

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

Факультет інформатики та обчислювальної техніки

Кафедра інформаційних систем та технологій

До захисту допущено:

Завідувач кафедри

_____ Олександр РОЛІК

« ____ » _____ 2025 р.

Дипломний проєкт

на здобуття ступеня бакалавра

**за освітньо-професійною програмою «Інформаційне забезпечення
робототехнічних систем»**

спеціальності 126 «Інформаційні системи та технології»

**на тему: «Підсистема комп'ютерного зору для визначення типу захвату
біонічного протезу»**

Виконала:

студентка IV курсу, групи ІК-12

Верховська Анна Сергіївна _____

Керівник:

доцент кафедри ІСТ, к.т.н., доц.

Олійник Володимир Валентинович _____

Рецензент:

Доцент кафедри ІСТ, к.т.н., доцент,

Лісовиченко Олег Іванович _____

Засвідчую, що у цьому дипломному
проєкті немає запозичень з праць інших
авторів без відповідних посилань.

Студентка _____

Київ – 2025 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 126 «Інформаційні системи та технології»

Освітньо-професійна програма «Інформаційне забезпечення робототехнічних систем»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олександр РОЛІК

« ___ » _____ 2025 р.

ЗАВДАННЯ

на дипломний проєкт студентці

Верховській Анні Сергіївні

1. Тема проєкту «Підсистема комп'ютерного зору для визначення типу захвату біонічного протезу», керівник проєкту Олійник Володимир Валентинович, доцент кафедри ІСТ, к.т.н., затверджені наказом по університету від «23» травня 2025 р. № 1705-с
2. Термін подання студентом проєкту: 9 червня 2025 року
3. Вихідні дані до проєкту: Підсистема комп'ютерного зору для автоматичного визначення типу захвату біонічного протезу, яка має забезпечувати точність класифікації не менше 85%. Алгоритми машинного навчання повинні працювати з вхідними зображеннями роздільною здатністю не менш ніж 224×224 пікселів, а обробка та класифікація захвату одного зображення не повинна перевищувати 0,5 секунди.
4. Зміст пояснювальної записки: задача визначення типу захвату біонічного протезу, інформаційне забезпечення, програмне забезпечення, результати роботи.

5. Перелік графічного матеріалу (із зазначенням обов'язкових креслеників, плакатів, презентацій тощо): блок-схема алгоритму роботи підсистеми, діаграма послідовності підсистеми, діаграма потоків даних, діаграма бізнес-процесів.

6. Дата видачі завдання: 7 березня 2025 року

Календарний план

№ з/п	Назва етапів виконання дипломного проєкту	Термін виконання етапів проєкту	Примітка
1	Аналіз предметної області	01.04.2025 – 04.04.2025	
2	Постановка мети, опис задачі та її особливостей	04.04.2025 – 06.04.2025	
3	Огляд існуючих реалізацій та підходів до вирішення задачі	06.04.2025 – 10.04.2025	
4	Вибір мови програмування та бібліотек для розробки підсистеми	10.04.2025 – 12.04.2025	
5	Огляд та вибір архітектур для визначення типу захвату	12.04.2025 – 16.04.2025	
6	Збір датасету	16.04.2025 – 20.04.2025	
7	Розробка методів визначення типу захвату	20.04.2025 – 12.05.2025	
8	Тестування та аналіз отриманих результатів	12.05.2025 – 23.05.2025	
9	Створення графічного інтерфейсу для візуалізації результатів роботи	23.05.2025 – 25.05.2025	
10	Оформлення пояснювальної записки	25.05.2025 – 31.05.2025	

Студентка

Анна ВЕРХОВСЬКА

Керівник

Володимир ОЛІЙНИК

АНОТАЦІЯ

Підсистема комп'ютерного зору для визначення типу захвату біонічного протезу.

Проект містить 68 с. тексту, 12 рисунків, 8 таблиць, посилання на 24 літературних джерела, 1 додаток та 4 конструкторських документи.

Ключові слова: біонічний протез, комп'ютерний зір, класифікація захвату, машинне навчання, нейронні мережі, few-shot навчання, zero-shot розпізнавання, Deep learning, CLIP, мережі відповідності, прототипічні мережі, fine-tuned моделі, ResNet, генеративні моделі, оптимізація алгоритмів, розпізнавання об'єктів, медичні протези, робототехніка, інтелектуальні системи, OpenRouter API, Python, PyTorch, згорткові нейронні мережі.

Об'єкт роботи — процес розпізнавання об'єктів для керування біонічним протезом.

Мета роботи — підвищення ефективності визначення типу захвату біонічного протезу за допомогою технологій комп'ютерного зору та методів машинного навчання.

Ця робота охоплює аналіз існуючих рішень, проектування алгоритмів обробки зображень та розпізнавання об'єктів, а також практичну реалізацію та тестування розробленої підсистеми. Для досягнення заданих цілей було поставлено ряд завдань, які відповідають структурі роботи. До них належить: постановка мети та задачі, огляд існуючих підходів та наявних реалізацій, аналіз та вибір технологій розробки, розробка програмного забезпечення та опис його структури, оцінка ефективності роботи підсистеми та результати проведених експериментів.

Розроблена підсистема комп'ютерного зору може бути використана компаніями та виробниками біонічних протезів для покращення їх функціональності. Оптимізація алгоритмів класифікації типу захвату дозволяє зробити такі пристрої більш ефективними, інтуїтивними та придатними для повсякденного використання.

SUMMARY

Computer Vision Subsystem for Determining the Grip Type of a Bionic Prosthesis

The project contains 68 pages of text, 12 figures, 8 tables, references to 24 literary sources, 1 appendix, and 4 design documents.

Keywords: bionic prosthesis, computer vision, grip classification, machine learning, neural networks, few-shot learning, zero-shot recognition, deep learning, CLIP model, Matching Networks, Prototypical Networks, fine-tuned models, ResNet, generative models, algorithm optimization, object recognition, medical prostheses, robotics, intelligent systems, OpenRouter API, Python, PyTorch, convolutional neural networks.

The object of this work is the process of object recognition for controlling a bionic prosthesis. The aim of this work is to improve the effectiveness of grasp type recognition for a bionic prosthesis using computer vision technologies and machine learning methods.

This work includes an analysis of existing solutions, the design of image processing and object recognition algorithms, as well as the practical implementation and testing of the developed subsystem. To achieve the set goals, a number of tasks were outlined, which correspond to the structure of the study. These include defining the purpose and objectives, reviewing existing approaches and implementations, analyzing and selecting development technologies, software development and structural description, evaluating the effectiveness of the subsystem, presenting the results of conducted experiments.

The developed computer vision subsystem can be used by companies and manufacturers of bionic prostheses to improve their functionality. Optimizing grasp classification algorithms enables more efficient, intuitive, and practical devices for daily use.

Номер рядка	Формат	Позначення	Найменування	Кільк. аркушів	Номер екзем.	Примітка
1			<u>Документація загальна</u>			
2						
3			Знову розроблена			
4						
5	A4	ІК12.030БАК.006 ПЗ	Пояснювальна записка	68		
6	A3	ІК12.030БАК.006 Д1	Підсистема комп'ютерного	1		
7			зору для визначення типу за-			
8			хвату біонічного протезу.			
9			Блок-схема алгоритму роботи			
10			підсистеми			
11	A3	ІК12.030БАК.006 Д2	Підсистема комп'ютерного	1		
12			зору для визначення типу за-			
13			хвату біонічного протезу.			
14			Діаграма послідовності			
15	A3	ІК12.030БАК.006 Д3	Підсистема комп'ютерного	1		
16			зору для визначення типу за-			
17			хвату біонічного протезу.			
18			Діаграма потоків даних			
19	A3	ІК12.030БАК.006 Д4	Підсистема комп'ютерного	1		
20			зору для визначення типу за-			
21			хвату біонічного протезу.			
22			Діаграма бізнес-процесів			
23						
24						
25						
26						
27						
28						

ІК12.030БАК.006 ТП

Зм.	Аркуш	№ докум.	Підпис	Дата				
Розроб.		Верховська А.С.			Підсистема комп'ютерного зору для визначення типу захвату біонічного протезу. Відомість проекту	Літ.	Аркуш	Аркушів
Керівн.		Олійник В.В.					1	1
						КПІ ім. Ігоря Сікорського		
Затв.								

Пояснювальна записка
до дипломного проєкту
на тему: «Підсистема комп'ютерного зору для визна-
чення типу захвату біонічного протезу»

Київ – 2025 року

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	4
ВСТУП.....	5
1 ЗАДАЧА ВИЗНАЧЕННЯ ТИПУ ЗАХВАТУ БІОНІЧНОГО ПРОТЕЗУ	7
1.1 Постановка мети і задач проєкту	7
1.2 Опис задачі створення підсистеми та її особливостей	8
1.2.1 Загальний огляд предмету дослідження	8
1.2.2 Особливості задачі визначення типу захвату.....	9
1.2.3 Актуальність розробки підсистеми	10
1.3 Існуючі підходи визначення захвату	12
1.3.1 Аналіз та вибір оптимального методу	14
1.3.2 Огляд гожих типів машинного навчання	15
1.3.3 Аналіз та вибір методів машинного навчання	16
1.4 Існуючі реалізації визначення типу захвату	17
1.4.1 Відмінність запропонованого рішення.....	19
1.5 Формулювання конкретних вимог до розробки підсистеми.....	19
Висновки до розділу 1	20
2 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ	22
2.1 Машинне навчання для автоматичного визначення захвату	22
2.1.1 Принцип роботи навчання з учителем та мета-навчання.....	22
2.1.2 Вибір вхідних та вихідних параметрів для нейромережі	23
2.2 Алгоритми, використані при розробці	24
2.2.1 Огляд доступних класів алгоритмів машинного навчання	24
2.2.2 Аналіз та вибір алгоритмів класифікації захвату	25
2.2.3 Особливості роботи мереж відповідності та прототипічних мереж.....	25
2.2.4 Особливості великої мовної моделі Gemma 3	26
2.3 Підготовка набору даних для підсистеми	27

					ІК12.030БАК.006 ПЗ					
З	Лист	№ докум.	Підпис	Дата						
Розробив	Верховська А.С.				Підсистема комп'ютерного зору для визначення типу захвату біонічного протезу. Пояснювальна записка			Літ.	Арк.	Аркушів
Перевірив	Олійник В.В.							Т	2	68
Затв.								КПІ ім. Ігоря Сікорського Група ІК-12		

2.4 Функціональна та структурна схема системи	31
2.4.1 Функціональна схема роботи підсистеми	31
2.4.2 Взаємодія підсистем та зв'язок з біонічним протезом	32
Висновки до розділу 2	34
3 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ	36
3.1 Вибір програмної платформи та мови програмування.....	36
3.1.1 Аналіз та вибір мови програмування	36
3.1.2 Огляд середовищ розробки для систем комп'ютерного зору	38
3.2 Архітектура програмного забезпечення.....	39
3.3 Інтерфейс для візуалізації роботи підсистеми	41
Висновки до розділу 3	43
4 РЕЗУЛЬТАТИ РОБОТИ	45
4.1 Постановка експериментальних задач	45
4.2 Опис експериментальних методів	46
4.3 Результати експериментів	48
4.3.1 Умови експериментів	48
4.3.2 Результати zero-shot методів.....	49
4.3.3 Результати методів few-shot навчання	50
4.3.4 Результати KNN та CLIP	53
4.4 Порівняння та аналіз результатів	55
4.4.1 Загальний аналіз продуктивності моделей.....	55
4.4.2 Порівняння сильних та слабких сторін методів у контексті використання в біонічному протезі.....	58
4.4.3 Вибір оптимального методу для біонічного протезу	60
Висновки до розділу 4.....	62
ВИСНОВКИ	64
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	67
ДОДАТОК А.....	69

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

API (Application Programming Interface) – інтерфейс прикладного програмування, через який клієнтська система може надсилати запити до зовнішньої моделі або сервісу.

CLIP (Contrastive Language-Image Pretraining) – нейромережева модель, що поєднує аналіз тексту та зображень для розпізнавання об'єктів на основі описів.

CNN (Convolutional Neural Network) – згорткова нейронна мережа, яка використовується для аналізу зображень.

GUI (Graphical User Interface) – графічний інтерфейс користувача, який забезпечує зручну візуальну взаємодію з програмним забезпеченням.

IDE (Integrated Development Environment) – інтегроване середовище розробки, що об'єднує редактор коду, налагоджувач і засоби запуску програм.

JSON (JavaScript Object Notation) – текстовий формат обміну даними, що представляє об'єкти у вигляді пар "ключ–значення".

KNN (K-Nearest Neighbors) – алгоритм класифікації, який порівнює нові зразки з найближчими сусідами у векторному просторі ознак.

LDA (Linear Discriminant Analysis) – алгоритм машинного навчання, який використовується для класифікації даних шляхом знаходження лінійних комбінацій ознак, що максимально розділяють класи.

LLM (Large Language Model) – велика мовна модель, яка обробляє природну мову та використовується для аналізу текстових описів у завданнях розпізнавання.

ResNet (Residual Neural Network) – глибока згорткова нейронна мережа, що використовує залишкові зв'язки для покращення навчання глибоких моделей.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		4

ВСТУП

У сучасному світі розвиток біонічних технологій є важливим напрямом інноваційної діяльності, що має значний вплив на медицину, реабілітацію та покращення якості життя людей з обмеженими можливостями. Біонічні протези дозволяють компенсувати втрату функцій кінцівок, забезпечуючи повернення пацієнтів до активного суспільного життя.

Одним із ключових аспектів ефективності біонічних протезів є надійність і точність виконання різних типів захватів. Завдяки інтеграції комп'ютерного зору в систему керування біонічним протезом з'являється можливість автоматичного розпізнавання об'єктів та оптимального вибору типу захвату. Це забезпечує не лише підвищення функціональних можливостей протезів, але й сприяє їх більш інтуїтивному використанню.

Об'єктом дослідження є процес розпізнавання об'єктів для керування біонічним протезом, а предметом – розробка підсистеми комп'ютерного зору для автоматичної класифікації типу захвату, що дозволяє здійснювати більш точну маніпуляцію з предметами у реальному середовищі.

Метою даного дипломного проєкту є підвищення ефективності визначення типу захвату біонічного протезу шляхом використання технологій комп'ютерного зору та методів машинного навчання.

Для її досягнення було поставлено ряд завдань, які відповідають структурі роботи. До них належить: постановка мети та задачі, огляд існуючих підходів та наявних реалізацій, аналіз та вибір технологій розробки, розробка програмного забезпечення та опис його структури, оцінка ефективності роботи підсистеми та аналіз результатів проведених експериментів.

Розроблена підсистема комп'ютерного зору може бути використана компаніями та виробниками біонічних протезів для покращення їх функціональності. Оптимізація алгоритмів класифікації типу захвату дозволяє зробити такі пристрої більш ефективними, інтуїтивними та придатними для повсякденного використання.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		5

Завдяки застосуванню штучного інтелекту, зокрема моделей CLIP та Gemma 3, та методів машинного навчання, таких як мережі відповідності та прототипічні мережі, вдалося створити систему, яка забезпечує автоматичне розпізнавання об'єктів та оптимальний вибір захвату. Це означає, що користувач біонічного протеза не мусить вручну налаштовувати пристрій для кожного предмета—система сама визначає відповідний спосіб захвату, спираючись на аналіз візуальних та текстових даних. Це також відкриває перспективи для застосування біонічних протезів у динамічно змінних умовах, коли користувач стикається з новими або нестандартними об'єктами у побуті чи професійній діяльності.

Таким чином, результати цього дипломного проєкту мають значний потенціал для вдосконалення біонічних протезів, підвищення їхньої автономності, точності та адаптивності, а також для розширення можливостей штучного інтелекту у сфері медичних та робототехнічних технологій.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		6

1 ЗАДАЧА ВИЗНАЧЕННЯ ТИПУ ЗАХВАТУ БІОНІЧНОГО ПРОТЕЗУ

1.1 Постановка мети і задач проєкту

Сформулюємо ціль даної роботи. Головною метою даного проєкту є підвищення ефективності визначення типу захвату. Дана мета була досягнута за рахунок знаходження та порівняння точності 8 методів визначення типу захвату, кожен з яких використовує той чи інший метод машинного навчання або модель штучного інтелекту.

