

Владислав Сторощук¹, Софія Спільник², Ірина Бодник³, Шварц Михайло⁴

¹ Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, 12, Львів, Україна, E-mail: storoshchuk.mknsp.2024@lpnu.ua, ORCID 0009-0002-7595-8582

² Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, 12, Львів, Україна, E-mail: sofia.spilnyk.mknsp.2024@lpnu.ua, ORCID 0009-0007-1737-0437

³ Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, 12, Львів, Україна, E-mail: iryna.bodnyk.mknsp.2024@lpnu.ua, ORCID 0009-0006-3576-3325

⁴ Кафедра систем автоматизованого проектування, Національний університет “Львівська політехніка”, вул. С. Бандери, 12, Львів, Україна, E-mail: mykhailo.y.shvarts@lpnu.ua, ORCID 0000-0002-5567-442X

ДОСЛІДЖЕННЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ РОЗПІЗНАВАННЯ ЖЕСТИВ РУК

Отримано: Серпень 08, 2025 / Переглянуто: Серпень 26, 2025 / Прийнято: Вересень 15, 2025

© Сторощук В., Спільник С., Бодник І., Шварц М., 2025

<https://doi.org/10.23939/cds2025.02.156>

Анотація. Розроблено інтерактивну систему розпізнавання та класифікації жестів рук людини на основі технологій машинного навчання. Запропоновано новий підхід до подання жестів, який поєднує просторові та часові характеристики розташування ключових точок руки, що забезпечує високу точність, стійкість до шумів та адаптивність системи до різних умов використання. Особливістю розробки є метод інтерактивного навчання, який дає змогу користувачам без спеціальних технічних знань швидко додавати нові жести за обмеженої кількості навчальних даних. Виконано серію експериментів, у ході яких протестовано різні стратегії машинного навчання, що дало змогу виявити оптимальні моделі та підтвердити ефективність запропонованого підходу. Розроблена система легко інтегрується у сучасні інтерфейси людино-комп'ютерної взаємодії та може застосовуватися у сфері віртуальної й доповненої реальності, медичних технологіях, освітніх системах і безконтактних сервісах управління. Отримані результати відкривають перспективи для подальшого вдосконалення системи, зокрема підвищення швидкодії, масштабованості та персоналізації під потреби користувачів.

Ключові слова: розпізнавання жестів, машинне навчання, комп'ютерний зір, класифікація, безконтактна взаємодія.

Вступ

Жести рук здавна є важливим засобом комунікації, а в поєднанні із сучасними технологіями вони набувають нового значення як основа для створення інтуїтивних та зручних систем взаємодії із цифровими пристроями. З розвитком комп'ютерного зору, штучного інтелекту та машинного навчання з'явилася можливість автоматизованого розпізнавання жестів, що відкрило перспективи для формування нових поколінь інтерфейсів людино-комп'ютерної взаємодії.

Особливої актуальності такі системи набули в умовах стрімкого розвитку технологій віртуальної та доповненої реальності, а також після пандемії COVID-19, коли безконтактні інтерфейси стали надзвичайно затребуваними в побуті, освіті та професійній діяльності. Попри значний прогрес у галузях комп'ютерного зору та машинного навчання, завдання розпізнавання жестів рук залишається актуальним. Це пояснюється необхідністю створення систем, здатних забезпечувати високу точність і стійкість до зовнішніх факторів, а також адаптацію до різноманітних умов використання й індивідуальних особливостей користувачів.

Метою цього дослідження є розроблення інтерактивної системи розпізнавання та класифікації жестів рук людини, яка поєднує високу точність, стабільність роботи та можливість навчання новим жестам у режимі реального часу. Система орієнтована на використання звичайних вебкамер, що робить її доступною та зручною у практичних сценаріях. Крім того, вона надає користувачам змогу самостійно розширювати набір жестів, їм не потрібні для цього спеціальні знання, що істотно підвищує її практичну цінність та універсальність застосування.

Постановка проблеми

Дослідження акустичних характеристик закритих приміщень є складною багатофакторною проблемою, що потребує застосування сучасних інформаційно-вимірювальних технологій. Пропоновані методи та системи не завжди відповідають вимогам точності, завадостійкості та автоматизованого опрацювання сигналів. У зв'язку з цим виникає необхідність розроблення нових апаратно-програмних засобів, здатних здійснювати точні вимірювання з можливістю гнучкого налаштування параметрів. Основні завдання дослідження такі:

- аналіз сучасних методів та засобів дослідження адаптивної акустики;
- розроблення архітектури програмованої вбудованої системи для вимірювання акустичних параметрів;
- реалізація методу змішаного сигнального перетворення на основі селективного підсилення заряду;
- експериментальна перевірка ефективності запропонованої системи.

