Human Activity Recognition

This report analyzes the PAMAP2 Physical Activity Monitoring dataset for Human Activity Recognition (HAR). It includes an exploratory analysis of sensor data, a literature review of existing approaches, a justified choice of model, experimental evaluation, and discussion of results. The report concludes with future research directions in wearable-sensor HAR.

Code is available at: https://github.com/martasumyk/activity_classification_project.

Introduction

Many modern smart devices continuously collect sensor data, yet the user is still often required to manually specify the type of activity being performed (e.g. cycling, running, walking and more). This highlights the need for reliable Human Activity Recognition (HAR) systems that can automatically infer activity categories directly from sensor signals.

HAR from wearable devices plays an increasingly important role in healthcare monitoring, rehabilitation, assistive technologies, sports analytics, and smart home environments. The PAMAP2 Physical Activity Monitoring dataset provides a comprehensive collection of multi-sensor recordings from participants performing a wide range of everyday and sports-related activities. Its richness makes it a valuable benchmark for developing and evaluating HAR algorithms.

The aim of this project is to analyze the PAMAP2 dataset, conduct exploratory data analysis, review existing research approaches, and design an effective machine-learning model for automatic activity classification using wearable sensor data.

Methods

Data

Dataset used:
https://archive.ics.uci.edu/dataset/231/pamap2+physical+activity+monitoring.

The PAMAP2 Physical Activity Monitoring dataset contains recordings from **9 subjects** performing **18 different activities**, including:

- resting (sitting, standing, lying)

- walking, running
- cycling
- household tasks such as ironing and vacuum cleaning
- rope jumping and Nordic walking

Sensors used

Each subject wore:

- **Three IMUs** (hand, chest, ankle)

  – accelerometer

  – gyroscope

  – magnetometer

  – temperature

- **Heart-rate monitor**

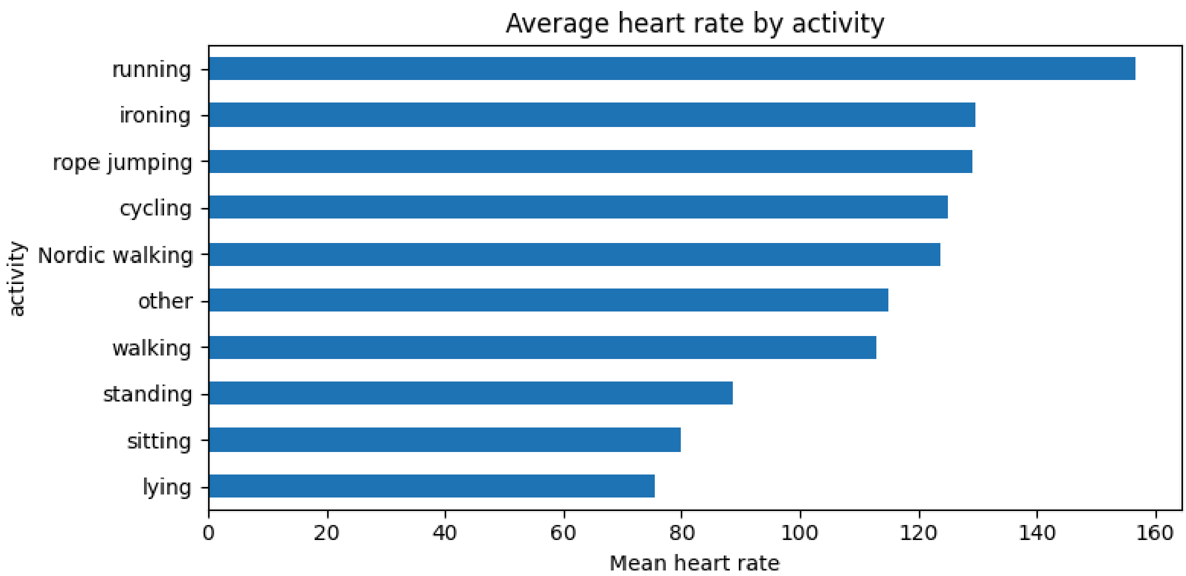
Sampling

- IMUs: 100 Hz
- Heart rate: ~9 Hz

Each sample contains 52 features.

