




ДОДАТОК А

Звіт результатів перевірки кваліфікаційної роботи на унікальність тексту

Дата звіту 6/9/2025
Дата редагування ---



Звіт не був оцінений

Звіт подібності

метадані

Назва організації
Kharkiv National University of Radio Electronics

Заголовок
2025_М_ПІ_ІПЗм-23-3_Гайдук_Д_А_скорочений

Автор Науковий керівник / Експерт
Гайдук Данііл Андрійович Олена Олійник

підрозділ
каф. ПІ

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.

0.99%
0.99%

КП 1

0.36%
0.36%

КЦ

25

Довжина фрази для коефіцієнта подібності 2

17633

Кількість слів

135131

Кількість символів

ДОДАТОК Б

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008:2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІІЗМ-23-3
(група)

Гайдук Данііл Андрійович

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт

(підпис)

10.06.2025

Олена ОЛІЙНИК

(прізвище, ініціали)

ДОДАТОК В

Апробація результатів роботи

 | **ISU**
INTERNATIONAL SCIENTIFIC UNITY

CERTIFICATE

of conference participant

it is hereby certified, that

ДАНІІЛ АНДРІЙОВИЧ ГАЙДУК

took part in the 1st International Scientific and Practical Conference
«MODERN PERSPECTIVES ON SCIENCE AND ECONOMIC PROGRESS»

June 4-6, 2025, Vilnius, Lithuania
 24 Hours of Participation
 (0,8 ECTS credits)








Head of the organizing committee
 
 Viktoriia Tsiundyk



isu-conference.com

ISU-25/0604-096



ДОСЛІДЖЕННЯ АЛГОРИТМІВ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ РОЗПІЗНАВАННЯ ЖЕСТІВ

Гайдук Данііл Андрійович

здобувач вищої освіти магістерського рівня

Голян Віра Володимирівна

доцент, кандидат технічних наук

Кафедра ПІ

Харківський національний університет радіоелектроніки, Україна

У сучасному технологічному суспільстві проблема створення природних інтерфейсів взаємодії між людиною та комп'ютером є особливо актуальною. Одним з найперспективніших напрямів у цій галузі є розпізнавання жестів – процес інтерпретації рухів тіла, зокрема рук, як інструментів комунікації з цифровими пристроями [1]. Системи такого типу відкривають нові можливості для безконтактного керування, зокрема у віртуальній і доповненій реальності, в медицині, інклюзивних технологіях, робототехніці та інших сферах. В умовах швидкого розвитку технологій комп'ютерного зору та глибокого навчання зростає потреба у створенні рішень, здатних розпізнавати як статичні, так і динамічні жести з високою точністю та стабільністю у змінних умовах зовнішнього середовища [2].

Метою цієї роботи стало всебічне дослідження алгоритмів комп'ютерного зору, які можна ефективно використати для розпізнавання людських жестів. Основний акцент було зроблено на порівняльному аналізі точності, швидкодії, стійкості до шумів та адаптивності моделей до нових користувачів і змін навколишнього середовища. Крім того, поставлено завдання розробити працездатну програмну систему, яка демонструє практичну реалізацію теоретичних рішень і функціонує в реальному часі без потреби у високопродуктивному обладнанні.

На початковому етапі дослідження було здійснено глибокий огляд літературних джерел, що охоплюють еволюцію методів комп'ютерного зору – від класичних технік фільтрації зображень та сегментації кольору до сучасних архітектур нейронних мереж. Було детально проаналізовано фундаментальні підходи до виділення ключових ознак, екстракції просторово-часової інформації, моделювання послідовностей та побудови моделей класифікації. У центрі уваги опинилися методи, які широко застосовуються в сучасних системах розпізнавання: згорткові нейронні мережі (CNN), рекурентні моделі з довготривалою пам'яттю (LSTM), трансформерні архітектури (ViT), а також класичні методи, такі як опорні вектори (SVM), алгоритми найближчих сусідів (KNN) і гістограми орієнтованих градієнтів (HOG).

