

МАГІСТЕРСЬКА РОБОТА

МР. ШМ - 29.00.00.000 ПЗ

Група ШМ-24-2

Микитенко Максим

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Микитенко Максим Володимирович

(прізвище, ім'я, по батькові)

УДК 004.9
(індекс)

МАГІСТЕРСЬКА РОБОТА

Моделі та методи рекурентних нейромереж для розпізнавання

активності людини на основі релевантних сенсорних даних

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Микитенко М.В.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Вовк Роман Богданович, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Нормоконтроль

доц. Вовк Р.Б.

(посада) (підпис) (дата) (ініціали та прізвище)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедрою

ІПЗ

доц.

В.В. Бандура

“ 04 ” вересня 2025 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Микитенку Максиму Володимировичу

(прізвище, ім'я, по-батькові)

1. Тема магістерської роботи “ **Моделі та методи рекурентних нейромереж для розпізнавання активності людини на основі релевантних сенсорних даних**”

керівник проекту (роботи) Вовк Роман Богданович, к.т.н., доцент

затверджені наказом закладу вищої освіти від “ 05 ” листопада 2025 р. № 695/7

2. Строк подання студентом проекту (роботи) 15 грудня 2025 р.

3. Вихідні дані до проекту (роботи) Теоретичні концепції та формальні моделі побудови рекурентних нейронних мереж для задач розпізнавання

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Аналіз предметної області використання нейромереж для розпізнавання активності

2. Методологія розпізнавання людської активності на основі сенсорів і LSTM моделі

3. Дослідження моделей та методів розпізнавання активності людини на основі сенсорних даних

4. Застосування рекурентних нейромереж для розпізнавання активності людини

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Розгортання рекурентного зв'язку РНМ вздовж часової осі (рис. 1.1)

2. Внутрішня архітектура та логіку роботи однієї комірки LSTM (рис. 1.2)

3. Графічне представлення архітектури BLSTM (рис. 1.3)

4. Огляд запропонованої моделі для розпізнавання активності (рис. 1.4)

5. Типові позиції розміщення інертних сенсорів (рис. 1.5)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2025 р.

Керівник

_____ (підпис)

Завдання прийняв до виконання _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	15.09.2025	виконано
2	Аналіз предметної області використання нейромереж для розпізнавання активності	29.09.2025	виконано
3	Методологія розпізнавання людської активності на основі сенсорів і LSTM моделі	15.10.2025	виконано
4	Дослідження моделей та методів розпізнавання активності людини на основі сенсорних даних	08.11.2025	виконано
5	Застосування рекурентних нейромереж для розпізнавання активності людини	20.11.2025	виконано
6	Архітектура BLSTM мережі для диференціації релевантних активностей	01.12.2025	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.12.2025	виконано

Студент – магістр

_____ (підпис)

Керівник роботи

_____ (підпис)

АНОТАЦІЯ

Магістерська робота: 82 с., 45 рис., 3 табл., 35 джерел.

Тема: Моделі та методи рекурентних нейромереж для розпізнавання активності людини на основі релевантних сенсорних даних

Мета магістерської роботи - розроблення та дослідження моделей і методів рекурентних нейронних мереж для ефективного розпізнавання активності людини на основі релевантних сенсорних даних.

Об'єкт дослідження - процеси розпізнавання активності людини на основі часових сенсорних даних, отриманих із інертних та мультимодальних сенсорів.

Предмет дослідження - моделі та методи рекурентних нейронних мереж для класифікації та інтерпретації часових сенсорних послідовностей у задачах розпізнавання людської активності.

Результати дослідження

В роботі розроблено архітектуру BLSTM-моделі, оптимізовану для класифікації людської активності, з урахуванням вибору функцій активації, технік регуляризації та алгоритмів оптимізації

Висновок

Розроблено удосконалену архітектуру двоспрямованої рекурентної мережі (BLSTM) для задач розпізнавання людської активності, що враховує часові залежності у двох напрямках та представлено методологію попередньої обробки та сегментації сенсорних даних за допомогою методу ковзного вікна.

РЕКУРЕНТНА НЕЙРОННА МЕРЕЖА, LSTM, BLSTM, РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ, СЕНСОРНІ ДАНІ, ГЛИБОКЕ НАВЧАННЯ, ЧАСОВІ РЯДИ, КЛАСИФІКАЦІЯ, ОПТИМІЗАЦІЙНІ АЛГОРИТМИ.

ABSTRACT

Master Thesis: 82 pp., 45 fig., 3 tab., 35 sources.

Topic: Models and methods of recurrent neural networks for recognizing human activity based on relevant sensory data

The purpose of the master's thesis is to develop and study models and methods of recurrent neural networks for effective recognition of human activity based on relevant sensory data.

The object of research is the processes of recognizing human activity based on temporal sensory data obtained from inert and multimodal sensors.

The subject of research is models and methods of recurrent neural networks for classifying and interpreting temporal sensory sequences in human activity recognition tasks.

Research results

The work developed the architecture of the BLSTM model optimized for human activity classification, taking into account the choice of activation functions, regularization techniques and optimization algorithms

Conclusion

An improved architecture of a bidirectional recurrent network (BLSTM) for human activity recognition tasks was developed, which takes into account time dependencies in two directions and a methodology for preprocessing and segmentation of sensory data using the sliding window method was presented.

RECURRENT NEURAL NETWORK, LSTM, BLSTM, HUMAN ACTIVITY RECOGNITION, SENSOR DATA, DEEP LEARNING, TIME SERIES, CLASSIFICATION, OPTIMIZATION ALGORITHMS.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	10
ВСТУП.....	11
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИКОРИСТАННЯ РЕКУРЕНТНИХ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ	15
1.1. Особливості, принцип роботи та класифікація рекурентних нейронних мереж	15
1.2. Методологія розпізнавання людської активності на основі сенсорів із застосуванням LSTM моделі	18
1.3. Опис архітектури BLSTM (Bidirectional Long Short-Term Memory)	20
1.3.1. Концепція LSTM.....	20
1.3.2. Принцип BLSTM	21
1.3.3. Переваги в процесах розпізнавання людської активності	22
1.4. Аналітичний огляд методології розпізнавання людської активності на основі глибокого навчання	23
1.4.1. Дані та сенсорна модальність.....	23
1.4.2. Розпізнавання активності як задача класифікації	24
1.4.3. Пропонована модель для розпізнавання активності	25
Висновки до розділу	27
РОЗДІЛ 2. ДОСЛІДЖЕННЯ МОДЕЛЕЙ ТА МЕТОДІВ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ НА ОСНОВІ РЕЛЕВАНТНИХ СЕНСОРНИХ ДАНИХ	29
2.1. Класифікація методів розпізнавання людської активності та деталізація сенсорно-орієнтованих підходів	29
2.1.1. Особливості інертних сенсорів	30
2.1.2. Опис набору даних UTD-MHAD	31

2.2. Еволюція методологій розпізнавання людської активності	34
2.2.1. Обмеження традиційних підходів.....	34
2.3. Огляд пов'язаних досліджень та методологій	35
2.3.1. Класичні методи машинного навчання	35
2.3.2. Методи глибокого навчання та гібридні архітектури.....	40
Висновки до розділу	50

РОЗДІЛ 3. ЗАСТОСУВАННЯ МОДЕЛЕЙ ТА МЕТОДІВ РЕКУРЕНТНИХ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ НА ОСНОВІ РЕЛЕВАНТНИХ СЕНСОРНИХ ДАНИХ	51
3.1. Моделі довгої короткочасної пам'яті та їх двоспрямовані варіанти (BLSTM) для задач розпізнавання активності людини	51
3.1.1. Односпрямована мережа LSTM.....	51
3.1.2. Двоспрямована мережа LSTM (BLSTM).....	53
3.2. Опис набору даних UTD-MHAD для дослідження	55
3.3. Архітектура BLSTM мережі для диференціації релевантних активностей	58
3.4. Методологія підготовки даних та експериментальна конфігурація	60
3.4.1. Підготовка даних та експериментальні розподіли.....	60
3.4.2. Програмне забезпечення та обчислювальне середовище.....	61
3.4.3. Метод ковзного вікна (sliding window).....	61
3.5. Функція активації та техніка регуляризації в моделі BLSTM.....	62
3.5.1. Функція активації	62
3.5.2. Техніка регуляризації	63
3.6. Функція втрат та оптимізаційні алгоритми для навчання моделі	65
3.6.1. Функція втрат. Категоріальна перехресна ентропія (Categorical Cross-Entropy)	65
3.6.2. Оптимізаційні алгоритми	66
3.7. Результати експериментальної оцінки моделей	68
3.7.1. Аналіз точності (Accuracy).....	68

3.7.2. Комплексні метрики оцінки (повнота, точність, F1-оцінка)	69
3.7.3. Аналіз матриць невідповідності (Confusion Matrix)	72
Висновки до розділу	75
ВИСНОВКИ	77
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	79

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

- AR - Activity Recognition - Розпізнавання активності
- ANN - Artificial Neural Network - Штучна нейронна мережа
- BRNN Bidirectional Recurrent Neural Network
- BLSTM - Bidirectional Long Short-Term Memory
- DNN - Deep Neural Network
- FFT - Fast Fourier Transform - Швидке перетворення Фур'є
- HAR - Human Activity Recognition - Розпізнавання людської активності
- HMM - Hidden Markov Model - Прихована Марковська модель
- LSTM - Long Short-Term Memory - Довга короткочасна пам'ять
- RBM - Restricted Boltzmann Machine
- SAE - Sparse Auto-Encoder
- RBF SVM - Radial Basis Function Support Vector Machine

ВСТУП

Актуальність теми.

Сучасний етап розвитку інформаційних технологій характеризується стрімким зростанням кількості інтелектуальних пристроїв, здатних збирати та передавати великі обсяги даних про фізичну активність користувачів. Зокрема, сенсори, вбудовані у смартфони, фітнес-браслети, смарт-годинники, а також системи “розумного дому”, формують нові можливості для аналізу поведінкових патернів людини. Одним із ключових напрямів обробки таких даних є розпізнавання активності людини (Human Activity Recognition, HAR), що має широкий спектр застосувань — від медичного моніторингу до автоматизації побутових і виробничих процесів.

Проте на практиці задача HAR ускладнюється низкою факторів: високою динамічністю сигналів, наявністю шумів у сенсорних даних, різноманітністю типів активностей та індивідуальними відмінностями користувачів. Традиційні методи машинного навчання, засновані на ручному виділенні ознак, не завжди забезпечують необхідну точність і узагальнювальну здатність моделей. У зв'язку з цим актуалізується застосування методів глибокого навчання, зокрема рекурентних нейронних мереж (RNN), здатних моделювати часові залежності в послідовних даних.

Особливе місце серед таких моделей посідають архітектури довгої короткочасної пам'яті (LSTM) та двоспрямовані LSTM (BLSTM), які демонструють високу ефективність у задачах аналізу часових рядів. Вони дають змогу не лише враховувати залежності між попередніми станами системи, а й використовувати контекст майбутніх спостережень, що особливо важливо для точного розпізнавання складних або комбінованих видів активності.

Таким чином, актуальним науковим завданням є розроблення та дослідження моделей і методів рекурентних нейромереж, здатних здійснювати автоматичне розпізнавання активності людини на основі

релевантних сенсорних даних, з урахуванням особливостей часової структури сигналів та варіацій між користувачами.

Актуальність теми дослідження зумовлена потребою в удосконаленні інтелектуальних систем аналізу людської поведінки, що використовують дані з мобільних та носимих сенсорів. Розпізнавання активності є базовим елементом у таких галузях, як охорона здоров'я (контроль фізичної активності, реабілітація, моніторинг стану пацієнтів), “розумні” середовища (smart home, smart office), спортивна аналітика, безпека та індустріальні системи контролю.

Зростання обсягів і складності сенсорних даних вимагає використання алгоритмів, здатних не лише класифікувати дії, але й адаптуватися до зміни контексту, шумів та неповних даних. Рекурентні нейромережі, зокрема моделі LSTM та BLSTM, відкривають нові можливості для побудови систем, що враховують часові залежності між вимірюваннями, забезпечують високу точність розпізнавання і є стійкими до спотворень у вхідних сигналах.

Попри значні досягнення у сфері глибокого навчання, питання інтеграції релевантних сенсорних модальностей, оптимізації архітектур RNN та підвищення ефективності їх навчання залишаються актуальними. Це визначає необхідність розроблення комплексних моделей, які поєднують аналітичну потужність рекурентних мереж із практичною придатністю до реального впровадження у системи моніторингу й підтримки прийняття рішень.

Метою магістерської роботи є розроблення та дослідження моделей і методів рекурентних нейронних мереж для ефективного розпізнавання активності людини на основі релевантних сенсорних даних.

Об'єкт дослідження - процеси розпізнавання активності людини на основі часових сенсорних даних, отриманих із інертних та мультимодальних сенсорів.

Предмет дослідження - моделі та методи рекурентних нейронних мереж для класифікації та інтерпретації часових сенсорних послідовностей у задачах розпізнавання людської активності.

Для досягнення поставленої мети в роботі вирішувалися такі основні завдання:

- Проаналізувати сучасні підходи, архітектури та методології розпізнавання активності людини на основі сенсорних даних.
- Дослідити принципи роботи та особливості рекурентних нейронних мереж, зокрема моделей LSTM і BLSTM.
- Провести аналіз існуючих наборів даних для задач HAR та визначити релевантні сенсорні модальності для навчання нейронних мереж.
- Розробити архітектуру двоспрямованої LSTM (BLSTM) для класифікації активності людини на основі часових рядів.
- Реалізувати методи підготовки, сегментації та нормалізації сенсорних даних для підвищення ефективності навчання.
- Здійснити порівняльний аналіз результатів із традиційними методами машинного навчання.

Методи дослідження

У роботі використано сукупність теоретичних, аналітичних та експериментальних методів, а саме:

- методи глибокого навчання — побудова та навчання моделей LSTM і BLSTM;
- методи статистичного аналізу для оцінки ефективності класифікації;
- алгоритмічні методи оптимізації (Adam, RMSProp) для налаштування параметрів моделі;
- методи обробки сигналів і часових рядів, включно з нормалізацією та фільтрацією даних;
- експериментальні методи — тестування моделей на публічних наборах даних і порівняння з альтернативними підходами.

Наукова новизна отриманих результатів

Розроблено удосконалену архітектуру двоспрямованої рекурентної нейромережі BLSTM, адаптовану для задач розпізнавання активності людини з урахуванням часових кореляцій між сенсорними спостереженнями. Показано, що використання двоспрямованої структури LSTM підвищує точність розпізнавання активності людини на 5–10% у порівнянні з односпрямованими мережами.

Практичне застосування результатів

Результати роботи можуть бути використані під час розроблення:

- інтелектуальних систем моніторингу фізичної активності та стану здоров'я користувачів;
- “розумних” середовищ (smart home, smart city) для автоматичного керування на основі поведінкових патернів;
- контекстно-залежних мобільних застосунків та IoT-рішень;
- систем безпеки для розпізнавання критичних або небезпечних рухів людини.

Структура магістерської роботи. Робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 82 сторінки, і містить 45 рисунків, 3 таблиці, список використаних джерел із 35 найменувань.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИКОРИСТАННЯ РЕКУРЕНТНИХ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ

1.1. Особливості, принцип роботи та класифікація рекурентних нейронних мереж

Рекурентна нейронна мережа (РНМ, Recurrent Neural Network, RNN) — це клас штучних нейронних мереж, спеціально розроблених для обробки послідовних даних або часових рядів, де порядок даних має вирішальне значення. На відміну від традиційних нейронних мереж прямого поширення (Feedforward Networks), РНМ містять петлі (loops) або зворотні зв'язки (recurrence) у своїх зв'язках.

Цей зворотний зв'язок дозволяє інформації з попередніх кроків часу впливати на поточний крок. Фактично, РНМ мають "пам'ять", яка дозволяє їм зберігати інформацію про попередні елементи послідовності під час обробки поточного елемента.

Основний принцип роботи РНМ полягає в наступному. На кожному часовому кроці (t), РНМ:

1. Приймає поточний вхід (X_t).
2. Приймає прихований стан (H_{t-1}) з попереднього кроку.
3. Обчислює новий прихований стан (H_t), який є функцією від X_t та H_{t-1} .

Генерує вихід (Y_t).

Математично прихований стан часто обчислюється за формулою:

$$H_t = \text{act}(W_{hh}H_{t-1} + W_{xh}X_t + b_h)$$

де:

H_t — новий прихований стан.

H_{t-1} — попередній прихований стан (пам'ять).

X_t — поточний вхід.

W_{hh} та W_{xh} — вагові матриці.

act — функція активації (наприклад, tanh).

Класифікацію РНМ можна проводити за архітектурою (типом зв'язків) та за шаблоном послідовності вхід-вихід.

1. Класифікація за шаблоном послідовності (Sequence-to-Sequence Patterns). РНМ можуть мати різну конфігурацію зв'язків між входом і виходом, залежно від завдання.

Таблиця 1.1.

Класифікація РНМ за шаблоном послідовності

Назва	Шаблон	Опис	Приклад застосування
Один-до-одного	Input → Output	Стандартна нейронна мережа без рекурсії (формально не є РНМ).	Класифікація зображень
Один-до-багатьох	Input → Sequence	Один вхід генерує послідовність виходів.	Створення музики, генерація підписів до зображення
Багато-до-одного	Sequence → Output	Вхідна послідовність генерує один вихід.	Класифікація часових рядів (наприклад, HAR), аналіз тональності тексту
Багато-до-багатьох (1)	Sequence → Sequence	Вхідна послідовність генерує вихідну послідовність однакової довжини.	Тегування частин мови, розпізнавання відеокадрів
Багато-до-багатьох (2)	Sequence → Sequence	Вхідна послідовність генерує вихідну послідовність іншої (часто невідомої) довжини (Encoder-Decoder).	Машинний переклад, автоматичне резюмування тексту

2. Класифікація за архітектурою (Види РНМ)

Найбільш поширені та вдосконалені види РНМ, розроблені для вирішення проблеми зникаючого градієнта та кращого захоплення довготривалих залежностей:

- Проста РНМ (Vanilla RNN) - базова архітектура зі зворотним зв'язком. Ефективна для коротких послідовностей, але схильна до проблеми зникаючого/вибухаючого градієнта при роботі з довгими залежностями.

- Довгострокова короткочасна пам'ять (Long Short-Term Memory, LSTM) - введена для вирішення проблеми зникаючого градієнта. Використовує складну внутрішню комірку пам'яті (cell state) та три шлюзи (забуваючий, вхідний, вихідний), які контролюють потік інформації та дозволяють зберігати релевантні дані протягом тривалого часу. Дуже ефективна для моделювання довготривалих залежностей.

- Вентильований рекурентний блок (Gated Recurrent Unit, GRU) – це спрощена версія LSTM. Використовує лише два шлюзи (шлюз оновлення та шлюз скидання), не має окремої комірки пам'яті. Є менш обчислювально витратна та швидша, часто демонструє порівнянну продуктивність з LSTM.

- Двоспрямована РНМ (Bidirectional RNN, BRNN) - може використовувати базові РНМ, LSTM або GRU. Складається з двох незалежних шарів, які обробляють послідовність у прямому та зворотному напрямках. Дозволяє моделі захоплювати контекст як із минулого, так і з майбутнього послідовності, що критично важливо для задач, де кінцевий результат впливає на інтерпретацію попередніх кроків (наприклад, BLSTM для NAR).

- Глибока РНМ (Deep RNN) - складається з кількох шарів рекурентних блоків, розташованих вертикально (один над одним), що дозволяє моделі вивчати ієрархічні представлення даних.

Довгострокова короткочасна пам'ять (LSTM) та вентильований рекурентний блок (GRU) є двома найпоширенішими та найефективнішими типами вентильованих рекурентних нейронних мереж. Обидві архітектури були розроблені для вирішення ключової проблеми зникаючого градієнта у традиційних РНМ, що дозволяє їм ефективно моделювати довготривалі залежності у послідовних даних.

На рис. 1.1 подано архітектуру основного блоку рекурентної нейронної мережі, розгорнутої в часі, яке відповідає опису її принципу роботи.