Було обрано наступні критерії ефективності підсистеми:

- точність (англ. Accuracy): частка правильно класифікованих типів захвату серед усіх запитів;
- час інференсу (англ. Inference time): час, який потрібен моделі, щоб видати результат на одне нове вхідне значення;
- адаптивність (англ. Adaptability): рівень здатності системи адаптуватися до нових, раніше не бачених об'єктів;
- масштабованість (англ. Scalability): наскільки легко підтримується додавання нових класів типів захвату.

Завдання дипломного проєкту:

- розкрити суть задачі, її актуальність та особливості;
- провести огляд літератури та проаналізувати існуючі рішення в області біонічних протезів і комп'ютерного зору;
- сформулювати вимоги до підсистеми комп'ютерного зору для біонічного протезу;
- обрати програмну платформу та мову програмування для розробки програмного забезпечення;
- розробити та реалізувати кілька методів класифікації типів захвату об'єктів;
- провести порівняльний експериментальний аналіз обраних методів за визначеними критеріями;
- оцінити ефективність розробленої підсистеми на практичних даних за

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		7

визначеними критеріями та визначити її придатність для використання у протезуванні;

– узагальнити результати, сформулювати висновки та надати рекомендації щодо подальшого вдосконалення системи.

Наукова новизна роботи полягає у розробці алгоритмів комп'ютерного зору, які дозволяють автоматично визначати типи захвату для біонічного протезу з урахуванням умов реального використання. Запропоноване рішення забезпечує підвищення точності, адаптивності та надійності роботи протезу, що сприяє його більш інтуїтивному використанню.

1.2 Опис задачі створення підсистеми та її особливостей

Відповідно до поставлених завдань, спочатку необхідно окреслити задачу дипломного проєкту.

Задача, що розглядається у даному проєкті, полягає у розробці підсистеми комп'ютерного зору для автоматичного визначення типу захвату біонічного протезу. Розробка такої підсистеми вимагає глибокого аналізу методів комп'ютерного зору та машинного навчання, а також дослідження їхньої ефективності у контексті біонічних протезів. Основним завданням є створення алгоритму, який дозволить розпізнавати об'єкти та їхні характеристики (форма, розмір, текстура) у реальному часі та визначати відповідний тип захвату.

Запропонована система може бути використана для вдосконалення сучасних біонічних протезів, підвищення їхньої функціональності та адаптивності до потреб користувачів. Вона також може слугувати основою для подальших досліджень у сфері інтеграції штучного інтелекту в медичні пристрої.

1.2.1 Загальний огляд предмету дослідження

Виходячи з постановки задачі, можна стверджувати, що предметом дослідження є підсистема комп'ютерного зору, яка забезпечує автоматичний аналіз зображень об'єктів для визначення оптимального типу захвату біонічного протезу.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		8

Предметна область дослідження охоплює алгоритми машинного навчання та їхню роль у розпізнаванні об'єктів, методи комп'ютерного зору, що дозволяють працювати з візуальними даними та екстрагувати ключові ознаки, а також інтеграцію цих методів у біонічні протези для адаптації алгоритмів до специфічних умов використання. У межах даної роботи увага зосереджується на порівнянні різних підходів до вирішення задачі, таких як використання згорткових нейронних мереж, моделей трансформерів та інших сучасних технологій.

Такі підходи мають кілька суттєвих переваг. По-перше, вони дозволяють автоматизувати процес вибору типу захвату, що знижує когнітивне навантаження на користувача й робить управління протезом більш інтуїтивним. По-друге, система може адаптуватися до змін у зовнішньому середовищі — таких як освітлення, варіативність об'єктів та індивідуальні особливості рухів. По-третє, застосування методів машинного навчання дає змогу досягти високої точності розпізнавання й забезпечити швидке прийняття рішень у реальному часі.

Так, дослідження у сфері комп'ютерного зору та його інтеграції в біонічні протези є перспективним напрямом, що дозволяє значно покращити якість життя людей з ампутаціями.

1.2.2 Особливості задачі визначення типу захвату

Задача визначення типу захвату для біонічного протезу стикається з низкою особливостей, які ускладнюють її вирішення. Основною проблемою є погана формалізація задачі через різноманіття форм і властивостей об'єктів, які необхідно розпізнати, що значно ускладнює створення універсального алгоритму. Крім цього, дані, що використовуються для аналізу, можуть містити шуми, бути низької якості або отримані в складних умовах освітлення, що впливає на точність розпізнавання.

В межах даного дослідження використовувалися вже навчені моделі, такі як CLIP та ResNet, які дозволяють ефективно витягувати ознаки (features) зображень. Це значно зменшує потребу у великому датасеті для навчання з нуля, оскільки ці моделі вже мають знання про загальні характеристики об'єктів. У процесі роботи

моделі донавчалися на невеликому спеціалізованому датасеті, що містить зображення об'єктів, з якими взаємодіє біонічний протез.

Реальне середовище функціонування протеза додає складності через необхідність взаємодії з рухомими або нестандартними об'єктами, а також через варіабельність даних. Обчислювальні ресурси, доступні в протезі, є обмеженими, що вимагає створення оптимальних рішень, здатних працювати у реальному часі. Окрім того, алгоритми повинні враховувати індивідуальні особливості користувачів, включаючи їх манери використання протезу та специфіку рухів.

Розробка системи вимагає балансу між точністю та швидкістю роботи для забезпечення комфорту і ефективності використання, що становить великий виклик. Ці фактори обумовлюють необхідність ретельного дослідження різних методів вирішення задачі, їх адаптації та порівняння в умовах, наближених до реального використання.

Таким чином, розробка підсистеми комп'ютерного зору для біонічного протезу є складною, але перспективною задачею, що вимагає поєднання сучасних методів машинного навчання, комп'ютерного зору та оптимізації обчислювальних процесів. Використання попередньо навчених моделей дозволяє значно підвищити ефективність системи та адаптувати її до конкретних умов експлуатації.

1.2.3 Актуальність розробки підсистеми

Для розуміння, чому задача є суттєвою та потребує вирішення саме зараз, розглянемо поточний стан справ у цій області.

Останні роки характеризуються стрімким розвитком біонічних технологій, спрямованих на покращення якості життя людей з ампутаціями. Біонічні протези стають дедалі функціональнішими, забезпечуючи користувачам можливість виконувати широкий спектр рухів. Однак, одним із ключових викликів залишається автоматичне визначення типу захвату, що дозволяє протезу адаптуватися до різних об'єктів та ситуацій без необхідності ручного налаштування.

Традиційні методи керування біонічними протезами, такі як міоелектричне

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		10

управління, хоча й забезпечують базову функціональність, часто вимагають значних зусиль з боку користувача для точного контролю рухів. Це може призводити до втоми, зниження ефективності використання протезу та обмеження його функціональності. Використання комп'ютерного зору та штучного інтелекту для автоматичного визначення типу захвату дозволяє значно спростити процес управління, зробити його більш природним та інтуїтивним.

Розробка такої підсистеми є актуальною з кількох причин:

– зростання потреби у високофункціональних біонічних протезах. Кількість людей, які потребують біонічних протезів зростає, а разом із цим підвищуються вимоги до їхньої точності та зручності використання;

– розвиток технологій штучного інтелекту. Сучасні моделі машинного навчання, такі як CLIP та ResNet, дозволяють ефективно розпізнавати об'єкти та витягувати їхні ознаки, що відкриває нові можливості для автоматизації процесу вибору захвату;

– підвищення рівня автономності біонічних пристроїв. Інтеграція комп'ютерного зору дозволяє зменшити залежність користувача від ручного налаштування, що робить протез більш адаптивним до реальних умов використання.

Запропонована підсистема може бути корисною не лише для кінцевих користувачів, а й для виробників біонічних протезів, які прагнуть вдосконалити свої пристрої та зробити їх більш конкурентоспроможними. Використання комп'ютерного зору та штучного інтелекту у біонічних протезах може стати важливим кроком у напрямку створення інтелектуальних адаптивних систем, що здатні самостійно аналізувати навколишнє середовище та приймати рішення щодо оптимального захвату.

Отже, задача розробки підсистеми комп'ютерного зору для визначення типу захвату біонічного протезу є дуже актуальною та важливою. Вирішення цієї проблеми може значно покращити якість життя людей з ампутаціями, зробити біонічні протези більш функціональними та зручними у використанні, а також сприяти подальшому розвитку технологій у сфері медичних пристроїв.

1.3 Існуючі підходи визначення захвату

За останні роки науковці й інженери розробили низку методів, які дозволяють автоматизувати процес вибору відповідного захвату залежно від контексту задачі й властивостей об'єкта.

Системи, які вже використовуються в сучасних біонічних протезах, демонструють важливий прогрес, проте кожен підхід має свої особливості, які слід враховувати при виборі оптимального рішення.

Одним із широко використовуваних методів є міоелектричне розпізнавання, що ґрунтується на аналізі електроміографічних сигналів, зчитаних з м'язів користувача. В поєднанні з ним використовуються алгоритми, такі як лінійний дискримінантний аналіз, для розпізнавання намірів руху та визначення типу захвату. Цей метод забезпечує високу точність і дозволяє адаптувати систему до індивідуальних особливостей користувача.

Інший підхід базується на алгоритмах машинного навчання, які використовуються для аналізу жестів та рухів руки. Сучасні системи використовують алгоритми машинного навчання, такі як нейронні мережі, для аналізу даних про рухи руки та визначення типу захвату. Це дозволяє враховувати складні патерни рухів і забезпечувати більш інтуїтивне управління.

Окрім цього, існують системи із сенсорним зворотним зв'язком, які інтегрують додаткові сенсори для тактильного або візуального сприйняття об'єкта. Це дозволяє користувачеві отримувати інформацію про об'єкт, який він захоплює, і коригувати свої дії відповідно.

Також існують синергетичні підходи для управління рухами, що дозволяють створювати більш інтуїтивні системи захвату, здатні адаптуватися до різних об'єктів. Новітні системи, такі як SoftHand Pro, використовують принципи постуральних синергій для управління рухами. Це дозволяє створювати більш природні та інтуїтивні системи захвату, які адаптуються до різних об'єктів.

Крім того, існують гібридні системи, які поєднують міоелектричне управління з іншими технологіями, такими як інерційні вимірювальні блоки або мобільні

додатки, що робить управління більш гнучким і точним.

Узагальнене порівняння порівняння існуючих підходів та методів для визначення типу захвату описано у таблиці 1.1.

Таблиця 1.1 – Порівняння існуючих підходів визначення захвату

Підхід	Опис методу	Переваги	Недоліки
Мікро-електричне розпізнавання	Аналіз електроміографічних сигналів, отриманих з м'язів, для визначення напрямів руху та класифікації типу захвату	Висока точність, адаптація до індивідуальних особливостей користувача	Чутливість до шумів, залежність від правильної установки електродів та потреба в періодичному калібруванні; використання дорогого спеціалізованого обладнання
Машинне навчання для розпізнавання жестів	Застосування алгоритмів машинного навчання для аналізу даних про рухи руки або відео з метою класифікації типу захвату	Інтуїтивне управління, можливість визначення складних патернів, автоматичне виділення ознак	Потребує великих наборів даних для навчання, може бути ресурсомістким і чутливим до змін умов зйомки; високий час обчислень при використанні складних моделей
Сенсорний зворотний зв'язок	Інтеграція додаткових сенсорів (тактичних, візуальних, тощо) для забезпечення зворотного зв'язку про характеристики об'єкта, що	Покращує точність і адаптивність системи, дозволяє миттєво коригувати позицію захвату	Вимагає додаткового апаратного забезпечення та складної інтеграції, може виникати затримка обробки сигналів; високі витрати на

Підхід	Опис методу	Переваги	Недоліки
	захоплюється, та коригування рухів		придбання та обслуговування сенсорів.
Синергетичні підходи	Використання принципів постуральних синергій, завдяки яким протез імітує природні патерни руху руки при захваті об'єктів	Забезпечують природну та інтуїтивну взаємодію з протезом, скорочують складність управління за рахунок використання природних закономірностей	Обмежена гнучкість при незвичних або індивідуальних варіаціях рухів, складно налаштувати оптимальні синергії для різних умов; висока вартість розробки системи
Гібридні системи	Комбінація декількох технологій для комплексного визначення типу захвату	Забезпечує підвищену точність, надійність та адаптивність завдяки мультисенсорному підходу.	Значне збільшення складності системи та вимог до апаратних засобів; високий час обчислень через обробку даних із різних джерел; зростання витрат на розробку та підтримку комплексного рішення.

1.3.1 Аналіз та вибір оптимального методу

На основі аналізу існуючих підходів до визначення типу захвату біонічного протезу можна зробити висновок, що кожен метод має свої переваги та недоліки. Вибір оптимального підходу залежить від таких ключових факторів, як точність

розпізнавання, швидкість обробки, адаптивність до змінних умов та вимоги до апаратного забезпечення.

Міоелектричне розпізнавання забезпечує високу точність, але має обмеження, пов'язані з необхідністю правильної установки електродів та періодичним калібруванням. Сенсорний зворотний зв'язок дозволяє отримувати додаткову інформацію про об'єкт, але потребує складної інтеграції та додаткових апаратних ресурсів[1]. Синергетичні підходи забезпечують природність рухів, але мають обмежену гнучкість при роботі з нестандартними об'єктами.

Враховуючи ці фактори, найбільш перспективним методом є використання машинного навчання для розпізнавання об'єктів та визначення типу захвату. Цей підхід дозволяє автоматично аналізувати візуальні дані, витягувати ознаки об'єктів та приймати рішення щодо оптимального захвату. Подібні підходи вже демонструють ефективність у завданнях реального часу, таких як аутентифікація з захистом від підробки [2], виявлення захисних масок [3], трекінг погляду [4] чи автоматичне паркування [5]. Це підтверджує універсальність і продуктивність комп'ютерного зору та глибокого навчання в складних прикладних задачах.

Таким чином, у межах даного дослідження було обрано комп'ютерний зір та машинне навчання як основний метод для визначення типу захвату біонічного протезу.

1.3.2 Огляд гожих типів машинного навчання

Машинне навчання можна розділити на кілька основних типів залежно від способу навчання моделі та наявності розмічених даних:

– навчання з учителем (Supervised Learning) – модель навчається на розмічених даних, де кожному вхідному прикладу відповідає правильний вихід. Це дозволяє досягти високої точності у задачах класифікації та регресії;

– навчання без учителя (Unsupervised Learning) – використовується для знаходження структур у даних без попередньої розмітки. Основні методи включають кластеризацію та зменшення розмірності;

– навчання з підкріпленням (Reinforcement Learning) – модель навчається шляхом взаємодії з середовищем, отримуючи винагороди за правильні дії. Використовується у складних задачах управління та автономних системах;

– мета-навчання (Meta Learning) – підхід, що дозволяє моделі швидко адаптуватися до нових задач, використовуючи попередній досвід навчання. Використовується у few-shot навчанні, коли доступна лише невелика кількість прикладів для кожного класу.

Враховуючи задачу автоматичного визначення захвату, у межах цього дослідження було обрано поєднання навчання з учителем та мета-навчання. Це дозволило використати попередньо навчені моделі, які ефективно витягують ознаки зображень, а також адаптувати систему до нових об'єктів без потреби у великому наборі навчальних даних.

З метою порівняння ефективності різних алгоритмів машинного навчання було протестовано кілька підходів, що включали згорткові нейронні мережі, мультимодальні моделі та few-shot класифікацію.

1.3.3 Аналіз та вибір методів машинного навчання

У межах даного дослідження було протестовано декілька підходів до класифікації типу захвату біонічного протезу. Основна мета порівняння — визначити метод, який забезпечує найкращу точність розпізнавання та адаптивність до нових об'єктів.

Для оцінки ефективності було обрано такі моделі та алгоритми:

– модель CLIP та KNN – класичний метод класифікації, який використовує векторну подібність між об'єктами на основі ознак, витягнутих CLIP;

– модель CLIP та zero-shot навчання – застосування попередньо навченого CLIP для класифікації об'єктів без додаткового донавчання, використовуючи текстові дескриптори;

– мережі відповідності (Matching Neural Networks, належить до Meta

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		16

Learning) з CLIP/ResNet – few-shot підхід, що дозволяє швидко адаптуватися до нових класів об’єктів, використовуючи невелику кількість прикладів;

– прототипічні мережі (Prototypical Neural Networks, належить до Meta Learning) з CLIP/ResNet – ще один метод few-shot навчання, який класифікує об’єкти за їхньою прототипічною близькістю;

– донавчена модель CLIP та KNN – модифікований варіант KNN, де CLIP був донавчений на спеціалізованому датасеті для покращення точності визначення типу захвату;

– велика мовна модель (Gemma 3) – використання великої мовної моделі для визначення типу захвату на основі текстових запитів.