Особливу увагу було приділено питанням попередньої обробки даних, адже саме якісна підготовка зображень суттєво впливає на ефективність подальшого розпізнавання. Було реалізовано етапи фільтрації шумів, нормалізації, масштабування зображень, корекції освітлення. Для точного виділення жестових

Modern Perspectives on Science and Economic Progress

областей використано бібліотеку MediaPipe, яка дозволяє локалізувати до 21 ключової точки на людській руці в реальному часі [3]. Завдяки цьому вдалося мінімізувати залежність розпізнавання від фону та покращити стійкість до шумів.

На етапі побудови моделей здійснено порівняльний аналіз різних підходів до класифікації жестів. Класичні методи, зокрема SVM і KNN, хоча й мають переваги у простоті реалізації та низьких апаратних вимогах, продемонстрували обмежену ефективність при роботі з динамічними жєстами, особливо у складних умовах фону або змінного освітлення. Метод HOG, який дозволяє витягувати інформацію про форму об'єкта, показав хорошу продуктивність у задачах зі статичними жєстами, проте виявився нестабільним при зміні перспективи або наявності тіней.

Суттєве зростання точності спостерігалось при використанні глибоких згорткових мереж (CNN), які здатні автоматично виявляти релевантні просторові ознаки. Однак для обробки послідовностей кадрів, у яких розпізнавання залежить не лише від форми, але й від змін у часі (наприклад, при маханні рукою), найбільш доцільним є використання рекурентних мереж, зокрема LSTM. Саме LSTM дозволяють зберігати інформацію про попередні кадри та враховувати часову динаміку, що є критичним для динамічних жестів. У дослідженні також перевірялись моделі Vision Transformer (ViT), які, хоч і продемонстрували високу точність при роботі зі статичними кадрами, виявилися менш оптимальними у задачах реального часу через високе навантаження на ресурси.

Найвищі результати було досягнуто за допомогою гібридної моделі, яка поєднує CNN для виділення просторових ознак з LSTM для аналізу їх послідовності. Такий підхід дозволив ефективно розпізнавати як фіксовані пози рук, так і складні жєсти, що реалізуються через серію рухів. Модель була навчена на власноруч зібраному наборі відеоданих, що включав жєсти 12 користувачів, виконані в умовах змінного освітлення, на різних фонах та з різними швидкостями виконання. Результати тестування засвідчили високу точність (асигасу понад 92%) та стабільність при практичному використанні. Модель також показала відмінні результати за метриками precision, recall і F1-score, що підтверджує її надійність і збалансованість.

У межах практичного етапу дослідження було реалізовано повноцінну програмну систему, яка функціонує в реальному часі. Архітектура системи була побудована за модульним принципом з розподілом на блоки: захоплення відео, попередня обробка, екстракція ознак, класифікація жестів, виведення результатів. Розробка здійснювалася на мові Python із залученням OpenCV, MediaPipe, PyTorch Lightning, NumPy, matplotlib. Система була протестована в режимі реального часу та не вимагала використання графічного процесора (GPU), що підтверджує її практичну придатність для широкого використання.

Значущим результатом є здатність запропонованої системи адаптуватися до різних користувачів без потреби в перенавчанні – завдяки використанню універсальних ознак, виділених на основі координат ключових точок.

Modern Perspectives on Science and Economic Progress

Враховуючи це, система може бути ефективно інтегрована в навчальні платформи, ігрові додатки, системи автоматизації, а також у медичні чи реабілітаційні середовища.

Узагальнюючи, результати дослідження підтвердили гіпотезу про перевагу гібридних нейромережових моделей для задач розпізнавання жестів у реальних умовах. Висока точність, стійкість до зовнішніх змін, здатність до обробки в реальному часі без складного апаратного забезпечення – усі ці характеристики забезпечують перспективність розробленої системи для практичного застосування. Наукова новизна полягає у поєднанні традиційної обробки зображень, бібліотеки MediaPipe для трекінгу та сучасних архітектур глибокого навчання у єдиному рішенні, яке продемонструвало ефективність на всіх етапах – від теорії до практики.

