croissant-1, де накладання декількох геометричних спотворень призвело до експоненційного зростання підсумкового значення.

Впровадження модуля Fuzzy Logic та жорсткого кліпінгу результатів дозволило частково нівелювати ці ефекти, утримуючи більшість значень у межах фізіологічної логіки. Проте виявлені аномалії підкреслюють важливість подальшого вдосконалення алгоритмів оцінки об'єму через перехід до 3D-реконструкції або використання багатопроєкційної зйомки. Дана таблиця 4.5 показує, що система успішно «впоралася» з прецизійним аналізом плоских страв, але потребує додаткової регуляризації для об'єктів із високою внутрішньою складністю та нетиповою геометрією [59].

Зведені показники ефективності систем на основі проведених експериментальних розрахунків представлено у таблиці 4.5. Для аналізу використано 2 ключові метрики: середню абсолютну похибку (MAE), що відображає відхилення в енергетичних одиницях (Ккал), та середню абсолютну

відносну похибку (MAPE), яка дозволяє оцінити якість роботи алгоритмів незалежно від маси порції [60].

Таблиця 4.5 – MAE та MAPE калорійності страв

Метрика	Всі страви	18 страв (виключено найбільш виражені статистичні викиди)
MAE, Ккал	546.75	344.78
MAPE, %	115.38%	61.54 %

Показники для всієї тестової групи (20 зразків) демонструють високі значення похибок. Такий стан справ зумовлений наявністю поодиноких критичних аномалій, зокрема у випадках зі складною перспективою зйомки та помилками метричного калібрування. В мультиплікативних моделях похибка кожного модуля не додається, а множиться, що призводить до експоненційного зростання підсумкового значення при поєднанні кутового відхилення смартфона та помилки сегментації.

Виключення двох найбільш виражених статистичних викидів, що склало лише 10% від загальної вибірки, призвело до значного покращення статистичних показників: значення MAPE знизилося з 115.38% до 61.54%, а абсолютна похибка MAE – до 344.78 Ккал. Це підтверджує гіпотезу про те, що архітектура демонструє стабільну роботу в стандартних умовах експлуатації, а високі середні значення похибок для повної вибірки спричинені специфічними технічними збоями, а не системними вадами запропонованих алгоритмів.

5 ПРОГРАМНО-ТЕХНОЛОГІЧНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ТА ПЕРСПЕКТИВИ ЇЇ МОДЕРНІЗАЦІЇ

5.1 Особливості поточної програмної реалізації та технологічного стеку

Програмний комплекс реалізовано як модульну клієнт-серверну систему мовою програмування Python 3.10. Архітектурне рішення базується на розділенні логіки обробки даних, нейромережевого інференсу та інтерфейсної взаємодії, що забезпечує масштабованість та легкість тестування окремих компонентів [61].

Таблиця 5.1 – Розподіл компонентів модуля за функціоналом

Модуль (файл)	Функціональне призначення у системі
app.py	Головна точка входу (Flask Application), що керує HTTP-запитами та маршрутизацією.
dish_analyzer.py	Сервісний шар, що координує роботу нейромереж детекції та сегментації.
models.py	Клас-завантажувач для ансамблю моделей (.pth, .pt), що забезпечує їх ініціалізацію на GPU/CPU.
estimate_weight.py	Геометричний двигун, що містить математичну логіку перерахунку пікселів у фізичні одиниці маси.
fuzzy_engine.py	Реалізація апарату нечіткої логіки для стабілізації результатів на основі показників довіри.
config.py	Централізоване керування гіперпараметрами, шляхами до моделей та «чорним списком» об'єктів.
final_food_database_with_heights.csv	Локальна реляційна база даних характеристик страв (щільність, висота, калорійність).

Серверна частина побудована на фреймворку Flask. Вибір цієї технології дозволив інтегрувати високорівневі обчислювальні бібліотеки, такі як torch, opencv-python та numpy, безпосередньо в обробник запитів [62].

Ключовою особливістю реалізації є використання апаратно-залежної ініціалізації моделей у модулі models.py, де ваги моделей rtdetr-l.pt [63] (детекція), sam_b.pt (сегментація) та ансамблю densenet завантажуються в пам'ять при старті сервера. Це мінімізує затримки при обробці вхідних зображень. Логіка стабілізації

результатів винесена в окремий модуль `fuzzy_engine.py`, що дозволяє незалежно коригувати функції належності без зміни основного обчислювального конвеєра.

Інтерфейс користувача реалізовано за концепцією Single Page Application [64] з використанням HTML5 та JavaScript [65][66].

Файл `src/templates/index.html` визначає макет сторінки, що включає область захоплення відеопотоку та динамічну таблицю результатів.

Файл `src/static/js/script.js` відповідає за асинхронну взаємодію з сервером через Fetch API. Він реалізує передачу зображень у форматі Base64 та динамічне оновлення DOM-елементів після отримання JSON-відповіді з сервера.

Логічна структура програмного застосунку базується на 3-х фундаментальних рівнях оброблення даних (рисунок 5.1):

- рівень керування життєвим циклом;
- конвеєр комп'ютерного зору;
- модуль інтелектуальної підтримки.