Огляд сучасних джерел інформації за тематикою публікації

У статті-огляді [1] “Survey on Hand Gesture Recognition from Visual Input” наведено систематичний огляд сучасного стану візуальних методів розпізнавання жестів, класифікуючи роботи за модальністю даних (RGB, глибина, RGB-D, мультивид), типом жестів (статичні, динамічні, безперервні) та застосуваннями (HCI, VR/AR, робототехніка). Автори підкреслюють, що перехід від класичних ознако-орієнтованих алгоритмів до архітектур на основі глибокого навчання істотно підвищив точність у контрольованих умовах, але вказують на стійку проблему перенесення моделей у реальні сценарії (освітлення, фон, різноманіття користувачів), а також на нестачу єдиних бенчмарків для безперервного розпізнавання.

У статті [2] Hakim et al. (2019) “Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM” наведено практичну комбінацію 3D-згортувачів (для вилучення просторово-часових ознак з відео) і LSTM (для моделювання довгих часових залежностей). Дослідження на прикладах динамічних наборів даних демонструє, що 3D-CNN дають змогу краще відображати рухову інформацію порівняно з 2D-CNN, а додавання LSTM поліпшує розпізнавання послідовних шаблонів. Проте автори вказують на високу обчислювальну вартість і труднощі з реальним часом на вбудованих пристроях, що спонукає до пошуку легших архітектур та компресії моделей (H. Hakim, D. Darmawan, і H. A. Nugroho, “Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM,” у Proc. Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS), 2019, pp. 93–98).

У статті [3] “Spatio-Temporal Dynamic Attention Graph Convolutional Network (STDA-GCN)” (2024) розглянуто підхід скелетно-орієнтованого розпізнавання жестів, де дані у вигляді ключових точок руки / пальців моделюються як граф із просторовими і часовими зв'язками. Автори доводять, що GCN з механізмами динамічної уваги краще захоплюють взаємозв'язки між суглобами та їхню еволюцію у часі, що підвищує точність за варіативності поз і швидкості жестів. Водночас робота підкреслює залежність від якості детекції ключових точок і вразливості до шуму в умовах низької якості відео, що робить важливою інтеграцію надійних 2D/3D-позних детекторів і підходів до фільтрації шуму (Y. Liu, F. Wang, і J. Chen, “Spatio-Temporal Dynamic Attention Graph Convolutional Network for Skeleton-based Gesture Recognition,” Pattern Recognition Letters, Vol. 178, pp. 85–92, 2024).

У статті-огляді та емпіричних дослідженнях щодо sEMG-підходів (поверхнева електроміографія) – зокрема систематичному огляді [4] та окремих 2024–2025 роботах – автори зазначають, що sEMG-сигнали є потужною альтернативою візуальним методам, особливо для застосувань у протезуванні та носимих інтерфейсах. Сучасні підходи комбінують багатопоточкові CNN / TCN / LSTM-архітектури або використовують трансфер-навчання для підвищення стабільності проти міжсуб'єктних варіацій; основні виклики – артефакти, потреба в калібруванні під конкретного користувача і нестача великих відкритих наборів sEMG для складних жестів.

У роботах, які стосуються датасетів та бенчмарків, останнім прикладом є [5] (Multi-View Hand gesture, 2025) – великий мультимодальний RGB-D набір із великою різноманітністю ракурсів і учасників. Автори MuViH обґрунтовують, що мультимодальне складання покращує стійкість моделей до зміни точки зору, але одночасно виявляють, що наявні архітектури не завжди ефективно використовують кореляцію між виглядами камер, тож потрібні спеціалізовані багаторівневі fusion-механізми. Крім того, багато публічних датасетів досі не відображають умов “в полі” – значних змін освітлення, часткового перекриття руки, різних фонів.

