

# **Тренування моделі нейронної мережі для розпізнавання різних видів фізичної активності**


## **Постановка задачі:**

- 1) Зібрати датасет, який містить дані 5-ти різних видів фізичної активності (стояння, ходіння, присідання і т.д.);
- 2) Натренувати на простій архітектурі класифікацію фізичної активності.

## **Критерії валідації:**

- 1) Тренувальна точність моделі нейронної мережі 80% і більше;
- 2) Точність тестування моделі нейронної мережі 70% і більше.

## Етап 1: Data Collection

Дані було зібрано за допомогою програми HyperIMU.   
Налаштування параметрів запису наведено на рис. 1.

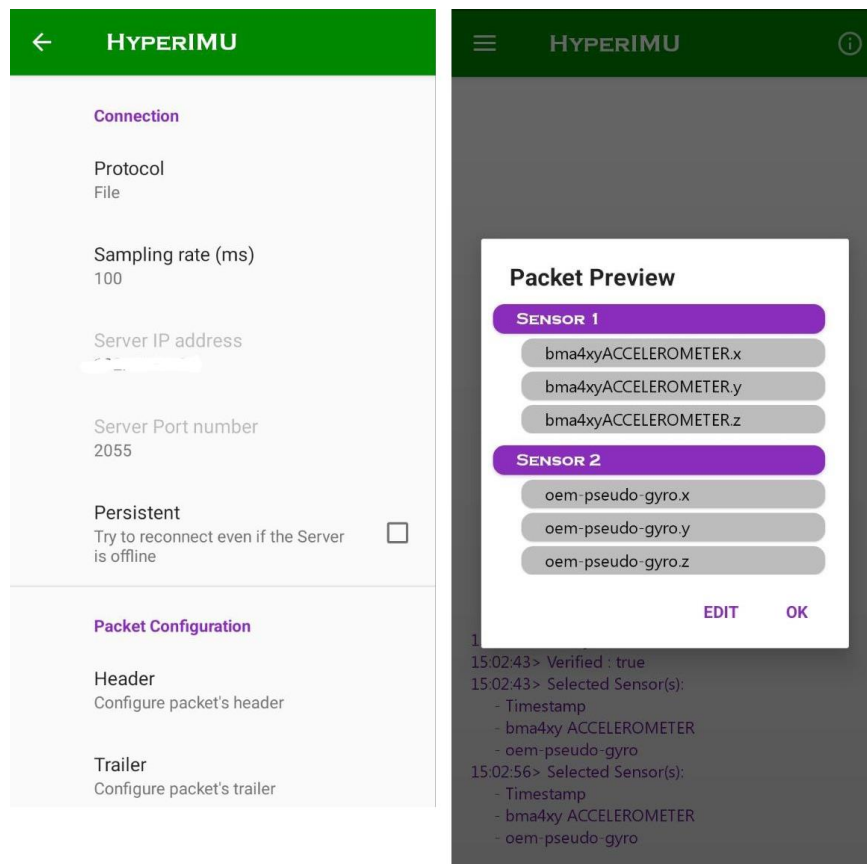


Рис. 1. Параметри запису даних акселерометра та гіроскопа

Як видно із рис. 1, записи робилися кожні 100 мс, тобто частота записів становить  $1/0.1 = 10$  Гц.

Дані записувалися для 5-ти різних активностей:

- 1) Присідання;
- 2) Ходьба;
- 3) Легкий біг (jogging);
- 4) Випади на кожную ногу (рис. 2);
- 5) Lateral Squat Slide (Переміщення центру маси з однієї ноги на іншу) (рис. 3).

Під час запису даних смартфон знаходився у правій руці у горизонтальному положенні (вісь OZ акселерометра смартфона спрямована перпендикулярно до горизонталі).

Для кожного виду активності записувався окремий датасет/-и тривалістю приблизно 5 хв. Також для тестування було записано окремий датасет, який містив дані всіх видів активності, оскільки вони виконувалися відразу одна після одної із невеликими перервами між вправами.