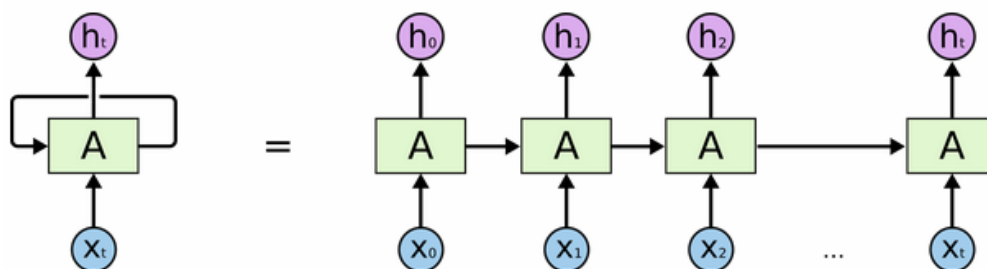


Рис. 1.1. Розгортання рекурентного зв'язку РНМ вздовж часової осі

Ця діаграма чітко ілюструє, як інформація (у вигляді прихованого стану H) передається від одного часового кроку до наступного, забезпечуючи "пам'ять" мережі для обробки послідовностей.

1.2. Методологія розпізнавання людської активності на основі сенсорів із застосуванням LSTM моделі

Розпізнавання людської активності (Human Activity Recognition, HAR) за допомогою методів глибокого навчання є критично важливим напрямом досліджень із значним потенціалом застосування у широкому спектрі галузей, включаючи спортивну аналітику, моніторинг пересування, системи спостереження, охорону здоров'я (зокрема, віддалений моніторинг пацієнтів та реабілітація) та мобільну робототехніку. Для сенсорно-орієнтованого HAR часто використовуються інертні сенсори (Inertial Measurement Units, IMU), які зазвичай складаються з акселерометрів та гіроскопів. Ці натільні пристрої забезпечують часові ряди даних, що відображають динаміку рухів тіла.

Це дослідження зосереджено на розробці та оцінці підходу для точного розпізнавання та класифікації близьких за активністю дій (activities with subtle differences) на основі даних, отриманих від натільних інертних

сенсорів. Для реалізації поставленої задачі застосовано архітектуру двоспрямованої довгострокової пам'яті (Bidirectional Long Short-Term Memory, BLSTM). Вибір BLSTM зумовлений її здатністю ефективно обробляти послідовні дані, враховуючи як попередній, так і наступний контекст у часовому ряді, що є особливо корисним для диференціації тонких відмінностей між схожими активностями.

Експериментальні дані були взяті з загальнодоступного набору даних UTD-MHAD (The University of Texas at Dallas Multimodal Human Activity Dataset). Цей набір даних містить інформацію про широкий спектр людських активностей, виконаних різними суб'єктами.

Модель BLSTM, розроблена в межах цього дослідження, продемонструвала високу ефективність у завданні HAR:

- Для підмножини з 15 різних активностей досягнуто загальної точності класифікації (overall accuracy) 98,05%.

- Для повного набору з 27 різних активностей, виконаних 8 особами з 4 спробами на кожну активність на особу, модель досягла загальної точності 90,87%.

Було проведено порівняльний аналіз продуктивності запропонованої моделі BLSTM з односпрямованою моделлю LSTM (Unidirectional LSTM). Результати порівняння засвідчили значне покращення точності розпізнавання всіх 27 активностей у випадку застосування архітектури BLSTM порівняно зі стандартною LSTM. Це підтверджує гіпотезу про перевагу двоспрямованої обробки часових рядів для задач, де контекст у майбутньому часі є інформативним для класифікації поточного моменту активності, особливо для дій, що мають схожі початкові чи кінцеві фази.

Отримані результати підтверджують високу ефективність архітектури BLSTM для точного та надійного розпізнавання комплексу людських активностей, зокрема близьких за природою, на основі даних, зібраних із натільних інертних сенсорів. Досягнуті показники точності є

конкурентоспроможними та свідчать про потенціал впровадження цієї методології в реальні системи моніторингу активності.

1.3. Опис архітектури BLSTM (Bidirectional Long Short-Term Memory)

1.3.1. Концепція LSTM

Довгострокова короткочасна пам'ять (LSTM) — це різновид рекурентних нейронних мереж (РНМ), спеціально розроблений для моделювання довготривалих залежностей у послідовних даних (таких як часові ряди сенсорів або текст).

На відміну від класичних РНМ, які страждають від проблеми зникаючого градієнта, блок LSTM має внутрішню структуру, що складається з комірки пам'яті (cell state) та трьох основних шлюзів (gates):

1. Забуваючий шлюз (Forget Gate) - визначає, яку частину інформації з попереднього стану комірки (C_{t-1}) слід ігнорувати або "забути".
2. Вхідний шлюз (Input Gate) - визначає, яка нова інформація (X_t) буде записана у поточний стан комірки (C_t).
3. Вихідний шлюз (Output Gate) - контролює, яку частину інформації з поточного стану комірки (C_t) слід вивести як прихований стан (h_t) для наступного кроку.

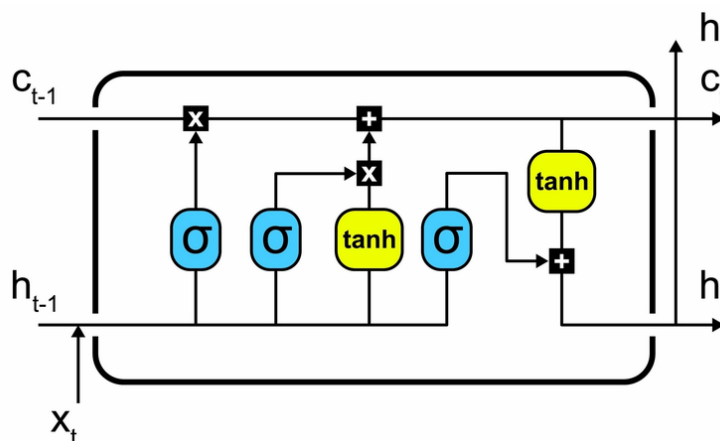


Рис. 1.2. Внутрішня архітектура та логіку роботи однієї комірки LSTM

На рис. 1.2 комірка LSTM деталізовано показує механізм, який дозволяє їй зберігати та керувати інформацією протягом тривалого часу:

- Лінія комірки (Cell State, C_t) - горизонтальна лінія, що проходить через верхню частину схеми. Вона несе інформацію через усю послідовність, забезпечуючи довготривалу пам'ять.

- Шлюзи (Gates) - чотири взаємодіючі нейронні шари (сігмоїдні σ та \tanh), які контролюють потік інформації.

1.3.2. Принцип BLSTM

Двоспрямована LSTM (BLSTM) являє собою розширення стандартної LSTM, яке значно підвищує її здатність до контекстуального розуміння послідовності, особливо у випадках, коли для класифікації поточного моменту (кроку t) потрібна інформація не лише з минулого, а й з майбутнього.

Архітектура BLSTM складається з двох незалежних шарів LSTM, які обробляють вхідну послідовність у протилежних напрямках:

- Прямий шар (Forward Layer): Обробляє вхідну послідовність від початку до кінця ($t=1 \rightarrow T$). Він захоплює інформацію про минулий контекст.

- Зворотний шар (Backward Layer): Обробляє ту ж вхідну послідовність від кінця до початку ($t=T \rightarrow 1$). Він захоплює інформацію про майбутній контекст.

На кожному часовому кроці (t) вихідний прихований стан BLSTM (H_t) формується шляхом конкатенації (об'єднання) вихідних даних обох шарів:

$$H_t = [\vec{H}_t; \overleftarrow{H}_t]$$

де:

H_t^{\rightarrow} — прихований стан прямого шару.

H_t^{\leftarrow} — прихований стан зворотного шару.

Графічно архітектура BLSTM може бути представлена як два дзеркальні шари LSTM, які працюють паралельно над однією вхідною послідовністю.

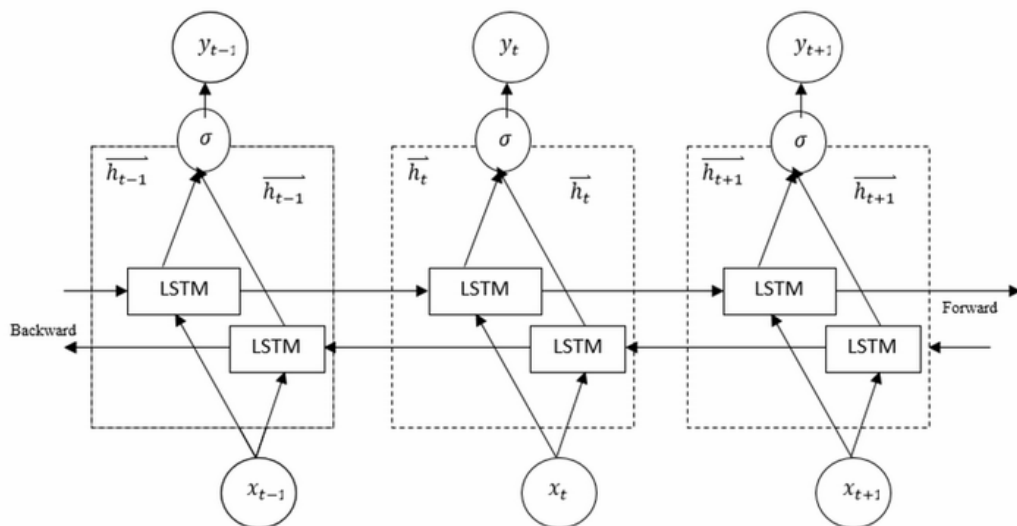


Рис. 1.3. Графічне представлення архітектури BLSTM

На рисунку 1.3:

- Верхній шар (Straight Arrow) - це прямий LSTM-шар (H_t), який обробляє послідовність у звичайному порядку, захоплюючи минулий контекст.

- Нижній шар (Reversed Arrow) - це зворотний LSTM-шар (H_t), який обробляє ту саму послідовність у зворотному напрямку, захоплюючи майбутній контекст.

- X_t - Вхідні дані (вектор сенсорів).

- Y_t - Кінцевий вихід, що утворюється шляхом об'єднання (конкатенації) вихідних даних обох шарів для класифікації.

1.3.3. Переваги в процесах розпізнавання людської активності

У контексті розпізнавання людської активності (HAR) на основі сенсорів, BLSTM є надзвичайно ефективною, оскільки:

- для класифікації активності, наприклад, "встати зі стільця" (класифікація відбувається в момент t), важливо знати, що цьому передувало ("сидів", минулий контекст) і що буде далі ("початок ходьби", майбутній контекст). Стандартна LSTM врахує лише "сидів", тоді як BLSTM врахує обидва аспекти.

- здатність поєднувати інформацію з обох напрямків дозволяє моделі краще розрізняти активності, які мають схожі початкові або кінцеві рухи.

1.4. Аналітичний огляд методології розпізнавання людської активності на основі глибокого навчання

Розпізнавання людської активності (Human Activity Recognition, HAR) є ключовою науково-технічною дисципліною, яка знаходить значуще застосування у диверсифікованому спектрі галузей. До них належать охорона здоров'я (зокрема, моніторинг стану пацієнтів та реабілітація), робототехніка, системи спостереження та відстеження руху, а також спортивна аналітика. Протягом багатьох років активно розвивається напрямок сенсорно-орієнтованого HAR. Завдяки інтенсивному технологічному прогресу, ефективність та точність сенсорного розпізнавання активності постійно підвищуються. Сучасні моделі глибокого навчання відкривають значні можливості для подальшої оптимізації цього процесу. Ці моделі мають фундаментальну перевагу, оскільки вони мінімізують необхідність ручного виділення ознак, що є трудомістким процесом у традиційних підходах. Вони здатні значно ефективніше розпізнавати складні та близькі за активністю дії порівняно з класичними алгоритмами машинного навчання.

1.4.1. Дані та сенсорна модальність

Основним джерелом даних для сенсорно-орієнтованого HAR є інертні сенсори (Inertial Measurement Units, IMU), що носяться на тілі.

Кожен IMU складається з:

- Тривісного акселерометра, що фіксує лінійне прискорення по осях x , y та z .

- Тривісного гіроскопа, що фіксує кутову швидкість обертання по осях x , y та z .

Ці сенсори генерують послідовності даних руху у форматі часового ряду. У цьому дослідженні використано дані інертних сенсорів, отримані з відкритого набору даних UTD-MHAD [5]. Цей набір даних є особливо цінним, оскільки включає групи близьких за активністю дій. Приклади таких дій включають:

- Схожі рухи верхньої кінцівки, зареєстровані сенсором на правій руці: замах бейсбольною биткою та замах тенісною ракеткою.

- Схожі рухи нижньої кінцівки, зареєстровані сенсором на правому стегні: випад та присідання.

1.4.2. Розпізнавання активності як задача класифікації

Мотивація дослідження зумовлена зростаючим попитом на економічно ефективні та високоточні методи розпізнавання людської активності для широкого спектру застосувань — від моніторингу літніх пацієнтів до детального аналізу рухів спортсменів. Хоча базові активності були успішно розпізнані раніше за допомогою глибокого навчання, розпізнавання складних та близьких за активністю дій залишається науково-технічним викликом.

Розпізнавання активності на основі натільних сенсорів, по суті, є задачею класифікації часових рядів.

Ключовий внесок цього проєкту полягає у наступному:

- Застосування вдосконаленого варіанта архітектури рекурентних нейронних мереж — двоспрямованої моделі довгострокової короткочасної пам'яті (BLSTM) — для задачі HAR.

- Фокус на розпізнаванні активностей, які є близькими одна до одної (тобто, мають схожість у часовій послідовності сенсорних даних).

Оцінка ефективності моделі на більшому пулі активностей (27 різних дій).

Передбачається, що модель BLSTM буде навчена, а її результати будуть ретельно проаналізовані та порівняні з показниками односпрямованої моделі довгострокової пам'яті (LSTM).

1.4.3. Пропонована модель для розпізнавання активності

Розглянемо модель для розпізнавання активності людини (Human Activity Recognition, HAR) за допомогою сенсорних даних та глибокого навчання, що подана на рис. 1.4:

1. Збір та підготовка даних

Дані збираються від інерційних сенсорів (inertial sensor), прикріплених до тіла (зокрема, до правого зап'ястя та правого стегна).

Кожен сенсор надає:

- 3 сигнали акселерометра - вимірювання прискорення по осях x,y,z.
- 3 сигнали гіроскопа - вимірювання кутової швидкості по осях x,y,z.

Загалом 6 сигналів з кожного місця, що дає 12 сигналів (або 6, якщо сенсор лише один).

Sliding Window - ,неперервні сигнали розбиваються на короткі сегменти (вікна), щоб підготувати їх для нейронної мережі. Наприклад, розмір вікна $T=128$ з певним перекриттям (Overlap).

2. Алгоритм глибокого навчання

Двонаправлена довга короткочасна пам'ять (Bi-directional Long Short-Term Memory, BLSTM) - це тип рекурентної нейронної мережі (RNN), який добре підходить для обробки послідовних даних (як-от часові ряди сенсорних сигналів).

Вхід (Input): оброблене вікно сенсорних даних.

Прямий шар (Forward Layer) - обробляє послідовність у звичайному порядку ($t_0 \rightarrow t_1 \rightarrow \dots$).

Зворотний шар (Backward Layer) - обробляє ту ж послідовність у зворотному порядку ($\dots \leftarrow tN-1 \leftarrow tN$).

Вихід (Output) - з'єднані виходи обох шарів для кожного часового кроку ($O_1, O_2, O_3, O_4, \dots$).

Перевагою є те, що BLSTM може враховувати як минулий, так і майбутній контекст у часовому ряді для кращого розпізнавання поточної активності.

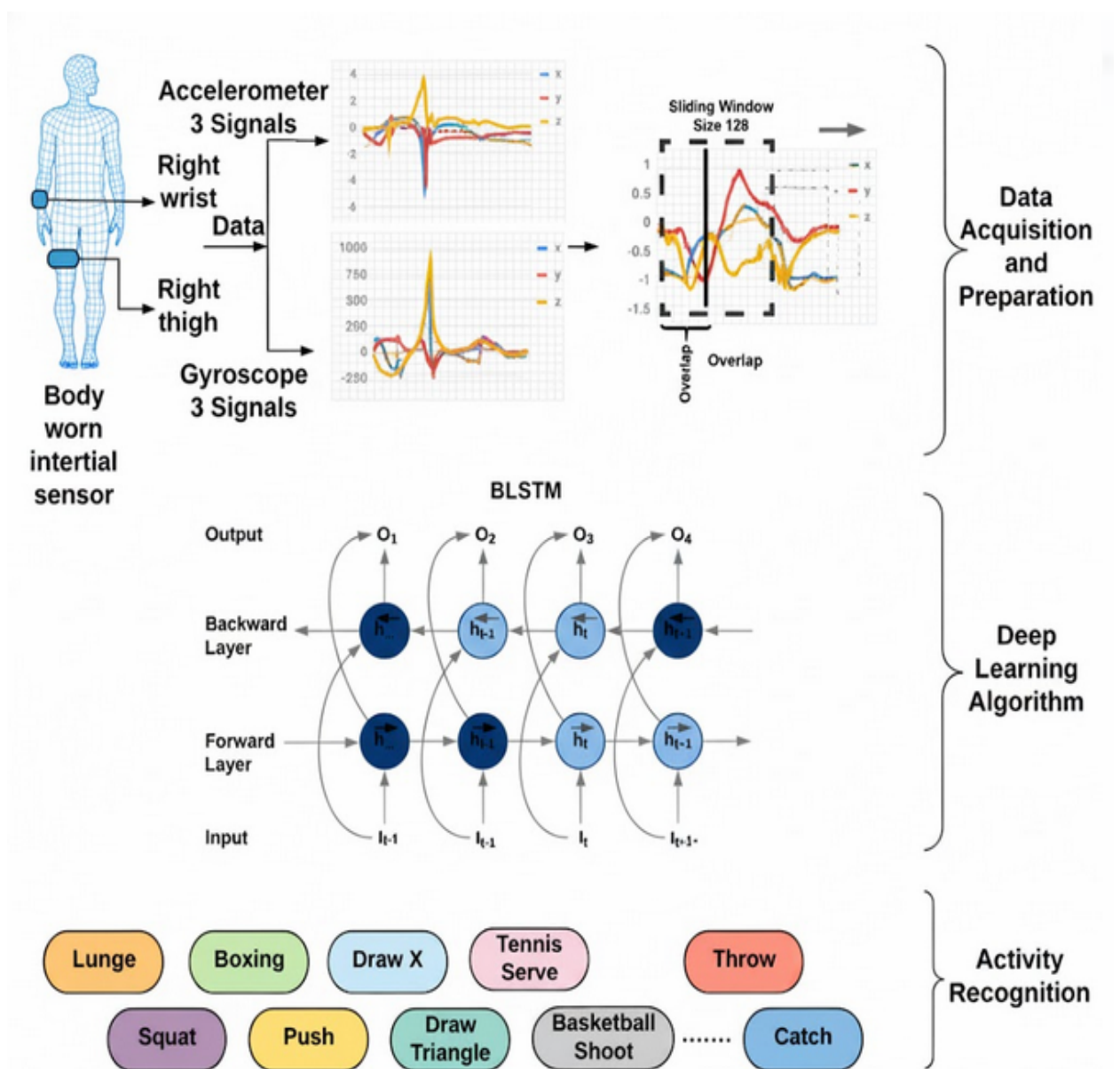


Рис. 1.4. Огляд запропонованої моделі для розпізнавання активності

3. Розпізнавання активності

Вихід BLSTM подається на останній шар (наприклад, Softmax), який класифікує вхідні сенсорні дані у одну з попередньо визначених категорій активності.

Модель навчена розпізнавати широкий спектр рухів:

- Випад (Lunge)
- Присідання (Squat)
- Бокс (Boxing)
- Поштовх (Push)
- Тенісна подача (Tennis Serve)
- Баскетбольний кидок (Basketball Shoot)
- Кидок (Throw)
- Ловіння (Catch)

Отже, представлена вище архітектура використовує інерційні сенсори для збору даних, ковзне вікно для сегментації, а мережу BLSTM для аналізу часових послідовностей і прогнозування того, яку активність виконує людина.

Висновки до розділу

У першому розділі здійснено аналітичний огляд предметної області використання рекурентних нейронних мереж у задачах розпізнавання людської активності. Розглянуто базові принципи роботи RNN, їхню класифікацію та специфічні властивості, що дозволяють ефективно моделювати часові залежності між сенсорними спостереженнями. Особливу увагу приділено архітектурам LSTM і BLSTM, які вирішують проблему затухання градієнта та забезпечують запам'ятовування довгострокових залежностей у послідовностях даних.

Охарактеризовано концепцію двоспрямованої пам'яті (BLSTM), що дозволяє враховувати контекст як з минулого, так і з майбутнього часових

кроків, підвищуючи точність розпізнавання складних видів активностей. Також проведено аналітичний огляд методологій HAR на основі глибоких нейронних мереж і сенсорних модальностей, зокрема акселерометрів, гіроскопів та мультимодальних датчиків. Доведено, що використання LSTM-архітектур у поєднанні з релевантними сенсорними даними дозволяє досягати вищих показників класифікаційної точності порівняно з традиційними методами машинного навчання.