Таким чином, в межах дослідження було проведено порівняння різних підходів машинного навчання, що дозволило оцінити їхню ефективність у задачі визначення типу захвату біонічного протезу.

1.4 Існуючі реалізації визначення типу захвату

У сфері біонічних протезів існує низка реалізацій, які використовують різні технології для визначення типу захвату. Кожна з них має свої особливості, переваги та обмеження, що впливає на точність, швидкість та зручність використання.

Одним із прикладів є SoftHand Pro, розроблений в Istituto Italiano di Tecnologia під керівництвом Антоніо Біккі. Ця система базується на синергетичному підході, що дозволяє користувачеві керувати протезом природним способом, використовуючи мінімальну кількість керуючих сигналів [6].

Інший цікавий підхід реалізований у Adaptive Grasping System, створеній у Beijing Institute of Technology під керівництвом Янчао Ванга. Тут використовується міоелектричне розпізнавання та алгоритм лінійного дискримінантного аналізу (LDA), що дозволяє ефективно визначати наміри користувача й адаптувати захват під його потреби [7].

Вдосконалення синергетичних підходів знайшло своє продовження в SoftHand 2 Pro, який є покращеною версією SoftHand Pro. Завдяки інтеграції двох

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		17

рівнів синергій ця система забезпечує вищу точність захвату та адаптивність до різних об'єктів [8].

Комерційні біонічні протези, такі як i-Limb, Bebionic, Össur та Touch Bionics, реалізують індивідуальне управління пальцями, а також використовують міоелектричні датчики для розпізнавання намірів руху. Це дає можливість кожному користувачеві налаштувати протез відповідно до своїх потреб, підвищуючи точність і комфорт під час використання [9].

Узагальнене порівняння існуючих реалізацій визначення типу захвату у біонічних протезах наведено у таблиці 1.2.

Таблиця 1.2 – Порівняння реалізацій біонічних протезів з визначенням захвату

Реалізація	Особливості	Переваги	Недоліки
SoftHand Pro (Istituto Italiano di Tecnologia)	Використання постуральних синергій для управління рухами	Інтуїтивне управління, механічна надійність	Обмежена гнучкість, неможливість точного контролю пальців
Adaptive Grasping System (Beijing Institute of Technology)	Міоелектричне розпізнавання з LDA	Висока точність, адаптація до користувача	Чутливість до шумів, потреба у правильному розташуванні електродів
SoftHand 2 Pro (Istituto Italiano di Tecnologia)	Два рівні синергій, міоелектричне управління	Покращена точність захвату, адаптивність	Висока вартість, складність інтеграції
Комерційні біонічні протези	Індивідуальне управління пальцями	Висока точність контролю кожного пальця	Висока вартість, складність навчання користувача

1.4.1 Відмінність запропонованого рішення

У межах роботи було досліджено кілька підходів до побудови підсистеми комп'ютерного зору, яка автоматично визначає тип захвату для біонічного протезу.

Розглянуті рішення мають низку важливих переваг:

- автоматичне розпізнавання об'єктів за допомогою комп'ютерного зору, що зменшує залежність від міоелектричних сигналів і полегшує керування протезом;
- використання попередньо навчених моделей (зокрема CLIP та ResNet) для витягування ознак, що дозволяє підвищити точність розпізнавання;
- адаптивність до нових або раніше не бачених об'єктів, яку забезпечують методи мета-навчання – зокрема, прототипічні мережі та мережі відповідності;
- зменшення потреби у складному калібруванні та підвищення зручності використання.

Розроблені рішення дозволяють автоматизувати процес вибору типу захвату, підвищити точність розпізнавання, а також забезпечити роботу в умовах обмеженої кількості прикладів для навчання. У разі неможливості визначення типу захвату система може застосовувати дефолтний тип захвату, що гарантує безпечне виконання дії.

Таким чином, запропоновані методи дають змогу створити більш автономну, точну та гнучку систему керування біонічним протезом, що усуває низку обмежень, притаманних традиційним підходам.

1.5 Формулювання конкретних вимог до розробки підсистеми

Після того, як було проаналізовано існуючі підходи та реалізації, необхідно сформулювати конкретні вимоги до власного програмного забезпечення, яке реалізує підсистему комп'ютерного зору для визначення типу захвату біонічного протезу.

Функціональні вимоги:

- автоматичне визначення типу захвату біонічного протезу на основі аналізу

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		19

зображень об'єктів;

- використання попередньо навчених моделей (CLIP, ResNet) для витягання ознак та аналізу візуальних даних;

- застосування нейронної мережі для класифікації типу захвату на основі витягнутих ознак;

- визначення дефолтного захвату у випадках, коли система не може точно ідентифікувати об'єкт або його характеристики;

- обробка вхідних даних у реальному часі для забезпечення швидкого прийняття рішень про тип захвату;

- легка інтеграція з іншими модулями біонічного протезу без необхідності змінювати базові механізми його роботи;

- відображення метрик ефективності, таких як точність класифікації та кількість невизначених об'єктів.

Нефункціональні вимоги:

- швидкість роботи: система повинна обробляти вхідні зображення та визначати тип захвату з мінімальною затримкою;

- адаптивність: можливість розширення та покращення алгоритмів шляхом донавчання моделі на нових даних;

- надійність: стабільність роботи підсистеми в умовах змінного освітлення, різних фонів та складної геометрії об'єктів;

- ефективність використання ресурсів: система має бути оптимізованою для роботи на пристроях з обмеженими обчислювальними можливостями;

- сумісність: можливість інтеграції з різними типами біонічних протезів без необхідності глибокого перепроектування.

Висновки до розділу 1

Виконавши аналіз предметної області автоматичного визначення типу захвату біонічного протезу, було встановлено, що традиційні методи управління біо-

нічними протезами мають низку обмежень, включаючи необхідність ручного налаштування та недостатню адаптивність до різних об'єктів. Це підтверджує актуальність дослідження у напрямку використання комп'ютерного зору для автоматизації процесу вибору захвату.

Постановка мети та опис задачі у підрозділах 1.1 та 1.2 дозволили зрозуміти основні критерії ефективності розроблюваної системи та сформулювати вимоги до її точності, швидкості та адаптивності. Дослідження існуючих підходів у підрозділі 1.3 продемонструвало, що кожен із розглянутих методів має свої переваги й недоліки, що вплинуло на вибір найбільш оптимального підходу.

Було проведено порівняння та аналіз типів машинного навчання у підрозділах 1.3.2 та 1.3.3, що дозволило обрати поєднання навчання з учителем та мета-навчання, оскільки цей підхід забезпечує можливість роботи із невеликими наборами даних та підвищує адаптивність моделі до нових об'єктів.

Також у пункті 1.4 було проведено порівняння існуючих реалізацій біонічних протезів, таких як SoftHand Pro, Adaptive Grasping System, SoftHand 2 Pro та комерційні біонічні протези, наприклад i-Limb чи Vebionic. Було встановлено, що більшість наявних систем мають обмеження у гнучкості, точності або потребують складного калібрування, що підкреслює необхідність використання комп'ютерного зору для покращення керування біонічним протезом.

На основі отриманої інформації у підрозділі 1.6 були сформульовані конкретні вимоги до програмної розробки, що включають використання попередньо навчених моделей, підтримку роботи у реальному часі, легку інтеграцію з біонічним протезом та можливість адаптації до змінних умов.

Отже, підсумовуючи проведений аналіз, у першому розділі було визначено предмет дослідження та його актуальність, сформульовано мету роботи та постановку задачі, розглянуто існуючі підходи та реалізації, проведено аналіз машинного навчання для визначення оптимальної моделі та визначено конкретні вимоги до розробки системи. Ці кроки дали змогу чітко визначити напрямок розробки та очікувані результати, які будуть описані у наступних розділах дипломного проєкту.

2 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

Проаналізувавши предметну область та визначивши конкретну задачу в попередньому розділі, перед початком розробки програмного забезпечення слід детально розглянути застосовані технології, алгоритми та програмні засоби, а також визначити ключові поняття, необхідні для глибшого розуміння процесу створення системи.

2.1 Машинне навчання для автоматичного визначення захвату

Розробка підсистеми комп'ютерного зору для біонічного протезу передбачає використання машинного навчання для автоматичного визначення типу захвату. Це дозволяє протезу адаптивно реагувати на об'єкти, з якими він взаємодіє, забезпечуючи зручність використання та точність виконання рухів.

На основі аналізу існуючих підходів у пункті 1.3 було обрано навчання з учителем та мета-навчання, які дозволяють ефективно використовувати попередньо навчені моделі та забезпечують адаптивність до нових об'єктів.

2.1.1 Принцип роботи навчання з учителем та мета-навчання

У процесі розробки системи визначення типу захвату біонічного протезу було проведено порівняльний аналіз методів машинного навчання, зокрема навчання з учителем (Supervised Learning) та мета-навчання (Meta Learning). Це дозволило оцінити їхню ефективність для класифікації захвату та адаптивність до нових об'єктів, які не входили до початкового навчального набору.

Одним із підходів стало навчання з учителем, де класифікація типу захвату здійснюється на основі розмічених даних. Для цього були протестовані нейронні мережі та алгоритми, що дозволяють аналізувати візуальні та текстові характеристики об'єктів. Використання моделей CLIP та ResNet забезпечило витягування ознак зображень, а алгоритм k-Nearest Neighbors допоміг проводити класифікацію шляхом порівняння схожих елементів у векторному просторі.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		22

Оскільки система повинна ефективно працювати навіть із незнайомими об'єктами, було проведене тестування мета-навчальних підходів, зокрема few-shot навчання. Використані моделі на основі мереж відповідності визначали клас об'єкта за принципом схожості між зображеннями, а прототипічні мережі класифікували об'єкти, орієнтуючись на відстань до середнього представника класу (прототипу).

Окрім цього, була протестована велика мовна модель Gemma 3, яка отримувала зображення разом із текстовими описами та здійснювала прогнозування оптимального типу захвату. Це дозволило оцінити можливості великих мовних моделей у визначенні захвату на основі контекстних ознак та мовної інтерпретації.

2.1.2 Вибір вхідних та вихідних параметрів для нейромережі

Правильний вибір вхідних та вихідних параметрів для нейромережі є критично важливим для ефективності класифікації типу захвату.

Вхідні параметри:

- зображення об'єкта – вхідні дані, які надходять у систему для аналізу;
- витягнуті ознаки – отримані за допомогою CLIP або ResNet, використовуються як основа для класифікації;
- контекстні характеристики – розмір, форма, текстура об'єкта.

Вихідні параметри:

- тип захвату – клас, до якого належить поточний об'єкт (гачковий, циліндричний, точковий тощо);
- ймовірність класифікації – рівень впевненості моделі у своєму рішенні;
- дефолтний захват – використовується у випадках, коли не було можливим визначити захват.

Цей підхід забезпечує гнучкість моделі, дозволяючи їй працювати в реальному середовищі, аналізувати широкий спектр об'єктів та автоматично адаптуватися до змінних умов експлуатації.

2.2 Алгоритми, використані при розробці

Автоматичне визначення типу захвату біонічного протезу потребує ефективних алгоритмів машинного навчання, які здатні витягувати ознаки об'єктів та виконувати їхню класифікацію. У межах даного дослідження було випробувано кілька класів алгоритмів, включаючи згорткові нейронні мережі, мультимодальні моделі та few-shot навчання.

2.2.1 Огляд доступних класів алгоритмів машинного навчання

У процесі вирішення задачі розпізнавання типу захвату були розглянуті різні класи алгоритмів, кожен з яких має свої особливості та переваги. Одним із найбільш ефективних підходів стали згорткові нейронні мережі, які широко застосовуються в комп'ютерному зорі. Вони дозволяють витягувати ознаки зображень і класифікувати об'єкти з високою точністю. У цьому дослідженні була використана модель ResNet, яка завдяки залишковим зв'язкам вирішує проблему зникання градієнта та покращує процес навчання [10].

Окрім класичних згорткових нейронних мереж, перспективним напрямом стали мультимодальні моделі, які поєднують текстові та візуальні дані. Використання CLIP дозволило асоціювати текстові описи із зображеннями, що дало можливість проводити класифікацію без додаткового навчання (zero-shot learning) [11].

Ще одним важливим підходом у дослідженні стало few-shot навчання, яке дає змогу моделі адаптуватися до нових класів навіть за мінімальної кількості навчальних прикладів. Система протестувала два основних типи таких моделей: мережі відповідності, які визначають клас об'єкта через механізм подібності між зображеннями, та прототипічні мережі, що класифікують об'єкти за відстанню до середнього представника класу (прототипу).

Крім того, у дослідженні були розглянуті класичні алгоритми машинного навчання, які не потребують значних обчислювальних ресурсів. Зокрема, KNN використовував подібність об'єктів у векторному просторі для швидкого визначення типу захвату, використовуючи наявні дані.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		24

Окремо була протестована велика мовна модель Gemma 3, яка аналізувала текстові характеристики об'єкта та прогнозувала відповідний тип захвату. Завдяки здатності працювати з контекстними ознаками вона продемонструвала ефективність у задачах класифікації захвату на основі текстових описів.

2.2.2 Аналіз та вибір алгоритмів класифікації захвату

У межах даного дослідження було обрано та реалізовано кілька підходів до класифікації типу захвату:

- модель CLIP та KNN – використовував попередньо навчені фічі CLIP для класифікації об'єктів за допомогою алгоритму k-Nearest Neighbors;
- модель CLIP та zero-shot навчання – застосував CLIP для безконтекстного прогнозування класу об'єкта без додаткового навчання;
- мережі відповідності з CLIP та ResNet – few-shot підхід, що використовував візуальні фічі з CLIP або ResNet для класифікації захвату;
- прототипічні мережі з CLIP та ResNet – альтернативний few-shot метод класифікації на основі відстані до прототипів класів;
- модель CLIP з KNN – KNN у поєднанні з донавченим на власному наборі даних CLIP;
- велика мовна модель (Gemma 3) для прогнозування захвату – використання великої мовної моделі для визначення типу захвату на основі текстових описів.

2.2.3 Особливості роботи мереж відповідності та прототипічних мереж

Навчання з малою кількістю прикладів (Few-shot learning) відіграє важливу роль у розпізнаванні нових об'єктів без потреби у великому навчальному наборі.

Мережі відповідності (Matching Networks) використовують принцип подібності між об'єктами для їх класифікації. Завдяки цьому методу можливо іденти-

фікувати нові об'єкти навіть за наявності мінімальної кількості навчальних прикладів. Однією з основних переваг мереж відповідності є врахування контекстних характеристик об'єкта, що дозволяє моделі адаптуватися до нових варіацій та забезпечує більш точне розпізнавання[12].

Прототипічні мережі (Prototypical Networks) працюють за принципом представлення кожного класу у вигляді середнього вектора, так званого "прототипу". Під час класифікації новий об'єкт порівнюється із цими прототипами, що дозволяє здійснювати швидке визначення класу навіть при обмеженій кількості навчальних прикладів. Завдяки такому підходу модель значно зменшує залежність від глибокого навчання, що робить її особливо ефективною для використання на пристроях із обмеженими обчислювальними ресурсами[13].

2.2.4 Особливості великої мовної моделі Gemma 3

В межах цього дослідження було протестовано велику мовну модель Gemma 3 для прогнозування типу захвату біонічного протезу.

Великі мовні моделі (LLM) є класом штучних нейронних мереж, які навчаються на величезних обсягах текстових даних для виконання завдань, пов'язаних із розумінням та генерацією природної мови. Вони здатні аналізувати текстові запити, знаходити закономірності у мовних структурах та формувати логічні й осмислені відповіді. Завдяки своїй архітектурі такі моделі добре справляються з завданнями машинного перекладу, генерації тексту та аналізу текстових даних.

Однією з ключових особливостей LLM є здатність працювати з довгим контекстом, що дозволяє моделі зберігати інформацію про попередні фрагменти тексту й формувати осмислені відповіді, зважаючи на контекст діалогу. Це особливо корисно у задачах прогнозування або складних аналітичних завданнях, де необхідно враховувати взаємозв'язки між окремими частинами тексту.

Ще однією важливою характеристикою таких моделей є мультимодальність — здатність працювати не лише з текстовою інформацією, а й із зображеннями,

аудіо чи іншими форматами даних. Завдяки цьому LLM можуть аналізувати та інтегрувати різні типи вхідної інформації, що значно розширює їхнє застосування у сфері штучного інтелекту.