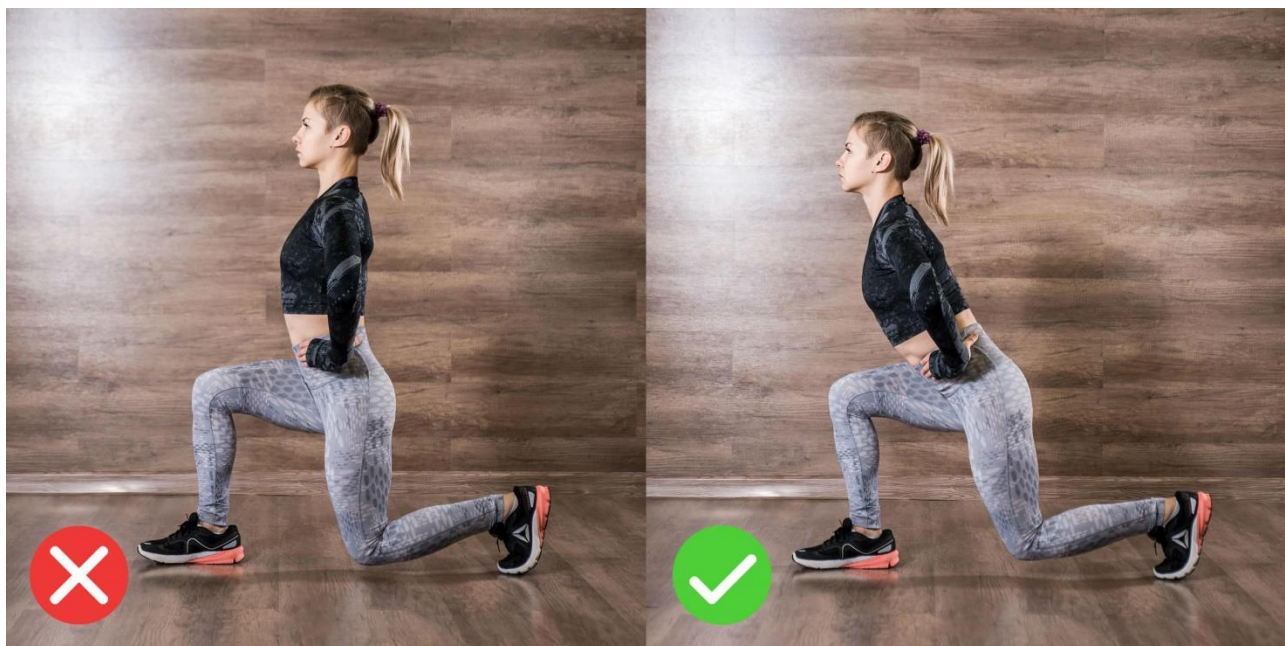


Рис. 2. Випади на кожну ногу

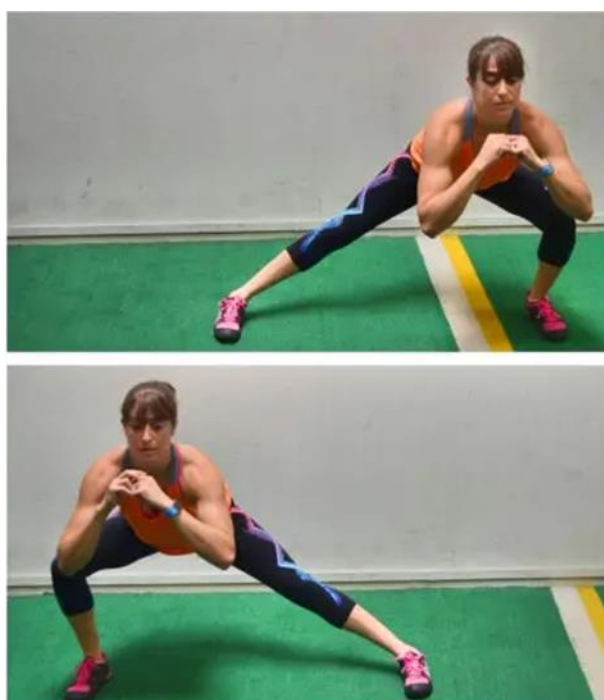


Рис. 3. Lateral squat slide

Розглянемо записані датасети.

```
columns = ['accX', 'accY', 'accZ', 'gyrX', 'gyrY', 'gyrZ']
squats_df = pd.read_csv('data/10_Hz/original_data/train/Squats_HIMU-2023-08-15_09-40-12.csv', names=columns, skiprows=4)
squats_df.head()
```