Таким чином, у розділі сформовано теоретичну основу для побудови моделі рекурентної нейромережі, оптимізованої для задач розпізнавання активності людини

РОЗДІЛ 2. ДОСЛІДЖЕННЯ МОДЕЛЕЙ ТА МЕТОДІВ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ НА ОСНОВІ РЕЛЕВАНТНИХ СЕНСОРНИХ ДАНИХ

2.1. Класифікація методів розпізнавання людської активності та деталізація сенсорно-орієнтованих підходів

Більшість існуючих методологій розпізнавання людської активності можуть бути категоризовані за двома основними напрямками:

- Розпізнавання активності на основі відео (Vision-based HAR) - також відоме як розпізнавання активності на основі зору [4,6]. Цей підхід використовує візуальні дані, отримані за допомогою відеокамер. Згенеровані дані зазвичай мають формат відеопослідовностей із визначеною частотою кадрів та/або даних глибини.

- Розпізнавання активності на основі сенсорів (Sensor-based HAR) - цей підхід використовує дані руху, зібрані різноманітними сенсорами, такими як акселерометри, гіроскопи, магнітометри, модулі Bluetooth, GPS-приймачі тощо.

Згенеровані дані є часовими рядами, які можуть бути аналізовані як у часовій, так і у частотній області. Сенсори, у свою чергу, поділяються на чотири широкі категорії [4]:

а) натільні сенсори (wearable sensors) - прикріплюються безпосередньо до людського тіла.

б) об'єктні сенсори (object sensors) - інтегровані в об'єкти навколишнього середовища.

в) навколишні сенсори (ambient sensors) - розміщуються у навколишньому середовищі для моніторингу активності.

г) гібридні сенсори (Hybrid Sensors) - комбінації вищезазначених типів.

2.1.1. Особливості інертних сенсорів

Інертні сенсори, що носяться на тілі (Wearable Inertial Sensors), є найпоширенішою категорією сенсорів, що використовуються для HAR. Їхня широка поширеність зумовлена значним розвитком носібельних обчислювальних пристроїв та доступністю недорогих, компактних інертних сенсорів. Ці сенсори, як правило, включають тривісний акселерометр та тривісний гіроскоп, а іноді й магнітометр. Вони прикріплюються до специфічних частин тіла, таких як кінцівки або тулуб, для реєстрації людських рухів. Портативність та компактність цих сенсорів роблять їх ідеальними для інтеграції у повсякденний одяг або аксесуари з метою безперервного збору даних руху.

У поточному проєкті для збору даних використовуються тривісні інертні сенсори, що носяться на тілі (акселерометр та гіроскоп), прикріплені до правого зап'ястя та правого стегна 8 різних суб'єктів. Типові позиції розміщення цих сенсорів на тілі проілюстровані на рис. 2.1.

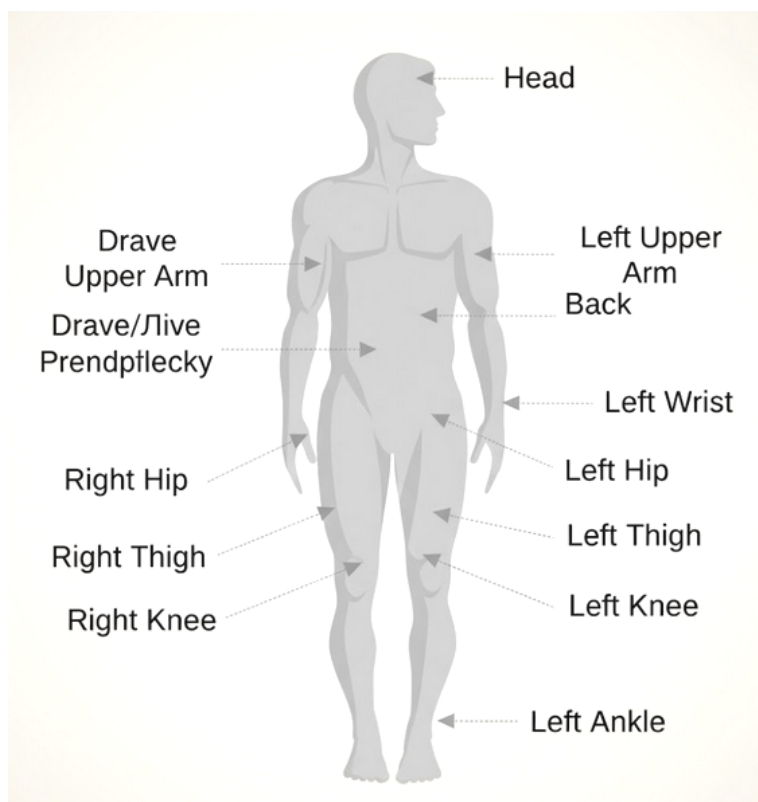


Рис. 2.1. Типові позиції розміщення інертних сенсорів

2.1.2. Опис набору даних UTD-MHAD

Ключовою особливістю UTD-MHAD є те, що він містить чотири часово синхронізовані модальності даних, зібрані одночасно:

1. Відео RGB - стандартні кольорові відеопослідовності.
2. Відео глибини (Depth Videos) - дані, отримані з камери глибини (Kinect), які надають інформацію про відстань до об'єктів у сцені.
3. Позиції скелета (Skeleton Positions) - дані про 20 ключових суглобів людського тіла, відстежених камерою Kinect.
4. Інертні сигнали (Inertial Signals) - дані, отримані з натільних інертних сенсорів (IMU), що складаються з тривісного акселерометра та тривісного гіроскопа.

Набір охоплює 27 різних людських активностей. Ці активності включають як прості базові рухи, так і складні, а також близькі за активністю дії (наприклад, "замах бейсбольною биткою" та "замах тенісною ракеткою"). Дані зібрані від 8 різних осіб і кожен суб'єкт виконав кожен активність 4 рази (спроби). Загальна кількість відео- та сенсорних послідовностей становить 27 активностей \times 8 осіб \times 4 спроби = 864 записи, кожен з яких містить чотири синхронізовані модальності.

UTD-MHAD має велике значення для досліджень оскільки дозволяє вивчати та порівнювати методи злиття інформації з різних сенсорних модальностей (наприклад, поєднання візуальних даних із даними руху IMU) для підвищення точності розпізнавання. Набір надає стандартизовану основу для тестування алгоритмів, що використовують виключно натільні інертні сенсори, як у даному проєкті з BLSTM.

Наявність близькоспоріднених активностей (таких як спортивні замахи або схожі рухи кінцівок) робить набір даних складним і реалістичним для оцінки здатності моделей глибокого навчання до тонкої диференціації рухів. UTD-MHAD є цінним ресурсом, що вільно розповсюджується, для досліджень мультимодальності та розробки надійних систем розпізнавання людської активності.

Рисунки 2.2 і 2.3 демонструють характерні покази акселерометра та гіроскопа для двох близьких за активністю дій з набору даних UTD-MHAD - замаху бейсбольною биткою та замаху тенісною ракеткою відповідно.

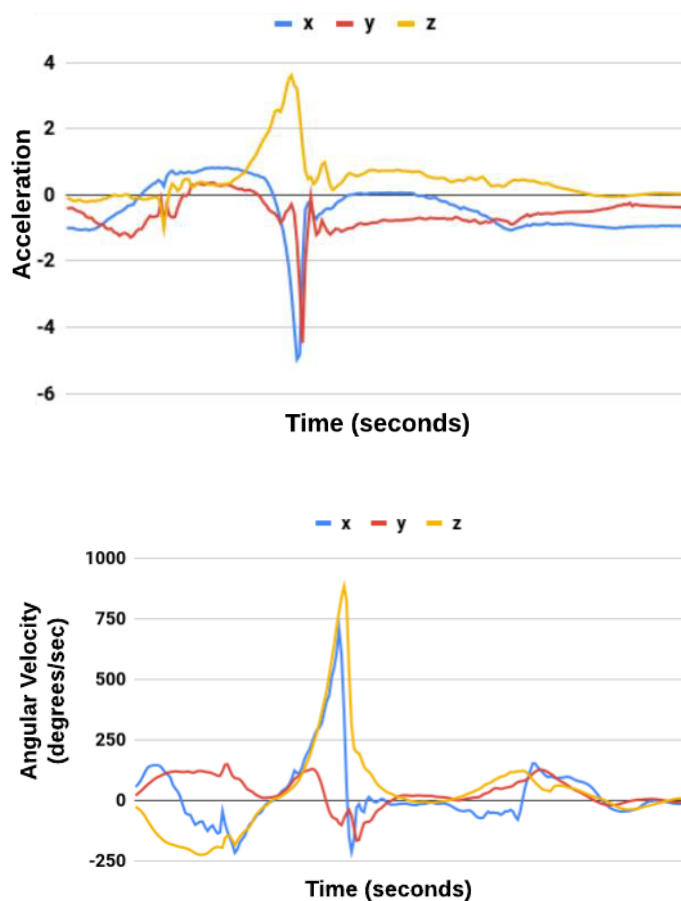
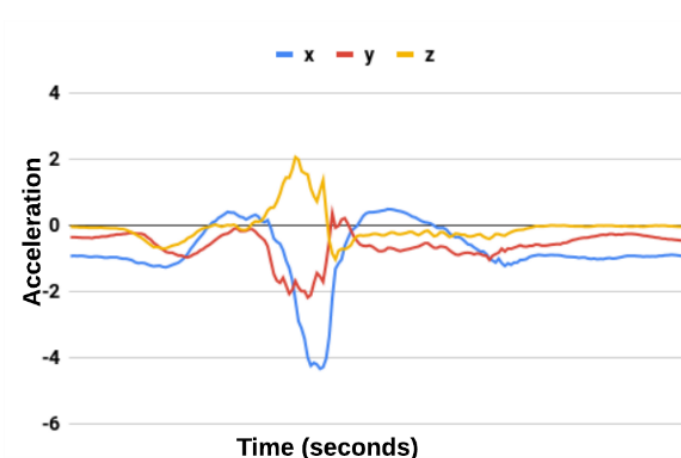


Рис. 2.2. Покази акселерометра та гіроскопа для замаху бейсбольною биткою



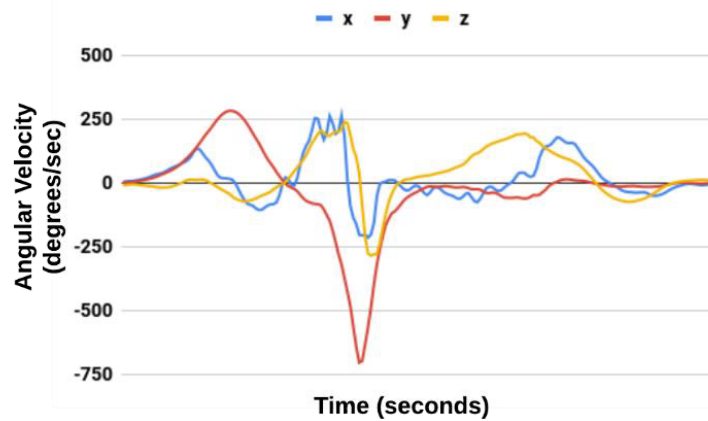


Рис. 2.3. Покази акселерометра та гіроскопа для замаху тенісною ракеткою

Рисунки 2.4 і 2.5 які надають деталізоване розуміння послідовностей руху покликані візуалізувати динаміку цих близьких за активністю дій. Метою цих зображень є виділення та порівняння схожих послідовностей рухів для активностей, таких як замах бейсбольною биткою та замах тенісною ракеткою, використовуючи кольорове кодування.

Важливо зазначити, що значення прискорення та/або кутової швидкості в цих виділених послідовностях є дуже близькими, що підкреслює складність їх диференціації.

Рисунки 2.4 і 2.5 взяті з набору даних UTD-MHAD.

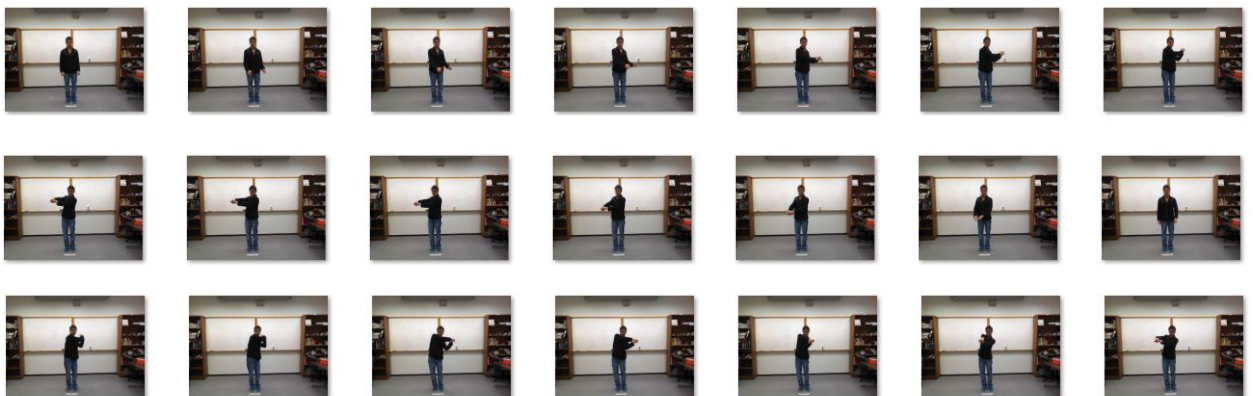


Рис. 2.4. Зображення послідовності рухів із набору даних для замаху бейсбольною биткою

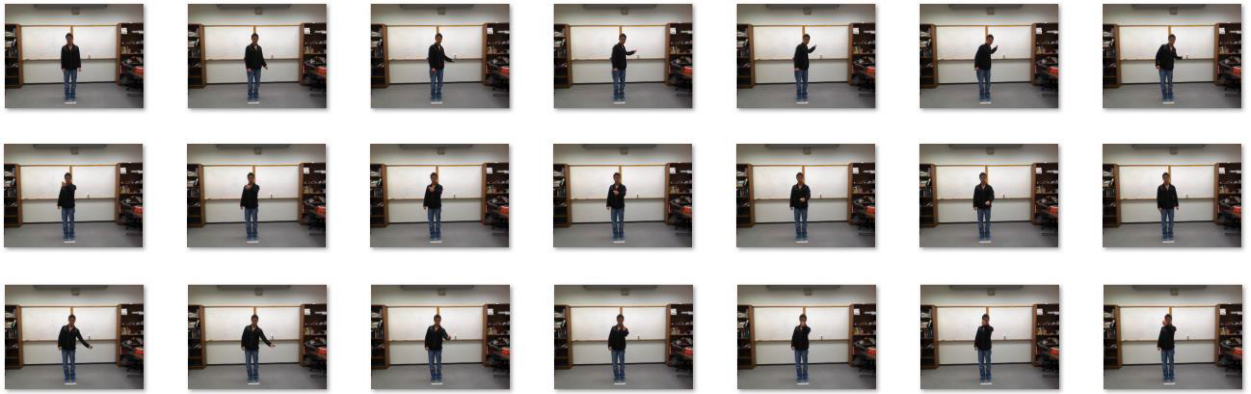


Рис. 2.5. Зображення послідовності рухів із набору даних для замаху тенісною ракеткою

2.2. Еволюція методологій розпізнавання людської активності

2.2.1. Обмеження традиційних підходів

Традиційні методи розпізнавання активності вимагають евристичного, ручного виділення ознак (*hand-crafted feature extraction*) та глибоких експертних знань про початкові сенсорні дані. Ці ознаки, необхідні для подальшого використання у статистичних моделях або класичних моделях машинного навчання, можуть бути отримані з часової області (наприклад, дисперсія, середнє значення) або з частотної області (наприклад, розподіл енергії сигналу, амплітудні спектри).

Оскільки виділені ознаки базуються переважно на математичних операціях, а не на контексті складного руху, процес виявлення відповідних ознак стає трудомістким і значно ускладнює розпізнавання складних, тонко диференційованих активностей [1]. Така залежність від експертних знань та ручного інжинірингу ознак обмежує як точність, так і загальність традиційних систем HAR.

2.2.2. Парадигма глибокого навчання

Саме в цьому контексті глибоке навчання (*Deep Learning, DL*) довело свою перевагу для розпізнавання активності [13]. Моделі глибокого навчання

здатні автоматично навчатися ієрархічним ознакам (feature learning) безпосередньо з сирих сенсорних даних.

Ключові переваги DL-підходів:

1. Автоматичне виділення ознак.

Усувається необхідність у трудомісткому ручному інжинірингу ознак, що підвищує гнучкість моделі.

2. Адаптивність до даних.

Моделі ефективно використовують великі та гетерогенні набори даних, включаючи комбінації з двох або більше сенсорних модальностей.

3. Високорівневі ознаки.

Здатність вивчати високорівневі, абстрактні ознаки, що є критично важливим для точного розпізнавання складних або схожих активностей.

2.3. Огляд пов'язаних досліджень та методологій

Проведено значний обсяг досліджень, що стосуються HAR, із застосуванням як візуальних, так і сенсорно-орієнтованих підходів. Серед використаних методологій машинного та глибокого навчання можна виділити наступні:

2.3.1. Класичні методи машинного навчання

Машина опорних векторів (Support Vector Machine, SVM) є дискримінаційним класифікатором, який визначає оптимальну гіперплощину для розділення простору зразків на категорії. Наприклад, в роботі [7] запропонували апаратно-орієнтовану багатокласову класифікацію на смартфонах з використанням SVM з класифікацією з фіксованою точкою.

Експерименти були проведені за участю групи з 30 волонтерів віком від 19 до 48 років. Кожна особа виконувала шість активностей, маючи смартфон закріпленим на поясі. Для спрощення процесу розмітки даних експерименти були записані на відео.

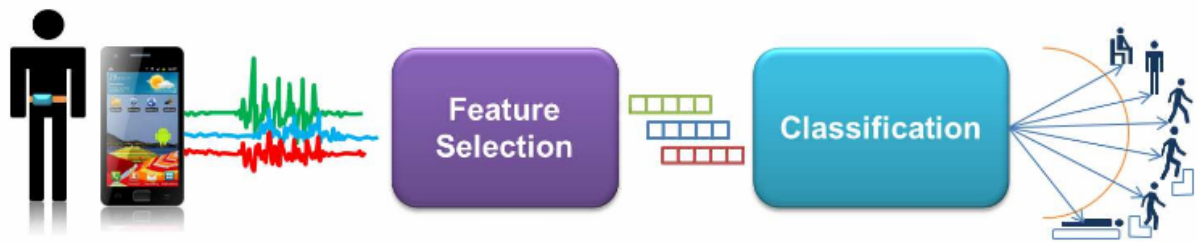


Рис. 2.6. Конвеєрний процес розпізнавання активності

Отримана база даних була випадковим чином розділена на дві частини: 70% зразків використано для навчання та 30% як тестові дані. Набір для навчання використовується для тренування багатокласового класифікатора SVM (Машина опорних векторів), описаного в наступному розділі.

Для експериментів використовувався смартфон Samsung Galaxy S2, оскільки він оснащений акселерометром та гіроскопом для вимірювання 3-осьового лінійного прискорення та кутової швидкості відповідно, з постійною частотою дискретизації 50 Гц, що є достатнім для захоплення рухів людського тіла.

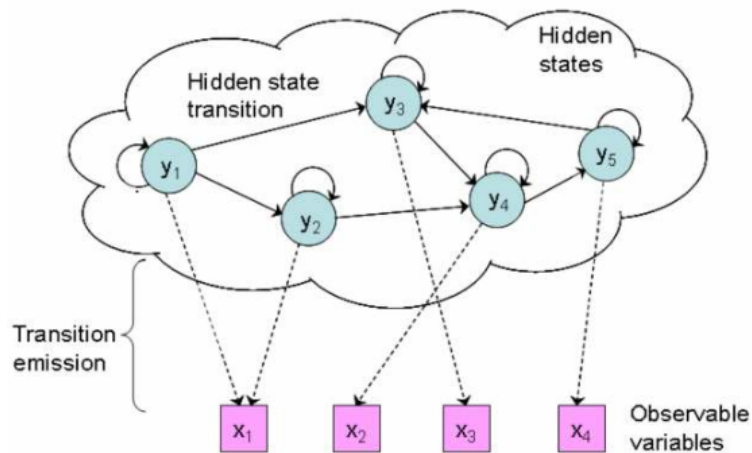
Для цілей розпізнавання активності (AR) було розроблено додаток для смартфона на базі операційної системи Google Android. Процес розпізнавання починається зі збору сенсорних сигналів, які потім проходять попередню обробку із застосуванням шумових фільтрів. Далі сигнали сегментуються за допомогою ковзних вікон фіксованої ширини 2,56 секунди із перекриттям 50%.