Останні кілька років характеризуються чітким зміщенням від ручних ознак до глибоких архітектур (3D-CNN, CNN+LSTM, GCN, трансформери) і до мультимодального ф'южену (RGB + depth + skeleton + sEMG). Водночас проблеми, які потребують подальшого розвитку, це (1) генералізація моделей у реальних, “шумних” умовах [6]; (2) ефективні легкі архітектури та методи компресії для роботи в реальному часі на вбудованих платформах [7]; (3) стандартизовані бенчмарки для безперервного розпізнавання та мультимодальної оцінки; (4) інтеграція сенсорних (sEMG, IMU) і візуальних потоків для підвищення надійності; (5) створення локальних / мовно-культурних датасетів (наприклад, для дактильної мови) – усе це прямо впливає з проаналізованих джерел і визначає перспективи розроблення інтерактивної системи розпізнавання та класифікації жестів [8, 9].

Постановка проблеми

Сучасні системи людино-комп'ютерної взаємодії все частіше потребують природних та інтуїтивних інтерфейсів [8–14], особливо в контексті стрімкого розвитку технологій віртуальної та доповненої реальності, а також зростання потреби у безконтактних рішеннях. Розпізнавання жестів рук є одним із найперспективніших напрямів створення таких інтерфейсів, проте наявні рішення мають істотні обмеження, що стримують їх широке практичне застосування.

Головною проблемою є необхідність забезпечити високу точність розпізнавання в умовах значної варіативності факторів навколишнього середовища. Зміна освітлення, кута знімання камери, відстані до користувача та індивідуальних анатомічних особливостей руки ускладнюють стабільну роботу систем. Крім того, більшість наявних рішень орієнтовані на розпізнавання лише статичних жестів або потребують використання спеціалізованого обладнання, що обмежує їх доступність та функціональні можливості для аналізу складних рухових шаблонів [14–20].

Особливо гострою є проблема розширюваності функціоналу. Традиційні системи працюють із фіксованим набором жестів і потребують значних технічних знань та ресурсів для додавання нових команд. Це створює бар'єр для користувачів, які хочуть адаптувати систему під специфічні потреби або культурні особливості. Водночас необхідність опрацювання як статичних поз руки, так і динамічних жестів, таких як рухи масштабування, кругові траєкторії або жести прокручування, в єдиній системі вимагає комплексного підходу до архітектури та алгоритмів машинного навчання.

Ще однією важливою проблемою є обчислювальна ефективність. Системи розпізнавання жестів повинні працювати в режимі реального часу на обладнанні з обмеженими ресурсами, використовуючи звичайні вебкамери без додаткових сенсорів глибини або спеціалізованих контролерів. Це потребує оптимізації алгоритмів для досягнення компромісу між точністю розпізнавання та швидкодією оброблення.

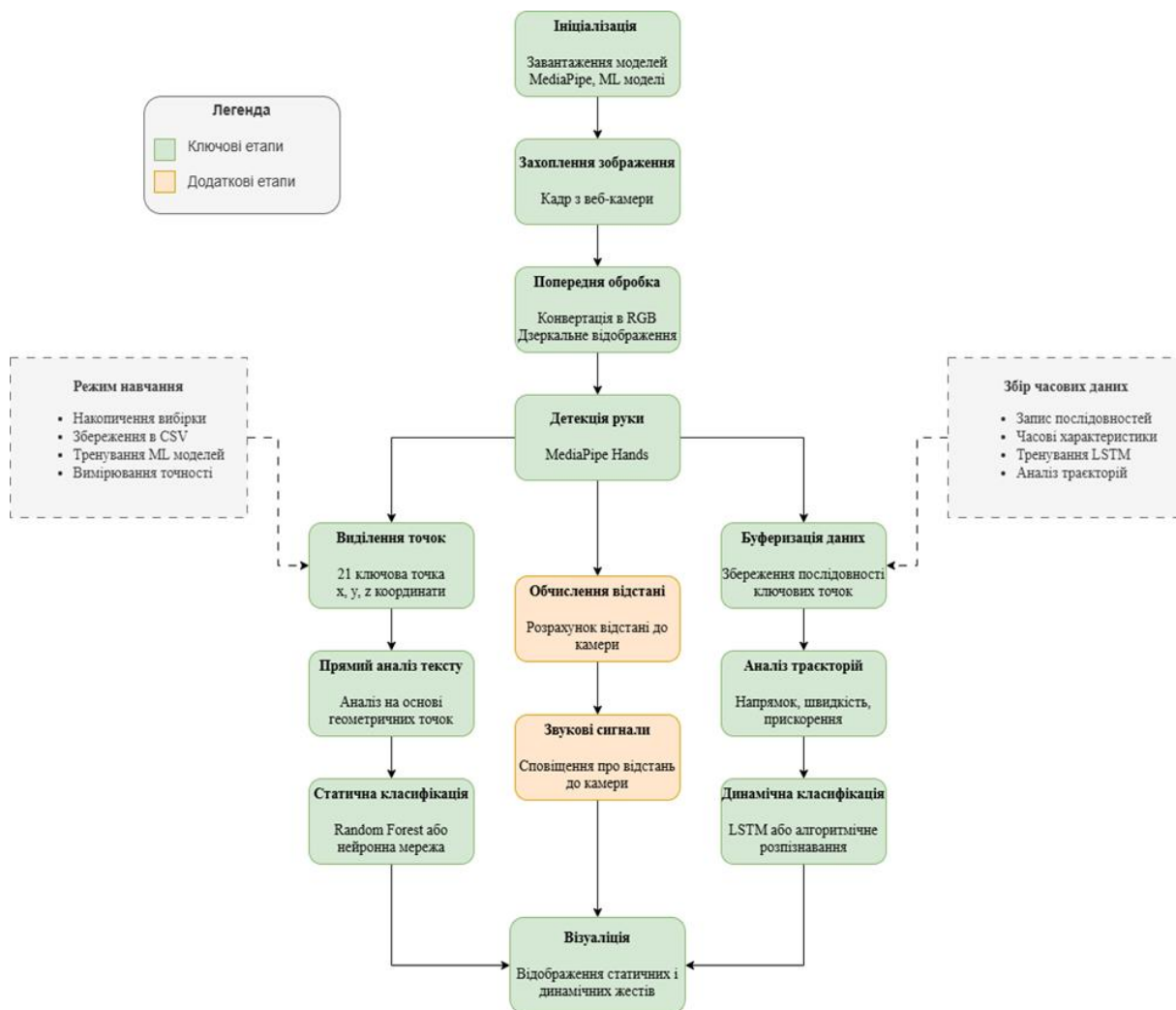
Отже, актуальним є завдання розроблення інтерактивної системи розпізнавання жестів, яка поєднує високу точність класифікації, стійкість до зовнішніх факторів, можливість інтерактивного