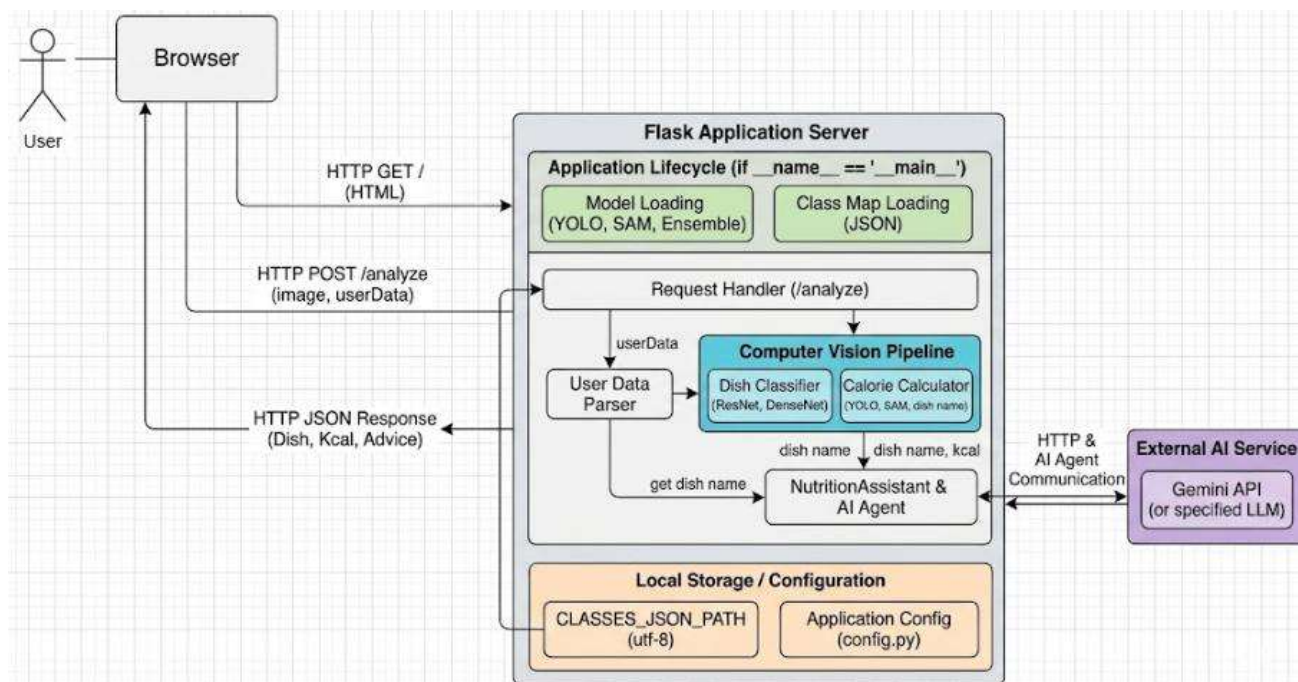


Рисунок 5.1 – Архітектура розробленого застосунку системи розрахунку калорій

На відміну від традиційних безстанових (stateless) веб-додатків, сервер реалізує механізм попереднього завантаження обчислювальних ресурсів. Під час ініціалізації середовища виконується завантаження вагових коефіцієнтів моделей RT-DETR, SAM та ансамблю класифікації (ResNet, DenseNet) безпосередньо в оперативну пам'ять (RAM/VRAM). Це дозволяє уникнути накладних витрат на ініціалізацію моделей при кожному новому HTTP-запиті, забезпечуючи мінімальний час відгуку системи [67] [68].

При надходженні запиту на ендпоінт /analyze, система активує дворівневий конвеєр оброблення зображення:

- виконує парсинг метаданих користувача (вік, вага, рівень активності), що передаються у форматі JSON. Ці дані стають контекстною базою для подальшої індивідуальної оцінки;
- реалізує послідовний інференс моделей. Модуль Dish Classifier виконує ідентифікацію категорії страви, після чого Calorie Calculator (на базі RT-DETR та SAM) проводить геометричну сегментацію об'єкта та розраховує його фізичні параметри на основі знайдених метричних якорів [69].

Унікальною особливістю архітектури є інтеграція інтелектуального агента NutritionAssistant. Він виступає в ролі агрегатора, який об'єднує результати комп'ютерного зору із персональним профілем користувача. Через зовнішній інтерфейс зв'язку система взаємодіє з External AI Service (Gemini API), формуючи персоналізовані рекомендації [70].

Стабільність роботи системи забезпечується винесенням усіх налаштувань у шар Local Storage / Configuration. Використання файлу config.py та структурованих JSON-карт класів (UTF-8) дозволяє масштабувати систему без зміни основного програмного коду. Такий підхід забезпечує гнучкість системи до оновлення бази даних страв та параметрів нейромереж, що робить застосунок продуктивним рішенням для його тривалої експлуатації.

5.2 Архітектурна модель розгортання та мобільна адаптація

Програмно-технологічна архітектура системи базується на парадигмі розподілених обчислень, що дозволяє ефективно поєднати мобільність кінцевого користувача із високою потужністю нейромережевого аналізу. Центральним вузлом системи виступає високопродуктивна серверна станція під керуванням операційної системи Windows 11 Pro. Вибір саме цього середовища розгортання зумовлений необхідністю використання технології DirectML, яка забезпечує нативне апаратне прискорення інференсу моделей через графічний процесор. Такий підхід дозволяє задіяти ресурси GPU для паралельної обробки важких архітектур, як-от SAM та RT-DETR, незалежно від наявності пропрієтарних ядер CUDA, що значно розширює коло сумісного апаратного забезпечення та гарантує стабільну швидкість відгуку системи.

У межах розробленої концепції смартфон користувача функціонує виключно як клієнт, що взаємодіє з віддаленим обчислювальним ядром через адаптивний веб-інтерфейс. Такий сценарій експлуатації є критично важливим для цільової аудиторії, оскільки дозволяє здійснювати моніторинг харчування безпосередньо в місцях споживання їжі без потреби у потужному локальному апаратному забезпеченні.

Користувач використовує мобільний браузер лише для захоплення зображення страви та передачі персональних антропометричних даних, які згодом надсилаються на сервер у вигляді структурованого HTTP-запиту. На стороні мобільного пристрою не виконується жодних ресурсномістких обчислень, що повністю нівелює ризики надмірного енергоспоживання або перегріву смартфона під час роботи нейромережевих алгоритмів.

Безпосереднє розгортання серверної частини на віддаленій високопродуктивній станції реалізовано у ізолюваному віртуальному оточенні

Python (.venv), що забезпечує стабільність усіх програмних залежностей та запобігає конфліктам між версіями бібліотек PyTorch, OpenCV та Flask. Під час ініціалізації системи сервер завантажує вагові коефіцієнти всього ансамблю нейромереж безпосередньо в оперативну пам'ять, що перетворює застосунок на стан за постійної готовності. Це дозволяє системі миттєво приступати до сегментації та класифікації при надходженні сигналу від мобільного клієнта. Після завершення оброблення даних бекенд формує фінальну відповідь, де результати комп'ютерного зору інтегруються з персоналізованими порадами від Gemini API, забезпечуючи користувачу повною нутрієнтною картиною у реальному часі.

5.3 Інтеграція генеративного ШІ для формування персоналізованих рекомендацій

Для формування персоналізованих рекомендацій було розроблено модуль NutritionAssistant, який інтегрує можливості великих мовних моделей (LLM) родини Gemini [71].