Одержані результати можуть слугувати основою для подальших досліджень у галузі мультимодальної взаємодії, адаптивного керування, розпізнавання мови жестів, а також для впровадження в інтерактивні системи з елементами штучного інтелекту, де природна взаємодія є пріоритетною.

Список використаних джерел

1. Mitchell, H. B. (2013). *Image Fusion: Theories, Techniques and Applications*. Springer.
2. Cali, M., & Conci, N. (2022). *Hand gesture recognition: A deep learning approach*. Springer.
3. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Ce Liu, Wu, Y., & The MediaPipe Team. (2019). *MediaPipe: A framework for building perception pipelines*. arXiv preprint arXiv:1906.08172. <https://arxiv.org/abs/1906.08172>

ДОСЛІДЖЕННЯ МЕТОДІВ КЛАСИФІКАЦІЇ АНОМАЛІЙ НА СУПУТНИКОВИХ ЗНІМКАХ ЗА ДОПОМОГОЮ КОМП'ЮТЕРНОГО ЗОРУ

Василець Олександр Олексійович

здобувач вищої освіти магістерського рівня

Голян Віра Володимирівна

доцент, кандидат технічних наук

Кафедра ПІ

Харківський національний університет радіоелектроніки, Україна

Автоматичне виявлення та класифікація аномалій на супутникових знімках є важливою складовою екологічного моніторингу, аграрного аналізу, планування територій, реагування на надзвичайні ситуації та дослідження змін навколишнього середовища [1]. В умовах зростання обсягів даних дистанційного зондування Землі, постає необхідність розробки ефективних, швидких і