навчання новим жестам без спеціальних технічних знань, та ефективну роботу на доступному обладнанні. Вирішення цієї проблеми відкриє нові можливості для створення природних інтерфейсів людино-комп'ютерної взаємодії в різних сферах застосування.

Виклад основного матеріалу

1. Архітектура системи розпізнавання жестів

Розроблена система має модульну архітектуру, що забезпечує гнучкість у виборі алгоритмів та можливість розширення функціоналу (див. рисунок). Система складається з двох основних гілок обробки: статичної та динамічної класифікації жестів, що дає змогу комплексно аналізувати як миттєві пози руки, так і рухи в часі.



Архітектура системи розпізнавання жестів рук

Перший етап опрацювання передбачає захоплення зображення із вебкамери та його попереднє оброблення – конвертацію в RGB формат і дзеркальне відображення для зручності користувача. Детектування руки здійснюється за допомогою неймережових алгоритмів, що виділяють 21 ключову точку руки з координатами (x, y, z) у нормалізованому просторі [20–27].

Основна особливість архітектури полягає в паралельній обробці даних двома незалежними модулями. Статичний модуль аналізує геометричні співвідношення між ключовими точками для миттєвого розпізнавання поз, тоді як динамічний модуль буферизує послідовності кадрів для виявлення рухових шаблонів. У систему також входить модуль контролю відстані, що забезпечує оптимальне позиціонування користувача відносно камери.

Класифікація методів розпізнавання жестів

Метод розпізнавання	Особливості	Переваги	Недоліки
На основі ключових точок	Використання 21 ключової точки MediaPipe	Висока точність, стійкість до часткових перекриттів	Залежність від якості виділення точок
На основі зображення руки цілком	Аналіз руки як цілісного зображення	Простота реалізації	Низька стійкість до зміни освітлення та ракурсу
На основі просторово-часових характеристик	Аналіз послідовності положень руки	Можливість розпізнавання динамічних жестів	Висока обчислювальна складність
На основі глибинних карт	Використання даних про глибину	Додаткова інформація про позицію руки	Потреба в спеціальному обладнанні

2. Математичний апарат витягування ознак

Для статичних жестів формування вектора ознак здійснюється лінеаризацією координат ключових точок:

$$F_static = [x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_{21}, y_{21}, z_{21}], \quad (1)$$

де (x_i, y_i, z_i) представляють нормалізовані координати i -ї ключової точки, що забезпечує інваріантність до розміру зображення.