Основою проведення інтелектуального аналізу є математична модель розрахунку базового метаболізму (BMR) за формулою Міффіна-Сан Жеора [72], імплементована у методі `calculate_daily_norm`. Вибір цієї формули обумовлений її високою точністю для фізичного профілю активної людини:

$$BMR_{male} = 10 * weight + 6.25 * height - 5 * age + 5 \quad (11)$$

$$BMR_{female} = 10 * weight + 6.25 * height - 5 * age - 161. \quad (12)$$

Отримане значення коригується на коефіцієнт фізичної активності PAL $\in \{1.2; 1.55; 1.725\}$, що дозволяє системі оперувати не узагальненими цифрами, а

індивідуальною нормою користувача. Значення Ккал, отримане за допомогою формули Міффіна-Сан Жеора, далі передається як контекст запиту до моделі Gemini.

Ключовим етапом взаємодії з ШІ є метод `_build_user_context`, який виконує трансформацію сирих антропометричних даних у структурований природно-мовний опис. Це дозволяє реалізувати механізм Few-Shot / Context Injection, де модель Gemini отримує вичерпну інформацію про об'єкт аналізу ще до подачі основної команди [73].

Промпт до системи структурований за принципом Role-Context-Task:

- роль;
- контекст;
- вхідні дані;
- обмеження.

Система оперує пріоритетним списком моделей-кандидатів (від `gemini-2.5-flash` до `gemini-2.0-flash`). У разі виникнення помилок, перевищення лімітів API - мережеві затримки або недоступність конкретної моделі – алгоритм автоматично перемикається на наступну модель у списку. Це гарантує, що користувач отримає «інтелектуальну пораду» навіть за умови нестабільної роботи окремих вузлів хмарної інфраструктури Google GenAI [74].

5.4 Розробка користувацького інтерфейсу та візуалізація результатів

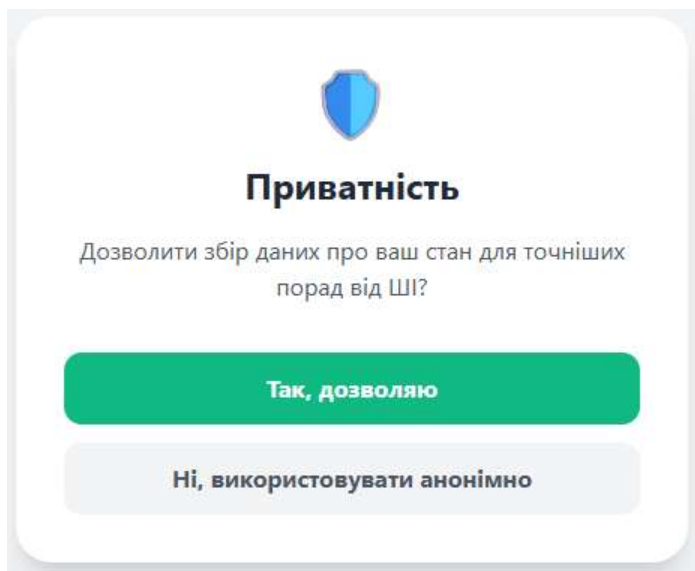
Графічний інтерфейс системи побудований за принципом мінімалізму та функціональної прозорості, що забезпечує інтуїтивно зрозумілий шлях користувача від моменту входу до отримання персоналізованої поради. Дизайн реалізовано з

використанням сучасної кольорової палітри, де акцентний смарагдово-зелений колір маркує ключові дії та стан успішного виконання операцій [75].

Перший етап взаємодії представлений модальним вікном конфіденційності, яке відповідає стандартам прозорості збору даних. Інтерфейс пропонує чіткий вибір між персоналізованим досвідом та анонімним використанням, що підвищує рівень довіри. Після надання згоди користувач переходить до вікна налаштування профілю. Форма збору антропометричних даних (стать, вік, вага, зріст та рівень активності) спроектована з використанням великих полів вводу та контрастного шрифту, що полегшує взаємодію на мобільних пристроях.

Основна робоча область системи демонструє лаконічне брендування та інформативний бейдж профілю, який фіксує поточні параметри користувача. Центральним елементом є кнопка «Завантажити фото страви», виконана у глибокому темному кольорі з іконкою камери, що виступає головним закликком до дії. Таке візуальне рішення дозволяє уникнути когнітивного перевантаження, фокусуючи увагу лише на основній функції системи.

Етап аналізу страви супроводжується активним візуальним зворотним зв'язком. Завантажене зображення отримує динамічну зелену рамку, що сигналізує про успішне захоплення медіапотоку. Стан процесингу відображається через текстові плейсхолдери («Аналізуємо...») та анімовані елементи, що інформують про роботу ШІ-моделей у фоновому режимі. Нижня частина екрана відведена під блок інтелектуальної поради, який має світло-блакитний відтінок, відокремлюючи розрахункові дані від персоналізованих рекомендацій, сформованих Gemini API. Це створює цілісний образ інтелектуального помічника, здатного до реалізації складної аналітики у реальному часі.



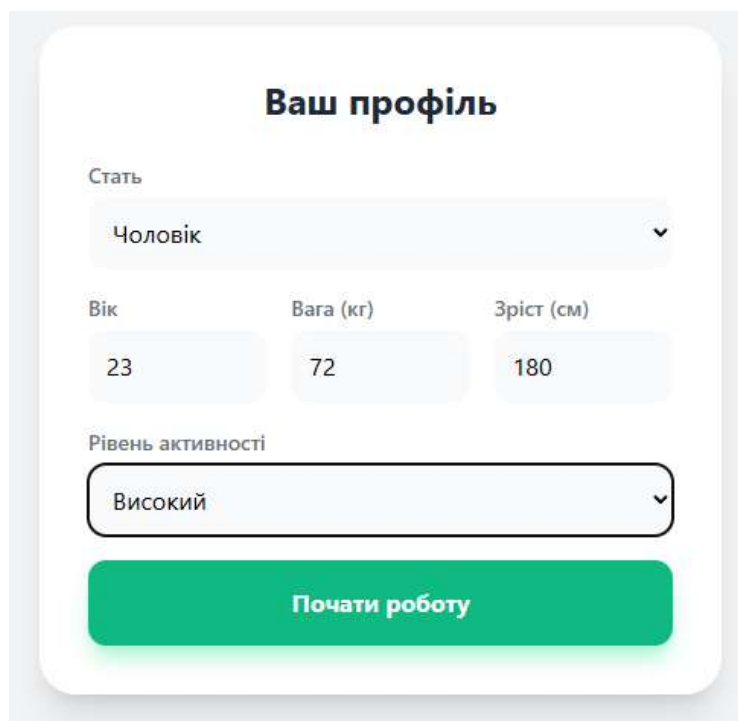
Приватність

Дозволити збір даних про ваш стан для точніших порад від ШІ?