З кожного вікна формується вектор із 17 ознак шляхом розрахунку змінних з сигналів акселерометра в часовій та частотній областях (наприклад, середнє значення, стандартне відхилення, площа магнітуди сигналу, ентропія, кореляція пар сигналів тощо). Для знаходження частотних компонентів сигналу використовується швидке перетворення Фур'є (Fast Fourier Transform, FFT). Зрештою, ці зразки подаються на вхід навченого класифікатора SVM для розпізнавання активностей. Весь конвеєрний процес розпізнавання активності зображено на рис. 2.6.

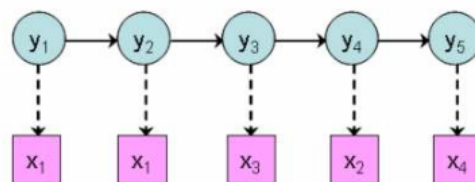
В роботі [3] використовували онлайн SVM у поєднанні з СТ-РСА, розробляючи систему, стійку до варіацій розміщення сенсорів та орієнтації.

Приховані марковські моделі (Hidden Markov Models, HMM) - створюють імовірнісний опис простору даних, де переходи між прихованими станами регулюються ймовірностями переходів. HMM використовувалися для HAR, наприклад, в роботі [20] застосували HMM для порівняльного аналізу одночасного та переміжного розпізнавання активності.

Прості активності можна з високою точністю моделювати як марковські ланцюги (Markov Chains). Однак, складні або нетипові (незнайомі) активності часто важко зрозуміти та відповідно змоделювати. Наприклад, досліднику, який вивчає повсякденну активність особи з деменцією, буде важко підібрати коректну модель, якщо він не є експертом у галузі поведінкових наук, пов'язаних із деменцією.



(a) A graphical representation of an HMM



(b) An observation sequence of an HMM

Legend: —→ State transition - - - - -> Transition emission

Рис. 2.7. Графічна інтерпретація Hidden Markov Models

Але для непрямого побудування моделі активності можна використати спостереження за сигналами, що виникають внаслідок таких складних чи нетипових дій. Для цього використовується прихована марковська модель. Спостерігаючи за наслідками або ефектами активності, НММ здатна поступово конструювати модель активності, яку потім можна допрацьовувати, розширювати та повторно використовувати в подібних дослідженнях.

НММ — це генеративна імовірнісна модель, яка використовується для генерації прихованих станів на основі спостережуваних даних. Зокрема, основна мета цієї моделі полягає у визначенні послідовності прихованих станів ($y_1 y_2 \dots y_t$), що відповідає спостережуваній послідовності виходів (x_1, x_2, \dots, x_t). Ще однією важливою метою є надійне навчання параметрів моделі на основі історії спостережуваних послідовностей виходів.

Рисунок 2.7 демонструє графічне представлення НММ, що складається з 5 прихованих станів та 4 спостережуваних змінних.

Альтернативний підхід до виявлення шаблонів активності полягає у візуальному спостереженні за діями та виділенні індивідуальних поз (individual poses) з відеоданих. Активності потім можуть бути представлені шляхом конструювання ймовірнісних контекстно-вільних граматики (Probabilistic Context-Free Grammars, PCFG), використовуючи пози як алфавіт граматики [5]. Наприклад, рисунок 2.8 демонструє зразки поз, які відповідають ключовим кадрам відеопослідовностей.

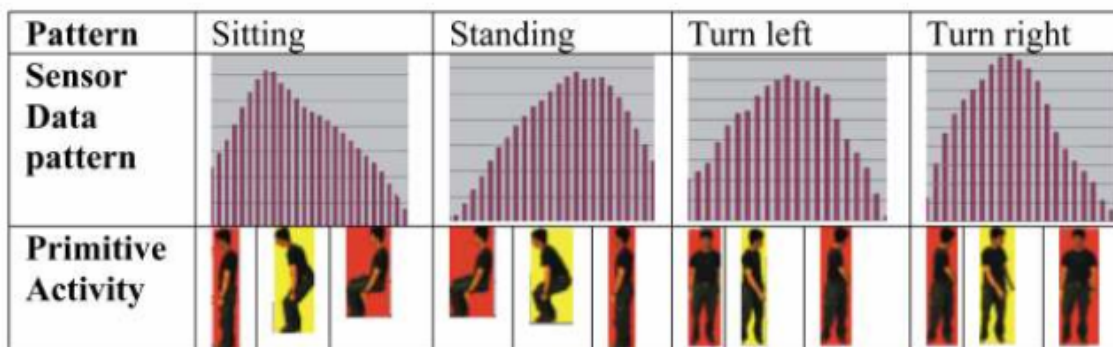


Рис. 2.8. Виділення шаблонів руху з даних сенсорів

Наступним кроком є виділення специфічних класів правил із даних, які відображають повторювані послідовності поз та їхні складніші комбінації, як показано на рисунку 2.9.

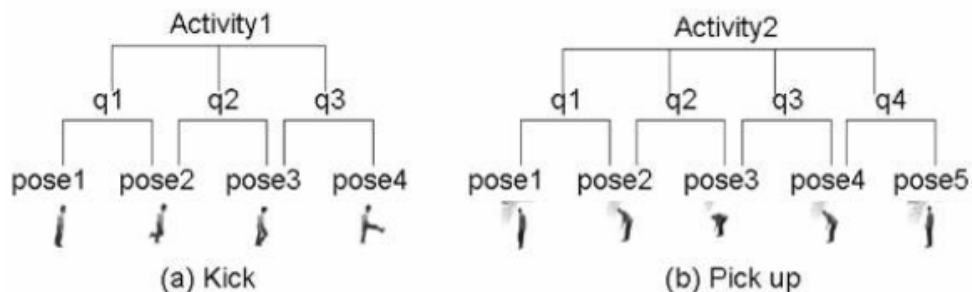


Рис. 2.9. Складові активності

Наприклад, дія "удар ногою" (kick) складається з трьох пар поз, позначених як q1 до q3 на рис. 2.9 (ліворуч). Якщо активність "удар ногою" комбінується з розпізнаним об'єктом, наприклад, "футбольним м'ячем", ця комбінація відповідає більш специфічній активності, як-от "грати у футбол".

Окрім кластеризації кількох примітивних активностей, їх комбінують з об'єктами для побудови більш специфічних складових активностей (composite activities). Наприклад, якщо активність "удар ногою" поєднується з об'єктом "футбольний м'яч", це генерує більш специфічну активність "грати у футбол".

Після виявлення активностей вони можуть стати основою для моделі, що дозволяє розпізнавати активність, відстежувати її прояви та навіть використовувати цю інформацію для оцінки самопочуття індивіда або надання контекстно-залежних послуг (activity-aware services). Таким чином, ці технології виявлення та розпізнавання активності є цінними для забезпечення повсюдної допомоги (pervasive assistance) в повсякденному середовищі людини.

В дослідженні [2] реалізували дискретну НММ на послідовностях поз. Реалізація системи складається з трьох частин, як показано на рис. 2.10. У цій роботі виходять з припущення про статичний фон (static background). Це

дозволяє нам виконати віднімання фону (background subtraction) шляхом порогової обробки різниці між поточним кадром і зображенням статичного фону.

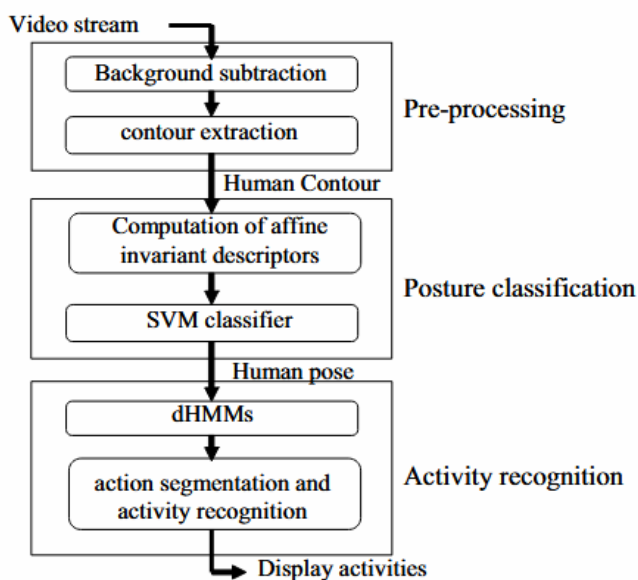


Рис. 2.10. Ілюстрація етапів обробки системи

Після віднімання фону ми отримуємо силует людини (human silhouette), і на фінальній фазі попередньої обробки ми виділяємо контур (contour) з цього силуету. У процесі класифікації пози (posture classification) спочатку обчислюють афінно-інваріантні дескриптори Фур'є (affine invariant Fourier descriptors) з контуру. Потім ці дескриптори використовуються як вектор ознак (feature vector) для класифікації пози за допомогою SVM з радіально-базисною функцією (Radial Basis SVM, RBF SVM).

Таким чином, виходом модуля класифікації пози є послідовність дискретних поз (sequence of discrete postures).

2.3.2. Методи глибокого навчання та гібридні архітектури

Глибоке навчання відкрило нові можливості завдяки використанню наступних архітектур.

Автокодувальники (Autoencoders) та обмежені машини Больцмана (Restricted Boltzmann Machines, RBM) - є неглибокими двошаровими нейронними мережами без внутрішньшарових зв'язків. Deep Belief Networks (DBN) - це клас попередньо навчених мереж без нагляду, де RBM слугують будівельними блоками.

В роботі [14] використовували автокодувальник із розрідженням, додаючи шум та розбіжність Кульбака-Лейблера (KL-divergence) до функції втрат, що покращило продуктивність HAR. Канально-орієнтований спосіб некерованого виділення ознак має такі переваги:

- зменшує складність моделі виділення ознак і, отже, є швидшим;
- більш гнучкий, оскільки різні канали можуть використовувати різні гіперпараметри і навіть різні моделі.

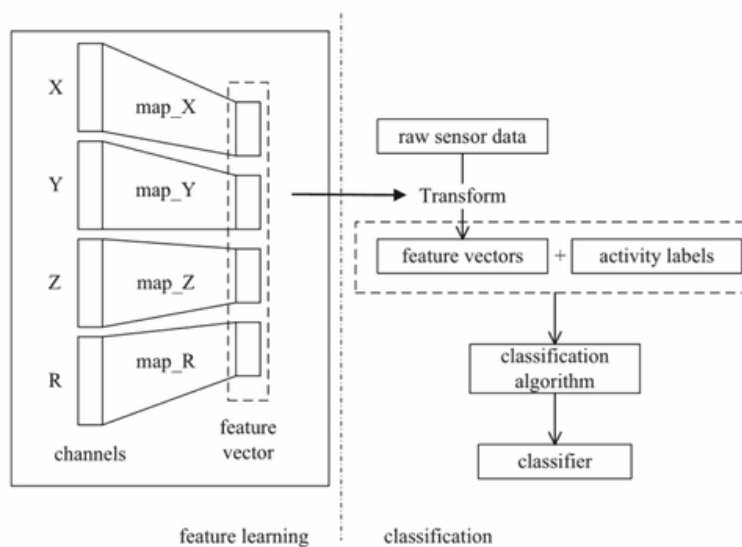


Рис. 2.11. Огляд Некерованого Виділення Ознак для HAR

В даному випадку зосереджуються на наступних трьох методах некерованого виділення ознак: розріджений автокодувальник (sparse auto-encoder, SAE), знешумлюючий автокодувальник (denoising auto-encoder, DAE) та аналіз головних компонент (principal component analysis, PCA).

В роботі [17] реалізували концепцію бінарного RBM для HAR. Розглянемо архітектуру моделі яку ту було запропоновано. Чотири внутрішні

моделі глибоких нейронних мереж (DNN), що входять до складу DeepEar, мають однакову базову мережеву архітектуру, оскільки всі вони, по суті, повинні обробляти один і той самий клас даних — аудіо. Однак, ці моделі навчаються різним представленням даних на кожному шарі залежно від конкретних аудіозадач (наприклад, ідентифікація різних мовців проти визначення загальних категорій звуку).

Ці відмінності проявляються, наприклад, у параметрах функцій активації кожного елемента (unit) та у тому, як елементи колективно обробляють дані, що поширюються мережею.

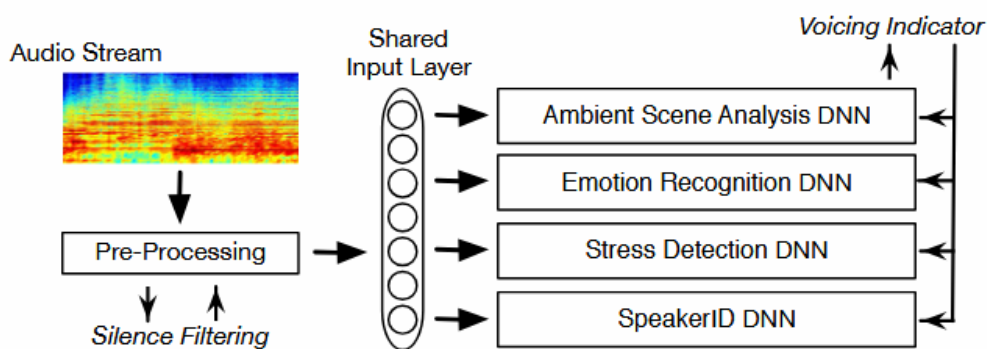


Рис. 2.12. Дизайн DeepEar

На рис. 2.13 проілюстровано архітектуру моделі для DNN DeepEar. Обмежені машини Больцмана (RBM) виступають як базовий будівельний блок для всієї мережі. RBM є типом марковських випадкових полів (Markov Random Field), які включають видимі та приховані елементи.

Шари мережі формуються шляхом накладання кількох RBM, де приховані елементи з одного набору RBM виступають як видимий шар для наступного.

Прихований елемент (k) обчислює свій власний стан (y_k), який передається наступним шарам, у два етапи.

По-перше, він обчислює проміжний стан (x_k) за формулою $b_k + \sum_i y_i w_{ik}$, де b_k — терм упередження, специфічний для цього елемента, y_i — стан

кожного елемента в попередньому шарі, а w_{ik} — вага зв'язку між елементом k і кожним елементом попереднього шару.

По-друге, він застосовує функцію активації до x_k . DeepEar переважно використовує випрямлені лінійні функції (rectified linear functions, ReLU), у цьому випадку $y_k = \max(0, x_k)$.

Вхідний шар конструюється з використанням гаусових видимих елементів, здатних працювати з дійснозначними даними (PLP ознаками), які використовуються для представлення сирих аудіокадрів — це є стандартною практикою при моделюванні аудіо. У всіх інших накладених RBM у мережі використовуються ReLU-елементи (тобто ті, що мають функцію активації ReLU).

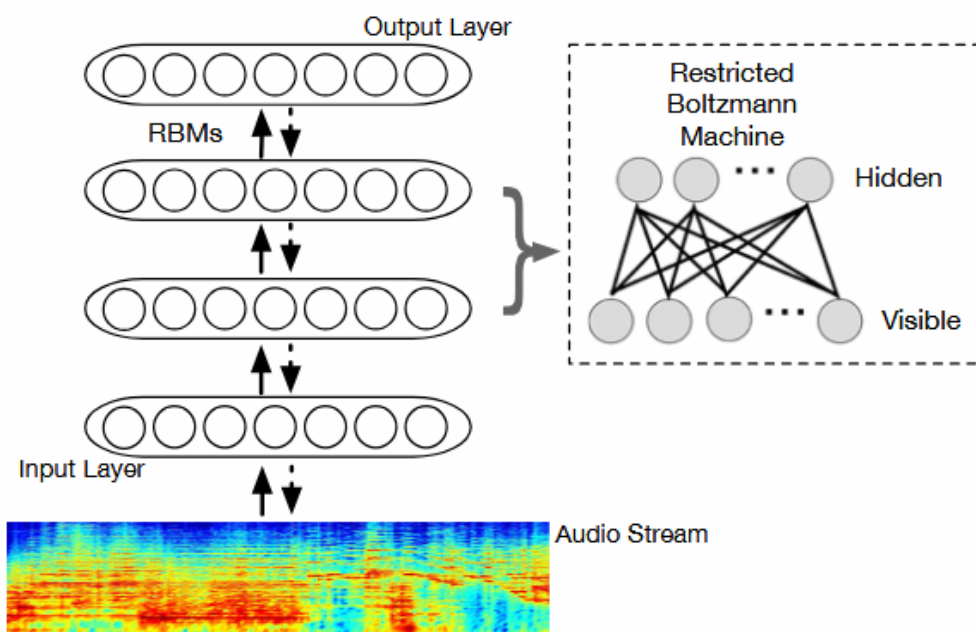


Рис. 2.13. Архітектура RBM, що використовується в DeepEar

Винятком є вихідний шар, де застосовуються softmax функції активації. Нагадаємо, що кожен елемент відповідає класу аудіовисновку, тому в такий спосіб стани елементів можуть бути інтерпретовані як апостеріорні ймовірності (тобто сукупно \sum дорівнює 1, а індивідуально знаходяться в діапазоні $[0,1]$).

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) ефективні для автоматичного виділення просторових ознак. В роботі [15] розглядали одновимірні сенсорні дані як одновимірне зображення для застосування CNN. У типовому підході глибокого навчання для розпізнавання активності (HAR) часові ряди, отримані за допомогою акселерометра та гіроскопа, подаються безпосередньо в інфраструктуру глибокого навчання. Глибока модель містить приховані зв'язки між парами вхідних сигналів; ці кореляції зазвичай ігноруються. Більше того, велика кількість шарів будується один на одному для автоматизації інжинірингу ознак шляхом спроби захопити всі можливі перестановки сигналу, використовуючи велику кількість вузлів. Кожен шар виконує нелінійне перетворення виходів попереднього шару, завдяки чому глибока модель моделює дані за допомогою ієрархії ознак — від низькорівневих до високорівневих.

Ці підходи мають високі обчислювальні вимоги, що робить їх непридатними для пристроїв із низьким енергоспоживанням.

У пропонованому підході зменшуються обчислювальні витрати шляхом обмеження зв'язків від вхідних вузлів та ефективного знаходження ознак за допомогою невеликої кількості вузлів і рівнів.

Лістинг 2.1. Пропонований алгоритм для HAR

Acc_x, Acc_y, Acc_z	▷ Raw Triaxial Acceleration data (Сирі тривісні дані прискорення)
Gyr_x, Gyr_y, Gyr_z	▷ Raw Triaxial Gyroscope data (Сирі тривісні дані гіроскопа)
k_w	▷ Kernel size for temporal convolution (Розмір ядра для часової згортки)
s_f	▷ Number of time-localized points in the spectrogram (Кількість локалізованих у часі точок у спектрограмі)
s_t	▷ Number of frequency points in the spectrogram (Кількість частотних точок у спектрограмі)
d_w	▷ Step of the convolution (Крок згортки)
w_p	▷ Number of filters (Кількість фільтрів)

```

1:   S.a[1], S.a[2], S.a[3] ← spectr(AccX), spectr(AccY), spectr(AccZ);
2:   S.g[1], S.g[2], S.g[3] ← spectr(GyrX), spectr(GyrY), spectr(GyrZ);
3:   for  $i = 1$  to  $w_p$  do
4:     for  $t = 1$  to  $s_f$  do
5:       for  $z = 1$  to 3 do
6:          $o[t][i] += \sum_{j=1}^{s_t} \sum_{k=1}^{k_w} w[i][j][k] \cdot S.g[z][d_w \cdot (t - 1) + k][j];$ 
7:          $o[t + s_f][i] += \sum_{j=1}^{s_t} \sum_{k=1}^{k_w} w[i][j][k] \cdot S.a[z][d_w \cdot (t - 1) + k][j];$ 
8:       end for
9:     end for
10:  end for
11:   $f_{c_o} \leftarrow \text{fully\_connected\_NN}(o);$ 
12:   $\text{result} \leftarrow \text{soft\_max}(f_{c_o});$ 

```

Наші явні зв'язки реалізовані за допомогою сум локальних часових згорток (local temporal convolutions) над перетвореним входом. Оскільки кожна людська активність має дискримінативний розподіл частот, сума виконується відповідно до кожної частоти, отриманої у різних локалізованих у часі значеннях. Деталі реалізації запропонованого підходу наведені в алгоритмі, що подано в лістингу 2.1.