Gemini 3 — це мультимодальна мовна модель, яка здатна аналізувати текст та зображення. Її особливістю є підтримка довгого контексту (128K токенів), що дозволяє враховувати складну інформацію під час обробки даних[14].

У цьому проєкті Gemini 3 використовувалася для прогнозування типу захвату на основі зображень об'єктів, доповнених текстовими описами можливих захватів. Модель отримувала вхідне зображення та супровідний промпт з варіантами та описами можливого захвату, після чого здійснювала аналіз і вибір оптимального варіанту.

Використання моделі здійснювалося через OpenRouter API, що дозволяло інтегрувати її в робочий процес без додаткового налаштування серверної інфраструктури. Однак такий підхід мав певні недоліки, зокрема повільну швидкість обробки та обмеження запитів. Імовірно, локальне завантаження моделі могло б суттєво прискорити її виконання.

2.3 Підготовка набору даних для підсистеми

Для ефективного функціонування підсистеми комп'ютерного зору з визначення типу захвату біонічного протезу було необхідно сформувати набір даних, який забезпечував би навчання та оцінку моделей класифікації. Оскільки класифікація здійснюється на основі порівняння нового зображення з уже відомими зразками, критично важливо було створити структурований набір референсних зображень, який слугуватиме основою для алгоритмів класифікації.

Було розроблено два основних набори: опорний набір (Support Set), який містить 8 класів захвату, по 5 зображень на кожен клас. Та запитний набір (Query Set), що мав 80 зображень та використовувався для перевірки роботи алгоритмів та розрахунку точності класифікації. Для оцінки точності класифікації було створено файл `true_labels.json`, що містить правильні відповідності між кожним зображенням

із запитного набору та його можливими типами захватів. Це дозволило об'єктивно порівняти результати класифікації алгоритмами із еталонними значеннями та автоматизувати процес розрахунку точності

Опорний набір виконує роль референсної бази, яка надає еталонні зразки для визначення типу захвату нових зображень. Кожен клас у наборі відповідає окремому типу захвату, а наявність кількох варіантів у кожному класі дозволяє враховувати можливу варіативність об'єктів у реальних умовах.

Запитний набір використовується для тестування моделі – вхідні зображення з цього набору проходять через алгоритм, який здійснює їхню класифікацію, порівнюючи зі зразками з опорного набору. У результаті отримані прогнози захватів порівнюються з відомими правильними відповідями, що дозволяє розрахувати точність роботи алгоритму.

При створенні наборів було приділено увагу таким аспектам:

а) кількість класів та їхнє представлення. У опорному наборі даних включено 8 основних типів захватів:

- 1) циліндричний захват;
- 2) гачковий захват;
- 3) долонний захват;
- 4) сферичний захват;
- 5) трьохпальцевий захват;
- 6) щипковий захват;
- 7) відкритий захват;
- 8) латеральний захват;

б) кожен клас містить 5 різних зображень, які демонструють варіативність об'єктів у межах одного типу захвату. Це дозволяє моделі враховувати невеликі візуальні відмінності та підвищує точність класифікації;

в) особливості зображень у запитному наборі: зображення у запитному наборі не входять до навчального набору, що забезпечує реалістичність тестування. Вони містять різні об'єкти, які можуть частково відповідати еталонним зразкам в опорному наборі, але мають інші параметри, наприклад освітлення чи ракурси;

г) формат та обробка даних: використовувалося масштабування та нормалізація, щоб забезпечити узгодженість характеристик при передачі в модель. Збереження набору здійснювалося у форматі PNG, що дозволяє уникнути значних втрат якості при обробці.

Приклади зображень з опорного набору даних наведено у таблиці 2.1.

Таблиця 2.1 – Приклади зображень в опорному наборі даних

Тип захвату	Приклад зображення 1	Приклад зображення 2
Циліндричний		
Гачковий		
Долонний		
Відкритий		

Зм.	Лист	№ докум.	Підпис	Дата

Тип захвату	Приклад зображення 1	Приклад зображення 2
Латеральний		
Щипковий		
Трьохпальцевий		
Сферичний		

Опорний набір слугував основою для алгоритмів few-shot навчання, а саме мереж відповідності та прототипічних мереж, які використовують концепцію схожості [15] для швидкої класифікації на основі опорних прикладів. У цьому випадку кожне нове зображення порівнювалося із наявними у опорному наборі, а відповідність визначалася за відстанню у векторному просторі.

Для роботи KNN опорний набір використовувався як референсна база: нове зображення класифікувалося на основі близькості до найближчих сусідів у навчальному наборі.

У випадку LLM (Gemma 3) та zero-shot CLIP робота з опорним набором здійснювалася не через пряме навчання, а через текстові підказки та семантичне зіставлення. У Gemma 3 вхідний запит формувався як зображення із текстовими промптами про можливі варіанти захвату, а zero-shot CLIP здійснював порівняння між зображенням та текстовими дескрипторами.

Таким чином, сформований набір даних забезпечив умови для тестування та порівняльного аналізу різних методів класифікації, що дозволило оцінити їхню точність та ефективність у визначенні типу захвату.

2.4 Функціональна та структурна схема системи

2.4.1 Функціональна схема роботи підсистеми

Робота підсистеми комп'ютерного зору реалізується через послідовну взаємодію кількох ключових модулів, починаючи з інтерфейсу користувача та завершуючи виведенням результату класифікації. Алгоритм роботи підсистеми комп'ютерного зору для визначення типу захвату представлено на кресленнику ІК12.030БАК.006 Д1. Функціональна схема, що описує основні етапи обробки даних у підсистемі комп'ютерного зору зображена на рисунку 2.1.



Рисунок 2.1 - Функціональна схема підсистеми

Взаємодію основних модулів підсистеми та їхню послідовність виконання запитів представлено у діаграмі послідовності, наведену в кресленнику ІК12.030БАК.006 Д2. Структурна схема підсистеми зображена на рисунку 2.2.

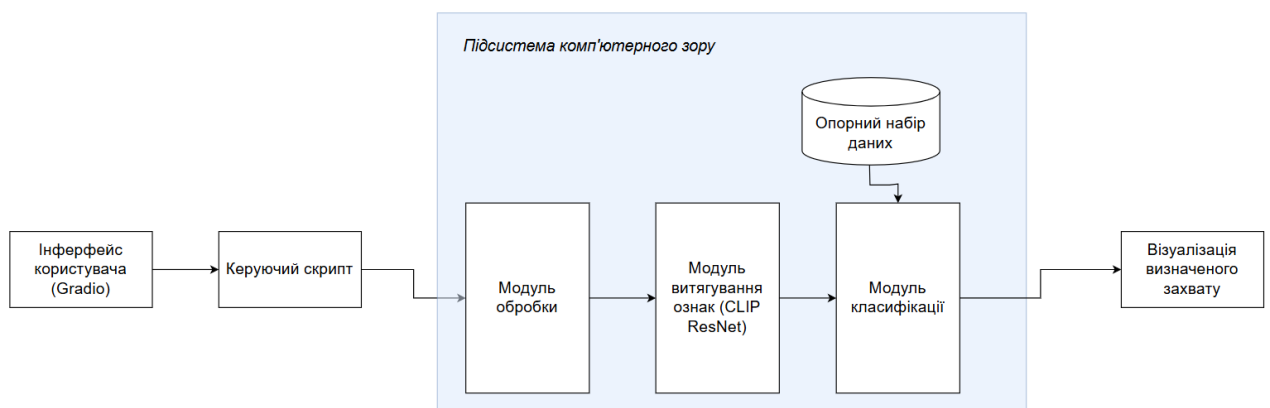


Рисунок 2.2 - Структурна схема підсистеми

2.4.2 Взаємодія підсистем та зв'язок з біонічним протезом

Інтеграція підсистеми у біонічний протез відбувається через програмне забезпечення, яке перетворює результати класифікації у конкретні механічні дії.

Компоненти взаємодії:

- сенсорний модуль – отримує зображення об'єкта;
- підсистема комп'ютерного зору – обробляє дані та визначає тип захвату;
- протезна система управління – отримує команду та виконує відповідний рух.

Механізм передачі даних:

- вхідне зображення → комп'ютерний зір → класифікація;
- результати класифікації → перетворення у команди;
- захват об'єкта за відповідною траєкторією руху.

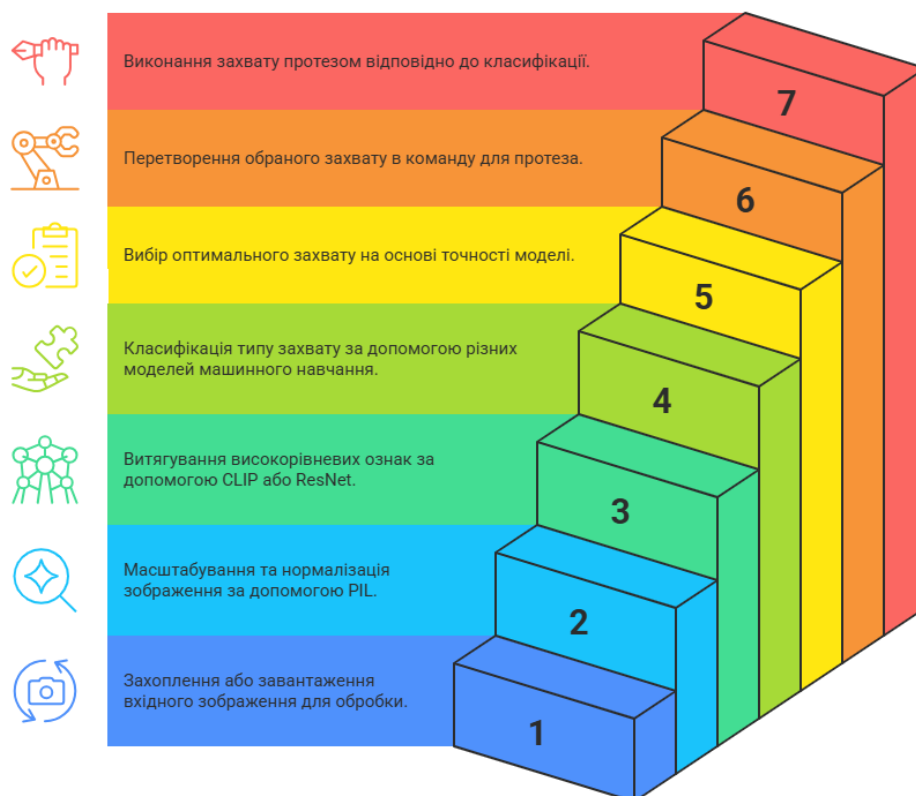
Загальна схема роботи системи протезу базується на послідовності кількох ключових етапів. Спочатку зображення передається у систему через камеру та піддається попередній обробці за допомогою РП, що включає масштабування та нормалізацію даних. Далі, за потреби, система проводить витягування ознак, або за допомогою CLIP, що аналізує високорівневі особливості об'єкта, або за допомогою ResNet, що розглядає його текстуру, форму та деталі.

Наступним етапом є класифікація типу захвату, що здійснюється на основі різних методів, таких як мережі відповідності та прототипічні мережі (few-shot навчання), алгоритм KNN, Zero-shot навчання за допомогою CLIP або аналіз мовної моделі Gemma 3. Після визначення класу об'єкта система обирає оптимальний захват, при цьому пріоритет надається варіанту з найвищою оцінкою. Якщо класифікація не дає чіткого результату, використовується дефолтний захват.

Завершальний етап передбачає передачу команди до біонічного протезу, де визначений тип захвату трансформується у керуючий сигнал для його двигунів. Після цього система виконує рух, узгоджений із результатами класифікації.

Такий алгоритм забезпечує високу гнучкість та дозволяє використовувати різні методи класифікації, адаптуючись до нових об'єктів. Схему загальної роботи системи протезу зображено на рисунку 2.3.

Етапи роботи алгоритму



Made with Napkin

Рисунок 2.3 - Візуалізація загальної роботи системи протезу

Ця інтеграція забезпечує автоматичне керування біонічним протезом без потреби у складному налаштуванні користувачем, що значно спрощує використання пристрою. Процеси передачі, обробки та класифікації даних у підсистемі представлено в діаграмі потоків даних, що зображена на кресленнику ІК12.030БАК.006 ДЗ. Послідовність виконання бізнес-процесів наведено у діаграмі бізнес-процесів на кресленнику ІК12.030БАК.006 Д4.

Висновки до розділу 2

Провівши детальний огляд принципів роботи машинного навчання у підрозділі 2.1, було описано основні концепції навчання з учителем та мета-навчання, а також особливості використання моделей CLIP, ResNet, мереж відповідності, прототипічних мереж, алгоритму KNN та великої мовної моделі Gemma. Це дозволило

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		34

сформувані загальну архітектуру системи та визначити ключові параметри, необхідні для точності класифікації захвату.

У підрозділі 2.2 було проведено огляд та аналіз доступних класів алгоритмів машинного навчання, включаючи згорткові нейронні мережі, мультимодальні моделі, few-shot навчання, класичні алгоритми та великі мовні моделі. У пункті 2.2.3 було розглянуто принцип роботи мереж відповідності та прототипічних мереж, які можуть бути корисними для розпізнавання нових об'єктів без великого обсягу навчальних даних.

У підрозділі 2.3 було описано підготовку набору даних, який слугує основою для навчання моделей машинного навчання та оцінки їхньої продуктивності. Було створено опорний набір даних, що включає 8 класів захвату, кожен із яких представлений п'ятьма зображеннями. Цей набір виконує роль референсної бази, яка дозволяє алгоритмам порівнювати нові об'єкти зі зразками. Також був розроблений запитний набір даних, що складається із 80 тестових зображень, які використовувалися для перевірки роботи моделей та розрахунку точності класифікації.

У підрозділі 2.4 було розглянуто функціональну схему підсистеми, яка описує ключові етапи роботи: отримання зображення, витягування ознак, класифікація типу захвату та візуалізація обраного захвату, а також структурну схему. Окрім цього, було проаналізовано механізм взаємодії підсистеми з біонічним протезом, що забезпечує автоматичне керування захватом без потреби у складному налаштуванні користувачем.

Таким чином, проведені дослідження та здійснені вибори забезпечили ефективне та оптимальне інформаційне забезпечення проєкту, що дозволило визначити найбільш продуктивні технології, алгоритми та середовища розробки для реалізації підсистеми комп'ютерного зору у біонічному протезі.

3 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

Лише після встановлення конкретного інформаційного забезпечення з повним розумінням засобів, алгоритмів та технологій, що будуть використовуватися, можна починати розробку підсистеми комп'ютерного зору для визначення типу захвату біонічного протезу.

У цьому розділі розглянуто вибір мови програмування та програмного середовища, архітектуру програмного забезпечення та інтерфейс користувача.

3.1 Вибір програмної платформи та мови програмування

Проектування системи для визначення типу захвату біонічного протезу вимагає використання ефективних інструментів, які забезпечують високу продуктивність при роботі з нейронними мережами, а також зручність при обробці зображень. Одним із ключових етапів розробки є вибір оптимального програмного середовища та мови програмування, які дозволять реалізувати необхідні функціональні можливості системи.

3.1.1 Аналіз та вибір мови програмування

У якості основної мови програмування для реалізації всієї системи було обрано Python. Цей вибір зумовлений кількома вагомими факторами.

По-перше, Python має широку підтримку в сфері штучного інтелекту та машинного навчання. Саме під цю мову розроблено більшість сучасних фреймворків, зокрема PyTorch, TensorFlow, Keras, HuggingFace Transformers, що забезпечує швидкий старт та доступ до великої кількості готових рішень. У рамках цього проєкту, використання Python дало змогу інтегрувати попередньо навчені моделі (CLIP, ResNet) без необхідності реалізації низькорівневого коду.

По-друге, Python має простий та зрозумілий синтаксис, що суттєво спрощує створення, тестування та налагодження коду, особливо у наукових дослідженнях та прототипуванні. Це дозволяє зосередитися на логіці алгоритмів, а не на мовних

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		36

особливостях.

По-третє, важливою перевагою Python є наявність системи віртуальних середовищ (venv, conda), які дають змогу точно відтворити середовище розробки на іншому комп'ютері, включаючи версії бібліотек та залежностей [16]. Це суттєво спрощує процес перенесення та тестування проєкту, особливо при використанні віддалених ресурсів з GPU.