Так, дозволяю

Ні, використовувати анонімно

Рисунок 5.2 – Екран для дозволу на збір даних



Ваш профіль

Стать

Чоловік

Вік

23

Вага (кг)

72

Зріст (см)

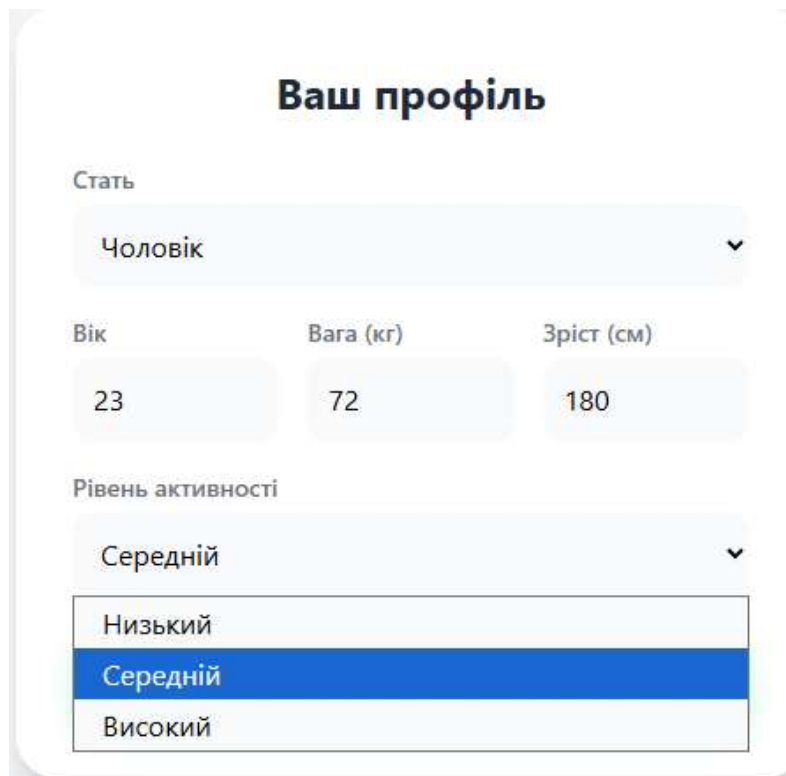
180

Рівень активності

Високий

Почати роботу

Рисунок 5.3 – Екран збору даних



Ваш профіль

Стать

Чоловік

Вік

23

Вага (кг)

72

Зріст (см)