Для підвищення стійкості системи до варіацій відстані реалізовано алгоритм оцінювання віддаленості руки від камери на основі геометричних властивостей долоні:

$$d = (R_ref \times D_ref) / H_palm, \quad (2)$$

де $R_ref = 200$ пікселів – еталонний розмір долоні; $D_ref = 50$ см – еталонна відстань; H_palm – вимірний розмір долоні в поточному кадрі.

Динамічні жести характеризуються часовими змінами положення ключових точок. Швидкість руху для кожної точки розраховують як:

$$V(t) = [P(t) - P(t-1)] / \Delta t. \quad (3)$$

Прискорення визначається диференціюванням швидкості:

$$A(t) = [V(t) - V(t-1)] / \Delta t. \quad (4)$$

Комплексний вектор ознак для динамічних жестів такий:

$$F_temporal = [P_last, \bar{V}, V_max, \bar{A}, D_total, T_direction], \quad (5)$$

де P_last – координати останнього кадру; $\bar{V}$ – середня швидкість на всіх точках; V_max – максимальна швидкість; $\bar{A}$ – середнє прискорення; D_total – загальна довжина траєкторії; $T_direction$ – вектор загального напрямку руху.

3. Алгоритми класифікації та машинного навчання

Для статичних жестів реалізовано два основні підходи: Random Forest та нейронні мережі. Random Forest забезпечує швидке навчання та інтерпретованість результатів:

$$\hat{G} = \text{mode}(h_1(F), h_2(F), \dots, h_n(F)), \quad (6)$$

де $h_i(F)$ – рішення i -го дерева з ансамблю; $\hat{G}$ – фінальне рішення на основі голосування.

Нейронна мережа для статичних жестів має архітектуру з двома прихованими шарами (128 та 64 нейрони) та dropout для запобігання перенавчанню. Для динамічних жестів використовується LSTM архітектура, що моделює часові залежності:

$$h_t = \text{LSTM}(F_t, h_{t-1}), \quad (7)$$

$$p = \text{softmax}(W_h \times h_T + b), \quad (8)$$

де h_t – прихований стан на кроці t ; F_t – вектор ознак поточного кадру; p – розподіл ймовірностей за класами жестів.

Таблиця 2

Порівняння алгоритмів машинного навчання для класифікації

Алгоритм	Точність на тестовому наборі, %	Швидкодія, мс	Складність реалізації	Вимоги до обчислювальних ресурсів
Random Forest	92	8	Низька	Низькі
Нейронна мережа (Dense)	94	15	Висока	Середні
LSTM (динамічні жести)	87	25	Висока	Високі

4. Оптимізація та фільтрація сигналів

Для забезпечення стабільності роботи системи реалізовано алгоритми згладжування та фільтрації сигналів. Відстань до камери згладжується за допомогою ковзного середнього:

$$d_{smooth}(t) = \frac{1}{n} \sum_{i=0}^{n-1} d(t-i), \quad (9)$$

де $n = 5$ – розмір вікна згладжування.

Детектування кругових рухів ґрунтується на аналізі консистентності напрямку обертання та стабільності радіальної відстані від центра траєкторії:

$$\alpha_{consistency} = \frac{\max(\Delta\theta^+, \Delta\theta^-)}{|\Delta\theta|}, \quad (10)$$

$$\sigma_{distance} = \frac{\sigma_r}{\mu_r}, \quad (11)$$

$$C_{circular} = (\alpha_{consistency} > 0,7) \cap (\sigma_{distance} < 0,3), \quad (12)$$

де $\Delta\theta^+$ та $\Delta\theta^-$ – кількість позитивних і негативних змін кута; σ_r та μ_r – стандартне відхилення та середнє значення реальних відстаней.

Жести масштабування детектуються через аналіз зміни евклідової відстані між ключовими точками великого та вказівного пальців:

$$Z_{detected} = |d_{final} - d_{initial}| > \epsilon. \quad (13)$$

Тип масштабування визначається знаком зміни відстані:

$$Z_{type} = \begin{cases} \text{Zoom In, якщо } d_{final} - d_{initial} < 0 \\ \text{Zoom Out, якщо } d_{final} - d_{initial} > 0 \end{cases}. \quad (14)$$

Хвилеподібні рухи класифікують на основі кількості локальних екстремумів у вертикальній проєкції траєкторії:

$$W_{detected} = (N_{peaks} + N_{valleys}) \geq 3, \quad (15)$$

де N_{peaks} та $N_{valleys}$ – кількість локальних максимумів і мінімумів у-координати траєкторії відповідно.