Для порівняння, було також розглянуто інші мови програмування, які іноді використовуються у сфері глибокого навчання. Результати порівняння наведено у таблиці 3.1.

Таблиця 3.1 – Порівняння альтернативних мов програмування

Мова програмування	Переваги	Недоліки
C++	Висока продуктивність та швидкість виконання. Ефективне використання пам'яті. Підтримка бібліотек, таких як TensorFlow C++ API.	Більш складний синтаксис, що уповільнює розробку. Висока складність реалізації нейромереж у порівнянні з Python. Менша кількість готових бібліотек для глибокого навчання.
Java	Стабільність та безпека платформи. Підтримка багатопотоковості, що може бути корисним для паралельних обчислень. Використання у великих корпоративних рішеннях.	Обмежена кількість бібліотек для сучасного машинного навчання. Відносно низька продуктивність у порівнянні з C++ при роботі з нейромережами. Складність інтеграції з популярними фреймворками для комп'ютерного зору.

На відміну від C++ та Java, Python забезпечує швидку розробку, багатий набір

інструментів для роботи з машинним навчанням та легкість інтеграції з середовищами PyTorch та PIL. Він також має велику спільноту розробників, що дозволяє швидко знаходити рішення проблем та використовувати готові моделі.

3.1.2 Огляд середовищ розробки для систем комп'ютерного зору

Під час проектування та реалізації алгоритмів класифікації типу захвату було протестовано кілька сучасних фреймворків машинного навчання. У результаті, для основної реалізації було обрано бібліотеку PyTorch, яка останніми роками стала одним із найбільш популярних інструментів у сфері глибокого навчання.

PyTorch є гнучким фреймворком з відкритим кодом, орієнтованим як на дослідницьке, так і на прикладне використання [17]. Його основною перевагою є динамічна обчислювальна графова модель, що дозволяє швидко змінювати архітектуру мережі, що особливо корисно у процесі експериментування з моделями, такими як CLIP, ResNet або кастомні енкодери. Крім того, PyTorch має зручний API, потужну підтримку CUDA для використання обчислювальних можливостей GPU, а також широку спільноту, що спрощує вирішення проблем і пошук документації.

Інші розглянуті середовища:

– TensorFlow — хоча ця бібліотека також є потужним інструментом для побудови нейронних мереж, її застосування у рамках цього проєкту виявилось менш доцільним через вищу складність інтеграції з попередньо навченими моделями типу CLIP, які у більшості випадків мають підтримку саме в PyTorch;

– OpenCV — бібліотека для обробки зображень, яка надає широкий набір інструментів для роботи з зображеннями, камерами, сегментацією тощо. Проте у поточному дослідженні вона не використовувалася. Замість неї було обрано бібліотеку PIL.

Для завантаження попередньо навчених моделей було використано додаткові бібліотеки, які інтегруються з PyTorch. Зокрема, модель CLIP імпортувалася за допомогою бібліотеки HuggingFace Transformers, яка забезпечує зручний доступ до великої кількості моделей із відкритим кодом [18]. Це дозволило використовувати

CLIP як у zero-shot режимі, так і в різних методах few-shot класифікації, зокрема у мережах відповідності, прототипічних мережах та KNN.

Модель ResNet, яка застосовувалася як енкодер зображень у мережах відповідності та прототипічних мережах, імпортувалася через бібліотеку torchvision — офіційне розширення PyTorch, що містить набір перевірених архітектур нейромереж і засобів попередньої обробки зображень.

Завдяки поєднанню transformers та torchvision з PyTorch вдалося забезпечити стабільну, ефективну та гнучку реалізацію функціональних компонентів системи.

Таким чином, середовищем розробки було обрано PyTorch для нейромереж та PIL для базової обробки зображень, а основною мовою програмування – Python, що забезпечує гнучкість та швидкість розробки.

У рамках дослідження навчання та тестування моделей виконувалося на віддаленому комп'ютері з GPU, що забезпечило стабільну продуктивність при роботі з CLIP, ResNet, KNN та іншими алгоритмами. Використання графічного процесора дозволило оптимізувати обчислювальні витрати та забезпечити швидке опрацювання вхідних даних.

3.2 Архітектура програмного забезпечення

Архітектура створеного програмного забезпечення побудована на модульному принципі, що дозволяє забезпечити зрозумілу логіку виконання, а також гнучкість і простоту подальшого розширення системи. Загальна структурна схема роботи системи була представлена раніше у підрозділі 2.4.1 (рисунок 2.2), і в цьому пункті наведено опис реалізації архітектури з програмної точки зору.

Основна мета програмної реалізації полягає в забезпеченні зручного інтерфейсу для завантаження зображень, запуску обраного методу класифікації, виводу результатів користувачу, а також у забезпеченні можливості дослідження точності різних підходів у рамках експериментального дослідження. Всі компоненти програми реалізовані мовою Python із використанням популярних бібліотек для обробки зображень, побудови моделей глибокого навчання та створення простого

графічного інтерфейсу.

Зокрема, графічний інтерфейс користувача реалізовано за допомогою бібліотеки Gradio. Інтерфейс дозволяє завантажити зображення, обрати метод класифікації (наприклад, KNN, мережі відповідності, прототипічні мережі, CLIP Zero-Shot або LLM), а також переглянути результат прогнозу. Інтерфейс мінімалістичний, оскільки основна функціональність зосереджена на обчислювальних модулях, які запускаються у фоновому режимі після взаємодії користувача з GUI.

Обробка зображення включає попередню обробку, таку як зміна розміру, нормалізація піксельних значень, а також формування тензорів для подачі в модель. Наступним етапом є обчислення або завантаження ознак, які було отримано за допомогою моделей CLIP або ResNet. У випадку few-shot методів та KNN також використовується попередньо підготовлений набір підтримуючих зразків, embedding яких зберігаються в окремих файлах [19].

Усі обчислювальні методи організовані у вигляді окремих функцій або класів. Вони приймають на вхід embedding зображення-запиту та набір підтримки, і повертають передбачений клас захвату. Метод KNN реалізований як базовий приклад методу на основі схожості. Мережі відповідності і прототипічні мережі реалізують підходи few-shot навчання, що дозволяють працювати з обмеженою кількістю даних. Крім того, реалізовано zero-shot підхід із використанням CLIP, де модель обирає найбільш релевантний текстовий опис без додаткового навчання. Також проводилися експерименти з використанням великої мовної моделі Gemma 3, яка прогнозувала тип захвату на основі текстових описів об'єкта.

Для взаємодії з великою мовною моделлю Gemma 3 у цьому дослідженні використовувався API OpenRouter, який дозволяє інтегрувати передові нейромережі без необхідності локального розгортання великих моделей. OpenRouter надає можливість виконувати запити до моделей через інтернет, що значно спрощує доступ до потужних мовних алгоритмів без необхідності самостійно завантажувати їх та підтримувати інфраструктуру для обчислень [20].

API OpenRouter забезпечує гнучкий формат запитів, що включає передавання текстових промптів та зображень у багатомодальних сценаріях. У рамках цього

проекту модель Gemma 3 отримувала поєднання текстового запиту та відповідного зображення, що дозволяло оцінювати її здатність правильно класифікувати тип захвату за комбінованими вхідними даними.

Запити до API формувалися у JSON-форматі, де вказувалася обрана модель, список повідомлень користувача та закодоване у base64 зображення для аналізу. Результати класифікації отримувалися у вигляді текстових відповідей, які потім оброблялися за допомогою Python для визначення відповідності передбаченого захвату до ground-truth міток у тестовому наборі.

Використання OpenRouter забезпечило універсальний доступ до моделі Gemma 3, дозволяючи її інтеграцію без необхідності розгортання на локальному обладнанні. Проте слід враховувати, що моделі великого розміру, такі як Gemma 3 12B або 27B, можуть потребувати значних обчислювальних ресурсів при виконанні запитів, що може впливати на затримки та продуктивність, якщо їх застосовувати у пристроях із обмеженими ресурсами, таких як біонічні протези.

Завдяки OpenRouter вдалося протестувати можливості Gemma 3 у режимі zero-shot класифікації, що дозволило оцінити її ефективність у контексті розпізнавання типів захвату без попереднього навчання на спеціалізованому датасеті.

Хоча порівняння ефективності методів не реалізовано у вигляді окремого модуля, воно було виконане шляхом запуску кожного методу на однаковому запитному наборі, фіксації точності прогнозування, та подальшого ручного занесення отриманих результатів до зведеної таблиці. Ці дані використовувалися для побудови підсумкових графіків, що порівнюють точність методів. Таким чином, програмне забезпечення є інструментом, який дозволяє не лише проводити індивідуальні прогнози, а й оцінювати якість роботи моделей у задачах few-shot та zero-shot класифікації захвату.

3.3 Інтерфейс для візуалізації роботи підсистеми

Інтерфейс для візуалізації роботи підсистеми реалізований з використанням бібліотеки Gradio, що забезпечує швидке створення веб-інтерфейсів [21]. Такий

підхід дозволив створити інтуїтивно зрозумілу та зручну оболонку для взаємодії з моделлю без потреби встановлення додаткового програмного забезпечення або складного розгортання.

Після запуску скрипту Gradio відкриває локальну веб-сторінку (рис. 3.1), де користувач може обрати або перетягнути зображення для класифікації, а також обрати один із доступних методів прогнозування з випадаючого меню. До таких методів належать: CLIP Zero-Shot, KNN з CLIP, Matching Networks, Prototypical Networks, та KNN з Fine-Tuned CLIP. Після натискання кнопки прогнозування система обробляє зображення й повертає передбачений(і) тип(и) захвату (рис. 3.2).

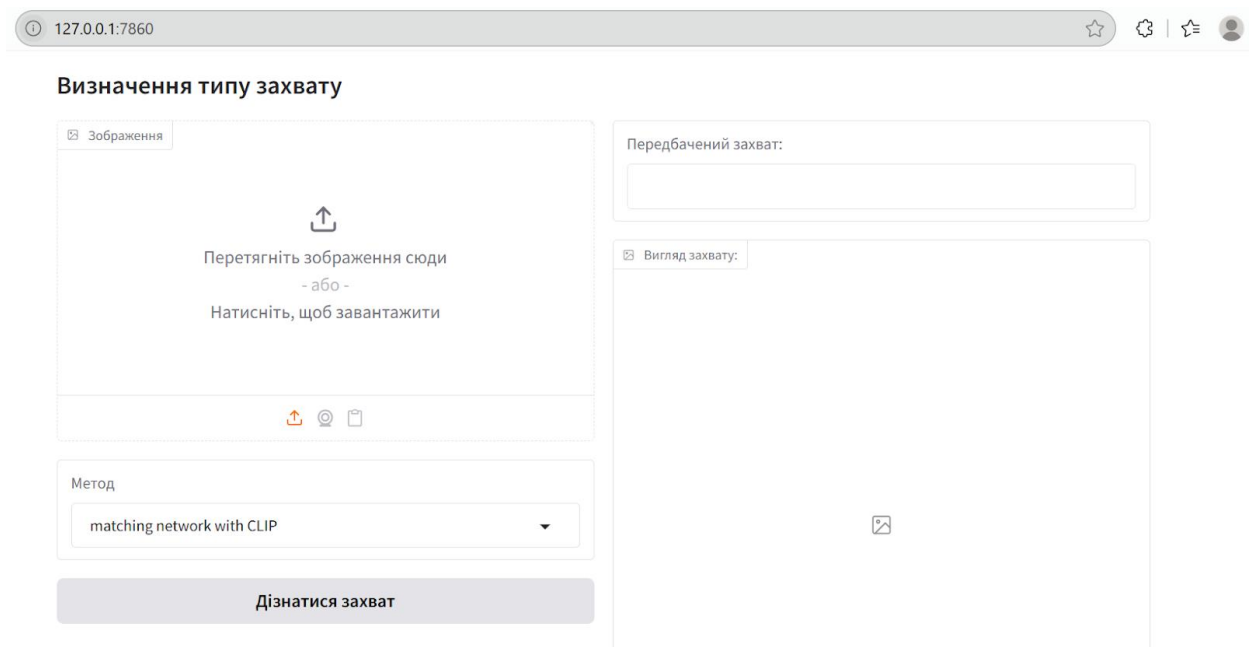


Рисунок 3.1 - Початковий вигляд вебсторінки

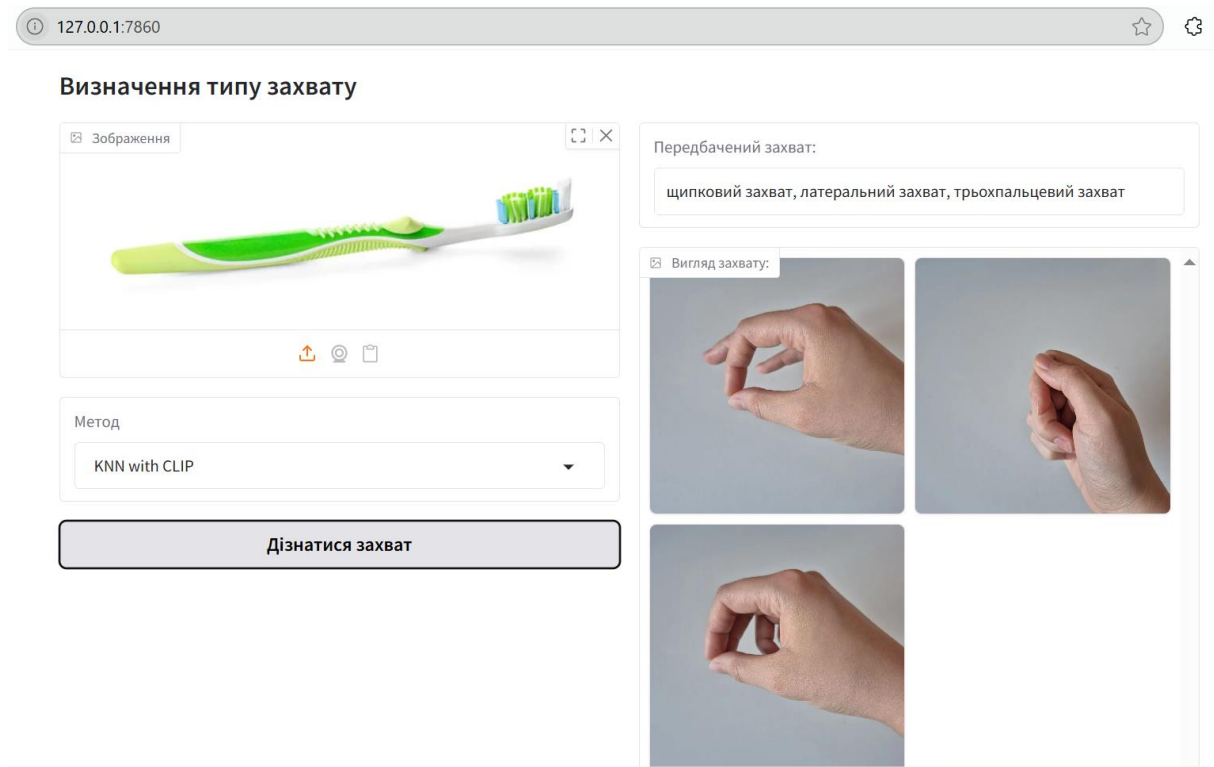


Рисунок 3.2 - Вигляд вебсторінки після визначення захвату

Інтерфейс розрахований переважно на демонстрацію роботи моделі та на зручність проведення одиничних тестів. Основна частина експериментальної роботи, включно з обробкою запитного набору зображень, обрахунком точності та побудовою порівняльних діаграм, виконується за допомогою окремих скриптів, без участі інтерфейсу. Водночас, Gradio дозволяє легко масштабувати або модифікувати інтерфейс у майбутньому, наприклад, для інтеграції з веб-додатком або демонстрації роботи моделі іншим користувачам.

Таким чином, реалізований веб-інтерфейс виконує функцію простого, але ефективного інструменту для взаємодії з моделлю, дозволяючи швидко перевірити результати її роботи на нових прикладах.

Висновки до розділу 3

У цьому розділі було здійснено вибір програмного середовища, мови програмування та ключових бібліотек для реалізації системи класифікації типу захвату.

На основі аналізу сучасних засобів розробки було обрано мову Python у поєднанні з фреймворками PyTorch та HuggingFace, що забезпечують гнучкість і високу продуктивність при роботі з моделями глибокого навчання. Для використання великої мовної моделі Gemma 3 використовувався API OpenRouter.