Конкретно, кожен фільтр w_i — розміром $k_w \times s_t$ — застосовується до спектрограми вертикально, як показано на рис. 2.14. Зважена сума згорнутого сигналу в момент часу t обчислюється наступним чином:

$$o[t][i] = \sum_{j=1}^{s_t} \sum_{k=1}^{k_w} w[i][j][k] \cdot \text{input}[d_w \cdot (t - 1) + k][j]$$

де d_w — це крок (stride) часової згортки.

Ці часові згортки формують вихідний шар о розміром $w_p \times \text{OutputFrame}$, де $\text{OutputFrame} = (\text{InputFrame} - k_w)/d_w + 1$, а w_p — кількість фільтрів.

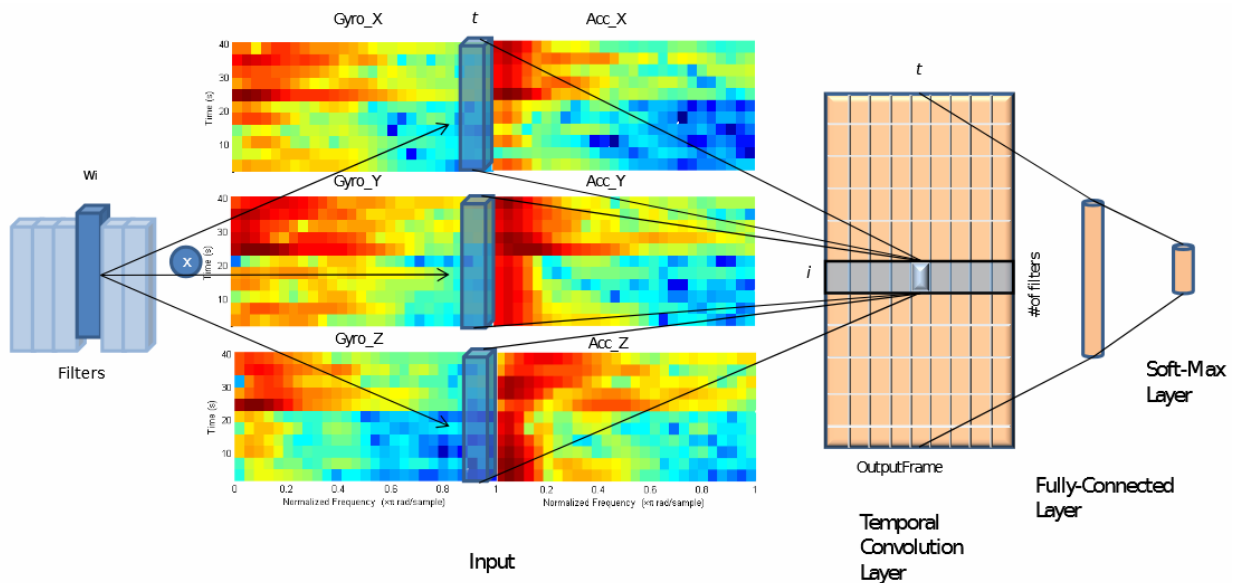


Рис. 2.14. Запропонована архітектура глибокого навчання

На рис. 2.14 фільтр w_i застосовується до попередньо організованих спектрограм входу, а зважені суми згорнутого сигналу в кожен момент часу t обчислюються у шарі часової згортки. Зрештою, повністю зв'язаний шар та шар soft-max використовуються для класифікації.

Для введення інваріантності до орієнтації (orientation invariance), результати часової згортки, отримані від різних осей інертного сенсора, інтегруються разом без будь-якої дискримінації, як показано в циклі for на рядку 5 в лістингу 2.1. Крім того, фільтри, застосовані до трьох осей, ділять одні й ті ж ваги (share the same weights). Це є важливим для зменшення кількості вузлів у шарі та для кращої узагальнювальної здатності (generalisation) моделі до даних.

Останні два шари запропонованого рішення — це повністю зв'язаний шар (fully-connected layer) і шар softmax. Вони необхідні, оскільки наша глибока модель не лише вивчає інформативні ознаки для розпізнавання різних активностей із даних, але й навчається вагам класифікації, які використовуються для розпізнавання класу активності.

Інтерпретація алгоритму лістингу 2.1 як псевдокод python подано в лістингу 2.2.

Лістинг 2.2. Інтерпретація псевдокоду Python запропонованого алгоритму

```
def proposed_har_algorithm(Acc, Gyr, kw, sf, st, dw, wp, w):
    """
    Виконує розпізнавання активності, використовуючи згортку на спектрограмах.

    Вхід:
    Acc: Сирі дані прискорення (3 осі: AccX, AccY, AccZ)
    Gyr: Сирі дані гіроскопа (3 осі: GyrX, GyrY, GyrZ)
    kw: Розмір ядра згортки
    sf: Кількість точок у часі (фреймів) у спектрограмі
    st: Кількість частотних точок у спектрограмі
    dw: Крок згортки
    wp: Кількість фільтрів
    w: Ваги фільтрів [wp, st, kw] - спільні для всіх 6 каналів
    """

    # 1. Попередня обробка: Перетворення на спектрограми
    # S.a[z] - Спектрограма прискорення по осі z (z=1..3)
    S_a = [spectr(Acc[i]) for i in range(3)]
    # S.g[z] - Спектрограма гіроскопа по осі z (z=1..3)
    S_g = [spectr(Gyr[i]) for i in range(3)]

    # Ініціалізація вихідного шару 'o'
    # o матиме розмір [2 * sf, wp] для прискорення та гіроскопа
    Output_Frame = 2 * sf
    o = initialize_output_layer(Output_Frame, wp)

    # 2. Часова згортка та Інтеграція осей (Лінії 3-10)

    for i in range(1, wp + 1): # Ітерація по фільтрах (wp)
        for t in range(1, sf + 1): # Ітерація по часових фреймах (sf)

            # Обчислення зміщення вхідного спектрограми
            time_offset = dw * (t - 1)

            for z in range(1, 3 + 1): # Ітерація по осях (z=1, 2, 3)

                # 6. Згортка ГІРОСКОПА та накопичення (орієнтаційна інваріантність)
                for j in range(1, st + 1): # Частотні точки
                    for k in range(1, kw + 1): # Розмір ядра
                        # Індекс time_index = time_offset + k
                        convolution_term = w[i-1][j-1][k-1] * S_g[z-1][time_offset
                        o[t-1][i-1] += convolution_term
```

```

# 7. Згортка ПРИСКОРЕННЯ та накопичення
for j in range(1, st + 1): # Частотні точки
    for k in range(1, kw + 1): # Розмір ядра
        # Вихідний індекс прискорення зі зміщенням sf
        convolution_term = w[i-1][j-1][k-1] * S_a[z-1][time_offset
o[t + sf - 1][i-1] += convolution_term

# 3. Класифікація

# 11. Повністю зв'язаний шар
fco = fully_connected_NN(o)

# 12. Softmax-класифікатор
result = soft_max(fco)

return result

```

В роботі [13] застосували двовимірну згорткову модель до даних сенсорів смартфонів.

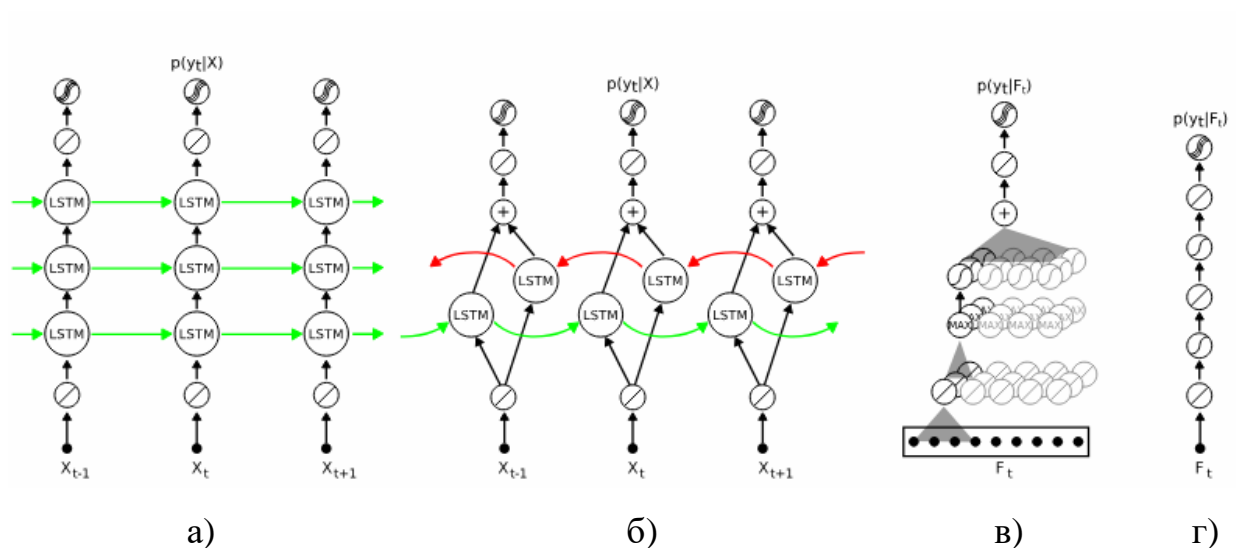


Рис. 2.15. Нейронні мережі що використані до даних сенсорів смартфонів

На рисунку 2.15 - а) Мережа LSTM із прихованими шарами, що містять LSTM-комірки, та фінальним шаром softmax на вершині, б) Двонаправлена мережа LSTM (bi-directional LSTM), що має дві паралельні доріжки: у напрямку майбутнього (зелена) та у напрямку минулого (червона). в) Згортковій мережі, які містять шари згортки і max-pooling, за якими йдуть

повністю зв'язані шари та група softmax. г) Повністю зв'язана мережа прямого поширення (feed-forward network) із прихованими шарами (ReLU).

Рекурентні нейронні мережі (Recurrent Neural Networks, RNN), зокрема LSTM ефективно моделює часові залежності в послідовностях.

Гібридні моделі (CNN-LSTM) - поєднують сильні сторони CNN (виділення локальних ознак) та RNN/LSTM (моделювання часових залежностей). В роботі [8] застосували гібридну модель CNN та LSTM для HAR. Завдяки високій модульності архітектур штучних нейронних мереж (ANN), також можливо додавати шари LSTM до згорткових блоків, як показано на рис. 2.16.

Останній згортковий блок мережі генерує n -вимірні часові ряди, де n — кількість нейронів у цьому згортковому шарі. Цей вихід потім нарізається вздовж часового виміру. Кожен зріз, позначений синім кольором на рис. 2.16, вирівнюється (flattened) і подається як вхід до однієї LSTM-комірки.

Кількість LSTM-комірок, яка дорівнює кількості зрізів, залежить від розмірів вхідних даних, а також від ядер згортки та пулінгу (pooling kernels) у згорткових блоках.

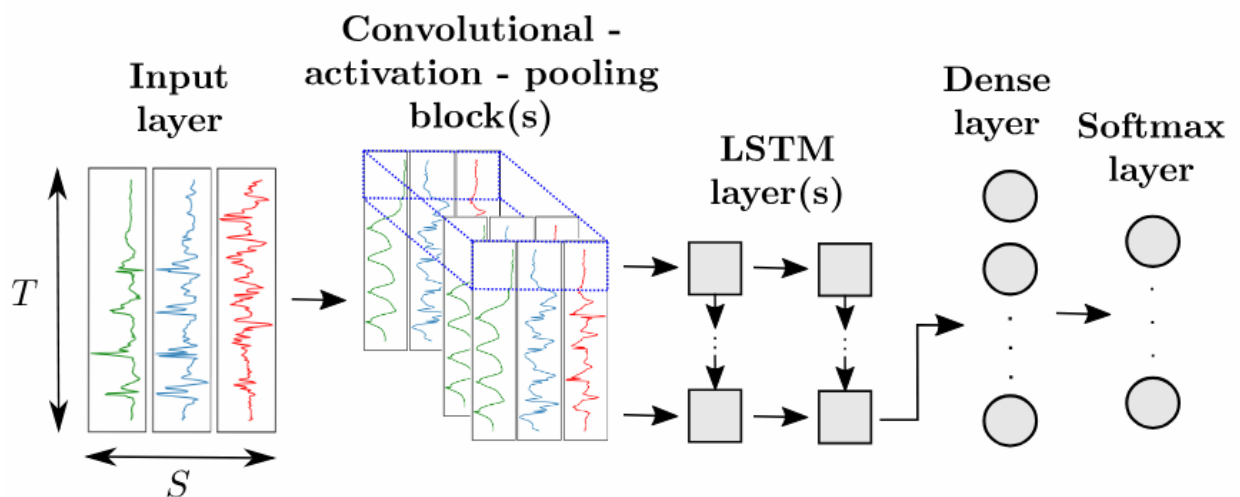


Рис. 2.16. Архітектура гібридної моделі CNN+LSTM для сенсорно-орієнтованого HAR

Кожен зріз вздовж часового виміру виходу згорткового блоку (синім кольором) подається до однієї LSTM-комірки. Усі шари LSTM організовані за схемою "багато-до-багатьох" (many-to-many), за винятком останнього, який відповідає схемі "багато-до-одного" (many-to-one).

В дослідженні [2] здійснили подальше розширення HAR за допомогою комбінації глибоких згорткових та LSTM рекурентних нейронних мереж (DeepConvLSTM). Цей підхід є одним із найефективніших для сенсорно-орієнтованого HAR.

Висновки до розділу

Другий розділ присвячено дослідженню та класифікації сучасних методів і підходів до розпізнавання активності людини з використанням релевантних сенсорних даних. Детально розглянуто інертні сенсори як основне джерело часових даних про рухи користувача, наведено опис публічного набору даних UTD-MHAD, який є еталонним у задачах HAR.

Проведено систематизацію досліджень, які продемонстрували переваги глибоких рекурентних мереж у контексті HAR. Встановлено, що використання двоспрямованих моделей типу BLSTM забезпечує підвищення точності розпізнавання на 5–10% порівняно з односпрямованими варіантами LSTM. Це підтверджує доцільність вибору BLSTM як базової архітектури для побудови запропонованої моделі.

Отже, у цьому розділі визначено ключові технологічні та алгоритмічні тенденції розвитку систем HAR, що ґрунтуються на глибокому навчанні, а також сформульовано вимоги до моделі, яка поєднує ефективність часових нейромереж та адаптивну обробку релевантних сенсорних даних.

РОЗДІЛ 3. ЗАСТОСУВАННЯ МОДЕЛЕЙ ТА МЕТОДІВ РЕКУРЕНТНИХ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АКТИВНОСТІ ЛЮДИНИ НА ОСНОВІ РЕЛЕВАНТНИХ СЕНСОРНИХ ДАНИХ

3.1. Моделі довгої короткочасної пам'яті та їх двоспрямовані варіанти (BLSTM) для задач розпізнавання активності людини

3.1.1. Односпрямована мережа LSTM

Завдяки своїй часовій природі, моделі на основі Довгої короткочасної пам'яті (Long Short-Term Memory, LSTM) здатні фіксувати динамічні часові зміни, що є критичним для точного розпізнавання та класифікації послідовностей активностей. LSTM можуть автономно вивчати контекст, що робить їх високопридатними для розпізнавання людської активності (HAR), дозволяючи моделі самостійно виявляти високорівневі ознаки.

Ключова перевага LSTM над класичними рекурентними нейронними мережами полягає в їхній здатності запам'ятовувати довгострокові часові залежності без виникнення проблеми зникнення або вибуху градієнтів (vanishing/exploding gradient problem). Ця нечутливість до довжини часового проміжку надає LSTM переваги над прихованими марковськими моделями (HMM), RNN та іншими моделями, орієнтованими на часові ряди/послідовності.

У цьому дослідженні проводиться порівняльний аналіз ефективності LSTM та двоспрямованої LSTM (BLSTM) у застосуваннях, що включають розпізнавання близьких за суттю (релевантних) активностей.

LSTM, як варіант RNN, була вперше запропонована Сеппом Хохрайтером та Юргеном Шмідхубером у 1997 році. Шар LSTM складається з множини блоків пам'яті. Кожен блок включає внутрішні вентилі (затвори): вхідний (input gate), забуваючий (forget gate) та вихідний (output gate). Комірки LSTM, які розділяють спільні затвори, утворюють блок LSTM. Ці

внутрішні вентиля виконують операції читання, запису та стирання інформації, дозволяючи успішне навчання моделі за допомогою зворотного поширення в часі (Backpropagation Through Time, BPTT), що ефективно вирішує проблему зникнення градієнтів.

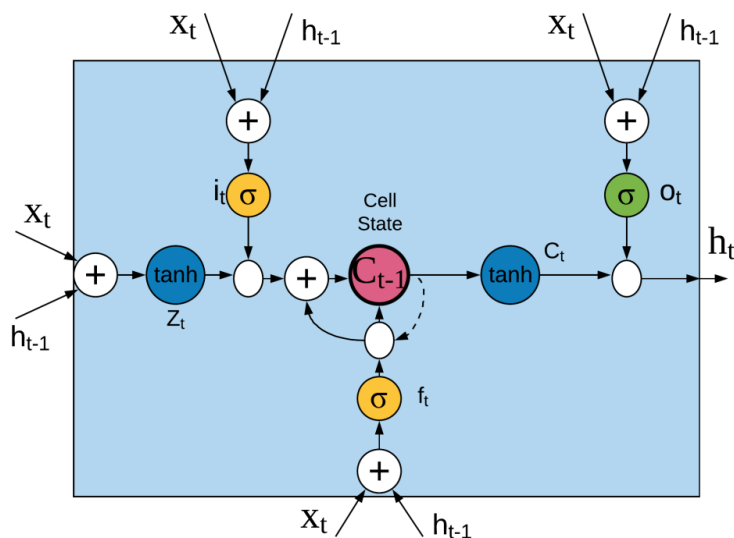


Рис. 3.1. Архітектура комірки LSTM

Рівняння функціонування комірки LSTM:

$$\begin{aligned}
 z_t &= \tanh(W_z x_t + V_z h_{t-1} + b_z) \\
 i_t &= \sigma(W_i x_t + V_i h_{t-1} + b_i) \\
 f_t &= \sigma(W_f x_t + V_f h_{t-1} + b_f) \\
 o_t &= \sigma(W_o x_t + V_o h_{t-1} + b_o) \\
 c_t &= c_{t-1} \circ f_t + i_t \circ z_t \\
 h_t &= o_t \circ \tanh(c_t)
 \end{aligned}$$

де:

x_t — вхід у момент часу t .

h_{t-1} — прихований стан попередньої комірки.

W та V — вагові матриці для входу x та попереднього прихованого стану h відповідно.

b — вектор зміщення.

◦ — добуток Адамара (поелементне множення).

σ та \tanh — сигмоїдна та гіперболічна функції активації.

z_t — кандидат на нову пам'ять (генерується на основі поточного входу та попереднього прихованого стану).

i_t — вхідний вентиль, який визначає, яка нова інформація буде записана.

f_t — забуваючий вентиль, який вирішує, яка інформація має бути видалена з попереднього стану пам'яті c_{t-1} .

o_t — вихідний вентиль, який контролює, яка частина поточного стану пам'яті c_t буде виведена як h_t .

c_t — стан пам'яті (остаточна пам'ять), що є сумою минулого стану (c_{t-1} , відфільтрованого f_t) та нової пам'яті (z_t , відфільтрованої i_t).

h_t — прихований стан (вихід комірки), що формується шляхом застосування \tanh до стану пам'яті та фільтрування через o_t .

3.1.2. Двоспрямована мережа LSTM (BLSTM)

Двоспрямована рекурентна нейронна мережа (BRNN) усуває обмеження стандартної RNN, дозволяючи враховувати контекст як з минулого, так і з майбутнього. Двоспрямована LSTM (BLSTM) є LSTM-варіантом BRNN.

У структурі BLSTM/BRNN використовуються дві незалежні рекурентні мережі — пряма (навчання від часу $t=1$ до $t=T$) та зворотна (навчання від $t=T$ до $t=1$) — які з'єднані з єдиним вихідним шаром. Це дозволяє одночасно оцінювати як майбутню, так і минулу інформацію послідовних входів у часовому проміжку.

Рівняння функціонування BRNN:

$$\begin{aligned}h_f &= H(W_{xh_f}x_t + W_{h_f h_f}h_{f_{t+1}} + b_{h_f}) \\h_b &= H(W_{xh_b}x_t + W_{h_b h_b}h_{b_{t+1}} + b_{h_b}) \\y_t &= W_{yh_f}h_{f_t} + W_{yh_b}h_{b_t} + b_y\end{aligned}$$

де:

h_f та h_b — приховані послідовності прямого та зворотного шарів відповідно.