ДОДАТОК Г
Слайди презентації

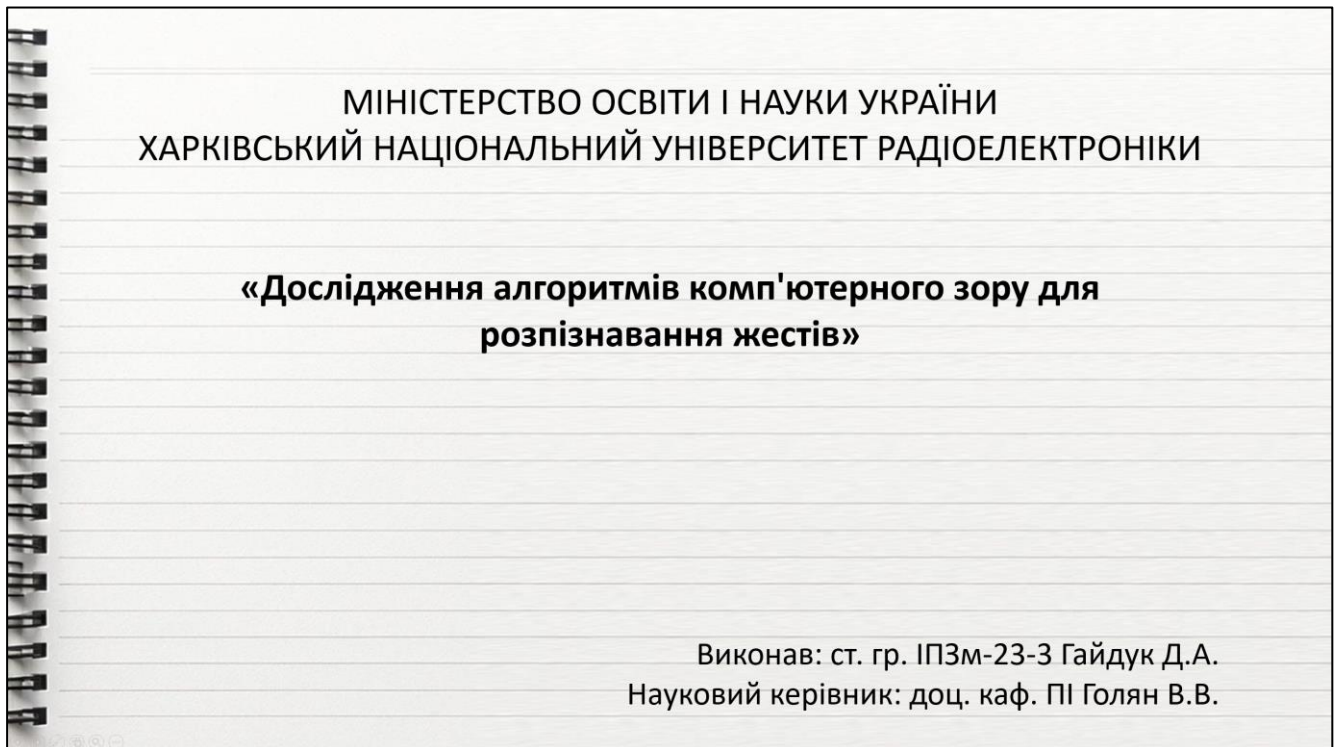


Рисунок Д.1 – Слайд 1

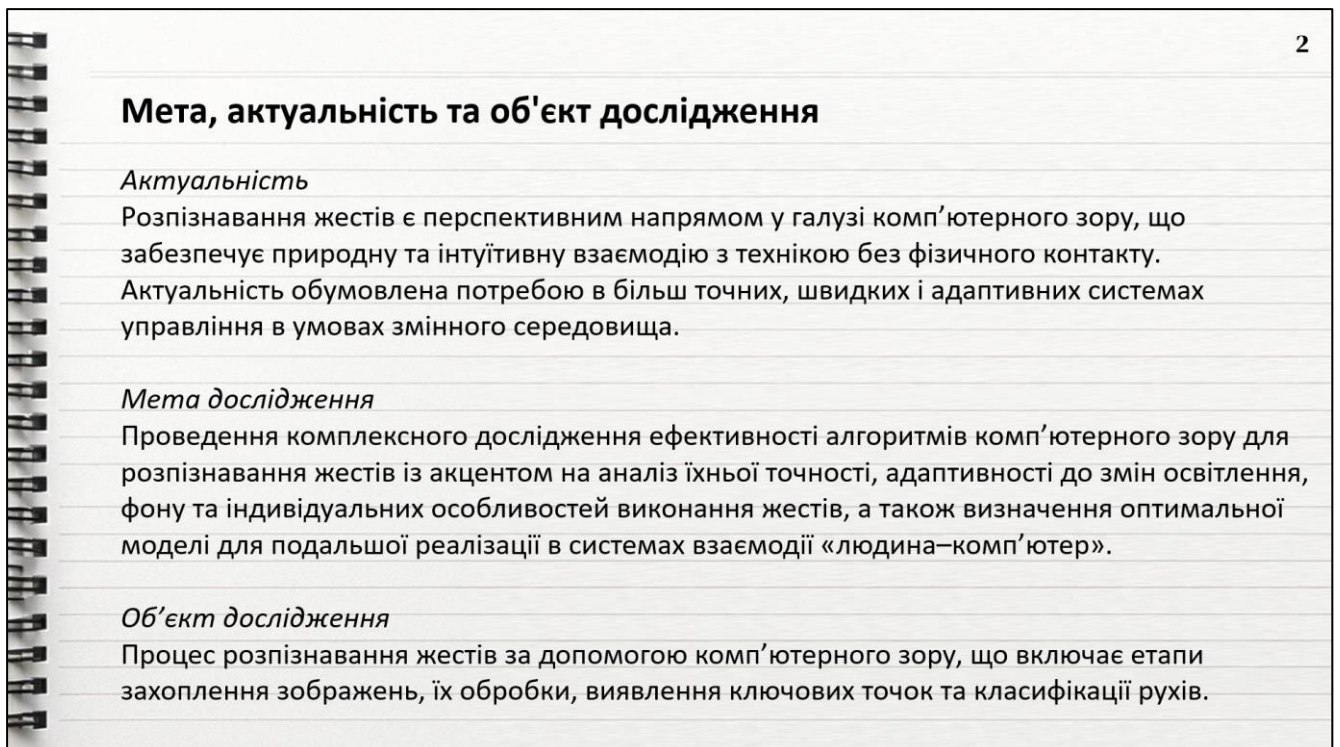


Рисунок Д.2 – Слайд 2

3

Постановка задачі

У зв'язку з актуальністю безконтактної взаємодії людини з комп'ютером, виникає потреба в розробці ефективних алгоритмів комп'ютерного зору для розпізнавання жестів. Сучасні підходи стикаються з проблемами адаптації до змін освітлення, фону, індивідуальних особливостей жестів і потребують високої точності та швидкості.