Також було розроблено загальну архітектуру програмного забезпечення, що включає модулі для обробки зображень, витягнення ознак, класифікації та інтерфейсу користувача. Особливу увагу приділено модульності коду та можливості розширення функціоналу в подальших експериментах.

Було створено простий і функціональний інтерфейс візуалізації роботи підсистеми, який дозволяє легко завантажувати зображення, обирати метод класифікації та кількість передбачень та переглядати результати, що забезпечує зручність взаємодії з системою для дослідників та розробників.

Таким чином, реалізована програмна основа є гнучкою платформою для проведення подальших експериментів, результати яких будуть проаналізовані у наступному розділі.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		44

4 РЕЗУЛЬТАТИ РОБОТИ

У цьому розділі представлено експериментальну оцінку розробленої системи класифікації типу захвату об'єкта за зображенням. Метою є порівняння різних методів машинного навчання та мультимодальних моделей у контексті обмеженої кількості даних на клас. Розглянуто як zero-shot та few-shot підходи, так і методи з донавчанням (fine-tuning).

Проаналізовано точність кожного методу, особливості їх використання, а також потенційні переваги та обмеження. Експерименти проводились на власноруч сформованому наборі даних, опис якого наведено у попередніх розділах.

4.1 Постановка експериментальних задач

Метою експериментального етапу дослідження є оцінка ефективності різних підходів до класифікації типу захвату об'єкта на зображенні в умовах обмеженого обсягу даних. З огляду на характер поставленої задачі та обмежену кількість прикладів у кожному класі, було обрано підходи, які включають zero-shot, few-shot та fine-tuning стратегії.

Для перевірки гіпотез про придатність різних архітектур та методів у сценаріях з малими даними, було сформульовано такі експериментальні задачі:

- порівняти якість zero-shot підходів на основі попередньо навчених мультимодальних моделей, зокрема CLIP та великої мовної моделі Gemma, що працює лише з текстовим описом;
- оцінити ефективність few-shot методів, зокрема мереж відповідності та прототипічних мереж, з різними фічекстракторами (CLIP та ResNet), у сценарії 5 прикладів на клас;
- перевірити вплив донавчання (fine-tuning) CLIP на точність класифікації у поєднанні з простим KNN-класифікатором;
- провести загальне порівняння результатів усіх підходів з метою виявлення найпридатнішого методу для класифікації типу захвату в умовах обмежених ресурсів.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		45

Кожен метод оцінювався за метриками Топ-3 точності, з використанням однакового розбиття на опорний та запитувану запитний набори. Далі у підрозділі 4.2 наведено стислий опис усіх експериментальних підходів, а в підрозділі 4.3 — їх результати та порівняльний аналіз.

4.2 Опис експериментальних методів

Обрані підходи, що були використані для класифікації типів захвату на основі зображень об'єктів, охоплюють як zero-shot стратегії, які не потребують навчання на даних, так і few-shot підходи, що працюють з обмеженою кількістю прикладів для кожного класу. Також було протестовано прості евристичні методи на основі найближчих сусідів (KNN) та метод із використанням великої мовної моделі для класифікації на основі текстового опису зображення.

Першу групу експериментів склали zero-shot підходи, серед яких ключовим є застосування моделі CLIP. Цей метод дозволяє проводити класифікацію без додаткового донавчання: зображення об'єкта кодується вектором, який порівнюється з векторами, отриманими з текстових описів кожного класу захвату [22]. Клас, чий опис має найвищу косинусну подібність до зображення, вважається передбаченням моделі. Окрім CLIP, було протестовано можливості великої мовної моделі Gemma 3, представленої у трьох версіях: Gemma 3 4B, Gemma 3 12B та Gemma 3 27B, кожна з яких була натренована на різній кількості даних. Останнє число в назві цих моделей вказує на кількість їх навчальних даних у мільярдах. Ці моделі використовують текстово-візуальну інтеграцію, проте їхній механізм класифікації суттєво відрізняється від CLIP.

У процесі експериментів Gemma 3 отримувала комбінацію зображення та текстового промπτу, який включав опис можливих типів захвату та питання щодо того, який саме захват є найбільш коректним для даного об'єкта. Це наближало процес аналізу до логічного міркування на основі контексту, а не лише візуального зіставлення.

Наступною групою методів були підходи, засновані на алгоритмі KNN. У

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		46

цьому випадку для кожного вхідного зображення здійснювався пошук найбільш схожих опорного зображень на основі косинусної подібності ембедінгів. Для витягання цих ембедінгів використовувалася модель CLIP. Було проведено два варіанти експериментів: перший — з використанням базової CLIP без додаткового донавчання, другий — із застосуванням моделі, попередньо донавченої на задачі класифікації типів захвату. Очікувалося, що другий варіант дасть вищу точність за рахунок адаптації ембедінг-простору до специфіки завдання.

Донавчання (fine-tuning) моделі CLIP на задачі класифікації типів захвату стало одним із ключових експериментів. Модель було адаптовано на основі невеликого навчального набору, що складався зі зображень, класифікованих за типами захоплення. Метою донавчання було оптимізувати embedding-простір моделі таким чином, щоб вона краще відокремлювала різні класи, специфічні для нашої задачі. Зокрема, використовувався класифікаційний projection head, який додавався до візуального енкодера CLIP і тренувався з використанням крос-ентропійної функції втрат.

Після завершення донавчання отримана модель використовувалася як фіксований енкодер для обчислення ембедінгів зображень. У методі KNN ці ембедінги порівнювалися з ембедінгами зображень із опорного набору. Таким чином, навіть без прямого використання класифікаційного head у фазі тестування, адаптований embedding-простір позитивно впливав на якість пошуку найближчих сусідів та класифікацію.

Такий підхід дозволив зберегти гнучкість KNN-методу, водночас підвищивши його ефективність завдяки попередньому навчанні моделі на релевантних прикладах.

Найчисленнішою категорією експериментів стали методи few-shot навчання, а саме мережі відповідності та прототипічні мережі. Обидва підходи передбачають наявність опорних зображень для кожного класу, що використовуються для передбачення класу нового зображення. У мережах відповідності класифікація виконується шляхом обчислення ваг до кожного опорного прикладу, а ймовірність кож-

ного класу визначається як зважена сума відповідних міток. У прототипічних мережах натомість для кожного класу розраховується «прототип» — середній вектор ембедінгів зображень з опорного набору, а потім визначається відстань запитного вектора до кожного з цих прототипів. В рамках кожного з цих двох підходів були протестовані два варіанти витягу ознак: з використанням CLIP та ResNet як моделей-енкодерів. Це дозволило порівняти ефективність сучасної мультимодальної моделі зі стандартною архітектурою для витягу ознак із зображень.

Таким чином, у проведених експериментах було охоплено широкий спектр стратегій класифікації, що дозволяє провести комплексне порівняння їх результатів на обмеженому наборі зображень з типами захватів.

4.3 Результати експериментів

Проведені експерименти дозволили отримати важливі дані щодо ефективності різних методів класифікації типу захвату об'єкта за зображенням. Оскільки основним завданням було визначити, який алгоритм є найбільш придатним у умовах обмеженої кількості навчальних прикладів, аналіз включав тестування zero-shot, few-shot та fine-tuned методів.

4.3.1 Умови експериментів

Для чесного порівняння методів визначення типу захвату варто перш за все визначити правила, що будуть стосуватися оцінки точності у всіх методах. Методи будуть порівнюватися за такими критеріями:

– топ-1 точність: точність передбачення моделлю одного типу захвату. Тобто, якщо передбачений до зображення захват також наявний у списку можливих захватів до нього у списку правильних захватів (ground truth labels), то передбачення вважається правильним;

– топ-3 точність: точність передбачення моделлю трьох типів захвату. Тобто, якщо хоча б один з передбачених захватів наявний у списку можливих захватів до зображення у списку істин то передбачення вважається правильним.

Також у методах, де було присутнє навчання, а саме – мережах відповідності, прототипічних мережах та донавчання моделі CLIP обчислювалися навчальні втрати.

4.3.2 Результати zero-shot методів

У ході експериментів zero-shot моделі демонстрували різний рівень точності залежно від принципу роботи. CLIP проводив класифікацію на основі зіставлення текстових дескрипторів із візуальними особливостями об'єкта, тоді як Gemma 3 отримувала зображення предмета разом із текстовими промптами, що описували можливі варіанти захвату, і визначала правильний клас на основі аналізу цих даних.

Модель CLIP у zero-shot режимі показала 45% точності за Top-1 та 85% за Top-3, підтверджуючи свою здатність здійснювати якісну класифікацію на основі текстово-візуальної подібності без додаткового донавчання.

У випадку з великою мовною моделлю Gemma 3, яка працювала лише з текстовими описами без аналізу візуального контенту, точність класифікації за Top-1 склала від 61.25% у Gemma 3 4B до 71.25% у Gemma 3 27B, а за топ-3 від 90% до 95%. Точні результати можна побачити в таблиці 4.1.

Таблиця 4.1 – Результати експериментів zero-shot методів

Метод	Топ-1 точність	Топ-3 точність	Середній час інференсу (мс)
CLIP	45%	85%	16,62
Gemma 3 4B	61,25%	90%	4474,95
Gemma 3 12B	66,25%	95%	3779,95
Gemma 3 27B	71,25%	93,75%	6882,97

Отримані результати демонструють, що CLIP ефективно використовує механізм зіставлення текстових дескрипторів із візуальними ознаками.

Найкращий показник серед усіх протестованих моделей у Top-3 метриці досягнуто версією Gemma 3 12B (95%), що свідчить про її здатність правильно включати альтернативні варіанти захвату. Водночас Gemma 3 27B має найвищий показник у Top-1 (71,25%), що робить її найбільш точною у виборі єдиного правильного класу.

Це підтверджує, що мовні моделі можуть бути ефективними у zero-shot класифікації, якщо вони отримують якісно структуровані текстові промпти. При цьому CLIP демонструє більш стабільну генералізацію, покладаючись на візуальні ознаки, що робить його надійним варіантом для завдань з класифікацією у реальному часі.

Детальний аналіз сильних і слабких сторін цих підходів буде представлено в розділі 4.4, де проведено порівняння всіх протестованих методів.

4.3.3 Результати методів few-shot навчання

Розглянемо результати few-shot класифікації типів захвату, отримані за допомогою двох популярних методів мета-навчання — мереж відповідності та прототипічних мереж — у поєднанні з двома типами екстракторів ознак: CLIP та ResNet. Мета експерименту полягала в оцінці здатності моделей узагальнювати інформацію про нові класи за обмеженої кількості прикладів. У кожному експерименті використовувалося однакове few-shot налаштування, при якому на клас подавалося лише кілька зразків, а тестування проводилося на небачених даних з тестового набору.

Результати цих експериментів подано у таблиці 4.1, де показано як top-1, так і top-3 точність кожної моделі. Найвищу точність top-1 продемонстрував метод мереж відповідності із ResNet-екстрактором, досягнувши значення 47,50%. Цей результат вказує на здатність ResNet формувати достатньо чіткі розділення між класами в ознаковому просторі, що є корисним у умовах few-shot навчання. Водночас, метод прототипічних мереж у поєднанні з CLIP показала найкращий результат у метриці top-3 — 82,50%, що свідчить про те, що мультимодальний ознаковий

простір CLIP добре вловлює семантичну близькість між різними типами захватів. Результати цих експериментів наведено повторно в таблиці 4.2.

Таблиця 4.2 – Результати експериментів few-shot методів

Метод	Топ-1 точність	Топ-3 точність	Середній час інференсу (мс)
Мережі відповідності + CLIP	38,75%	80%	9,18
Мережі відповідності + ResNet	47,50%	77,50%	2,7
Прототипічні мережі + CLIP	43,75%	82,50%	13,61
Прототипічні мережі + ResNet	36,25%	70%	13,16

Для кращого розуміння процесу навчання моделей на рисунках 4.1 – 4.4 зображено графіки навчальних втрат для кожної з чотирьох конфігурацій:

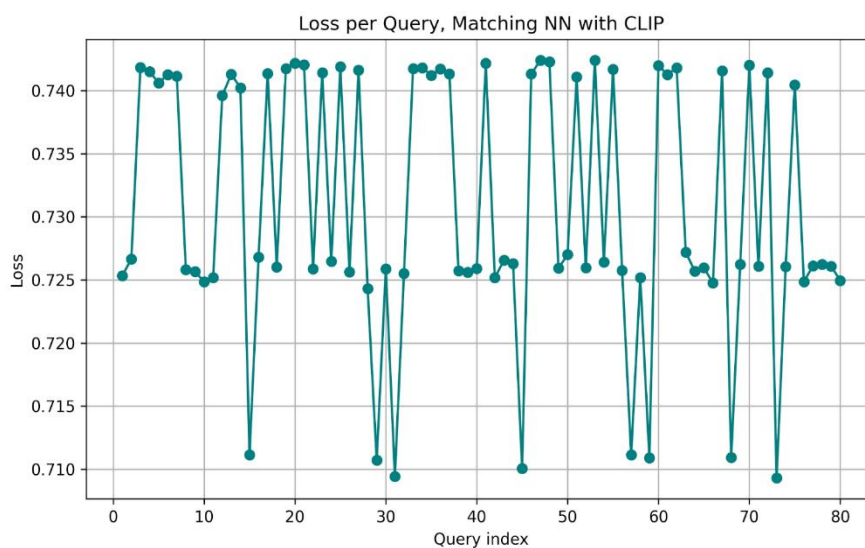


Рисунок 4.1 - Графік навчальних втрат у мережах відповідності з CLIP

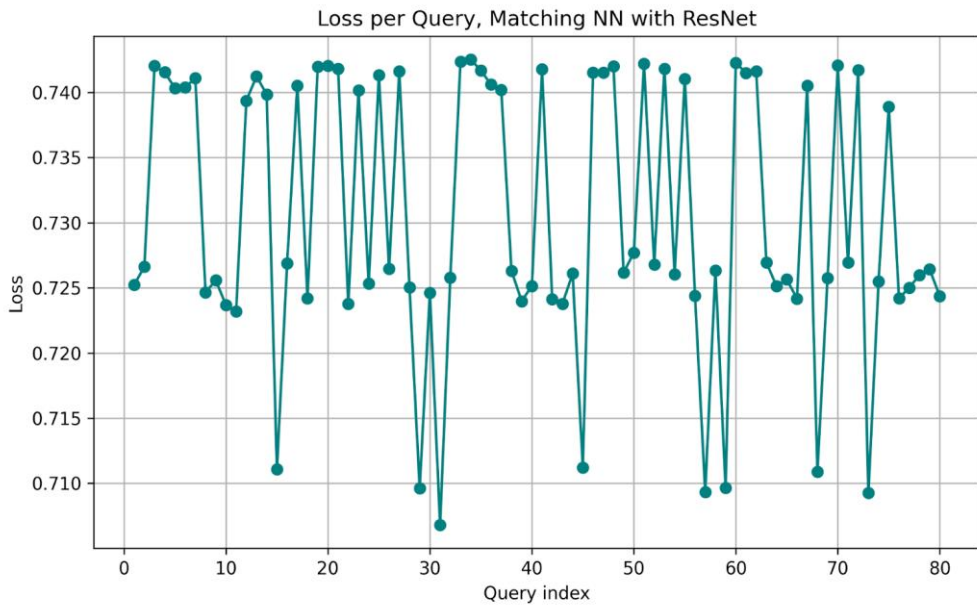


Рисунок 4.2 - Графік навчальних втрат у мережах відповідності з ResNet

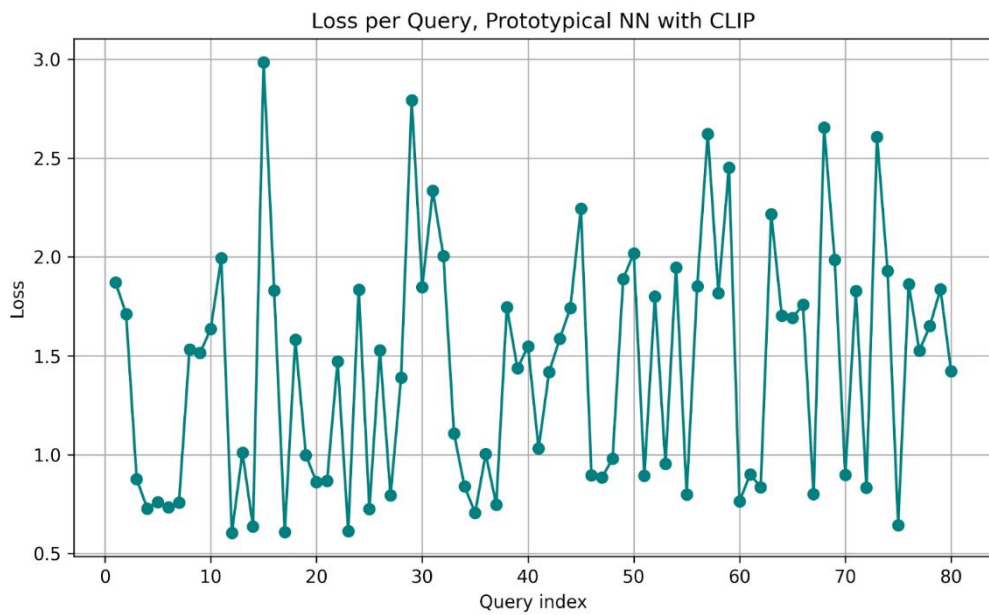


Рисунок 4.3 - Графік навчальних втрат у прототипічних мережах з CLIP

Зм.	Лист	№ докум.	Підпис	Дата

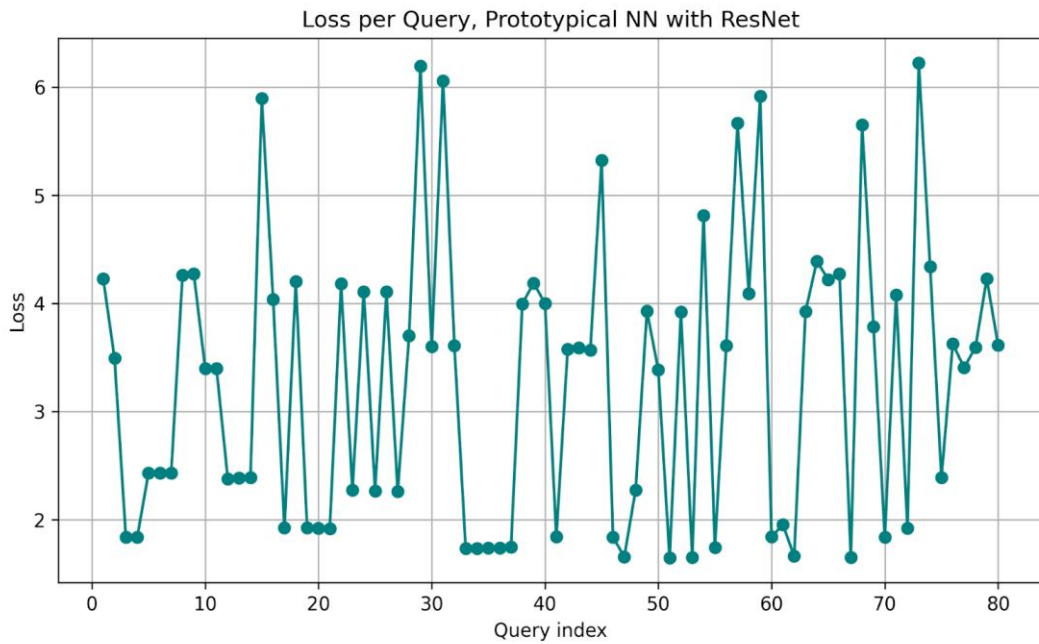


Рисунок 4.4 - Графіки навчальних втрат у прототипічних мережах з ResNet

Ці графіки дають змогу оцінити стабільність навчання та темп збіжності моделей, а також співвіднести перебіг тренування із фінальними результатами точності.

4.3.4 Результати KNN та CLIP

У цьому експерименті було досліджено ефективність класифікації типів захвату з використанням алгоритму К найближчих сусідів (KNN), що працює на основі ембеддингів, отриманих із моделі CLIP. Розглядалися два варіанти: KNN, який використовує ознаки з попередньо натренованого CLIP, та KNN, який працює з ембеддингами, отриманими з CLIP після fine-tuning на задачі класифікації захватів.

Комбінація KNN + CLIP (без fine-tuning) досягла top-1 точності 50,00% та top-3 точності 81,25%. Це свідчить про здатність нефайнтюненого CLIP генерувати доволі виразні представлення зображень, які добре розподіляються у просторі ознак і дозволяють KNN правильно ідентифікувати клас об'єкта в більшості випадків.

У випадку використання CLIP після fine-tuning, модель показала дещо нижчі результати: top-1 точність становила 45,00%, а top-3 точність — 80,00%. Це може

свідчити про часткову втрату генералізаційної здатності ембеддингів після додаткового навчання на малому датасеті. Імовірно, *fine-tuning* призвів до адаптації представлень під навчальні приклади, що знизило ефективність пошуку найближчих сусідів для нових прикладів. Результати експериментів з класичною та донавченою моделлю CLIP наведено в таблиці 4.5.

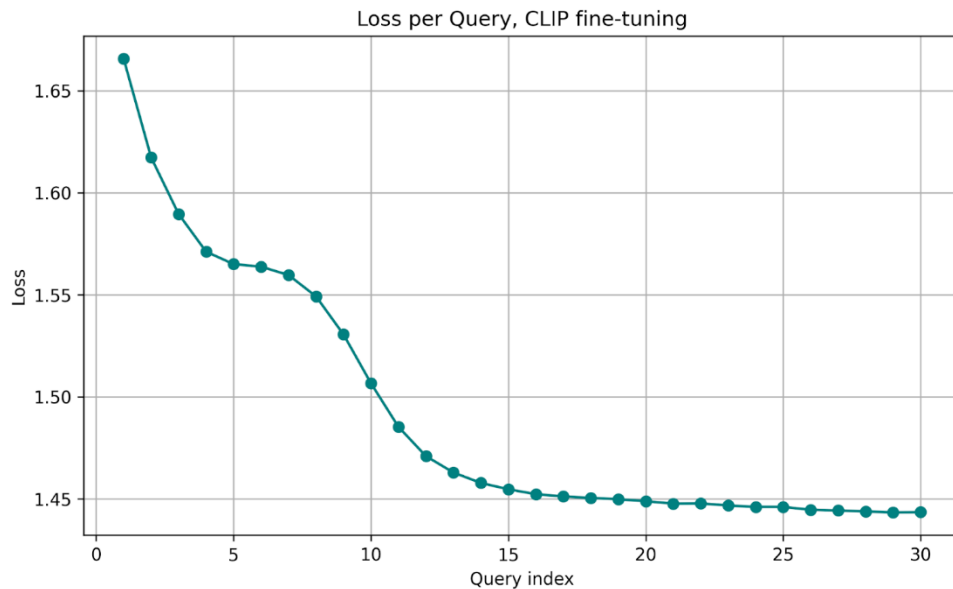


Рисунок 4.5 - Графік втрат моделі CLIP під час *fine-tuning*

Таблиця 4.3 – Результати експериментів з KNN та CLIP

Метод	Топ-1 точність	Топ-3 точність	Середній час інференсу (мс)
KNN + CLIP	50%	81,5%	17,27
KNN + <i>fine-tuned</i> CLIP	45%	80%	17,26

Загалом, результати демонструють, що навіть без додаткового навчання модель CLIP може бути ефективною у задачах класифікації типу захвату за допомогою KNN. Проте при *fine-tuning* потрібно враховувати ймовірність зниження здат-

ності до генералізації, особливо за наявності невеликого обсягу тренувальних даних.

4.4 Порівняння та аналіз результатів

Проведені експерименти дозволили оцінити ефективність різних методів класифікації типу захвату об'єкта за зображенням, кожен з яких має свої сильні та слабкі сторони. У цьому розділі розглядається детальне порівняння отриманих результатів, а також аналіз їхнього впливу на вибір оптимального підходу для роботи з обмеженою кількістю навчальних прикладів.

4.4.1 Загальний аналіз продуктивності моделей

Результати показали суттєву варіативність точності між zero-shot, few-shot та fine-tuned методами. Zero-shot моделі, зокрема CLIP та Gemma 3, продемонстрували, що навчання на великих наборах даних дозволяє здійснювати класифікацію без додаткового тренування на специфічному датасеті. CLIP показав стабільний результат із 85% точності за топ-3, що підтверджує його ефективність у визначенні класу захвату через текстово-візуальну подібність. Gemma 3 27B, найбільша з мовних моделей у тестуванні, перевершила CLIP за топ-1 метрикою, досягнувши 71.25% точності, що свідчить про високу здатність цієї моделі до узагальнення, в той час як Gemma 3 12B досягла топ-3 точності 95%.

У порівнянні з zero-shot підходами, few-shot методи, зокрема мережі відповідності та прототипічні мережі показали конкурентні результати, особливо при використанні мультимодальних моделей на етапі витягування ознак. Прототипічні мережі у поєднанні з CLIP досягли Top-3 точності 82,5%, що вказує на здатність моделі адаптуватися до малих навчальних наборів. Використання ResNet як альтернативного екстрактора ознак також демонструє непогані результати, проте його продуктивність трохи поступається мультимодальним методам.

Окремо варто розглянути донавчену модель CLIP у поєднанні з KNN, який, попри очікуваний приріст продуктивності після додаткового навчання, показав

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		55

45% точності за топ-1 та 80% за Топ-3, що трохи поступається оригінальному варіанту. Це підтверджує, що надмірне донавчання може знижувати генералізаційну здатність ембеддинг-простору.

Діаграму з порівнянням точностей досліджених методів визначення захвату зображено на рисунку 4.6.

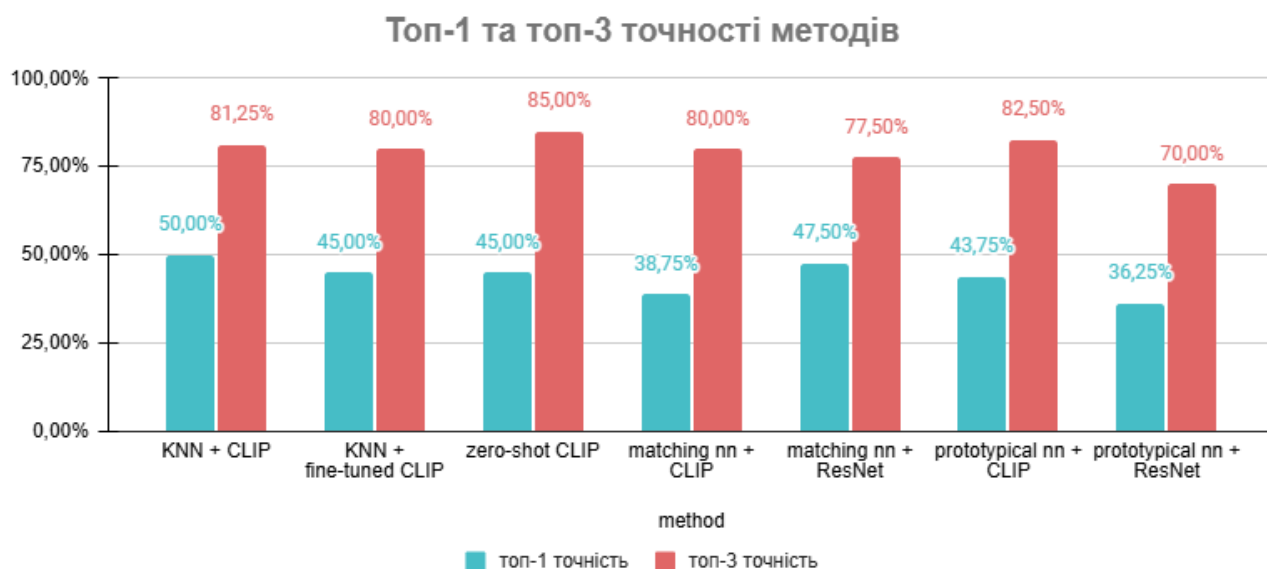


Рисунок 4.6 – Порівняння точностей обраних методів

Діаграму з порівнянням точностей визначення захвату різних версій моделі Gemma зображено на рисунку 4.7.

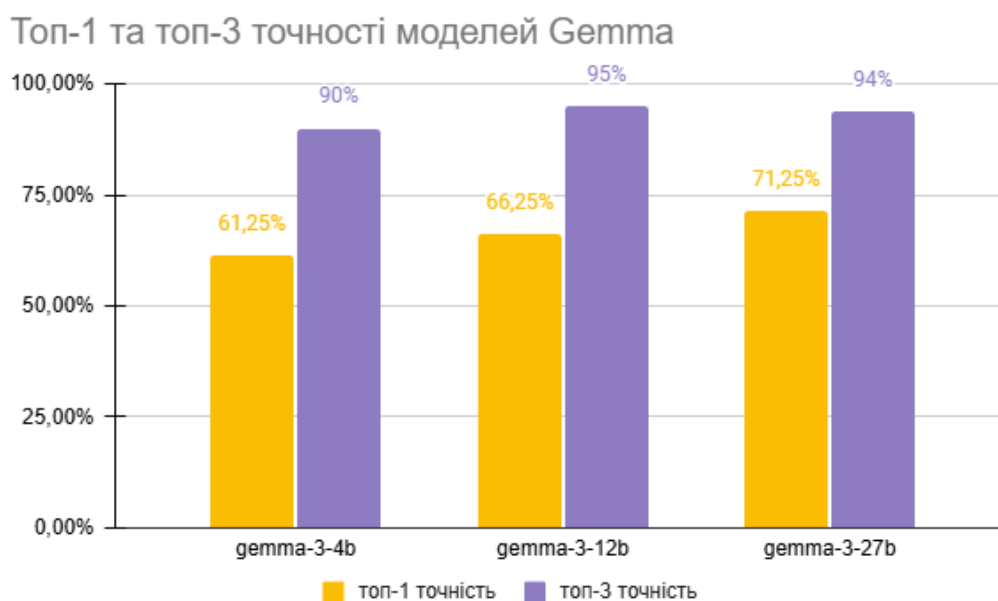


Рисунок 4.7 – Порівняння точності версій Gemma

Зм.	Лист	№ докум.	Підпис	Дата

Окрім точності, не менш важливим показником для застосування в реальному часі є середній час інференсу. Найшвидшим виявився метод мережі відповідності з ResNet, що продемонстрував середній час інференсу 2.7 мс. Він підходить для систем із жорсткими обмеженнями по затримці. Натомість, zero-shot CLIP працює повільніше — 21.79 мс, оскільки виконує обчислення подібності між ембеддингом зображення та всіма текстовими запитами.

Компромісним варіантом виявився метод KNN + CLIP, який забезпечує непогану точність при середньому часі інференсу 17.27 мс. Таким чином, у разі пріоритету реального часу доцільно обирати моделі з меншою затримкою, навіть якщо це означає незначне зниження точності.

З точки зору адаптивності — тобто здатності правильно класифікувати нові, раніше не бачені об'єкти до вже існуючих класів — найкращі результати показують мультимодальні моделі, зокрема CLIP у zero-shot та few-shot режимах. Завдяки навчанню на великій кількості текстово-візуальних пар, CLIP демонструє високу узагальнювальну здатність, що дозволяє йому коректно зіставляти нові вхідні зображення з відомими класами на основі семантичної подібності. Це особливо помітно у zero-shot режимі, де модель може класифікувати незнайомі предмети, не маючи їх у тренувальному наборі.

Few-shot методи, зокрема прототипічні мережі з CLIP або ResNet-ембеддингами, також демонструють певну адаптивність, але залежать від того, наскільки добре нові об'єкти узгоджуються з тими прикладами, що були в опорному наборі. Мультимодальні ознаки (CLIP) забезпечують кращу адаптивність, ніж виключно візуальні (ResNet), оскільки дають змогу краще зрозуміти суть класу, а не лише схожість на рівні пікселів.

Донавчені моделі мають найнижчу адаптивність: вони схильні запам'ятовувати конкретні приклади і часто помиляються при зустрічі з новими варіаціями відомих класів, особливо якщо ті відрізняються за ракурсом, освітленням або кольором.