H — функція прихованого шару (для BLSTM це функція комірки LSTM).

y_t — вихід у момент часу t .

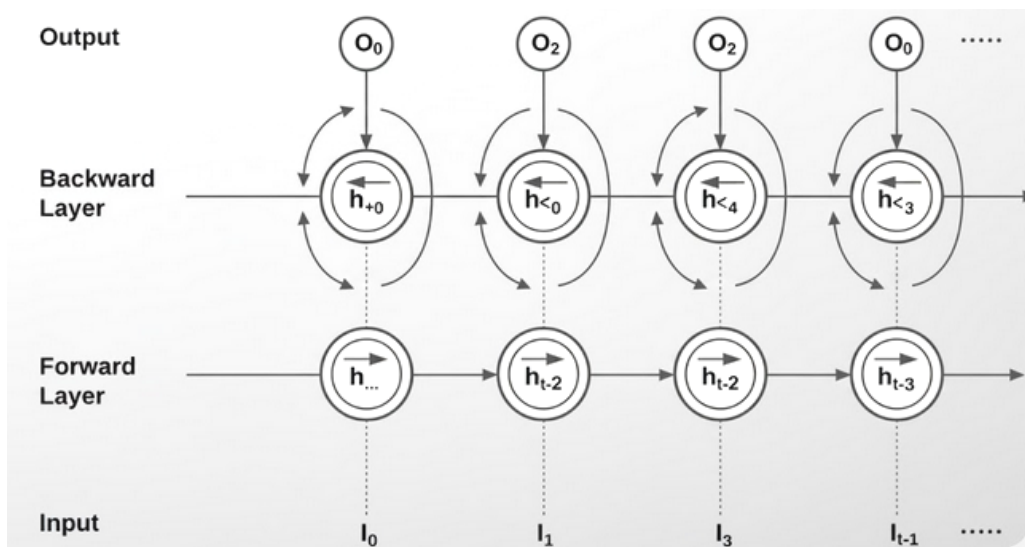


Рис. 3.2. Двоспрямована рекурентна нейронна мережа

Хоча BRNN споживає вдвічі більше пам'яті через підтримку двох прихованих шарів, це забезпечує двосторонній доступ до довгострокового контексту. BLSTM успадковує цю здатність і може продемонструвати покращення продуктивності у класифікаційних процесах, особливо для розпізнавання схожих активностей, оскільки модель навчається в обох напрямках: вперед та назад.

Нейронні вузли у прямому та зворотному напрямках двоспрямованої рекурентної нейронної мережі функціонують як односпрямовані елементи. Оскільки ці дві рекурентні мережі не мають внутрішньшарових зв'язків між собою, процес їхнього навчання аналогічний процесу навчання звичайної RNN, але з необхідністю обробки двох незалежних прихованих послідовностей.

1. Етап прямого поширення (Forward Pass)

На етапі прямого поширення виконується ітеративна обробка всієї вхідної послідовності від часу $t=1$ до $t=T$.

Прямий прохід - обчислення прихованих станів здійснюється у часовому напрямку (від $t=1$ до $t=T$).

Зворотний прохід - обчислення прихованих станів здійснюється у зворотному часовому напрямку (від $t=T$ до $t=1$).

Вихідний шар - після завершення обох проходів (прямого та зворотного) для всієї послідовності, здійснюється фінальний прямий прохід для обчислення передбачуваних виходів y_t на вихідному шарі.

2. Етап зворотного поширення (Backward Pass)

На етапі зворотного поширення обчислюються градієнти для оновлення параметрів мережі. Для кожного моменту часу $t \in [1, T]$:

Обчислення градієнтів помилки - визначається похідна функції помилки (loss function), яка була використана на етапі прямого поширення.

Градієнти для прямих станів поширюються у зворотному часовому напрямку (від $t=T$ до $t=1$). Градієнти для зворотних станів поширюються у прямому часовому напрямку (від $t=1$ до $t=T$).

Після завершення зворотного поширення градієнтів для всіх часових кроків, оновлюються усі вагові коефіцієнти та зміщення мережі.

У цьому проекті властивість двостороннього навчання BLSTM використовується для покращення розпізнавання близько споріднених дій, де модель BLSTM розроблена для доступу до довгострокового контексту в обох напрямках.

3.2. Опис набору даних UTD-MHAD для дослідження

Для дослідження та оцінки ефективності двоспрямованої мережі довгої короткочасної пам'яті (BLSTM) у контексті розпізнавання релевантних за

суттю активностей (closely related activities) було використано публічно доступний набір даних UTD-MHAD.

UTD Multimodal Human Action Dataset (UTD-MHAD)

Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz

[Signal and Image Processing \(SIP\) Lab](#) & [Embedded Systems and Signal Processing \(ESSP\) Lab](#)

Department of Electrical Engineering, University of Texas at Dallas

Introduction

This dataset was collected as part of our research on human action recognition using fusion of depth and inertial sensor data. The objective of this research has been to develop algorithms for more robust human action recognition using fusion of data from differing modality sensors. (Note: A more recent dataset using the second generation of Kinect camera is provided at this link <https://drive.google.com/drive/folders/1iMZ68ASWTVBomP2U10qhAlsaQ2z-XTIa?usp=sharing> or at this link <http://www.utdallas.edu/~kehtar/Kinect2Dataset.zip> and its description at <http://www.utdallas.edu/~kehtar/Kinect2DatasetReadme.pdf> .

A multiView action dataset is also provided at <http://www.utdallas.edu/~kehtar/MultiViewDataset.zip> with a description at <http://www.utdallas.edu/~kehtar/MultiViewDataset.pdf>

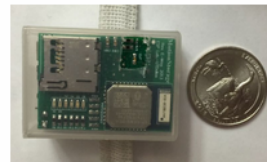
Sensors

For our multimodal human action dataset reported here, only one Kinect camera and one wearable inertial sensor were used. This was intentional due to the practicality or relatively non-intrusiveness aspect of using these two differing modality sensors. Both of these sensors are low cost, easy to operate, and do not require much computational power for the real-time manipulation of data generated by them. A picture of the Kinect camera is shown below. It can capture a color image with a resolution of 640×480 pixels and a 16-bit depth image with a resolution of 320×240 pixels. The frame rate is approximately 30 frames per second.

The wearable inertial sensor used here was the low-cost wireless inertial sensor built in the ESSP Laboratory at the University of Texas at Dallas. This sensor consists of (i) a 9-axis MEMS sensor which captures 3-axis acceleration, 3-axis angular velocity and 3-axis magnetic strength, (ii) a 16-bit low power microcontroller, (iii) a dual mode Bluetooth low energy unit which streams data wirelessly to a laptop/PC, and (iv) a serial interface between the MEMS sensor and the microcontroller enabling control commands and data transmission. This wearable inertial sensor is shown in the figure below. The sampling rate of this wearable inertial sensor is 50 Hz. The measuring range of the wearable inertial sensor is $\pm 8g$ for acceleration and ± 1000 degrees/second for rotation.



Kinect camera



Wearable inertial sensor

Рис. 3.3. Інформація про датасет

Набір даних UTD-MHAD містить інформацію, зібрану від натільного інертного сенсора, який фіксує 3-вісні показники прискорення та 3-вісні показники кутової швидкості (гіроскоп).

- Частота дискретизації: 50 Гц.
- Діапазон вимірювання: $\pm 8g$ для прискорення та ± 1000 градусів/секунду для обертання.
- Розміщення сенсора: інертний сенсор був закріплений на правому зап'ясті кожної особи.

Експериментальна база охоплює 27 різних активностей у приміщенні.

Суб'єкти: У дослідженні брали участь 8 осіб (4 жінки та 4 чоловіки).

Спроби: Кожна особа виконувала кожен активність 4 рази.

Загальний обсяг даних: Після видалення 3 пошкоджених послідовностей, було отримано загалом 861 послідовність $[(27 \times 8 \times 4) - 3]$.

Набір даних було використано у двох конфігураціях:

- Експеримент з 15 активностями - використовувалися перші 15 активностей зі списку.

- Експеримент з 27 активностями - використовувався повний набір із 27 активностей.

Перелік 27 активностей, розділених за їхньою природою, виглядає наступним чином (таблиця 3.1).

Таблиця 3.1.

Перелік досліджуваних активностей

Категорія руху	Активності
Жести/ маніпуляції руками	Свайп вліво, свайп вправо, махання правою рукою, плескання в долоні двома руками, штовхання (двома руками), схрещення рук на грудях, згинання рук (двома руками).
Малювальні/ точні рухи	Малювання хрестика, малювання кола (за годинниковою стрілкою), малювання кола (проти годинникової стрілки), малювання трикутника.
Спортивні/ кидкові рухи	Боулінг, боксування вперед, замах бейсбольною биткою правою рукою, замах тенісною ракеткою форхендом, кидок баскетбольного м'яча, подача тенісного м'яча, кидок.
Взаємодія з об'єктами	Стукання у двері, ловля об'єкта, підняття та кидок (об'єкта).
Пересування/ постуральні зміни	Інертний сенсор, що носить на правому стегні для: біг на місці, ходьба на місці, сидання-вставання, вставання-сидання, випад лівою ногою вперед, присідання (з витягнутими вперед двома руками).

3.3. Архітектура BLSTM мережі для диференціації релевантних активностей

Застосування запропонованої моделі орієнтоване на розпізнавання активності у великій множині, що налічує 27 близько споріднених дій. Ці активності характеризуються або високою подібністю у послідовностях сенсорних даних, або візуальною схожістю при спостереженні.

Наприклад, дії «свайп вправо» та «махання правою рукою» можуть мати візуальну схожість, тоді як «замах бейсбольною биткою» та «замах тенісною ракеткою» демонструють високоподібні патерни прискорення вздовж осей X, Y та Z.

Для ефективної класифікації таких активностей із застосуванням глибокого навчання, модель повинна володіти здатністю виявляти незначні, дискримінативні відмінності у динаміці руху.

У структурі двоспрямованої LSTM вхідні дані навчаються у прямому та зворотному напрямках у двох окремих прихованих шарах, які використовують спільний вхідний шар. Ця властивість дозволяє моделі враховувати повний часовий контекст (минуле та майбутнє відповідно) послідовності, що є ключовим для точного розрізнення подібних активностей і забезпечує кращі результати порівняно з односпрямованими нейронними мережами.

Архітектура BLSTM-моделі, розробленої в цьому проєкті, складається з двох шарів:

Перший шар - шар двоспрямованої LSTM із конфігурацією «багато-до-багатьох» (many-to-many). Вихід усіх комірок цього шару використовується як вхід для наступного шару.

Вихідний шар (щільний шар) - Повністю зв'язаний шар, який виконує класифікацію. Для виходу цього шару застосовується сигмоїдна функція активації.

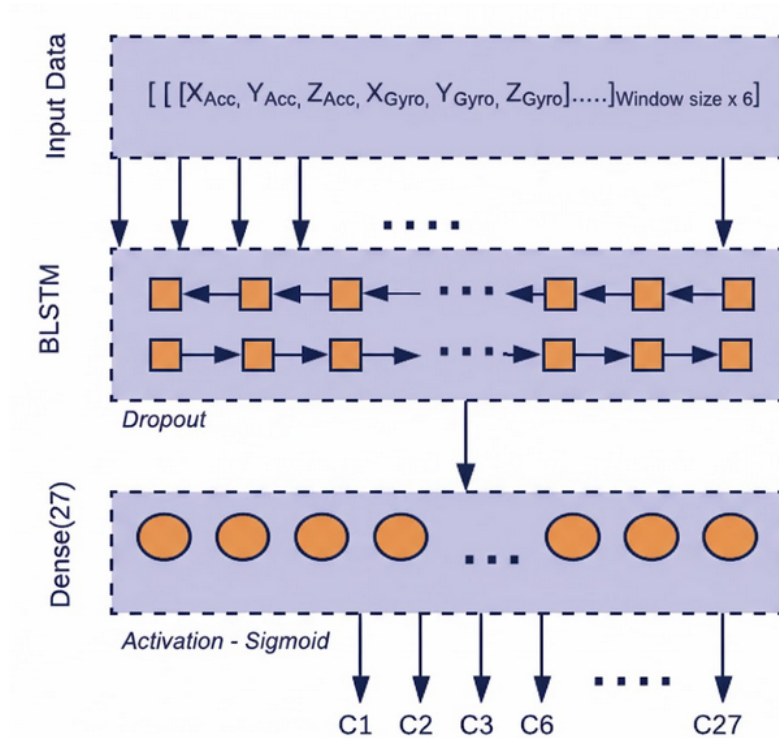


Рис. 3.4. Архітектура двоспрямованого LSTM

Кількість комірок BLSTM у кожному шарі фіксується і визначається на основі кількості часових зразків для однієї спроби активності в наборі даних. Ця кількість залишається незмінною як для експериментів з 15, так і з 27 активностями. Застосування методу ковзного вікна до початкових даних збільшує загальну кількість послідовностей (зразків) для навчання. Для запобігання перенавчанню (overfitting) моделі використовувалася техніка регуляризації dropout.

Проаналізовано оптимізаційні алгоритми, найбільш ефективні для цих випадків, залежно від розміру навчальної вибірки:

- Для меншої вибірки (15 активностей) ефективним виявився оптимізатор RMSProp.
- Для більшої вибірки (27 активностей) краще підходив оптимізатор Adam.

Спостерігається необхідність адаптації оптимізатора відповідно до зростання розміру вибірки. Крім того, проводиться порівняльний аналіз кінцевих результатів розпізнавання між розробленою моделлю BLSTM та

односпрямованою моделлю LSTM, архітектура якої використовується для порівняння.

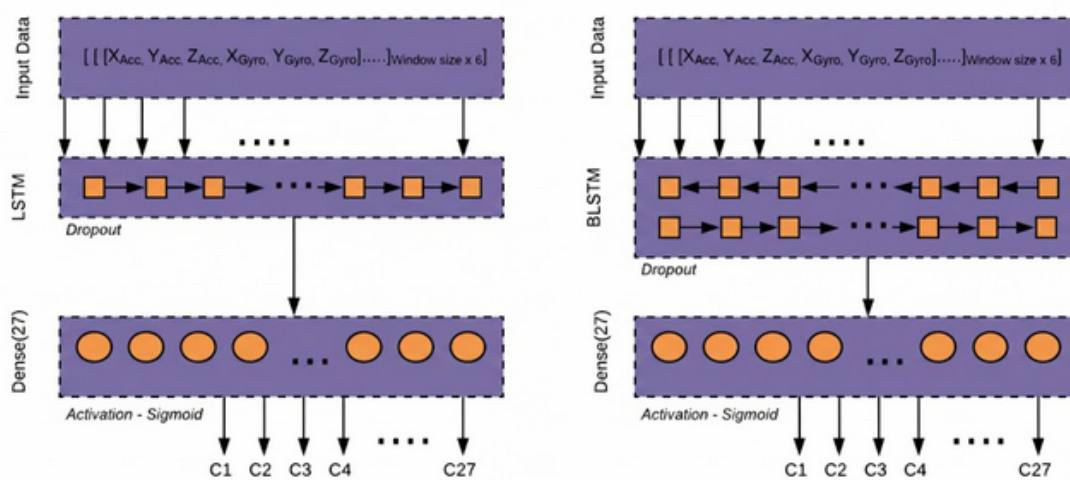


Рис. 3.5. Моделі BLSTM та LSTM для порівняння

3.4. Методологія підготовки даних та експериментальна конфігурація

3.4.1. Підготовка даних та експериментальні розподіли

Для цілей навчання та експериментальної оцінки моделі BLSTM вибраний набір даних було розділено на дві окремі множини активностей: 15 активностей та 27 активностей. Це поділ виконано для проведення глибшого аналізу моделі BLSTM та для виявлення потреби в оптимізації методів та/або гіперпараметрів при зростанні складності активностей.

Кожна з цих множин даних була стандартно розділена на навчальний набір (80%) та тестовий набір (20%). Тестування тестового набору даних проводилося згідно з двома основними стратегіями розподілу:

1. Специфічний для суб'єкта розподіл (Subject-Specific Split) - для тестування було виділено останні два зразки кожної активності, виконані кожною особою. Таким чином, кожний суб'єкт вніс свій внесок у формування тестової вибірки.

2. Загальний тестовий розподіл (General Test Split) - для тестування були зарезервовані повні зразки активностей останніх двох осіб у наборі даних. Повні зразки активностей решти шести осіб використовувалися виключно для навчання.

3.4.2. Програмне забезпечення та обчислювальне середовище

Реалізація моделі здійснювалася на обчислювальній платформі з процесором Intel Core i7 та 16 ГБ оперативної пам'яті. Використовувалися наступні програмні фреймворки та бібліотеки глибокого навчання:

- Фреймворк TensorFlow
- Бібліотеки – Keras, Pandas, Numpy та Scikit-Learn.

3.4.3. Метод ковзного вікна (sliding window)

Нерівна довжина послідовностей у часових рядах є типовою проблемою. У даному випадку, тривалість послідовності для кожної спроби активності, виконаної конкретною особою, є різною.

Традиційні методи нормалізації (наприклад, обрізання або доповнення нулями) часто призводять до втрати критичної часової інформації. На противагу цьому, застосовується метод ковзного вікна, який дає змогу розділити часові дані на блоки фіксованої довжини без втрати контексту.

Принцип ковзного вікна полягає в наступному. Часові дані діляться на послідовні блоки. Вікно переміщується до наступного блоку (з певним перекриттям або кроком), і цей блок додається до послідовності зразків. Це забезпечує фіксовану довжину послідовності для всіх вхідних даних, зберігаючи часову інформацію.

Якщо часові дані представлені як послідовність $(x_0, x_1, x_2, \dots, x_{n-1}, x_n, x_{n+1}, \dots)$, а розмір вікна фіксовано на рівні k , інтервал даних перетворюється на зразок $(x_{i-k}, x_{i-k+1}, \dots, x_i, x_{i+1})$.

Параметри вхідного вектора:

- Розмір вікна (W) встановлено на 128 часових кроків.

- Форма вхідного вектора: $N \times W \times 6$.
 - Розмірність каналу (6) включає 3 осі прискорення та 3 осі кутової швидкості (гіроскоп).
 - N : Загальна кількість вікон, згенерована для всього простору зразків.
- Оскільки частота дискретизації інертного сенсора становить 50 Гц, часовий інтервал між послідовними точками даних дорівнює 0,02 секунди.

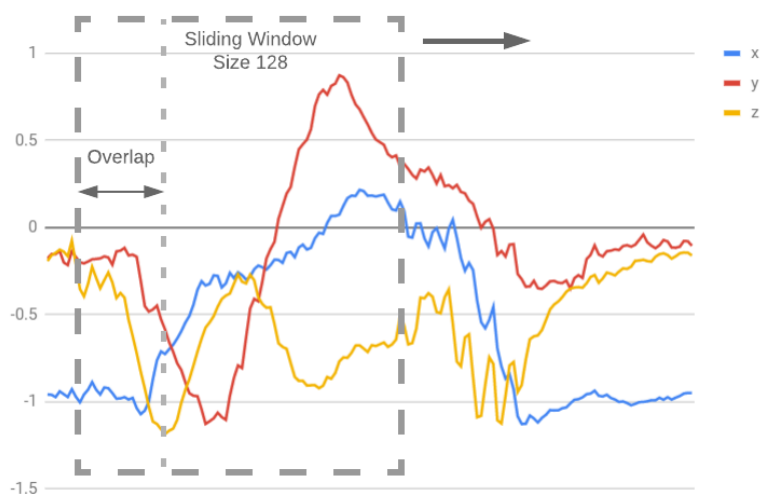


Рис. 3.6. Ковзне вікно зі 128 тимчасовими кроками

3.5. Функція активації та техніка регуляризації в моделі BLSTM

3.5.1. Функція активації

У запропонованій моделі двоспрямованої мережі довгої короткочасної пам'яті як функція активації використовується сигмоїдна функція (Sigmoid Activation Function), яка також відома як логістична функція. Ця функція застосовується у фінальному щільному (fully-connected) вихідному шарі моделі.

Сигмоїдна функція $f(x)$ є дійсною та диференційованою, що є критично важливим для обчислення градієнтів під час зворотного поширення помилки. Вона математично імітує поведінку біологічного нейрона, де її вихід вказує на імовірність спрацьовування.

Функція задається рівнянням:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Експериментальне порівняння з іншими функціями активації продемонструвало, що сигмоїда є найбільш ефективною для даної конфігурації моделі.