Для вирішення цієї проблеми поставлено такі завдання:

- провести огляд існуючих алгоритмів розпізнавання жестів;
- порівняти їх за точністю, швидкістю та адаптивністю;
- удосконалити алгоритми для складних умов (освітлення, фон, жести);
- реалізувати програмну систему для тестування;
- оцінити ефективність у практичних сценаріях.

Рисунок Д.3 – Слайд 3

4

Протиріччя та невирішені питання

- *Класичні методи vs. Глибоке навчання*
Прості, але чутливі до зовнішніх умов проти потужних, але ресурсоємних алгоритмів.
- *Універсальність vs. Спеціалізація*
Універсальні моделі не завжди точні у специфічних сценаріях, спеціалізовані — менш гнучкі.
- *Стійкість до зовнішніх факторів*
Освітлення, фон, швидкі жести знижують точність; сенсори дорогі й мають обмеження.
- *Індивідуальні відмінності користувачів*
Варіативність форм рук і стилів жестів ускладнює розпізнавання.
- *Обмеженість датасетів*
Брак реалістичних даних та висока вартість їх збору.
- *Компроміс між точністю та зручністю пристроїв*
Точні сенсори часто незручні або дорогі у використанні.
- *Проблема інтеграції*
Поєднання різних технологій дає найкращі результати, але вимагає складного проектування.

Рисунок Д.4 – Слайд 4

5

Обґрунтування методології дослідження

Фактори вибору алгоритмів:
точність, швидкодія, адаптивність до умов (освітлення, шум), ефективність при обмежених ресурсах.

Класичні методи:
SVM, KNN, HOG — ефективні для класифікації та виявлення, але обмежені відсутністю глибинної обробки даних.

Методи глибинного навчання:
CNN, RNN, LSTM, ViT — автоматичне виділення ознак та обробка великих даних для складних завдань (розпізнавання жестів, динамічні сцени).

Гібридні підходи:
CNN + RNN (LSTM) — поєднання просторових ознак і тимчасових залежностей для розпізнавання складних динамічних жестів.

Рисунок Д.5 – Слайд 5

6

Порівняльна оцінка алгоритмів

Для порівняння алгоритмів використовувався метод PROMETHEE, що дозволяє оцінювати алгоритми за кількома критеріями, такими як точність, швидкодія, стійкість, адаптація та інші. Зібрані дані, отримані з відкритих баз даних та результатів тестувань, допомогли порівняти алгоритми з різних аспектів ефективності. Ключові критерії включали:

- точність (Accuracy): отримана з наборів даних Sign Language MNIST та ChaLearn Gesture Dataset;
- швидкодія (Latency): час обробки кадру на стандартному обладнанні;
- стійкість (Robustness): вплив зовнішніх умов (шум, освітлення);
- обчислювальна складність (Complexity): потреби в обчислювальних ресурсах;
- адаптація до користувача (Adaptability): оцінка необхідності додаткового навчання;
- гнучкість (Flexibility): здатність працювати з різними умовами та типами даних.

Результати нормалізовані для забезпечення справедливого порівняння. Потоки переваг для кожного критерію обчислені за допомогою формули PROMETHEE, що дозволяє отримати комплексну оцінку ефективності алгоритмів у різних умовах.