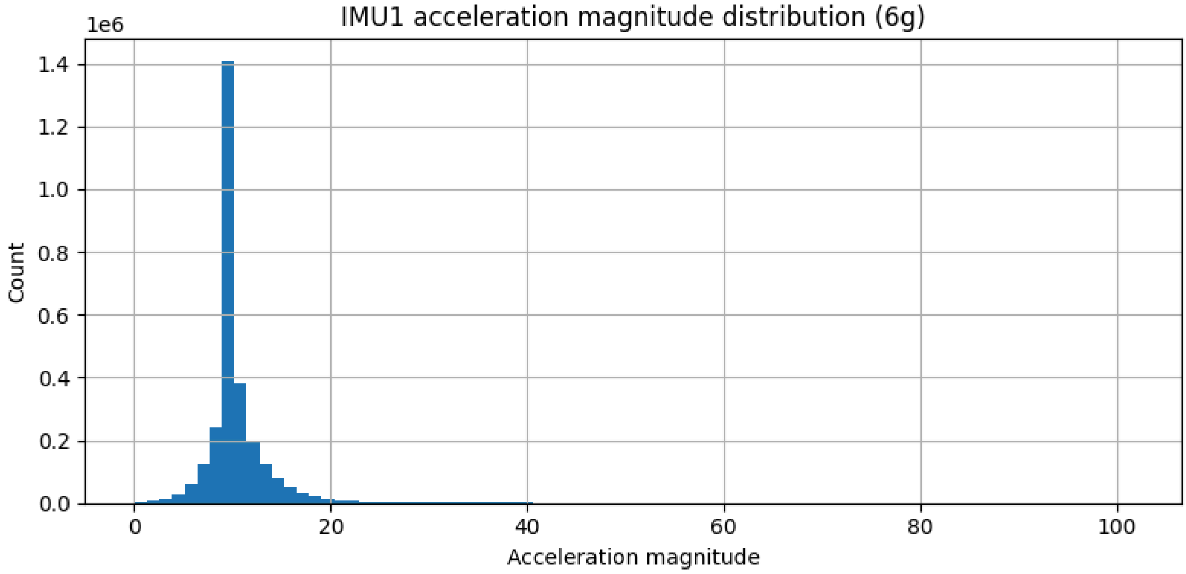
EDA

Average heart rate by activity:



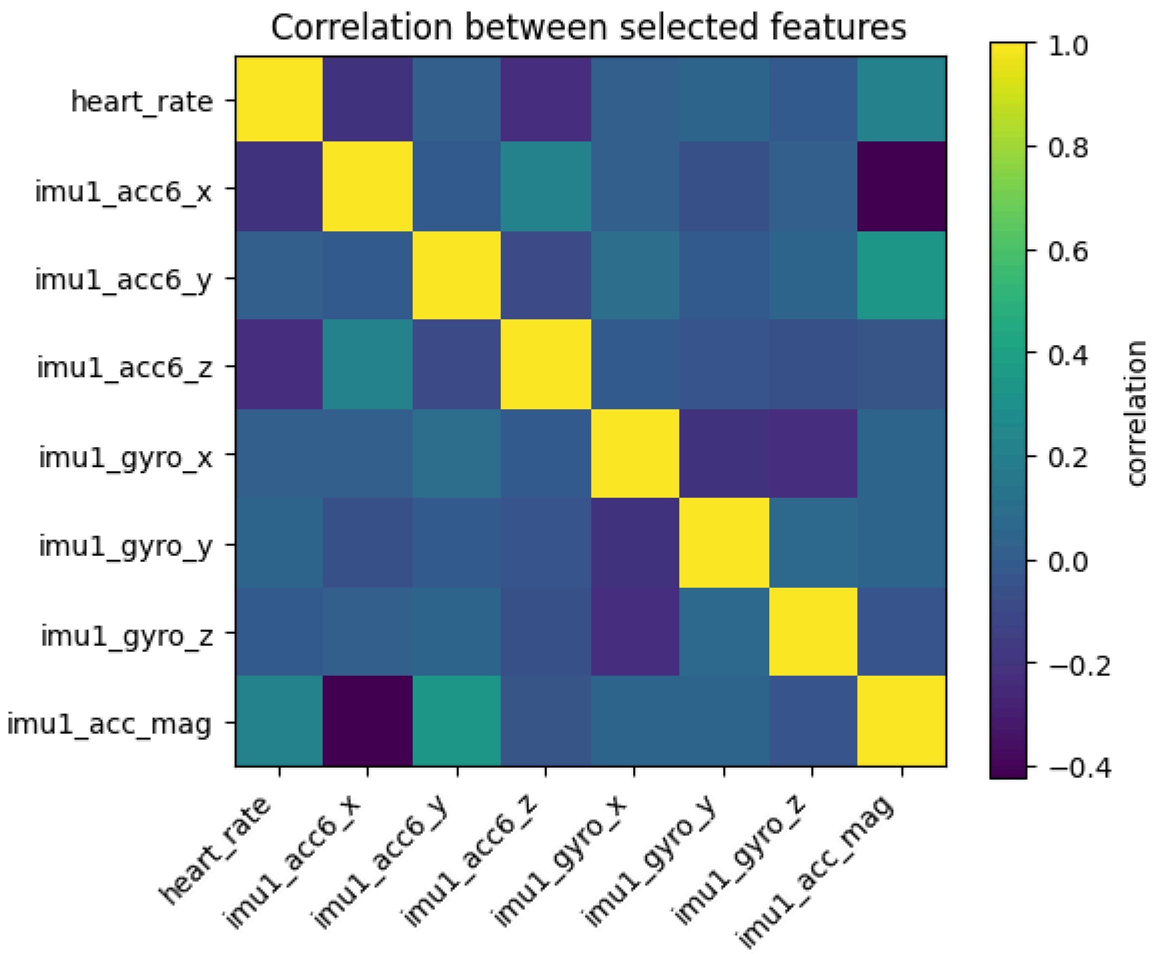
It is pretty obvious that for lying the heart rate is the smallest, while for running the biggest.

Acceleration magnitude:



The histogram shows the distribution of acceleration magnitude measured by the IMU placed on the hand (IMU1). Most values fall between 5–20 m/s^2, forming a sharp peak that corresponds to normal daily movements such as sitting, standing, or walking. A long but sparse tail extends toward higher magnitudes, reflecting short periods of more intense activities like running, stair climbing, or rope jumping.

Correlation between features:



A correlation heatmap of IMU1 features shows only weak to moderate correlations among accelerometer and gyroscope axes. Also, we can see that heart rate shows almost no correlation with inertial sensors.

Task that can be done with this data

Using PAMAP2 sensor data, several tasks can be solved. The most popular is Human Activity Recognition (HAR) - to classify activities from wearable sensors.

Another is Health and rehabilitation monitoring: tracking gait, movement patterns, and recovery progress. Also, real-time activity segmentation to detect transitions between activities.

And lastly, anomaly detection: to identify irregular movement patterns (useful for elderly fall detection). My project focuses on supervised activity classification.

Literature review

HAR has been extensively studied in wearable-sensor and mobile computing research.

Classical ML approaches

Traditional approaches to Human Activity Recognition typically rely on handcrafted feature engineering followed by classical machine-learning classifiers. Common models used in early HAR research include Decision Trees and Random Forests, which handle nonlinear relationships and provide interpretable decision boundaries; Support Vector Machines (SVMs), which perform well on high-dimensional feature spaces; k-Nearest Neighbors (k-NN), a simple yet effective distance-based method; and Hidden Markov Models (HMMs). These methods often achieve strong baseline performance as can be seen in related works.

The advantages of these methods are obvious: they are pretty interpretable and fast. The limitation is that they require handcrafted features, however we already have them in the dataset.