Таблиця 3

Характеристики розробленої системи розпізнавання жестів

Характеристика	Значення
Кількість базових жестів	8
Можливість додавання нових жестів	Так
Мінімальна кількість зразків для навчання	50
Частота кадрів під час роботи в реальному часі	30 FPS
Середня точність розпізнавання	94 %
Затримка розпізнавання	< 100 мс
Підтримувані платформи	Windows, Linux, macOS
Необхідність спеціального обладнання	Ні (достатньо вебкамери)

5. Експериментальні результати та аналіз продуктивності

Систему протестовано на наборі із восьми базових статичних жестів та восьми динамічних жестів із різною кількістю навчальних зразків для кожного класу. Загальна точність системи становила 94 % для статичних жестів та 87 % для динамічних.

Таблиця 4

Аналіз точності розпізнавання жестів за типами

Тип жесту	Точність (Random Forest), %	Точність (нейронна мережа), %	Основні причини помилок	Рекомендації для підвищення точності
Статичні прості жести (кулак, відкрита долоня)	96	98	Схожість із іншими жестами	Збільшення кількості навчальних зразків
Статичні складні жести (із переплетінням пальців)	88	95	Складність виділення точок	Додаткові ознаки на основі відстаней між точками
Динамічні жести (рухи рукою)	85	94	Непослідовність виконання	Використання рекурентних нейронних мереж
Жести з частковим перекриттям	82	91	Втрата даних про ключові точки	Імпуція даних з використанням попередніх кадрів
Жести на великій відстані від камери	78	85	Низька роздільна здатність	Масштабування та додаткова нормалізація

Аналіз факторів впливу (табл. 5) показав, що найбільше впливають на точність освітлення та відстань до камери, проте розроблені алгоритми компенсації дають змогу істотно знизити їх негативний ефект.

Таблиця 5

Аналіз впливу різних факторів на точність розпізнавання

Фактор	Ступінь впливу	Методи компенсації	Залишковий вплив після компенсації
Освітлення	Високий	Нормалізація яскравості, контрасту	Низький
Кут знімання	Середній	Просторова нормалізація	Низький
Відстань до камери	Високий	Масштабування координат	Низький
Індивідуальні особливості руки	Середній	Нормалізація відносних пропорцій	Середній
Швидкість виконання жесту	Низький	Часова нормалізація	Дуже низький
Фон зображення	Низький	Сегментація руки	Дуже низький
Якість камери	Середній	Підвищення роздільної здатності	Середній

Обчислювальна складність системи становить $O(n \times d)$ для статичних жестів, де n – кількість класифікаторів; $d = 63$ – розмірність вектора ознак; та $O(T \times H^2)$ для динамічних жестів, де T – довжина часової послідовності; H – розмір прихованого стану LSTM. Середній час відгуку системи не перевищує 100 мс, що забезпечує комфортну роботу в режимі реального часу.

Висновки

У результаті виконання дослідження розроблено комплексну систему розпізнавання жестів рук, яка успішно поєднує аналіз статичних поз та динамічних рухів у єдиній архітектурі. Система демонструє високу ефективність у розпізнаванні як миттєвих жестів, так і складних рухових шаблонів, забезпечуючи середню точність 94 % для статичних жестів та 87 % для динамічних.

Ключовою перевагою розробленого рішення є модульна архітектура, що дає змогу комбінувати різні алгоритми машинного навчання відповідно до специфіки завдання. Порівняльний аналіз показав, що нейронна мережа Dense забезпечує найвищу точність для статичних жестів (94 %), тоді як Random Forest демонструє оптимальний баланс між точністю (92 %) та швидкістю (8 мс), тому він придатний для застосувань реального часу з обмеженими ресурсами. Для динамічних жестів LSTM архітектура забезпечила точність 87 % за часу відгуку 25 мс, що підтверджує ефективність рекурентних мереж для аналізу часових послідовностей рухових шаблонів.

Математичний апарат системи містить комплексний набір просторово-часових ознак, що забезпечує стійкість до варіацій зовнішніх факторів. Алгоритми згладжування сигналів та фільтрації шумів істотно підвищили стабільність роботи в реальних умовах використання. Реалізовані методи компенсації впливу освітлення, кута знімання та відстані до камери знизили їх негативний ефект до мінімального рівня.