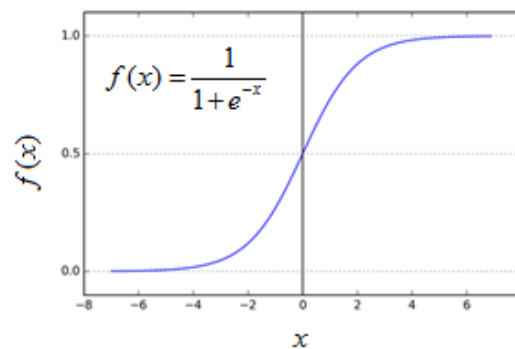


Рис. 3.7. Функція активації сигмоїда

3.5.2. Техніка регуляризації

Для запобігання перенавчанню (overfitting) моделі BLSTM застосовується техніка регуляризації Dropout. Dropout передбачає випадкове вимкнення (ігнорування) певної частки нейронів під час кожного кроку навчання. Це запобігає формуванню складних коадаптацій (спільної залежності) на тренувальних даних. Ефективність Dropout доведена у багатьох складних задачах, включаючи розпізнавання мовлення та зображень.

Було запропоновано кілька варіантів Dropout, адаптованих для рекурентних нейронних мереж (RNN) та їх варіантів (LSTM):

- Нерекурентний Dropout: пропонується застосовувати Dropout лише до нерекурентних з'єднань RNN.

- Варіаційний (рекурентний) Dropout: потрібно застосовувати одну й ту ж маску Dropout на кожному часовому кроці як до рекурентних, так і до прямих з'єднань.

- Dropout стану комірки: необхідно застосовувати Dropout до станів клітин.

- Dropout векторів оновлення: необхідно застосовувати Dropout до векторів оновлення прихованого стану.

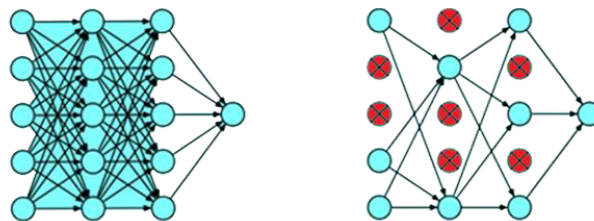


Рис. 3.8. Звичайний (регулярний) dropout у нейронній мережі прямого поширення

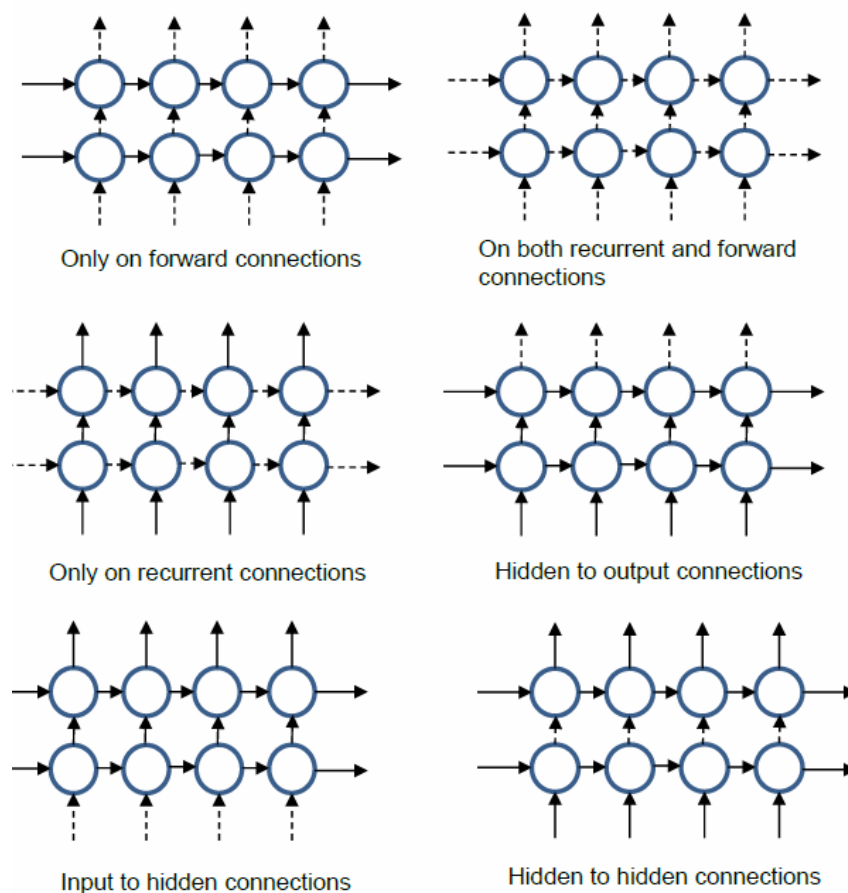


Рис. 3.9. Dropout у рекурентних нейронних мережах та їхніх варіантах

На рис. 3.9 суцільними лініями позначено відсутність Dropout, а пунктирними - застосування Dropout (між вхід-прихований, прихований-вихід або прихований-прихований).

У даному проєкті використовується звичайний Dropout (regular dropout), який застосовується до прямих з'єднань між шаром BLSTM та фінальним щільним вихідним шаром.

Причиною було те, що оскільки застосування методу ковзного вікна значно збільшує кількість послідовностей у вибірці, звичайний Dropout виявився ефективним для узагальнення.

Додаткові експерименти, що включали комбінування цього звичайного Dropout із рекурентним Dropout у шарі BLSTM, показали, що звичайний Dropout самостійно забезпечує кращу продуктивність та призводить до значно нижчої помилки узагальнення (generalisation error).

3.6. Функція втрат та оптимізаційні алгоритми для навчання моделі

3.6.1. Функція втрат. Категоріальна перехресна ентропія (Categorical Cross-Entropy)

Для кількісної оцінки розбіжності між прогнозованими та істинними мітками в даній моделі використовується категоріальна перехресна ентропія (Categorical Cross-Entropy). Вибір цієї функції втрат обумовлений потребою в імовірнісній інтерпретації оцінок моделі, де вихідний шар генерує вектор S ймовірностей, по одній для кожної категорії.

Мета навчання полягає у мінімізації цієї функції втрат протягом ітерацій тренування. Категоріальна перехресна ентропія вимірює розбіжність (dissimilarity) між прогнозованим розподілом міток та істинним розподілом міток.

Функція задається наступним чином:

$$-\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C 1_{y_i \in C_c} \log p_{\text{model}} [y_i \in C_c]$$

де:

N — загальна кількість спостережень (зразків).

C — загальна кількість категорій класів.

$1_{y_i \in C_c}$ — індикаторна функція, яка дорівнює 1, якщо істинна мітка y_i для спостереження i належить до категорії c , і 0 в іншому випадку.

$p_{\text{model}} [y_i \in C_c]$ — ймовірність, передбачена моделлю для приналежності спостереження i до категорії c .

3.6.2. Оптимізаційні алгоритми

Для ефективного оновлення параметрів моделі та мінімізації функції втрат були досліджені два адаптивні оптимізаційні алгоритми.

RMSProp (Root Mean Square Propagation)

Оптимізатор RMSProp підтримує індивідуальні швидкості навчання для кожного параметра, які адаптуються на основі експоненційно зменшеного середнього квадрата градієнтів за останні ітерації. RMSProp ефективно працює у випадках, коли ваги змінюються з високою швидкістю. У даній роботі RMSProp показав високу ефективність для навчальної вибірки з 15 активностями.

RMSProp автоматично зменшує розмір кроку градієнта, коли кроки стають надмірно великими.

Оновлювальні рівняння RMSProp:

$$v_t = \gamma v_t + (1 - \gamma) g_t^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} g_t$$

Де:

η — початкова швидкість навчання (типово 0,001).

g_t — градієнт у момент часу t .

v_t — експоненційно зменшуване середнє минулого квадрата градієнтів.

ε — мала константа для уникнення ділення на нуль.

γ — параметр зменшення (зазвичай встановлюється на 0,9).

Алгоритм Adam (Adaptive Moment Estimation)

Adam заснований на адаптивних оцінках моментів нижчого порядку. На відміну від RMSProp, Adam зберігає не лише експоненційно зменшуване середнє квадрата градієнтів (v_t), але й експоненційно зменшуване середнє минулого градієнта (m_t) — так звані перший та другий моменти відповідно. Adam добре підходить для нестационарних цільових функцій і продемонстрував кращу продуктивність, коли складність набору даних була збільшена до 27 активностей.

Оскільки m_t та v_t ініціалізуються нульовими векторами, вони мають зміщення до нуля (bias towards zero). Для коригування цього зміщення застосовуються виправлені оцінки \bar{m}_t та \bar{v}_t .

Обчислення та корекція моментів:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t && \text{(Перший момент)} \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 && \text{(Другий момент)} \\ \bar{m}_t &= \frac{m_t}{1 - \beta_1^t} && \text{(Корекція першого моменту)} \\ \bar{v}_t &= \frac{v_t}{1 - \beta_2^t} && \text{(Корекція другого моменту)} \end{aligned}$$

Правило оновлення Adam:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\bar{v}_t + \varepsilon}} \bar{m}_t$$

де β_1 (типово $\approx 0,9$) та β_2 (типово $\approx 0,99$) — коефіцієнти зменшення. Спостережено, що зі зростанням розміру та складності вибірки існує необхідність у зміні оптимізатора.

3.7. Результати експериментальної оцінки моделей

3.7.1. Аналіз точності (Accuracy)

Експериментальна оцінка продуктивності моделі двоспрямованої мережі довгої короткочасної пам'яті продемонструвала високі показники точності, особливо на наборі даних, специфічному для суб'єкта:

- 15 Активностей (специфічний для суб'єкта): BLSTM досягла загальної точності 98,05%.

- 27 Активностей (специфічний для суб'єкта): BLSTM досягла загальної точності 90,87%.

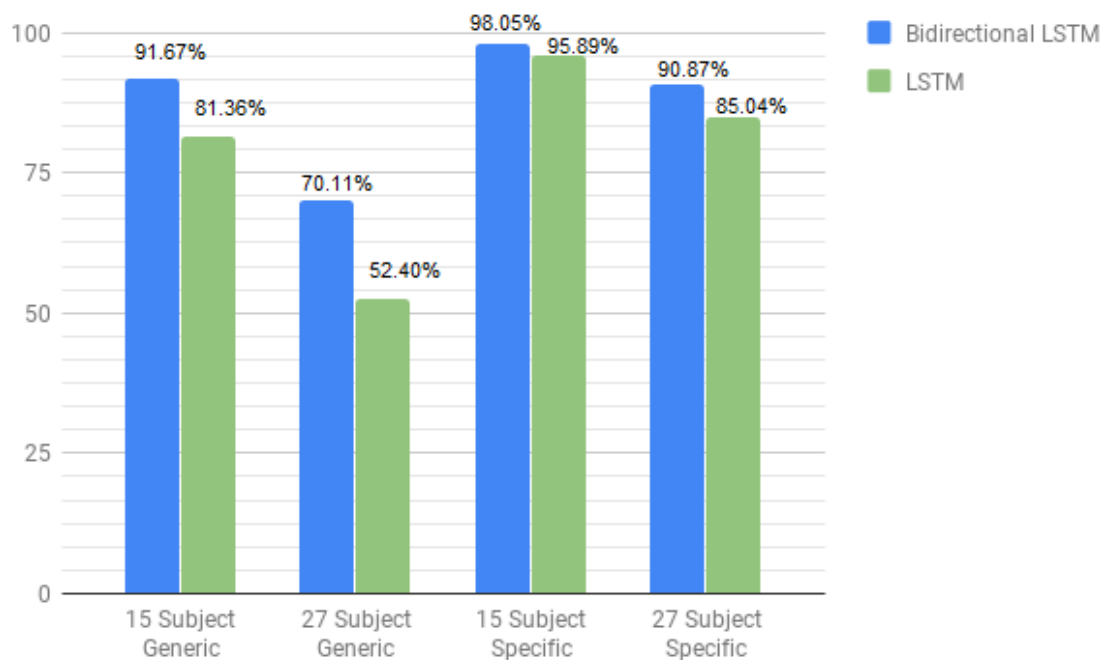


Рис. 3.10. Точність BLSTM та LSTM відображає порівняння точності між моделями BLSTM та LSTM для обох наборів активностей (15 та 27) у конфігураціях загального та специфічного для суб'єкта тестування

Як висновок, зі збільшенням розміру вибірки від 15 до 27 активностей продуктивність LSTM значно знизилася, наприклад, у випадку загального тестування — з 81,36% до 52,40%. Проте, зі зростанням складності завдання

(збільшення до 27 активностей), BLSTM демонструє кращу стійкість та ефективність порівняно з LSTM.

Різниця в точності (специфічний для суб'єкта): для 15 активностей різниця між BLSTM та LSTM становила 2,16%; для 27 активностей ця різниця зросла до 5,83%, підкреслюючи перевагу BLSTM у розпізнаванні близько споріднених дій.

3.7.2. Комплексні метрики оцінки (повнота, точність, F1-оцінка)

Для всебічної оцінки класифікаторів були використані метрики повнота (Recall), точність (Precision) та F1-оцінка.

Повнота (Recall) відображає частку істинно позитивних випадків, які були правильно ідентифіковані (справжній позитивний показник).

$$\text{Повнота} = \frac{\text{Істинно позитивний}}{\text{Істинно позитивний} + \text{Хибно негативний}}$$

Точність (Precision) відображає частку правильно ідентифікованих позитивних випадків серед усіх елементів, класифікованих як позитивні.

$$\text{Точність} = \frac{\text{Істинно позитивний}}{\text{Істинно позитивний} + \text{Хибно позитивний}}$$

F1-оцінка (F1-Score) - гармонійне середнє повноти та точності, що забезпечує міру балансу між цими двома метриками.

$$\text{F1-оцінка} = \frac{2 \times \text{Повнота} \times \text{Точність}}{\text{Повнота} + \text{Точність}}$$

Результати метрик наведено в таблиці 3.2.

Результати метрик

Набір активностей / модель	Тип Тестування	Повнота (%)	Точність (%)	F1-оцінка (%)
15 активностей LSTM	Загальний	82,54	82,83	78,98
15 активностей BLSTM	Загальний	90,99	91,81	90,92
15 активностей LSTM	Специфічний	95,87	95,25	95,45
15 активностей BLSTM	Специфічний	97,90	97,67	97,74
27 активностей LSTM	Загальний	53,84	55,73	54,77
27 активностей BLSTM	Загальний	72,03	71,17	69,55
27 активностей LSTM	Специфічний	84,90	83,72	83,78
27 активностей BLSTM	Специфічний	90,65	90,58	90,34

Отже, BLSTM демонструє найвищу продуктивність у специфічному для суб'єкта тестуванні, досягаючи F1-оцінки 97,74% для 15 активностей та 90,34% для 27 активностей. Це підтверджує, що BLSTM краще використовує двонаправлений контекст для розрізнення складних часових патернів.

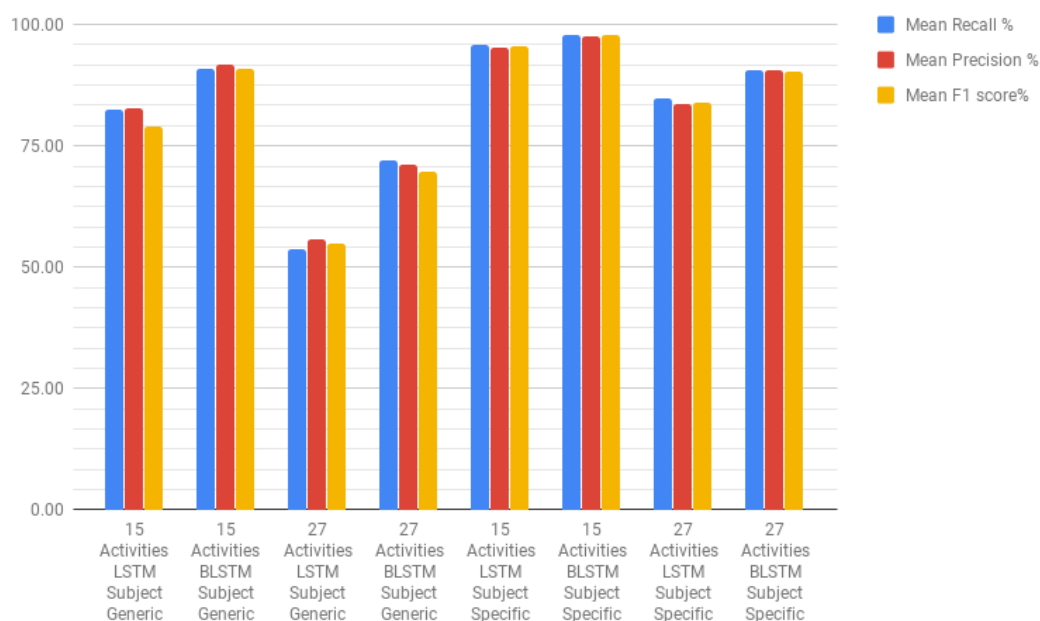


Рис. 3.11. Повнота, точність та F1-оцінка для різних комбінацій моделі

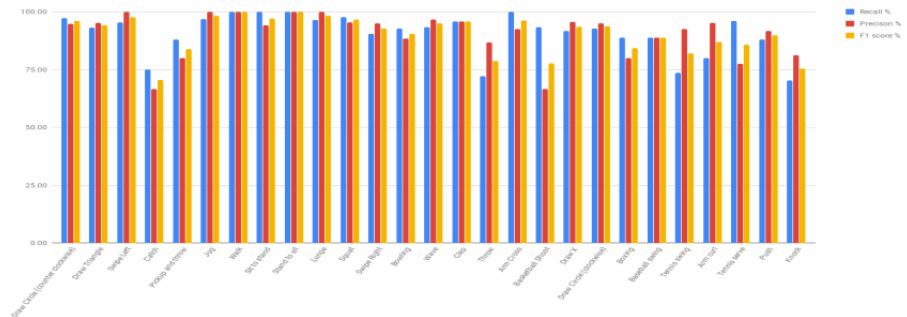
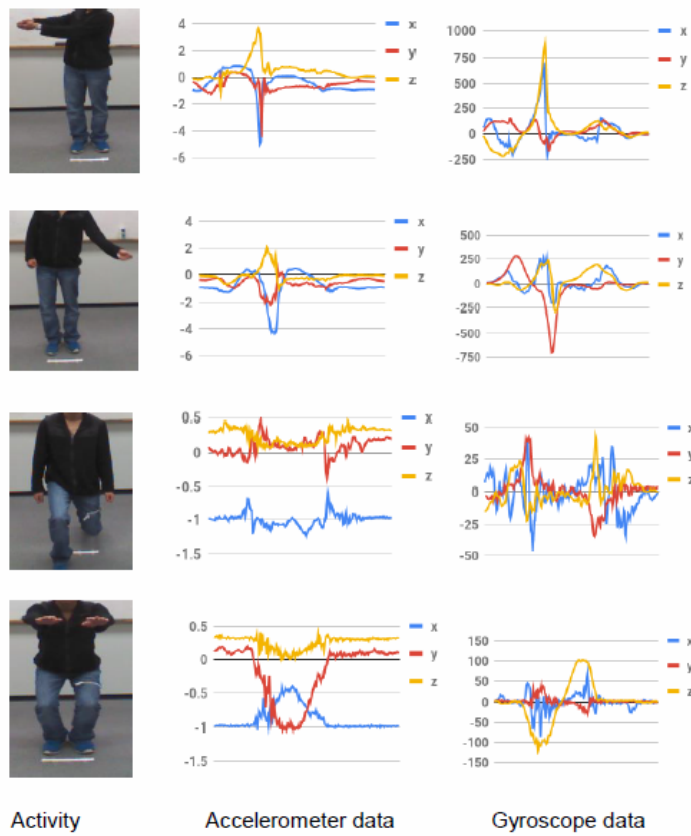


Рис. 3.12. Активність, вхідні дані та вихідні результати

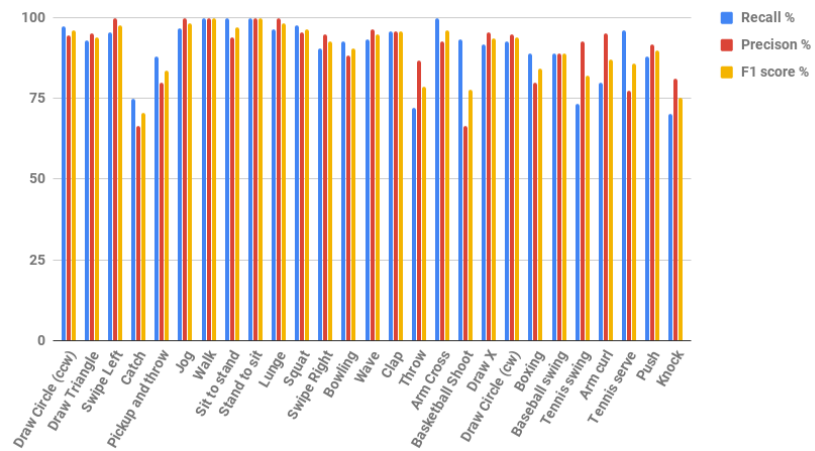


Рис. 3.13. Метрики точності для моделі BLSTM на тестовому наборі даних, специфічному для суб'єкта, для 27 активностей