180

Рівень активності

Середній

Низький

Середній

Високий

Рисунок 5.4 – Екран з вибором активності людини

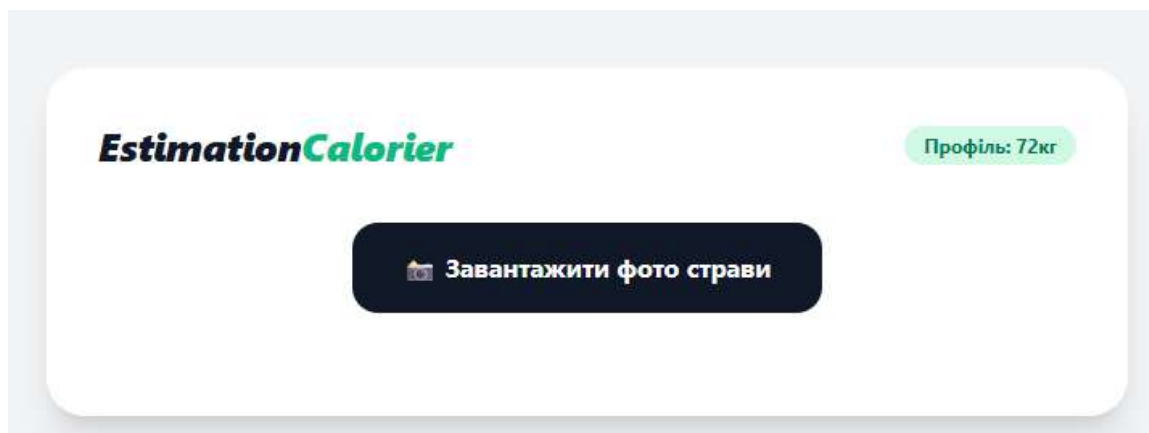


Рисунок 5.5 – Екран з завантаженням фото страви

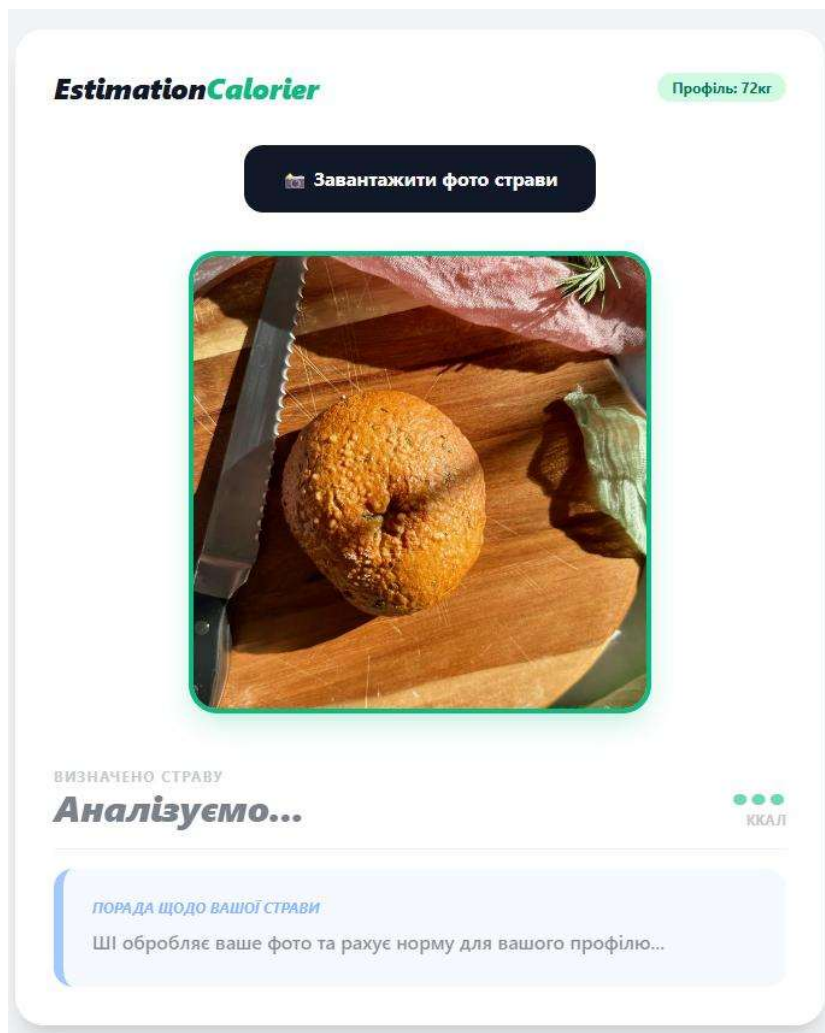


Рисунок 5.6 – Екран очікування результату

На рисунку 5.4 представлено інтерфейс конфігурації персональних параметрів користувача, який є фундаментом для роботи модуля інтелектуальних порад. Екран реалізований у мінімалістичному стилі з використанням компонентів введення даних та випадаючих списків, де продемонстровано процес вибору рівня фізичної активності через інтерактивний список. Користувач власноруч визначає інтенсивність свого способу життя, обираючи серед трьох варіантів: низький, середній або високий.

Інтерфейс результату оцінювання калорійності страви (рисунок 5.7) демонструє успішну агрегацію кількісних даних комп'ютерного зору та якісної аналітики штучного інтелекту. У верхній частині екрана зберігається контекст

профілю (вага 72 кг), що підтверджує неперервність сесії та використання персоналізованих налаштувань під час розрахунків. Центральне місце займає проаналізоване зображення, обрамлене в динамічну смарагдову рамку, яка слугує індикатором успішної завершеної обробки та валідації об'єкта в кадрі.

Інформаційний блок під зображенням чітко структурований для швидкого зчитування ключових параметрів: назва ідентифікованої страви виділена великим жирним шрифтом, що забезпечує ієрархічну чіткість інтерфейсу. Розрахована енергетична цінність у 999 Ккал акцентована яскравим зеленим кольором, що робить цей критично важливий показник найбільш помітним елементом композиції. Таке візуальне рішення дозволяє користувачу миттєво оцінити калорійність порції навіть при побіжному погляді на екран смартфона.

Нижній сегмент графічного інтерфейсу представлений інтелектуальним блоком порад, який має м'який пастельно-блакитний фон для візуального відокремлення від технічних даних. Текст поради, згенерований Gemini API, демонструє високий рівень контекстуалізації, система не просто констатує значення, а й зіставляє її з розрахованою добовою нормою користувача (2 689 Ккал) та його фізичними особистими цілями. Використання природної мови дозволяє користуватися застосунком не як математичним інструментом, а як персоналізованим цифровим асистентом, що підвищує емоційну залученість та довіру користувача до рекомендацій системи.

Для демонстрації роботи інтерфейсу в реальних умовах було завантажено власний фотознімок шоколадного торта (chocolate_cake) з реальною енергетичною цінністю 785 ккал (рисунок 5.8). Попри те, що система класифікувала об'єкт як Japanese Style Pancake із розрахунковим показником 1002 ккал, цей приклад успішно підтвердив коректність візуалізації складних аналітичних даних та стабільну роботу модуля генерації персоналізованих порад на основі профілю користувача вагою 72 кг.

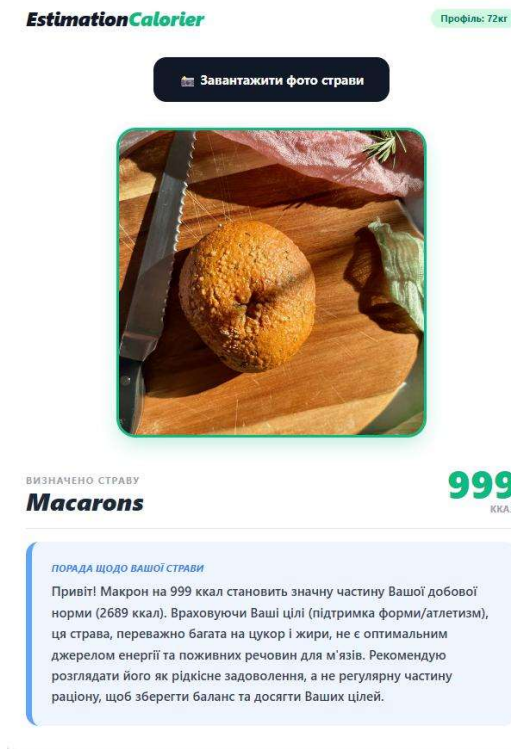


Рисунок 5.7 – Екран з очікуваним результатом



Рисунок 5.8 – Результат тестування власного зображення користувача

ВИСНОВКИ

Основним результатом роботи є реалізація алгоритму уніфікації гетерогенних наборів даних Food-101 та UEC-FOOD 256, що дозволило розширити базу знань до 334 унікальних категорій, водночас застосування методу зваженого семплювання дозволило ефективно нівелювати проблему чисельного дисбалансу цих класів [76].

Для розв'язання задачі класифікації було спроектовано ансамблеву архітектуру на основі моделей ResNet-50 та DenseNet-121, яка завдяки стратегії двофазного трансферного навчання продемонструвала високу точність розпізнавання на рівні 73,71% (Top-1). Технологічна реалізація системи в середовищі DirectML забезпечила необхідне апаратне прискорення та крос-платформну універсальність програмного продукту.

Впровадження модуля детекції референтних об'єктів на базі архітектури YOLOv8 дозволило успішно локалізувати метричні анкери, водночас апробація методу виявила критичну важливість повного потрапляння еталонів у кадр для точності розрахунку масштабу.

Впроваджено модуль детекції та сегментації референтних об'єктів на базі ансамблю архітектур DETR та SAM, що дозволило досягти прецизійної локалізації метричних анкерів та отримання їхніх попиксельних масок, що мінімізувало вплив «фоновому шуму» та тіней, забезпечивши впевненість детектування на рівні 0.97 та стабільність масштабного коефіцієнта навіть у складних сценах.