Deep learning approaches

More recent HAR systems increasingly rely on deep learning architectures such as CNNs, LSTMs, and Transformers (Ordóñez & Roggen (2016), “Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition”). Convolutional Neural Networks (CNNs) are effective at automatically extracting local spatial, temporal patterns from multichannel IMU signals, reducing the need for manual feature engineering. Recurrent models like LSTMs and GRUs capture longer-term temporal dependencies, making them well suited for activities with sequential structure. Hybrid CNN–LSTM architectures combine the strengths of both approaches by first learning local motion features and then modeling their evolution over time. Most recently, Transformer-based models have shown strong performance in HAR tasks due to their ability to model long-range relationships and capture complex dependencies across entire activity sequences. However, these methods are less interpretable yet.

Chosen approach

For this project, I chose a **Random Forest classifier** with **window-based feature extraction**. This choice is justified by the dataset size, nonlinear relationships in it and also sliding windows reflect natural temporal structure of activities.

Feature extraction included:

- mean and standard deviation of each signal
- window size: 5 seconds (500 samples)
- step size: 2 seconds

Experimental setup

Preprocessing the data

- Loaded 9 subject files from the Protocol set.
- Removed activity “0” (undefined).
- Forward/backward filled missing sensor values per subject.
- Normalized sampling inconsistencies.
- Generated overlapping windows of raw data.
- Extracted statistical features (mean + std).

Train-test split

- 80/20 stratified split across windows

Classifier

- Random Forest (200 trees)
- Max-depth: unlimited
- n_jobs = −1 (full CPU)

Evaluation metrics

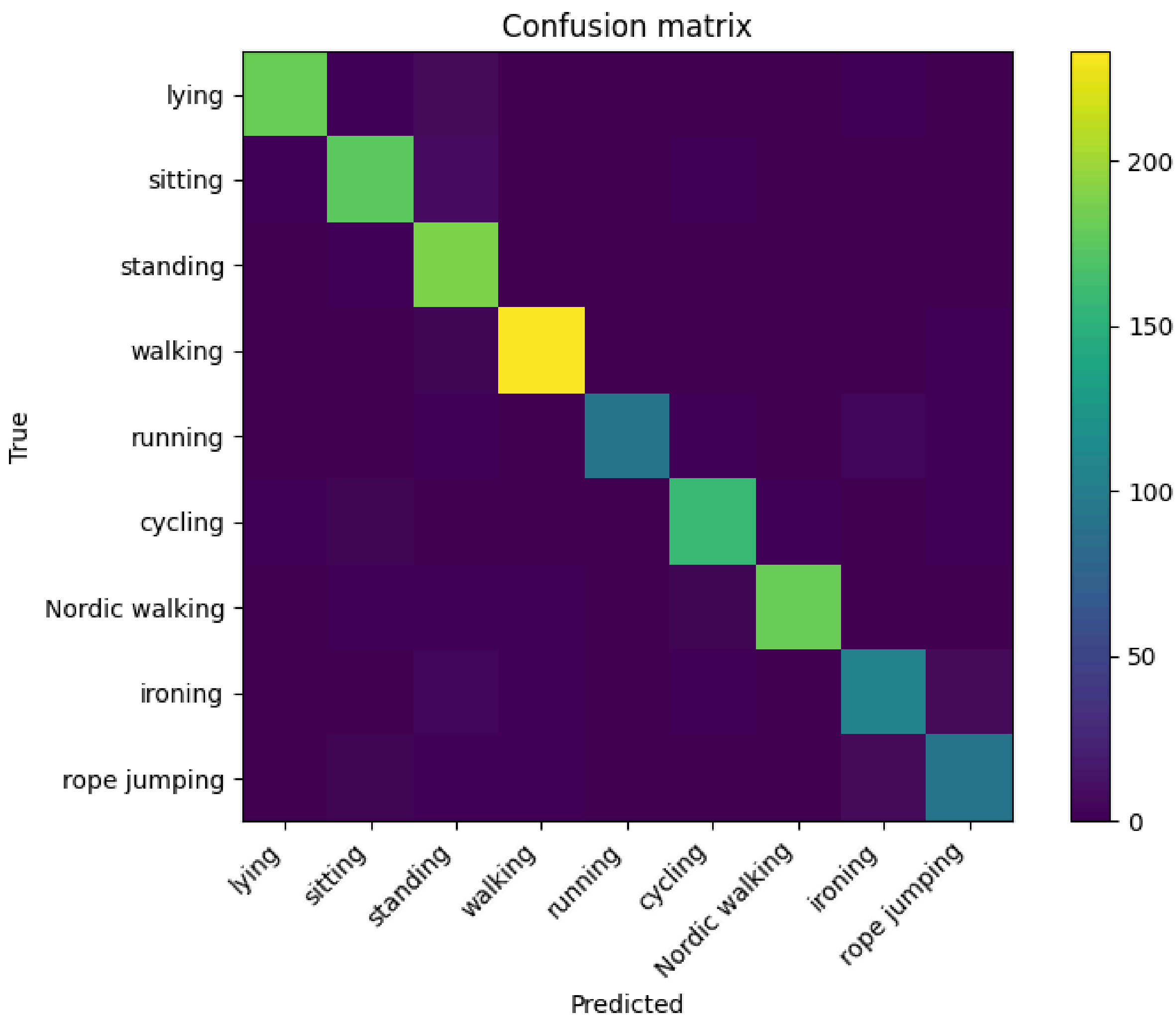
- Accuracy
- Macro F1-score
- Confusion matrix