Щодо масштабованості, найбільш гнучкими є zero-shot моделі, зокрема CLIP

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		57

Група методів	Переваги	Недоліки
Gemma 3	Gemma 3 демонструє найвищу точність як у топ-1 так і у топ-3 порівнянні, що означає, що вона може давати найбільш точний прогноз при унікальних або складних об'єктах, якщо правильно сформований текстовий запит.	Має дуже великий розмір моделі: її 12B та 27B версії потребують значних обчислювальних ресурсів, що може ускладнити її інтеграцію безпосередньо у процесор протеза. Довгий час інференсу при роботі через API.
Few-shot методи	Мережі відповідності показали конкурентні результати, особливо у топ-3 оцінці, що дозволяє моделі легко адаптуватися до нових об'єктів у ситуаціях, де протез працює зі змінними умовами. Швидка обробка запитів. Прототипічні мережі використовують середні вектори для класифікації, що зменшує обчислювальну складність і може бути ефективним для роботи у малопотужних пристроях.	Менша точність у топ-1, що означає, що модель може плутати об'єкти, коли необхідна висока точність. У біонічному протезі це може бути критично для точних маніпуляцій (наприклад, захоплення маленьких предметів). Залежність від кількості навчальних прикладів: якщо набір даних містить недостатньо репрезентативних зображень, ефективність моделі значно знижується.
Fine-tuned CLIP та KNN	Підходить для спеціалізованих датасетів, якщо протез використовується у завданнях зі специфічними предметами, наприклад,	Fine-tuning може зменшити генералізаційну здатність моделі, що може бути проблемою, якщо протез працює у середовищах із непередбачуваними об'єктами.

Зм.	Лист	№ докум.	Підпис	Дата

Група методів	Переваги	Недоліки
	у медичних або промислових роботизованих пристроях. Висока продуктивність у топ-3 (80%), що робить його надійним у ситуаціях, де важливо забезпечити гнучкість вибору захвату.	Класифікація через KNN потребує збереження достатнього набору еталонних ембеддингів, що може вимагати більшої пам'яті у пристрої.
CLIP та KNN	Може працювати у реальному часі, оскільки не потребує навчання. Хороша генералізація, що дозволяє протезу адаптуватися до нових об'єктів. Висока точність у топ-3, що забезпечує гнучкість у виборі захвату.	Залежність від якості опорного набору, тобто якщо набір еталонних зображень недостатньо репрезентативний, точність може знижуватися. Не завжди чітко розділяє класи, що іноді призводить до плутанини між подібними захватами.

4.4.3 Вибір оптимального методу для біонічного протезу

На основі отриманих результатів можна зробити висновки щодо вибору оптимального методу класифікації типу захвату для біонічного протезу залежно від його призначення та умов використання.

Якщо пристрій має працювати в режимі реального часу, тобто виконувати миттєву класифікацію без попереднього навчання, найбільш ефективним є zero-shot підхід із використанням моделі CLIP. Ця модель показала високу точність у топ-3 сценаріях, що дозволяє зберегти гнучкість у виборі захвату у непередбачуваних ситуаціях.

Якщо ж у роботі протезу критично важлива точність вибору одного правильного класу, наприклад, у випадку складних медичних маніпуляцій або точної

взаємодії з дрібними предметами, найбільш продуктивною є мовна модель Gemma 3. Її найбільш потужні версії, Gemma 3 27B та Gemma 3 12B, демонструють найкращі показники у топ-1 та у топ-3 оцінках, що робить їх придатними для ситуацій, де висока точність має перевагу над швидкістю. Водночас, варто враховувати, що такі моделі є надто великими для прямої інтеграції у процесор біонічного протеза, що може обмежити їх практичне використання.

У випадках, коли біонічний протез має працювати у змінних умовах та адаптуватися до нових об'єктів, перспективним варіантом є few-shot метод мереж відповідності із CLIP. Ця модель добре працює при обмеженій кількості навчальних прикладів і показала ефективність у сценаріях, де доступний лише невеликий набір опорних зображень. Такий підхід дозволяє протезу швидко адаптуватися до нових умов без потреби у великих наборах даних для попереднього навчання.

Якщо ж система використовується в специфічному середовищі з чітко визначеним набором об'єктів, наприклад, у роботизованих виробничих процесах або медичних пристроях, ефективним варіантом буде fine-tuned CLIP у поєднанні з KNN. Доновчання дозволяє оптимізувати представлення зображень у векторному просторі, підлаштовуючи модель під специфічні завдання, проте варто враховувати можливе зниження генералізаційної здатності, якщо набір навчальних даних надто обмежений.

Таким чином, найбільш придатними методами для біонічного протезу в умовах обмежених даних є zero-shot CLIP, що забезпечує досить швидку реакцію та високу точність; мовна модель Gemma 3 для задач, що вимагають високої точності у визначенні одного правильного класу, але при наявності великих обчислювальних даних; а також few-shot мережі відповідності для сценаріїв, де потрібна адаптація до нових предметів та швидка обробка. Fine-tuned CLIP + KNN може бути ефективним у вузькоспеціалізованих застосуваннях, проте його використання потребує ретельного контролю над обсягом навчальних даних та їх змістом.

Такий вибір методів забезпечує необхідний баланс між точністю, адаптивністю до нових об'єктів, та масштабуванням з можливістю додавання нових типів захвату, що є ключовими характеристиками для інтеграції у сучасні біонічні

протези [23].

Висновки до розділу 4

Проведене дослідження дозволило оцінити ефективність різних методів класифікації типу захвату для біонічного протезу, зокрема zero-shot, few-shot та fine-tuned підходів. Результати експериментів показали, що кожен метод має свої сильні та слабкі сторони, залежно від поставлених вимог до точності, швидкості та адаптивності системи.

Модель CLIP у zero-shot режимі підтвердила свою здатність швидко класифікувати об'єкти без додаткового навчання, досягнувши 85% точності у топ-3, що робить її перспективним варіантом для реального часу. Водночас, мовна модель Gemma 3 продемонструвала найвищу точність у топ-1, що підтверджує її ефективність для задач, де важливий вибір одного правильного класу, особливо у медичних маніпуляціях та точному захопленні предметів.

Few-shot методи, зокрема мережі відповідності та прототипічні мережі, виявилися адаптивними у змінних умовах, що є корисним для сценаріїв, коли протез стикається з новими об'єктами [24]. Найкращі результати показали моделі із CLIP як екстрактором ознак, що дозволяє ефективно використовувати візуальну інформацію для аналізу захвату.

Fine-tuned CLIP у поєднанні з KNN забезпечив стабільні результати, проте спостерігалось зниження генералізаційної здатності моделі після навчання на спеціалізованому датасеті. Це підтверджує, що fine-tuning слід застосовувати обережно, особливо у випадках, коли система повинна працювати з неочікуваними предметами.

Отримані результати демонструють, що поєднання комп'ютерного зору та штучного інтелекту може суттєво покращити ефективність біонічних протезів, роблячи їх інтуїтивними у використанні. Використання API OpenRouter для інтеграції мовних моделей дозволило оптимізувати обробку даних, забезпечуючи гнучку та масштабовану архітектуру системи класифікації.

Таким чином, найбільш придатними методами для біонічного протезу в умовах обмежених навчальних даних є:

- zero-shot CLIP для сценаріїв, де важлива висока адаптивність до нових, раніше не бачених об'єктів, навіть без додаткових прикладів;
- модель Gemma 3 для задач, де критично важлива точність вибору одного класу, та є можливість довше чекати на виконання захвату;
- методи few-shot навчання для випадків, де необхідна гнучка адаптивність до нових об'єктів у змінному середовищі, та швидка класифікація в реальному часі;
- донавчана модель CLIP в поєднанні з KNN для вузькоспеціалізованих сценаріїв, де допустиме нижче узагальнення, але потрібна висока точність на знайомих об'єктах.

Проведене дослідження підтвердило, що нейромережеві алгоритми можуть бути успішно застосовані для підвищення точності визначення захвату та адаптивності біонічних протезів, відкриваючи нові перспективи для розробки асистивних технологій та робототехніки.

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		63

ВИСНОВКИ

Протягом виконання дипломного проєкту було розроблено підсистему комп'ютерного зору для визначення типу захвату біонічного протезу. Ціллю проєкту було розробити підсистему, що зможе ефективно визначати тип захвату без участі користувача, тим самим полегшуючи адаптацію до протезу та його використання.

У процесі виконання дипломного проєкту було розроблено підсистему комп'ютерного зору, призначену для автоматичного визначення типу захвату, який повинен бути виконаний біонічним протезом. Основною метою цього проєкту було створення інтелектуального компонента, здатного автономно аналізувати зображення об'єкта, що захоплюється, і на цій основі визначати відповідний тип захвату без потреби в прямому керуванні з боку користувача. Такий підхід покликаний підвищити зручність використання біонічного протезу, зменшити когнітивне навантаження на людину і сприяти швидшій адаптації до пристрою.

На початковому етапі було проведено аналіз предметної області, в межах якого визначено актуальність проблеми інтелектуального управління біонічними протезами. Під час аналізу літератури та огляду існуючих технічних рішень було виявлено, що хоча деякі сучасні протези здатні підтримувати кілька типів захватів, вибір типу переважно залишається на користувачеві. Це потребує або активного перемикання режимів, або передбачає використання допоміжних сенсорів, які не завжди дають стабільні результати в реальних умовах експлуатації. Таким чином, було підтверджено потребу в розробці нової підсистеми, яка б дозволяла визначати тип захвату безпосередньо на основі візуальної інформації, що отримується з камери, встановленої на або поблизу протезу.

Після вивчення наукових і прикладних підходів до розпізнавання об'єктів та дій у зображеннях, було визначено архітектурні вимоги до підсистеми. Вона мала об'єднувати простоту реалізації, високу узагальнюваність на нові об'єкти, стабільність у прогнозах і швидкодію. У процесі розробки було вирішено використовувати

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		64

можливості мультимодальних моделей комп'ютерного зору, зокрема CLIP, як базовий компонент для отримання векторних ознак зображень та текстових описів. Цей вибір обумовлено тим, що CLIP дозволяє поєднувати візуальну та текстову інформацію, що є особливо корисним для задач класифікації з обмеженою кількістю прикладів.

Під час формулювання вимог до системи було визначено її основні функціональні компоненти: модуль обробки зображень, модуль порівняння ознак і модуль класифікації. Система повинна бути здатною адаптуватися до нових об'єктів та класів захватів при наявності лише кількох візуальних або текстових прикладів.

У процесі реалізації було зібрано невеликий опорний датасет зображень об'єктів, що потребують різні типи захвату, та проведено низку експериментів з поєднанням візуальних ознак, отриманих зображень, з текстовими описами, які відповідають кожному класу. Основною задачею було не лише точне визначення класу, а також досягнення достатньої швидкодії для потенційної роботи в реальному часі.

У рамках експериментальної частини дипломного проєкту було проведено низку досліджень для оцінки працездатності підсистеми, перевірки її адаптивності та порівняння різних способів організації класифікації типів захвату. Було протестовано декілька методів, зокрема zero-shot навчання, де класифікація відбувається на основі текстових описів, few-shot навчання, де для класифікації з використовується опорний набір даних, а також комбіновані підходи, що поєднували семантичну подібність з векторними представленнями зображень та тексту. Експерименти дозволили не лише отримати формальні метрики якості, а й виявити практичні переваги та недоліки кожного з підходів.

У ході досліджень з'ясувалося, що різні методи можуть бути корисними в різних контекстах використання біонічного протезу. Наприклад, zero-shot класифікація на основі текстових підказок демонструє хорошу універсальність та здатність до генералізації, особливо у випадках, коли користувач не має можливості завчасно надати приклади усіх можливих типів об'єктів або ситуацій. Такий підхід може бути доцільним у побутовому середовищі, де об'єкти, з якими працює

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		65

користувач, можуть змінюватися щодня, а сам протез має підлаштовуватися до нових умов у напівавтоматичному режимі.

З іншого боку, few-shot підходи, класичні алгоритми машинного навчання, такі як KNN та донавчання моделі CLIP показали високу ефективність у середовищах з обмеженою кількістю класів і чітко визначеним набором об'єктів. У таких умовах, як виробничі лінії або ситуації з фіксованим набором інструментів, система може бути попередньо налаштована на відповідні захвати з мінімальною кількістю прикладів і демонструвати стабільні результати навіть при суттєвому зменшенні обсягу навчальних даних.

У результаті проведеної роботи було створено актуальну та функціональну підсистему комп'ютерного зору для визначення типу захвату біонічного протезу. Розроблена система здатна здійснювати розпізнавання об'єкта та пропонувати найбільш доречний варіант захвату, що дозволяє автоматизувати один із ключових етапів взаємодії користувача з протезом. Завдяки використанню сучасних методів комп'ютерного зору та можливості адаптації до нових умов, підсистема може бути застосована як у повсякденному житті, так і в умовах виробництва. Вона має практичну цінність для виробників біонічних протезів завдяки своїй гнучкості, доступності й потенціалу для подальшої інтеграції в апаратно-програмні комплекси керування протезами. Отримані результати підтвердили перспективність обраного напрямку дослідження та створили основу для подальшого розвитку інтелектуальних систем підтримки керування біонічними пристроями.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Antfolk, C., D'Alonzo, M., Rosén, B., Lundborg, G., Sebelius, F., & Cipriani, C. (2013). Sensory feedback in upper limb prosthetics. Expert Review of Medical Devices.
2. Hatsan S., Oliinyk V. Computer vision based authentication model with spoofing protection // The International Conference on Security, Fault Tolerance, Intelligence ICSFTI2024 (June 07, 2024, Kyiv, Ukraine), 2024. P. 1-12. URL: <https://icsfti-proc.kpi.ua/article/view/308401>
3. Oliinyk, V. An efficient face mask detection model for real-time applications / V. Oliinyk, A. Ryzhiy // Адаптивні системи автоматичного управління: міжвідомчий науково-технічний збірник. – 2022. – № 1 (40). – С. 54-64.
4. Oliinyk, V. An efficient real-time gaze tracking method for browser-based applications / Oliinyk V., Korol S. // Адаптивні системи автоматичного управління: міжвідомчий науково-технічний збірник. – 2025. – № 2 (47).
5. Oliinyk V. Autonomous car parking model for different types of parking lots using deep reinforcement learning / Oliinyk V., Danyliuk Y. // Адаптивні системи автоматичного управління: міжвідомчий науково-технічний збірник. – 2025. – № 1 (46). – С. 237-246.
6. SoftHand Pro [Електронний ресурс] – URL: <https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-023-01130-x>
7. Adaptive Grasping System [Електронний ресурс] – URL: <https://www.mdpi.com/2072-666X/13/2/219>
8. SoftHand 2 Pro [Електронний ресурс] – URL: https://www.academia.edu/113917836/The_SoftHand_Pro_Translation_from_Robotic_Hand_to_Prosthetic_Prototype
9. Össur [Електронний ресурс] – URL: <https://www.ossur.com/en-gb/prosthetics/arms/i-limb-ultra>
10. ResNet (глибока згортова нейронна мережа) [Електронний ресурс] – URL: <https://viso.ai/deep-learning/resnet-residual-neural-network/>

					ІК12.030БАК.006 ПЗ	Арк.
Зм.	Лист	№ докум.	Підпис	Дата		67

11. CLIP [Електронний ресурс] – URL: <https://openai.com/index/clip/>
12. Matching Networks and Few-shot Learning [Електронний ресурс] – URL: <https://arxiv.org/abs/1905.06331>
13. Prototypical Networks and Few-shot Learning [Електронний ресурс] – URL: <https://arxiv.org/abs/1703.05175>
14. Gemma 3 [Електронний ресурс] – URL: <https://ai.google.dev/gemma>
15. Cosine Similarity [Електронний ресурс] – URL: <https://www.geeksforgeeks.org/cosine-similarity/>
16. About Python [Електронний ресурс] – URL: <https://www.python.org/about/>
17. PyTorch Documentation [Електронний ресурс] – URL: <https://pytorch.org/docs/stable/index.html>
18. Hugging Face Transformers: Accessing Pretrained Models [Електронний ресурс] – URL: <https://huggingface.co/docs/transformers/index>
19. Few-shot Learning: Challenges and Future Directions [Електронний ресурс] – URL: <https://arxiv.org/abs/2007.05135>
20. OpenRouter [Електронний ресурс] – URL: <https://openrouter.ai/>
21. Gradio [Електронний ресурс] – URL: <https://www.gradio.app/docs/python-client/introduction>
22. A. Radford, J. W. Kim, C. Hallacy та ін. “Learning Transferable Visual Models From Natural Language Supervision” [Електронний ресурс] – URL: <https://arxiv.org/pdf/2103.00020v1>
23. Deep Learning for Robotics [Електронний ресурс] – URL: <https://arxiv.org/abs/1803.00134>
24. Meta-Learning Survey: Challenges and Advances [Електронний ресурс] – URL: <https://arxiv.org/abs/2004.11149>