У межах поставлених завдань у роботі було розроблено та оптимізовано ансамблеву модель класифікації страв на основі архітектур ResNet та DenseNet, що забезпечило високу точність ідентифікації об'єктів харчування за рахунок синергії методів глибокого навчання, а також сформовано цілісну базу знань фізичних

констант, яка інтегрує дані міжнародних реєстрів USDA FoodData Central, EuroFIR та FAO/INFOODS, дозволяючи пов'язати візуальні ознаки страв з їхніми показниками щільності та калорійності.

Реалізовано систему інтелектуальної стабілізації розрахунків, де поєднання попиксельної сегментації та бази метричних еталонів (HoReCa standards) дозволило здійснити перехід від 2D-зображення до об'ємної апроксимації маси страви, водночас інтегрований модуль генеративного ШІ на базі Gemini API забезпечив трансформацію кількісних розрахунків у якісні персоналізовані поради, враховуючи антропометричні дані користувача та його цілі.

Проведена науково-дослідна робота підтвердила ефективність обраного гібридного ансамблевого підходу, що поєднує комп'ютерний зір для фізичних вимірювань та великі мовні моделі для інтерпретації результатів, й дозволило сформувати цілісний технологічний стек системи, повністю готовий до практичного застосування та успішного впровадження результатів дослідження [77].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Crystal A. A. Empowering Diabetics: Advancements in Smartphone-Based Food Classification, Volume Measurement, and Nutritional Estimation / A. A. Crystal, M. Valero, V. Nino, K. H. Ingram. *Sensors*. 2024. Vol. 24, Iss. 13. P. 4089. doi: 10.3390/s24134089.
2. Knez S. Food Object Recognition Using a Mobile Device: State of the Art / S. Knez, L. Šajn. *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. Cham: Springer, 2015. Vol. 9281. P. 366–374. doi: 10.1007/978-3-319-23222-5_45.
3. Fakhrou A. Smartphone-based food recognition system using multiple deep CNN models / A. Fakhrou, J. Kunhoth, S. Al Maadeed. *Multimed Tools Appl*. 2021. Vol. 80. P. 33011–33032. doi: 10.1007/s11042-021-11329-6.
4. Vu T. Wearable Food Intake Monitoring Technologies: A Comprehensive Review / T. Vu, F. Lin, N. Alshurafa, W. Xu. *Computers*. 2017. Vol. 6, Iss. 1. P. 4. doi: 10.3390/computers6010004.
5. Usman M. Recent trends in food intake monitoring using wearable sensors / M. Usman, H. Chen. *IEEE Communications Surveys & Tutorials*. 2021. P. 1–27. doi: 10.48550/arXiv.2101.01378.
6. Kiourt C. Deep learning approaches in food recognition / C. Kiourt, G. Pavlidis, S. Markantonatou. *Machine Learning Paradigms*. Berlin: Springer, 2020. P. 83–108. doi: 10.48550/arXiv.2004.03357.
7. Min W. A Survey on Food Computing / W. Min, S. Jiang, L. Liu et al. *ACM Comput. Surv.* 2019. Vol. 52. P. 1–36. doi: 10.1145/3329168.
8. Мінухін С., Шапошник М. Гібридний підхід до візуально-орієнтованої генерації кулінарних рецептів на основі згорткових нейронних мереж та великих

мовних моделей // Herald of Khmelnytskyi national university. Issue2. 2026 (363). С. 418-434. <https://doi.org/10.31891/2307-5732-2026-363-57>.

9. Ege T. Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions / T. Ege, K. Yanai. *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. New York: ACM, 2017. P. 367–375. doi: 10.1145/3126686.3126742.

10. Мінухін С. В. Використання CNN для багатокласової класифікації та класифікації з багатьма мітками зображення їжі / С. В. Мінухін, М. В. Шапошник // Інформаційно-комунікаційні технології та кібербезпека (ІКТК-2025) : матеріали Міжнародної науково-технічної конференції, 04 – 05 грудня 2025 р. / О.В. – Харків, ХНУРЕ, 2025. – С. 96-100. <https://repository.hneu.edu.ua/handle/123456789/38398>.

11. Wang J. Learning to hash for indexing big data: a survey / J. Wang, W. Liu, S. Kumar, S. F. Chang. *Proceedings of the IEEE*. 2016. Vol. 104, Iss. 1. P. 34–57. doi: 10.48550/arXiv.1509.05472.

12. Wu J. An Introduction to Convolutional Neural Networks. *Nanjing University, China*. 2017. Vol. 5. P. 495. URL: <https://cs.nju.edu.cn/wujx/paper/CNN.pdf> (дата звернення: 08.05.2026).

13. Subhi M. A Deep Convolutional Neural Network for Food Detection and Recognition / M. Subhi, S. Ali. *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. 2018. P. 284–287. doi: 10.1109/IECBES.2018.8626720.

14. Zhou L. Application of Deep Learning in Food: A Review / L. Zhou, C. Zhang, F. Liu et al. *Comprehensive Reviews in Food Science and Food Safety*. 2019. Vol. 18. P. 1793–1811. doi: 10.1111/1541-4337.12492.

15. Bossard L. Food-101 – Mining Discriminative Components with Random Forests / L. Bossard, M. Guillaumin, L. Van Gool. *Computer Vision – ECCV 2014*. Cham: Springer, 2014. Vol. 8694. P. 446–461. doi: 10.1007/978-3-319-10599-4_29.

16. Krizhevsky A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G. Hinton. *Neural Information Processing Systems*. 2012. doi: 10.1145/3065386.
17. Kagaya H. Food Detection and Recognition Using Convolutional Neural Network / H. Kagaya, K. Aizawa, M. Ogawa. *Proceedings of the 22nd ACM international conference on Multimedia*. New York: ACM, 2014. P. 1085–1088. doi: 10.1145/2647868.2654970.
18. Kawano Y. Food image recognition with deep convolutional features / Y. Kawano, K. Yanai. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York: ACM, 2014. P. 589–593. doi: 10.1145/2638728.2641339.
19. Kawano Y. Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation / Y. Kawano, K. Yanai. *Computer Vision – ECCV 2014 Workshops*. Cham: Springer, 2014. Vol. 8927. doi: 10.1007/978-3-319-16199-0_1.
20. Kagaya H. Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network / H. Kagaya, K. Aizawa. *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. Cham: Springer, 2015. Vol. 9281. doi: 10.1007/978-3-319-23222-5_43.
21. Zhang Y. Deep learning in food category recognition / Y. Zhang, L. Deng, H. Zhu et al. *Inf. Fusion*. 2023. Vol. 98. doi: 10.1016/j.inffus.2023.101859.
22. Ciocca G. State Recognition of Food Images Using Deep Features / G. Ciocca, G. Micali, P. Napoletano. *IEEE Access*. 2020. Vol. 8. P. 32003–32017. doi: 10.1109/ACCESS.2020.2973704.
23. Salim N. O. Study for food recognition system using deep learning / N. O. Salim, S. R. Zeebaree, M. A. Sadeeq et al. *Journal of Physics: Conference Series*. 2021. Vol. 1963. doi: 10.1088/1742-6596/1963/1/012014.

