Human Activity Recognition with HARTH Dataset

Shervan Shahparnia & Sharon Le

Introduction -

Human Activity Recognition (HAR) is a rapidly growing area in machine learning with important applications in health monitoring, fitness tracking, and smart environments. The HARTH dataset allows us to explore how wearable sensors can be used to recognize various human activities in a natural, free-living setting. For this project, our goal was to preprocess and analyze the HARTH dataset, extract features from sensor data, and train machine learning models to predict activity labels with high accuracy.

Dataset Overview -

The HARTH dataset contains professionally annotated activity data from 22 participants, each wearing two Axivity AX3 accelerometers for about 2 hours. The sensors were placed on the lower back and right front thigh and recorded acceleration data at a 50 Hz sampling rate.

Each subject's data is stored in a .csv file, with the following columns:

- timestamp: sample time
- back_x, back_y, back_z: 3D acceleration from the back sensor
- thigh_x, thigh_y, thigh_z: 3D acceleration from the thigh sensor
- label: the activity code

Activity labels include:

Walking, running, shuffling, ascending/descending stairs, sitting, standing, lying down, and several variations of cycling. Some activities, such as cycling while standing (inactive), are relatively rare, which introduces some class imbalance.

This dataset was funded by NTNU Helse and introduced in the paper *HARTH: A Human Activity Recognition Dataset for Machine Learning* by Logacjov et al. (2021).

Data Preprocessing and Segmentation -

To convert the continuous sensor data into individual training samples, we used a sliding window approach:

- Window size: 100 samples (2 seconds)
- Step size: 50 samples (50% overlap)

Each window was labeled based on the most frequent activity label in that segment. This method preserved temporal information while reducing the complexity of full time-series modeling.

Feature Extraction -

1. Time-Domain Features

For each window, we extracted:

- Mean, Median, Standard Deviation
- Minimum, Maximum
- Root Mean Square (RMS)
- Signal Magnitude Area (SMA)
- Entropy

2. Frequency-Domain Features

- Applied FFT (Fast Fourier Transform)
- Extracted energy and dominant frequency components

Model Training -

To evaluate the effectiveness of our extracted features, we trained and tested four machine learning models: Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), and Support Vector Machine (SVM). These models were selected for their proven performance in classification tasks and their compatibility with structured feature data. Each model was trained on the segmented and feature-rich dataset produced from our preprocessing and sliding window pipeline.

We used two cross-validation strategies to assess model performance. First, we applied 10-Fold Cross-Validation, where the data is split into ten parts, and each part is used once as the test set while the others are used for training. This method provides a strong indication of overall model reliability and is commonly used to prevent overfitting. Second, we implemented

Leave-One-Subject-Out (LOSO) validation, where each model is trained on the data from 21 subjects and tested on the remaining one. This simulates real-world scenarios where models must generalize to new, unseen users.

Model performance was evaluated using standard classification metrics: Accuracy, F1 Score, Precision, and Recall. These metrics give a balanced view of model effectiveness, accounting for both overall correctness and the ability to correctly identify each activity class. Through these evaluation strategies, we were able to compare and identify the most robust models for human activity recognition based on our dataset.

Model Results -

Model	Evaluation	Accuracy	F1 Score	Precision	Recall
Random Forest	10-Fold CV	0.9913	0.9911	0.9915	0.9914
Decision Tree	10-Fold CV	0.987	0.987	0.987	0.987
K-Nearest Neighbors	10-Fold CV	0.9646	0.9646	0.9647	0.9646
Support Vector Machine	10-Fold CV	0.8663	0.827	0.8016	0.8663
Random Forest	LOSO	0.8229	0.8286	0.8407	0.8229
K-Nearest Neighbors	LOSO	0.8174	0.8486	0.9375	0.8174

Decision Tree	LOSO	0.8132	0.8227	0.8884	0.8132
Support Vector Machine	LOSO	0.7453	0.7613	0.8406	0.7453

Key Insight**:**

Random Forest consistently outperformed all other models. While the 10-Fold CV results were extremely high, the LOSO accuracy dropped, showing that cross-user generalization is more difficult and needs further tuning or data.

Challenges and Insights -

Empty feature matrix issue: During demo2, we encountered an issue where windowed feature extraction returned empty matrices. We fixed this by debugging window lengths and ensuring proper overlap.

SVM performance: SVM underperformed, possibly due to difficulty separating overlapping classes or lack of feature scaling.

Class imbalance: Some rare activities had very few samples (e.g., cycling inactive), making them hard to classify. This could be mitigated in future with oversampling or class weighting.

Conclusion -

This project successfully demonstrated how classical machine learning techniques can be applied to sensor-based time-series data for human activity recognition. Using the HARTH dataset, we built a full machine learning pipeline from data understanding and preprocessing to model training and evaluation. Our work focused on converting continuous sensor streams into meaningful features using sliding window segmentation and statistical analysis, then training models to classify activities such as walking, sitting, and cycling.

In Demo 1, we focused on understanding the structure of the HARTH dataset. We explored how the placement and orientation of sensors on the lower back and right thigh influence the data and how each activity is represented. This step helped us grasp the importance of clear annotations and sensor positioning for building accurate recognition models.

Demo 2 emphasized the transformation of raw data into usable input for modeling. We applied windowing techniques and calculated time and frequency domain features such as mean, standard deviation, entropy, and signal strength. A major takeaway from this stage was recognizing how design choices like window size and overlap can significantly affect model inputs. We also encountered and solved an issue where some windows produced no data, reinforcing the importance of debugging during preprocessing.

In Demo 3, we trained and compared several models including Random Forest, K-Nearest Neighbors, Support Vector Machines, and Decision Trees. We evaluated their performance using 10-fold cross-validation and Leave-One-Subject-Out (LOSO) testing. While Random Forest achieved over 99 percent accuracy in 10-fold CV, its performance dropped under LOSO, showing the difficulty of generalizing across different users. This taught us the value of realistic testing strategies and the challenges in building models that work reliably for new individuals.

Overall, this project helped us develop a solid understanding of the end-to-end process of building an activity recognition system using wearable sensor data. It showed the importance of thoughtful preprocessing, meaningful feature extraction, and careful model evaluation. It also highlighted the impact of subject variability and the need for robust validation methods when applying machine learning to real-world data.