Особливо цінна система завдяки можливості інтерактивного навчання новим жестам без потреби у спеціальних технічних знаннях. Мінімальна кількість навчальних зразків (50 для кожного класу) та інтуїтивний інтерфейс роблять систему доступною для широкого кола користувачів. Це відкриває перспективи для персоналізації та адаптації під специфічні потреби різних сфер застосування.

Обчислювальна ефективність системи забезпечує роботу в режимі реального часу на стандартному обладнанні без додаткових сенсорів. Затримка розпізнавання менше ніж 100 мс та підтримка частоти 30 FPS створюють комфортні умови для взаємодії користувача із системою. Кросплатформність рішення гарантує широкі можливості інтеграції у різні програмні та апаратні середовища.

Експериментальна перевірка підтвердила ефективність запропонованих алгоритмів детекції специфічних жестових патернів, урахування кругові рухи, масштабування та хвилеподібні траєкторії. Розроблені методи правило-орієнтованого розпізнавання забезпечують функціональність системи навіть за відсутності попередньо навчених моделей.

Практична значущість роботи полягає у створенні готового до використання програмного рішення із відкритим кодом, що може бути інтегровано у системи віртуальної та доповненої реальності, медичні технології, освітні платформи та безконтактні інтерфейси управління. Модульна архітектура та документований код полегшують подальший розвиток та адаптацію системи.

Перспективи подальших досліджень охоплюють розширення набору розпізнаваних жестів, оптимізацію алгоритмів для мобільних платформ, інтеграцію із додатковими сенсорними модальностями та розроблення спеціалізованих рішень для людей з обмеженими можливостями. Потенціал системи для створення природних людино-комп'ютерних інтерфейсів відкриває широкі можливості для інновацій у сфері взаємодії людини з цифровими технологіями.

Перелік використаних джерел

- [1] M. Linardakis, I. Varlamis, i G. Th. Papadopoulos, "Survey on Hand Gesture Recognition from Visual Input", arXiv preprint *arXiv:2501.11992*, 2025.
- [2] H. Hakim, D. Darmawan, i H. A. Nugroho, "Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM", у Proc. Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS), 2019, 93–98.
- [3] Y. Liu, F. Wang, i J. Chen, "Spatio-Temporal Dynamic Attention Graph Convolutional Network for Skeleton-based Gesture Recognition", *Pattern Recognition Letters*, Vol. 178, 85–92, 2024.