24. Srigurulekha K. Food image recognition using CNN / K. Srigurulekha, V. Ramachandran. *2020 International Conference on Computer Communication and Informatics (ICCCI)*. 2020. P. 1–7. doi: 10.1109/ICCCI48352.2020.9104078.
25. Pandey P. FoodNet: Recognizing Foods Using Ensemble of Deep Networks / P. Pandey, A. Deepthi, B. Mandal, N. B. Puhan. *IEEE Signal Processing Letters*. 2017. Vol. 24, Iss. 12. P. 1758–1762. doi: 10.1109/LSP.2017.2758862.
26. Aguilar E. Food recognition using fusion of classifiers based on CNNs / E. Aguilar, M. Bolaños, P. Radeva. *Int. Conf. Image Anal. Proc.* Cham: Springer, 2017. P. 213–224. doi: 10.1007/978-3-319-68548-9_20.
27. Pan L. DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning / L. Pan, S. Pouyanfar, H. Chen et al. *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*. 2017. P. 181–189. doi: 10.1109/CIC.2017.00033.
28. Sarwinda D. Automatic Multi-class Classification of Indonesian Traditional Food using Convolutional Neural Networks / D. Sarwinda et al. *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*. 2020. P. 43–47. doi: 10.1109/IC2IE50715.2020.9274636.
29. Arnita A. Selection of Food Identification System Features Using Convolutional Neural Network (CNN) Method / A. Arnita et al. *Scientific Journal of Informatics*. 2023. Vol. 10 (2). P. 205–216. doi: 10.15294/sji.v10i2.44059.
30. Alahmari S. Food State Recognition Using Deep Learning / S. Alahmari, T. Salem. *IEEE Access*. 2022. P. 130048–130057. doi: 10.1109/ACCESS.2022.3228701.
31. Liu Y. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices / Y. Liu, H. Pu, D.-W. Sun. *Trends in Food Science & Technology*. 2021. Vol. 113. doi: 10.1016/j.tifs.2021.04.042.

32. Ciocca G. CNN-based features for retrieval and classification of food images / G. Ciocca, P. Napoletano, R. Schettini. *Computer Vision and Image Understanding*. 2018. P. 176–177. doi: 10.1016/j.cviu.2018.09.001.
33. Pappas N. A Survey on Language Modeling using Neural Networks. 2012. doi: 10.48550/arXiv.1906.03591.
34. Wang H. A Short Text Classification Method Based on N-Gram and CNN / H. Wang, J. He, X. Zhang, S. Liu. *Chinese Journal of Electronics*. 2020. Vol. 29, Iss. 2. P. 248–254. doi: 10.1049/cje.2020.01.001.
35. Jing K. A Survey on Neural Network Language Models. 2019. doi: 10.48550/arXiv.1906.03591.
36. Jiang T. Classifying Organizations for Food System Ontologies using Natural Language Processing / T. Jiang et al. *2023 Integrated Food Ontology Workshop (FOIS)*. 2023. doi: 10.48550/arXiv.2309.10880.
37. Ambikapathi R. Global food systems transitions have enabled affordable diets but had less favourable outcomes for nutrition, environmental health, inclusion and equity / R. Ambikapathi et al. *Nat Food*. 2022. Vol. 3. P. 764–779. doi: 10.1038/s43016-022-00588-7.
38. Ковальов М. І., Мінухін С. В. Застосування глибокого навчання для дієтологічного контролю продуктів харчування // Харківський національний університет радіоелектроніки. Тези доповідей 30-го Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті». – Харків: ХНУРЕ, 2026. С.
39. Arora N. Machine learning and natural language processing models to predict the extent of food processing / N. Arora, S. Bhagat, R. Dhama, G. Bagler. 2024. doi: 10.48550/arXiv.2412.17217.
40. Hu G. Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods /

G. Hu, M. Ahmed, MR. L'Abbé. *Am J Clin Nutr.* 2023. Vol. 117 (3). P. 553–563. doi: 10.1016/j.ajcnut.2022.11.022.

41. Sirawan P. Food Image Classification with Improved MobileNet Architecture and Data Augmentation / P. Sirawan, S. Olarik. *Proceedings of the 3rd International Conference on Information Science and Systems (ICISS '20)*. New York: ACM, 2020. P. 51–56. doi: 10.1145/3388176.3388179.

42. Leveraging automatic personalised nutrition: food image recognition benchmark and dataset based on nutrition taxonomy / S. Romero-Tapiador, R. Tolosana, A. Morales et al. *Multimedia Tools and Applications*. 2025. Vol. 84. P. 1945–1966. DOI: <https://doi.org/10.1007/s11042-024-19161-4>.

43. Xiao Z., Ling R., Deng Z. FoodCSWin: A High-Accuracy Food Image Recognition Model for Dietary Assessment. School of Artificial Intelligence and Computer Science, Jiangnan University. 2024. 14 p.

44. Rokhva S. Computer Vision in the Food Industry: Accurate, Real-time, and Automatic Food Recognition with Pretrained MobileNetV2 / S. Rokhva, B. Teimourpour, A. Soltani. 2024. doi: 10.48550/arXiv.2405.11621.

45. Metwalli A.-S. Food Image Recognition Based on Densely Connected Convolutional Neural Networks / A.-S. Metwalli, W. Shen, C. Q. Wu. *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*. 2020. P. 027–032. doi: 10.1109/ICAIIIC48513.2020.9065281.

46. Karmakar T. Introduction to DenseNet-121. 2024. URL: <https://www.kaggle.com/code/iamtapendu/introduction-to-densenet-121> (дата звернення: 08.05.2026).

47. Memiş S. A Comparative Study of Deep Learning Methods on Food Classification Problem / S. Memiş, B. Arslan, O. Z. Batur, E. B. Sönmez. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. 2020. P. 1–4. doi: 10.1109/ASYU50717.2020.9259904.