	accX	accY	accZ	gyrX	gyrY	gyrZ
0	-0.40800	-0.14595	9.844050	4.987537	-6.732138	5.511550
1	-0.36495	-0.12195	9.811050	3.540487	-2.135788	4.027513
2	-0.42705	-0.15795	10.014001	2.065800	-0.555775	1.789150
3	-0.40800	-0.12195	9.655050	0.998250	-0.017188	0.431750
4	-0.43005	-0.21000	10.048051	0.362725	0.018975	0.395037

Рис. 4. Вміст датасету із даними присідань

```
squats_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3072 entries, 0 to 3071
Data columns (total 6 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   accX    3072 non-null   float64
 1   accY    3072 non-null   float64
 2   accZ    3072 non-null   float64
 3   gyrX    3072 non-null   float64
 4   gyrY    3072 non-null   float64
 5   gyrZ    3072 non-null   float64
dtypes: float64(6)
memory usage: 144.1 KB
```

Рис. 5. Інформація про датасет із даними присідань

Отже, записані датасети містять шість стовпців: перші три відповідають даним трьох осей (OX, OY, OZ) акселерометра і виміряні у  $\text{м/с}^2$ , наступних три стовпці – три осі гіроскопа (рад/с).

Розглянемо також поведінку, наприклад, акселерометра під час запису даних присідань.