Рисунок Д.6 – Слайд 6

Підсумки порівняння алгоритмів для розпізнавання жестів

Таблиця 1 – Підсумкова таблиця порівняння алгоритмів

Алгоритм	ϕ^+ (Вихідний потік)	ϕ^- (Вхідний потік)	ϕ (Чиста перевага)	Ранг
CNN + RNN (LSTM)	0.85	0.15	0.70	1
CNN	0.75	0.20	0.55	2
LSTM	0.65	0.25	0.40	3
VIT	0.60	0.30	0.30	4
RNN	0.55	0.35	0.20	5
SVM	0.30	0.50	-0.20	6
KNN	0.25	0.55	-0.30	7
HOG	0.20	0.60	-0.40	8

Висновок: алгоритм CNN + RNN (LSTM) є найкращим варіантом для задач розпізнавання жестів, оскільки поєднує переваги глибокого навчання для ефективного аналізу зображень та рекурентних нейронних мереж для обробки послідовних даних. Цей алгоритм забезпечує високу точність, швидкодію та стабільність, що робить його оптимальним вибором для реальних застосувань.

Рисунок Д.7 – Слайд 7

Програмна реалізація системи розпізнавання жестів

Опис процесу розробки

Для створення ефективної системи розпізнавання жестів було обрано технологічний стек, що поєднує комп'ютерний зір, нейронне навчання та управління пристроєм. Мова програмування Python була вибрана через свою зручність та підтримку бібліотек машинного навчання.

Вибрані мови програмування та фреймворки:

- Python – основна мова розробки, завдяки зручному синтаксису та великій кількості бібліотек;
- OpenCV – для захоплення відеопотоку з вебкамери та попередньої обробки зображень;
- MediaPipe – для відстеження положень руки та отримання координат ключових точок кисті;
- PyTorch Lightning – для побудови та тренування нейронної мережі;
- AutoPy, pynput, pycaw – для управління системними подіями, такими як рух миші, зміна гучності та натискання клавіш.

Рисунок Д.8 – Слайд 8

9

Програмна реалізація системи розпізнавання жестів

Основні етапи розробки:

1. збір та анотування даних: відеозапис жестів, витяг координат через MediaPipe, анотація та нормалізація даних;
2. навчання нейронної мережі: побудова моделі в PyTorch Lightning, підбір параметрів, навчання 100 епох, точність до 96%;
3. інтеграція компонентів: об'єднання відеопотоку та відстеження координат для передачі в модель, відповідність жестів подіям;
4. розробка інтерфейсу: консольний та графічний інтерфейс для запуску системи та візуалізації жестів;
5. підготовка до тестування: тестування на точність, швидкість і стійкість до зовнішніх чинників.

Висновок: розроблена система розпізнавання жестів забезпечує високу точність та ефективність завдяки поєднанню MediaPipe для відстеження рухів рук та нейронної моделі CNN + LSTM. Використання PyTorch Lightning полегшує навчання моделі, а інтеграція з бібліотеками для управління системними подіями створює зручний інтерфейс для користувачів.

Рисунок Д.9 – Слайд 9

10

Архітектура та функціональні можливості системи

Система для розпізнавання жестів включає кілька основних компонентів:

- модуль захоплення відео – підключає камеру і захоплює відеопотік у реальному часі (OpenCV);
- модуль обробки зображень – перетворює кадри для подальшого виявлення руки за допомогою MediaPipe;
- модуль трекінгу ключових точок – виявляє та нормалізує 21 ключову точку на руці (MediaPipe);
- модуль класифікації жестів – передає координати до нейронної мережі (CNN + LSTM) для розпізнавання статичних та динамічних жестів (PyTorch Lightning).

Ця модульна архітектура забезпечує гнучкість і масштабованість системи.

Рисунок Д.10 – Слайд 10

Тестування та оптимізація системи

Таблиця 2 – Тестування стандартних жестів

Жест	Точність (Accuracy)	Час відгуку (ms)	Нотації
Open Palm	98.5%	85	Висока точність при різних умовах освітлення.
Thumbs Up	96.3%	90	Точність трохи знижується при слабкому освітленні.
Four Fingers Up	95.1%	92	Важко розпізнати при швидких рухах.
Up Pinch One	97.8%	87	Добре працює з низьким рівнем шуму.
Palm Rotated Inward	94.2%	95	Вимагає стабільної руки для точного розпізнавання.