48. Albelwi S. Deep Architecture based on DenseNet-121 Model for Weather Image Recognition. *International Journal of Advanced Computer Science and Applications*. 2022. doi: 10.14569/ijacsa.2022.0131065.
49. Chen C. Deep Learning on Computational-Resource-Limited Platforms: A Survey / C. Chen et al. *Mobile Information Systems*. 2020. P. 1–19. doi: 10.1155/2020/8454327.
50. Leveraging Natural Language Processing and Machine Learning for Evidence-Based Food Security Policy Decision-Making. 2026. URL: <https://arxiv.org/pdf/2603.20425> (дата звернення: 08.05.2026).
51. Application of machine learning in food science and nutrition. *MDPI Foods*. 2025. Vol. 14, Iss. 6. URL: <https://www.mdpi.com/2304-8158/14/6/922> (дата звернення: 08.05.2026).
52. Advances in food recognition systems using deep learning. *MDPI Foods*. 2025. Vol. 14, Iss. 19. URL: <https://www.mdpi.com/2304-8158/14/19/3424> (дата звернення: 08.05.2026).
53. Comprehensive Dataset for Food Computing. 2024. URL: <https://link.springer.com/article/10.1007/s44187-024-00253-x> (дата звернення: 08.05.2026).
54. Machine Learning in Nutrition Research. WUR, 2021. URL: <https://edepot.wur.nl/584260> (дата звернення: 08.05.2026).
55. Precision nutrition driven by machine learning: a paradigm evolution in nutritional assessment and intervention. *PMC*. 2026. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12787445/> (дата звернення: 08.05.2026).
56. Applications of AI, ML and DL in nutrition: a systematic review. *MDPI Nutrients*. 2024. Vol. 16, Iss. 7. URL: <https://www.mdpi.com/2072-6643/16/7/1073> (дата звернення: 08.05.2026).

57. Understanding ML with DL: architectures, workflow and applications. *MDPI Computers*. 2023. Vol. 12, Iss. 5. URL: <https://www.mdpi.com/2073-431X/12/5/91> (дата звернення: 08.05.2026).
58. AI and ML technologies for personalized nutrition: a review. *MDPI Informatics*. 2024. Vol. 11, Iss. 3. URL: <https://www.mdpi.com/2227-9709/11/3/62> (дата звернення: 08.05.2026).
59. Improving nutrition management and nutrition analysis using deep learning: a multi-model approach. *Nature Scientific Reports*. 2025. URL: <https://www.nature.com/articles/s41598-025-32610-w> (дата звернення: 08.05.2026).
60. Chen W. Large-Scale Food Image Datasets and Food Recognition Systems: A Survey / W. Chen, R. Song. *The International Journal of Intelligent Control and Systems*. 2024. Vol. 29, Iss. 2. P. 95–100. doi: 10.62678/IJICS202403.10113.
61. Niu C. R. Review of the deep learning for food image processing / C. R. Niu et al. *International Journal of Agricultural and Biological Engineering*. 2024. Vol. 17, Iss. 5. P. 15–30. doi: 10.25165/j.ijabe.20241705.8975.
62. Lopez Yse D. Text Normalization for Natural Language Processing (NLP). *Medium*. 2021. URL: <https://medium.com/data-science/text-normalization-for-natural-language-processing-nlp-70a314bfa646> (дата звернення: 08.05.2026).
63. Deep learning for food image analysis: a review. *Nature Medicine*. 2025. URL: <https://www.nature.com/articles/s41591-025-04027-5> (дата звернення: 08.05.2026).
64. Jia W. Estimation of Food Volume from 2D Digital Images with Perspective Transformation / W. Jia et al. *Nutrients*. 2017. Vol. 9, Iss. 3. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5359369/> (дата звернення: 08.05.2026).
65. Bach M. Delboeuf Illusion: Optical Illusions & Visual Phenomena. 2026. URL: <https://michaelbach.de/ot/cog-Delboeuf/> (дата звернення: 08.05.2026).

66. Albattah W. Impact of imbalanced features on large datasets / W. Albattah, R. U. Khan. *Frontiers in Big Data*. 2025. Vol. 8. doi: 10.3389/fdata.2025.1455442.
67. Chukwuemeka A. What is Image Preprocessing and Why it Matters in Deep Learning. *Medium*. 2023. URL: <https://medium.com/@anthonychukwuemeka48/what-is-image-preprocessing-and-why-it-matters-in-deep-learning-df7ec3a37260> (дата звернення: 08.05.2026).
68. He K. Deep Residual Learning for Image Recognition / K. He et al. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 770–778.
69. Huang G. Densely Connected Convolutional Networks / G. Huang et al. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 4700–4708.
70. Zhou Z. H. Ensemble Methods: Foundations and Algorithms. CRC Press, 2012. 236 p.
71. ReLU Activation Function in Deep Learning. *GeeksforGeeks*. 2025. URL: <https://www.geeksforgeeks.org/deep-learning-relu-activation-function-in-deep-learning/> (дата звернення: 08.05.2026).
72. Herath H. M. S. S. Comparative Analysis of Deep Learning Architectures for Macular Hole Segmentation in OCT Images / H. M. S. S. Herath et al. *J. Imaging*. 2025. Vol. 11, Iss. 2. P. 53. doi: 10.3390/jimaging11020053.
73. Shahriar T. Comparative Analysis of Lightweight Deep Learning Models for Memory-Constrained Devices. *arXiv:2505.03303*. 2025.
74. PyTorch with DirectML. *Microsoft Learn*. 2024. URL: <https://learn.microsoft.com/en-us/windows/ai/directml/gpu-pytorch-windows> (дата звернення: 08.05.2026).

75. ДСТУ 3008:2015. Державний стандарт України. Інформація та документація. Звіти у сфері науки і техніки. Структура та правила оформлення. – К.: ДП «УкрНДНЦ», 2016. – 31 с.

76. ДСТУ 8302:2015. Бібліографічне посилання. Загальні положення та правила складання. – К.: ДП «УкрНДНЦ», 2016. – 17 с.

77. Методичні вказівки до організації виконання та захисту атестаційної роботи на здобуття другого (магістерського) рівня вищої освіти спеціальності 122 Комп'ютерні науки освітньо-наукова програма «Системне проектування» / Упорядники: І.В. Гребеннік, Н.І. Калита, І.М. Рябченко, З.А. Імангулова, О.Б. Колесник, М.Ю. Вишняк. Харків: ХНУРЕ, 2021. – 54 с.