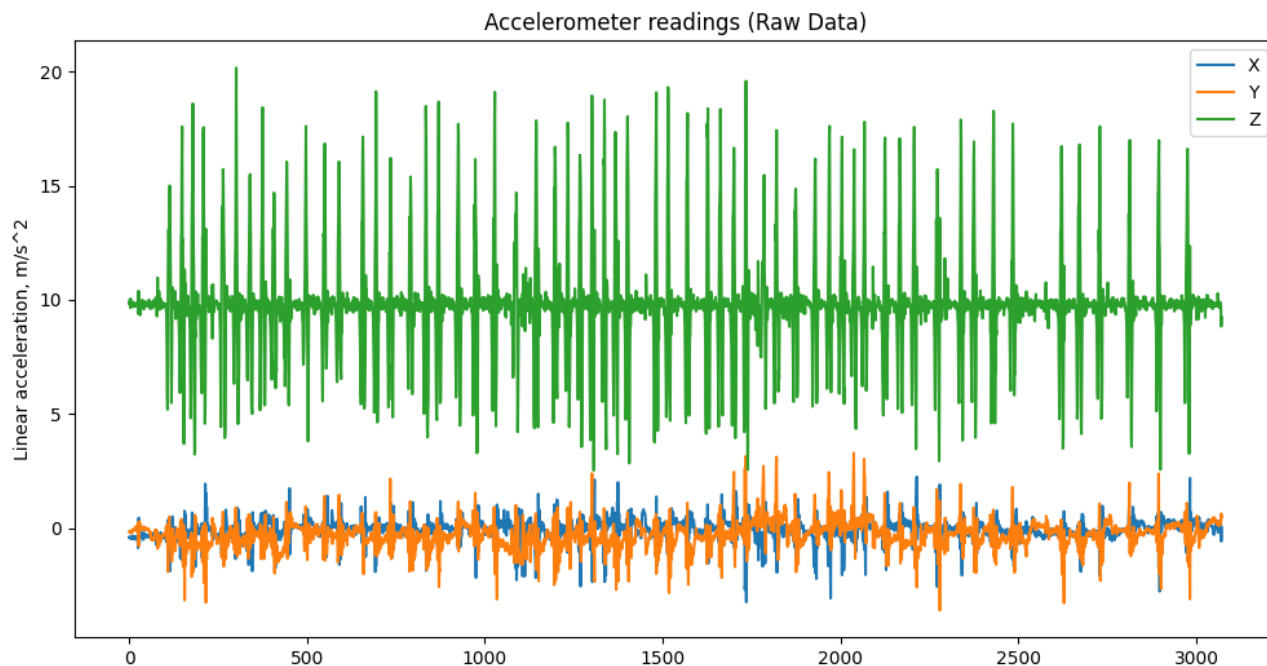


Рис. 6. Результати вимірювання акселерометра (присідання)

Розглянувши рис. 6, можна побачити, що на початку та в кінці запису відбувається стабілізація датчиків. Окрім того, між самими присіданнями теж зустрічаються моменти стабілізації датчиків (їх особливо добре видно на осі OZ). Такі моменти відповідають відпочинку між присіданнями. Оскільки дані початкової та кінцевої стабілізації пристрою, а також перерви між присіданнями не можна промаркувати як "присідання", а датасет невеликий, всього 3072 елементи, отримаємо недостатню кількість даних для тренування моделі нейронної мережі.

```
leg_land_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3635 entries, 0 to 3634
Data columns (total 6 columns):
#   Column  Non-Null Count  Dtype  
---  -
0   accX    3635 non-null   float64
1   accY    3635 non-null   float64
2   accZ    3635 non-null   float64
3   gyrX    3635 non-null   float64
4   gyrY    3635 non-null   float64
5   gyrZ    3635 non-null   float64
dtypes: float64(6)
memory usage: 170.5 KB
```

Рис. 7. Інформація про датасет із даними випадів на ноги

```

walk_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3508 entries, 0 to 3507
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0    accX        3508 non-null   float64
1    accY        3508 non-null   float64
2    accZ        3508 non-null   float64
3    gyrX        3508 non-null   float64
4    gyrY        3508 non-null   float64
5    gyrZ        3508 non-null   float64
dtypes: float64(6)
memory usage: 164.6 KB

```

Рис. 8. Інформація про датасет із даними ходьби

```

lateral_squat_slide_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2860 entries, 0 to 2859
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0    accX        2860 non-null   float64
1    accY        2860 non-null   float64
2    accZ        2860 non-null   float64
3    gyrX        2860 non-null   float64
4    gyrY        2860 non-null   float64
5    gyrZ        2860 non-null   float64
dtypes: float64(6)
memory usage: 134.2 KB

```

Рис. 9. Інформація про датасет із даними вправи lateral squat slide

```

jogging_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3154 entries, 0 to 3153
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0    accX        3154 non-null   float64
1    accY        3154 non-null   float64
2    accZ        3154 non-null   float64
3    gyrX        3154 non-null   float64
4    gyrY        3154 non-null   float64
5    gyrZ        3154 non-null   float64
dtypes: float64(6)
memory usage: 148.0 KB

```

Рис. 10. Інформація про датасет із даними легкого бігу (jogging)

```
test_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2857 entries, 0 to 2856
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0    accX        2857 non-null   float64
1    accY        2857 non-null   float64
2    accZ        2857 non-null   float64
3    gyrX        2857 non-null   float64
4    gyrY        2857 non-null   float64
5    gyrZ        2857 non-null   float64
dtypes: float64(6)
memory usage: 134.0 KB
```

Рис. 11. Інформація про тестовий датасет

Розглянувши інформацію про інші датасети, можна помітити, що всі вони є відносно невеликими (2800-3500 елементів). Як уже згадувалося раніше, цих обсягів може бути недостатньо для тренування моделі нейронної мережі. Окрім того, існує ще один недолік: оскільки у датасетах відсутній стовпець, що відповідає за час вимірювання (наприклад, 'time' або 'timestamp'), неможливо оцінити стабільність частоти забору даних.

Враховуючи усі вищезгадані недоліки, було прийнято рішення перезаписати датасети.