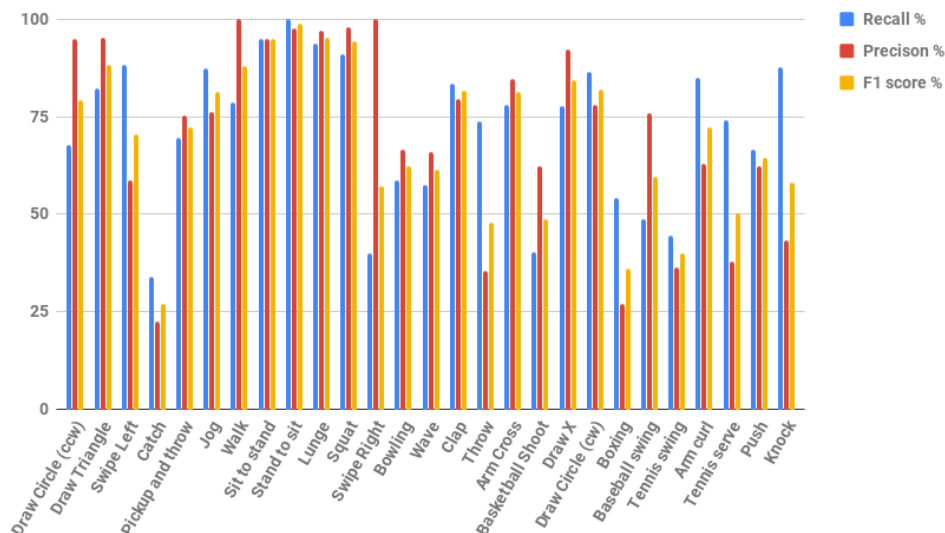


Рис. 3.14. Метрики точності для моделі BLSTM на загальному тестовому наборі даних для 27 активностей

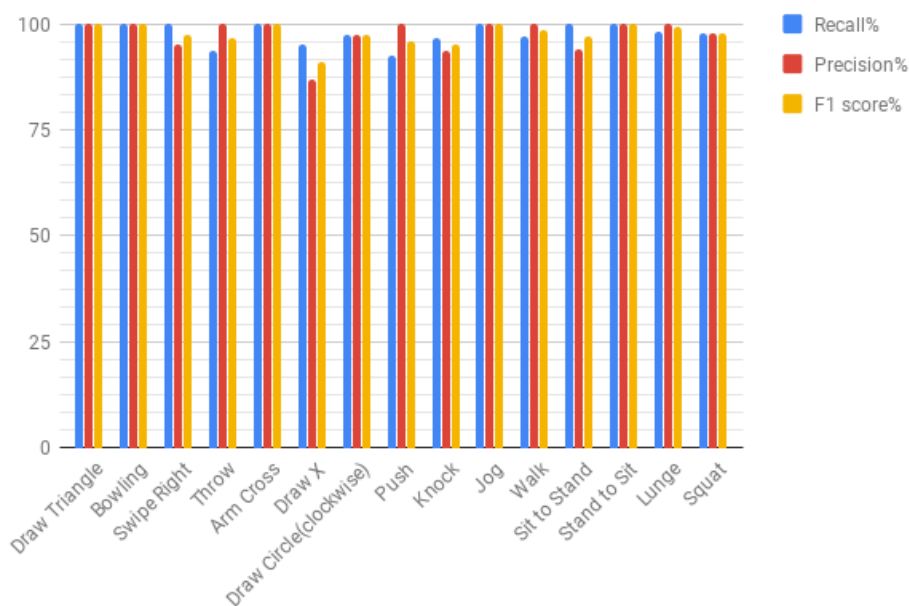


Рис. 3.15. Метрики точності для моделі BLSTM на тестовому наборі даних, специфічному для суб'єкта, для 15 активностей

3.7.3. Аналіз матриць невідповідності (Confusion Matrix)

Матриці невідповідності були використані для детального аналізу помилок класифікації та ідентифікації пар активностей, які є найбільш схильними до змішування.

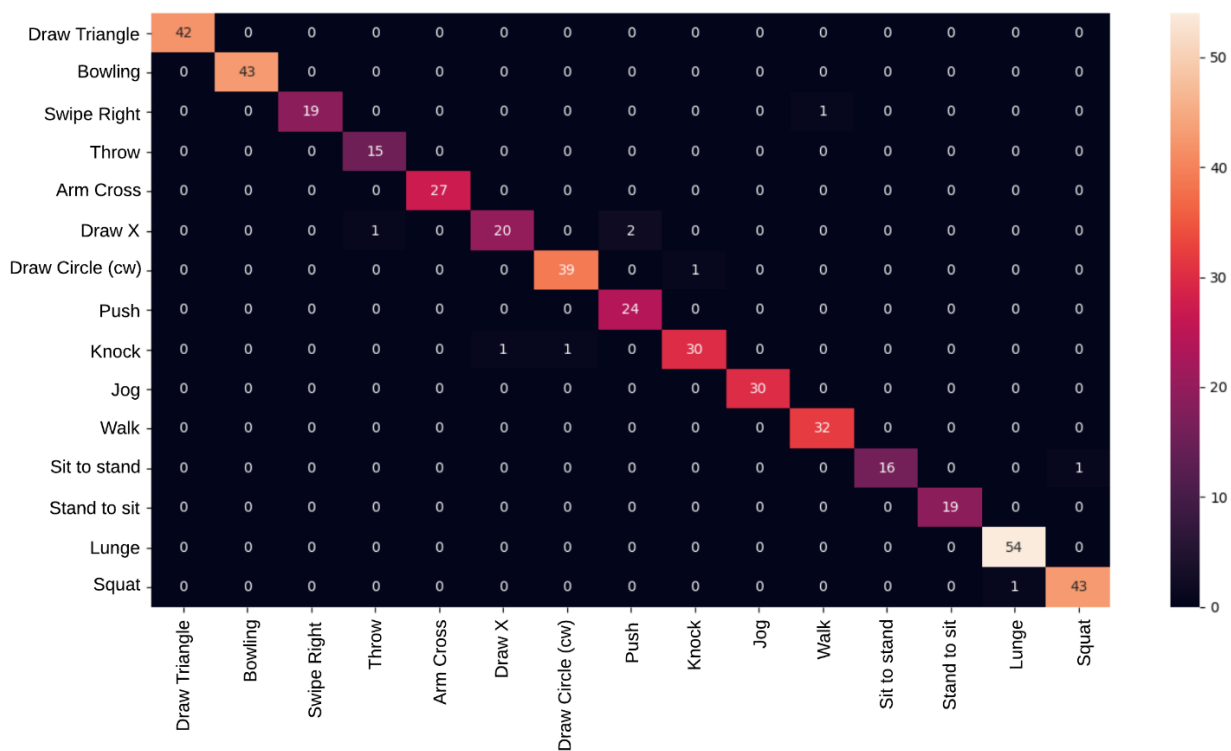


Рис. 3.18. Матриця невідповідності для 15 активностей у BLSTM, специфічному для суб'єкта

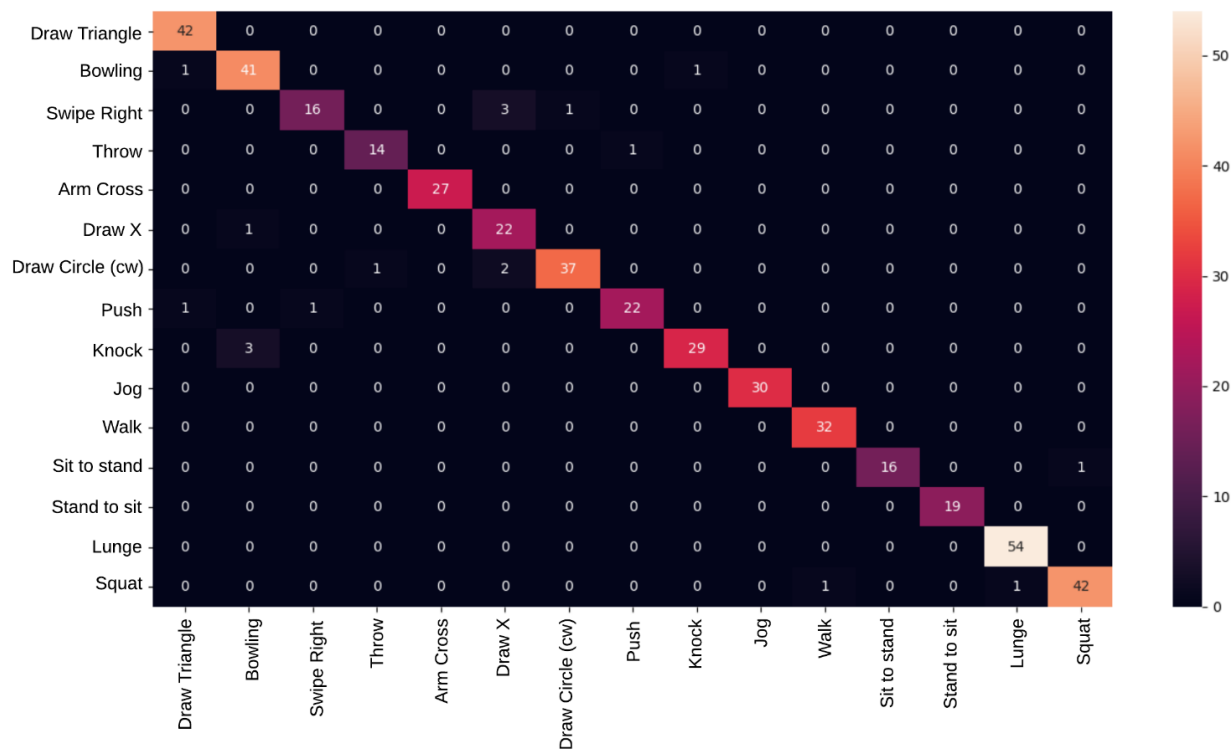


Рис. 3.19. Матриця невідповідності для 15 активностей у LSTM, специфічному для суб'єкта

Ці матриці є необхідними для візуалізації того, як BLSTM успішно диференціює близько споріднені активності (наприклад, замах бейсбольною битою проти замаху тенісною ракеткою), на відміну від LSTM, де помилки класифікації, ймовірно, зростають.

Отже, двоспрямована мережа довгої короткочасної пам'яті продемонструвала високу ефективність при безпосередній обробці сирих даних інертних сенсорів, що носяться на тілі, у задачі класифікації часових рядів.

Розроблена модель BLSTM досягла значної точності на тестових наборах даних, специфічних для суб'єкта:

- 98,05% для множини, що включає 15 активностей.

- 90,87% для складнішої множини з 27 активностей.

Навіть при роботі з великим пулом із 27 близько споріднених активностей, середня F1-оцінка для більшості класів активності становила близько 90%. Цей показник підтверджує, що модель є високопридатною для розпізнавання людської активності (HAR) та ефективно розрізняє дії зі схожими патернами руху.

Проведене дослідження чітко встановило, що BLSTM забезпечує вищу точність порівняно з односпрямованою моделлю LSTM у розпізнаванні близько споріднених активностей. Ця перевага пояснюється здатністю BLSTM інтегрувати контекстуальну інформацію як із минулого, так і з майбутнього часових послідовностей, що є критичним для диференціації тонких відмінностей у динаміці руху.

Висновки до розділу

У третьому розділі розроблено та реалізовано практичну модель двоспрямованої рекурентної нейромережі типу BLSTM для задачі розпізнавання активності людини. Представлено архітектурну схему моделі,

методологію підготовки та попередньої обробки даних, а також описано процеси навчання і тестування моделі на наборі даних UTD-MHAD.

Застосування методу ковзного вікна (sliding window) забезпечило ефективне сегментування часових рядів та покращення якості навчання мережі. Для оптимізації параметрів використано функцію втрат Categorical Cross-Entropy та оптимізаційний алгоритм Adam, що сприяло прискоренню збіжності моделі без втрати узагальнювальної здатності. У роботі також розглянуто вплив функцій активації (ReLU, tanh) та технік регуляризації (Dropout) на стабільність і точність навчання.

Отримані результати підтверджують ефективність використання двоспрямованих рекурентних мереж у задачах HAR і демонструють перспективність їх застосування в системах моніторингу здоров'я, "розумних" середовищах та контекстно-орієнтованих мобільних сервісах.

ВИСНОВКИ

У магістерській роботі розв'язано науково-прикладну задачу підвищення точності та надійності систем розпізнавання активності людини на основі релевантних сенсорних даних шляхом розроблення, дослідження та впровадження моделей і методів рекурентних нейронних мереж типу LSTM та BLSTM.

Проведене дослідження дозволило комплексно проаналізувати сучасний стан і тенденції розвитку методів розпізнавання людської активності (Human Activity Recognition, HAR), визначити їхні обмеження та виявити переваги підходів глибокого навчання, здатних ефективно працювати з часовими послідовностями сенсорних даних. На основі аналізу сучасних наукових праць доведено, що рекурентні нейронні мережі, особливо архітектури з довгою короткочасною пам'яттю (LSTM), забезпечують істотно вищу продуктивність порівняно з класичними алгоритмами машинного навчання (SVM, Decision Trees, kNN), оскільки здатні моделювати нелінійні залежності та контекстні зв'язки у часових рядах.

Проведено систематизацію принципів функціонування рекурентних нейронних мереж, визначено особливості архітектур LSTM та BLSTM у контексті задач HAR. Проаналізовано наявні підходи до збирання, обробки та нормалізації сенсорних даних із різних модальностей (акселерометр, гіроскоп, RGB, глибина), що є основою для формування тренувальних вибірок;

Визначено обмеження традиційних методів класифікації людської активності та обґрунтовано доцільність застосування глибоких рекурентних моделей для покращення результатів розпізнавання. Досліджено ефективність архітектур LSTM та BLSTM при роботі з даними набору UTD-MHAD, що містить мультисенсорні спостереження різних типів активностей.

Розроблено архітектуру BLSTM-моделі, оптимізовану для класифікації людської активності, з урахуванням вибору функцій активації, технік регуляризації та алгоритмів оптимізації (зокрема Adam).

Практичне значення отриманих результатів полягає у можливості впровадження запропонованих моделей у системи моніторингу фізичної активності, “розумні” середовища (smart homes), медичні та реабілітаційні IoT-рішення, а також у мобільні контекстно-залежні додатки. Запропонована модель може бути використана як базовий компонент інтелектуальних систем аналізу поведінки користувачів у режимі реального часу.

Таким чином, результати магістерської роботи підтверджують ефективність застосування рекурентних нейронних мереж типу LSTM та BLSTM у задачах розпізнавання активності людини. Запропоновані моделі та методи забезпечують підвищення точності класифікації, стабільності навчання та адаптивності до різних типів сенсорних даних, що робить їх перспективними для подальшого використання в інтелектуальних системах аналізу поведінки та моніторингу діяльності людини.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. A Recurrent Neural Network Music Generation Tutorial - <https://magenta.withgoogle.com/2016/06/10/recurrent-neural-network-generation-tutorial>
2. Deep Learning for Human Activity Recognition: A Resource Efficient Implementation on Low-Power Devices / Daniele Rav`1, Charence Wong, Benny Lo and Guang-Zhong Yang. - <https://files.core.ac.uk/download/pdf/77013774.pdf>
3. DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments using Deep Learning / Nicholas D. Lane, Petko Georgiev, Lorena Qendro. - https://userpages.umbc.edu/~nroy/courses/shhasp18/papers/DeepEar_Ubicomp2015.pdf
4. Human Activity Recognition Using Sequences of Postures / Vili Kellokumpu. - <https://www.mva-org.jp/Proceedings/CommemorativeDVD/2005/papers/2005570.pdf>
5. Human Activity Recognition and Pattern Discovery / Eunju Kim, Sumi Helal, and Diane Cook. - <https://scispace.com/pdf/human-activity-recognition-and-pattern-discovery-12e07ae22p.pdf>
6. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine / Davide Anguita¹, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz - <https://files.core.ac.uk/download/pdf/81581215.pdf>
7. Mining Intelligence and Knowledge Exploration: Second International ... - Google книги - https://books.google.com.ua/books?hl=uk&lr=&id=y8ujB=fnd&pg=PA99&ots=AazBPUQg-c&sig=wuo5sSTwqQ6_tUG-
8. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables / Nils Y. Hammerla, Shane Halloran, Thomas Pf`otz. - <https://arxiv.org/pdf/1604.08880>

9. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*, 1997, Vol. 9, No. 8, pp. 1735-1780.
10. Graves A., Liwicki M., Fernandez S., Bertolami R., Bunke H., Schmidhuber J. A novel connectionist system for improved unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, Vol.31, No.5.
11. Cho K., van Merriënboer B., Gülçehre Ç., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078, 2014.
12. Anguita D., Ghio A., Oneto L., Parra X., Reyes-Ortiz J. A public domain dataset for human activity recognition using smartphones. *ESANN*, 2013.
13. Kwapisz J.R., Weiss G.M., Moore S.A. Activity recognition using cell phone accelerometers. *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (KDD-10)*, 2010.
14. Lara O.D., Chambers J.C. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 2013.
15. Hammerla N.Y., Halloran S., Preece S.J. Deep, convolutional, and recurrent models for human activity recognition using wearables. *IJCAI 2016 Workshop on “Big Data from Sensors to Knowledge”*, 2016.
16. Guan Y., Ploetz T. Ensemble of deep LSTM learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017.
17. Ordóñez F.J., Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 2016, Vol.16, No.1.
18. Yao S., Wang Z., He J., Zhang X. Modeling sensor data using deep residual networks for human activity recognition. *Proceedings of the 2017 IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2017.
19. Hammerla N.Y., Halloran S., Preece S. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. *IJCAI*, 2016.

- 20.Zhang M., Sawchuk A.A. Human daily activity recognition with sparse representation using wearable sensors. Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012.
- 21.Ravi N., Dandekar N., Mysore P., Littman M.L. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), 2016.
- 22.Wang J., Chen Y., Hu L., Peng X., Zhou X. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 2019, Vol.119, pp.3-11.
- 23.Atzori M., Giacomini V., Basile F. A new dataset of labeled activity data acquired on wearable sensors. Sensors, 2016, Vol.16, No.7.
- 24.Chen Y., Zhang L., Tang L. A multi-sensor fusion approach for human activity recognition via CNN-LSTM neural networks. Sensors, 2018.
- 25.Zhu J., Gao Y., Shi Q. Multi-modal human activity recognition via deep learning and sensor fusion. IEEE Access, 2018.
- 26.Hammerla N.Y. et al. On-device activity recognition using deep recurrent neural networks. Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2017.
- 27.Banos O., Galvez J.M., Díaz R., Poyade M., Ruiz J. Window size impact in human activity recognition for wearable devices. Sensors, 2014, Vol.14, No.4, pp.6474-6493.
- 28.Reyes-Ortiz J.L., Oneto L., Ghio A., Anguita D. Transition Aware Human Activity Recognition using smartphones. Neurocomputing, 2016, Vol.171, pp.754-767.
- 29.Gjoreski M., Gjoreski H., Lajner K., Gams M. Deep and Transfer Learning for Activity Recognition in Evolving Environments. Sensors, 2018, Vol.18, No.10.
- 30.Wang Z., Xia T., Ma G., Liu G. A survey of wearable sensor based human activity recognition. IEEE Communications Surveys & Tutorials, 2020.

31. Yang J.T., Wu C.H., Yang B., Wu H.J. Human activity recognition via wearable sensing: A survey. *IEEE Internet of Things Journal*, 2019, Vol.6, No.5.
32. Bacciu D., Boldrini P., Lepri B., Stella L. Modeling human activity with recurrent neural networks. *ACM SIGKDD Workshop on Machine Learning for Human Activity Recognition*, 2015.
33. Tran D., Bourdev L., Fergus R., Torresani L., Paluri M. Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.
34. Molchanov P., Tyree S., Krasin I., Huang J., Kautz J. Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
35. Yao Z., Ren Y., Zhu J., Yang R., Liu J. Deep recurrent neural networks for human activity recognition using wearable sensors. *Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2018.