Results

Model performance

	precision	recall	f1-score	support
lying	0.98	0.95	0.97	191
sitting	0.94	0.94	0.94	186
standing	0.87	0.99	0.93	190
walking	0.98	0.98	0.98	238
running	1.00	0.92	0.96	97
cycling	0.96	0.96	0.96	164
Nordic walking	0.99	0.96	0.98	189
ironing	0.90	0.89	0.89	117
rope jumping	0.89	0.87	0.88	105
accuracy			0.95	1477
macro avg	0.95	0.94	0.94	1477
weighted avg	0.95	0.95	0.95	1477

We have high F1-score on distinct activities (lying, sitting, standing, running) - all 88%+



We have moderate confusion between activities with similar dynamics, for example:

- walking and Nordic walking
- ascending and descending stairs
- ironing and vacuum cleaning

These confusions reflect the natural similarity of the motions.

Interpretation

- IMU acceleration features dominate predictive power.
- Heart rate provides additional separation for intensity-based activities.
- Window-based mean + std features capture most relevant patterns but may miss timing-specific variations, this is where CNN/LSTM models could improve performance.

Future directions

Several extensions could significantly improve the performance and scope of this work. First, training deep learning models such as CNNs, LSTMs, or Transformers would allow the system to capture richer temporal dynamics present in IMU signals. A leave-one-subject-out evaluation could be introduced to assess how well the model generalizes to unseen individuals, which is crucial for real-world applications. Further improvements could be achieved through enhanced feature engineering, including frequency-domain descriptors, wavelet transforms, or jerk-based features. Another valuable extension would be real-time activity segmentation (for example, with some fitness-specific devices), developing models that detect transitions and operate in streaming settings.

Conclusion

This project demonstrates a complete HAR pipeline based on the PAMAP2 dataset, including EDA, literature review, model justification, experimental design, and evaluation of random forest model for this task. This classifier with window-based features provides strong and interpretable performance, highlighting the potential of wearable-sensor HAR.

References

1. Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). *Activity recognition using cell phone accelerometers.*

2. Lara, O. D., & Labrador, M. A. (2013). *A survey on human activity recognition using wearable sensors.*

3. Ordóñez, F. J., & Roggen, D. (2016). *Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition.*

4. Hammerla, N. et al. (2016). *Deep, convolutional, and recurrent models for human activity recognition using wearables.*

5. Tang, et al. (2020). *Deep learning for sensor-based activity recognition: A survey.*
6. *https://medium.com/@pacosun/one-out-all-in-leave-one-out-cross-validation-explained-409df5ff6385*