Таблиця 3 – Тестування режимів управління

Жест (режим)	Точність (Accuracy)	Час відгуку (ms)	Нотації
Open Palm (Mouse Mode)	98.2%	88	Швидке реагування на позицію пальця, мінімальні збої.
Open Pinch (Volume Mode)	96.9%	91	Легка втрата точності при різких рухах.
Shaka (Scroll Mode)	95.7%	93	Найбільша точність при постійному використанні одного жесту.

Рисунок Д.11 – Слайд 11

Тестування та оптимізація системи

- Адаптація до умов освітлення: застосування методів попередньої обробки зображень для покращення точності при змінних умовах освітлення.
- Швидкодія: використання технік для зниження часу обробки, таких як видалення фону і фільтрація малих об'єктів.
- Управління шумами: фільтраційні методи для зменшення впливу шуму та непотрібних рухів, забезпечуючи точність навіть при фонових шумах.

Висновок: оптимізація покращила точність та швидкодію системи в умовах змінного освітлення та шуму.

Рисунок Д.12 – Слайд 12

Інтерфейс та застосування жестів: розпізнавання і управління

Рисунок 1 – Розпізнавання моделлю різних варіацій жестів

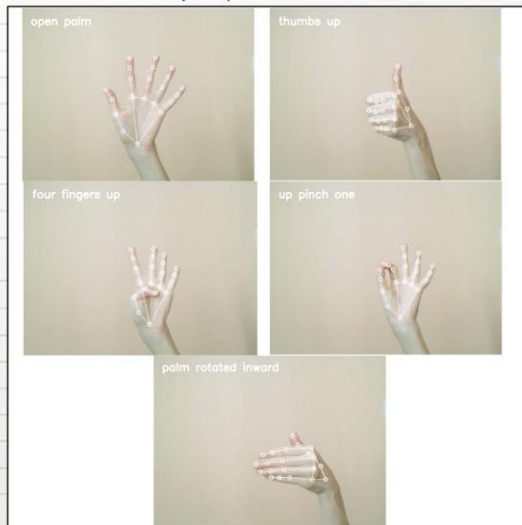


Рисунок 2 – Управління мишею, зміна гучності та навігація в браузері

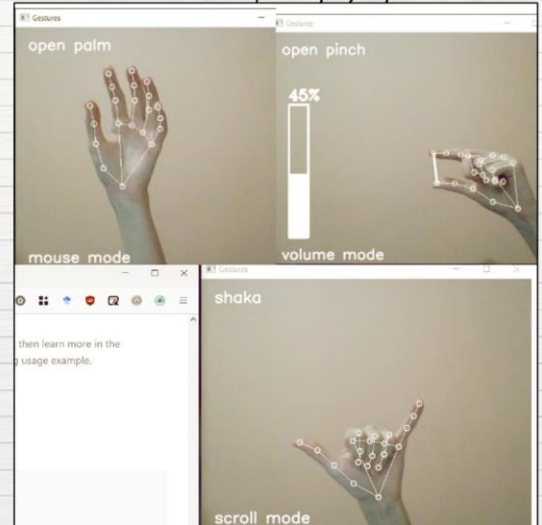


Рисунок Д.13 – Слайд 13

Підсумки дослідження

Реалістичність та корисність отриманих результатів

Розроблена система для розпізнавання жестів ефективно функціонує в умовах змінного освітлення, фонових перешкод і варіативності жестів. Показники точності моделі (92.0% accuracy, 91.7% precision, 90.9% recall та 91.3% F1-score) підтверджують її здатність працювати в реальному часі та з різними умовами, що робить систему корисною для безконтактного керування технічними пристроями.

Можливий розвиток досліджень

Подальші дослідження можуть включати розширення навчальних вибірок, використання мультимодальних трансформерів для покращення обробки відео та сенсорних даних, а також підтримку персоналізації моделей за допомогою few-shot learning, що підвищить адаптивність системи.

Висновок

Розробка та впровадження системи на основі гібридної архітектури CNN + RNN (LSTM) продемонстрували високі результати в точності та швидкодії. Це підтверджує перспективність технології для практичного використання у різних сферах, таких як VR/AR, робототехніка та медичні інтерфейси. Подальші вдосконалення дозволять ще більше підвищити ефективність системи.

Рисунок Д.14 – Слайд 14