- [4] A. Sultana, “A Systematic Review on sEMG-based Hand Gesture Recognition Using Deep Learning Methods”, *Bioengineering*, Vol. 10, No. 3, 225–239, 2023. <https://doi.org/10.3390/bioengineering10030225>
- [5] K. Zhang, Y. Zhao, i L. Huang, “MuViH: A Multi-View Hand Gesture Dataset for Robust Recognition”, *IEEE Access*, Vol. 13, 45012–45025, 2025. <https://doi.org/10.1109/ACCESS.2025.102957>
- [6] Zhang F., Bazarevsky V., Vakunov A., et al. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv preprint *arXiv:2006.10214*, 2020.
- [7] Chang W., Hsu F., Cherng S. Deep Learning Recognition Method for Hand Gestures. *Applied Sciences*, 2021, 11(6), 2735.
- [8] Wang Y., Lin Z., Liang X., et al. GestureRecognizer: A Toolbox for Gesture Recognition Built with Temporal Convolution Networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–4.
- [9] Li Y., Sun X., Zhang Y., et al. Human-computer interaction using dynamic hand gestures based on deep learning. *Information Sciences*, 2021, 545, pp. 713–730.
- [10] Pomorova O.V., Hovorushchenko T.O. *Intelektualnyi analiz ta obrobka danykh*. Khmelnytskyi: KhNU, 2018. 236 s.
- [11] MediaPipe – ofitsiyni sait biblioteki dlia rozpoznavannia obektiv [Elektronnyi resurs]. Rezhym dostupu: <https://developers.google.com/mediapipe>
- [12] TensorFlow – ofitsiina dokumentatsiia freimvorku dlia mashynnoho navchannia [Elektronnyi resurs]. Rezhym dostupu: https://www.tensorflow.org/api_docs
- [13] OpenCV – biblioteka kompiuternoho zoru z vidkrytym kodom [Elektronnyi resurs]. Rezhym dostupu: <https://opencv.org/>
- [14] Scikit-learn – biblioteka mashynnoho navchannia dlia Python. [Elektronnyi resurs]. Rezhym dostupu: <https://scikit-learn.org/>
- [15] PyTorch – biblioteka dlia mashynnoho navchannia z vidkrytym kodom. [Elektronnyi resurs]. Rezhym dostupu: <https://pytorch.org/>
- [16] Molchanov P., Yang X., Gupta S., et al. Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4207–4215.
- [17] GitHub repozytorii projektu MediaPipe [Elektronnyi resurs]. Rezhym dostupu: <https://github.com/google/mediapipe>
- [18] Kaggle – dataset zhestiv ruk dlia mashynnoho navchannia [Elektronnyi resurs]. Rezhym dostupu: <https://www.kaggle.com/datasets/gti-upm/leapgestrecog>
- [19] NumPy – biblioteka dlia naukovykh obchyslen u Python [Elektronnyi resurs]. Rezhym dostupu: <https://numpy.org/>
- [20] Martynenko S. S., Kot T. M. *Metody rozpoznavannia dynamichnykh zhestiv dlia keruvannia interfeisamy. Systemy obrobky informatsii*, 2019, 2(157), 74–82.
- [21] Python Software Foundation – ofitsiyni sait movy prohramuvannia Python [Elektronnyi resurs]. Rezhym dostupu: <https://www.python.org/>
- [22] Natsionalna biblioteka Ukrainy imeni V. I. Vernadskoho – elektronnyi arkhiv naukovykh publikatsii [Elektronnyi resurs]. Rezhym dostupu: <http://www.nbuv.gov.ua/>
- [23] Stack Overflow – forum dlia rozrobnykiv prohramnogo zabezpechennia [Elektronnyi resurs]. Rezhym dostupu: <https://stackoverflow.com/>
- [24] Kharchenko V. S., Yakovlev S. V. *Metody mashynnoho navchannia dlia kiberfizychnykh system: analiz ta perspektyvy. Radioelektronni i kompiuterni systemy*, 2020, 1(93), 24–36.
- [25] Real Python – Resurs dlia vyvchennia Python ta yoho zastosuvannia v realnykh projektakh [Elektronnyi resurs]. Rezhym dostupu: <https://realpython.com/>
- [26] GitLab – platforma dlia rozrobky prohramnogo zabezpechennia [Elektronnyi resurs]. Rezhym dostupu: <https://about.gitlab.com/>
- [27] W3Schools – Resurs dlia vyvchennia veb-tekhnologii [Elektronnyi resurs]. Rezhym dostupu: <https://www.w3schools.com/>

Vladyslav Storoshchuk¹, Spilnyk Sofiia², Iryna Bodnyk³, Shvarts Mykhaylo⁴

¹ Department of Computer Design Systems, Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: vladyslav.storoshchuk.mknsnp.2024@lpnu.ua, ORCID 0009-0002-7595-8582

² Department of Computer Design Systems, Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: sofii.spilnyk.mknsnp.2024@lpnu.ua, ORCID 0009-0007-1737-0437

² Department of Computer Design Systems, Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: iryna.bodnyk.mknsnp.2024@lpnu.ua, ORCID 0009-0006-3576-3325

² Department of Computer Design Systems, Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: mykhailo.y.shvarts@lpnu.ua, ORCID 0000-0002-5567-442X

RESEARCH AND SOFTWARE IMPLEMENTATION OF HAND GESTURE RECOGNITION METHODS

Received: August 08, 2025 / Revised: August 18, 2025 / Accepted: September 15, 2025

© *Storoshchuk V., Spilnyk S., Bodnyk I., Shvarts M., 2025*

Abstract. The article presents the development of an interactive system for recognizing and classifying human hand gestures based on machine learning technologies. A new approach to gesture representation is proposed, combining spatial and temporal characteristics of the location of key points of the hand, which ensures high accuracy, noise resistance, and adaptability of the system to various conditions of use. A distinctive feature of the development is the interactive learning method, which allows users without special technical knowledge to quickly add new gestures even with a limited amount of training data. A series of experiments was conducted to test various machine learning strategies, which made it possible to identify optimal models and confirm the effectiveness of the proposed approach. The system can be easily integrated into modern human-computer interaction interfaces and can be used in virtual and augmented reality, medical technologies, educational systems, and contactless control services. The results obtained open up opportunities for further improvement of the system, in particular in terms of increasing performance, scalability, and personalization to meet user needs.

Keywords: gesture recognition, machine learning, computer vision, classification, contactless interaction.